EXTENSION AND GENERALIZATION OF NEWELL'S SIMPLIFIED THEORY OF KINEMATIC WAVES

A Thesis
Presented to
The Academic Faculty

by

Daiheng Ni

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Civil and Environmental Engineering

Georgia Institute of Technology
December 2004

EXTENSION AND GENERALIZATION OF NEWELL'S SIMPLIFIED THEORY OF KINEMATIC WAVES

Approved by:

Dr. John D. Leonard II, Advisor
Civil and Environmental Engineering

Dr. Michael Hunter
Civil and Environmental Engineering

Dr. Karen Dixon
Civil and Environmental Engineering

Dr. David Goldsman
Industrial and Systems Engineering

Dr. Adjo Amekudzi
Civil and Environmental Engineering

November 9, 2004

This dissertation is dedicated to:

my parents

their love and support make this possible

This dissertation is also dedicated to

and my wife and sons

for their endless love

ACKNOWLEDGEMENT

I would like to express deepest gratitude to my advisor, **Dr. John D. Leonard II**, for his full support, expert guidance, understanding and encouragement throughout my study and research. Without his incredible patience and timely wisdom and counsel, my dissertation would have been a frustrating and overwhelming pursuit.

In addition, I express my appreciation to Dr. Karen Dixon, Dr. Michael Hunter, Dr. Adjo Amekudzi, and Dr. David Goldsman for having served in my defense committee. Their thoughtful questions and comments were valued greatly.

I would also like to thank Dr. Michael D. Meyer, Dr. Peter Parsonson, Dr. Randall Guensler, and Dr. Billy M. Williams for helping me with my coursework and academic research during my Ph.D. study at School of Civil and Environmental Engineering, Georgia Institute of Technology.

Special thanks also go to my fellow graduate students at the Transportation System Group and staff of the School for their help during my study at Georgia Institute of Technology.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Flow of traffic on freeways and limited access highways can be represented as a series of kinemetic waves. Solutions to these systems of equations become problematic under congested traffic flow conditions, and under complicated (real-world) networks. A simplified theory of kinematics waves was previously proposed. Simplifying elements includes translation of the problem to moving coordinate system, adoption of bi-linear speed-density relationships, and adoption of restrictive constraints at the on- and off-ramps. However, these simplifying assumptions preclude application of this technique to most practical situations.

This research explores the limitations of the simplified theory of kinematic waves. First this research documents a relaxation of several key constraints. In the original theory, priority was given to on-ramp merging vehicles so that they can bypass any queue at the merge. This research proposes to relax this constraint using a capacity-based weighted fair queuing (CBWFQ) merge model. In the original theory, downstream queue affects upstream traffic as a whole and exiting traffic can always be able to leave as long as it gets to the diverge. This research proposes that this diverge constraint be replaced with a contribution-based weighted splitting (CBWS) diverge model. This research proposes a revised notation system, permitting the solution techniques to be extended to freeway networks with multiple freeways and their ramps. This research proposes a generalization to permit application of the revised theory to general transportation networks. A generalized CBWFQ merge model and a generalized CBWS diverge model

are formulated to deal with merging and diverging traffic. Finally, this research presents computational procedure for solving the new system of equations.

Comparisons of model predictions with field observations are conducted on GA 400 in Atlanta. Investigations into the performance of the proposed CBWFQ and CBWS models are conducted. Results are quite encouraging, quantitative measures suggest satisfactory accuracy with narrow confidence interval.

CHAPTER 1

OVERVIEW OF THE RESEARCH

This chapter gives an overview of the research and issues discussed include the motivations of the research, major contributions of the research, and the organization of the dissertation.

1.1. RESEARCH MOTIVATIONS

A decade ago, Gordon F. Newell (1993a, 1993b, and 1993c) proposed a simple solution, referred to as the simplified theory thereafter, to a special case of a traffic model (Lighthill and Whitham 1955; Richards 1956) proposed nearly half a century ago. It was recognized, and will be demonstrated in Chapter 3, that the simplified theory addresses the traffic queuing problem on a freeway mainline, while it would make more sense to view such a problem at a system-wide perspective, e.g., a regional freeway network. Except for a few research efforts (Son 1996; Hurdle and Son 2000; Leonard 1997, 1998; Banks 2000) that supported or applied Newell's work, the simplified theory has remained in its original form since its publishing. Much the same as what Newell mentioned in his historical nodes (Newell 1993a) "In retrospect, one can't help wonder why something as simple as this has gone unnoticed for more than 35 years or, if it was noticed, that is was

not exploited." We share the same feeling and are curious why a solution as simple and promising (Son 1996; Hurdle and Son 2000) as the simplified theory has remained restricted for yet another decade knowing that there is a direction to make the solution more practically appealing. This curiosity is, perhaps, the primary reason that motivates this research.

Just after Newell's original work, Daganzo (1994, 1995a, 1995b, 1999) proposed a related model, called the Cell Transmission Model (CTM), which has been widely acknowledged and essentially solves the same type of problem as that of Newell's except that the latter is able to address network traffic. In principle, a freeway network in the CTM is partitioned into a series of small segments, called cells, and the model keeps track of the contents (or vehicles) in these cells. However, a major drawback of the CTM is its strict requirement on the size of the cells which is recommended (Daganzo 1994) as the distance a vehicle traverses at the free flow speed during a time increment to achieve best accuracy (Daganzo 1995a). The implication of this requirement is that, as a freeway network gets large which is a situation that a macroscopic traffic model typically excels, modeling such a network will be cumbersome. In contrast, the simplified theory does not pose such a requirement on link length. As a result, modeling the same network with the simplified theory will consume much less computational resources provided that the simplified theory can be extended to deal with network traffic. With the assistance of properly proposed merging model and diverging model, we believe that the extension of the simplified theory is not impossible. This belief is, probably, another reason that motivates this research.

1.2. RESEARCH CONTRIBUTIONS

As mentioned above, the primary limitation of the simplified theory lies in its inability to model network traffic. This limitation is the result of Newell's restrictive assumptions on merging and diverging behavior. A reasonable starting point of extending the simplified theory is to relax these restrictive assumptions by proposing a merge model and a diverge model to allow traffic queuing at any merging and diverging branch, respectively. Then we can extend the simplified theory by incorporating the proposed merge model and diverge model to address network traffic in a freeway system where multiple freeways with their on- and off-ramps are allowed. If the proposed merge model and diverge model are able to incorporate multiple (more than 2) merging and diverging branches, respectively, we can further generalize the simplified theory to address a general transportation network where a link in the network can have multiple upstream and/or downstream branches. More specifically, the contributions of this research include the following:

(1) A model for traffic merging behavior

The simplified theory assumes that ramp entering flow can always bypass a queue, if any, at a merge and experiences no delay. Therefore, there is virtually no queuing at the on-ramp because, whenever there is any demand, it is satisfied without delay. To relax this assumption, we developed a merge model, as well as its generalized form, that allows traffic queuing at all merging branches.

(2) A model for traffic diverging behavior

When dealing with diverging traffic, the simplified theory assumes that the travel time of all vehicles in a freeway section is independent of their destinations. On the other

hand, exiting vehicles can always exit without any delay as long as they have successfully arrived at the diverge, i.e., there is virtually no queuing at the off-ramp, either. To relax this assumption, we developed a diverge model, as well as its generalized form, that allows queuing at any diverging branch and queues from the diverging branches can back onto the upstream mainline, further constraining traffic there.

(3) Extension of the simplified theory to address a freeway system

As mentioned above, the major limitation of the simplified theory lies in its inability to address network traffic. With the assistance of the above merge and diverge models, we extended the simplified theory to address a freeway system where multiple freeways with their on- and off-ramps are allowed. We formulated the extension based on 5 basic building blocks, i.e., an entrance, an exit, a mainline, a merge, and a diverge, and we assume that a freeway system can be represented by certain combination of these basic building blocks.

(4) Generalization of the simplified theory to address a general transportation network

With the assistance of the generalized forms of the above merge and diverge models, we generalized the simplified theory to address a general transportation network where a link in the network can have multiple upstream and/or downstream branches. We formulated the generalization based on a generic building block which may incorporate multiple upstream and/or downstream branches and we assume that a general transportation network can be represented by certain combination of some special cases of the generic building block.

1.3. ORGANIZATION OF THIS DISSERTATION

This dissertation can be roughly divided into four major parts: introduction, theoretical development, model validation, and conclusion. The introduction includes this chapter and literature review (Chapter 2). The theoretical development consists of review of the simplified theory (Chapter 3), modeling merging and diverging behavior (Chapter 4), and the proposed extension and generalization (Chapter 5). The model validation includes methodology of model validation (Chapter 6), test sites and test data (Chapter 7), dynamic origin-destination estimation (Chapter 8), and results of model validation (Chapter 9). The conclusion is Chapter 10 which presents summary, conclusion, and future directions.

More specifically, Chapter 2 presents a literature review on macroscopic approaches of modeling traffic flow with a special interest in identifying the position of the simplified theory in this context. Chapters 3, 4, and 5 progressively develop the proposed theory. Chapter 3 briefly introduces the simplified theory and summarizes the theory in a computational procedure which serves as the major routine in the theoretical development of the proposed extension and generalization. Chapter 4 presents the proposed capacity-based weighted fair queuing (CBWFQ) merge model and the proposed contribution-based weighted splitting (CBWS) diverge model as well as their generalized forms. These two models will be incorporated in the proposed extension and generalization to help address network traffic. Chapter 5 presents the theoretical development of the proposed extension and generalization of the simplified theory. We formulated the extension based on 5 basic building blocks, i.e., an entrance, an exit, a mainline, a merge, and a diverge, and we assume that a freeway system can be

represented by certain combination of these basic building blocks. We further formulated the generalization based on a generic building block which may incorporate multiple upstream and/or downstream branches and we assume that a general transportation network can be represented by certain combination of some special cases of the generic building block. Chapters 6, 7, 8, and 9 present different aspects of model validation, a formal procedure to empirically test the model by comparing model output against field observations. Chapter 6 deals with methodology issues to guide subsequent steps to prepare and conduct the model validation. The main focuses of this chapter are data preparation and validation scheme. The data preparation part discusses data needs in the model validation and how to collection and process data to meet the data needs. The validation scheme part presents a systematic procedure to evaluate model performance. Chapter 7 presents the test sites and, for each of them, prepares two of the three pieces of information that are required as the model input. Chapter 8 prepares the last piece of information, i.e., the time-varying origin-destination (O-D) flows for one of the test sites (the other two sites do not require O-D estimation). With all the preparations, we conduct model validation by means of empirical tests and Chapter 9 presents test results of the comparing the model output against field observations. We evaluate the model performance based on the validation scheme developed in Chapter 6. Chapter 10 concludes the dissertation by summarizing this research, drawing conclusions, and identifying future research directions.

CHAPTER 2

LITERATURE REVIEW

The purpose of this chapter is to provide a historical perspective on the evolution of macroscopic traffic models (in this dissertation, we do not make any difference between a macroscopic traffic model and a continuum flow model) with a special interest in identifying the position of this research as well as the simplified theory in this context.

2.1 AN OVERVIEW OF THE HISTORY

Before we present more details of the literature, it might be helpful to provide an overview of the history regarding the evolution of macroscopic traffic models and Figure 2-1 serves such a purpose. In this figure, we place traffic models (represented by bold boxes) and their solutions (represented by regular boxes) on the same page in relative chronological order. The vertical axis represents the time and the axis runs from top to bottom with a lower position meaning more recent in the history. The figure consists of three panes with the top pane representing the area of hydrodynamics, the left pane representing the area of first-order continuum flow models, and the right pane representing the area of high-order continuum flow models. A line connecting two boxes represents the relation between these boxes with the one that the arrow points to being

derived (or originated) from the one at the other end of the line. Interpretation of the figure can be as follows. Most of the models can be traced, directly or indirectly, back to a common origin – the conservation law, from which first-order models, such as LWR (Lighthill and Whitham 1955; Richards 1956), Bick and Newell (1960), and Munjal and Pipes (1971), and high-order models, such as Payne (1971) and Whitham (1974), Phillips (1979), Kühne (1984, 1989), Kerner and Konhäuser (1993), Michalopoulos, et al (1993), Zhang (1998), and Treiber, et al. (1999), are derived. An other origin of Payne (1971) and Whitham (1974) is Prigogine (1961) which does not seem to have a connection with the conservation law. As far as model solution is concerned, Shock Waves is given by the LWR model itself. KRONOS (Michalopoulos 1984), KWaves (i.e., the simplified theory, Newell 1993a, 1993b, and 1993c), and CTM (Daganzo 1994, 1995a, 1995b, 1999) are derived from the LWR model. FREFLO (Payne 1971) is derived from Payne (1971) and Whitham (1974). Leonard (1997, 1998), Son (1996) and Hurdle and Son (2000), Banks (2000), and Ni and Leonard (2004a, 2004b, 2004c) are derived from KWaves. FREQ (May, et al 1991; May 1998) and CORQ (Yagar 1975) do not seem to have any connection with other models or solutions, so they are just put somewhere in the middle of the figure. Note that we only differentiate the terms “model” and “solution” in this chapter with a “model” focusing more on analytical formulations and a “solution” focusing more on computational procedures. Beyond this chapter, we do not make such a differentiation and a solution such as KWaves is also referred to as a model.

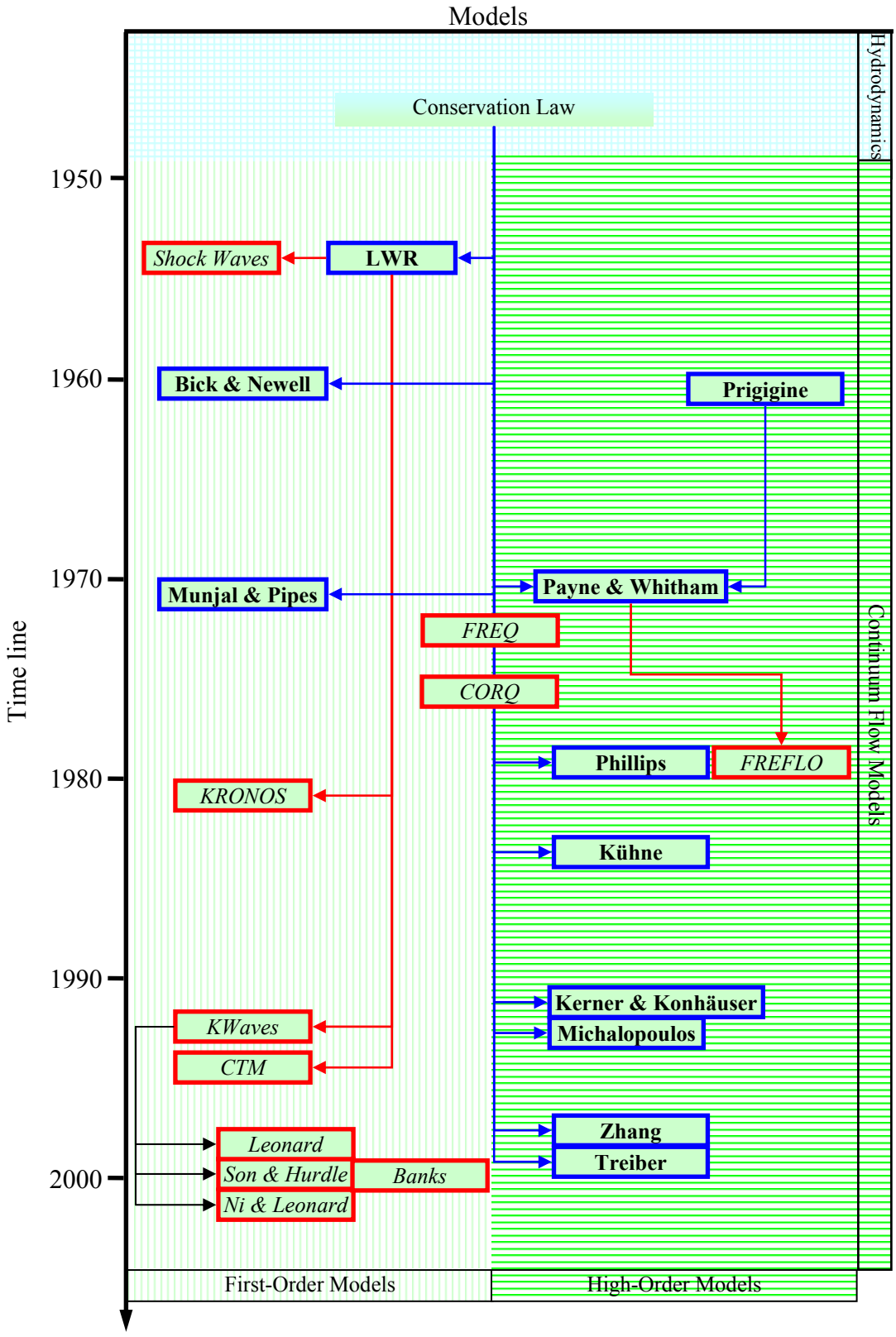


FIGURE 2-1 Taxonomy of continuum flow models

2.2 CONTINUUM FLOW MODELS

Originated from hydrodynamics, continuum flow models are typically associated with the following two assumptions:

i. Traffic flow is conserved, and

ii. There is a functional relationship between flow (or speed) and density.

The first assumption is expressed by the conservation or continuity equation. The conservation equation implies that in any traffic system inflow is equal to outflow plus storage. This principle was generally accepted without any question as to its validity. However, there was a large controversy surrounding the second assumption, partly because it was not well-understood or partly because of contradicting measurements. Modifications to the second assumption have led to a variety of continuum flow models, as will be reviewed shortly.

2.2.1 The Mass (or Vehicle) Conservation Equation

Consider a generic problem where a section of highway consists of an on-ramp and an off-ramp, as shown in Figure 2-2. If detectors are placed at an upstream station, a downstream station, the entrance, and the exit so that cumulative vehicle counts, $N_1(t)$, $N_2(t)$, $N_{on}(t)$, $N_{off}(t)$, respectively, are obtained at these stations as functions of time t. Let x_1 and x_2 denote the distances of the upstream and the downstream detectors, respectively, measured from some reference point; n(t) denote the content (number of vehicles) between the upstream and the downstream detectors.

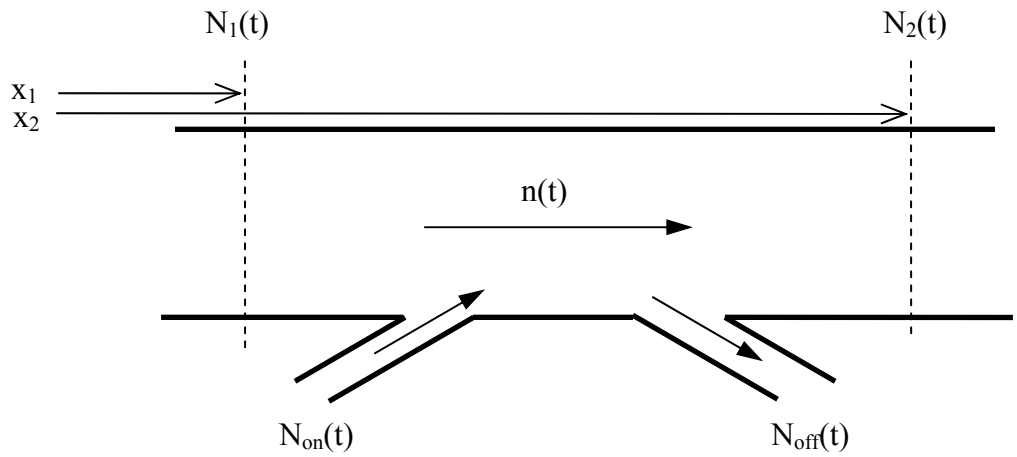


FIGURE 2-2 A section of highway

The law of vehicle conservation says that inflow is equal to outflow plus any storage, i.e., from time t_1 to time t_2, the following equation holds:

$$n(t_2) = n(t_1) + \left(N_1(t_2) - N_1(t_1)\right) + \left(N_{on}(t_2) - N_{on}(t_1)\right) - \left(N_2(t_2) - N_2(t_1)\right) - \left(N_{off}(t_2) - N_{off}(t_1)\right)$$

Let q(x,t) and k(x,t) denote the flow rate and traffic density, respectively, at location x and time t; l denote the number of lanes on the highway. Flows and densities can be determined as:

$$q(x_1, t_1, t_2) = \frac{N_1(t_2) - N_1(t_1)}{(t_2 - t_1) \times l}$$

$$q(x_2, t_1, t_2) = \frac{N_2(t_2) - N_2(t_1)}{(t_2 - t_1) \times l}$$

$$k(x_1, x_2, t) = \frac{n(t)}{(x_2 - x_1) \times l}$$

$$q_{on}(t_1,t_2)=\frac{N_{on}(t_2)-N_{on}(t_1)}{(t_2-t_1)}$$

$$q_{off}(t_1,t_2)=\frac{N_{off}(t_2)-N_{off}(t_1)}{(t_2-t_1)}$$

Then the equation of vehicle conservation can be re-written as:

$$\frac{k(x_1,x_2,t_2)-k(x_1,x_2,t_1)}{t_2-t_1}+\frac{q(x_2,t_2,t_1)-q(x_2,t_2,t_1)}{x_2-x_1}$$
$$=\frac{q_{on}(t_2,t_1)-q_{on}(t_2,t_1)}{(x_2-x_1)\times l}-\frac{q_{off}(t_2,t_1)-q_{off}(t_2,t_1)}{(x_2-x_1)\times l}$$

After taking limits $t_2\to t_1$ and $x_2\to x_1$ successively, the following partial differential equation is resulted:

$$\frac{\partial k(x,t)}{\partial t}+\frac{\partial q(x,t)}{\partial x}=g(x,t)$$

If no traffic enters and exits from ramps, the above equation reduces to:

$$\frac{\partial k(x,t)}{\partial t}+\frac{\partial q(x,t)}{\partial x}=0$$

This equation is the fundamental equation that appears in virtually all continuum flow models and it can be used to determine the flow at any section of the roadway. The attractiveness of this equation is that it relates two fundamental dependent variables, density and flow rate, with the two independent variables, time t and distance x. However, Solution of this equation is impossible because it contains two unknowns. Therefore, additional equation or assumption has to be supplied and this has spawned a variety of continuum flow models. For example, the assumption of functional flow-density relationship leads to the first-order models, while the addition of momentum equations results in high-order continuum models. These models are reviewed below.

2.2.2 First-order Models

First-order models concern with mass conservation but not momentum conservation.

LWR Model (1955)

Probably the first continuum flow model appeared in literature was developed by Lighthill and Whitham (1955) and, independently, Richards (1956), the so-called LWR model. This model was based on a hyperbolic system of conservation laws (Daganzo 1995a; LeVeque 1992; Whitham 1974) and assumed that the speed (or flow) of traffic is a function of density only.

For a single one-way highway without entrances or exits, the LWR model can be completely defined, using three variables that vary in time t and space x: flow, q(x, t), density, k(x, t), and speed, v(x, t), by a first-order partial differential equation and the functional relationship between flow and density or speed and density:

$$\begin{cases} \dfrac{\partial k(x,t)}{\partial t} + \dfrac{\partial q(x,t)}{\partial x} = 0 \\ q(x,t) = Q(k(x,t)) \end{cases} \text{ or } \begin{cases} \dfrac{\partial k(x,t)}{\partial t} + \dfrac{\partial q(x,t)}{\partial x} = 0 \\ v(x,t) = V_e(k(x,t)) \end{cases}$$

The most striking feature of this model is the existence of shock wave solution. However, this particular feature, in turn, is also one of the shortcomings that are mostly criticized because vehicles change speed abruptly when passing through shockwaves. This shortcoming was confirmed by the first attempt of applying this model to real world traffic operation (Newell 1962; Edie and Foote 1958; Edie and Baverez 1965)

Bick and Newell's model (1960)

Bick and Newell (1960) extended the LWR model by investigating two lanes of traffic flowing in opposite directions. The proposed model uses two equations of

continuity and assumes that the average speeds are functions of the densities in both lanes. They found that, even if the speed in one lane depends only weekly on the density in the other lane, the equations are of elliptic rather than the expected hyperbolic type for a certain small range of densities. For densities outside the range, solutions of the equations can be found for various special types of initial conditions.

Munjal and Pipes (1971)

Instead of dealing with the propagation of disturbance of traffic along freeway mainlines, Munjal and Pipes (1971) investigated the propagation of perturbation introduced by on- ramp flow on unidirectional two and three-lane freeways. The perturbed traffic density on the freeway initiated by the on-ramp flow is defined as the difference in the actual instantaneous prevailing traffic density and the equilibrium traffic density which would prevail under steady-state traffic conditions. The basis of this model is the use of separate equations of continuity of an idealized traffic fluid under equilibrium conditions for each lane and a hypothesis for lane-interactions.

2.2.3 High-order Models

High-order models concern with both mass conservation and momentum conservation.

Prigogine's model (1961)

To find meaningful refinement to overcome the apparent shortcomings of the LWR model, researchers began to consider connections between continuum models to the development of car-following models, and many high-order continuum flow models are developed. The first of such models was developed by Prigogine (1961). When modeling

the interactions among vehicles, Prigogine hypothesized that a faster vehicle is slowed down by a slower leader only if the faster one reaches the position of the slower one in the same lane and is meanwhile blocked from changing lanes by the presence of vehicles in adjacent lanes. It was also assumed that, whenever a chance comes, the trapped vehicle will free to pass and instantaneously resume its desired speed. In this model, vehicles are assumed zero length and zero reaction time.

PW model (1971)

Proposed by Payne (1971) and, independently, Whitham (1974), the PW model is a system of two equations, the first of which is the conservation of mass equation given in the LWR model. The second equation in the system is derived from the Navier-Stokes equation of motion for a one-dimensional compressible flow with a pressure and a relaxation term.

$$\begin{cases} \dfrac{\partial k}{\partial t} + \dfrac{\partial q}{\partial x} = 0 \\ \dfrac{\partial v}{\partial t} = -v\dfrac{\partial v}{\partial x} - \lambda\left(v - V_e(k)\right) - \dfrac{1}{k}\dfrac{dP}{dk}\dfrac{\partial k}{\partial x} \end{cases}$$

where P(k) is the traffic pressure.

Unfortunately, the PW model does not describe the well-known self-organization phenomenon of stop-and-go waves.

Phillips' model (1979)

Considering that a faster vehicle has to reduce its speed before it reaches the position of the slower leader, Prigogine (1961) over simplified the situation. Phillips (1979) developed a statistical kinetic model for the flow on an n lane highway. The model is based on the assumption that each vehicle will maintain a minimum space between itself

and the leading vehicle proportional to its speed. Phillips' model consisted of a continuity equation, a momentum equation, and a specification of the traffic pressure as a function of density:

$$\begin{cases} \dfrac{\partial k}{\partial t}+\dfrac{\partial q}{\partial x}=0 \\ \dfrac{\partial v}{\partial t}+v\dfrac{\partial v}{\partial x}=\lambda\left(V_e-v\right)-\dfrac{1}{k}\dfrac{dP}{dk}\dfrac{\partial k}{\partial x} \\ P=P(k) \end{cases}$$

where symbols are defined as before and V_e is the equilibrium mean speed.

Kühne's model (1984)

Kühne (1984, 1989) also proposed a model by considering sound speed and viscosity:

$$\begin{cases} \dfrac{\partial k}{\partial t}+\dfrac{\partial q}{\partial x}=0 \\ \dfrac{\partial v}{\partial t}+v\dfrac{\partial v}{\partial x}=-\dfrac{1}{\tau}\left(v-V_e(k)\right)-\dfrac{c_0^2}{k}\dfrac{\partial k}{\partial x}+\eta\dfrac{\partial^2 v}{\partial x^2} \end{cases}$$

where where c_0 is the sound speed, and η is the viscosity constant.

Kerner and Konhäuser's model (1993)

Kerner and Konhäuser (1993) showed that given an initially homogeneous traffic flow, regions of high density and low average speed (clusters of cars) can spontaneously appear. These high-density regions can move either with or against the flow of traffic, and two clusters with different speeds, widths, and amplitudes merge when they meet, resulting in a single cluster. The continuum flow model adopted is in the following form:

$$\begin{cases} \dfrac{\partial k}{\partial t}+\dfrac{\partial q}{\partial x}=0 \\ \dfrac{\partial v}{\partial t}=-\dfrac{c_0^2}{k}\dfrac{\partial k}{\partial x}+\dfrac{1}{\tau}\left(V_e(k)-v\right)+\dfrac{1}{k}\dfrac{\partial(\eta\partial v/\partial x)}{\partial x} \end{cases}$$

where η is the viscosity.

Michalopoulos' model (1993)

Michalopoulos, et al (1993) proposed a model which does not require the use of an equilibrium speed-density relationship. Traffic friction at interrupted flows and changing geometries is also addressed through the use of a viscosity term. Tests with field data and comparison with existing models suggested that the proposed mode is more accurate and computationally more efficient.

$$\begin{cases} \dfrac{\partial k}{\partial t}+\dfrac{\partial q}{\partial x}=0 \\ \dfrac{\partial v}{\partial t}+v\dfrac{\partial v}{\partial x}=\dfrac{1}{\tau}\left(V_f-v\right)-G\dfrac{\partial v}{\partial t}-\upsilon k^{\beta}\dfrac{\partial k}{\partial x} \end{cases}$$

where $G=\mu k^{\varepsilon} g$, and μ, υ, ε, and β are all constant parameters, and V_f is the free flow speed.

Zhang's model (1998):

Zhang (1998) proposed a non-equilibrium traffic flow model which is based on both empirical evidence of traffic flow behavior and basic assumptions on drivers' reaction to stimuli. By assuming an equilibrium speed-density relationship and introducing a disturbance propagation speed, the model includes the LWR model as a special case and removes some of its deficiencies. Unlike existing high-order continuum models, this model does eliminates "wrong-way travel" because in this model traffic disturbances are always propagated against the traffic stream.

$$\begin{cases} \dfrac{\partial k}{\partial t}+\dfrac{\partial q}{\partial x}=0 \\ \dfrac{\partial v}{\partial t}+v\dfrac{\partial v}{\partial x}=\dfrac{1}{\tau}\left(V_e(k)-v\right)-k(V_e'(k))^2\dfrac{\partial k}{\partial x} \end{cases}$$

Treiber's model (1999)

Treiber, et al. (1999) derived macroscopic traffic equations from specific gas-kinetic equations and the resulting partial differential equations for the vehicle density and average speed contain a non-local interaction term which is very favorable for a fast and robust numerical integration, so that several thousand freeway kilometers can be simulated in real time.

$$\begin{cases} \dfrac{\partial k}{\partial t}+\dfrac{\partial q}{\partial x}=0 \\ \dfrac{\partial v}{\partial t}+v\dfrac{\partial v}{\partial x}=\dfrac{1}{k}\dfrac{\partial(kAv^2)}{\partial x}+\dfrac{1}{\tau}(V_e-v)-\dfrac{V_eA(k)}{\tau A(k_j)}\left(\dfrac{k_aTv}{1-k_a/k_j}\right)^2 B(\delta_v) \end{cases}$$

where A=A(k) is a density-dependent function, k_a is the density at point x_a ahead of x, $B(\delta_v)$ is a macroscopic interaction term, and $V_e(k)$ is the normal equilibrium relationship between v and k.

2.2.4 Relation among the Continuum Flow Models

Starting from the vehicle conservation equation, a variety of continuum flow models are developed by supplying additional assumptions. Generally, these models can be summarized by the following model (Yi, et al 2003):

$$\begin{cases} \dfrac{\partial k}{\partial t}+\dfrac{\partial q}{\partial x}=g(x,t) \\ \dfrac{\partial v}{\partial t}+v\dfrac{\partial v}{\partial x}=-\dfrac{1}{k}\dfrac{\partial P}{\partial x}+\dfrac{1}{\tau}(V_e-v) \end{cases}$$

where $V_e(k,v)$ is the generalized equilibrium speed, which is given by the steady-state relationship between highway speed v and density k. P(k,v) is the traffic pressure, and τ is the relaxation time, which is the time constant of the regulating traffic speed v to the equilibrium speed V_e. For a highway without on- or off-ramps, g(x,t)=0.

Each of the above-mentioned continuum flow models can be viewed as a special case of the general model when applying different traffic pressure P, relaxation time τ, and the generalized equilibrium speed Ve. For example,

1) if τ=0 and P=0, LWR model is resulted;

2) if $P = -\frac{V_e(k)}{2\tau}$ with $V_e(k,v) = V_e(k)$, PW model is resulted;

3) if $P = k\Theta$ with $\Theta = \Theta_0(1 - \frac{k}{k_j})$ and k_j the jam density, Phillips' model is resulted.

4) if $P = k\Theta_0 - \eta\frac{\partial v}{\partial x}$, Kerner and Konhäuser model is resulted

5) Michalopoulos' model is resulted if $P = \frac{v}{\beta+2}k^{\beta+2}$, where v is an anticipation parameter and β a dimensionless constant; and $V_e(k) = V_f$, where V_f is the free flow speed.

6) Zhang's model is resulted if $P = \frac{1}{3}k^3 V'^2_e(k)$ with $V'^2_e(k) = \frac{dV_e(k)}{dk}$.

7) Treiber's model is resulted if $P = kAv^2$ where A=A(k) is a density-dependent function, and $V_e(k,v) = V_e(k)\left[1 - \frac{A}{A(k_j)}\left(\frac{k_a T v}{1 - k_a/k_j}\right)^2 B(\delta_v)\right]$.

2.2.5 Relative Advantages of the Continuum Flow Models

As a first-order continuum flow model, LWR model is proposed on dense traffic with equilibrium and it is flawed for light traffic. This is because, when passing is allowed, LWR model fails to recognize that the preferred speed for each vehicle varies over time and the desired speeds among a group of vehicles vary as well. These variations can cause a platoon to disperse in a way that is not predicted by LWR model. When passing is restricted, LWR model produces unsatisfactory results in the following three aspects (Daganzo 1995c). First, LWR model predicts abrupt speed change when a vehicle passes through a shockwave, an action that is unrealistic in real situation. Second, LWR model fails to predict instabilities of the stop-start traffic. Third, LWR model assumes zero reaction time which does not happen in real world.

Given these deficiencies, continuum flow models developed thereafter try to fix them and almost all these models follow a direction of incorporating a momentum conservation equation. An early attempt to fix the deficiencies in LWR model was made by Prigogine (1959) who proposed a kinetic model incorporating a speed distribution to address platoon dispersion. A decade later, Payne (1971) and Whitham (1974) proposed a dynamic model, the so-called PM model, trying to smooth out the discontinuity in speed change across shockwaves. A momentum equation was introduced in this model to describe the structure of a shockwave. This seminal work has inspired many thoughts in analytical explanation of shockwave behavior and, thus, has spawned several variants, among which are Phillips (1979), Kühne (1984, 1989), Kerner and Konhäuser (1993), Michalopoulos, et al (1993), Zhang's model (1998), Treiber, et al. (1999), etc.

Several deficiencies (Daganzo 1995c) are found in PW model. First, it does not remove all the shockwaves. Second, as reported by de1 Castillo, et al. (1993), vehicles in PW model can adjust their speeds in response to disturbance from behind, while in real situation vehicles typically respond to their leaders. Third, PW model incorporates a momentum equation which is derived from a car-following model. This momentum equation neglects second- and higher-order terms of spacings and speeds which may not be negligible when spacings and speeds are not slowly varying. Fourth, PM model as well as other high-order model always produce wave speeds that are greater than traffic speeds. This is an unattractive feature to macroscopic models because it implies that the future conditions of a vehicle are partially decided by what happens behind it. Fifth, the strength that high-order models smooth out shocks turns out to be these models' weakness. This is because any model that attempts to smooth all the discontinuities must sometimes predict negative speeds and such negative speeds observed in computer models cannot be removed by convergent numerical approximation methods. Sixth, but probably not the last, high-models involve more complex partial differential equations and more variables which increases computational complexity and are more difficult to calibrate and implement. Given these shortcomings, it is therefore recognized that, using well-established techniques, high-order models (despite their added complexity and additional parameters) do not improve the LWR model (Michalopoulos, et al., 1987; Leo and Pretty, 1992; Daganzo 1995c).

2.3 SOLUTIONS TO THE CONSERVATION EQUATION

In general, solutions to the conservation equation can be classified as analytical and numerical. If a solution is analytical, we mean it is continuous in either time or space or

both. If a solution is numerical, we mean is it discrete in both time and space. In the light of this, the solutions identified can be summarized in the following table:

TABLE 2-1 Classification of solutions to the conservation equation

Analytical Solutions	**Numerical Solutions**
Shock Waves	FREFLOE
KWaves	FREQ
	KRONOS
	CORQ
	CTM
	Extension to the simplified theory

The table shows that Shock Waves (Lighthill and Whitham 1955; Richards 1956) and KWaves (Newell 1993a, 1993b, and 1993c) are analytical solutions. FREFLO (Payne 1971), FREQ (May, et al 1991; May 1998), KRONOS (Michalopoulos 1984), CORQ (Yagar 1975), CTM (Daganzo 1994, 1995a, 1995b, 1999), and this research, i.e., the extension to the simplified theory (Ni and Leonard 2004a, 2004b, 2004c) are numerical solutions.

2.3.1 Analytical Solutions

We discuss here the shock waves solution and Newell's simplified theory of kinematic waves.

Shock Waves (1955)

Shock waves solution works with LWR model by means of characteristics, a family of curves along which density is constant. Starting from the boundaries of the time-space domain, the characteristics emanate at a slope equal to the tangent of the flow-density

curve at the point representing the flow conditions at the boundary from which the characteristic emanates. The density at any point (x, t) of the time space domain is found by drawing proper characteristic passing through that point. When two characteristics intersect, density at this point is multi-valued, a shockwave is generated and the characteristics terminate. A shockwave represents the abrupt change in speed, flow, and density, and the speed of the shockwave can be found by connecting the two points on the flow-density curve representing the upstream and downstream conditions. A positive speed means that the shockwave moves along the direction of traffic, while a negative speed means the shockwave travels against traffic.

The Simplified Theory / KWaves (1993)

In contrast to shock waves solution, Newell's simplified theory of kinematic waves (1993a, 1993b, and 1993c), is continuous in time but discrete in space. Newell has made some simplifying assumptions to achieve computational efficiency. The most striking one among these assumptions is the triangular (or trapezoidal) flow-density relationship, by which there are only two wave speeds, one traveling with the traffic at the free flow speed and the other moving against the traffic as a function of capacity and jam density. The nice feature of the simplified theory solution is that one does not worry about the movements of the finite elements of traffic in the time-space domain nor what happens in between two nodes (typically locations on the highway where physical conditions or traffic characteristics change). It only counts vehicles at a selection of nodes along the highway and traffic states (e.g., flow, speed, density) can be retrieved by analyzing these cumulative vehicle counts. At each node, the amount of vehicles that can pass is

determined by applying the minimum principle, i.e., the departure count is the minimum of upstream arrival and amounts allowed by local capacity and downstream congestion.

2.3.2 Numerical solutions

FREFLO (1971)

FREFLO is a discretized version of the continuum model that Payne (1971) developed in the early 1970s. FREFLO simulates traffic flow on freeways using a formulation of aggregate variables based on suitably modified analogies of fluid flow. Initial work with the FREFLO revealed that the model was limited in its ability to realistically simulate congested flow conditions (Rathi, et al 1987). Many efforts were made to address this problem, including the development of another freeway model, FRECON (Babcock, et al 1984), which adopted the heuristic scheme in FREFLO. FREFLO itself was modified to resolve the difficulties in representing congested conditions and was incorporated into TRAF. TRAF allows FREFLO to interface with other simulation models that can simulate the neighboring urban surface street systems. Within TRAF, an equilibrium traffic assignment model exists that may be used to provide volume and routing information to FREFLO. One of the shortcomings of FREFLO is that it has more difficulty dealing with the time/space nature of traffic flow because traffic flow must be dealt within aggregate time slices.

FREQ (early 1970s)

FREQ (May, et al 1991; May 1998) was developed at the University of California, Berkeley in the early 1970s. FREQ consists of a family of freeway simulation models and it can be described as quasi-static macroscopic: quasi-static because changes in demand

levels can only be input at specific times; macroscopic because the movement of individual vehicles is not modeled. Operating on the basis of speed/volume and demand/capacity relationships, FREQ calculates freeway and ramp capacities using the Highway Capacity Manual, and bottleneck capacities checked in the field. The FREQ model uses the standard shockwave theory to model the sudden drop of lane capacities that may be caused by traffic incidents. The macroscopic and deterministic mature of the FREQ model significantly simplifies the estimation of parameters.

Over the years, FREQ has gone through many upgrades and the most recent version is FREQ12 which includes an interactive graphical interface with comprehensive input checking, carefully selected default values, on-screen graphic representation of the simulation results, and user-selected output options including traffic performance contour maps. The two models contained within the FREQ12 system are: FREQ12PE, an entry control macroscopic model for analyzing ramp metering; and FREQ12PL, an on freeway priority macroscopic model for analyzing HOV facilities.

The primary weakness of FREQ is the over-simplification of arterial alternatives and lack of full diverting/rerouting techniques (Van Aerde, et al 1987). On the other hand, FREQ's two programs (FREQPL and FREQPE) can not be run concurrently and there is no direct interaction between the two programs.

KRONOS (early 1980s)

KRONOS was developed in the early 1980s at the University of Minnesota by Michalopoulos (Michalopoulos 1984). This model has been under continuous refinement since its inception and several versions are described in the literature (Michalopoulos and Lin 1986; Michalopoulos, et al 1991; Kwon and Michalopoulos 1995). KRONOS uses a

simple continuum model based on the LWR model. A finite differencing scheme with discretization over space (Δx =100 ft.) and time (Δt =1 sec.) enables it to solve 1-dimensional, time-dependent, compressible flow containing shock waves. In KRONOS, the flow-density relationship can be linear in light-to-moderate traffic conditions. KRONOS explicitly models interrupted flow behaviors such as accelerating, decelerating, lane changing, merging, diverging, weaving, and spillback, which were not taken into account by other macroscopic freeway programs. Disadvantages include unforgiving user-input overrides and the tendency to overestimate the benefit of freeway lane addition, particularly for cases where the bottleneck is mostly attributable t heavy weaving among neighboring on-and off-ramps (Prevedouros and Li 2000).

CORQ (1975)

CORQ (CORridor Queuing), developed by Yagar (Yager 1975), is a freeway corridor simulation/assignment model. The corridor consists of a directional freeway, its ramps, major cross streets, and any competing alternative surface streets. Traffic flows are approximated as fluids, and travel times are calculated as simple step functions for both free-flowing and congested conditions. A key element of CORQ is the dynamic assignment technique for allocating time-slice O-D demands to a time-dependent traffic network. However, the travel time relationship is expressed as a static step function of link flows and intersection delays: this is a drawback of CORQ (Van Aerde, et al 1987). The time relationship is insensitive to changes in signal timings on parallel arterials. Because CORQ was perhaps the most detailed corridor-level model throughout the 1980s, parts of its modeling approach were modified and incorporated into the design of

the area-level integrated network simulation model INTEGRATION which was developed in the late 1980s by Van Aerde.

CTM (1995)

Developed by Daganzo (1994, 1995a, 1995b, 1999), CTM is a finite difference approximation of the LWR model. In the proposed approximation, the network is represented as a set of cells with the lengths usually set equal to the distance traveled in light traffic by a typical vehicle in one clock tick. In each cell in each time interval, it is assumed that there is a continuous flow of traffic and that the state is known in terms of a point in its speed-flow-density diagram. The flow between two neighboring lattice points is the minimum of the two values returned by: a "sending" function evaluated at the density prevailing at the upstream lattice point, and a "receiving" function evaluated at the downstream lattice point. The sending and receiving functions correspond to the increasing and decreasing branches of the freeway's flow-density curve. The CTM then updates the state of a cell based on a recursion where the cell occupancy at the next tick of time is equal to the number of vehicles that was in that cell before, plus the number of vehicles that entered, minus the number of vehicles that left.

Extension to the Simplified Theory (2004)

Newell's simplified theory of kinematic waves was extended and generalized in this research to address network traffic (Ni and Leonard 2004a, 2004b, 2004c) and the resulting model is a numerical approximation in discrete time and space. Key features of this model include a computational procedure derived from the simplified theory and a capacity-based weighted fair queuing model and a contribution-based weighted splitting model to facilitate handling of merging and diverging traffic.

2.3.3 Comparison of CTM and the Extension to the Simplified Theory

This comparison highlights the similarities and differences between CTM and the extension to the simplified theory.

2.2.3.1 The similarities

Both are originated from the LRW model

Both CTM and the extension to the simplified theory are numerical approximations to LWR model based on some simplifying assumptions

Both are pure macroscopic models

Both CTM and the extension to the simplified theory are pure macroscopic traffic simulation models. "Pure macroscopic" means traffic is treated as a continuous compressible fluid whose aggregate measures such as flow, speed, density are of primary interest. Driver personalities and individual vehicle characteristics are ignored in such a model.

Both solve the same type of problems

Both CTM and the extension to the simplified theory address the well-posed problem of traffic evolution in a highway network given time-dependent origin-destination flows.

Both yield similar results

As parallel models approximating the LWR model, CTM and the extension to the simplified theory theoretically guaranteed to generate similar results for the same problem.

2.2.3.2 The differences

Flow-density relation

Though the finite difference scheme of approximating LRW model is proposed on a general flow-density curve, the implementation of this scheme, the CTM, is based on a trapezoidal flow-density curve. The extension to the simplified theory deals with a triangular (or trapezoidal) flow-density curve as was the case in the simplified theory.

Working units

CTM works on finite elements of traffic, called cells, of size Δx by Δt, where Δx is the distance traveled at free flow speed during Δt. CTM keeps track of cell contents by means of vehicle conservation. The extension to the simplified theory works on links and nodes. Links are homogeneous sections of roadway and the geometric properties and traffic characteristics remain the same in a link. The extension to the simplified theory keeps track of the cumulative number of vehicles past every node by means of the minimum principle. There are several advantages of introducing cumulative vehicle counts in solving LWR problems, as can be seen in state update, vehicle conservation, and shock conditions.

State update

CTM advances system state in a recursive manner such that, at each clock tick, the flow in each cell is updated based on the ability to send at the current cell and the ability to receive at the next cell. Then the density of the cell is updated based on vehicle conservation. Finally, speed is updated based on the fundamental relationship among flow, speed, and density.

In the extension to the simplified theory, system state is stored in cumulative vehicle counts a selection of nodes. At each node and each clock tick, the departure count takes the minimum of upstream arrival, the count allowed by local capacity, and the count allowed by downstream congestion, if any, and all these counts are in cumulative terms.

Rather than working on flow, speed, density as in CTM, the extension to the simplified theory does nothing more than just counting vehicles at a selection of nodes and, therefore, is simpler in principle.

Vehicle conservation

To conserve vehicles, CTM has a specific equation to ensure the conservation based on the density, the inflow and the outflow at a cell, while, in the extension to the simplified theory, vehicle conservation is automatically guaranteed by the introduction of cumulative vehicle counts/curves. Intuitively, if each node is monitored by an observer who attaches numbers consecutively to vehicles passing by him, it is guaranteed that no vehicle will disappear and no vehicle will suddenly appear – vehicles are conserved. Theoretically, the second derivative of cumulative curve to space and then to time is equal to the second derivative of cumulative curve to time and then to space, if these second derivatives exist, and this relation is equivalent to the conservation equation.

Shock conditions

In CTM, shock waves are captured by discontinuity in flows, speeds, and densities of two consecutive cells and the shock can be made arbitrarily sharp by using finer meshes, while, in the extension to the simplified theory, one just counts vehicles at both ends of a link and does not have to worry about where and when a shock appears. To determine the

time at which a shock passes any location x* or the location of a shock at any time t*, one needs not to follow the actual path of the shock. It suffices to construct the cumulative curve versus t for fixed x* and see where the curve has discontinuity in the slope q(x*, t), or construct the cumulative curve versus x for fixed time t* and see where it has a discontinuity in slope k(x, t*).

Wave propagation

In CTM, propagation of disturbances in traffic flow (i.e., waves) is not explicitly modeled, but can be retrieved from the traffic states (flow, speed, and density) of the cells. In the extension to the simplified theory, wave propagation is explicitly modeled by means of translation, i.e., a forward wave at an upstream node x_1 will arrive at a downstream node x_2 after a free trip time, while a backward wave at x_2 will be sensed by x_1 after a backward wave travel time and an adjustment of jam storage. The advantage of such the translation is that the impact of any wave on a location can be readily determined by examining the known conditions at its upstream and downstream nodes.

Merge model

The merge model proposed in CTM gives a complete solution in that it incorporates all the possibilities of merge queuing and allows analysts to adjust priority factors of the merging branches. However, this advantage happens to be to drawback of the model because not every possibility is equally likely and, to obtain the priority factors that make the most sense, analysts have to go through calibration process which can be time- and resource-consuming. Developed independently, the merge model in the extension to the simplified theory is a special case of Daganzo's, but the former saves the costly

calibration process by assigning the most realistic values to the priority factors of the merging branches. On the other hand, the merge model is proposed on such a scheme that it is readily extensible at no additional cost.

Diverge model

Both the diverge models in CTM and the extension to the simplified theory recognize that queues can be developed from either of the diverging branches. However, their difference lies in how the queues affect the upstream common link. In CTM, it is assumed that a downstream queue will block the entire upstream link, while, in the extension to the simplified theory, a downstream queue blocks only part of the upstream link and an optional time-varying constraint on capacity can be applied on the upstream link to capture the possible friction between the congested and uncongested flows on the link. For example, if a queue spills back from an off-ramp, there is no blockage of the main line at the point of the off-ramp. As one moves upstream, the capacity constraint may move along a triangle from the diverge across the multiple mainline lanes until fully encompass these lanes. The slope of this triangle depends on the blockage of the off-ramp and the mainline flows. This model updates upstream capacities depending on the ratio of flows at the off-ramp, and, therefore, is more realistic.

Memory consumption

To achieve fine resolution, CTM needs to keep cell sizes small and this inevitably renders the model memory intensive. In the extension to the simplified theory, there is virtually no constraint on link lengths as long as the links remains homogeneous, so

traffic states can be stored in a selection of nodes and this makes the extension to the simplified theory more memory efficient.

Computation efficiency

Unlike CTM which keeps track of traffic states by going through a recursion and working on the finite difference equation of vehicle conservation, the extension to the simplified theory just counts vehicles past some nodes and, when a departure count is multi-valued, just takes the minimum of them. This is much simpler in concept and makes the model more computationally efficient.

Origin-destination flows

Both CTM and the extension to the simplified theory assume time-varying origin-destination (O-D) flows. CTM advances traffic to diverging branches based on some time-varying turning percentages derived from the time-varying O-D flows. In contrast, the extension to the simplified theory advances vehicles through the network according to their predetermined O-D path, i.e., a sequence of links that a vehicle travels along from it enters the network until it exits the network.

Real-time application

One of the promising features of the extension to the simplified theory is that this model is capable of applying in real time. This is because this model works faster than real time and, at each step, the model can be stopped to allow analysts to modify future demand and/or create incidents and the model is able to continue based on the updated

information. This is different than traditional simulation models which typically have to run to the very end before one can modify input data and/or view simulation results.

Application to General Transportation Networks

The extension to the simplified theory has been theoretically demonstrated that it also applies to general transportation networks because it is capable of dealing with general merges and diverges with multiple merging and diverging branches.

2.3 THE SIMPLIFIED THEORY AND ITS RELATED WORKS

There are several endeavors to further Newell's work. Banks (2000) proposed an on-ramp queuing model by constraining ramp departure by arrival, capacity, and metering rate. This model implicitly assumes that ramp traffic is always dictated by forward waves and backward waves never back onto the ramp, an assumption that is always true for meter-controlled on-ramps. Hurdle and Son (Son 1996; Hurdle and Son 2000) conducted a validation of Newell's theory and tested the accuracy of Newell's theory and the adequacy of its underlying assumption, the triangular flow-density relation, with real data collected from freeways in the San Francisco Bay Area. The test results supported the validity of Newell's theory, and showed that the theory works best under over-saturated conditions. Leonard (1997, 1998) coded Newell's theory into software GTWaves, which bridged the theory and its application.

For some reason, Newell confined the theoretical presentation of his theory to a freeway mainlines, i.e., no queuing on ramps, which greatly limits its practical appeal. The extension to the simplified theory attempts to remove the limitation by extending and generalizing the original theory so that the simple and elegant algorithm applies to a

freeway system and even a general transportation network. Figure 2-3 summarizes the above discussion and highlights the position of this research, i.e., the extension to the simplified theory.

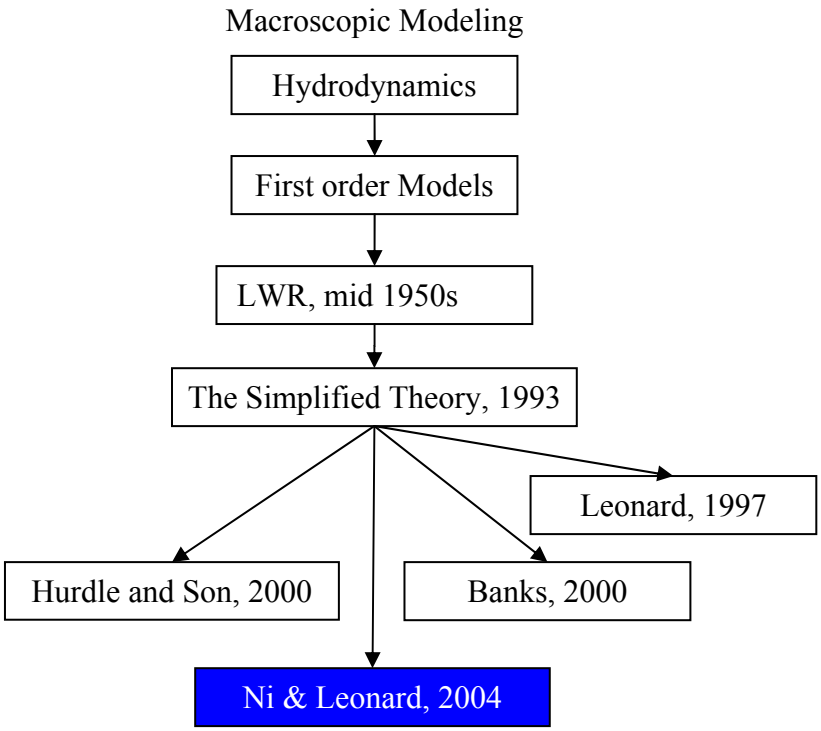


FIGURE 2-3 The simplified theory and its related works

2.4 SUMMARY

This chapter provided a historical perspective on the evolution of continuum flow models with a special interest in identifying the position of this research in the historical context. We identified the existing continuum models including both first-order and high-order models and discussed the relative strengths of the first-order models vs. the high-order models. Some studies showed that, though the LWR model as a first-order model

provides only a coarse description of traffic behavior, the gains of the high-order models such as the PW model do not seem to justify their complexity. This partly explains why the efforts of solving the traffic models are primarily attempted on the LWR model and the simplified theory is such an example. Being discrete approximations of the LWR model, the CTM and the extension to the simplified theory bear some similarities and differences, both of which are discussed in this chapter. In addition, this chapter identified some research efforts closely related to the simplified theory and positioned this research in a more specific context.

CHAPTER 3

REVIEW OF THE SIMPLIFIED THEORY

This section summarizes the key assumptions and modeling approaches of Newell's simplified theory of kinematic waves. To be consistent with the original papers, this section follows Newell's notation. To facilitate the development of the proposed extension and generalization, the simplified theory highlighted in this chapter is a discrete version in time-space domain.

3.1. ASSUMPTIONS

The following assumptions (or simplifications) are made (or implied) in the simplified theory:

Assumption 1: To facilitate handling of wave propagation, Newell postulated a triangular (or trapezoidal) flow-density relationship, as shown in Figure 3-1. Therefore, there are only two constant wave speeds: a forward wave speed, v, in under-saturated flow, which happens to be the free flow speed and is independent of flow, and a backward wave speed, u, in congested flow, which is a simple function of free flow speed, v, capacity, Q, and jam density, k_j. To be able to apply this simplification, the

freeway mainline is subdivided until each freeway section can be considered homogeneous.

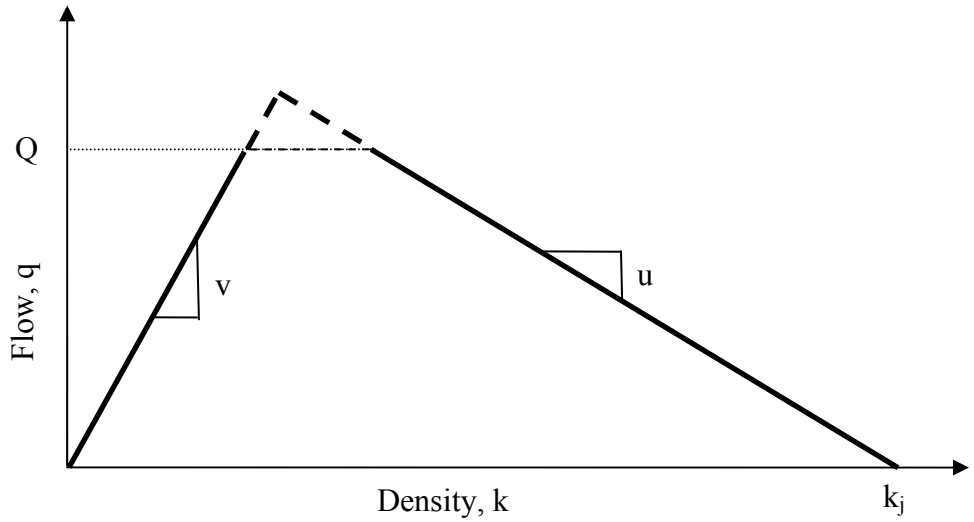


FIGURE 3-1 Triangular (or trapezoidal) flow-density relationship

The only use of the triangular (or trapezoidal) flow-density relationship in the simplified theory is to cleanly propagate waves (disturbances in traffic) forward and backward so that one can relate the conditions at an upstream location and a downstream location to a location between them. However, this triangular (or trapezoidal) flow-density relationship does not necessarily imply that traffic stream characteristics are constrained to this curve, as can be seen in the comparison of observed and predicted flow-density relationship in Chapter 9.

Assumption 2: When dealing with on-ramps, Newell assumed that on-ramp traffic can always bypass a queue, if any, at the merge and experiences no delay, see Figure 3-2.

Therefore, there is actually no queuing at the on-ramp because, whenever there is any demand, it is satisfied without delay.

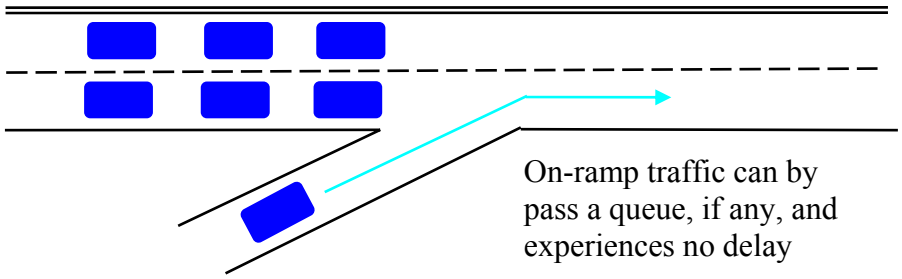


FIGURE 3-2 Assumption on on-ramp traffic

Assumption 3: Travel time of all vehicles in a freeway section is independent of their destinations. Therefore exiting vehicles experience the same travel time as through vehicles in this section. On the other hand, exiting vehicles can always be able to exit without delay as long as they have successfully arrived at the diverge, i.e., there is actually no queuing at the off-ramp, either, see Figure 3-3.

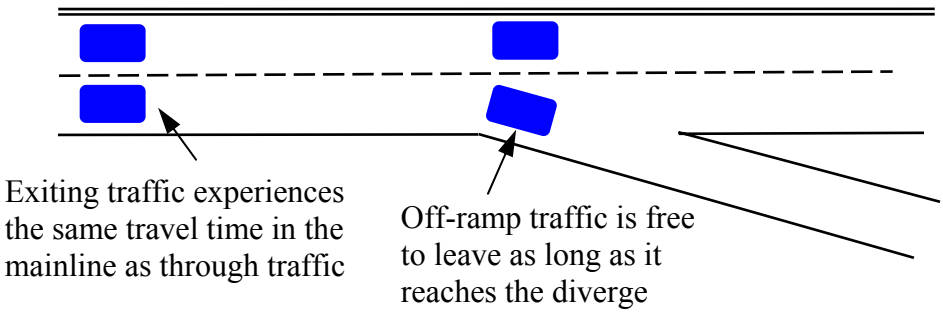


FIGURE 3-3 Assumption on off-ramp traffic

Assumption 4: To keep his theory focused, Newell assumed that the time-dependent origin-destination (O-D) flows are given and they are consistent with any system constraint.

3.2. WAVE PROPAGATION RULES

With the simplified flow-density relation, wave propagation is simple and can be accomplished by some translations.

3.2.1. Forward Wave Propagation Rule

If there is no congestion, i.e., traffic is dictated by forward waves, traffic demand at an upstream location x_0 will arrive at a downstream location x after a free trip time, i.e., the time to traverse this section at the free flow speed, $(x - x_0)/v$. Therefore, the cumulative number of vehicles versus t curve at location x, $N(x,t)$, can be obtained by translating the cumulative curve $N(x_0,t)$ versus t at the upstream location x_0 horizontally to the right by a time displacement $(x - x_0)/v$, where v is the forward wave speed and $x > x_0$. The trip time of any vehicle from x_0 to x is $(x - x_0)/v$, independent of t. This rule is illustrated in Figure 3-4.

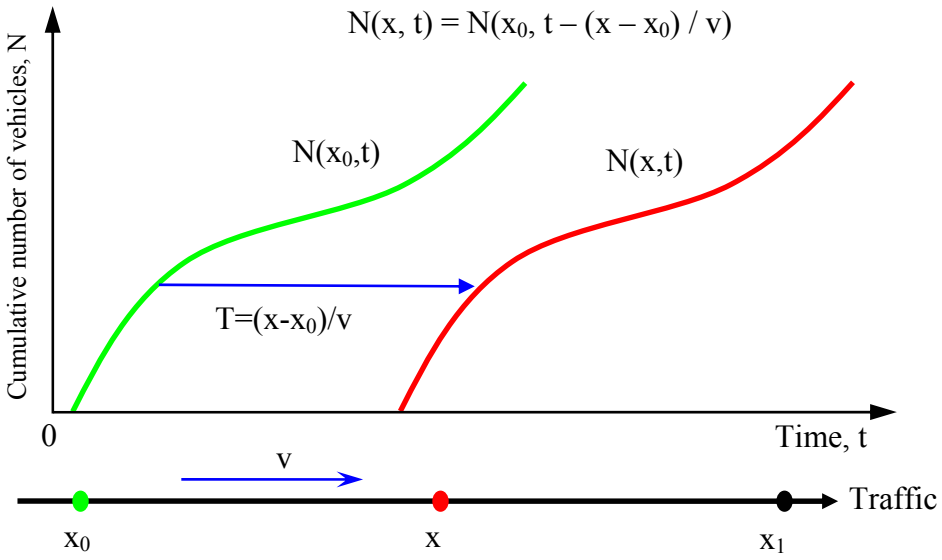


FIGURE 3-4 Forward wave propagation rule

3.2.2. Backward Wave Propagation Rule

If there is congestion, i.e., traffic is dictated by backward waves, the disturbance at a downstream location x_1 will be sensed at an upstream location x after a backward wave travel time, i.e., the time for the backward wave to traverse this section at the backward wave speed u, $(x_1 - x)/u$. Therefore, the cumulative curve $N(x,t)$ versus t at location x can be obtained by translating the cumulative curve $N(x_1,t)$ at the downstream location x_1 horizontally to the right by a time displacement $(x_1 - x)/u$ and vertically upward by a jam storage $k_j(x_1 - x)$, where $x_0 < x < x_1$ and k_j is the jam density of the section. This rule is illustrated in Figure 3-5.

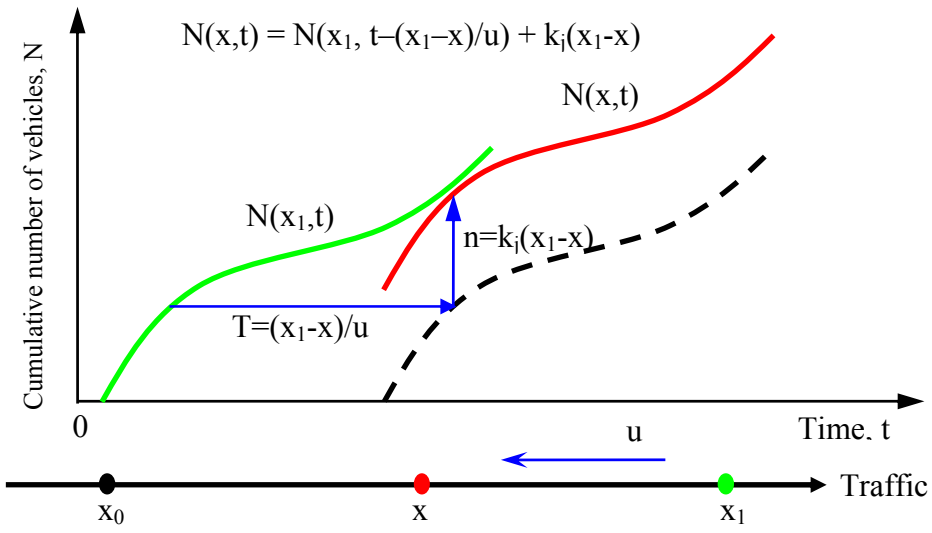


FIGURE 3-5 Backward wave propagation rule

3.3. THE COMPUTATIONAL PROCEDURE

The computational procedure of the simplified theory works as follows. A freeway is divided into a series of homogeneous sections delineated by consecutively indexed points $(x_1,...,x_{i-1},x_i,x_{i+1},...x_n)$ and the time-location diagram is partitioned into a lattice if the time increment is τ, see the top part of Figure 3-6. The algorithm works on the cumulative numbers of vehicles that arrive at and depart from point x_i by time t, $A(x_i, t)$ and $D(x_i, t)$, respectively, where i = 1, 2, …, n. For easy reference, $A(x_i, t)$ and $D(x_i, t)$ will be called the cumulative arrival and departure counts (at a time instant) or curves (over the time), respectively, thereafter. At each time increment τ, the cumulative arrival and departure curves at all locations, from the upstream end to the downstream end, are evaluated one by one, and then the system time advances one step. The above process repeats until the whole simulation period is traversed. The following procedure summarizes how the algorithm works at a lattice point (x_i,t) assuming that the

cumulative arrival and departure curves at all previous lattice points have been processed. The bottom part of Figure 3-6 sketches location x_i with an optional on- or off-ramp.

Before we proceed, it is helpful to repeat Newell's notation. As is revealed above, capital letters A and D denote cumulative arrival and departure curves, respectively, which are functions of time t and space x. Superscript "+ (plus)" and "- (minus)" mean "slightly downstream (or to the right) of x_i" and "slightly upstream (or to the left) of x_i," respectively. $A_i(t)$ denotes the net ramp flow at x_i. Q_i denotes the capacity at x_i.

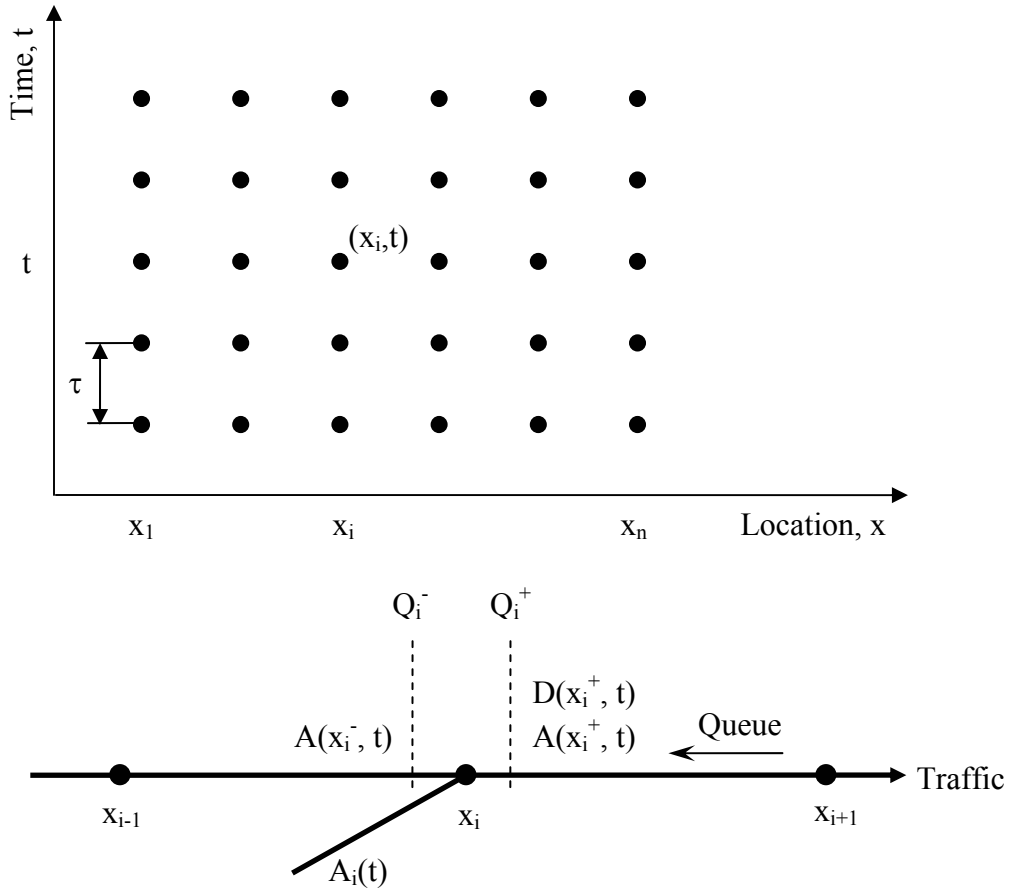


FIGURE 3-6 The computational procedure

A. Departure to the right

Note that the following discussion focuses on location x_i, so the references "to the right", "to the left", "upstream arrival", "left capacity", "right capacity", and "downstream queue" are all relative to x_i unless specifically noted otherwise. The cumulative departure curve past point x_i by time t, $D(x_i^+, t)$, is constrained by the following 4 aspects:

a. Upstream arrival

$$A(x_i^+, t) = A(x_i^-, t) + A_i(t) = D(x_{i-1}^+, t - (x_i - x_{i-1})/v_{i-1}) + A_i(t)$$

Where $A_i(t)$ is cumulative net ramp entering flow, if any, and v_{i-1} is the forward wave speed in section (x_{i-1}, x_i). Here is the place where the forward wave propagation rule applies.

b. Left capacity

$$D(x_i^+, t-\tau) + Q_i^- \tau + A_i(t) - A_i(t-\tau)$$

Where τ is time increment, and Q_i^- is the capacity to the left of x_i.

c. Right capacity

$$D(x_i^+, t-\tau) + Q_i^+ \tau$$

Where Q_i^+ is the capacity to the right of x_i.

d. Downstream Queue

$$D(x_{i+1}^{-}, t-(x_{i+1}-x_i)/u_i) + k_i(x_{i+1}-x_i)$$

Where u_i is the backward wave speed in section (x_i, x_{i+1}), and k_i is the jam density of this section. Here is the place where backward wave propagation rule applies.

The cumulative departure at x_i by t, $D(x_i^{+}, t)$, is then determined as the minimum of the above four.

B. Departure to the left

The purpose of determining cumulative departure curve to the left of x_i at time t, $D(x_i^{-}, t)$, is to facilitate finding travel time in the upstream section (x_{i-1}, x_i), which, in turn, is used to move flows of other destinations forward. The cumulative departure curve to the left of x_i by t is simply $D(x_i^{+}, t)$ minus net ramp entering flow, if any, at this lattice point, i.e., $D(x_i^{-}, t) = D(x_i^{+}, t) - A_i(t)$. If multiple destinations are considered, $D(x_i^{-}, t)$ should be labeled as $D_{i+1}(x_i^{-}, t)$, where subscription $i+1$ means "destined for x_{i+1} and beyond."

C. Section travel time

Suppose the cumulative departure curve past the right of the upstream point x_{i-1} by time t destined for x_{i+1} and beyond, $D_{i+1}(x_{i-1}^{+}, t)$, is known (from earlier steps or boundary conditions). The travel time of vehicles destined for x_{i+1} and beyond in section (x_{i-1}, x_i) can be determined as illustrated in Figure 3-7. In this figure, the bottom part sketches the highway and the cumulative departure curves with origins and destinations

shown. The above part of the figure illustrates the procedure of determining section travel time. Starting from the current time t, run a vertical line and find the intersections 1 and 2 with curves $D_{i+1}(x_i^-,t)$ and $D_{i+1}(x_{i-1}^+,t)$, respectively. Curve $D_{i+1}(x_{i-1}^+,t)$ is then traced back to some prior time t' such that $D_{i+1}(x_{i-1}^+,t')$ is equal to $D_{i+1}(x_i^-,t)$ and this determines point 3. Then distance between point 1 and 3, $T_i(t)=t-t'$, is the travel time for the vehicle bearing the "number" $D_{i+1}(x_{i-1}^+,t)$ in this section.

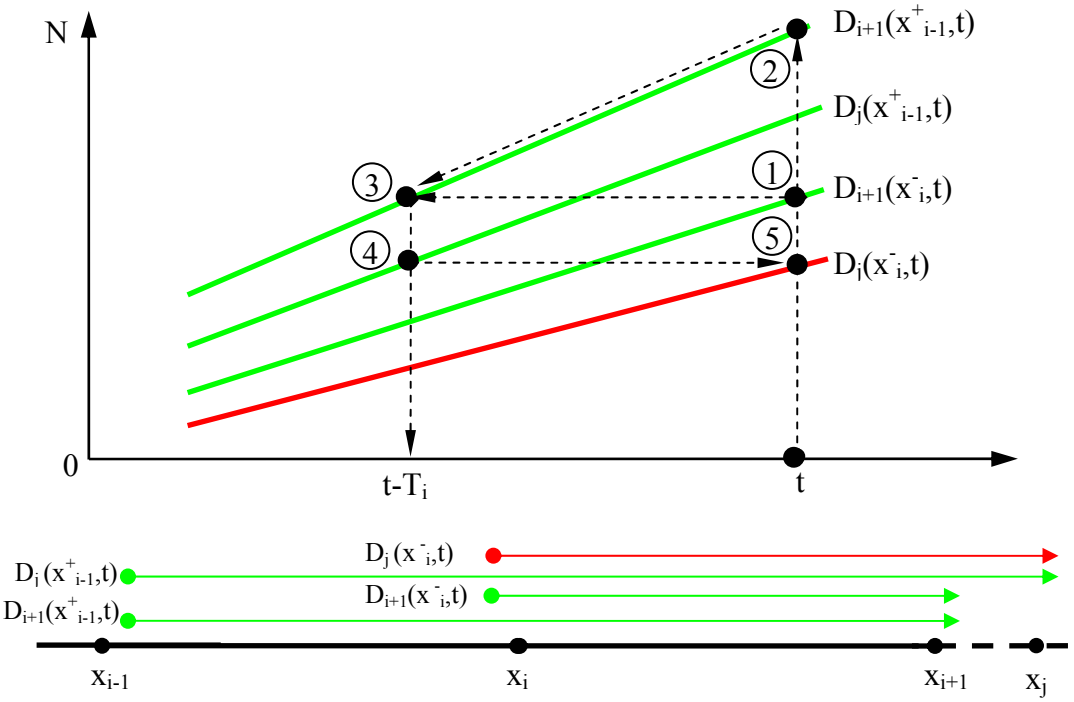


FIGURE 3-7 Section travel time and multi-destination flows

D. Departure to the left – multi-destinations

Based on Newell's assumption, all vehicles experience the same travel time in a section regardless of their destinations. Since vehicles destined for x_{i+1} and beyond traverse section (x_{i-1},x_i) in time $T_i(t)$, vehicles for other destinations, such as

$x_i, x_{i+2}, x_{i+3}, ...,$ etc., also experience the same travel time. If the cumulative departure curves past the right of x_{i-1} by t destined for other destinations $D_j(x_{i-1}^+, t)$, $(j = i+2, i+3, ...)$, are known, then cumulative departure curves past the left of x_i by t destined for other destinations, $D_j(x_i^-, t)$, is simply a translation of $D_j(x_{i-1}^+, t)$ to the right by $T_i(t)$, i.e., $D_j(x_i^-, t) = D_j(x_{i-1}^+, t - T_i(t))$. This is also illustrated in Figure 3-7. Continuing the procedure in Section Travel Time, now run a vertical line at point 3 and this line intersects curve $D_j(x_{i-1}^+, t)$ at point 4. Run a horizontal line at point 4 and this line intersects the line determined by points 1 and 2 at point 5. As time moves on, the trajectory of point 5 uniquely determines curve $D_j(x_i^-, t)$.

E. Departure to the right – multi-destinations

With $D_j(x_i^-, t)$, the cumulative departure curve past the right of x_i by t destined for other destinations, $D_j(x_i^+, t)$, is simply $D_j(x_i^-, t)$ plus net ramp entering flow, if any, at this lattice point, i.e., $D_j(x_i^+, t) = D_j(x_i^-, t) + A_j(t)$.

Now, the procedure is ready to go to the next lattice point, the order of which has been defined above.

3.4 SUMMARY

Newell formulated the simplified theory based on four major assumptions and two wave propagation rules. The first assumption postulates a triangular (or trapezoidal) flow-density relationship. The second assumption specifies merging behavior at on-ramps. The

third assumption governs diverging behavior at off-ramps. The fourth assumption provides input flow to the model. Under uncongested conditions, traffic is dominated by the wave forward propagation rule, i.e., traffic will arrive at a downstream point of a highway after a free trip time. When traffic is congested, the wave backward propagation rule applies, i.e., a downstream disturbance will be sensed by an upstream point after some time. These two rules constitute the basis on which the simplified theory is formulated.

To facilitate subsequent theoretical development, this chapter presented a discrete version of the simplified theory, i.e. , the computational procedure. The procedure keeps track of the cumulative number of vehicles that arrive at and depart from a set of pre-determined locations along a highway and the resulted arrival and departure curves are the final product of the simplified theory.

To facilitate the development of the propose extension and generalization of the simplified theory, we propose the CBWFQ merge model and the CBWS diverge model in the next chapter. By incorporating the two models, we formulate the proposed extension and generalization in Chapter 5.

CHAPTER 4

MODELING MERGING AND DIVERGING BEHAVIOR

One major limitation of the simplified theory is its restrictive assumptions on merging and diverging behavior. For merging, Newell assumed that ramp entering flow can always bypass any queue at the merge and experiences no delay. Therefore, there is actually no queuing at the on-ramp because, whenever there is any demand, it is satisfied without delay. For diverging, Newell assumed that exiting vehicles can always exit without delay as long as they have successfully arrived at the diverge, i.e., there is actually no queuing at the off-ramp, either. These assumptions are not realistic considering the facts that a queue may block traffic from both the upstream mainline and the on-ramp for the merge case and a queue may build up at an off-ramp and spill back onto the freeway mainline for the diverge case. Therefore, a good starting point to address the limitation is to establish models to deal with merging and diverging behavior of freeway traffic. In this chapter, we review literature focusing on the specific topics of merging and diverging models. We then propose a capacity-based weighted fair queuing (CBWFQ) merge model and a contribution-based weighted splitting (CBWS) diverge model, as well as their generalized forms.

4.1 THE CBWFQ MERGE MODEL

This sub-section reviews existing merge models and proposes the CBWFQ merge model. The model is then generalized to deal with a merge with multiple merging branches.

4.1.1. Review of Existing Merge Models

To facilitate future discussion, a merge is sketched below which has two upstream links and one downstream link. A merge can be a junction where an on-ramp joins a freeway, two freeways or highways come into one, or even a multi-legged intersection if properly defined. Suppose at time t, link 1 can send S_1 vehicles, link 2 can send S_2 vehicles, and the downstream link can receive R vehicles. Note that when discussing merges and diverges, S means the number of vehicles that an upstream link is able to send and R means the number of vehicles that a downstream link is able to receive. Denote d_1 the outflow (i.e., departure count) of link 1, d_2 the outflow of link 2, and d the inflow of the downstream link, where $d = d_1 + d_2$. Denote also p_1 the priority factor or splitting coefficient of link 1 and p_2 the priority factor or splitting coefficient of link 2. See the sketch of a merge in Figure 4-1.

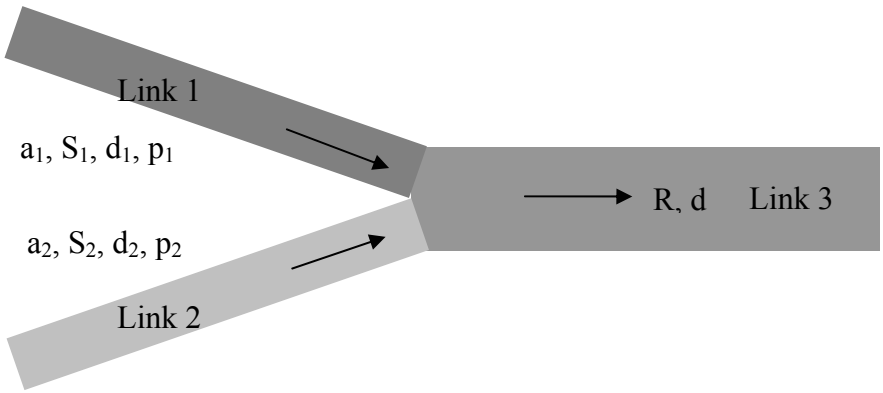


FIGURE 4-1 Sketch of a merge

4.1.1.1 Lebacque's Lane-Based Merge Model

This model (Lebacque, 1996) is based on user optimal strategy, i.e., the model tries to maximize outflows at the merging branches and sum them up to obtain the downstream inflow. The model works as follows:

$$\begin{cases} d_1 = \min\{S_1, p_1 R\} \\ d_2 = \min\{S_2, p_2 R\} \\ d = d_1 + d_2 \end{cases}$$

The splitting coefficients, defined as the ratios of maximum densities of the upstream links to that of the downstream link, often translate to the ratios of their respective number of lanes. There is no guarantee that the splitting coefficients sum up to 1.

Solution space of Lebacque's model is illustrated by the shaded area in Figure 4-2. Notice that, in addition to the upsloping solid line through the origin, the dashed lines indicate other possibilities of the splitting coefficients. If R is fixed and $S_1 + S_2 \le R$, the solution is the intersection of line 0-5 and bold line S_2-2-S_1. As one changes the splitting coefficients, line 0-5 rotates with point 0 fixed and points on bold line S_2-2-S_1 are all potential solutions. If $S_1 + S_2 > R$, the solution is the intersection of line 0-5 and bold line

S_2-1-6-4-S_1. As one changes the splitting coefficients, potential solutions lie on bold line S_2-1-6-4-S_1. The above cases hold when $p_1 + p_2 = 1$. If, however, of $p_1 + p_2 > 1$, a solution found may not be realistic because the sum of upstream outflows can be greater than downstream supply, and something has to be done to make it feasible. Nevertheless, this model is a comprehensive one and yields the largest solution space.

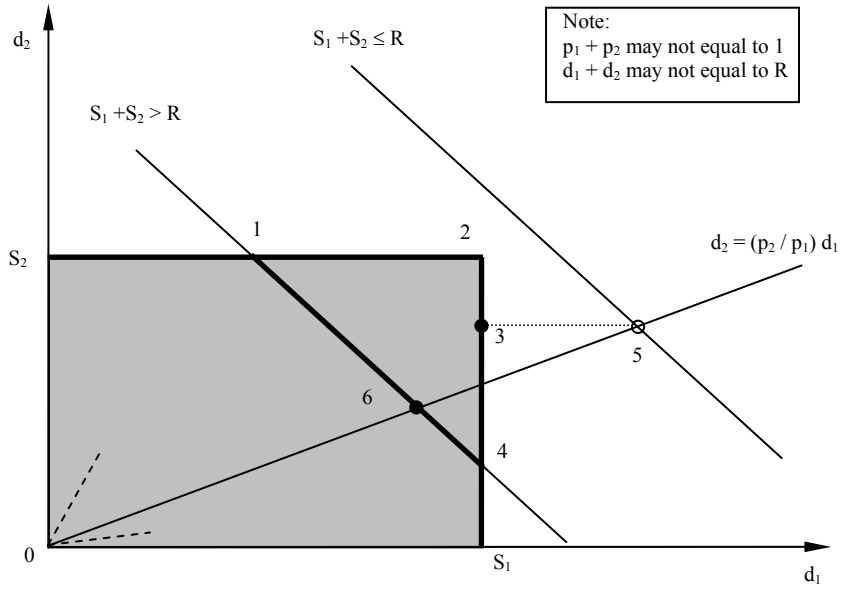


FIGURE 4-2 Lebacque's lane-based merge model

4.1.1.2 Daganzo's Priority-Based Merge Model

Unlike Lebacque's model, this model (Daganzo, 1995b) is based on a system optimal strategy, i.e., the model tries to maximize the downstream inflow while keeping upstream outflows feasible. The model works as follows:

$$\begin{cases} d_1 = S_1 \\ d_2 = S_2 \end{cases} \quad \text{if} \quad R \geq S_1 + S_2$$

$$\begin{cases} d_1 = mid\{S_1, R - S_2, p_1 R\} \\ d_2 = mid\{S_2, R - S_2, p_2 R\} \end{cases} \quad \text{if } R < S_1 + S_2$$

where the *mid* operator takes the middle value of all the members.

Again, solution space of this model is shown by the shaded area in Figure 4-3 and the dashed lines indicate the many possibilities of the priority factors. The solution under certain supply and demand conditions can be multiple and all the points in the shaded are feasible, though they might not be equally likely. To find a unique solution, some additional constraints must be provided. For example, the priority constraint assumes that flow on link 1, d_1, has higher priority than flow on link 2, d_2, i.e., $p_1 > p_2$. This eliminates virtually half of the solution space. If p_1 and p_2 are fixed, the solution space reduces to the bold line 1-2-3-4. Depending on the values of sending flows and receiving flow, unique solution can be found at point 1 if both merging branches are constrained by backward waves, or at point 3 if link 1 is constrained by forward wave and link 2 is constrained by backward wave, or at point 4 if both branches are dictated by forward waves.

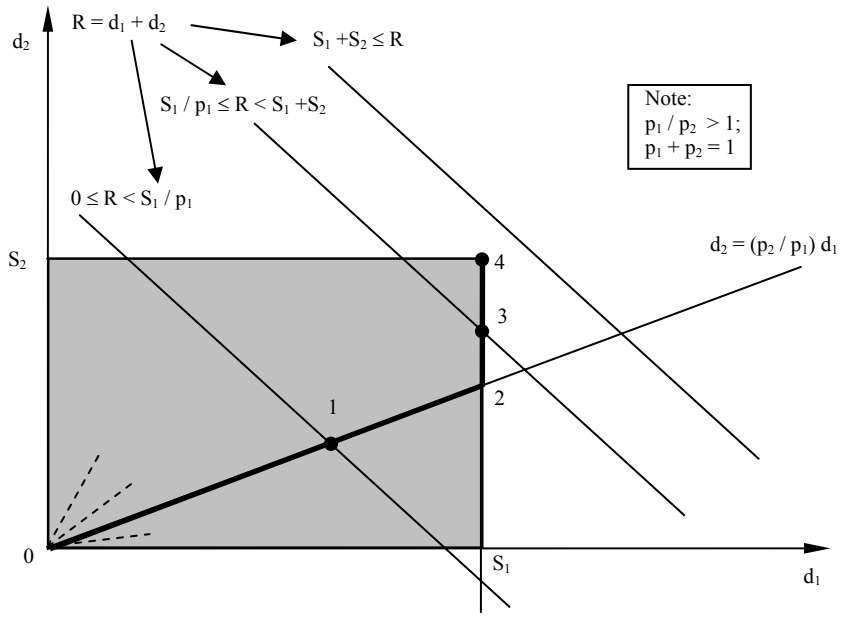


FIGURE 4-3 Daganzo's priority-based merge model

4.1.1.3 Jin and Zhang's Simplest Distribution Scheme

In an attempt to develop a special and simplest case of Daganzo's priority constraint, Jin and Zhang (2003) proposed a distribution scheme based on contributions of upstream demands, i.e.,

$$\begin{cases} d_1 = S_1 \\ d_2 = S_2 \end{cases} \text{if} \quad S_1 + S_2 < R$$

$$\begin{cases} d_1 = R \times S_1 / (S_1 + S_2) \\ d_2 = R \times S_2 / (S_1 + S_2) \end{cases} \text{if} \quad S_1 + S_2 \geq R$$

Based on this assumption the solution space of Daganzo's reduces to the bold line in Figure 4-4.

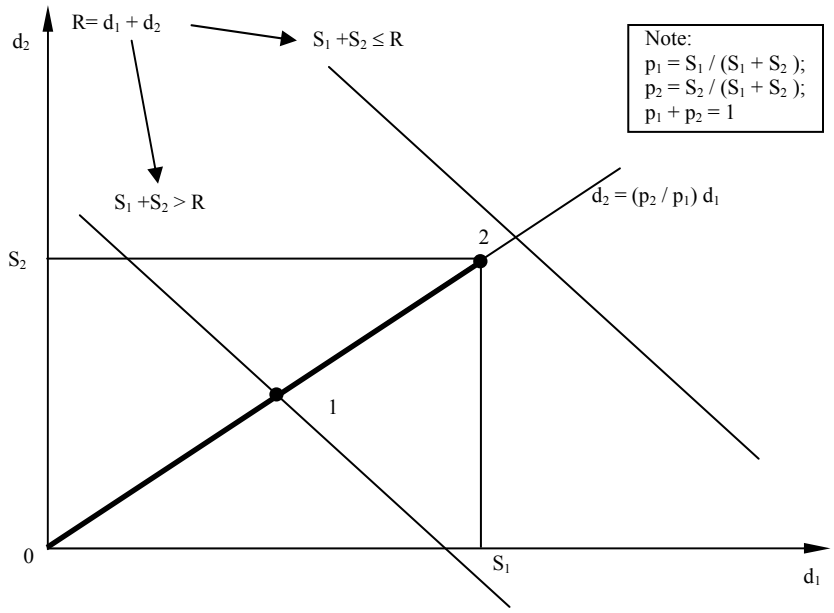


FIGURE 4-4 Jin and Zhang's demand-based merge model

Unfortunately, this model may yield unrealistic results under certain conditions. For example, if upstream mainline demand is S_1=2000, on-ramp demand is S_2=200, and downstream supply is R=2000, this distribution scheme suggests a solution of $d_1 \approx 1818$ and $d_2 \approx 182$. This implies that the on-ramp can depart only 1 vehicle at the departure of every 10 vehicles at the upstream mainline, a situation that is very rare in real life.

4.1.1.4 Newell's Simplified Merge Model

As one of his underlying assumptions of the simplified theory, Newell (1993a, 1993b, 1993c) assumed that ramp-entering vehicles can always bypass a queue, if any, and experience no delay. Unlike Daganzo's model, here the full priority is given to on-ramp traffic, i.e., $p_2 = 1$ and $p_1 = 0$. The model works as follows:

$$\begin{cases} d_2 = S_2 \\ d_1 = S_1 \end{cases} \text{ if } S_1 + S_2 \le R$$

$$\begin{cases} d_2 = S_2 \\ d_1 = R - S_2 \end{cases} \text{ if } S_1 + S_2 > R$$

With this, the solution space reduces to the bold line indicated in Figure 4-5. The solution is point 1 if link 1 is dictated by backward waves or point 2 if link 1 is dictated by forward waves. In either case, link 2 is always dictated by forward waves.

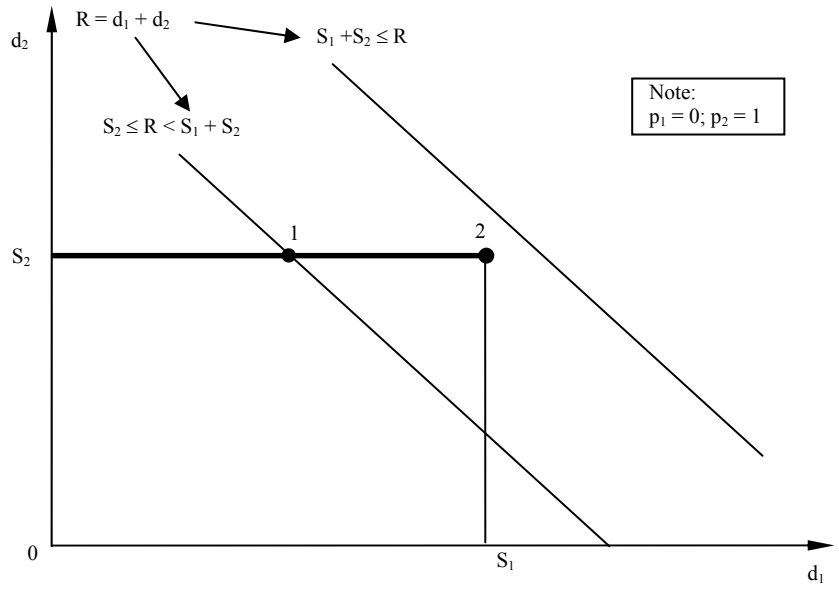


FIGURE 4-5 Newell's simplified merge model

4.1.1.5 Banks' Ramp-Metering Merge Model

As part of his effort to analyze the effect of ramp-metering in reducing traffic delay, Banks (2000) relaxed Newell's assumption on a merge such that ramp outflow is constrained by its demand S_2, capacity Q_2, and metering-rate M_2, i.e.,

$d_2 = \min\{S_2, Q_2, M_2\}$. This model is essentially the same as Newell's, i.e., full priority is still given to link 2 and backward waves never reach this branch. If we combine constraints S_2, Q_2, and M_2 into a new demand S_2', Bank's model yields exactly the same solution space as Newell's.

In summary, Lebacque's model yields larger solution space than Daganzo's due to the former's relaxation of constraint $p_1 + p_2 = 1$. Newell's model, Banks' model, and Jim's model can be considered as special cases of the first two models and the relation of these models is illustrated in Figure 4-6. As a point of interest, the CBWFQ model that is going to be proposed can also be viewed as a special case since its priority factor is based on capacity.

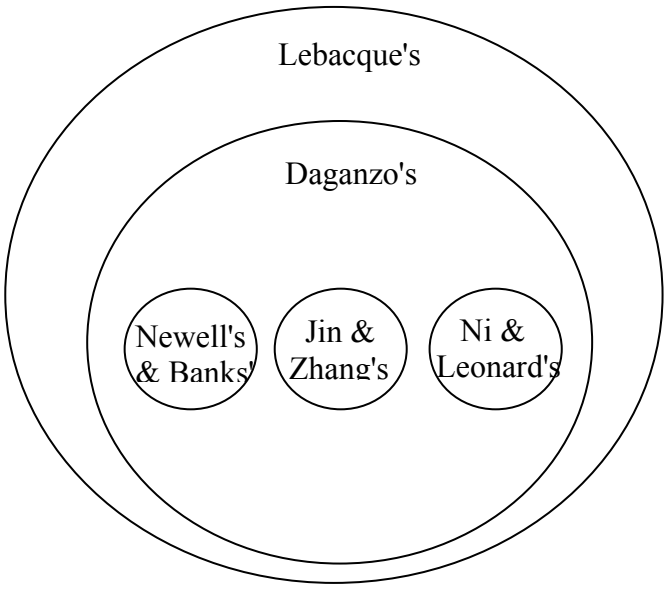


FIGURE 4-6 Relation of merge models under analysis

As can be seen from the analysis above, Lebacque's and Daganzo's models are comprehensive. It is informative to know the entire solution space, but not all solutions are feasible (for Lebacque's model), nor are feasible solutions equally likely (for both). The physically meaningful, and thus highly likely, solutions are only a small subset of the entire solution space. To define priority factors properly, additional effort of calibration, which may be costly, is required. The other three models are very simple and easy to implement and require no calibration. However, these models are subject to over-simplification and may yield unrealistic result under certain conditions. Therefore, a merge model that is realistic yet still simple to implement deserves further investigation and the CBWFQ merge model below serves such a purpose.

4.1.3. The CBWFQ Model and its Generalized Form

As part of the effort to extend and generalize the simplified theory, it is desirable for a model to deal with kinematic waves at a merge. This section proposes a CBWFQ merge model, as well as its generalized form, which preserves the advantages of the existing models while addressing their unattractive features. Figure 4-7 shows a general merge where there are α merging branches.

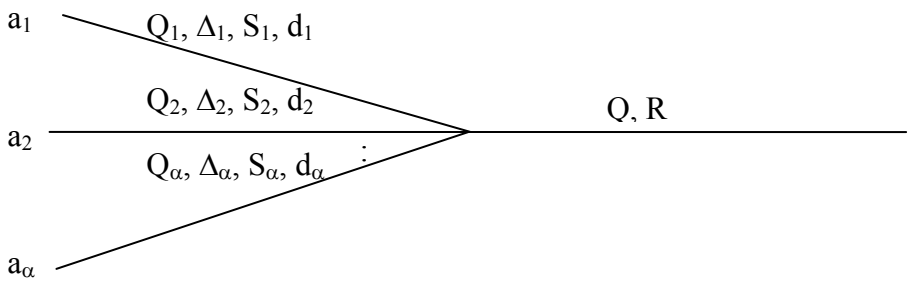


FIGURE 4-7 A general merge

Let:

a_i denotes upstream arrival of branch i, i.e., the number of vehicles that are expected to arrive at branch i based on further upstream conditions. In this and the definitions below, i = 1, 2, …, α .

Q_i denotes the capacity of the branch i and Q denotes the capacity of the downstream common link.

Δ_i denotes the fair share of downstream supply by branches i. $\Delta_i = R \times \frac{Q_i}{\sum_{j=1}^{\alpha} Q_j}$.

S_i denotes the number of vehicles that branches i is able to send. S_i takes the minimum of upstream arrival and the amount allowed by local capacity, i.e., $S_i = \min\{a_i, Q_i \tau\}$ where τ is time increment.

d_i denotes the outflow of branches i, i.e., the number of vehicles that branch i can actually send based on local capacity and upstream and downstream conditions.

R denotes the downstream supply, i.e., number of vehicles that can be received by the downstream link.

B denotes the set of branches $\{1,2,...,\alpha\}$.

The proposed CBWFQ model works with a merge with two merging branches and the model determines their outflows as illustrated in the left part of Figure 4-8, i.e., the two-dimensional (2D) case:

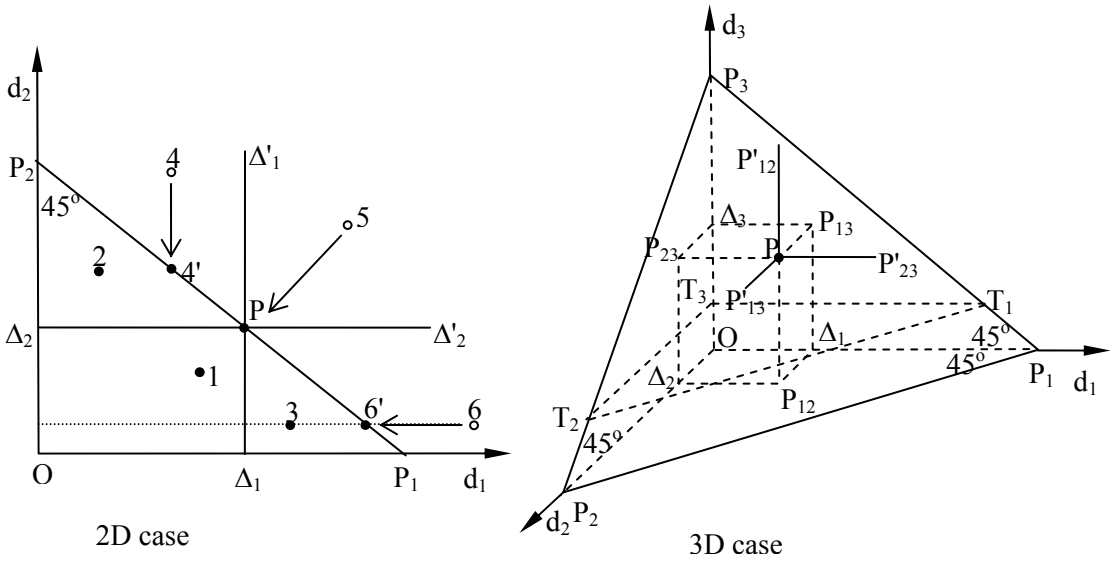


FIGURE 4-8 Generalized CBWFQ merge model

The two axes represent the outflows of the two merging branches 1 and 2. After determining the fair shares of the branches, we then construct a fair share rectangle O-Δ_1-P-Δ_2 by running two lines based on Δ_1 and Δ_2. We also construct a supply line P_1-P_2 with the coordinates of every point on the line summing up to R.

If the demands on the two branches, S_1 and S_2 are greater than the corresponding fair shares, Δ_1 and Δ_2, the demand point must fall in the half space Δ'_1-P-Δ'_2, as is the case of point 5. The solution is point P, i.e., $d_1 = \Delta_1$ and $d_2 = \Delta_2$.

Otherwise, there must be at least one branch, e.g., branch 2, whose demand S_2 is less than the corresponding fair share Δ_2. We satisfy this demand and allocate the remaining supply to the other branch. For example, if the demand is point 6, the solution is point 6', the intersection of the horizontal line through point 6 and the supply line. If the demand is point 3, the solution is this point itself. Similar treatment applies if branch 1 has more demand than its fair share, e.g., points 4 and 2.

If both branches have lower demands than their corresponding fair shares, we satisfy these demands immediately, as is the case of point 1.

The above CBWFQ model is readily extensible at no extra cost. Take the three-dimensional (3D) case for example, as shown in the right of Figure 4-8 where the three axes represent the outflows of the three branches, respectively. We construct a fair share box (a rectangular prism to be precise) O-Δ_2-P_{12}-Δ_1-Δ_3-P_{23}-P-P_{13}. and then a supply plane $P_1P_2P_3$ with the coordinates of every point on it summing up to R. Obviously, the fair share box and the supply plane contact at point P. If all three branches have higher demands than their corresponding fair shares, the demand point must fall in the half space $PP'_{12}P'_{13}P'_{23}$ pointing outward and the solution is point P. Otherwise, there must be at least one branch, say branch 3, whose demand is less than its fair share. We satisfy this demand, subtract this amount from the downstream supply, and remove this branch from further consideration. This is equivalent to cutting the picture with a horizontal plane at height S_3, shown as the dotted triangle $T_1T_2T_3$. This essentially reduces the problem to a 2D case whose solution is not repeated.

Summing up, the algorithm for the generalized CBWFQ model works as follows:

Step 1: compute the fair share of the downstream supply for each of the merging branches proportional to its capacity.

Step 2: for each merging branch, if its demand is less than or equal to its fair share, set its outflow to its demand S, subtract this amount from the downstream supply, and remove this branch from the set. Repeat this step until all merging branches have been processed. i.e.:

$$\text{IF } a_i \le \Delta_i \quad \text{THEN } \begin{cases} d_i = S_i \\ R = R - d_i \\ B = B - \{i\} \end{cases}$$

Step 3: for the remainder of the downstream supply and the remainder of the branches, repeat steps 1 and 2 until no new branch's demand is satisfied or B is empty.

Step 4: based on the remainder of the downstream supply and the remainder of the branches, recalculate the fair share of the remaining supply for each of the remaining branches and set their outflows to their fair shares.

The CBWFQ model, as well as its generalized form, merits the following advantages: (1) It deals with both forward and backward waves, i.e., it considers both upstream and downstream conditions when determining outflows; (2) It yields unique solution, so it is well-formulated; (3) The solution is physically meaningful and highly likely, so the model eliminates many unnecessary possibilities that are both unlikely and costly; (4) The model is simple to understand and easy to implement, so it makes much sense to traffic engineers; (5) The model is able to account for many factors related to

traffic operation, such as demand, supply, road geometry, capacity, ramp-metering strategies, etc., at no extra cost. (6) The model is readily extensible at no extra cost.

4.2 THE CBWS DIVERGE MODEL

This sub-section reviews existing diverge models and proposes the CBWS diverge model. The model is then generalized to deal with a diverge with multiple diverging branches.

4.2.1. Review of Existing Diverge Models

A diverge comes into play because two or more of its downstream links can affect traffic on its upstream common link. Very often, one is faced with splitting traffic between the diverging branches and a demand-supply framework is typically used to explain queuing at a diverge.

In his simplified theory, Newell (1993a, 1993b, 1993c) assumed unlimited supply on exits, so there is essentially no congestion from off-ramps. If, however, vehicles are prevented from departing at a diverge either because of insufficient capacity at the diverge point or congestion from downstream mainline, Newell assumed that all vehicles upstream will be affected regardless of their destinations, i.e., traffic state (e.g., speed, flow, density, etc.) is uniform over all lanes at the upstream link. This diverge model serves as one of the underlying assumptions based on which the simplified theory is formulated.

In one of his later papers as an attempt to estimate delays caused by an off-ramp queue, Newell (1999) slightly relaxed the above assumption and tended to believe that queues from different diverging branches need to be treated separately. This implies that

traffic states can be different for traffic bound for different destinations. However, to be consistent with his original theory and avoid going into the detail of lane-by-lane difference, Newell limited the differences in the vicinity of the diverge by assuming vertical queues. If, however, one feels that a vertical queue might over-simplify things and a physical queue makes more sense, there will be issues of more than one queue and hence more than one traffic state on a link, which is typically not captured in most macroscopic traffic models.

Daganzo (1995b) proposes a diverge model similar to Newell's first model but with finite supply at an off-ramp. The model assumes that, if either diverging branch is blocked, all upstream traffic will be restricted regardless of destination. In another paper (Daganzo 1997), Daganzo proposes a second diverge model to deal with freeways with special lanes. This model assumes two vehicle types and three traffic regimes such that, upstream of a diverge, there are "two pipes" carrying two sets of fluids with different speeds and, further upstream, there is "one pipe" carrying the mixed fluid with uniform speed. This is a closer approximation of the reality at the cost of allowing more than one traffic state on a link concurrently.

Lebacque (1996) proposes two diverge models. The first one determines the upstream demand and then divides it among the diverging branches based on turning proportions. This mode is essentially the same as Daganzo's first model, i.e., if one of the diverging branches is unable to provide any supply, no vehicle can depart at the upstream link. The second model determines the supplies at the diverging branches and sums them up to get the upstream departure. This model implies that, if there is congestion from one of the diverging branches, a hypothetical storage is needed to store the vehicles that are

unable to depart while traffic on other lanes are not affected. It also implies that traffic state on different lanes may not be the same, though the lane-by-lane difference is not explicitly modeled.

In summary, there are generally two modeling strategies at a diverge: one that allows multiple concurrent traffic states on the upstream link and the other that does not. The former treats queues from different diverging branches separately and a queue only affects its corresponding part of upstream traffic. However, this strategy may involve lane-by-lane difference and is hard to capture at a macroscopic level. This is probably part of the reason why Newell assumes vertical queue and Lebacque assumes hypothetical storage. The latter spreads downstream congestion over all lanes of the upstream link and all traffic there will be affected regardless of destinations. This is relatively easy to model but at some cost of realism. The CBWS diverge model that is going to be proposed takes advantages of the above seemingly competing strategies such that the former is used to split flow and move traffic forward at a diverge and the latter is used to update traffic state for each link.

4.2.2. Analysis of Diverging Behavior and the CBWS model

This subsection develops the CBWS diverge model based on an analysis of traffic diverging behavior. Figure 4-9 shows a sketch of a diverge. There are two types of vehicles, type 1 vehicles which are destined for the left branch (branch 1) of the diverge and type 2 vehicles which are destined for the right branch (branch 2) of the diverge. Our discussion here addresses a generic case that incorporates or can be extended to off-ramps, diverging highways or freeways, as well as intersections.

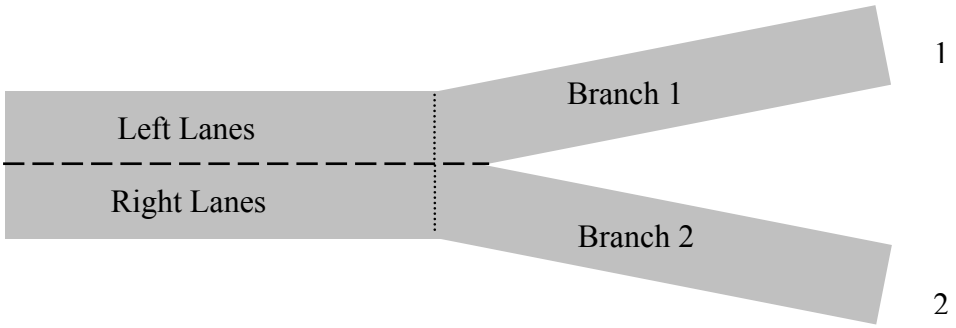


FIGURE 4-9 Sketch of a diverge

Before any queue backs up from downstream of the diverge, let us assume that traffic upstream is operating in the free-flow regime and traffic state is uniform over all lanes. Though lane variation exists, this is neither the main focus of nor typically captured at a macroscopic level. Under this assumption, there are some 1-vehicles traveling on the right lanes and some 2 vehicles traveling on the left lanes and these drivers have the piece in mind that they are able to change to their desired lanes whenever needed. When a queue backs up from downstream, it may come from branch 1 or 2 or both. Let's assume it comes from branch 2 and base our following discussion on this assumption. The same discussion also applies if the queue comes from branch 1. Several pictures are proposed to model queuing at the diverge.

Figure 4-10 illustrates the 2-vehicle queue that dictates the overall traffic condition upstream. This is the diverge model assumed in Daganzo (1995b) and also the first diverge model proposed by Lebacque (1996). As the 2-vehicle queue backs up, all vehicles upstream will be affected and delay is experienced by all vehicles regardless of their destinations. This model is more appropriate if one is willing to model traffic at a

higher level and achieves efficiency at some cost of realism. This model also makes more sense if the upstream link has only one lane. However, this model might not be realistic when the upstream link has multiple lanes.

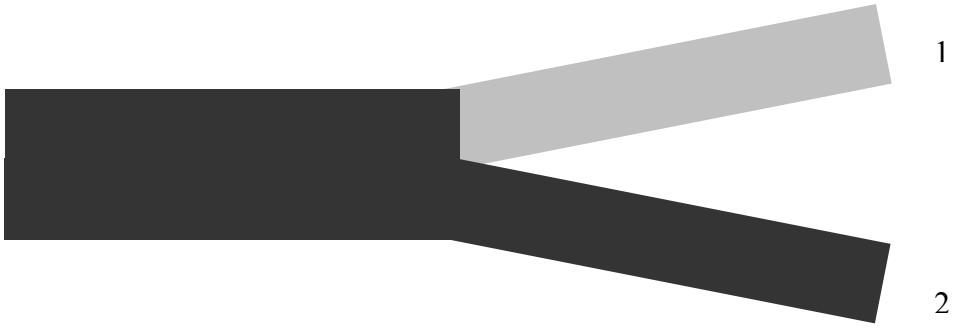


FIGURE 4-10 Diverge model 1

Newell (1996) proposed another model, shown in Figure 4-11, where the queue is assumed vertical and confined somewhere near the diverge at the side of branch 2. Upstream of the 2-vehicle queue, there could be another queue mainly formed by 1-vehicles if their arrival rate is higher than the capacity of the left lanes. Further upstream of the 1-vehicle queue, traffic states tend to be uniform over all lanes. This model is proposed primarily for evaluating delays caused by congestion at an off-ramp and the model applies to isolated exits without interaction with further upstream and downstream links, features that render the model unsuitable for extending the simplified theory to network application. Despite of this, Newell did imply that queues from different branches need to be treated differently, which is the idea adopted in the CBWS diverge model.

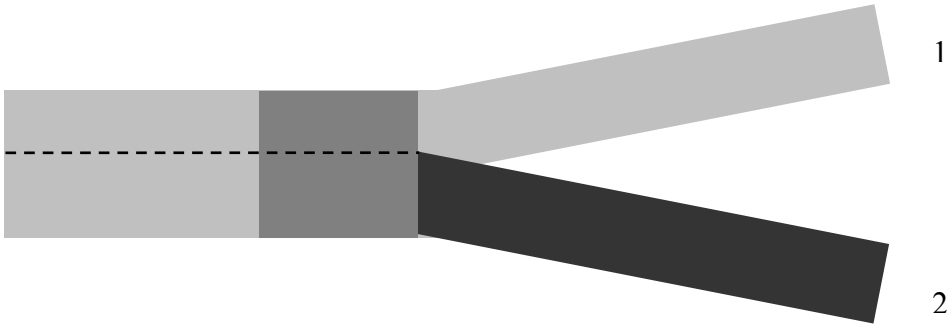


FIGURE 4-11 Diverge model 2

Figure 4-12 probably gives a more realistic representation of queuing at a diverge. As 2-vehicle queue backs onto the upstream mainline link, traffic at branch 2 is dictated by the queue and, hence, exhibits a high density. Slightly upstream of the queue, i.e., the area consisting of several right-most lanes on the mainline near the diverge, traffic state is almost the same as that of branch 2 because the former is a natural backward extension of the latter. Also the difference of traffic states between the right lanes and the left lanes becomes sharper as one gets closer to the diverge. This is especially true when origin-destination flow is predetermined and route choice is absent, as in most macroscopic models. Several reasons help to maintain the 2-vehicle queue on the right lanes. First, most 1-vehicles, noticing that the 2-vheicle queue is building up, tend to change to the left lanes not only because they are not intended to exit via branch 2 but also because traffic is able to move at higher speeds on the left lanes. Second, those who are bound to exit via branch 2 have to stay in the queue even though the adjacent lanes exhibit higher

speeds. Third, given the short distance and the high speed difference, queued vehicles that are close to the diverge may not be able to change to the left lanes even though they want to divert. However, considering that there are some 2-vehicles who bypass the queue and try to squeeze in at the head of the queue at the expense of others' delay, these vehicles can block some 1-vehicles behind and the blockage can gradually grow to encompass more lanes upstream, forming a triangular shape, as shown by the dotted lines, along which the capacity constraint moves. This “congestion diffusion” and other phenomena such as some 1-vehicles trapped in the queue who want to squeeze out can be modeled by applying a friction at the interface between the left and right lanes. This friction, in turn, acts as a time-varying constraint on the effective capacity based on the ratio of the flows destined for the two diverging branches.

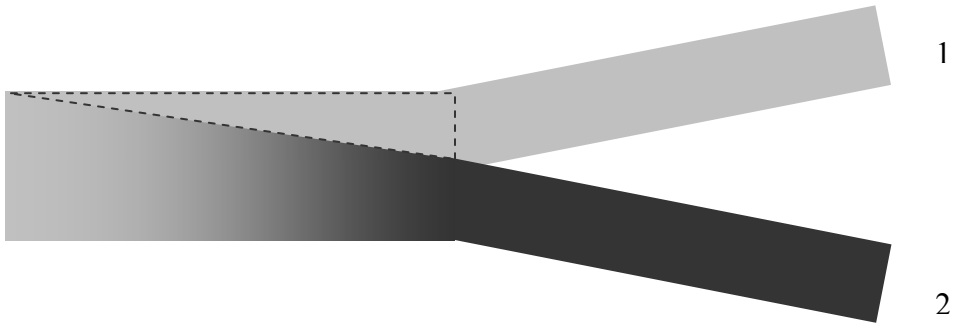


FIGURE 4-12 Diverge model 3

As one goes further upstream, the difference in traffic states between the left lanes and the right lanes may diminish due to a transition from congested to uncongested condition on the right lanes if upstream arrival is light. Traffic density in the transition area, i.e., the middle part of the mainline link, is lower then that of the congested area,

i.e., the downstream part of the mainline link, but higher than the uncongested area, i.e., the upstream part of the mainline link. Still in the transition area, downstream congestion becomes foreseeable and vehicles bound for different destinations begin to consider changing lanes with 2-vehicles changing to or staying at the right lanes even though they can travel faster at adjacent lanes. Upstream of the transition area, traffic is free-flowing (or nearly so), and downstream congestion has not been perceived by travelers here, so traffic tends to be distributed uniformly over all lanes.

Considering that, technically, it is difficult to model the state transition and congestion diffusion effects at a macroscopic level, Figure 4-13 is an approximation of Figure 4-12. In Figure 4-13, there is a discrete change of traffic states between branch 2 and the right mainline lanes and this simulates the transition effect. Depending on where the congestion diffusion effect exists, there is also an optional friction, represented as the two arrows, between the left lanes and the right lanes on the mainline near the diverge and this friction is implemented as a time-varying capacity constraint based on the ratio of the flows destined for the two branches. If there is a concurrent queue from branch 1, similar treatment applies. It is interesting to note that the proposed diverge model agrees with Daganzo's second diverge model, i.e., the one with two vehicle types and three traffic states.

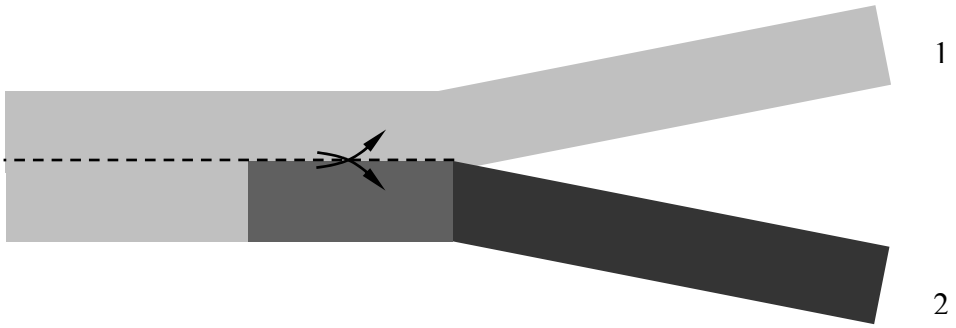


FIGURE 4-13 The proposed diverge model

Figure 4-13 is the proposed contribution-based weighted splitting (CBWS) diverge model. When splitting the number of vehicles, arrived at the diverge, between the two diverging branches, we distribute the number proportionally to the numbers of vehicles that the two branches can receive, i.e., downstream supplies or contributions. Once the cumulative departure curves, with branch 2 as their destinations, at both ends of the upstream mainline link are determined, the queue tail, i.e., where the discrete change of traffic states takes place, can be found by using the queue tail tracking procedure proposed in Son (1996). The next subsection will present the formulation of the CBWS diverge model in its generalized form.

As an interesting point, one might argue that treating queues from different branches separately violates the FIFO (first-in-first-out) assumption of a queuing system. This is not necessarily the case because FIFO still holds if the two queues from the two diverging branches are evaluated individually. On the other hand, vehicles for different destinations will operate independently once they have past the diverge and FIFO loses its meaning for them.

4.2.3 The Generalized Model

To facilitate future generalization of the simplified theory, a generalized CBWS model is formulated. Figure 4-14 sketches a general diverge with β diverging branches and one upstream common link. Let:

S denotes the upstream demand, i.e., the number of vehicles that can be sent to the diverging branches. S takes the minimum of the upstream arrival, the amount allowed by local capacity Q, and the sum of downstream supplies.

R_i denotes the supply/contribution of diverging branch i, i.e., the number of vehicles that branch i can receive, where $i = 1,2,...,\beta$

d_i denotes the outflow at the upstream link to diverging branch i, where, $i = 1,2,...,\beta$.

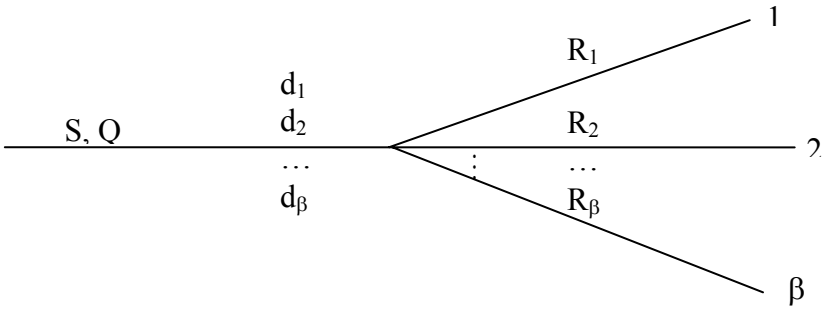


FIGURE 4-14 Generalized CBWS diverge model

The generalized CBWS diverge model determines the outflows at the upstream link destined for the diverging branches based on the following algorithm.

$$d_i = S \times \frac{R_i}{\sum_{j=1}^{\beta} R_j}, \quad i = 1,2,...,\beta$$

4.3 SUMMARY

The simplified theory assumes that ramp entering flow can always bypass any queue at a merge and experiences no delay. Therefore, there is actually no queuing at the on-ramp because, whenever there is any demand, it is satisfied without delay. As a relaxation to this assumption, we proposed a CBWFQ model, as well as its generalized form, where queuing is allowed at any of the merging branches and vehicles on these merging branches compete for downstream supply. The basic idea of this model is that every merging branch is associated with a fair share of the downstream supply. Merging branches with less-demanding traffic are first satisfied. Downstream supply is then recalculated and the fair shares of the remaining merging branches are redistributed accordingly. If a merging branch depletes its fair share, it has to wait until other less-demanding merging branches have been satisfied and then it makes use of whatever downstream supply leftover.

When dealing with diverging behavior, the simplified theory assumes that travel time of all vehicles in a freeway section is independent of their destinations. On the other hand, exiting vehicles can always exit without delay as long as they have successfully arrived at the diverge, i.e., there is actually no queuing at the off-ramp, either. As a relaxation to this assumption, we proposed a CBWS model, as well as its generalized form, where queuing is allowed at all diverging branches and queues from any of the diverging branches can back onto the upstream mainline, further constraining traffic

there. The CBWS model splits the number of vehicles that the upstream link can send among the diverging branches proportionally to the supplies/contributions of these branches.

The next chapter proposes the extension of the simplified theory, based on the CBWFQ merge model and the CBWS diverge model, and the generalization of the theory, based on the generalized CBWFQ and CBWS models.

CHAPTER 5

THE PROPOSED EXTENSION AND GENERALIZATION

As mentioned before, the major limitation of the simplified theory lies in its inability to address network traffic. With the assistance of the merge model and the diverge model proposed in the previous chapter, we extend the simplified theory to address a freeway system where multiple freeways with their on- and off-ramps are allowed. We formulate the extension based on 5 basic building blocks, i.e., an entrance, an exit, a mainline, a merge, and a diverge, and we assume that a freeway system can be represented by certain combination of these basic building blocks. With the assistance of the generalized forms of the proposed merge and diverge models, we generalize the simplified theory to address a general transportation network where a link in the network can have multiple upstream and/or downstream branches. We formulate the generalization based on a generic building block which may incorporate multiple upstream and/or downstream branches and we assume that a general transportation network can be represented by certain combination of some special cases of the generic building block.

5.1 EXPANSION OF K-WAVE DOMAIN

In addition to modeling merging and diverging behavior, several issues need to be addressed before extending and generalizing the simplified theory, among which representation of the more complicated space domain demands special attention. This section serves as a preparation for subsequent discussion.

5.1.1 K-Wave Domain

The solution space of a kinematic wave problem, here referred to as the K-Wave domain, consists of a time domain and a space domain. The original simplified theory essentially deals with a one-dimensional (1D) space domain where a freeway is typically represented by an array of nodes that are consecutively numbered/indexed. In particular, two consecutive nodes are physically adjacent and a node can have up to two adjacent nodes. On the other hand, the index also serves as the sequence based on which the nodes are processed by applying the computational procedure.

However, as we go from a freeway mainline situation to a freeway system or general transportation network, we are migrating from a 1D space domain to 2D space domain. Representation of roadway network becomes more complicated because a node may have multiple upstream nodes and/or multiple downstream nodes. In this case, the old way of representing nodes in an array is not sufficient and the concept of a *set* has to be used where the order of members in the set may or may not bear spatial relation. To establish a loose spatial relation among the nodes, nodes in a set are sorted such that all the upstream nodes of a node bear lower indices and all the downstream nodes of the node bear higher indices. This actually stipulates a spanning tree structure of roadway topology. On the

other hand, some mechanism has to be devised for each node to remember its adjacent nodes and potential destinations. By this way, it is guaranteed that, before processing a lattice point, all the necessary information has been ready been prepared.

5.1.2 O-D Flows

O-D flows in Newell's theory are quite simple due to its simple 1D space domain. However, in a general transportation network, O-D flows can be increasingly complicated and difficult to obtain when the network size grows. However, it is still reasonable to assume that entrance-exit (E-E) flows can somehow be estimated based on traffic counts from traffic surveillance systems. With a well-defined network topology and some simple synthesis, it is possible to obtain flows from each entrance to its potential destinations (E-D flows), and this is the starting point of the proposed extension and generalization. The goal of our model is to determine the cumulative arrival and departure curves at every node destined for all its potential destinations.

5.1.3 Revised Notation

Due to the expansion from 1D to 2D space domain, Newell's original notation is insufficient and the enhanced notation is presented below with the aid of a sketch of a generic building block in Figure 5-1.

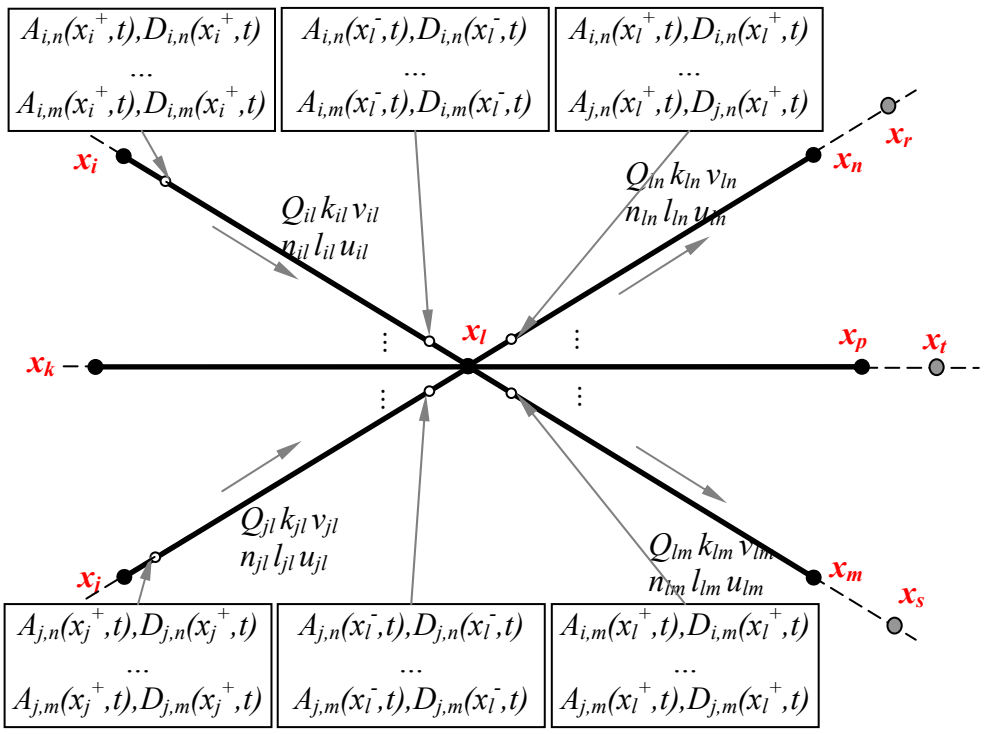


FIGURE 5-1 A generic building block of transportation network

x denotes a node and its subscription indexes the node. For example, x_l denotes the current node, $x_i,...,x_k,...,x_j$ denote its α upstream nodes, $x_n,...,x_p,...,x_m$ denote its β downstream nodes, x_r denotes any further downstream nodes of x_l via x_n, and x_s denotes any further downstream nodes of x_l via x_m. Nodes are sorted and indexed such that all potential origins of a node bear lower indices and all potential destinations of the node bear higher indices. On the other hand, a node remembers its adjacent nodes as well as its potential destinations. However, no implication is made on the relationship between x_i and x_j, nor on x_n and x_m.

A and D denote cumulative arrival and departure curves/counts, respectively. Considering that a node may have multiple upstream nodes and/or downstream nodes,

additional information, such as origin and destination, has to be supplied to clarify the A's and D's. For example, $A_{i,n}(x_l^-,t)$ denotes the cumulative arrival curve past the left of node x_l at time t originated from node x_i and destined for node x_n and beyond, and $D_{i,n}(x_l^+,t)$ denotes the cumulative departure curve past the right of node x_l at time t originated from node x_i and destined for node x_n and beyond.

Q denotes capacity. For example, Q_{il} denotes the capacity of link (x_i, x_l).

k denotes jam density. For example, k_{il} denotes the capacity of link (x_i, x_l).

v denotes forward wave propagation speed (i.e., free flow speed under the assumption of triangular or trapezoidal flow-density relationship). For example, v_{il} denotes the forward wave propagation speed of link (x_i, x_l).

u denotes backward wave propagation speed (it is a constant under the assumption of triangular or trapezoidal flow-density relationship). For example, u_{il} denotes the backward wave propagation speed of link (x_i, x_l).

n denotes number of lanes of a link. For example, n_{il} denotes the number of lanes of link (x_i, x_l). Notice that, when n appears in subscription, it means the index of a node, as defined above.

l denotes the length of a link. For example, l_{il} denotes the length of link (x_i, x_l). Again, when l appears in subscription, it means the index of a node, as defined above.

5.2 EXTENSION TO THE SIMPLIFIED THEORY

In this section, we extend the simplified theory to address a freeway system where multiple freeways and their on- and off-ramps are allowed. We formulate modeling procedures to each of the following basic building blocks: an entrance, an exit, a mainline, a merge, and a diverge. We assume that a freeway system can be represented by certain combination of these basic building blocks.

5.2.1 Entrance

In this case, we consider a node that is linked with only one downstream node and no upstream node. This case is often seen at the upstream end of a freeway or an on-ramp. Entrances serve as sources of freeway traffic such that they have infinite ability to generate traffic, but traffic is released into the freeway system at the rate specified by time-varying O-D tables.

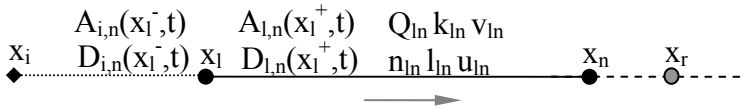


FIGURE 5-2 The sketched of an entrance

An entrance is sketched in Figure 5-2, where x_l is the entrance node, x_n is the downstream node, and x_r represents any further downstream node. Though there is nothing upstream of x_l, a dummy link (x_i, x_l) with infinity capacity is added to facilitate

subsequent discussion. A computational procedure for the entrance is formulated as follows.

A. Departure to the right

Since the discussion that follows has x_l as the current node, the references "to the left/upstream" and "to the right/downstream" are all relative to x_l unless explicitly noted otherwise. The cumulative departure curve to the right of x_l originated from x_l destined for x_n and beyond, $D_{l,n}(x_l^+, t)$ is constrained by the following:

a. Upstream arrival

Since there is no ramp flow, the cumulative arrival curve to the right of x_l originated from x_l destined for x_n and beyond, $A_{l,n}(x_l^+, t)$ is the same as $A_{i,l}(x_l^-, t)$ which is known from boundary conditions.

b. Right capacity

$$D_{l,n}(x_l^+, t-\tau) + \tau Q_{\ln}$$

c. Downstream queue

According to Newell's backward wave propagation rule, the cumulative departure curve at x_l constrained by downstream condition can be obtained by translating its downstream cumulative departure curve to the right by a backward wave propagation time and upward by a jam storage, i.e.:

$$D_{l,n}(x_n^-, t - l_{\ln} / u_{\ln}) + l_{\ln} k_{\ln}$$

d. Left capacity

Since the capacity of the dummy link upstream of node x_l is assumed to be infinity, this constraint actually does not exist.

Therefore, according to the minimum principle, $D_{l,n}(x_l^+,t)$ is the minimum of the amount determined in a, b, c, and d, i.e.:

$$D_{l,n}(x_l^+,t) = \min\left\{A_{l,n}(x_l^+,t), D_{l,n}(x_l^+,t-\tau)+\tau Q_{\ln}, D_{l,n}(x_n^-,t-l_{\ln}/u_{\ln})+l_{\ln}k_{\ln}\right\}$$

For a well-posed problem, congestion should not reach the upstream end of a freeway, i.e., $D_{l,n}(x_l^+,t)$ should not be dictated by downstream queue. Nor should a well-posed problem have input flow rates greater than capacity and causes delay for traffic waiting to enter the system. In this case, $D_{l,n}(x_l^+,t)$ is dictated by capacity constraint and the entrance itself is a possible bottleneck. Given these, a general procedure is still formulated here for completeness. In case of congestion at the entrance, only limited confidence can be given to the result and a warning should be registered to call for attention.

B. Departure to the left

Since there is no ramp traffic entering or exiting here, the cumulative departure curve to the left of x_l originated from x_i destined for x_n and beyond, $D_{i,n}(x_l^-,t)$, is the same as $D_{l,n}(x_l^+,t)$, i.e.: $D_{i,n}(x_l^-,t)=D_{l,n}(x_l^+,t)$

C. Link travel time

There is no link upstream of node x_l, but we still define a link travel time here to be consistent with other building blocks. Actually, the purpose of computing link travel time

is to come up with a mechanism to move multiple-destination flow forward because we have assumed that vehicles on the same link experience the same travel time regardless of destinations. For the same reason, if a vehicle has to wait some time before entering the system, vehicles for other destinations will experience the same delay. The waiting time is obtained by comparing curve pair $A_{i,n}(x_l^-,t)$ and $D_{i,n}(x_l^-,t)$ such that curve $A_{i,n}(x_l^-,t)$ is traced back until some earlier time t' when $A_{i,n}(x_l^-,t')$ is equal to $D_{i,n}(x_l^-,t)$, and the difference of t and t' is the waiting time experienced by the vehicle bearing "number" $D_{i,n}(x_l^-,t)$, i.e.,

$$T_{ij}(t) = t - t'$$

Graphical illustration of determining link travel time, $T_{ij}(t)$, is shown in Figure 5-3.

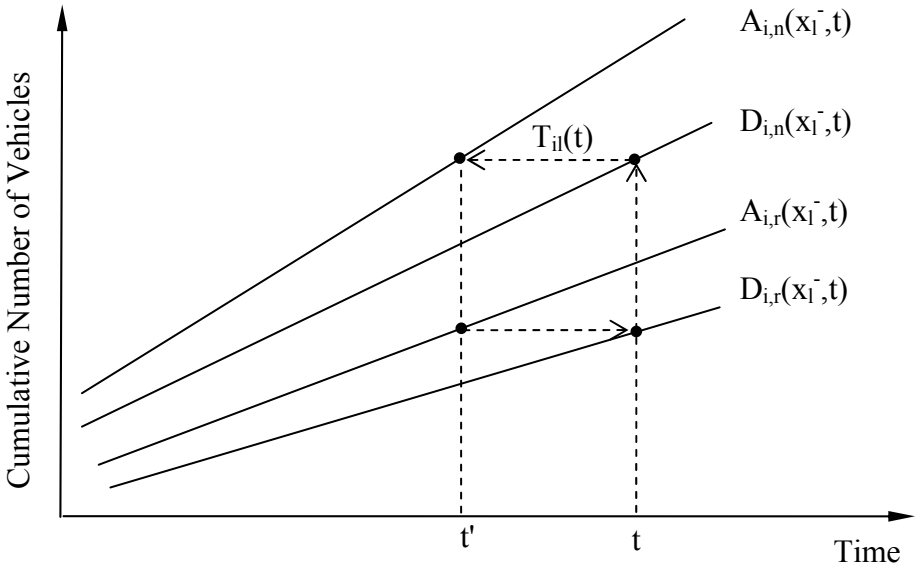


FIGURE 5-3 Solution of link travel time

D. Departure to the left – multi-destinations

As assumed, the waiting time, $T_{ij}(t)$,determined above is experienced by all vehicle regardless of destinations, the cumulative departure curves to the left of x_l originated from x_i destined for other destinations x_r and beyond, $D_{i,r}(x_l^-,t)$, can be obtained by simply translating $A_{i,r}(x_l^-,t)$ to the right by a waiting time, $T_{il}(t)$, i.e.:

$$D_{i,r}(x_l^-,t) = A_{i,r}(x_l^-,t-T_{il}(t))$$

This step is also illustrated in Figure 5-2.

E. Departure to the right – multi-destinations

Since there is no ramp flow entering or exiting here, the cumulative departure curve past the right of x_l originated from x_l destined for other destinations x_r and beyond, $D_{l,r}(x_l^+,t)$, is simply $D_{i,r}(x_l^-,t)$, i.e.:

$$D_{l,r}(x_l^+,t) = D_{i,r}(x_l^-,t)$$

This concludes the computational procedure for the entrance, and one can proceed to the next lattice point.

5.2.2 Exit

In this case, we consider a node that is linked with only one upstream node and no downstream node. This case is often seen at the downstream end of a freeway or an off-

ramp. Exits serve as sinks of freeway traffic such that they have infinite ability to absorb traffic and remove it out of the system.

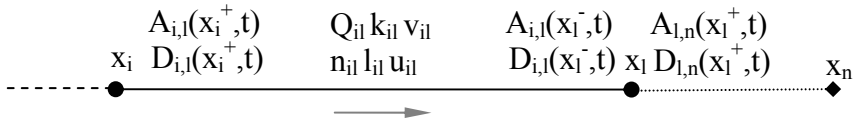


FIGURE 5-4 The sketched of an exit

An exit is sketched in Figure 5-4, where x_l is the exit node, x_i is the upstream node. Though there is nothing downstream of x_l, a dummy link (x_l, x_n) with infinity capacity is added to facilitate discussion. A computational procedure for the exit is formulated as follows.

A. Departure to the right

The cumulative departure curve to the right of x_l originated from x_l destined for x_n and beyond, $D_{l,n}(x_l^+, t)$ is constrained by the following:

a. Upstream arrival

Since there is no ramp flow, the cumulative arrival curve to the right of x_l originated from x_l destined for x_n and beyond, $A_{l,n}(x_l^+, t)$ is the same as $A_{i,l}(x_l^-, t)$ which is obtained by translating $D_{i,l}(x_i^+, t)$ horizontally to the right by a free trip time l_{il} / v_{il}, i.e.:

$$A_{l,n}(x_l^+,t)=A_{i,l}(x_l^-,t)=D_{i,l}(x_i^+,t-l_{il}/v_{il})$$

b. Right capacity

Since the capacity of the dummy link downstream of node x_l is assumed to be infinity, this constraint actually does not exist.

c. Downstream queue

A queue never builds up here because of the infinite capacity of the dummy link, so this constraint does not exist either.

d. Left capacity

$$D_{l,n}(x_l^+,t-\tau)+\tau\times Q_{il}$$

Therefore, according to the minimum principle, $D_{l,n}(x_l^+,t)$ is the minimum of the above, i.e.:

$$D_{l,n}(x_l^+,t)=\min\left\{A_{l,n}(x_l^+,t),D_{l,n}(x_l^+,t-\tau)+\tau\times Q_{il}\right\}$$

B. Departure to the left

Since there is no ramp traffic entering or exiting here, the cumulative departure curve to the left of x_l originated from x_i destined for x_l and beyond, $D_{i,l}(x_l^-,t)$, is the same as $D_{l,n}(x_l^+,t)$, i.e.:

$$D_{i,l}(x_l^-,t)=D_{l,n}(x_l^+,t)$$

C. Link travel time

Now, we are working on a real link travel time, i.e., the time needed for the vehicle bearing "number" $D_{i,l}(x_l^-,t)$ to traverse link (x_i,x_l). This is obtained by comparing curve

pair $D_{i,l}(x_i^+,t)$ and $D_{i,l}(x_l^-,t)$ such that curve $D_{i,l}(x_i^+,t)$ is traced backwards until some prior time t' when $D_{i,l}(x_i^+,t')$ is equal to $D_{i,l}(x_l^-,t)$, and the difference of t and t' is the desired link travel time, i.e.,

$$T_{ij}(t) = t - t'$$

Since x_l has no downstream, all vehicles on link (x_i, x_l) are destined for x_l. There is no need to deal with multiple-destination flows. However, it is still helpful to compute link travel time because it is useful for reporting measures of effectiveness (MOEs).

D. Departure to the left – multi-destinations

This does not apply to an exit.

E. Departure to the right – multi-destinations

This does not apply to an exit.

This concludes the computational procedure for the exit, and one can proceed to the next lattice point.

5.2.3 Mainline

In this case, we consider a node that is linked with one upstream node and one downstream node. This case is often seen at any ordinary point on freeway mainline.

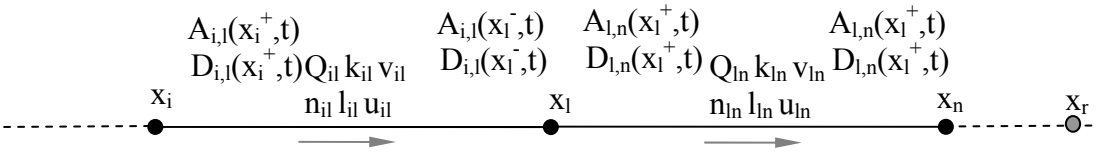


FIGURE 5-5 The sketched of a mainline

A mainline is sketched in Figure 5-5, where x_l is the mainline node, x_i is the upstream node, x_n is the downstream node, and x_r represents any further downstream node via x_n. A computational procedure for the mainline is formulated as follows.

A. Departure to the right

The cumulative departure curve to the right of x_l originated from x_l destined for x_n and beyond, $D_{l,n}(x_l^+,t)$ is constrained by the following:

a. Upstream arrival

Since there is no ramp flow here, upstream arrival is determined by applying forward wave propagation rule, i.e., traffic will arrive at a downstream location after a free trip time if there is no congestion:

$$A_{l,n}(x_l^+,t) = A_{i,l}(x_l^-,t) = D_{i,l}(x_i^+, t - l_{il}/v_{il})$$

b. Right capacity

$$D_{l,n}(x_l^+, t-\tau) + \tau Q_{\ln}$$

c. Downstream queue

Applying backward wave propagation rule, we have:

$$D_{l,n}(x_n^-, t - l_{\ln}/u_{\ln}) + l_{\ln}k_{\ln}$$

d. Left capacity

$$D_{l,n}(x_l^+, t-\tau) + \tau Q_{il}$$

Therefore, according to the minimum principle, $D_{l,n}(x_l^+, t)$ is the minimum of the above, i.e.:

$$D_{l,n}(x_l^+, t) = \min\left\{A_{l,n}(x_l^+, t), D_{l,n}(x_l^+, t-\tau) + \tau Q_{\ln}, D_{l,n}(x_n^-, t - l_{\ln}/u_{\ln}) + l_{\ln}k_{\ln}, D_{l,n}(x_l^+, t-\tau) + \tau Q_{il}\right\}$$

B. Departure to the left

Since there is no ramp traffic entering or exiting here, the cumulative departure curve to the left of x_l originated from x_i destined for x_l and beyond, $D_{i,l}(x_l^-, t)$, is the same as $D_{l,n}(x_l^+, t)$, i.e.:

$$D_{i,l}(x_l^-, t) = D_{l,n}(x_l^+, t)$$

C. Link travel time

Compare curve pair $D_{i,l}(x_i^+, t)$ and $D_{i,l}(x_l^-, t)$ in the same manner as before and get a prior time t' such that $D_{i,l}(x_i^+, t') = D_{i,l}(x_l^-, t)$. The link travel time is:

$$T_{ij}(t) = t - t'$$

D. Departure to the left – multi-destinations

As assumed, the link travel time, $T_{ij}(t)$, determined above is experienced by all vehicle regardless of destination, the cumulative departure curves to the left of x_l originated from x_i destined for other destinations x_r and beyond, $D_{i,r}(x_l^-, t)$, can be obtained by simply translating $A_{i,r}(x_l^-, t)$ to the right by a link travel time, $T_{il}(t)$, i.e.:

$$D_{i,r}(x_l^-, t) = D_{i,r}(x_i^+, t - T_{il}(t))$$

E. Departure to the right – multi-destinations

Since there is no ramp flow entering or exiting here, the cumulative departure curve past the right of x_l originated from x_l destined for other destinations x_r and beyond, $D_{l,r}(x_l^+, t)$, is simply $D_{i,r}(x_l^-, t)$, i.e.:

$$D_{l,r}(x_l^+, t) = D_{i,r}(x_l^-, t)$$

This concludes the computational procedure for the mainline, and one can proceed to the next lattice point.

5.2.4 Merge

In this case, we consider a node that is linked with two upstream nodes and one downstream node. This case is often seen at a point on a freeway where an on-ramp or a merging freeway joins.

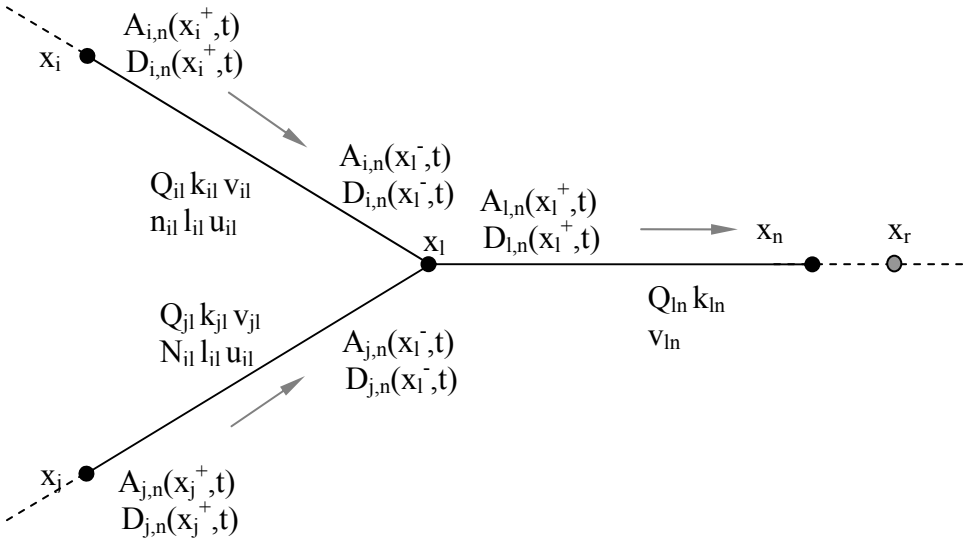


FIGURE 5-6 The sketched of a merge

A merge is sketched in Figure 5-6, where x_l is the merge node, x_i and x_j are two upstream nodes, x_n is the downstream node, and x_r represents any further downstream node via x_n. A computational procedure for the merge is formulated as follows.

A. Departure to the right

The cumulative departure curve to the right of x_l originated from x_l destined for x_n and beyond, $D_{l,n}(x_l^+,t)$, is constrained by the following:

a. Upstream arrival

There are two upstream links and upstream arrival consists of two parts. According to the forward wave propagation rule, the cumulative arrival curve to the left of x_l originated from x_i destined for x_n and beyond, $A_{i,n}(x_l^-,t)$, can be obtained by translating the cumulative departure curve to the right of x_i originated form x_i destined for x_n and

beyond, $D_{i,n}(x_i^+,t)$, by a free trip time $T_{il} = l_{il}/v_{il}$, i.e., $A_{i,n}(x_l^-,t) = D_{i,n}(x_i^+,t-l_{il}/v_{il})$. Similarly, $A_{j,n}(x_l^-,t)$ is obtained from $D_{j,n}(x_j^+,t)$, i.e., $A_{j,n}(x_l^-,t) = D_{j,n}(x_i^+,t-l_{jl}/v_{jl})$. The cumulative arrival to the right of x_l, $A_{l,n}(x_l^+,t)$, is then the sum of $A_{i,n}(x_l^-,t)$ and $A_{j,n}(x_l^-,t)$, i.e.,

$$A_{l,n}(x_l^+,t) = A_{i,n}(x_l^-,t) + A_{j,n}(x_l^-,t) = D_{i,n}(x_i^+,t-l_{il}/v_{il}) + D_{j,n}(x_j^+,t-l_{jl}/v_{jl})$$

b. Right capacity

$$D_{l,n}(x_l^+,t-\tau) + \tau Q_{\ln}$$

c. Downstream queue

$$D_{l,n}(x_n^-,t-l_{\ln}/u_{\ln}) + l_{\ln}k_{\ln}$$

d. Left capacity

The maximum number of vehicles that are allowed to depart from link (x_i,x_l) at current time step is $\tau \times Q_{il}$ and the maximum number that is allowed to depart from link (x_j,x_l) is $\tau \times Q_{jl}$. So, the left capacity constraint can be: $D_{l,n}(x_l^+,t-\tau) + \tau Q_{il} + \tau Q_{jl}$

Therefore, Cumulative departure to the right of x_l, $D_{l,n}(x_l^+,t)$, is the minimum of the above four, i.e.,

$$D_{l,n}(x_l^+,t) = \min\left\{A_{l,n}(x_l^+,t), D_{l,n}(x_l^+,t-\tau) + \tau Q_{\ln}, D_{l,n}(x_n^-,t-l_{\ln}/u_{\ln}) + l_{\ln}k_{\ln}, D_{l,n}(x_l^+,t-\tau) + \tau Q_{il} + \tau Q_{jl}\right\}$$

B. Departure to the left

Now, we are interested in knowing, of $D_{l,n}(x_l^+,t)$, how much is contributed by link (x_i,x_l) and how much by (x_j,x_l). To split $D_{l,n}(x_l^+,t)$, we apply the capacity-based weighted fair queuing (CBWFQ) merge model proposed in Chapter 4. Let:

$a_{i,n}(x_l^-,t)$ be the upstream arrival at current step to the left of x_l originated from x_i destined for x_n and beyond. $a_{i,n}(x_l^-,t) = A_{i,n}(x_l^-,t) - D_{i,n}(x_l^-,t-\tau)$.

$a_{j,n}(x_l^-,t)$ be the upstream arrival at current step to the left of x_l originated from x_j destined for x_n and beyond. $a_{j,n}(x_l^-,t) = A_{j,n}(x_l^-,t) - D_{j,n}(x_l^-,t-\tau)$.

$d_{i,n}(x_l^-,t)$ be the amount at current step that can actually depart to the left of x_l originated from x_i destined for x_n and beyond. This is an unknown variable.

$d_{j,n}(x_l^-,t)$ be the amount at current step that can actually depart to the left of x_l originated from x_j destined for x_n and beyond. This is the other unknown variable.

S_{il}^n be the amount at current step that link (x_i,x_l) can send. $S_{il}^n = \min\{a_{i,n}(x_l^-,t), \tau Q_{il}\}$.

S_{jl}^n be the amount at current step that link (x_j,x_l) can send. $S_{jl}^n = \min\{a_{j,n}(x_l^-,t), \tau Q_{jl}\}$.

$R_{\ln}$ be the supply of link (x_l,x_n) at current step. $R_{\ln} = D_{l,n}(x_l^+,t) - D_{l,n}(x_l^+,t-\tau)$.

$\Delta_{\ln}^i$ be link (x_i,x_l)'s fair share of downstream supply. $\Delta_{\ln}^i = R_{\ln} \times \dfrac{Q_{il}}{Q_{il}+Q_{jl}}$.

$\Delta_{\ln}^j$ be link (x_j,x_l)'s fair share downstream supply. $\Delta_{\ln}^j = R_{\ln} \times \dfrac{Q_{jl}}{Q_{il}+Q_{jl}}$

We can determine $d_{i,n}(x_l^-,t)$ and $d_{j,n}(x_l^-,t)$ by applying the CBWFQ model. Therefore, the cumulative departure curve to the left of x_l originated from x_i destined

for x_n and beyond, $D_{i,n}(x_l^-,t)$, and the cumulative departure curve to the left of x_l originated from x_j destined for x_n and beyond, $D_{j,n}(x_l^-,t)$, are:

$$D_{i,n}(x_l^-,t) = D_{i,n}(x_l^-,t-\tau) + d_{i,n}(x_l^-,t)$$

$$D_{j,n}(x_l^-,t) = D_{j,n}(x_l^-,t-\tau) + d_{j,n}(x_l^-,t)$$

Since this is a merge, no traffic exits. The cumulative departure curve to the left of x_l originated from x_i destined for x_l and beyond, $D_{i,l}(x_l^-,t)$, is the same as $D_{i,n}(x_l^-,t)$, and $D_{j,l}(x_l^-,t)$ is the same as $D_{j,n}(x_l^-,t)$, i.e.,

$$D_{i,l}(x_l^-,t) = D_{i,n}(x_l^-,t)$$

$$D_{j,l}(x_l^-,t) = D_{j,n}(x_l^-,t)$$

C. Link travel time

As before, link travel time is obtained by comparing the cumulative departure curves at both ends of a link. Therefore, travel time on link (x_i,x_l), $T_{il}(t)$, can be found by comparing curve pair $D_{i,l}(x_i^+,t)$ vs. $D_{i,l}(x_l^-,t)$ such that $D_{i,l}(x_i^+,t)$ is backwards until some prior time t' when $D_{i,l}(x_i^+,t')$ is equal to $D_{i,l}(x_l^-,t)$. Then $T_{il}(t) = t - t'$ is the link travel time for the vehicle bearing the "number" $D_{i,l}(x_i^+,t)$, and the link travel times for other vehicles at the same link are assumed to be the same regardless of their destinations.

In a similar fashion, travel time on link (x_j,x_l), $T_{jl}(t)$, can be found by comparing curve pair $D_{j,l}(x_j^+,t)$ vs. $D_{j,l}(x_l^-,t)$.

D. Departure to the left – multi-destinations

Based on Newell's assumption that vehicles on the same link experience the same link travel time regardless of their destinations, the cumulative departure curves to the left of x_l originated from x_i destined for other destinations x_r, $D_{i,r}(x_l^-,t)$, can be obtained by simply translating $D_{i,r}(x_i^+,t)$ to the right by a link travel time, $T_{il}(t)$, and the same applies to other multi-destination cumulative departure curves $D_{j,s}(x_l^-,t)$, i.e.,

$$D_{i,r}(x_l^-,t) = D_{i,r}(x_i^+,t-T_{il}(t))$$

$$D_{j,s}(x_l^-,t) = D_{j,s}(x_j^+,t-T_{jl}(t))$$

E. Departure to the right – multi-destinations

The cumulative departure curve past the right of x_l originated from x_l destined for other destination x_r, $D_{l,r}(x_l^+,t)$, is simply:

$$D_{l,r}(x_l^+,t) = D_{i,r}(x_l^-,t) + D_{j,r}(x_l^-,t)$$

This concludes the computational procedure for the merge, and one can proceed to the next lattice point.

5.2.5 Diverge

In this case, we consider a node that is linked with one upstream node and two downstream nodes. This case is often seen at a point on a freeway where an off-ramp or a diverging freeway leaves.

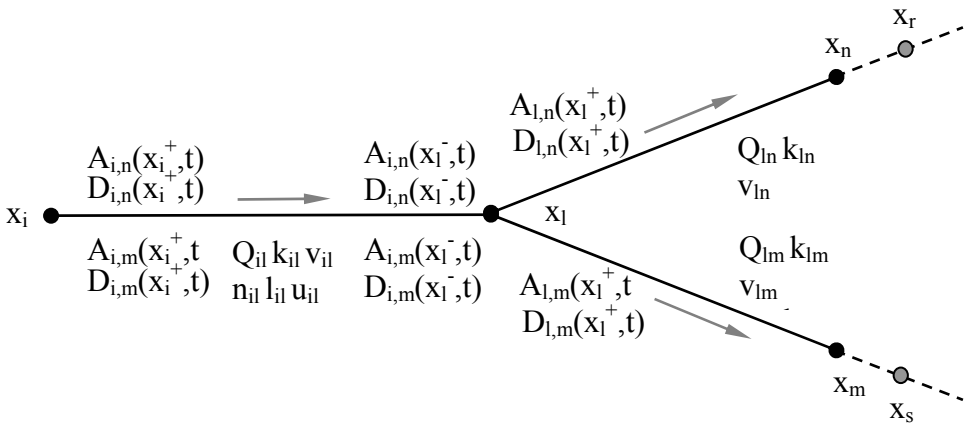


FIGURE 5-7 The sketched of a diverge

A diverge is sketched in Figure 5-7, where x_l is the diverge node, x_i is the upstream node, x_n and x_m are two downstream nodes, and x_r and x_s represent any further downstream nodes via x_n and x_m, respectively. A computational procedure for the diverge is formulated as follows.

A. Departure to the right

There are two links to the right of x_l, (x_l, x_n) and (x_l, x_m), so cumulative departure curves $D_{l,n}(x_l^+, t)$ and $D_{l,m}(x_l^+, t)$ are evaluated individually. According to Newell, the cumulative departure curve to the right of x_l originated from x_l destined for x_n and beyond, $D_{l,n}(x_l^+, t)$, is constrained by the following:

a. Upstream arrival

$$A_{l,n}(x_l^+,t) = A_{i,n}(x_l^-,t) = D_{i,n}(x_i^+,t-l_{il}/v_{il})$$

b. Right capacity

$$D_{l,n}(x_l^+,t-\tau)+\tau\times Q_{\ln}$$

c. Downstream queue

$$D_{l,n}(x_n^-,t-l_{\ln}/u_{\ln})+l_{\ln}\times k_{\ln}$$

d. Left capacity

There is a problem here. Usually the capacity to the left of x_l is enough to handle traffic destined for x_n and beyond. However, this capacity is, at the same time, shared by traffic destined for x_m and beyond. The question is, how much of the capacity can be utilized by the former? It is hard to answer at this point and let us leave it for a second. For now, $D_{l,n}(x_l^+,t)$ is simply the minimum of a, b, and c, i.e.,

$$D_{l,n}(x_l^+,t) = \min\left\{A_{l,n}(x_l^+,t), D_{l,n}(x_l^+,t-\tau)+\tau\times Q_{\ln}, D_{l,n}(x_n^-,t-l_{\ln}/u_{\ln})+l_{\ln}\times k_{\ln}\right\}$$

Similarly, we can obtain the cumulative departure curve to the right of x_l originated from x_l destined for x_m and beyond, $D_{l,m}(x_l^+,t)$:

$$D_{l,m}(x_l^+,t) = \min\left\{A_{l,m}(x_l^+,t), D_{l,m}(x_l^+,t-\tau)+\tau\times Q_{lm}, D_{l,m}(x_m^-,t-l_{lm}/u_{lm})+l_{lm}\times k_{lm}\right\}$$

B. Departure to the left

The cumulative departure curve to the left of x_l originated from x_i destined for x_l and beyond, $D_{i,l}(x_l^-,t)$, is simply the minimum of:

a. Upstream arrival

$$A_{i,l}(x_l^-, t) = D_{i,l}(x_i^+, t - l_{il} / v_{il})$$

b. Left capacity

$$D_{i,l}(x_l^-, t - \tau) + \tau \times Q_{il}$$

Recall that, in Chapter 4, we proposed a CBWS model and suggested an optional friction to address the congestion diffusion effect, i.e., some drivers may want to squeeze in at the head of a queue and this behavior can make the congestion gradually spread to the entire mainline lanes. We also suggested that the friction is a function of the ratio of the flows destined for the diverging branches. Let $f_{il}(\phi(t))$ denote the friction coefficient and $\phi(t)$ denote the time-varying flow ratio. $f_{il}(\phi(t))$ is typically unknown and has to be treated as a design parameter. The effect of a friction on traffic operation is the reduction of effective capacity and the effective capacity is computed as $Q_{il}' = Q_{il}(1 - f_{il}(\phi(t)))$. Now the capacity constraint becomes $D_{i,l}(x_l^-, t - \tau) + \tau \times Q_{il}'$.

c. Downstream departure

$$D_{l,n}(x_l^+, t) + D_{l,m}(x_l^+, t)$$

Notice that the destination of $D_{i,l}(x_l^-, t)$ is x_l, not x_n or x_m. We are considering the aggregate flow at link (x_i, x_l).

Sum up, $D_{i,l}(x_l^-, t)$ is the minimum of a, b, and c, i.e.,

$$D_{i,l}(x_l^-, t) = \min\left\{A_{i,l}(x_l^-, t), D_{i,l}(x_l^-, t - \tau) + \tau \times Q_{il}', D_{l,n}(x_l^+, t) + D_{l,m}(x_l^+, t)\right\}$$

In response to the problem of left capacity in A, this step guarantees that the cumulative departure destined for x_l (i.e., the sum of those destined for x_n and x_m) will not exceed the capacity to the left of x_l.

Now, a new question arises. Of the amount $D_{i,l}(x_l^-,t)$ determined above, how much is destined for x_n, i.e. $D_{i,n}(x_l^-,t)$, and how much is destined for x_m, i.e., $D_{i,m}(x_l^-,t)$? This involves splitting $D_{i,l}(x_l^-,t)$ among the two diverging branches and the CBWS model solves the problem. Let:

S denote the upstream supply. $S = D_{i,l}(x_l^-,t) - D_{i,l}(x_l^-,t-\tau)$.

Downstream supplies: $R_1 = D_{l,n}(x_l^+,t) - D_{l,n}(x_l^+,t-\tau)$ and $R_2 = D_{l,m}(x_l^+,t) - D_{l,m}(x_l^+,t-\tau)$

The CBWS model computes outflows d_1 and d_2, so

$$D_{i,n}(x_l^-,t) = D_{i,n}(x_l^-,t-\tau) + d_1$$

$$D_{i,m}(x_l^-,t) = D_{i,m}(x_l^-,t-\tau) + d_2$$

C. Link travel time

At previous step, we determined the cumulative departure curves based on aggregate flow for the upstream link. At this step, we evaluate the queues for the two diverging branches individually. This is done by computing link travel time for vehicles destined for each diverging branch.

Travel time for 1-vehicles at link (x_i,x_l) can be obtained by comparing curve pair $D_{i,n}(x_i^+,t)$ and $D_{i,n}(x_l^-,t)$ to find some prior time t' such that $D_{i,n}(x_i^+,t')$ is equal

to $D_{i,n}(x_l^-,t)$. Then the link travel time for 1-vehicles is $T_{il}^n(t)=t-t'$, where the superscript n means "destined for x_n and beyond".

Similarly, travel time for 2-vehices at link (x_i,x_l), $T_{il}^m(t)$, can be obtained by curve pair $D_{i,m}(x_i^+,t)$ and $D_{i,m}(x_l^-,t)$.

D. Departure to the left – multi-destinations

With link travel time $T_{il}^n(t)$ obtained above, the cumulative departure curve to the left of x_l originated from x_i destined for x_r and beyond, $D_{i,r}(x_l^-,t)$, is determined as

$$D_{i,r}(x_l^-,t)=D_{i,r}(x_i^+,t-T_{il}^n(t))$$

Similarly, the cumulative departure curve to the left of x_l originated from x_i destined for x_s and beyond, $D_{i,s}(x_l^-,t)$ is determined as

$$D_{i,s}(x_l^-,t)=D_{i,s}(x_i^+,t-T_{il}^m(t))$$

E. Departure to the right – multi-destinations

Since this is a diverge, no traffic enters from any on-ramp. The cumulative departure curve to the right of x_l originated from x_i destined for x_n and beyond, $D_{i,n}(x_l^+,t)$ is the same as $D_{i,n}(x_l^-,t)$, and the cumulative departure curve to the right of x_l originated from x_i destined for x_m and beyond, $D_{i,m}(x_l^+,t)$ is the same as $D_{i,m}(x_l^-,t)$, i.e.,

$$D_{i,n}(x_l^+,t)=D_{i,n}(x_l^-,t)$$

$$D_{i,m}(x_l^+,t)=D_{i,m}(x_l^-,t)$$

Notice that, at step A, we have preliminarily determined $D_{l,n}(x_l^+,t)$ and $D_{l,m}(x_l^+,t)$, which are the equivalent of $D_{i,n}(x_l^+,t)$ and $D_{i,m}(x_l^+,t)$, respectively. As the procedure goes on, those preliminary values are fine-tuned and updated.

Similarly, for other destinations x_r and x_s, we have:

$$D_{i,r}(x_l^+,t) = D_{i,r}(x_l^-,t)$$

$$D_{i,s}(x_l^+,t) = D_{i,s}(x_l^-,t)$$

This concludes the computational procedure for the diverge, and one can proceed to the next lattice point.

5.3 GENERALIZATION OF THE SIMPLIFIED THEORY

In this section, we generalize the simplified theory to address a general transportation network where a link in the network can have multiple upstream and/or downstream branches. We formulate the generalization based on a generic building block which may incorporate multiple upstream and/or downstream branches and we assume that a general transportation network can be represented by certain combination of some special cases of the generic building block.

The generic building block is sketched in Figure 5-1, where x_l denotes the current node, $x_i,...,x_k,...,x_j$ denote its α upstream nodes, $x_n,...,x_p,...,x_m$ denote its β downstream nodes, x_r denotes any further downstream nodes of x_l via x_n, and x_s denotes any further downstream nodes of x_l via x_m. Nodes are sorted and indexed such that all potential origins of a node bear lower indices and all potential destinations of the node bear higher

indices. On the other hand, a node remembers its adjacent nodes as well as its potential destinations. However, no implication is made on the relationship between x_i and x_j, nor on x_n and x_m. A computational procedure for the general case is formulated as follows.

A. Departure to the right

Goal: determining the β downstream cumulative departure curves:

$$D_{l,n}(x_l^+,t),\ \ldots,\ D_{l,m}(x_l^+,t)$$

Since there are β links to the right of x_l, $\{(x_l,x_n),\ldots,\ (x_l,x_m)\}$, cumulative departure curves $D_{l,n}(x_l^+,t),\ \ldots,\ D_{l,m}(x_l^+,t)$ are evaluated individually. Take one of them for example, the cumulative departure curve to the right of x_l originated from x_l destined for x_n and beyond, $D_{l,n}(x_l^+,t)$, is constrained by the following:

a. Upstream arrival

According to the forward wave propagation rule, the cumulative arrival curve to the left of x_l originated from x_i destined for x_n and beyond, $A_{i,n}(x_l^-,t)$, can be obtained by translating the cumulative departure curve to the right of x_i originated form x_i destined for x_n and beyond, $D_{i,n}(x_i^+,t)$, by a free link travel time $T_{il}=l_{il}/v_{il}$. The same applies to the remaining $(\alpha-1)$ upstream links, and the cumulative arrival to the right of x_l originated from x_l destined for x_n and beyond, $A_{l,n}(x_l^+,t)$, is the sum of those of the α upstream links, i.e.,

$$A_{l,n}(x_l^+,t) = A_{i,n}(x_l^-,t) + ... + A_{j,n}(x_l^-,t) = D_{i,n}(x_i^+,t - l_{il}/v_{il}) + ... + D_{j,n}(x_j^+,t - l_{jl}/v_{jl})$$

b. Right capacity

$$D_{l,n}(x_l^+,t-\tau) + \tau Q_{\ln}$$

c. Downstream queue

$$D_{l,n}(x_n^-,t - l_{\ln}/u_{\ln}) + l_{\ln}k_{\ln}$$

d. Left capacity

The existence of multiple upstream links makes the left capacity constraint complicated, but it will be easier to take care of this constraint later on when we determine cumulative departures to the left of x_l. For now, $D_{l,n}(x_l^+,t)$ is simply the minimum of I, II, and III, i.e.,

$$D_{l,n}(x_l^+,t) = \min\{A_{l,n}(x_l^+,t), D_{l,n}(x_l^+,t-\tau) + \tau Q_{\ln}, D_{l,n}(x_n^-,t - l_{\ln}/u_{\ln}) + l_{\ln}k_{\ln}\}$$

Similarly, the cumulative departure curves to the right of x_l originated from x_l destined for the remaining $\beta - 1$ nodes and beyond can be determined, i.e.,

…

$$D_{l,m}(x_l^+,t) = \min\{A_{l,m}(x_l^+,t), D_{l,m}(x_l^+,t-\tau) + \tau Q_{lm}, D_{l,m}(x_m^-,t - l_{lm}/u_{lm}) + l_{lm}k_{lm}\}$$

B. Departure to the Left

Goal: determining the α upstream cumulative departure curves:

$$D_{i,l}(x_l^-,t),\ \ldots,\ D_{j,l}(x_l^-,t)$$

Since there are α links to the left of x_l, $\{(x_i, x_l), \ldots, (x_j, x_l)\}$, cumulative departure curves $D_{i,l}(x_l^-, t)$, …, $D_{j,l}(x_l^-, t)$ are evaluated individually. Take one of them for example, the cumulative departure curve to the left of x_l originated from x_i destined for x_l and beyond, $D_{i,l}(x_l^-, t)$, is simply the minimum of:

a. Upstream arrival

$$A_{i,l}(x_l^-, t) = D_{i,l}(x_i^+, t - l_{il} / v_{il})$$

b. Left capacity

$$D_{i,l}(x_l^-, t - \tau) + \tau \times Q_{il}$$

c. Downstream departure

$$D_{i,n}(x_l^+, t) + \ldots + D_{i,m}(x_l^+, t)$$

There are β terms in the downstream departure constraint and each of them represents the contribution of a downstream link to the cumulative departure curve originated from x_i. Notice that none of the β terms is known in advance, so they have to be computed from the known.

Up to now, the cumulative departure curve to the right of x_l originated from x_l destined for x_n and beyond, $D_{l,n}(x_l^+, t)$, is known and it is the sum of cumulative departure curves to the right of x_l originated from all α upstream nodes, i.e.,

$$D_{l,n}(x_l^+, t) = D_{i,n}(x_l^+, t) + \ldots + D_{j,n}(x_l^+, t)$$

None of the α terms on the right hand side is known in advance either, so they have to be computed first. This is equivalent to distributing a downstream departure/supply

among its α upstream merging branches, which can be accomplished by applying the generalized capacity-based weighted fair queuing (CBWFQ) merge model with the following relations in mind:

Notation in CBWFQ	Corresponds to	Notation in K-Waves
R	$\rightarrow$	$R_{\ln} = D_{l,n}(x_l^+, t) - D_{l,n}(x_l^+, t-\tau)$
a_1	$\rightarrow$	$a_{i,n}(x_l^-, t) = A_{i,n}(x_l^-, t) - D_{i,n}(x_l^-, t-\tau)$
	……	
a_α	$\rightarrow$	$a_{j,n}(x_l^-, t) = A_{j,n}(x_l^-, t) - D_{j,n}(x_l^-, t-\tau)$
Δ_1	$\rightarrow$	$\Delta_{\ln}^i = R_{\ln} \times \frac{Q_{il}}{Q_{il} + Q_{jl}}$
	……	
Δ_α	$\rightarrow$	$\Delta_{\ln}^j = R_{\ln} \times \frac{Q_{jl}}{Q_{il} + Q_{jl}}$
d_1	$\rightarrow$	$d_{i,n}(x_l^+, t)$
	……	
d_α	$\rightarrow$	$d_{j,n}(x_l^+, t)$

Continue applying the generalized CBWFQ model to the remaining $\beta - 1$ downstream links $\{...,(x_l, x_p),...,(x_l, x_m)\}$, and we eventually obtain the following (1) $\alpha \times \beta$ matrix whose elements are current departure counts to the right of x_l:

$$d_{i,n}(x_l^+, t),..., d_{k,n}(x_l^+, t), ... d_{j,n}(x_l^+, t)$$

…

$$d_{i,p}(x_l^+, t),..., d_{k,p}(x_l^+, t), ... d_{j,p}(x_l^+, t)$$

…

$$d_{i,m}(x_l^+, t),..., d_{k,m}(x_l^+, t), ... d_{j,m}(x_l^+, t)$$

Adding the cumulative departure counts of the last step to the corresponding elements in the above matrix, we get a new $\alpha \times \beta$ matrix whose elements are cumulative departure counts to the right of x_l:

$$D_{i,n}(x_l^+,t),\ldots,\ D_{k,n}(x_l^+,t),\ \ldots D_{j,n}(x_l^+,t)$$

…

$$D_{i,p}(x_l^+,t),\ldots,\ D_{k,p}(x_l^+,t),\ \ldots D_{j,p}(x_l^+,t) \qquad \text{(ii)}$$

…

$$D_{i,m}(x_l^+,t),\ldots,\ D_{k,m}(x_l^+,t),\ \ldots D_{j,m}(x_l^+,t)$$

Notice that the first row of the matrix sums up to $D_{l,n}(x_l^+,t)$ and the first column sums up to constraint III (Downstream departure). Therefore, back to our question before, $D_{i,l}(x_l^-,t)$ is determined as:

$$D_{i,l}(x_l^-,t) = \min\left\{A_{i,l}(x_l^-,t), D_{i,l}(x_l^-,t-\tau)+\tau \times Q_{il}, D_{i,n}(x_l^+,t)+\ldots+D_{i,m}(x_l^+,t)\right\}$$

In a similar way, departures to the left of x_l for other upstream links are:

……

$$D_{j,l}(x_l^-,t) = \min\left\{A_{j,l}(x_l^-,t), D_{j,l}(x_l^-,t-\tau)+\tau \times Q_{jl}, D_{j,n}(x_l^+,t)+\ldots+D_{j,m}(x_l^+,t)\right\}$$

If we have a matrix of current departure to the left of x_l stratified by origins and destinations, as shown in matrix (iii), it is interesting to know whether it is equal to matrix (i).

$$d_{i,n}(x_l^-,t),\ldots,\ d_{k,n}(x_l^-,t),\ \ldots d_{j,n}(x_l^-,t)$$

…

$$d_{i,p}(x_l^-,t),\ldots,\ d_{k,p}(x_l^-,t),\ \ldots d_{j,p}(x_l^-,t) \qquad \text{(iii)}$$

…

$$d_{i,m}(x_l^-,t),\ldots,\ d_{k,m}(x_l^-,t),\ \ldots d_{j,m}(x_l^-,t)$$

When all the departures to the left are dictated by downstream departure (constraint B.III), the two matrices might be equal. Otherwise, there should be some differences between them and this is especially true when some departures to the left are dictated by their respective capacities. So, something has to be done to update matrix (i) based on the latest information we get, i.e., the departures to the left. This is equivalent to splitting an upstream departure among several diverging branches, which can be accomplished by applying a generalized contribution-based weighted splitting (CBWS) diverge model to determine the first column of matrix III using $D_{i,l}(x_l^-,t)$ with the following relations in mind:

Notation in CBWFQ	Corresponds to	Notation in K-Waves
S	$\rightarrow$	$D_{i,l}(x_l^-,t)$ - $D_{i,l}(x_l^-,t-\tau)$
$R_1,\ldots,R_\beta$	$\rightarrow$	$d_{i,n}(x_l^+,t),\ldots\ldots d_{j,n}(x_l^+,t)$, respectively
$d_1,d_2,\ldots d_\beta$	$\rightarrow$	$d_{i,n}(x_l^-,t),\ldots\ldots d_{j,n}(x_l^-,t)$, respectively

Similarly, other columns of matrix (iii) can be determined and the whole matrix is known. Matrix (i) is then set equal to matrix (iii) and matrix (ii) can be updated accordingly.

C. Link travel time

Goal: determining link travel times at each of the α upstream links:

$$T_{il}^{n}(t),\ldots,T_{il}^{m}(t)$$

……

$$T_{jl}^{n}(t),\ldots,T_{jl}^{m}(t)$$

On link (x_i,x_l), travel times for vehicles destined for different destinations may not be the same, so link travel times for different destinations are computed separately. For example, link travel time for traffic destined for x_n and beyond is determined by comparing curve pair $D_{i,n}(x_i^+,t)$ vs. $D_{i,n}(x_l^-,t)$ such that $D_{i,n}(x_i^+,t)$ is traced back to an earlier time t' when $D_{i,n}(x_i^+,t')=D_{i,n}(x_l^-,t)$. Then, link travel time for the above traffic is $T_{il}^{n}(t)=t-t'$. Similarly, we get the link travel times on link (x_i,x_l) for other destinations:

$$..., T_{il}^{m}(t).$$

In the same fashion, travel times on other links can be determined:

……

$$T_{jl}^{n}(t),\ldots,T_{jl}^{m}(t)$$

D. Departure to the left – multi-destinations

Goal: determining the α sets of upstream cumulative departure curves for further destinations:

$$D_{i,r}(x_l^-,t),\ \ldots,\ D_{i,s}(x_l^-,t)$$

……

$$D_{j,r}(x_l^-,t),\ \ldots,\ D_{j,s}(x_l^-,t)$$

The cumulative departure curves to the left of x_l originated from x_i destined for x_r and beyond, $D_{i,r}(x_l^-,t)$, can be obtained by simply translating $D_{i,r}(x_i^+,t)$ to the right by a link travel time, $T_{il}^n(t)$, and similar processing applies to other multi-destinations cumulative departure curves , i.e.,

$$D_{i,r}(x_l^-,t) = D_{i,r}(x_i^+,t-T_{il}^n(t)),\ \ldots,\ D_{i,s}(x_l^-,t) = D_{i,s}(x_i^+,t-T_{il}^m(t))$$

……

$$D_{j,r}(x_l^-,t) = D_{j,r}(x_j^+,t-T_{jl}^n(t)),\ldots,D_{j,s}(x_l^-,t) = D_{j,s}(x_j^+,t-T_{jl}^m(t))$$

E. Departure to the right – multi-destinations

Goal: determining the β downstream cumulative departure curves for further destinations:

$$D_{l,r}(x_l^+,t),\ \ldots,D_{l,s}(x_l^+,t)$$

The cumulative departure curve past the right of x_l originated from x_l destined for x_r and beyond, $D_{l,r}(x_l^+,t)$, is simply the sum of cumulative departure curves past the left of x_l originated from all upstream nodes destined for x_r and beyond. Similarly, we obtain other downstream cumulative departure curves, i.e.,

$$D_{l,r}(x_l^+,t) = D_{i,r}(x_l^-,t)+\ldots+D_{j,r}(x_l^-,t)$$

…

$$D_{l,s}(x_l^+,t) = D_{i,s}(x_l^-,t)+\ldots+D_{j,s}(x_l^-,t)$$

This concludes processing a lattice point of a general network, and the algorithm is ready to proceed.

5.4 SUMMARY

Based on the computational procedure of the simplified theory and the proposed models for merging and diverging behavior, this chapter extends and generalizes the simplified theory so that it is capable of dealing with a freeway system and a general transportation network, respectively.

Before we propose the extension and generalization of the simplified theory, we discussed some general issues regarding the expansion of K-Wave domain. This includes how to represent the expanded 2D space domain, the assumption on O-D flows, and the revised notation.

We proposed the extension to the simplified theory based on 5 basic building blocks of a freeway system, i.e., an entrance, an exit, a mainline, a merge, and a diverge, and we assumed that a freeway system can be represented by certain combination of these basic building blocks. We proposed the generalization of the simplified theory based on a generic building block where multiple upstream and/or downstream branches are allowed and we assumed that a general transportation network can be represented by certain combination of some basic building blocks which are special cases of the generic building block.

The rest part of this dissertation deals with empirical tests of the extension to the simplified theory (referred to as the "proposed model" thereafter), a formal process known as model validation.

CHAPTER 6

METHODOLOGY OF MODEL VALIDATION

The previous chapters have completed the theoretical development which includes an extension case that applies to freeway networks and a generalization case which applies to general transportation networks. In this chapter and the chapters that follow, we discuss the validation of the extension case (referred to as the proposed model thereafter). The purpose of this chapter is to present a methodology to guide subsequent steps to validate the model. We focus on data preparation and validation scheme in this chapter.

6.1 DATA REQUIREMENTS AND PREPARATION

Model validation involves empirical tests that compare the model output against field observations. Therefore, we need to determine a study site, collect data at this site (i.e., field observations), choose some data of acceptable quality from the data collected, prepare the information that is needed to run the model, implement and run the model, and compare the model output with the field observations.

6.1.1 Information Flow in Model Validation

It is helpful to understand how information flows in the process of model validation and Figure 6-1 serves such a purpose and we explain it in the paragraphs that follow.

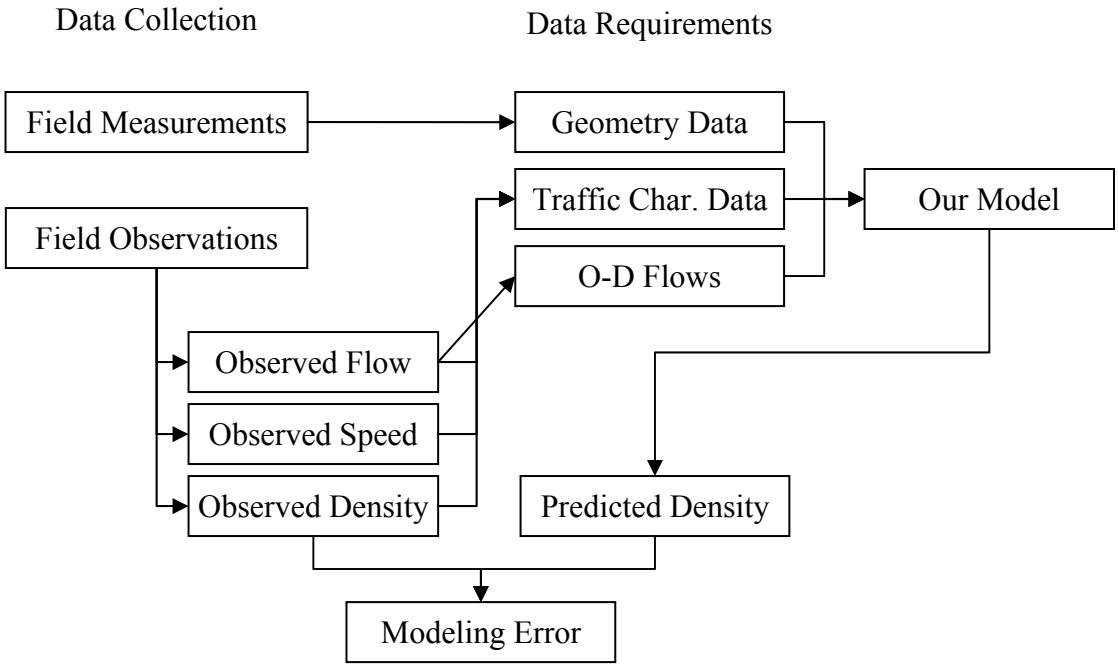


FIGURE 6-1 Information flow in model validation

The model we are trying to validate is the extension to the simplified theory and the model applies to a freeway system. Three pieces of data are required as the input to the model, i.e.,

i. **Geometry data** of the freeway system including location of entrances and exits, and length and number of lanes of each link.

ii. **Traffic characteristics data** including capacity, free-flow speed (FFS), and jam density of each link. These data uniquely define the triangular or trapezoidal flow-density curve assumed in the simplified theory.

iii. **Time-varying origin-destination (O-D) flows** of the freeway system, which is represented as the traffic demands at one of a set of entrances destined for one of a set of exits.

Data collection efforts for the above data are carried out in two parts: one is field measurements and the other is field observations. Field measurements can be obtained by some trips to the study site and the measurements provide the source of information to the geometry data. Field observations involve systematic traffic data collection at the study site over time and space. It is desirable to have the study site covered by traffic surveillance systems because such systems typically provide automated surveillance data. Field observations, including observed flow, speed, and density, are the source of information to the traffic characteristics data. The observed flow is also the source of information to the time-varying O-D flows. With the necessary input information, we are able to code the freeway into computer and run our model. As a result, we obtain the predicted traffic density as part of the model output. The model validation is a formal process that compares the predicted and the observed densities and the evaluation of the model performance is made on the modeling error, i.e., the difference between the two densities.

In subsection 6.1.2, we discuss the methodology of preparing the input information to the model, and, in subsection 6.1.3, we discuss our considerations in choosing density as the measure of evaluating model performance.

6.1.2 Methodology of Preparing the Input Data

In this subsection, we mainly discuss the methodology of determining the traffic characteristics data and briefly mention the methodology of preparing the geometry data and the time-varying O-D flows.

6.1.2.1 Methodology of Determining Geometry Data

The geometry data of a freeway system include

- Representation of the freeway system by a link-node structure,
- The locations of nodes
- The locations of entrances,
- The locations of exits,
- The lengths of links
- The numbers of lanes of links,
- Whether a link is a mainline section or a ramp.

These data can be determined by a few trips to the freeway system. First, one needs to get an overview of the freeway system and identify its homogeneous sections which are important in representing the freeway system by the link-node structure. When determining the locations of nodes, entrances, and exits, one needs to be consistent in choosing delineating points. The numbers of lanes can be spotted by the observer or obtained from aerial photos, if any. Link lengths can be measured by in-vehicle distance measurement devices.

6.1.2.2 Methodology of Determining Traffic Characteristics Data

Traffic characteristics data can be summarized by the relation among flow, speed, and density. In simplified theory, a triangular (or trapezoidal) flow-density relation is assumed and three parameters determine such a relation. One option is to use capacity, critical density, and jam density. It seems that this option is not practically attractive because critical density is typically difficult to obtain. An alternative, but more practically appealing, is to use capacity, FFS, and jam density and the methodology of determining these data is presented as follows.

Determining FFS

In the simplified theory, FFS coincides with the wave forward speed. HCM 2000 states that "FFS is the mean speed of passenger cars measured during low to moderate flows. For a specific segment of freeway, speeds are virtually constant in this range of flow rates. Two methods can be used to determine the FFS of a basic freeway segment: field measurement and estimation. … If field-measured data are used, no adjustments are made to the free-flow speed". In this study, the FFS is determined as the mean speed in the uncongested region. Figure 6-2 shows an example, based on field observations (We will discuss the field observations in the next chapter, but some examples are borrowed in advanced here). Part A of the figure plots speed against flow. It shows that the FFS starts somewhere around 60 mi/hr at low volume, and then gradually decreases to somewhere around 55 mi/hr when approaching capacity. Part B plots flow against density. The plot shows almost a straight line starting from the origin up to the capacity, and the slope is roughly 58 mi/hr, consistent with the result of Part A.

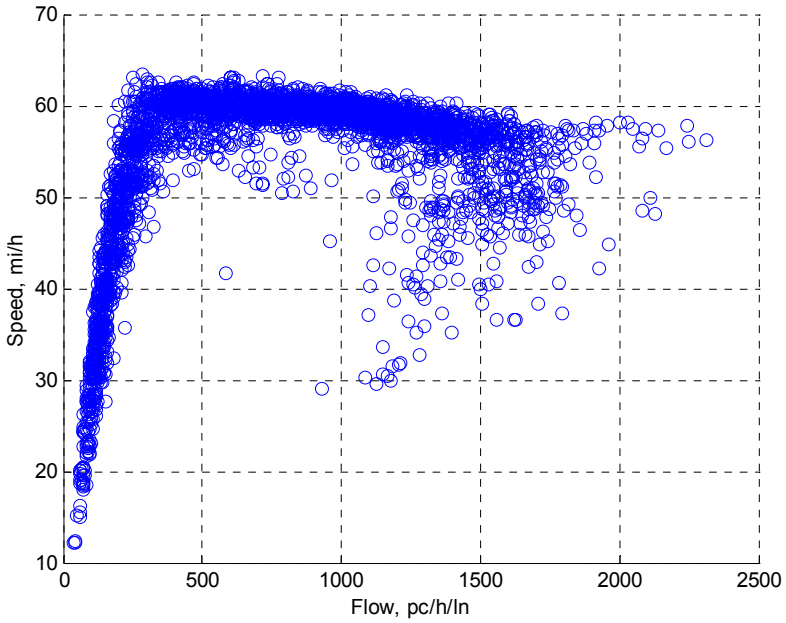


FIGURE 6-2A Plot of observed speed vs. observed flow

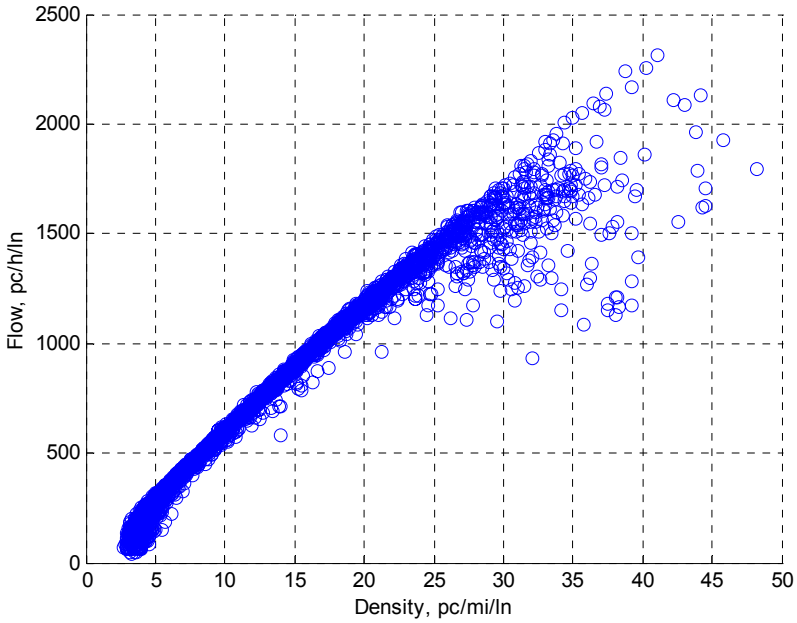


FIGURE 6-2B Plot of observed flow vs. observed density

Determining capacity

By definition, capacity is the maximum sustainable flow rate at which a segment of roadway can reasonably accommodate. HCM 2000 provides a procedure to estimate capacity based on various factors including FFS. Our discussion on model development reveals that the proposed model is sensitive to link capacity, especially at bottleneck(s). On the other hand, field measure suggests that traffic breaks down well before "theoretical" capacity, i.e., the capacity estimated from HCM procedure. Hurdle and Son (2000) also reported a similar finding. Therefore, an effective capacity that reflects the real traffic operation is more appropriate. HCM procedure, however, does not give an answer to this, so field measurement is used to determine such capacity. For each link, observed data points are carefully studied and observations during incidents are removed. The data are then merged into 5-minute intervals to reduce data variation, a common practice in capacity analysis. The resulting observation points are then plotted in a speed-flow diagram. A time series analysis (by means of a time series scatter plot) is also performed to analyze the pattern of the points. Link capacities are then read directly from the speed-flow diagram.

Determining jam density

Jam density, by definition, is the density measured when all vehicles are stopped because of congestion. Jam density is best measured from aerial photograph which is unfortunately unavailable in this study. An alternative way is to determine the density from observed flow-density diagram as the maximum density observed when flow drops to zero. However, an observed flow-density diagram typically fails to reveal such

information, as can be seem in Figure 6-2B. As a last resort, engineering judgment has to be exercised. Fortunately, the model is not very sensitive to jam density as reported in Hurdle and Son (2000) and this gives us the liberty to choose a jam density based on our best knowledge. From the findings reported for North American freeways (Ross 1988; Payne 1971; Michalopoulos, et al 1984, 1993), jam density generally falls in the range of 160 ~ 200 veh/mi/ln. As a reasonable starting point, we choose a jam density of 180 veh/mi/ln for every link.

Once FFS, capacity, and jam density are determined, a triangular (or trapezoidal) flow-density curve can be constructed. It should be noted that, in real traffic operation, there is so much random variation in these parameters that the simplified theory, as a deterministic macroscopic simulation model, is incapable of capturing all of them with a single setting. Therefore, these parameters, especially bottleneck capacities, are treated as "design parameters" which are subject to calibration and fine tuning.

6.1.2.3 Methodology of Determining Time-Varying O-D Flows

If the configuration of a site is simple, such as a merge or a diverge, it is easy to determine the time-varying O-D flows for the site. For example, the O-D flows for a merge can be synthesized from the observed link flows at the merging branches and the O-D flows for a diverge can be synthesized from the observed link flows at the diverging branches. However, the O-D flows for a freeway network is complicated and a suitable O-D estimator has to be employed to estimate the time-varying O-D flows from the observed link flows. This topic will be discussed in detail in Chapter 8.

6.1.3 Choosing Density as the Measure of Model Performance

In subsection 6.1.1, we propose using density as the measure of model performance i.e., we compare the predicted density against the observed density and draw conclusions based on the result of the comparison. Several reasons drive the use of density. First, density is a direct output of the model and density is not involved in model input. This means that the predicted density and the observed density come from different sources and thus are independent. This is important because the existence of correlation between the two variables that one tries compare will undermine the comparison result. In the light of this, flow and speed are not good candidates because they are part of the model input, i.e., the predicted flow or speed is partly determined by the observed flow or speed. Second, density is a continuous variable and it is sensitive to changes in model input, i.e., the density is able to reflect the full range within which the model behaves. In contrast, model outputs such delay and queue length are not good candidates because they vary only under congested conditions, i.e., one loses the opportunity to evaluate model performance in uncongested conditions. Third, the observed density is typically provided in the field observations and thus is readily available.

6.2 MODEL VALIDATION SCHEME

Before performing model validation, we need to determine a procedure to guide the evaluation of model performance. This section develops a formal scheme (Ni, et al 2004) by compiling a list of qualitative evaluation measures that are typically employed in traffic model validation and proposing a quantitative evaluation measure called

simultaneous statistical inference technique which performs statistical tests on both the mean and variance of the modeling error.

6.2.1 Qualitative Evaluation

Qualitative evaluation is mainly based on visual comparison of the predicted and observed data by means of graphs and plots. This is a commonly accepted practice to evaluate model performance and identify problems. However, there are some drawbacks with this approach, the most prominent of which is that the result it gives is also qualitative and fuzzy. Therefore, a formal test needs to incorporate quantitative evaluation to provide complementary information.

The means of qualitative evaluation include the following:

- Time series plot. It compares the observed and the predicted density in time series. This plot can be used to visually exam the difference between observed and predicted density over time.
- Contour plot. It compares the observed and the predicted density in terms of congestion (congestion is defined at a specific level of density). While a time series plot gives temporal information, a contour provides both temporal and spatial information. For example, if a density of 45 veh/mi/ln is assumed to delineate the congested and uncongested regions, evolution of congestion in time-space domain can be visualized by a density contour at this level.
- Surface plot. While a contour provides 2 dimensional (2D) information in time-space domain, a surface plot provides a 3D visualization by incorporating depth information in addition to temporal and spatial information. For example, an

analyst is able to visually compare the difference of the predicted and observed density at all levels over time and space.

- Diagonal plot. This plots the predicted density against the observed density. An ideal fit would be a 45 degree line. This plot provides an opportunity to evaluate the goodness of fit of a target variable.
- Histogram. This plot the frequency of modeling error and the plot helps to check whether modeling error is balanced at both sides of 0 and gives a general idea of the distribution of modeling error.
- Plot of the predicted and the observed flow-density relationship. This plot reveals whether the traffic in a simulated environment behaves in a similar way that the traffic in a real system does. The plot is also helpful to explain that the triangular (or trapezoidal) flow-density relationship assumed in the simplified theory is only a convenient way of propagating waves (disturbances in traffic) and it does not necessarily imply that one can use the relation as a lookup table to solely determine traffic density given a flow.

6.2.2 Quantitative Evaluation

This subsection develops a simultaneous statistical inference technique to quantitatively evaluate model performance.

6.2.2.1 Issues in traffic model validation

Traffic simulation model validation is not about selecting the best among several alternatives, nor about testing the goodness of fit between two random samples. It is about comparing two processes, the simulated and the observed, and checking how one

approximates the other. In quantitative traffic model validation, there are at least three salient problems. First, in some cases the timing, or ordering, of the samples is very important. For example, traffic density of an urban freeway is typically high around 8:00AM, but a peak in the night is rarely seen. This implies that we are very often concerned with testing the goodness of fit (GOF) between two time series processes, i.e., the simulation and the observation over time. Second, in some cases the samples are correlated. For example, the traffic density at the next step is a result, at least in part, of the traffic density at previous steps. Correlation is usually a problem in performing statistical tests which typically require random samples. Third, a traffic model can be a deterministic one in which no randomness is involved. Unlike stochastic models where statistical analysis is often employed to study the average behavior of the model, our interest here is to check whether the model makes proper responses at the right time.

The above problems show that regular statistical tests may not be sufficient and a quantitative evaluation procedure has to be carefully devised to deal with these problems. The first problem pertains to the ordering of data points, which means the sequence of simulated results has to match the observed results. This problem can be solved by pairing the two groups of data. For example, rather than working on the raw data as two groups, we work on the residuals after pairing the simulated and the observed. By this way, the ordering of data has been automatically accounted for. The second problem is the correlation/autocorrelation in the raw data as well as the residuals. Correlation/autocorrelation is usually a problem in performing statistical tests because they typically require random/independent samples. To circumvent the problem of autocorrelation, the effective sample size has to be reduced (Dawdy and Matalas 1964)

and the reduction in number of independent observations has implications for hypothesis tests. Fortunately, the batch means technique (Goldsman, 2000) discussed below can decently address autocorrelation while still making full use of the raw data. The last problem pertains to the meaning of goodness-of-fit (GOF). GOF may have different meanings in different contexts and the nuance may not be easily captured. For example, in model validation, GOF often means how close the model output approximates the observation, while, in statistics, GOF typically means whether two sets of observations could reasonably have come from the same distribution. What makes things even more complicated is that model validation often involves statistical tests and the two meanings of GOF are easily mixed. In traffic simulation model validation, one is frequently working with the former meaning of GOF and the simultaneous statistical inference technique discussed below is developed with this in mind.

6.2.2.2 The Batch Means Technique

Generally, the objectives of quantitative model validation are the following:

A. Test whether the model is capable of replicating the real system with sufficient accuracy, i.e., the simulation is unbiased. This translates to testing whether the mean of modeling error is statistically different from zero.

B. It should be kept in mind that the passing of the test in A is only a necessary condition for a good model because large positive and negative errors can cancel each other and still yield a zero mean. Hence, it is important to check whether the variance of the modeling error is reasonably small, i.e., whether the model is capable of replicating the real system with sufficient precision.

C. With test (1) and (2) combined, we are confident about drawing a conclusion. However, something has to be done to address the problem of correlation/autocorrelation before performing the above tests. Otherwise their validity can be undermined substantially.

With these considerations in mind, simultaneous statistical tests are devised based on the following two hypotheses:

Hypothesis 1: the prediction is unbiased. This is intended to address objective A.

Hypothesis 2: modeling error is reasonably small. This is intended to address objective B.

To address objective C, the above tests are going to be performed based on the batch means technique (Goldsman, 2000) which is particularly designed to handle correlation/autocorrelation in samples. To facilitate subsequent discussion, the batch means technique is presented below.

Suppose we have two processes as above where X is the steady-state simulation output and Y is the observation in the field. Their residuals are computed as $Z_i = X_i - Y_i$, $i = 1,2,...,n$. Notice that Z might be a non-identically independently distributed (non-IID) process. For example, it can be correlated so that the samples are not independent, or it can be non-stationary because its variability increases as X and Y get large. Generally, there is no uniform treatment to turn a non-stationary process stationary. However, if the process exhibits some special pattern, a log transformation might be helpful to serve our purpose. Considering that traffic flow measurements such as density, flow, speed, queue

length, etc., often take positive values and a log transformation happens to lead to a nice feature of percentage error measurement at predetermined confidence level, this transformation deserves special note and hence serves as the basis of the following discussion. If, luckily, process Z is already stationary before transformation, general idea about the following procedure still applies, except that no transformation is needed and the measurement of error is in absolute terms rather than percentage.

Let $Z_i = \ln X_i - \ln Y_i$, $i = 1,2,...,n$, and assume this makes process Z stationary. Next, let us address the problem of non-independence by batching the data:

$$\underbrace{Z_1, Z_2,...,Z_m}_{\text{batch 1}} \,|\, \underbrace{Z_{m+1}, Z_{m+2},...,Z_{2m}}_{\text{batch 2}} \,|\, \ldots\ldots \,|\, \underbrace{Z_{(b-1)m+1}, Z_{(b-1)m+2},...,Z_{bm}}_{\text{batch b}}$$

where m is the batch size, b is the number of batches. Obviously, $n = m \times b$.

For each batch, we use its batch mean as the "representative" of the batch:

$$\overline{Z}_{1,m} = \frac{1}{m}\sum_{j=1}^{m} Z_j$$

$$\overline{Z}_{2,m} = \frac{1}{m}\sum_{j=1}^{m} Z_{m+j}$$

…

$$\overline{Z}_{b,m} = \frac{1}{m}\sum_{j=1}^{m} Z_{(b-1)m+j}$$

If m is large enough, batch means are approximately normally distributed, i.e., $\overline{Z}_{1,m}, \overline{Z}_{2,m},..., \overline{Z}_{b,m} \approx N\left(E\left[\overline{Z}_{1,m}\right], Var\left(\overline{Z}_{1,m}\right)\right)$. Here, independence is roughly achieved because most Z_i's in one batch are nearly independent of most Z_i's in another batch for large m.

Identical distribution is also achieved by Central Limit Theorem considering that we are in steady state by assumption and Z_i's are stationary after transformation.

Next, let us estimate the mean of process Z, $\mu = E[Z]$, and construct a batch means confidence interval for it. The batch means estimator for μ is the grand mean $\overline{Z}_n$:

$$\overline{Z}_n = \frac{1}{b}\sum_{i=1}^{b} \overline{Z}_{im}$$

The batch means estimator for variance σ^2 is $\hat{V}_B$:

$$\hat{V}_B = \frac{m}{b-1}\sum_{i=1}^{b} \left(\overline{Z}_{im} - \overline{Z}_n\right)^2$$

As we can see that

$$\frac{\overline{Z}_n - \mu}{\sqrt{\sigma^2/n}} \approx N(0,1)$$

$$\hat{V}_B \approx \frac{\sigma^2 \chi^2(b-1)}{b-1}$$

and they are nearly independent. Therefore,

$$\frac{\dfrac{\overline{Z}_n - \mu}{\sqrt{\sigma^2/n}}}{\sqrt{\hat{V}_B/\sigma^2}} = \frac{\overline{Z}_n - \mu}{\sqrt{\hat{V}_B/n}} \approx t(b-1)$$

The batch means confidence interval is constructed as:

$$\mu \in \overline{Z}_n \pm t_{\frac{\alpha}{2},b-1}\sqrt{\hat{V}_B/n}$$

It can be shown that, as $m \to \infty$ while b is held constant, the coverage of the confidence interval $\to 1-\alpha$. On the other hand, to minimize the mean squared error of $\hat{V}_B$ as an estimator of σ^2, $b >> m$ is desirable. Therefore, when choosing m and b, there

are two competing criteria and one has to properly trade off between the two based on his/her goals.

6.2.2.3 Simultaneous Statistical Inference Technique

Based on the two hypotheses presented above, the simultaneous hypotheses are formally made below:

Hypothesis 1: $H_o^1 : \mu_Z = 0$

Hypothesis 2: $H_o^2 : \sigma_Z^2 \le \varepsilon$

where μ and σ_Z^2 are the mean and variance of process Z and ε is a pre-selected number that is reasonably small based on the specific problem. Since Z might be a non-IID process, the batch means technique is applied and a new process $\overline{Z} = \{\overline{Z}_{1m}, \overline{Z}_{2m}, ..., \overline{Z}_{bm}\}$ with batch size m and number of batches b is resulted. After applying the batch means technique, we are assumed to have obtained a (nearly) IID stationary process $\overline{Z}$ and the simultaneous hypotheses are redefined as follows:

Hypothesis 1: $H_o^1 : \mu_{\overline{Z}} = 0$

Hypothesis 2: $H_o^2 : \sigma_{\overline{Z}}^2 \le \varepsilon$

To test Hypothesis 1, a regular Student's t-Test suffices the need. The test statistic t_0 is computed as:

$$t_0 = \frac{\overline{Z}_n}{\sqrt{\hat{V}_B / b}}$$

where $\overline{Z}_n$ is the grand mean of the batch means and $\hat{V}_B$ is the batch means estimator for variance $Var(\overline{Z}_{im})$ as defined above. If $|t_0| > t_{\frac{\alpha}{2},b-1}$, we reject the null hypothesis H_o^1. Otherwise, there is a lack of evidence that $E[\overline{Z}_{im}]$ is statistically different from 0.

For hypothesis 2, since $\hat{V}_B \sim \varepsilon \frac{\chi^2(b-1)}{b-1}$ under null hypothesis H_o^2, a Chi-squared test serves our need. The test statistic χ_0^2 is computed as:

$$\chi_0^2 = \frac{(b-1)\hat{V}_B}{\varepsilon}$$

which is $\chi^2(b-1)$ distributed. If $\chi_0^2 > \chi_{\alpha,b-1}^2$, we reject H_o^2. Otherwise, there is a lack of evidence that $Var(\overline{Z}_{im})$ is statistically greater than ε.

In order to interpret the above tests, let us plug X_i's and Y_i's in the above test statistics. Suppose we have made a log transformation on X_i's and Y_i's, i.e., $Z_i = \ln X_i - \ln Y_i$. From the definition of batch means, we have:

$$\overline{Z}_{im} = \frac{1}{m}\sum_{j=1}^{m}\left(\ln X_{(i-1)m+j} - \ln Y_{(i-1)m+j}\right) = \ln\left(\prod_{j=1}^{m}\frac{X_{(i-1)m+j}}{Y_{(i-1)m+j}}\right)^{1/m} = \ln\left(\frac{\left(\prod_{j=1}^{m} X_{(i-1)m+j}\right)^{1/m}}{\left(\prod_{j=1}^{m} Y_{(i-1)m+j}\right)^{1/m}}\right) = \ln\left(\frac{\widehat{X}_{im}}{\widehat{Y}_{im}}\right)$$

where $\widehat{X}_{im}$ denotes the geometric mean of the i-th batch of the simulation output, and $\widehat{Y}_{im}$ denotes the geometric mean of the i-th batch of the field observation.

$$E\left[\ln\frac{\hat{X}_{im}}{\hat{Y}_{im}}\right]=\ln E\left[\frac{\hat{X}_{im}}{\hat{Y}_{im}}\right]=E\left[\overline{Z}_{im}\right]$$

$$E\left[\frac{\hat{X}_{im}}{\hat{Y}_{im}}\right]=e^{E\left[\overline{Z}_{im}\right]}$$

Notice that the above equations are derived based on geometric means of the simulated and observed batches, so interpretation of the simultaneous statistical test has to be based on batches. The t-test implies that, if $|t_0|>t_{\frac{\alpha}{2},b-1}$, there is strong evidence that $E[\overline{Z}_{im}]$ is statistically different than 0, i.e., the expected value of the geometric mean of the simulated is statistically different than that of the observed. Otherwise, we accept the opposite. The Chi-squared test can be translated to a confidence interval. Since we can reasonably assume that $\overline{Z}_{im}$'s are normally distributed, the confidence interval for the ratio of the geometric mean of the simulated and the observed can be constructed as $e^{\overline{Z}_n\pm t_{\frac{\alpha}{2},b-1}\sqrt{\varepsilon}}$ with a pre-determined small number ε. This actually gives us a sort of percentage error with confidence level of $(1-\alpha)$.

The last question left is how to choose m and b which are competing parameters if the sample size is fixed. There is no fixed answer and the decision largely depends on the actual situation and the goal one tries to achieve. For example, if the goal is to minimize the mean squared error of $\hat{V}_B$ as an estimator of $Var(\overline{W}_i)$, it may be better to have $b>>m$. If, on the other hand, the goal is to increase the change for the confidence interval to cover the true mean, choosing an m with $m\to\infty$ might be more appropriate.

6.3 SUMMARY

In this chapter, we discussed the methodology that guides the subsequent model validation effort. More specifically, we focused on two issues: one is data preparation and the other is validation scheme.

For the first issue, we gave a general framework showing how information flows in a model validation process. From this framework, we identify the data requirements and data collection efforts. Based on the data requirements, we discuss the methodology in preparing the required data as well as considerations on using density as the measure of model performance.

For the second issue, we developed a formal scheme for model validation which involves means of qualitative evaluation and quantitative evaluation. For the qualitative means, we proposed using time series plots, contour plots, surface plots, diagonal plots, histograms, and the plot of predicted and observed flow-density relationships to evaluate model performance. For the quantitative means, we formulated a simultaneous statistical inference technique to evaluate both the accuracy and the precision of the model.

Each of the subsequent chapters deals with a piece of the tasks in model validation. Chapter 7 presents the test sites and prepares two of the three pieces of input information to the model: the geometry data and the traffic characteristics data of these sites. Chapter 8 prepares the last piece of input information to the model: the time-varying O-D flows of one of the test sites (the other two do no require O-D estimation). Chapter 9 presents the results of model validation.

CHAPTER 7

TEST SITES AND TEST DATA

Model validation involves empirical tests which compare the model output against field observations. We select three test sites, each of which is used to test a scenario of the proposed model. We will test a merge scenario, a diverge scenario, and a network scenario. The first two scenarios do not require O-D estimation but the last scenario does. For each test site/scenario, the proposed model requires three pieces of information as its input, i.e., geometry data, traffic characteristics data, and time-varying O-D flows. This chapter prepares the first two pieces.

7.1 SITE SELECTION

Before preparing the two pieces of data, we need to select a study site. During site selection, we have the following considerations:

1) The site needs to be a freeway with on- and off-ramps because we are testing the extension case which applies to freeways.
2) It is better to have a freeway that is covered by automated traffic surveillance systems because such systems are likely to execute mass data collection over time and space.

3) The data collected need to include both mainline and ramp observations because the proposed model involves both of them.
4) It is helpful to have both sites that require and do not require O-D estimation because O-D estimation error is exogenous to our model and we want to get some insights in our model performance with and without the presence the O-D estimation error.

In this study we choose a section of GA400 as our study site because it satisfies the first three considerations as will be discussed in the next section. We further select three smaller portions from the study site as our test sites to satisfy the fourth consideration and we use the three test sites to test three scenarios, respectively, of the model: a merging scenario and a diverging scenario which do not require O-D estimation and a network scenario which does. The three test sites will be presented in detail in section 7.4.

7.2 THE STUDY SITE AND THE DATA SOURCE

7.2.1 The Location

GA 400 is a long highway that runs between I-85 to the south and S Chestatee St., Dahlonega GA, to the north. Currently, automated traffic surveillance systems cover the section between I-285 and Old Milton Parkway, a stretch of road of approximately 12.56 miles long, and this serves as our study site, as circled in Figure 7-1. Though GA 400 is not a freeway, our study site is considered a "freeway by design." First, there is neither access to abutting lands nor driveway or intersection at grade. Second, all access and egress are via ramps that are intended to allow merging and diverging to take place at reasonable speeds. These are in response to the first two considerations mentioned above.

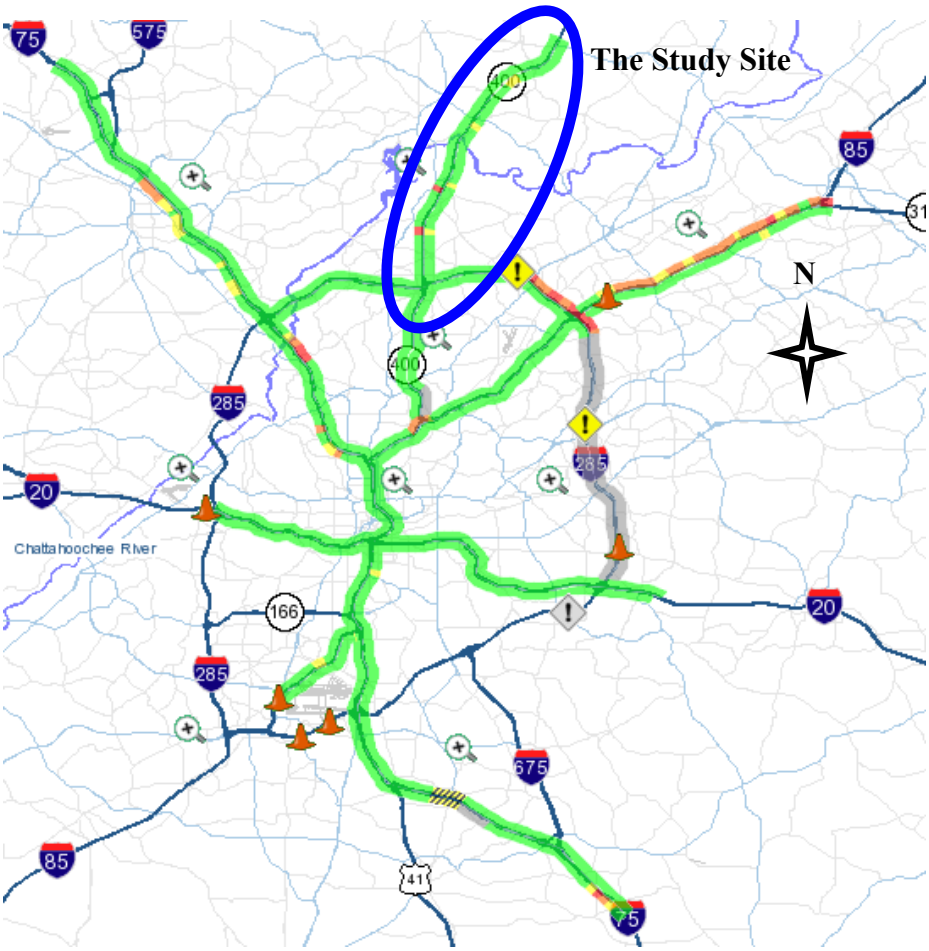


FIGURE 7-1 The study site – a section of GA 400

7.2.2 The Sources of Data

Two sources of data are available for this study. One is the data collected by the automated traffic surveillance systems and these data provide traffic observations over time and space at the study site, from which we prepare the traffic characteristics data mentioned in the beginning of this chapter. The other is the data we collected during our

field trips to the study site and these data provide mainly the geometry information of the study site, from which we prepare the geometry data mentioned in the beginning of this chapter.

7.2.3 Automated Data Collection Stations

Automated data collection stations are deployed along the study site approximately every one third mile in each direction, and ramps are also monitored. This amounts to 100 observation stations in total. All stations are numbered with a leading "400". Mainline stations bear numbers between 4000025 and 4001139, in which 4000025 - 4000063 are for northbound direction and 4001101 - 4001139 are for southbound direction. Each on-ramp bears a number of 5000 level, e.g., 4005005, 4005102, etc., and each off-ramp bears a 6000 level number, e.g., 4006006, 4006103, etc. Mapping of observation stations is sketched in Figure 7-2. In addition to the stations, this figure also shows the numbers of lanes and locations of ramps and observation stations. The annotation of a station typically contains the following information: station number, milepost, and nearby street or freeway exit.

Traffic is monitored by video cameras at the observation stations. Video images are then processed by a Belgium-based Traficon detection system. For each station, traffic data flow in every 20 seconds and items collected include sample start time, station status, data confidence, vehicle counts (auto, van, truck, and other), time occupancy, time mean speed, density, etc. Five month worth of data, i.e., from September 1st, 2002 to January 25th, 2003, are available for the study.

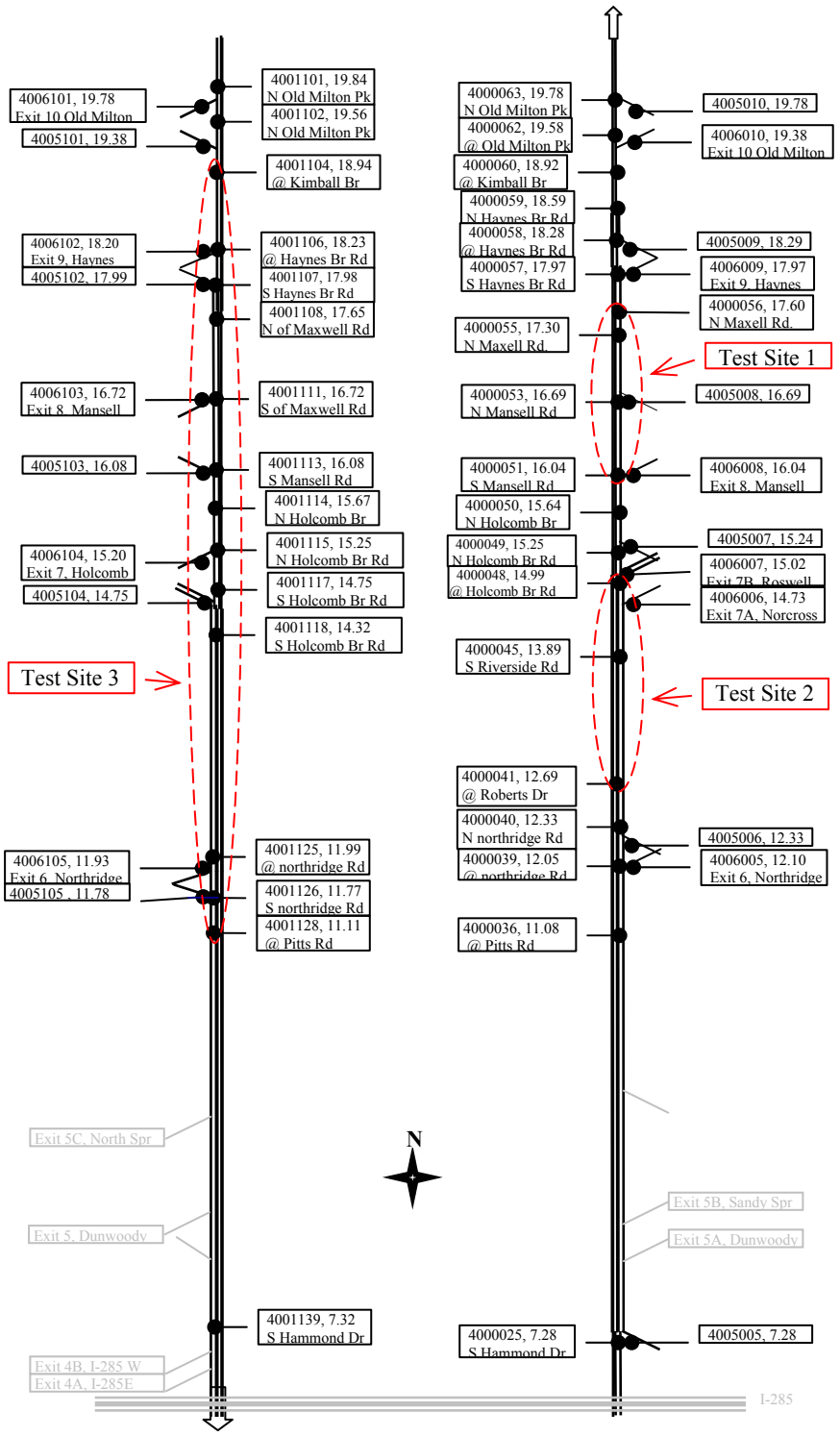


FIGURE 7-2 Observation stations and geometry of the study Site

7.2.4 Description of the Automated Data

The data collected by the automated surveillance systems are archived every day in the form of a single compressed file. This archived file contains observations at each station during the day. The starting and ending times of observations vary from day to day, and from station to station even on the same day. An example of the data at station 4000025 is displayed (after a transpose) in Table 7-1 in which reported data items include the following:

TABLE 7-1 Sample data recorded at an observation station

Station ID	4000025	4000025	4000025	…
Sample Start	2002/09/01 23:50:40 EDT	2002/09/01 23:51:00 EDT	2002/09/01 23:51:20 EDT	…
Status	|OK|OK|OK	|NO_ACT|OK|OK	|OK|OK|OK	…
Confidence	1	2	1	…
Volume Auto	7	9	5	…
Volume Van	0	0	0	…
Volume Truck	0	0	0	…
Volume Other				…
Time Occupancy	0.0127	0.0227	0.0267	…
Space Occupancy				…
Time Speed	56.379	56.9966	57.7874	…
Space Speed				…
Length	8.4646	8.2021	8.2021	…
Level of Service				…
Flow				…
Density	8.9333	33.8182	8.8889	…
Gap	2469.5333	2016	1194.1111	…
Headway				…
Alarms				…

Station ID: The ID of the observation station, e.g., 4000055. A station consists of one or several detectors.

Sample Start: Start time of the sample, e.g., 2002/08/31 23:50:40 EDT

Status: Lane by lane status of detectors, e.g., for a 3 lane section, status |OK|NO_ACT|SENSOR_FAILURE means that the detector of left lane is good, there is no activation at the detector of the middle lane, and there is failure with the detector of the right lane.

Confidence: The confidence of the current sample, taking a value between 0 and 10 with 0 for no confidence and 10 for full confidence.

Volume Auto: Count of passenger cars detected in current 20-second interval.

Volume Van: Count of vans detected in current 20-second interval.

Volume Truck: Count of trucks detected in current 20-second interval.

Volume Other: Count of other vehicles detected in current 20-second interval.

Time Occupancy: Percent of time in an hour that the detector is calling

Space Occupancy: Unknown

Time Speed: Time mean speed (TMS), in mi/h

Space Speed: Space mean speed (SMS), in mi/h

Length: Unknown

Level of Service: LOS

Flow: traffic flow converted from the 20-second sample, in vehicle/hr.

Density: traffic density measured in vehicle/mi.

Gap: Unknown

Headway: Average headway of the current sample.

Alarms: Unknown

Note: Unknown means the definition of this item is not clear.

7.2.5 Sources of Error

Generally, we identified the following sources of error in the automated data:

Data not Available

This includes the types of data that are desirable but not collected by the automated surveillance systems, such as space occupancy, space mean speed, level of service, flow, and headway.

Machine or communication failure

This is indicated as "NO_ACT" or "SENSOR_FAILURE" in the Status. If these indicators appear in a data entry, this entry is generally useless and a gap is resulted.

Sensor error

This is indicated in the Confidence as a number between 0 and 10 with 10 being the best data quality. Such an error can take the forms of gaps (missing data) or erroneous data due to bad sensor or bad conditions to collect data. Gaps and erroneous data are big issues because they may render a set of data useless. For example, one has a site consisting 20 stations and have collected a day worth of data at the site. If gaps and erroneous data account for 5% of the data at a station, the worst case will be no usable data because one will have difficult to find a common time period during which the data are good throughout all the stations.

Conversion error

This happens when converting the field observations to some measures based on flawed equations. For example, during a sampling interval at a station, 5 vehicles are observed and their speeds are recorded. One can compute the time mean speed as the arithmetic mean of these speeds or the space mean speed as the harmonic mean of

these speeds. The conversion error is resulted when one uses the time mean speed (which is provided by the surveillance system) in place of the space mean speed.

The purpose that we discuss the potential sources of error here is to remind their existence. Our strategies to the error are the following: we try to avoid data with missing or erroneous values and data that are flawed in other forms; however, if we decide to use a set of data, we have to assume they are accurate.

7.3 THE TEST SITES AND THE DATA OF THE TEST SITES

This section presents the test sites and their geometry data and traffic characteristics data which are part of the input to the proposed model.

7.3.1 The Test Sites

In response to Section 7.1, we need three test sites to test the three scenarios of the proposed model, respectively. Considering the data availability and the data quality issues, we determine the test sites as follows:

- Test site 1, for testing the merge scenario, lies on the northbound direction between stations 4000051 and 4000056 with an on-ramp station 4005008.
- Test site 2, for testing the diverge scenario, lies between stations 4000041 and 4000048 with an off-ramp station 4006006.
- Test site 3, for testing the network scenario, lies between stations 4001104 and 4001128 with 4 on-ramp stations and 4 off-ramp stations. Maps and data about these sites will be presented in the next section.

These test sites are labeled with dashed circles in Figure 7-2.

7.3.2 Data of Test Site 1

Figure 7-3 illustrates the map of test site 1 which is part of northbound GA 400. The site begins at station 4000051 near the south of Mansell Road, and runs through stations 4000052 at Mansell Road, 4000053 near the north of Mansell Road, 4000054 near the south of Maxwell Road, 4000055 near the north of Maxwell Road, and ends at station 4000056 near 1/2 mile to the north of Maxwell Road. An on-ramp with station 4005008 joins the freeway near the north of Mansell Road at node 5008 which is not a real station and is so numbered for only coding purpose. Note that, in future simulation model, coding of the test site may not literally follow this map and some links may be skipped or merged depending on actual needs.

Geometry data of this site are summarized in Table 7-2 and traffic characteristics data are listed in Table 7-3.

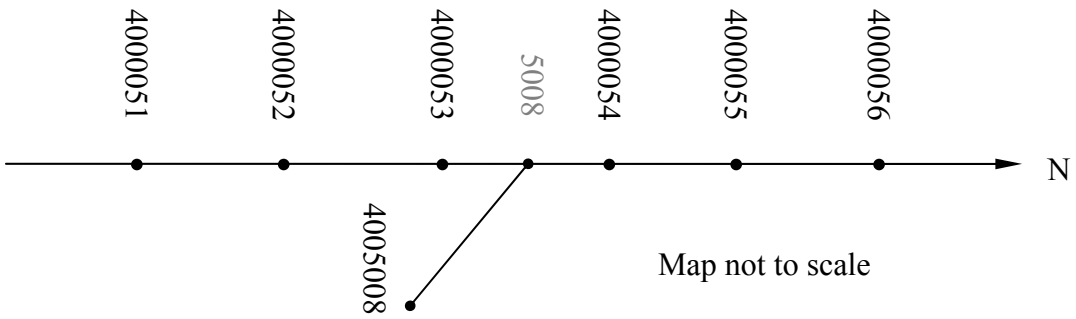


FIGURE 7-3 Test Site 1 – merging scenario (GA 400 northbound)

TABLE 7-2 Geometry data of test site 1

Link	Up Node	Down Node	Length (mi)	# of Lanes	Description
1	4000051	4000052	0.37	3	Mainline
2	4000052	4000053	0.28	3	Mainline
3	4000053	5008	0.16	3	Mainline
4	5008	4000054	0.17	3	Mainline
5	4000054	4000055	0.28	3	Mainline
6	4000055	4000056	0.31	3	Mainline
7	4005008	5008	0.50	1	On-ramp

TABLE 7-3 Traffic characteristics data of test site 1

Link	Up Node	Down Node	FFS (mi/h)	Capacity (veh/h/ln)	Jam Density (veh/mi/ln)
1	4000051	4000052	65	2200	180
2	4000052	4000053	65	2200	180
3	4000053	5008	65	2200	180
4	5008	4000054	60	2200	180
5	4000054	4000055	60	2200	180
6	4000055	4000056	61	2200	180
7	4005008	5008	20	1800	180

In this site, section between stations 4000054 and 4000056 is a possible bottleneck, especially considering that an on-ramp joins before station 4000054. In addition, queues can also originate from further downstream of station 4000056.

7.3.3 Data of Test Site 2

Figure 7-4 illustrates the map of test site 2 which is part of northbound GA 400. The site begins at station 4000041 at Roberts Drive, and runs through stations 4000042 near the north of Roberts Drive, 4000043 near half mile to the north of Roberts Drive, 4000044 near one mile to the north of Roberts Drive, 4000045 near the south of

Riverside Road, 4000046 near 3/4 mile to the south of Holcomb Bridger Road, 4000047 near the south of Holcomb Bridger Road, and ends at station 4000048 at Holcomb Bridger Road. An off-ramp with station 4006006 leaves the freeway at node 6006 near the south of Holcomb Bridger Road. Again, this node is not an actual station.

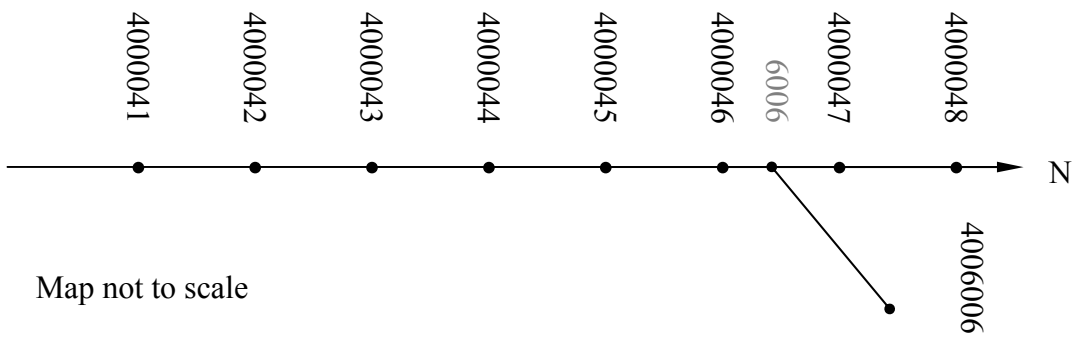


FIGURE 7-4 Test Site 2 – diverging scenario (GA 400 northbound)

Geometry data of this site are summarized in Table 7-4 and traffic characteristics data are listed in Table 7-5.

TABLE 7-4 Geometry data of test site 2

Link	Up Node	Down Node	Length (mi)	# of Lanes	Description
1	4000041	4000042	0.28	4	Mainline
2	4000042	4000043	0.33	4	Mainline
3	4000043	4000044	0.32	4	Mainline
4	4000044	4000045	0.27	4	Mainline
5	4000045	4000046	0.35	4	Mainline
6	4000046	6006	0.23	4	Mainline
7	6006	4000047	0.24	4	Mainline
8	4000047	4000048	0.28	4	Mainline
9	6006	4006006	0.50	1	Off-ramp

TABLE 7-5 Traffic characteristics data of test site 2

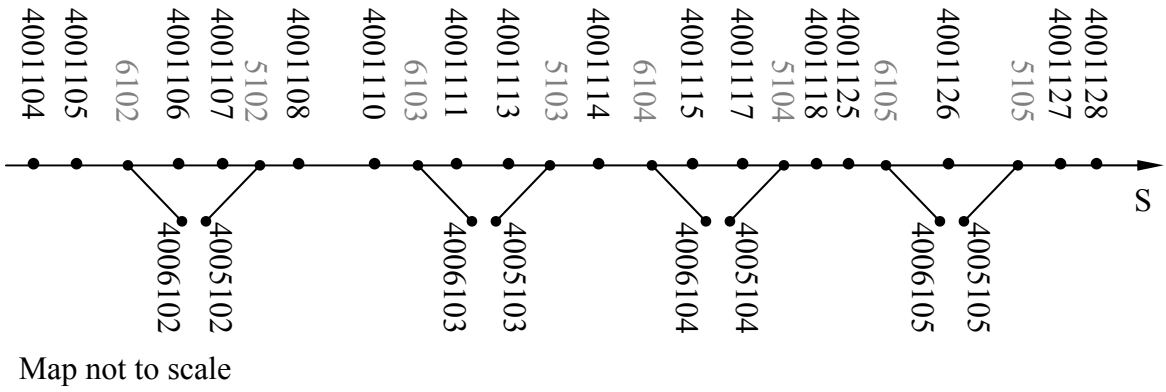
Link	Up Node	Down Node	FFS (mi/h)	Capacity (veh/h/ln)	Jam Density (veh/mi/ln)
1	4000041	4000042	68	2200	180
2	4000042	4000043	68	2200	180
3	4000043	4000044	68	2200	180
4	4000044	4000045	68	2200	180
5	4000045	4000046	68	2200	180
6	4000046	6006	68	2200	180
7	6006	4000047	60	2200	180
8	4000047	4000048	65	2200	180
9	6006	4006006	60	2000	180

Notice that mainline links in this site have the same capacity, so it is unlikely that a queue is developed and dissipated within this site. However, queues can possibly originate from further downstream of the off-ramp or the mainline where there might be insufficient capacity.

7.3.4 Data of Test Site 3

Figure 7-5 illustrates the map of test site 3 which is more complicated and is part of southbound GA 400. The site begins at station 4001104 at Kimball Bridge Road, and passes station 4001105 near the north of Haynes Bridge Road all the way to station 4001128 at Pitts Road. Four on-ramps join and four off-ramps leave the freeway. The off-ramp with station 4006102 leaves the freeway at Haynes Bridge Road, the off-ramp with station 4006103 leaves the freeway near the south of Maxwell Road, the off-ramp with station 4006104 leaves the freeway near the north of Holcomb Bridge Road, and the off-

ramp with station 4006105 leaves the freeway at Northridge Road. The on-ramp with station 4005102 joins the freeway near the south of Haynes Bridge Road, the on-ramp with station 4005103 joins the freeway near the south of Mansell Road, the on-ramp with station 4005104 joins the freeway near the south of Holcomb Bridge Road, and the on-ramp with station 4005105 joins the freeway near the south of Northridge Road.



Map not to scale

FIGURE 7-5 Test Site 3 – network scenario (GA 400 southbound)

Geometry data of this site are summarized in Table 7-6 and traffic characteristics data are listed in Table 7-7.

TABLE 7-6 Geometry data of test site 3

Link	Up Node	Down Node	Length (mi)	# of Lanes	Description
1	4001104	6102	1	2	Mainline
2	6102	5102	0.4	2	Mainline
3	5102	6103	1.1	3	Mainline
4	6103	5103	0.9	3	Mainline
5	5103	6104	0.88	3	Mainline
6	6104	5104	0.5	3	Mainline
7	5104	6105	2.9	4	Mainline
8	6105	5105	0.5	4	Mainline
9	5105	4001128	0.5	4	Mainline
10	6102	4006102	0.5	1	Off-ramp
11	4005102	5102	0.5	1	On-ramp
12	6103	4006103	0.5	1	Off-ramp
13	4005103	5103	0.5	1	On-ramp
14	6104	4006104	0.5	1	Off-ramp
15	4005104	5104	0.5	2	On-ramp
16	6105	4006105	0.5	1	Off-ramp
17	4005105	5105	0.5	1	On-ramp

TABLE 7-7 Traffic characteristics data of test site 3

Link	Up Node	Down Node	FFS (mi/h)	Capacity (veh/h/ln)	Jam Density (veh/mi/ln)
1	4001104	6102	56	2200	180
2	6102	5102	57	2200	180
3	5102	6103	61	2200	180
4	6103	5103	65	2100	180
5	5103	6104	61	2200	180
6	6104	5104	60	1900	180
7	5104	6105	64	2100	180
8	6105	5105	60	2100	180
9	5105	4001128	56	2200	180
10	6102	4006102	22	1000	180
11	4005102	5102	32	2000	180
12	6103	4006103	20	1000	180
13	4005103	5103	22	1250	180
14	6104	4006104	40	1200	180
15	4005104	5104	40	2000	180
16	6105	4006105	20	1500	180
17	4005105	5105	20	1700	180

Mainline link 6104 – 5104 is a possible bottleneck which may initiate a queue. In addition, other factors, such as sudden inflow from on-ramps and in sufficient capacity at further downstream of off-ramps or mainline, may also cause congestion.

7.4 EXPERIMENT DESIGN

So far, we have determined the test sites for the model validation and have prepared geometry data and traffic characteristics data for the test sites. We need one more piece of data, i.e., the time-varying O-D flows, for the test sites, and this requires some specific days' data of acceptable quality to: (1) synthesize the O-D flows and (2) to provide the field measurement against which the model output is compared. We then associate these days' data to the test sites and call this process experiment design. When selecting which day's data to use, we have the following considerations:

- It is desirable to include congestions in the selected data because whether the model is able to replicate congestion with reasonable accuracy is the major concern of our model.
- Avoid data with congestions on entry links because we use the observed link flows synthesize O-D flows. If a queue reaches an entry link, the observed link flow at this time no longer represents the true demand at the link. As a result, the synthesized O-D flows are not accurate
- Avoid data with machine/communication failure and gaps in the selected data because they represent missing values in the raw data.

- Avoid data with incidents or road work because they result in non-recurrent congestions which are not the major interest of the model. In fact, the model is intended to capture recurrent and sustained congestions.

Based on the above considerations, 2 days' data are selected for each test site. For easy reference, Table 7-8 summarizes the experiment design.

TABLE 7-8 Summary of the experiment design

Test Site	Testing Scenario	Day of Week	Date	Duration	O-D estimation
1	Merging	Monday	10/14/2002	00:01:00 – 23:51:00	Not Required
1	Merging	Friday	9/6/2002	00:00:40 – 23:50:40	Not Required
2	Diverging	Thursday	9/12/2002	00:01:00 – 23:51:00	Not Required
2	Diverging	Monday	12/9/2002	00:00:00 – 23:45:00	Not Required
3	Network	Friday	10/11/2002	05:50:00 – 19:20:00	Required
3	Network	Monday	10/14/2002	05:51:40 – 19:56:40	Required

The table shows that test site 1 is for testing the merging scenario and we use the data collected on Monday, October 14, 2002 from 00:01:00 to 23:51:00 and the data collected on Friday, September 6, 2002 from 00:00:40 to 23:50:40 for this test site. No O-D estimation is required at this site because O-D flows can be synthesized from entry link flows. Test site 2 is for testing the diverging scenario and we use the data collected on Thursday, September 12, 2002 from 00:01:00 to 23:51:00 and the data collected on Monday, December 9, 2002 from 00:00:00 to 23:45:00 for this test site. Also, no O-D estimation is required at this site because O-D flows can be synthesized from exit link flows. Test site 3 is for testing the network scenario and we use the data collected on

Friday, October 11, 2002 from 05:50:00 to 19:20:00 and the data collected on Monday, October 14, 2002 from 05:51:40 to 19:56:40 for this test site. This test site requires O-D estimation and, in the next chapter, we will estimate O-D flows for this test site based on each of the two days' data.

We recognize that the data in the selected days may not account for the daily variation in a week. This is partly because we have been unable to find data of acceptable quality on Tuesdays and Wednesdays. Fortunately, this is not an important issue for model validation purpose, as oppose to traffic design purpose. More specifically, traffic design considers a "representative" condition where daily variation plays an important role. For example, traffic pattern and driver behavior may vary from day to day, but the design result applies to all situations. In contrast, the daily variation is not important issue in model validation because the goal now is to evaluate the model's ability to replicate a given a condition with reasonable accuracy. Whether the model is able to successfully replicate congestions on Tuesdays or Wednesdays is of minor interest.

7.5 SUMMARY

In this chapter, we prepared for model validation by selecting tests sites and the data for each of the test sites. We used a section of GA 400 as our study site because it meets the considerations we set forth in site selection. We discussed the possible sources of error in the data collected at the study site and further selected three portions of the study site as our test sites, each of which is intended to test a scenario of the model.

For each of the test sites, we need three pieces of data, i.e., geometry data, traffic characteristics data, and time-varying O-D flows. The first two pieces are prepared in this

chapter. For the third piece, we only need to estimate the O-D flows for the test site intended to test the network scenario because the O-D flows of the other two tests sites can be synthesized easily from the observed link flows. The above-mentioned O-D estimation will be the topic of the next chapter.

To facilitate the preparation of time-varying O-D flows, we selected two days' data for each of the test sites and developed an experiment design, based on which we perform the model validation.

CHAPTER 8

DYNAMIC ORIGIN-DESTINATION (O-D) ESTIMATION

The proposed model requires three pieces of information as its input: (1) geometry data; (2) traffic characteristics data; (3) time-varying O-D flows. This chapter deals with the third piece. The model assumes accurate O-D flows. However, such information is rarely readily available and we have to estimate it somehow. Considering that O-D estimation is prone to estimation error which is exogenous to our model. This error will eventually be explained by our modeling error since it is difficult to separate the two sources of errors. Therefore, it is helpful to evaluate the performance of the model with and without the presence of O-D estimation and this motivates the idea of empirical tests on three scenarios: a merging scenario and a diverging scenario which do not require O-D estimation and a network scenario which does.

In the previous chapter, we have chosen three test sites (one for each of the scenarios mentioned above) as well as two sets of test data for each of the test sites. The purpose of this chapter is to estimate the time-varying O-D flows for the last test site (i.e., the one for testing the network scenario) using the selected data sets. The estimated O-D flows will be used as part of the input to the proposed model to test the network scenario in the next chapter.

8.1 REVIEW AND SELECTION OF EXISTING O-D ESTIMATORS

To prepare O-D flows, we need an O-D estimator. Since O-D estimation is not part of our model development but a necessary step to provide input data to the model, we do not develop a new O-D estimator. With a limited goal, we examine the existing O-D estimators and select one that fits our need and implement the selected estimator to prepare the O-D flows.

8.1.1 Early O-D Estimation Approaches

There are many approaches to estimate O-D flows. Early approaches include household and/or roadside survey, license plate survey, aerial photographs, etc. These direct approaches are typically time- and resource-consuming and have generally given their way to indirect methods which estimate O-D flows from link traffic counts. Since the number of possible O-D pairs is typically much more than available constraints, the O-D estimation problem is usually under-determined and additional information or assumption has to be employed to find unique solution to the problem and this partly explains why there exists a rich literature in this field. O-D estimators can be roughly classified as in 8.1.2 – 8.1.8 based on their estimation approaches:

8.1.2 Least Squares

In least squares (LS) estimation, the unknown values of the split parameters (the proportions, in percent, of flow from each of a set of origins to each of a set of destinations) are estimated by finding numerical values for the parameters that minimize the sum of the squared deviations between the observed responses and the functional

portion of the model. Cremer and Keller (1987), Davis and Yu (1994), Hai. (1995), Sherali, et al (1997), Sun and Porwal (2000), Dixon (2000), Sherali and Park (2001), and Ashok and Ben-Akiva (2002) are examples of this approach.

8.1.3 Kalman Filtering

The Kalman filter is a set of mathematical equations that provides an efficient computational (recursive) means to estimate the state of a process, in a way that minimizes the mean of the squared error. Nihan and Davis (1987), Cremer and Keller (1987), Zijpp and Hamerslag (1993, 1994), Madanat, et al (1995), Ikenoue, et al (1995), Hu (1996), Bhattacharjee, et al (1998a), (Bhattacharjee, et al (1998b), Kang (1999), Dixon. (2000), and Sun and Porwal (2000) are examples of this approach.

8.1.4 Bayesian Estimation

The Bayesian estimation seeks to optimally combine information from information contained in the data in the form of a likelihood function and knowledge that is known at the beginning of the research in the form of a prior. Geva, et al (1982), Lo, et al (1996), and Zijpp (1997) are examples of this approach.

8.1.5 Maximum Likelihood Estimation

In maximum likelihood estimation, the desired probability distribution is the one that makes the observed data most likely, which is obtained by seeking the value of the split parameters that maximizes the likelihood function. Geva, et al (1982), Nihan and Davis (1989), Davis and Nihan (1991), Lo, et al (1996), and Zhang and Maher (1998) are examples of this approach.

8.1.6 Programming and Optimization

O-D estimation problem can be expressed as an optimization problem where one tries to minimize an objective function subject to a set of constraints.

Turnquist and Gur (1979), Cremer and Keller (1987), Brenninger-Gothe, et al (1989), Florian and Chen (1991), Sherali, et al (1994), Hai (1995), Sherali, et al (1997), Li and Moor (1999), Sivanandan, et al (1999), Sherali and Park (2001), and Maher, et al (2001) are examples of this approach.

8.1.7 Neural Networks

This approach uses the learning capability of neural networks to associate origin flows with destination flows and learn the split parameters by means of the weights of the neurons. Chin, et al (1994), Fusco and Recchia (1997), and Yang, et al (1998) are examples of this approach.

8.1.8 Statistical Modeling

Examples of this approach include Zijpp and Hamerslag (1994), Lo, et al (1996), and Hazelton (2000), etc.

8.1.9 Selection of an O-D Estimator

With a limited goal, we select an O-D estimator from the available ones. Several considerations drive the selection process. First, the estimator needs to have reasonable accuracy to reproduce O-D pattern from link traffic counts. Second, the estimator needs to be relatively easy to implement because O-D estimation is not the main objective of

this study. Third, the estimator needs to be suited for real-time operation since our proposed model is expected to migrate to real-time application in the future.

Among the many options, we choose the recursive predictions error (RPE) estimator proposed by Nihan and Davis (1987). The estimator, combining features of recursive method and Kalman filtering method, has many advantages and fits our needs well. First, it is reasonably accurate. Second, it is computationally efficient and easy to implement. Third, it is suited for real-time application because it is self-updated and poses almost no requirement on sampling period which translates to the level of aggregation of sampling data.

8.2 THE RPE O-D ESTIMATOR

For easy reference, the algorithm of the RPE estimator is briefly repeated here. For detailed information, please refer to the original paper (Nihan and Davis 1987).

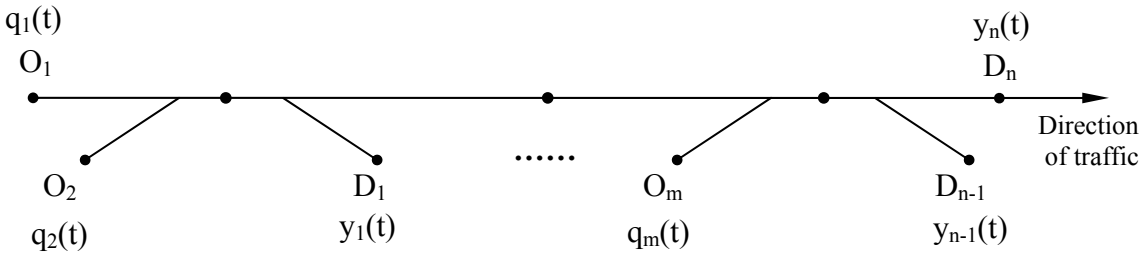


FIGURE 8-1 The sketch of a freeway

Figure 8-1 shows a freeway having m origins (labeled as $O_1, O_2,$ and O_m) and n destinations (labeled as $D_1, D_2,$ and D_n). Let $\vec{y}(t) = (y_1(t), y_2(t), ..., y_n(t))$ is a vector of traffic counts at each of the destinations during time interval t. $\vec{q}(t) = \{q_1(t), q_2(t), ..., q_m(t))$ is a vector of traffic counts at each of the origins during time interval t. $B(t) = \left(b_{i,j}(t)\right)_{m \times n}$ is an $m \times n$ matrix of proportions of trip from origin i to destination j, where $i = 1,2,...,m$ and $j = 1,2,...,n$. Let $\vec{b}_j(t)$ denote the jth column of $B(t)$.

Under ideal conditions, i.e., time interval is sufficiently long and no measurement errors exist, we have:

$$\vec{y}(t) = \vec{q}'(t)B(t)$$

However, ideal conditions are rarely attained in real world, so an error term, $\vec{e}(t) = (e_1(t), e_{12}(t), ..., e_n(t))$, is added:

$$\vec{y}(t) = \vec{q}'(t)B(t) + \vec{e}'(t)$$

This can be viewed as a set of n linear regression equations if we assume $\vec{B}(t)$ constant but unknown. A set of N observations $(\vec{y}(t), \vec{q}(t)), t = 1,2,...,N$ enable least square estimate for $B(t)$, $\hat{B}(t) = (\hat{b}_1(t), \hat{b}_2(t), ...\hat{b}_n(t))$:

$$\hat{b}_j(t) = \frac{\sum_{k=1}^{t} \hat{q}(k) y_j(k)}{\sum_{k=1}^{t} \hat{q}(k)\hat{q}'(k)}, j = 1,2,...,n$$

Since the above estimator involves matrix inversion, which is typically computationally inefficient, the estimate can be replaced by a recursive one if we denote $A(t) = \frac{1}{t}\sum_{k=1}^{t} \vec{q}(t)\vec{q}'(t)$:

$$\hat{b}_j(t) = \hat{b}_j(t-1) + \frac{1}{t} A^{-1}(t)\vec{q}(t)[y_j(t) - \vec{q}'(t)\hat{b}_j(t-1)]$$
$$A(t) = A(t-1) + \frac{1}{t}[\vec{q}(t)\vec{q}'(t) - A(t-1)]$$

Let $P(t) = \frac{1}{t} A^{-1}(t)$, the above equations are transformed to be:

$$\hat{b}_j(t) = \hat{b}_j(t-1) + K_j(t)[y_j(t) - \vec{q}'(t)\hat{b}_j(t-1)]$$
$$K_j(t) = \frac{P_j(t-1)\vec{q}(t)}{1 + \vec{q}'(t)P_j(t-1)\vec{q}(t)}$$
$$P_j(t) = P_j(t-1) - \frac{P_j(t-1)\vec{q}(t)\vec{q}'(t)P_j(t-1)}{1 + \vec{q}'(t)P_j(t-1)\vec{q}(t)}$$

Estimate of this form is sometimes called recursive least-squares, or RPE. It is also related to Kalman filter approach if one defines state equations as:

$$\begin{aligned} \vec{b}_j(t) &= \vec{b}_j(t-1) \\ y_j(t) &= \vec{q}'(t)\vec{b}_j(t) + e_j(t) \end{aligned} \quad j = 1,2,...,n$$

where $e_j(t)$ are independent random variables with 0 mean and variances $r_j(t)$. The least-square estimator can be transformed to a Kalman filter estimator:

$$\hat{b}_j(t) = \hat{b}_j(t-1) + K_j(t)[y_j(t) - \vec{q}'(t)\hat{b}_j(t-1)]$$
$$K_j(t) = \frac{P_j(t-1)\vec{q}(t)}{r_j(t) + \vec{q}'(t)P_j(t-1)\vec{q}(t)}$$
$$P_j(t) = P_j(t-1) - \frac{P_j(t-1)\vec{q}(t)\vec{q}'(t)P_j(t-1)}{r_j(t) + \vec{q}'(t)P_j(t-1)\vec{q}(t)}$$

So far, the estimators discussed deal with constant proportion matrix $B(t)$, which is not realistic in real world. A slight modification enables the above estimators to track time variations. The extension to Kalman filter estimator starts with generalizing state equations:

$$\begin{aligned} \vec{b}_j(t) &= \vec{b}_j(t-1) + \vec{s}_j(t) \\ y_j(t) &= \vec{q}'(t)\vec{b}_j(t) + e_j(t) \end{aligned} \quad j = 1,2,...,n$$

where $\vec{s}_j(t)$ is a vector of random variables with 0 mean and covariance matrices $R_j(t)$. Then the above static Kalman filter estimator can be turned into a time-varying one:

$$\begin{aligned} \hat{b}_j(t) &= \hat{b}_j(t-1) + K_j(t)[y_j(t) - \vec{q}'(t)\hat{b}_j(t-1)] \\ K_j(t) &= \frac{P_j(t-1)\vec{q}(t)}{r_j(t) + \vec{q}'(t)P_j(t-1)\vec{q}(t)} \\ P_j(t) &= P_j(t-1) + R_j(t) - \frac{P_j(t-1)\vec{q}(t)\vec{q}'(t)P_j(t-1)}{r_j(t) + \vec{q}'(t)P_j(t-1)\vec{q}(t)} \end{aligned}$$

Note that, to initialize and run the recursive estimator, this estimator requires *a priori* information about $\hat{b}_j(0)$, $P(0)$, and $R_j(t)$ which is typically unavailable. A common practice is to treat them as design parameters and choose their values empirically.

As a last point, there are some natural constraints that the O-D proportions of $B(t)$ must satisfy:

$$\begin{aligned} &\sum_{j=1}^{n} b_{ij}(t) = 1 && i = 1,2,...,m \\ &b_{ij}(t) \geq 0 && i = 1,2,...,m;\ j = 1,2,...,n \\ &b_{ij}(t) = 0 && (i,j) \in Z \end{aligned}$$

The first natural constraint requires that proportions from all origins to a destination sum up to 1. The second natural constraint eliminates negative flow from any origin to any destination. The third natural constraint removes infeasible O-D flows by enforcing 0 to those proportions which represent flows that can never happen, such as flow from a downstream entrance to an upstream exit.

8.3 APPLICATION TO TEST CASES

In this section, we implement the RPE estimator and perform O-D estimation based on the two data sets for testing the network scenario.

8.3.1 The Test Site

As discussed in Chapter 7, there are three test sites in this study. Only test site 3 requires O-D estimation. To facilitate discussion that follows, the geometry of the test site is repeated in Figure 8-2.

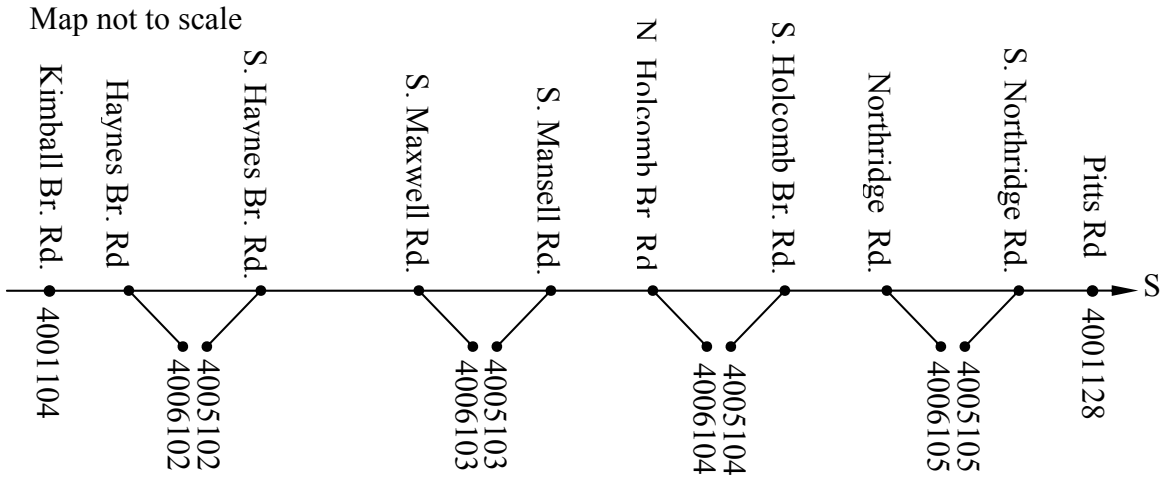


FIGURE 8-2 Origins and destinations of test site 3 (GA 400 southbound)

There are 5 origins/entrances and 5 destinations/exits at this test site. The origins are: Station 4001104 at Kimball Br. Rd, Station 4005102 at S. Haynes Br. Rd., Station 4005103 at S. Mansell Rd, Station 4005104 at S. Holcomb Br. Rd., and Station 4005105 at S. Northridge Rd. The destinations are: Station 4006102 at Haynes Br. Rd, Station 4006103 at S. Maxwell Rd, Station 6006104 at N. Holcomb Br. Rd., Station 4006105 at Northridge Rd., and Station 4001128 at Pitts Rd..

As one of the natural constraints, the possible O-D flows are indicated in Table 8-1 where a check mark represents feasible O-D pair and a cross represents infeasible O-D pair. The estimation result is discussed as follows.

TABLE 8-1 Possible O-D flows of test site 3

Origin. \ Destin.	4006102	4006103	4006104	4006105	4001128
4001104	√	√	√	√	√
4005102	×	√	√	√	√
4005103	×	×	√	√	√
4005104	×	×	×	√	√
4005105	×	×	×	×	√

8.3.2 Initial Estimation Results

In Chapter 7, two data sets are identified for this test site, i.e., the data collected on Friday, Oct. 11, 2002 from 05:50:00 to 19:20:00 and the data collected on Monday, Oct. 14, 2002 from 5:52 to 19:57. The original data were in 20-second intervals. We merged the data into 5-minute intervals primarily for smoothing out local variations. We perform O-D estimation based on the merged data sets.

The estimation result is a time-varying matrix of proportions, in percent, of flow from each of a set of origins to each of a set of destinations. Based on the flows at the entrances as well as this matrix, the flows at the exits can be synthesized. The performance of the estimator is assessed by comparing the estimated and observed exit flows. Table 8-2 lists a segment of sample comparison of observed and estimated exit flows at exit 4006104. In this table, each time step represents 5 minutes. The percent error is the result of dividing the absolute error (the predicted minus the observed) by the observed flows.

TABLE 8-2 Sample comparison of observed and estimated exit flows

Data are for exit 4006104 and each time step represents 5 minutes

Time Step	Predicted (veh/hr)	Observed (veh/hr)	% Error
…	…	…	…
52	395	528	-0.25
53	484	684	-0.29
54	586	780	-0.25
55	533	756	-0.29
56	561	744	-0.25
57	494	624	-0.21
58	539	684	-0.21
59	509	888	-0.43
60	398	672	-0.41
61	656	768	-0.15
62	607	696	-0.13
63	652	792	-0.18
64	583	780	-0.25
65	410	696	-0.41
66	766	816	-0.06
67	700	912	-0.23
68	609	852	-0.29
…	…	…	…

Figure 8-3 shows the full picture at the station where the X axis is time step and the Y axis is flow in veh/hr. The solid line is the observed flow and the dashed line is the estimated flow.

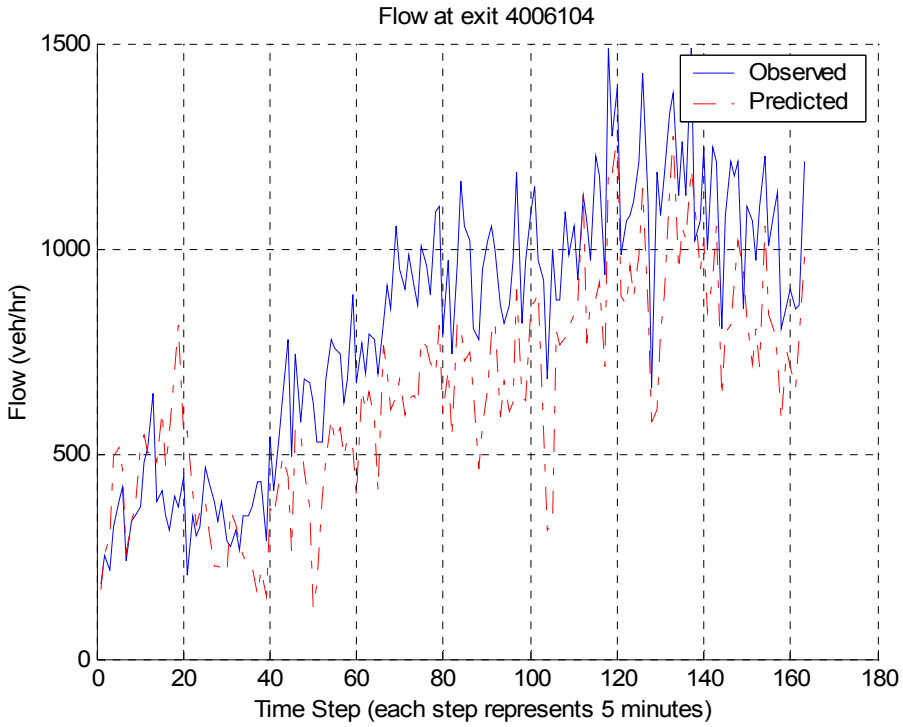


FIGURE 8-3 Sample comparison of observed and estimated exit flows

Both the table and the figure show that the estimated curve captures the variation of the observed curve well. We also find that the predicted curve is lower than the observed curve and this is consistent at other stations. Therefore, we must correct this systematic error before applying the estimation results to model validation.

8.4 SYSTEMATIC ERROR CORRECTION STRATEGY

This section presents our proposed strategy and the results of applying the strategy to the initially estimated O-D flows.

8.4.1 The Proposed Strategy

To deal with the O-D estimation error, we propose a post-processing calibration-based strategy, as shown in Figure 8-4. The observed link flows, including the entry flows and the exit flows, are used to estimate the time-varying O-D matrix. The O-D flows, which are the input to our model, can be computed based on the observed entry flows and the O-D matrix. After a synthesis, we obtain the predicted exit flows. Then the difference between the observed and the predicted exit flows is used as a feedback to adjust the O-D matrix. The O-D flows are then updated based on the adjusted O-D matrix and the resulted O-D flows are what we actual used to input to our model in model validation. Though the O-D estimation error may not be totally eliminated, this adjustment helps minimize the impact of the O-D estimation error on subsequent model validation.

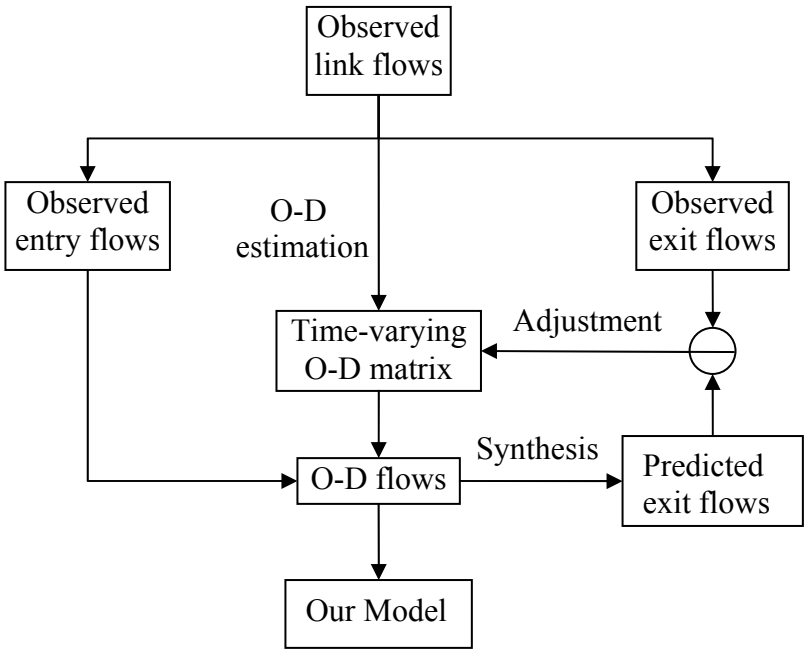


FIGURE 8-4 Strategy to minimize the impact of o-d estimation error

8.4.2 Adjustment Results

The comparison of the observed and the predicted (after adjustment) exit flows are presented as follows. Figure 8-5 shows the result for data collected on Friday, Oct. 11, 2002 from 05:50:00 to 19:20:00 and Figure 8-6 shows the result for data collected on Monday, Oct. 14, 2002 from 5:52 to 19:57.

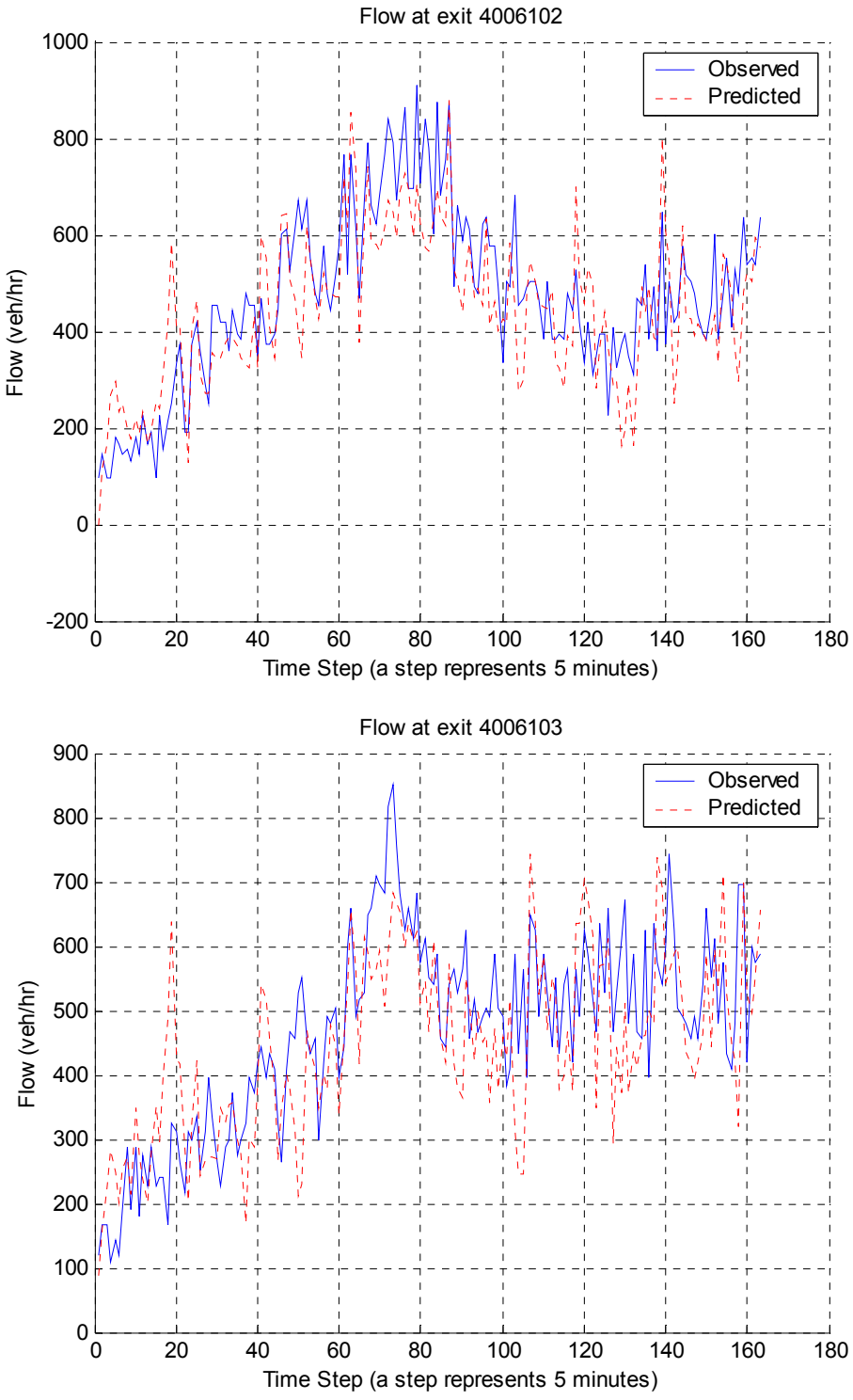
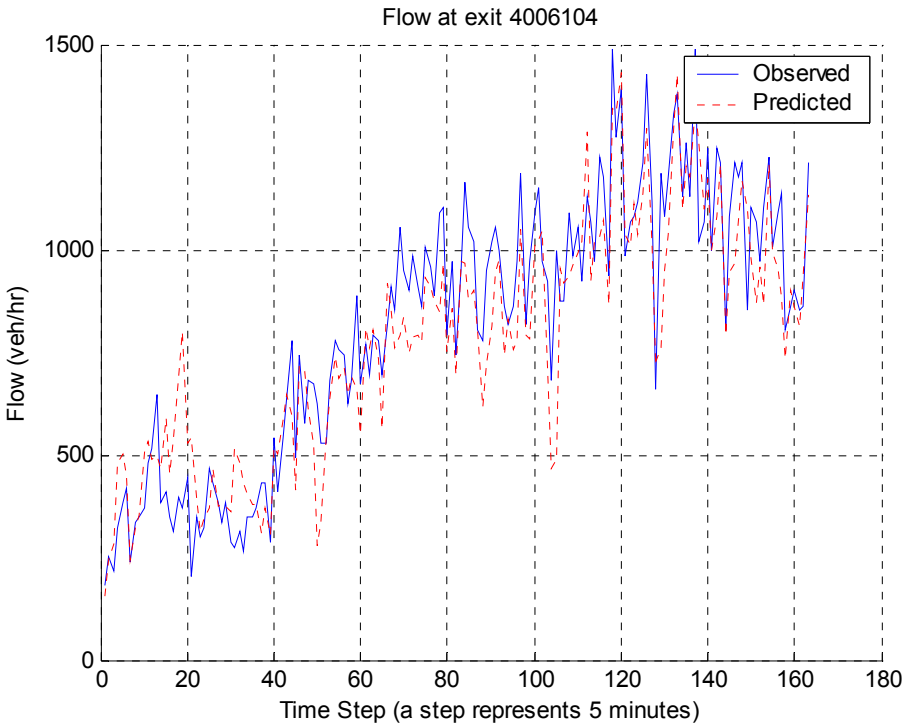


FIGURE 8-5 O-D estimation result (Oct. 11, 2002)



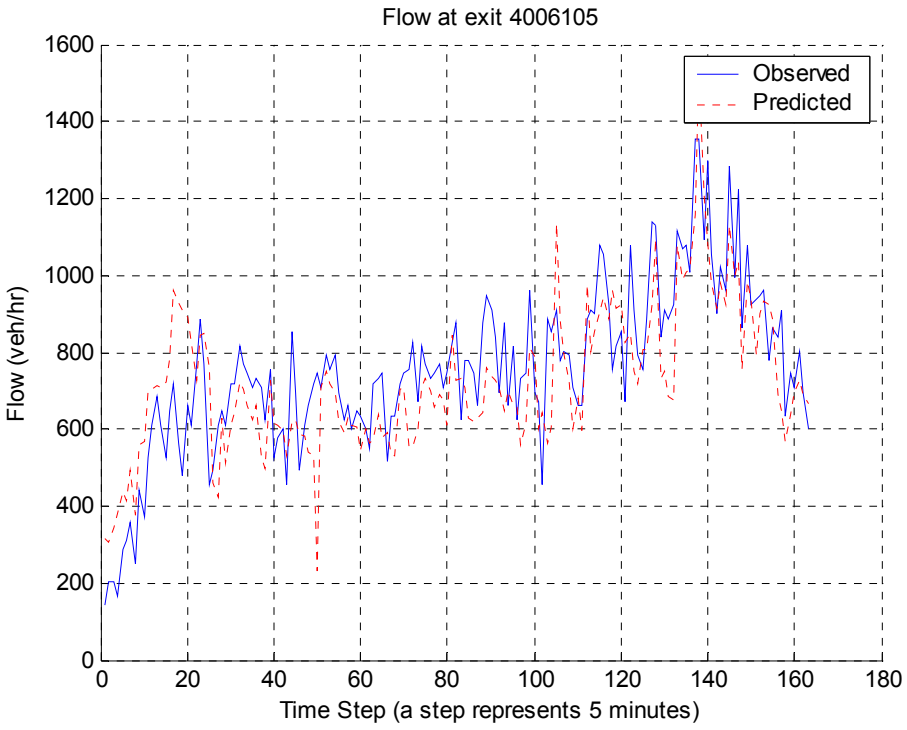


FIGURE 8-5 Continued

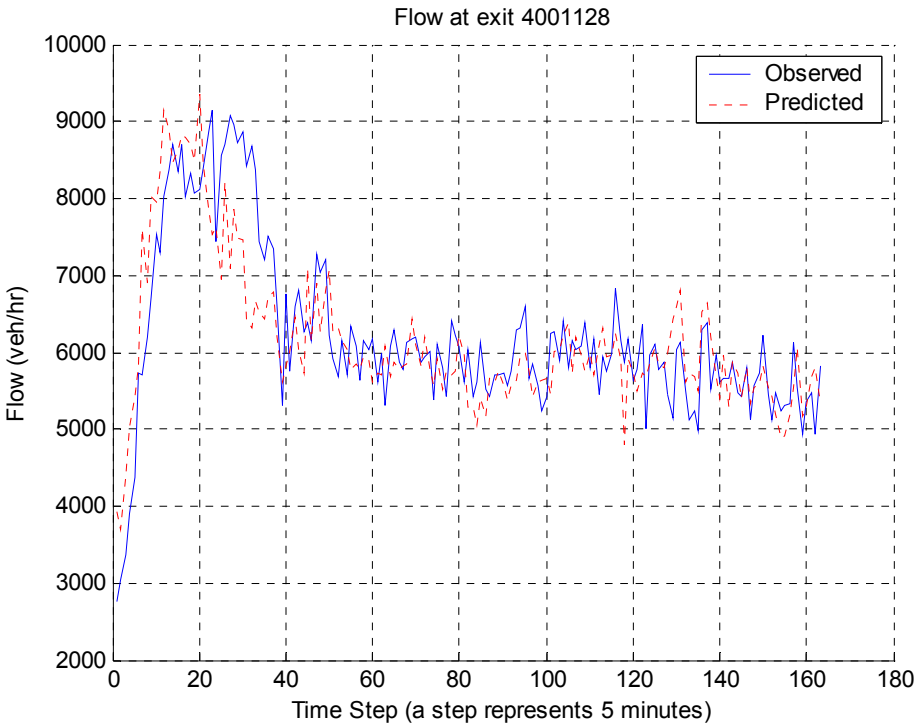


FIGURE 8-5 Continued

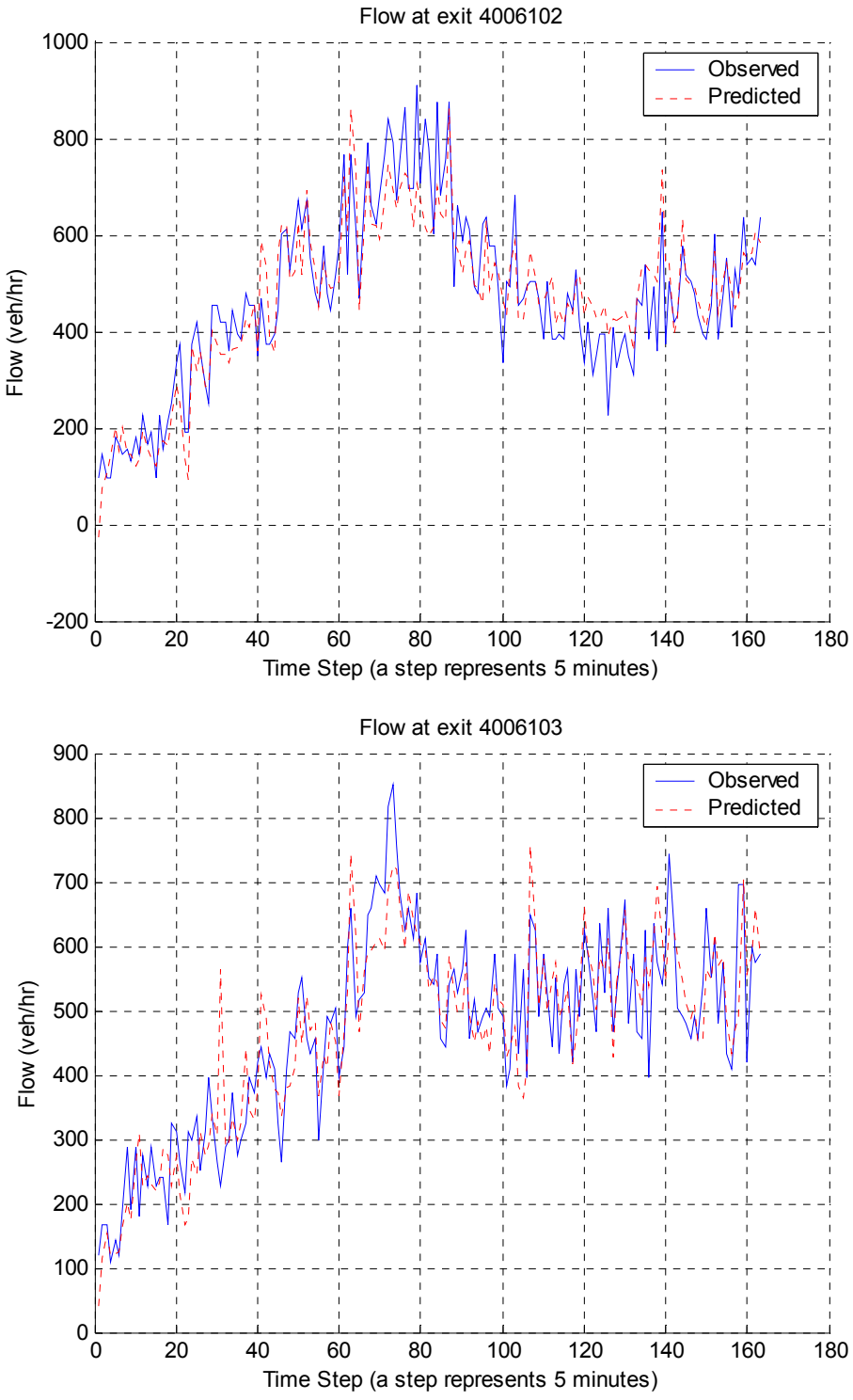
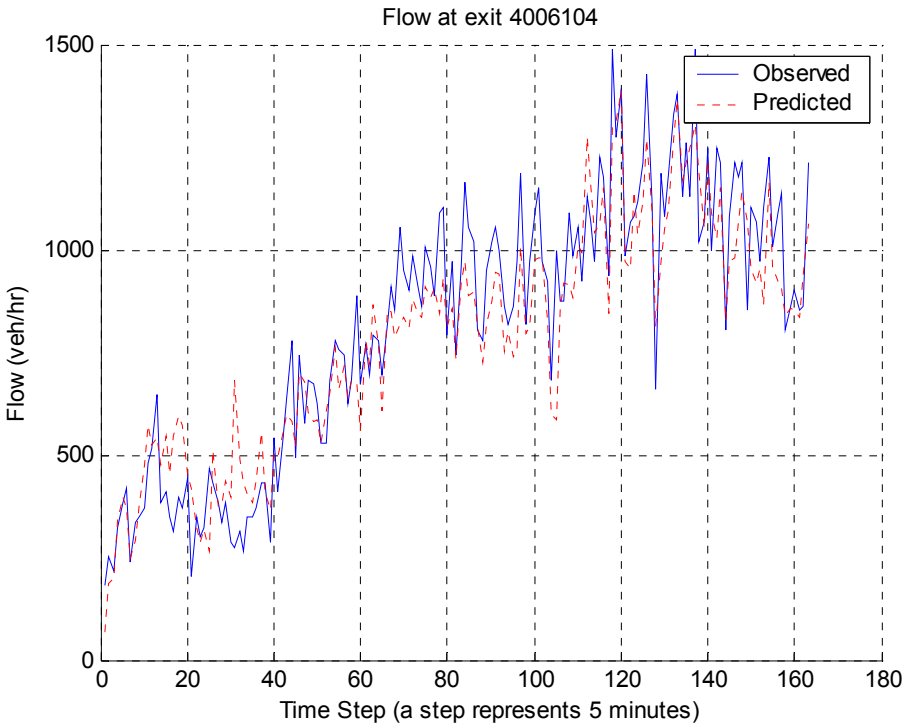


FIGURE 8-6 O-D estimation result (Oct. 14, 2002)



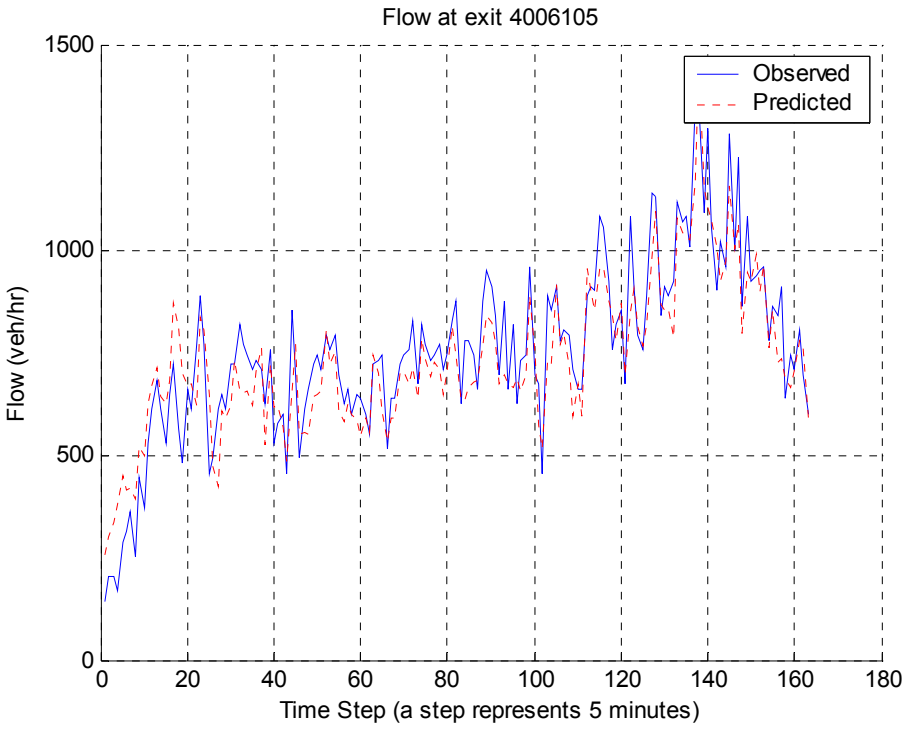


FIGURE 8-6 Continued

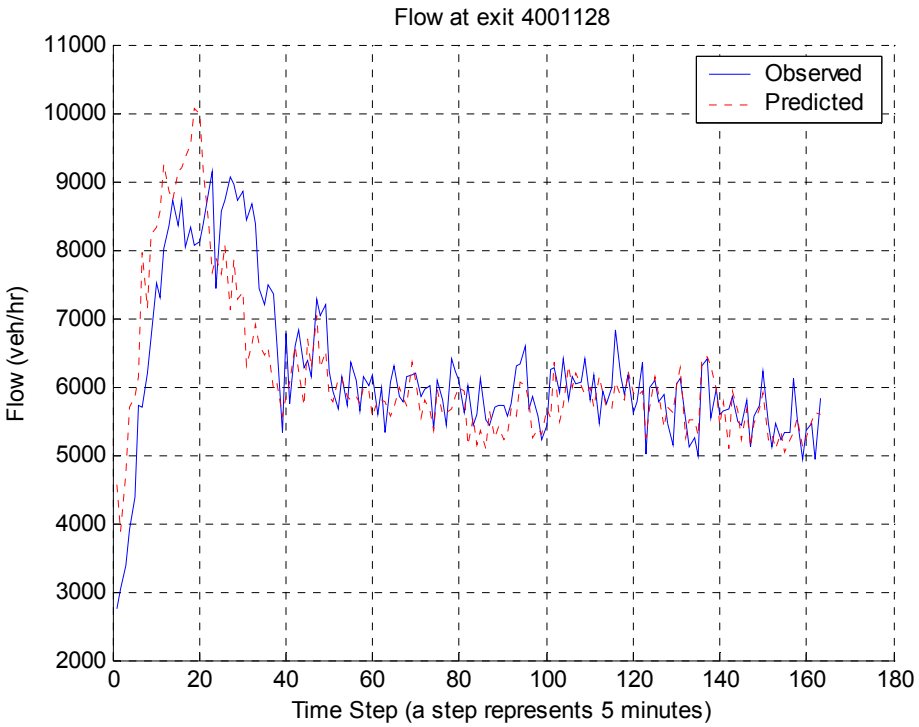


FIGURE 8-6 Continued

Statistical analysis shows that, in all cases, the observed and the predicted exit flows are not statistically different and the 95% confidence interval falls within ±4.0%. This result is quite satisfactory and makes the adjusted O-D flows suitable for feeding into the model validation process. As a reminder, the above O-D flows are calibrated/adjusted based on field observations and this calibration is not part of the selected O-D estimator.

8.5 SUMMARY

With a limited goal in this chapter, we selected an existing O-D estimator, implemented it, and used it to estimate the O-D flows of the test site for the network scenario. The estimated O-D flows are part of the input information to our proposed model.

Based on the estimation results, we found that estimation error is considerable and systematic. We proposed a strategy to minimize the impact of the O-D estimation error by making adjustments on the initial estimation results using the difference between the observed and the predicted exit flows. The adjusted O-D flows will be used to perform model validation in the next chapter.

CHAPTER 9

RESULTS OF MODEL VALIDATION

The purpose of this chapter is to validate the proposed model by performing the empirical tests described in Chapter 6 to the test sites and test data prepared in Chapter 7 and 8. These empirical tests compare the model output against field observations and this chapter presents the results obtained for each of the testing scenarios, i.e., the merging scenario, the diverging scenario, and the network scenario. The first two scenarios do not involve O-D estimation, but the last scenario does. The last scenario will offer insight in model performance with O-D estimation.

9.1 INTRODUCTION

To highlight the structure of our presentation of the empirical test results, the experiment design of the empirical tests is partly repeated here (please refer to Section 7.5 in Chapter 7 for complete information).

Test Site	Testing Scenario	Date	O-D estimation
1	Merging	10/14/2002, 9/6/2002	Not Required
2	Diverging	9/12/2002, 12/9/2002	Not Required
3	Network	10/11/2002, 10/14/2002	Required

We have three test sites, i.e., test sites 1, 2 and 3. We also have three scenarios to test, i.e., the merging, the diverging, and the network scenarios. We use the three tests sties to test the three scenarios, respectively. For each scenario, we perform two tests based on the two days' data, respectively. There is no O-D estimation involved in the tests for merging and diverging scenarios, but O-D estimation is involved in the tests for the network scenario.

When presenting the test results, we follow the following sequence: For each test site/scenario, we first briefly repeat the topological structure of the test site. For each test day of the scenario, we present qualitative results and quantitative results. For qualitative results, we present the following items:

- Comparison of prediction and observation in time series
- Comparison of prediction and observation in terms of congestion
- Comparison of prediction and observation in 3D space
- Plot of prediction against observation
- Frequency of modeling error
- Comparison of Predicted and Observed Flow-Density Relationships

For quantitative results, we present the results of the simultaneous statistical test.

To keep this chapter concise and focused, we only discuss the results of the first empirical test in detail and we generally highlight the results of other empirical tests since the latter follow the same pattern of presentation as the former.

9.2 EMPIRICAL TEST - MERGING SCENARIO

Information regarding test site 1 (for testing the merging scenario) has been detailed in Section 7.3.2 of Chapter 7 and is not repeated here. The link-node structure of the test site illustrated in Figure 7-3 is based on the locations of surveillance stations. In simulation runs, this structure is not literally followed because some of the adjacent links are similar in roadway geometry and traffic characteristics and hence can be merged.

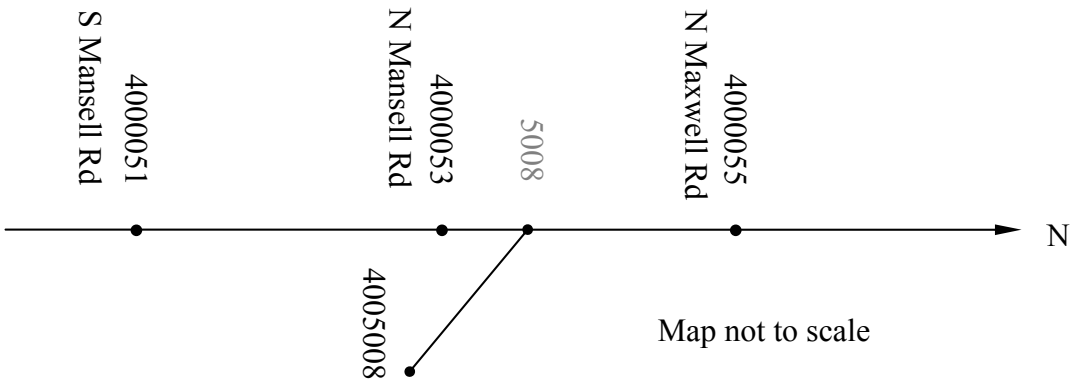


FIGURE 9-1 Structure of test site 1 - merging scenario (GA 400 northbound)

Figure 9-1 shows the structure of the test site that is actually used in the tests. It consists of two upstream mainline links (4000051-4000053 and 4000053-5008), one downstream mainline link (5008-4000055), and an on-ramp link (4005008-5008). All mainline links have 3 lanes with approximately the same capacity. The merging point, node 5008, might be a bottleneck because the capacity of its downstream link (5008-4000055) is less than the sum of its upstream links (4000053-5008 and 4005008-5008).

Another potential bottleneck is the downstream of node 4000055 because queues might build up from further downstream and back up into our test site which is allowed in the simplified theory as long as they do not exceed the upstream end of the site.

9.2.1 Test Day 1

The data for test day 1 was collected on Monday, October 14, 2002 from 5:52 to 19:57. Comparison of model prediction and field observation, based on qualitative as well as quantitative measures, are presented in the following subsections.

9.2.1.1 Qualitative evaluation

Qualitative assessment of the performance of the model can be made based on visual comparison of the observed and predicted densities. This section presents a series of plots to graphically illustrate the observed and predicted densities from various perspectives. Again, our goal here is to check whether the model is able to predict traffic density in time-space domain with acceptable accuracy, especially the ability to capture major peaks, if any.

Comparison of prediction and observation in time series

Figure 9-2 shows the density vs. time curves at each link in test 1. The solid line is observed density and the dashed line is predicted density. The X axis is time of day and the Y axis is density in veh/mi/ln. Each part of the figure consists of two subplots with the density plot being the upper one and the schematic test site map the lower one. The schematic map shows the current link with circles at both ends of the link.

The simulation starts at 5:52 and ends at 19:57. There were two peaks on this day, one in the morning and the other in the afternoon. Both originated from the downstream of node 4000055, probably due to insufficient capacity there. Since the real bottleneck was located outside of our test site, a dummy link is added at the downstream of node 4000055 and a time varying capacity was applied at the dummy link to simulate the effect of the real bottleneck. If one views the plot in a reverse order from Part C to Part A, the impact of the peaks decreases and eventually disappears. Notice that there was no peak at the on-ramp link 4005008-5008 because its share of downstream capacity sufficed its demand, which is exactly what the merging scenario predicts.

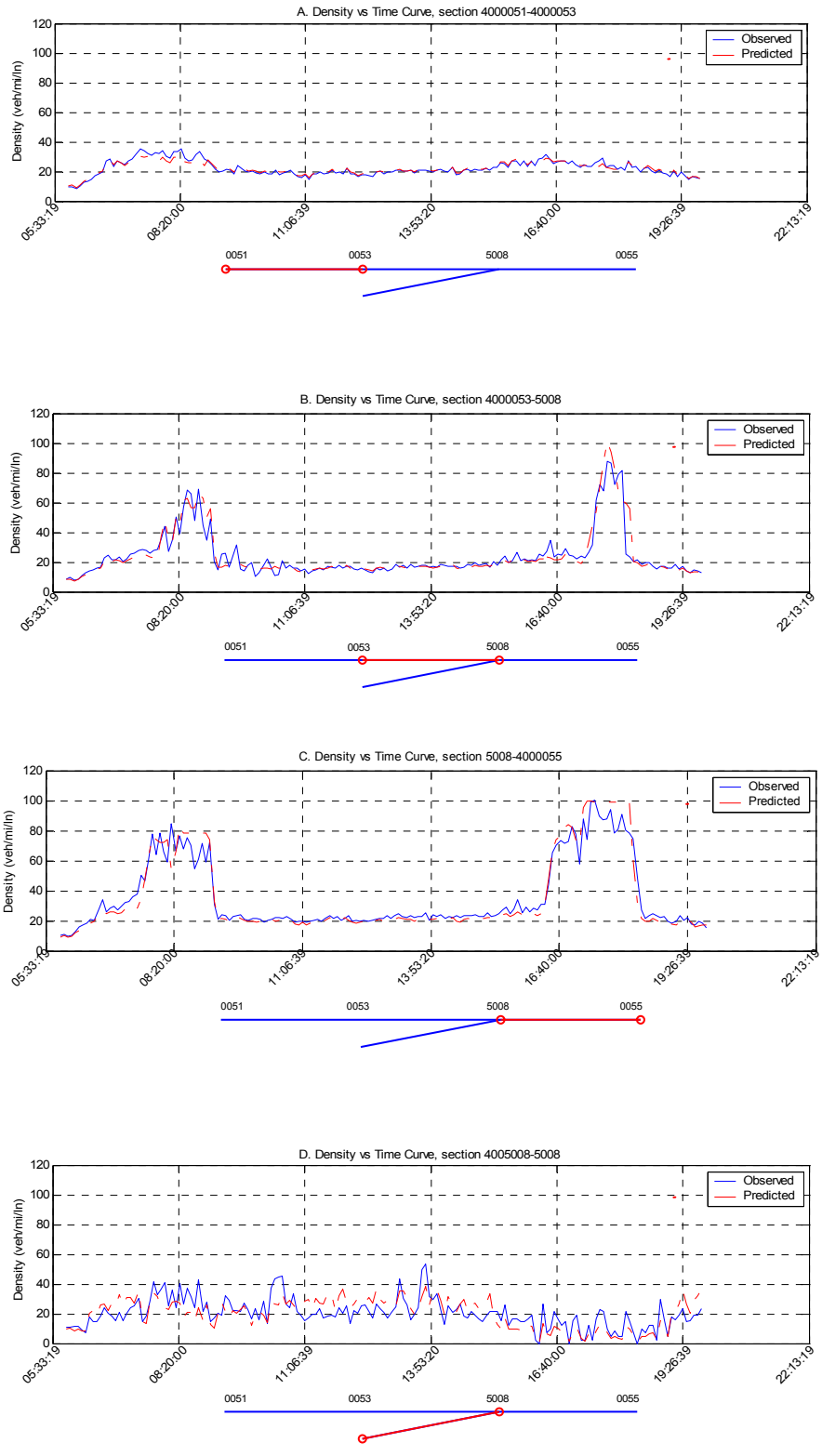


FIGURE 9-2 Time series comparison of density (merging scenario, day 1)

Comparison of prediction and observation in terms of congestion

Figure 9-3 shows the density contour of the day based on a density level of 45 veh/mi/ln. This density level is used in this study to delineate the boundary of congestion primarily due to the following three reasons. First, HCM 2000 uses a density level of 45 pc/mi/ln as the upper bound of level of service (LOS) F, and this density level typically signifies the breakdown of traffic operation. Second, Son (1996) suggest that the actual boundary of the congested region is most likely to be somewhere between 40 and 50 veh/mi/ln and 45 veh/mi/ln is the middle of this range. Third, the contour lines of 40 and 50 veh/mi/ln are very close to each other in most of the cases, and plotting contours of both levels does not make too much sense.

In Figure 9-3, the X axis represents location (nodes along the freeway mainline) and the Y axis represents time of day. There are two congested regions backing up from downstream node 4000055 towards upstream node 4000051. The solid line is the observed contour and the dashed line is the predicted contour. In this figure, we know the exact information at observation stations (e.g., 4000051, 4000052, 4000053, etc.) Information regarding queue development and queue tail, i.e., the portion of the lines between stations, is an approximation. In general, the figure shows a very good fit between the prediction and the observation.

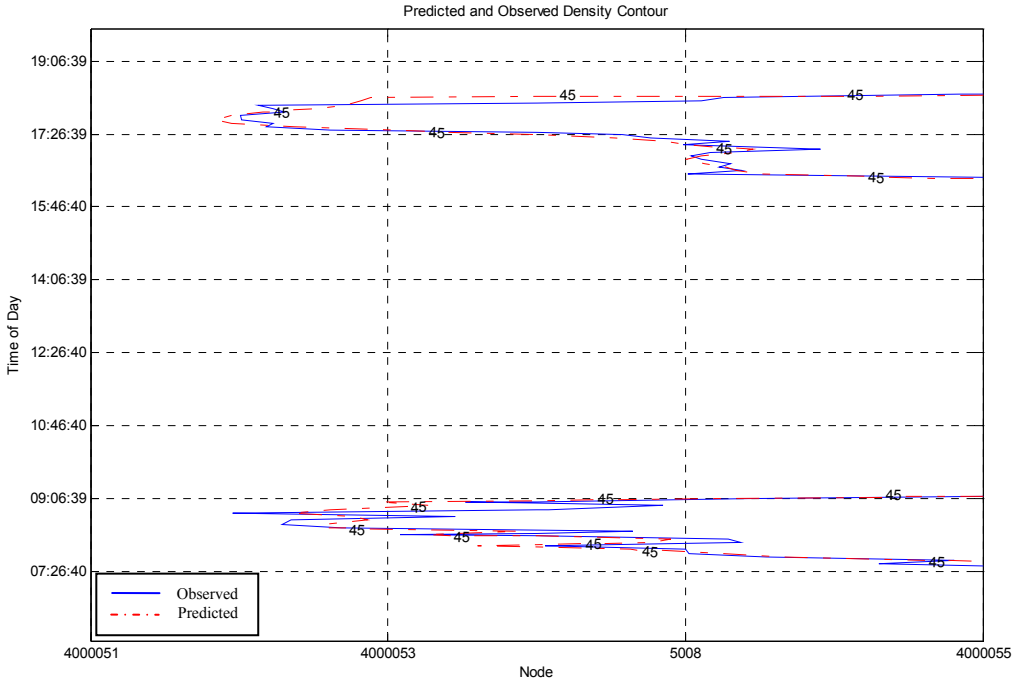


FIGURE 9-3 Comparison of congestion (merging scenario, day 1)

Generally, it is preferable to have the congested regions contained by the bounding box of a time-space diagram, i.e., no congestion appears at the upstream end and downstream end and no congestion occurs at the beginning and the end of the test. Such examples, however, are difficult to find in real world because it generally requires observation of a larger area over longer period so that congestion is unlikely to reach the boundaries. This is particularly difficult for merging and diverging scenarios since these scenarios deal only with a ramp junction. Fortunately, the simplified theory does not prohibit congestion backing up from downstream end as long as it does not exceed upstream end. Otherwise, there is a problem because traffic is now operating at the congested side of the underlying flow-density curve and the arrival flow no longer

represents the true demand. If this is the case, one generally goes further upstream, trying to find a node where congestion never reaches, so that the congestion is still contained in the new bounding box.

Comparison of prediction and observation in 3D space

If a time series plot gives one-dimensional information of density in time domain and a contour gives two-dimensional information of density in time-space domain, density surface presents three-dimensional information of density in time-space-depth domain. In Figure 9-4, the X axis represents location, the Y axis represents time of day, and the Z axis represents depth (density levels) in veh/hr/ln. The first plot shows the predicted density surface and the second shows the observed. The Figure generally confirms our findings in previous steps – the timing, the location, and the shape of the peaks agree well.

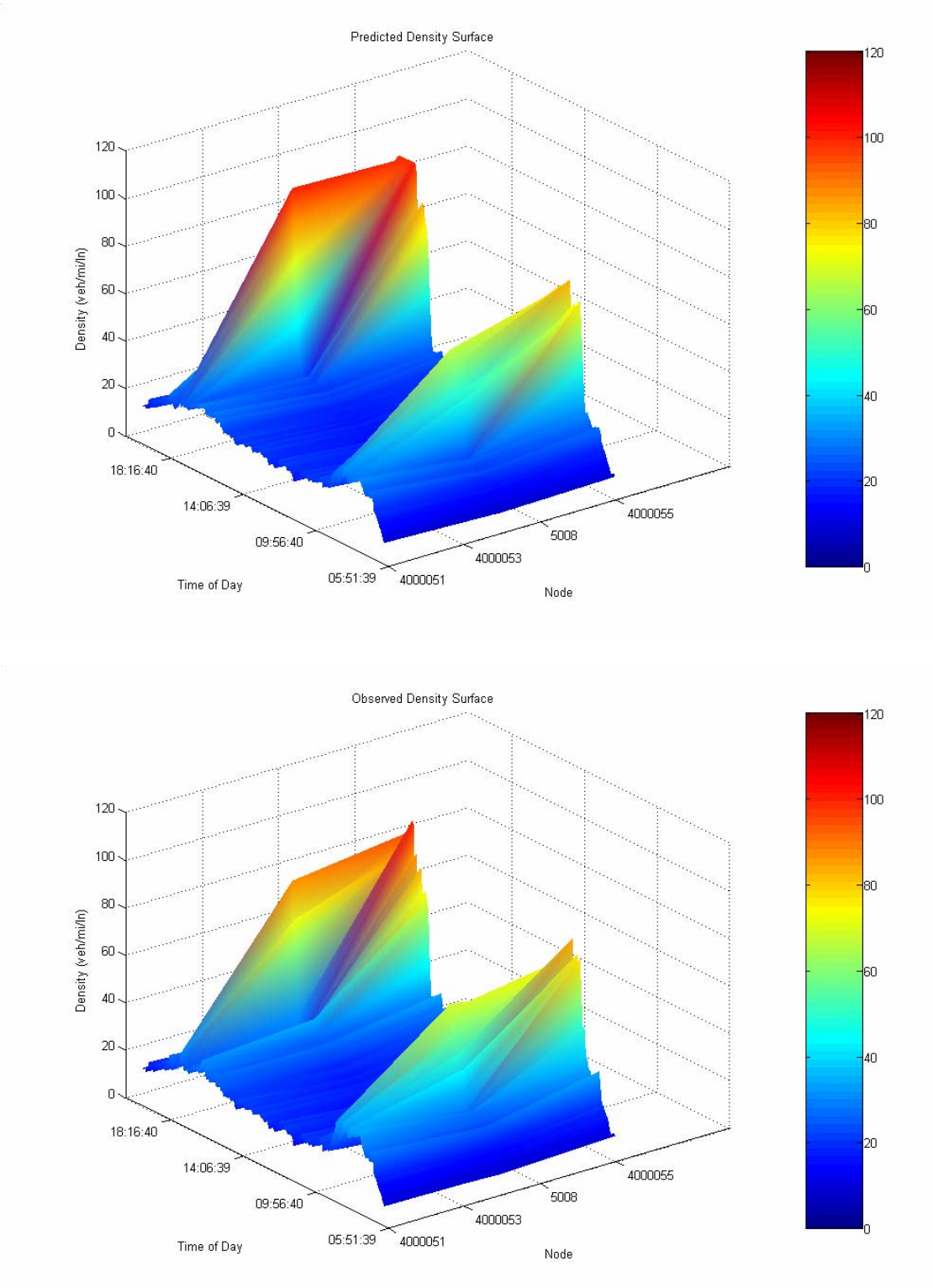


FIGURE 9-4 Comparison of density in 3D space (merging scenario, day 1)

Plot of prediction against observation

In a diagonal plot, all samples are put together and the predicted density values are plotted against the observed density values. An ideal fit in the diagonal plot would be a 45 degree line running from the origin to the upper right corner. Figure 9-5 generally shows such a trend. There are 676 samples in total, most of which densely distribute along the line and a linear relationship is quite apparent. At higher density values, the trend still exists but the points are a little bit dispersed. However, this does not necessarily imply that the model performs inferior at high density. First, there are only a few samples at high density and they might not be enough to make their statistical property pronounced. Second, it is likely to get larger modeling error at higher density level given that the accuracy of the model does not deteriorate as density increases.

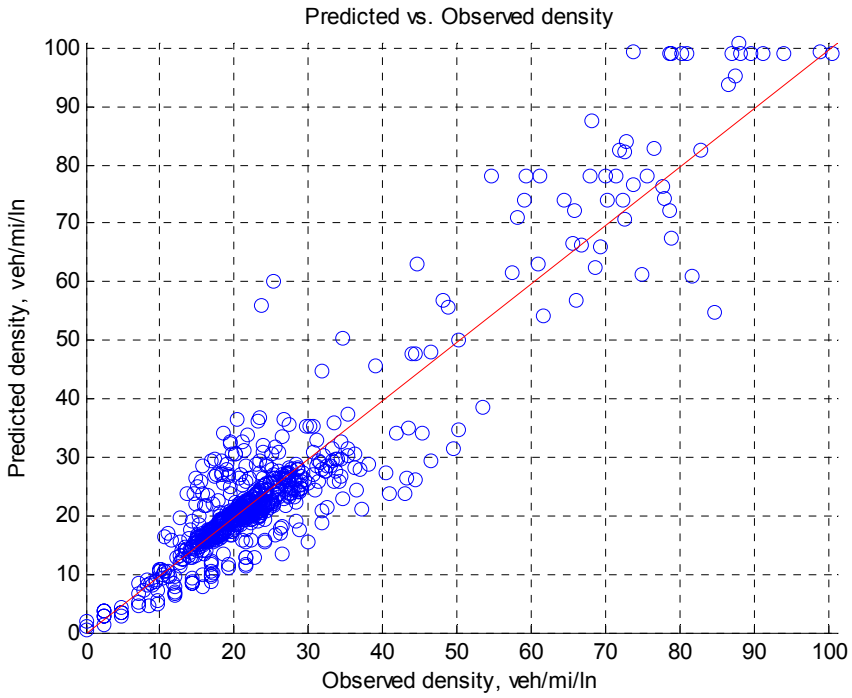


FIGURE 9-5 Plot of prediction against observation (merging scenario, day 1)

Frequency of modeling error

Figure 9-6 illustrates the frequency of modeling error, i.e., the residuals resulted after subtracting the observed values from the predicted values. As expected, residuals are densely concentrated around zero with the rest balanced at both sides – a bell-shaped distribution. There are 676 samples in total. The highest bar represents 466 samples, accounting for 69% of the total. The number between ±10 makes up approximately 90% of the total samples.

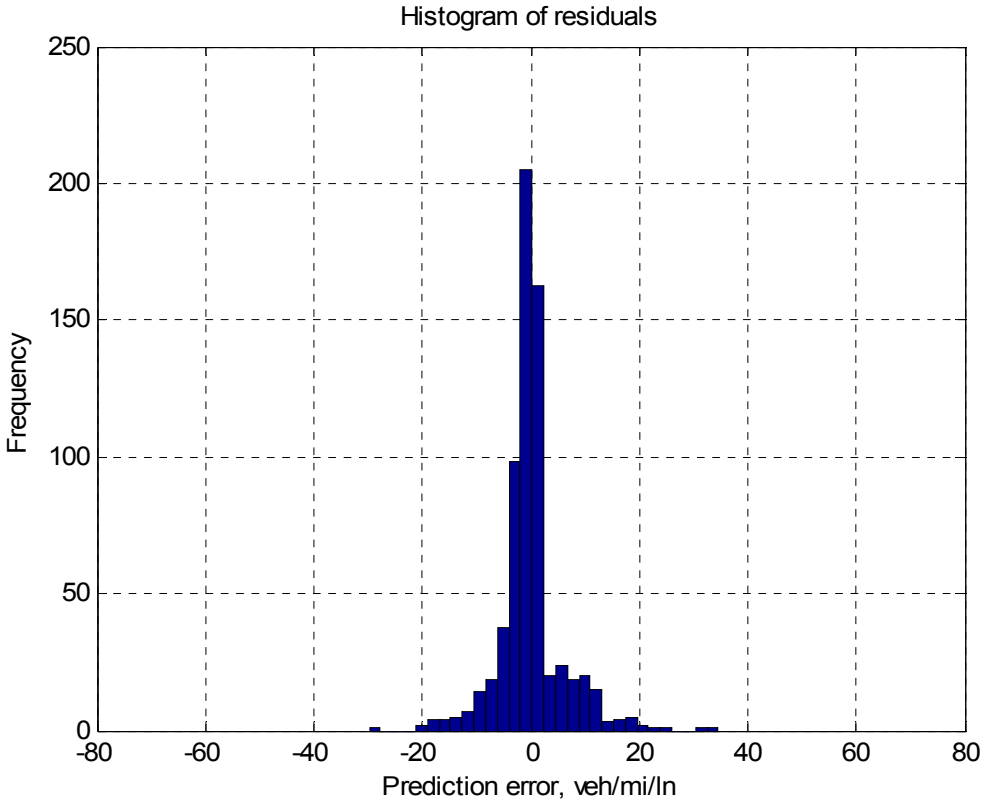


FIGURE 9-6 Frequency of modeling error (merging scenario, day 1)

Comparison of Predicted and Observed Flow-Density Relationships

Figure 9-7 plots the predicted and the observed flow-density relationships on the same figure. It reveals that the traffic in a simulated environment behaves in a similar way that the traffic in a real system does and it confirms that, though the input flow-density relationship is a triangular (or trapezoidal) type, the output relationship is not.

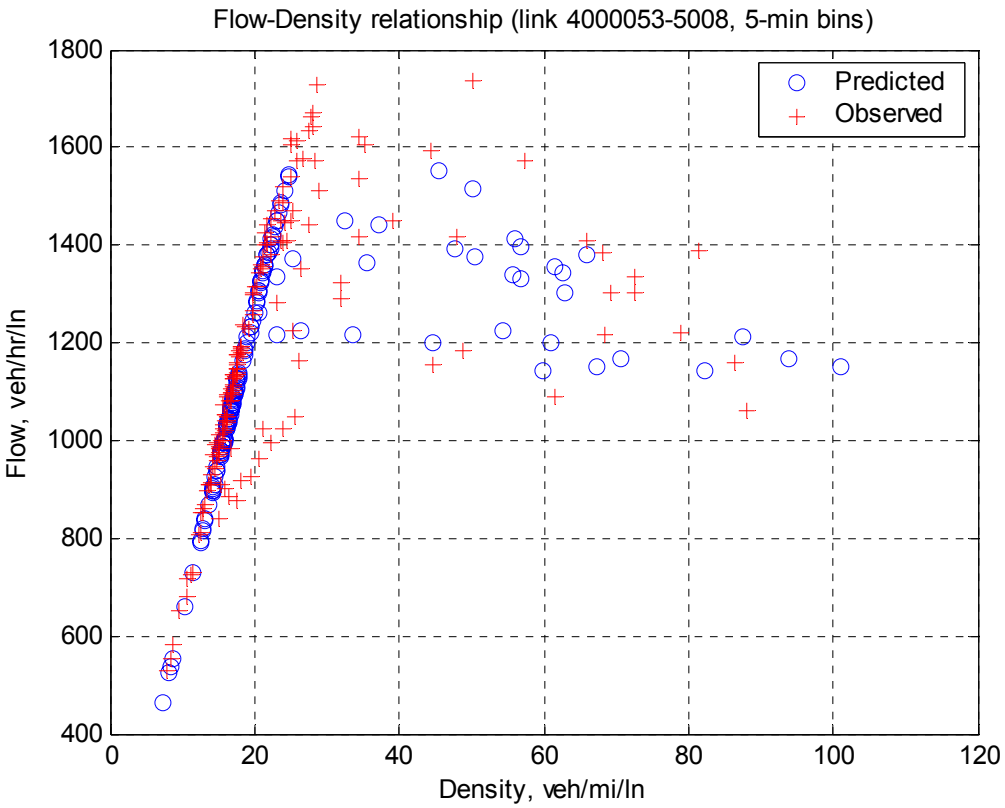


FIGURE 9-7 Comparison of flow-density relationships (merging scenario, day 1)

9.2.1.2 Quantitative evaluation

The simultaneous statistical inference technique discussed above is applied. To conduct the test, values of number of batches and batch sizes needs to be determined.

Here, our major goal is to have a good coverage of confidence interval, so a large batch size is desirable. On the other hand, minimizing the estimation error of batch means variance generally requires large number of batches. After many experiments and comparison, we choose to have 8 batches with each batch having 84 samples, and this setting seems to take care of both of the competing criteria well. Table 9-1 summarizes the result of the simultaneous statistical test.

TABLE 9-1 Simultaneous statistical test result of test day 1 (merging scenario)

Item	Value
Batch size (m)	84
Batches (b)	8
Total samples (n)	676
Significance level (α)	0.05
Grand mean ($\overline{W}$)	-0.025097
Variance ($\hat{V}_B$)	0.001954
T-test Statistic (t_0)	-1.605861
Critical Value ($t_{0.025,7}$)	2.36462
T-test result	Fail to reject H_0, i.e., $E[\overline{W}_i] = 0$.
Small number (ε)	0.00097
Critical χ^2 value ($\chi^2_{0.05,7}$)	14.07
95 % confidence interval for percentage modeling error	(-0.084, 0.038) $\times 100\%$

Interpreting of the result can be as follows. If we compare the predicted density with the observed density based on batches of the specified size, the difference of the two will be zero approximately 95 out of every 100 trials, i.e., the mean of the modeling errors is not statistically different than 0. On the other hand, if we accept that ε is a reasonably

small number, the variance of the ratio of the prediction and observation will be less than or equal to this number approximately 95 out of every 100 trials. This also translates to a 95% confidence interval of (-0.084, 0.038) $\times$100% for percentage modeling error.

9.2.2 Test Day 2

The data for test day 2 of the merging scenario was collected on Friday, September 6, 2002 from 00:01 to 23:51, almost a whole day. Description of the tests thereafter generally follows the format of the test presented above. To be concise, we only highlight the results of the tests that follow and skip much of their discussion.

9.2.2.1 Qualitative evaluation

Comparison of prediction and observation in time series

Figure 9-8 shows two major peaks, one in the morning and the other in the afternoon, both originated from the downstream of node 4000055.

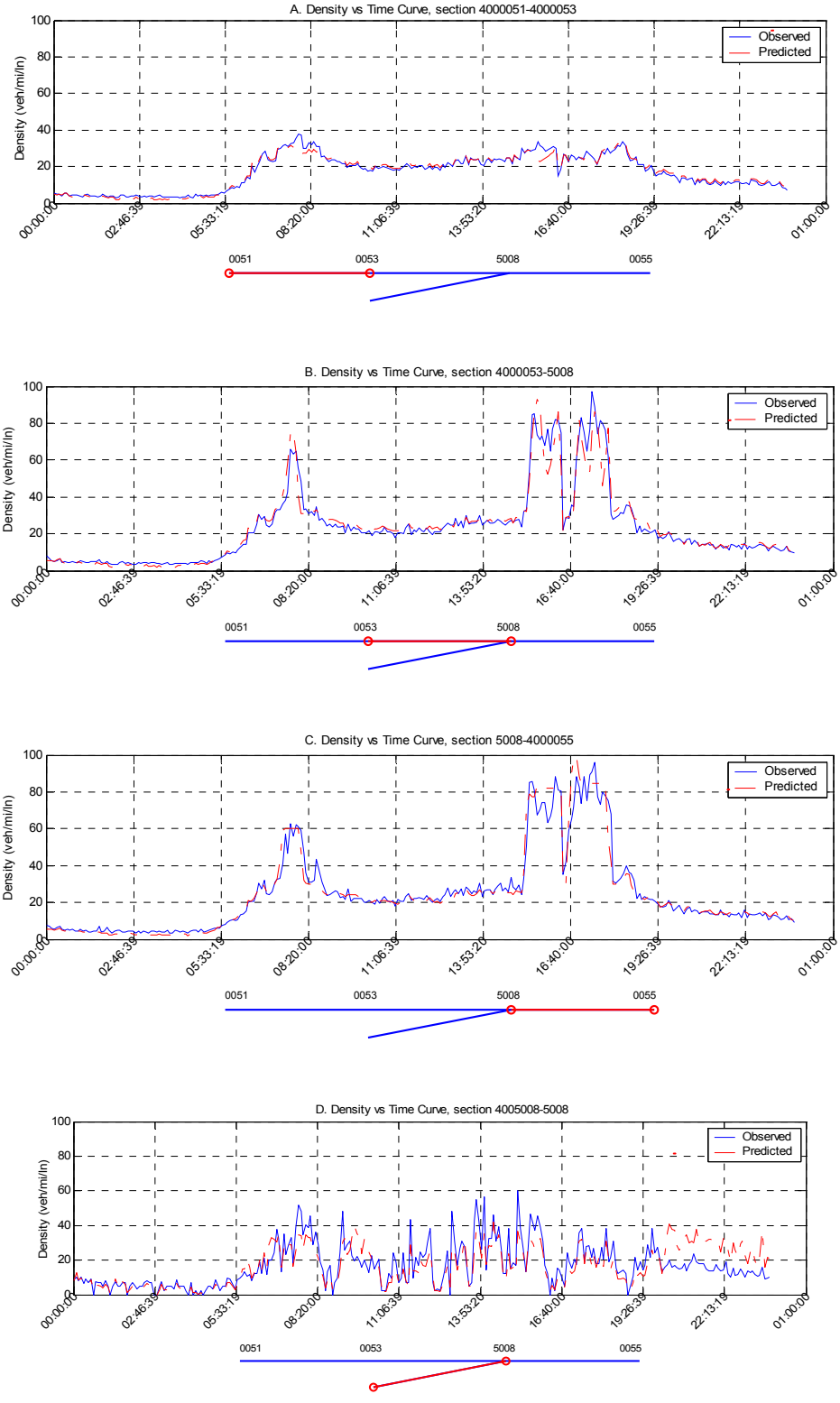


FIGURE 9-8 Time series comparison of density (merging scenario, day 2)

Comparison of prediction and observation in terms of congestion

Figure 9-9 indicates that the prediction matches the observation well in queue formation and dissipation, but less satisfactory in transition area between queued and unqueued traffic.

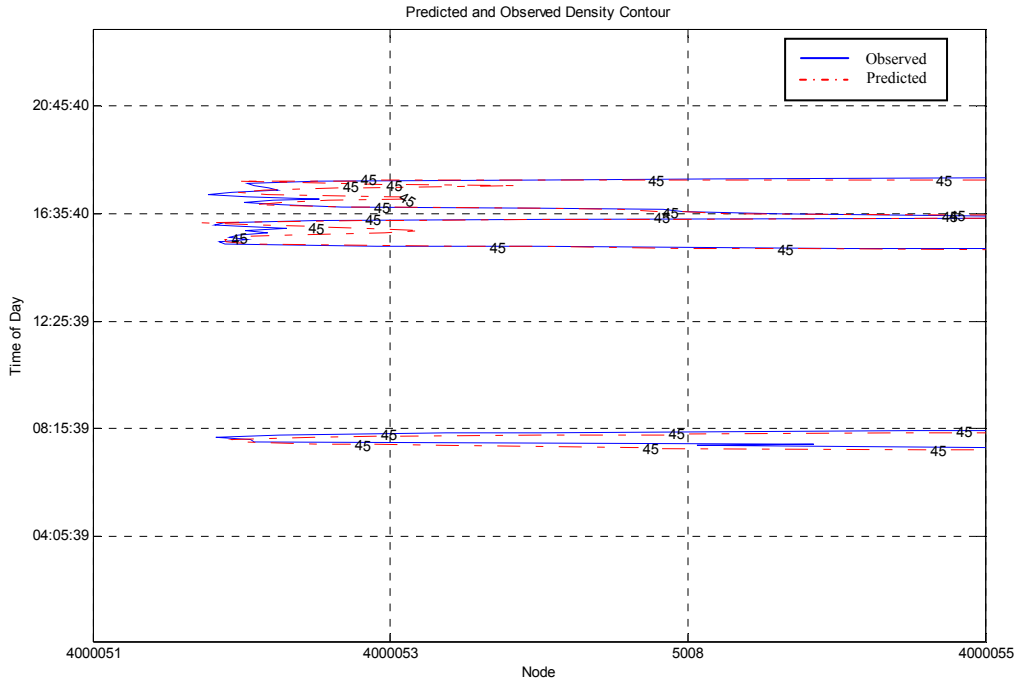


FIGURE 9-9 Comparison of congestion (merging scenario, day 2)

Comparison of prediction and observation in 3D space

Figure 9-10 generally confirms that the two peaks are very similar in terms of timing, location and shape.

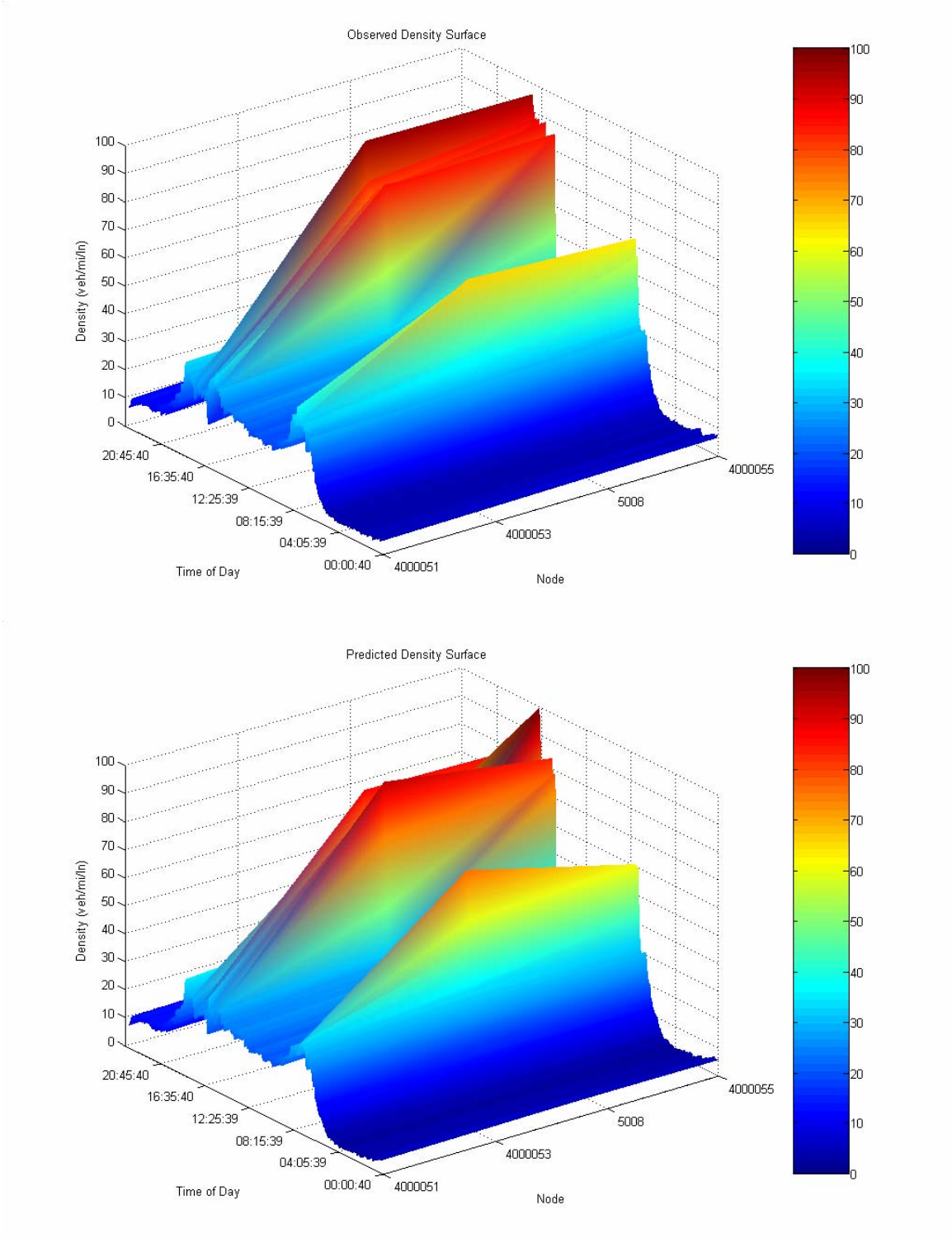


FIGURE 9-10 Comparison of density in 3D space (merging scenario, day 2)

Plot of prediction against observation

Figure 9-11 shows a linear relation between predicted and observed density as indicated by the diagonal line which is a sign of good fit.

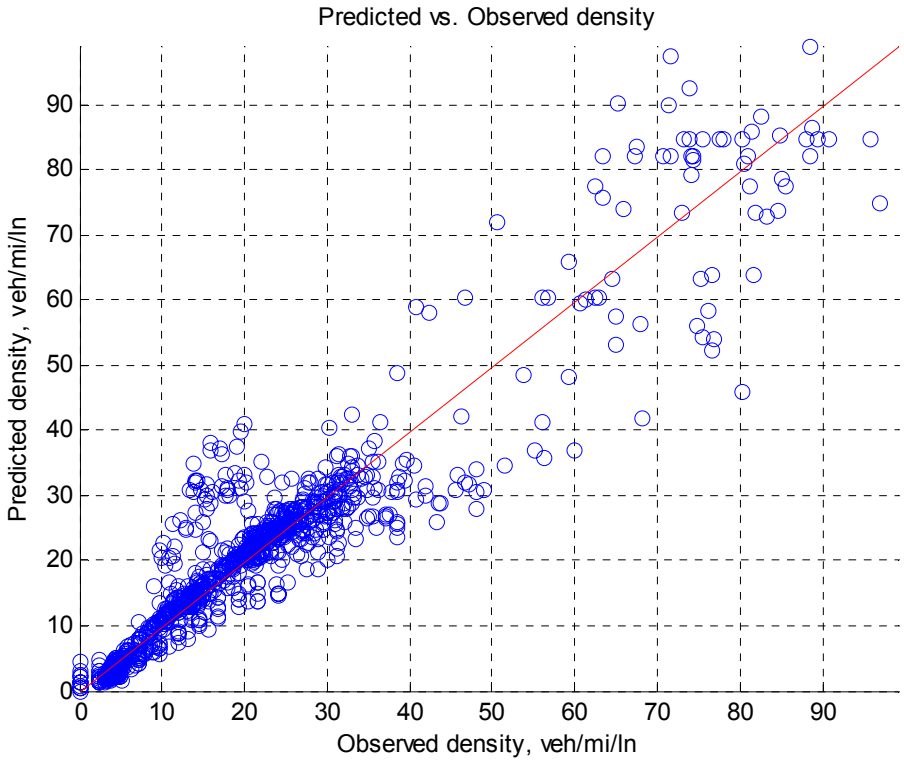


FIGURE 9-11 Plot of prediction against observation (merging scenario, day 2)

Frequency of modeling error

Figure 9-12 also confirms a good fit between the prediction and the observation.

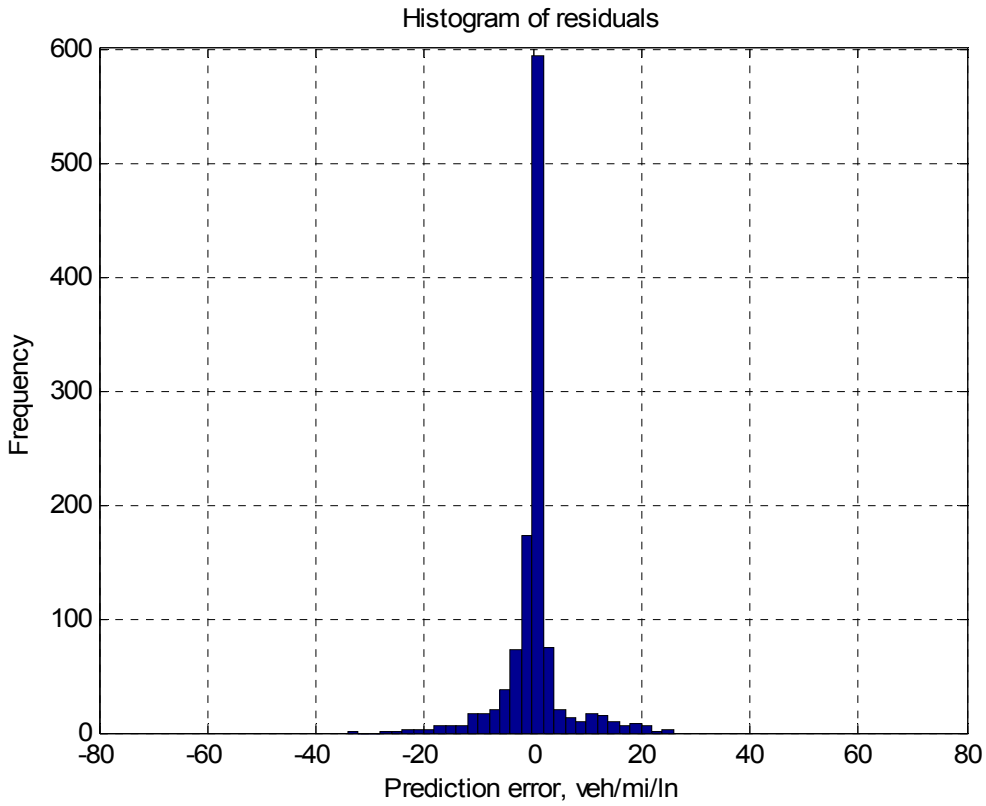


FIGURE 9-12 Frequency of modeling error (merging scenario, day 2)

Comparison of Predicted and Observed Flow-Density Relationships

Figure 9-13 shows the predicted and the observed flow-density relationships on the same figure.

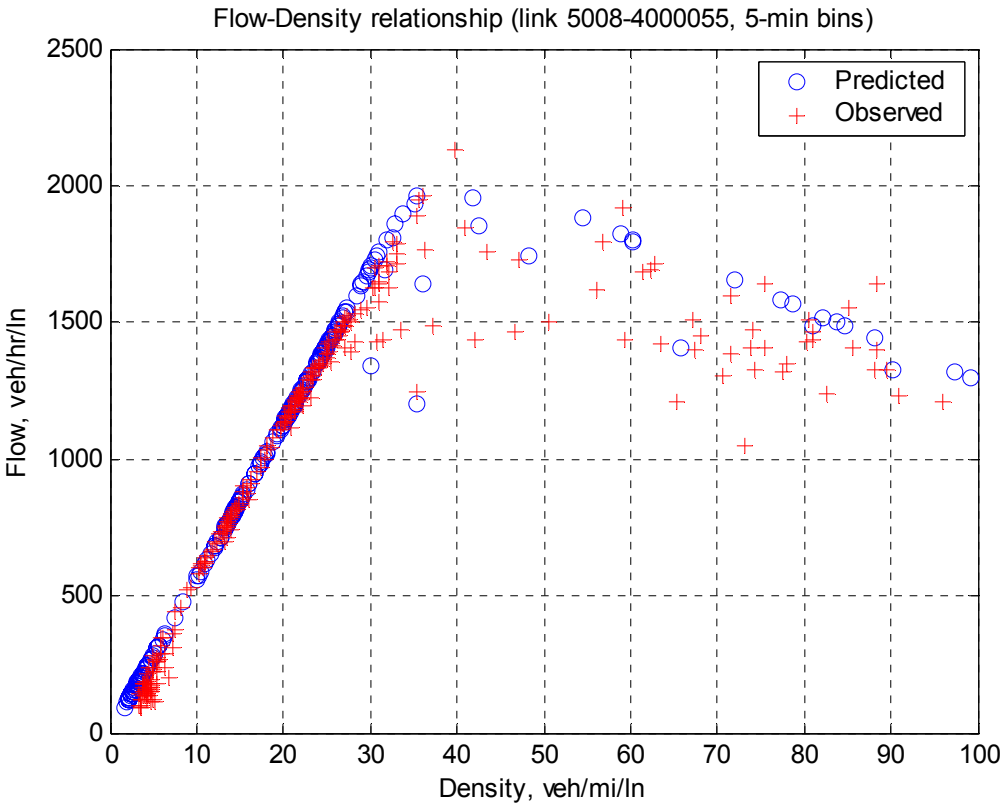


FIGURE 9-13 Comparison of flow-density relationships (merging scenario, day 2)

9.2.2.2 Quantitative evaluation

Simulation of this day spans almost a whole day, starting from 00:01 and ending at 23:51. The result of statistical tests is listed in Table 9-2.

TABLE 9-2 Simultaneous statistical test result of test day 2 (merging scenario)

Item	Value
Batch size (m)	143
Batches (b)	8
Total samples (n)	1144
Significance level (α)	0.05
Grand mean ($\overline{W}$)	0.010975
Variance ($\hat{V}_B$)	0.001116
T-test Statistic (t_0)	0.929335
Critical Value ($t_{0.025,7}$)	2.36462
T-test result	Fail to reject H_0, i.e., $E[\overline{W}_i] = 0$.
Small number (ε)	0.00056
Critical χ^2 value ($\chi^2_{0.05,7}$)	14.07
95 % confidence interval for percentage modeling error	(-0.035, 0.060) $\times 100\%$

Interpreting the results is as before. Briefly, the result suggests that the mean of the modeling error is not statistically different than 0 and the 95% confidence interval for percentage modeling error is (-0.035, 0.060) $\times 100\%$.

9.3 EMPIRICAL TEST - DIVERGING SCENARIO

Information regarding test site 2 (for testing the diverging scenario) has been detailed in Section 7.3.3 of Chapter 7 and is not repeated here. Figure 9-14 shows the structure of the test site that is actually used in the tests. All the mainline links have 4 lanes with approximately the same capacity. Queues can back up from either of the downstream links. The optional friction mentioned in section 4.2.2 of Chapter 4 is not implemented in this scenario.

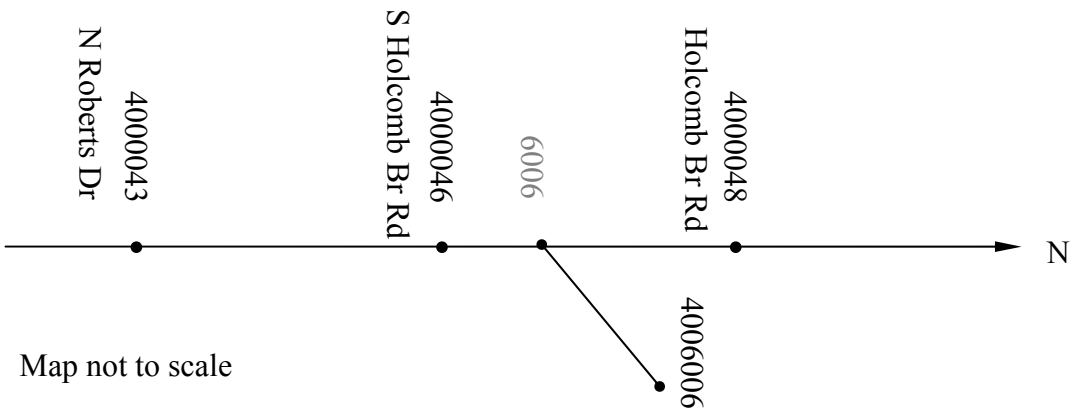


FIGURE 9-14 Structure of test site 2 - diverging scenario (GA 400 northbound)

9.3.1 Test Day 1

Data for test day 1 of diverging scenario was collected on Thursday, September 12, 2002 from 00:01 to 23:51, almost a whole day

9.3.1.1 Qualitative evaluation

Comparison of prediction and observation in time series

Figure 9-15 shows two peaks during this day: a morning peak originated from the downstream of node 4000048 probably caused by insufficient capacity and an afternoon peak backed up from the downstream of node 4006006 due primarily to high exit volume.

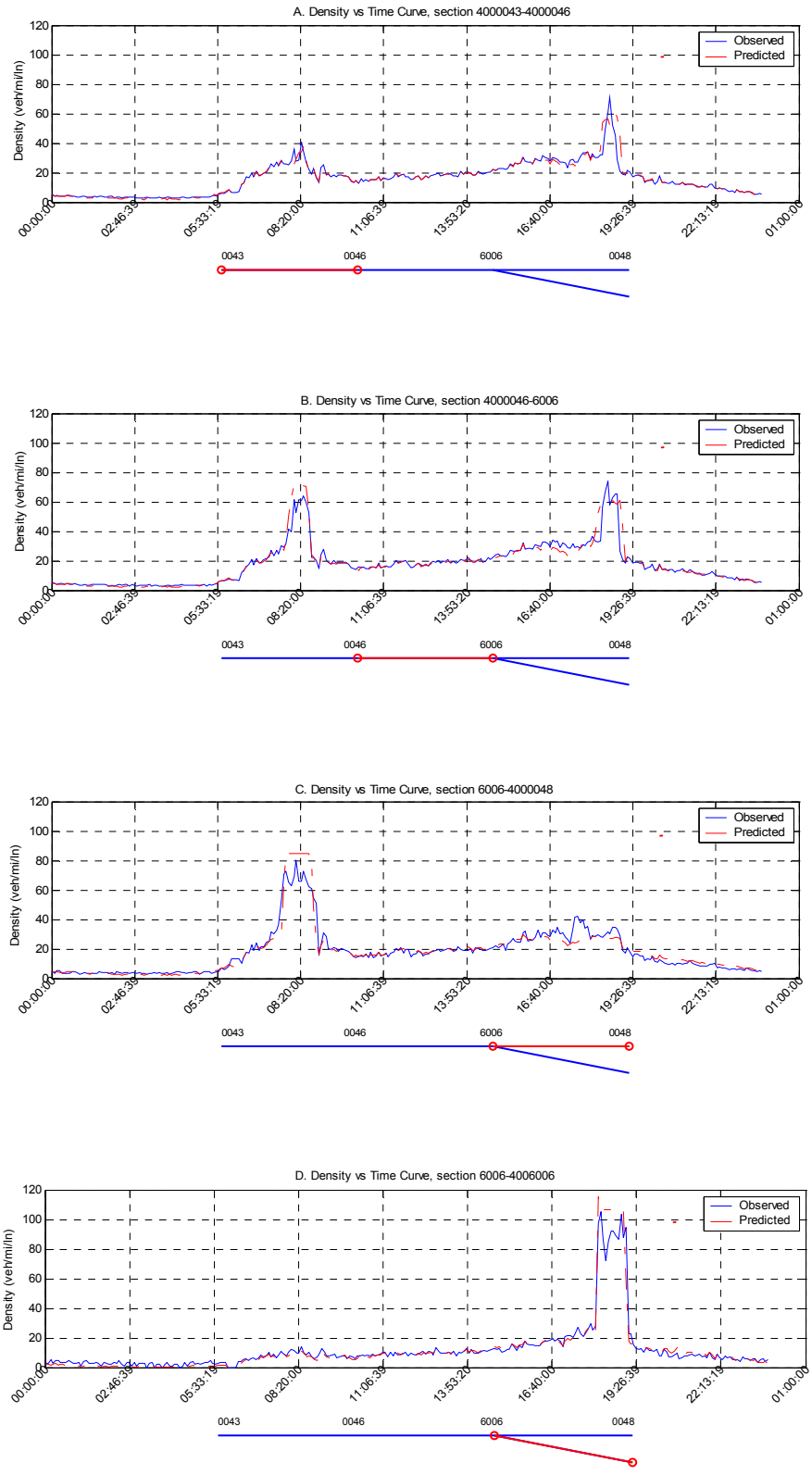


FIGURE 9-15 Time series comparison of density (diverging scenario, day 1)

Comparison of prediction and observation in terms of congestion

Figure 9-16 shows a good fit between the prediction and the observation.

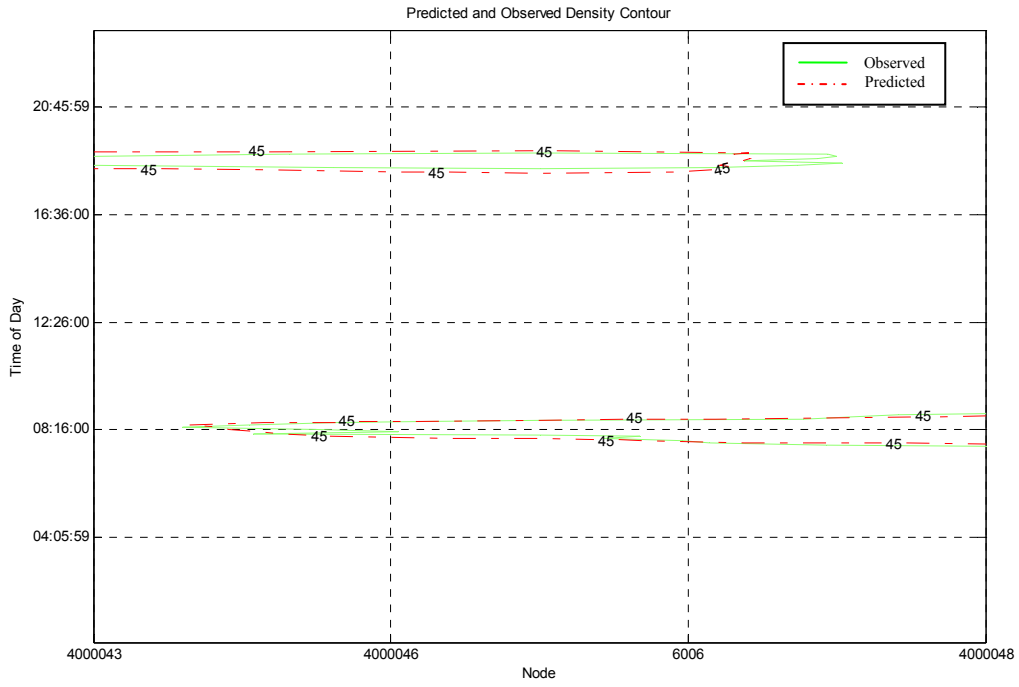


FIGURE 9-16 Comparison of congestion (diverging scenario, day 1)

Comparison of prediction and observation in 3D space

Figure 9-17 generally confirms our findings above.

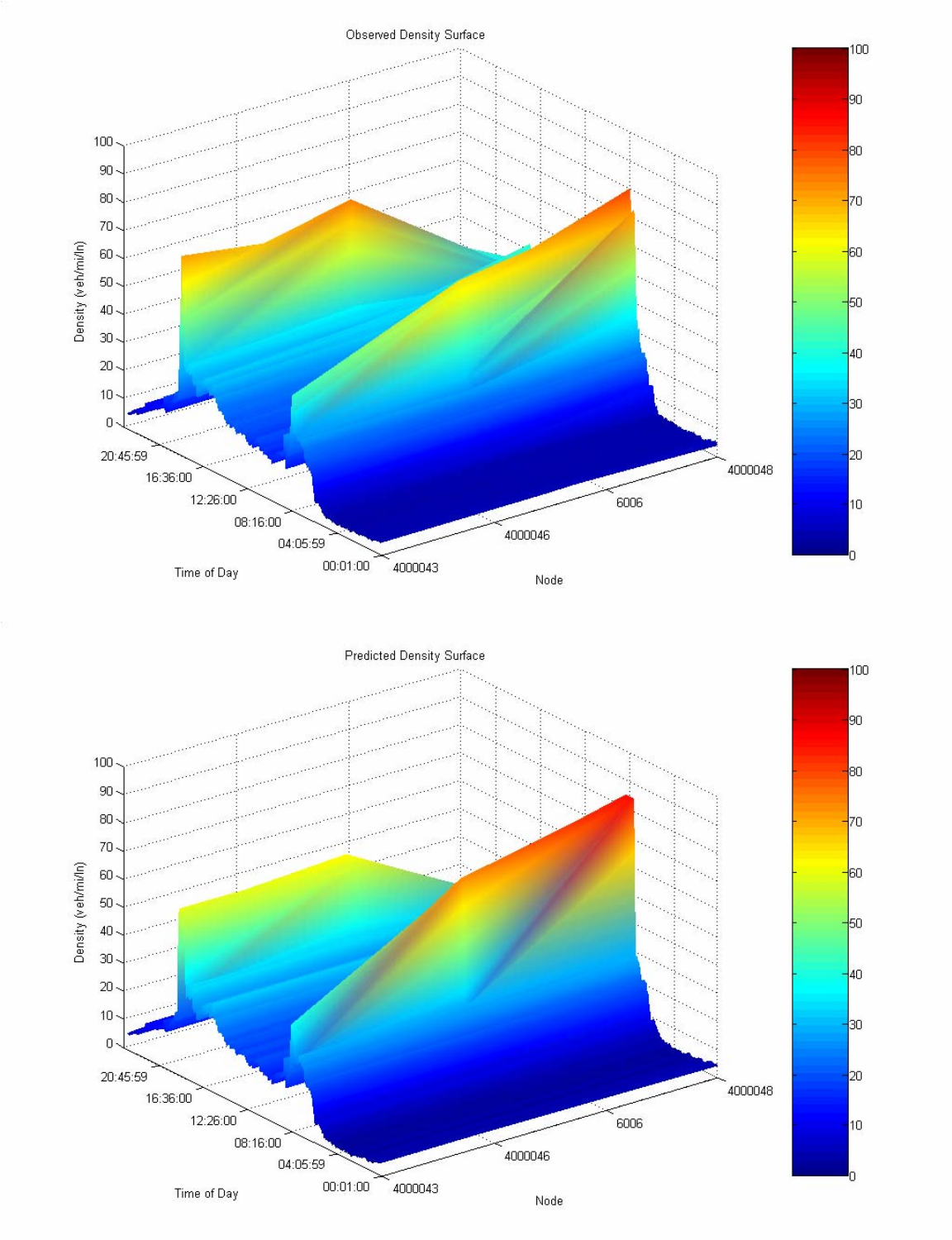


FIGURE 9-17 Comparison of density in 3D space (diverging scenario, day 1)

Plot of prediction against observation

Figure 9-18 shows a good fit between the prediction and the observation. There are 1144 samples in total.

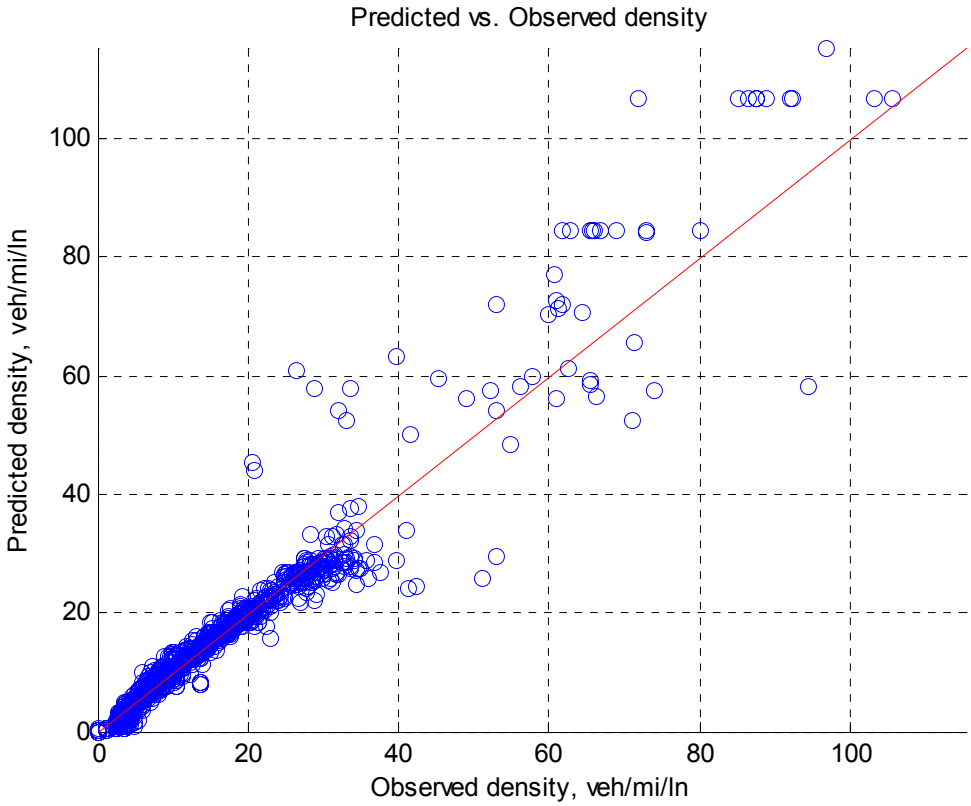


FIGURE 9-18 Plot of prediction against observation (diverging scenario, day 1)

Frequency of modeling error

In Figure 9-19, the highest bar represents 846 samples and the second highest represents 249 samples. Therefore, modeling error within ±10 makes up 96% of the total.

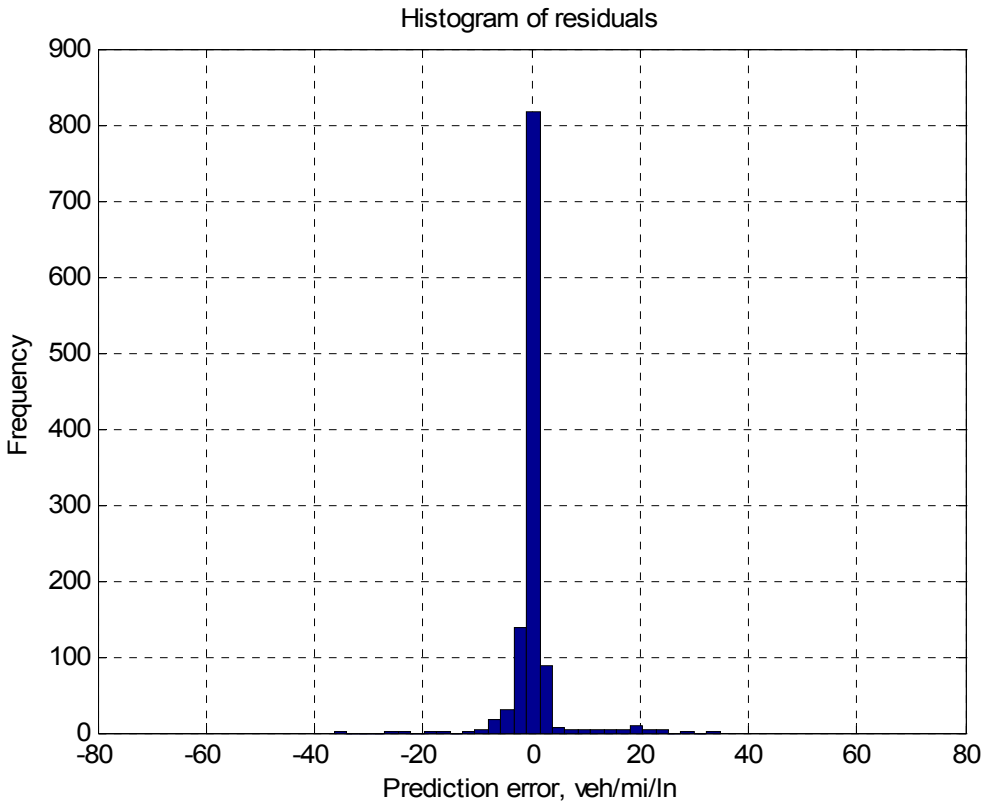


FIGURE 9-19 Frequency of modeling error (diverging scenario, day 1)

Comparison of Predicted and Observed Flow-Density Relationships

Figure 9-20 shows the predicted and the observed flow-density relationships on the same figure.

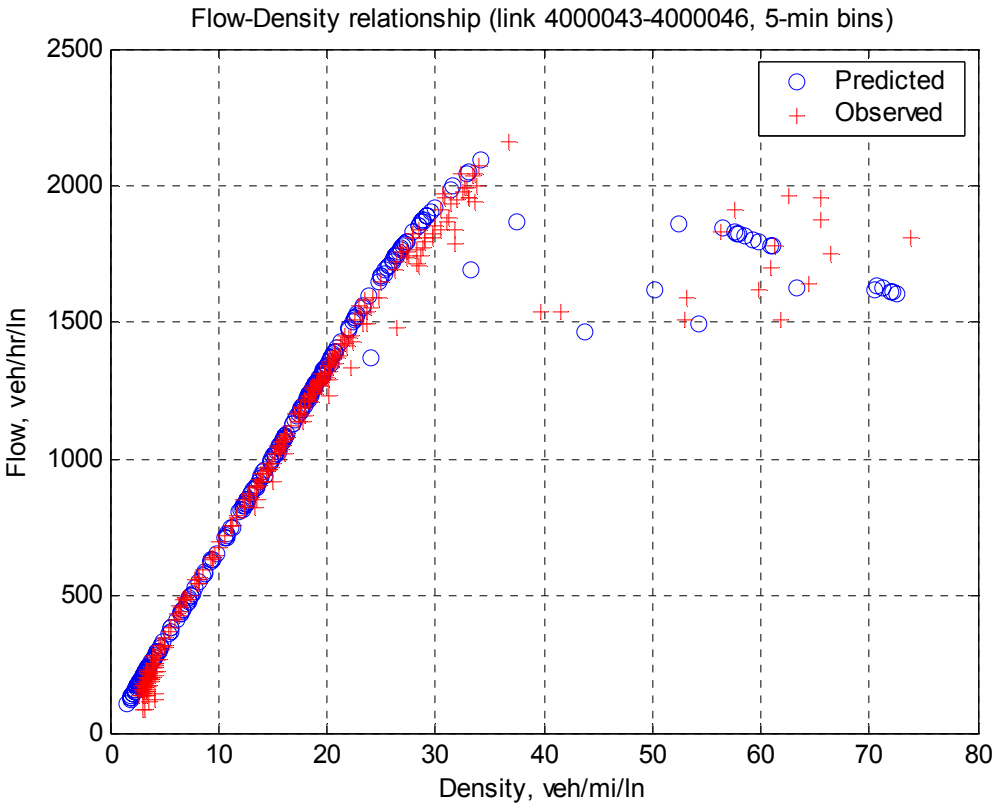


FIGURE 9-20 Comparison of flow-density relationships (diverging scenario, day 1)

9.3.1.2 Quantitative evaluation

The result of simultaneous statistical test is listed in Table 9-3.

TABLE 9-3 Simultaneous statistical test result of test day 1 (diverging scenario)

Item	Value
Batch size (m)	143
Batches (b)	8
Total samples (n)	1144
Significance level (α)	0.05
Grand mean ($\overline{W}$)	0.004786
Variance ($\hat{V}_B$)	0.000787
T-test Statistic (t_0)	0.482557
Critical Value ($t_{0.025,7}$)	2.36462
T-test result	Fail to reject H_0, i.e., $E[\overline{W}_i] = 0$.
Small number (ε)	0.000391
Critical χ^2 value ($\chi^2_{0.05,7}$)	14.07
95 % confidence interval for percentage modeling error	(-0.034, 0.045) $\times 100\%$

Interpreting the results is as before. Briefly, the result suggests that the mean of the modeling error is not statistically different than 0 and the 95% confidence interval for percentage modeling error is (-0.034, 0.045) $\times 100\%$.

9.3.2 Test Day 2

Data for empirical test day 2 of diverging scenario came from Monday, December 9, 2002 from 00:00 to 23:45, almost a full day.

9.3.2.1 Qualitative evaluation

Comparison of prediction and observation in time series

Figure 9-21 generally shows a good fit between the prediction and the observation.

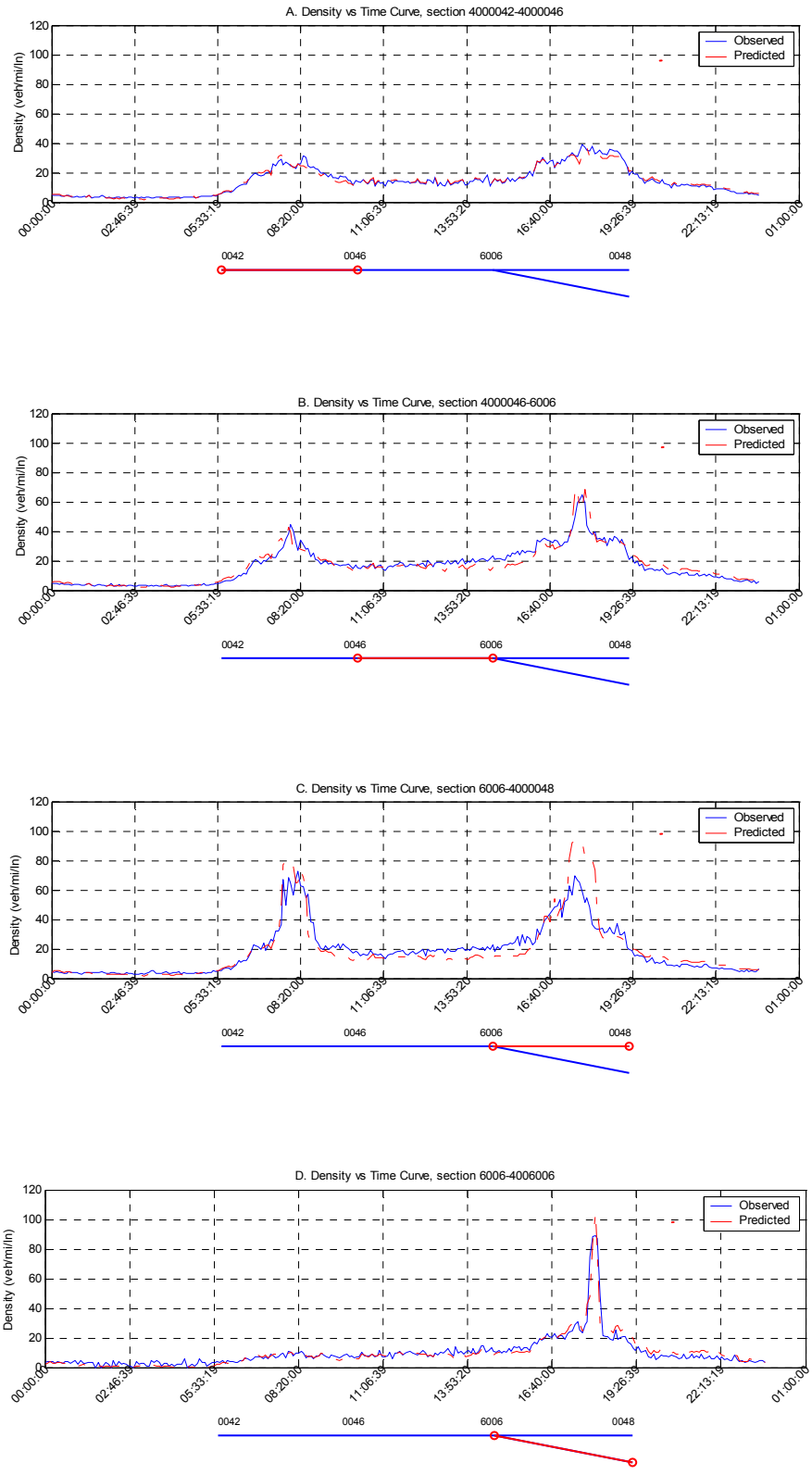


FIGURE 9-21 Time series comparison of density (diverging scenario, day 2)

Comparison of prediction and observation in terms of congestion

Figure 9-22 shows a good agreement between the prediction and the observation as far as congested region is concerned.

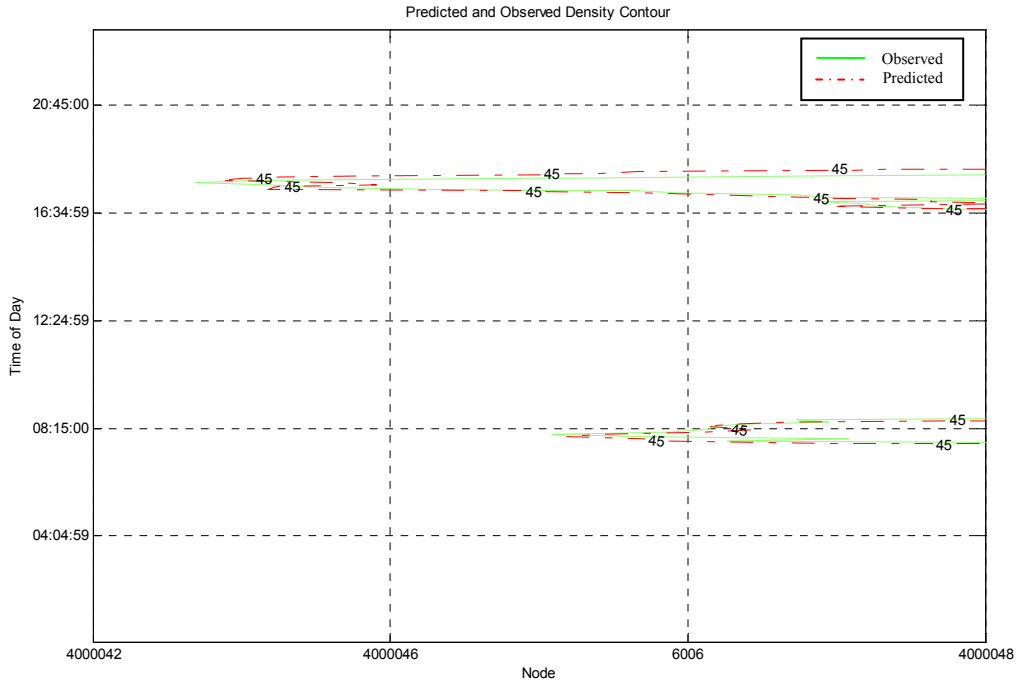


FIGURE 9-22 Comparison of congestion (diverging scenario, day 2)

Comparison of prediction and observation in 3D space

Figure 9-23 generally confirms our findings above.

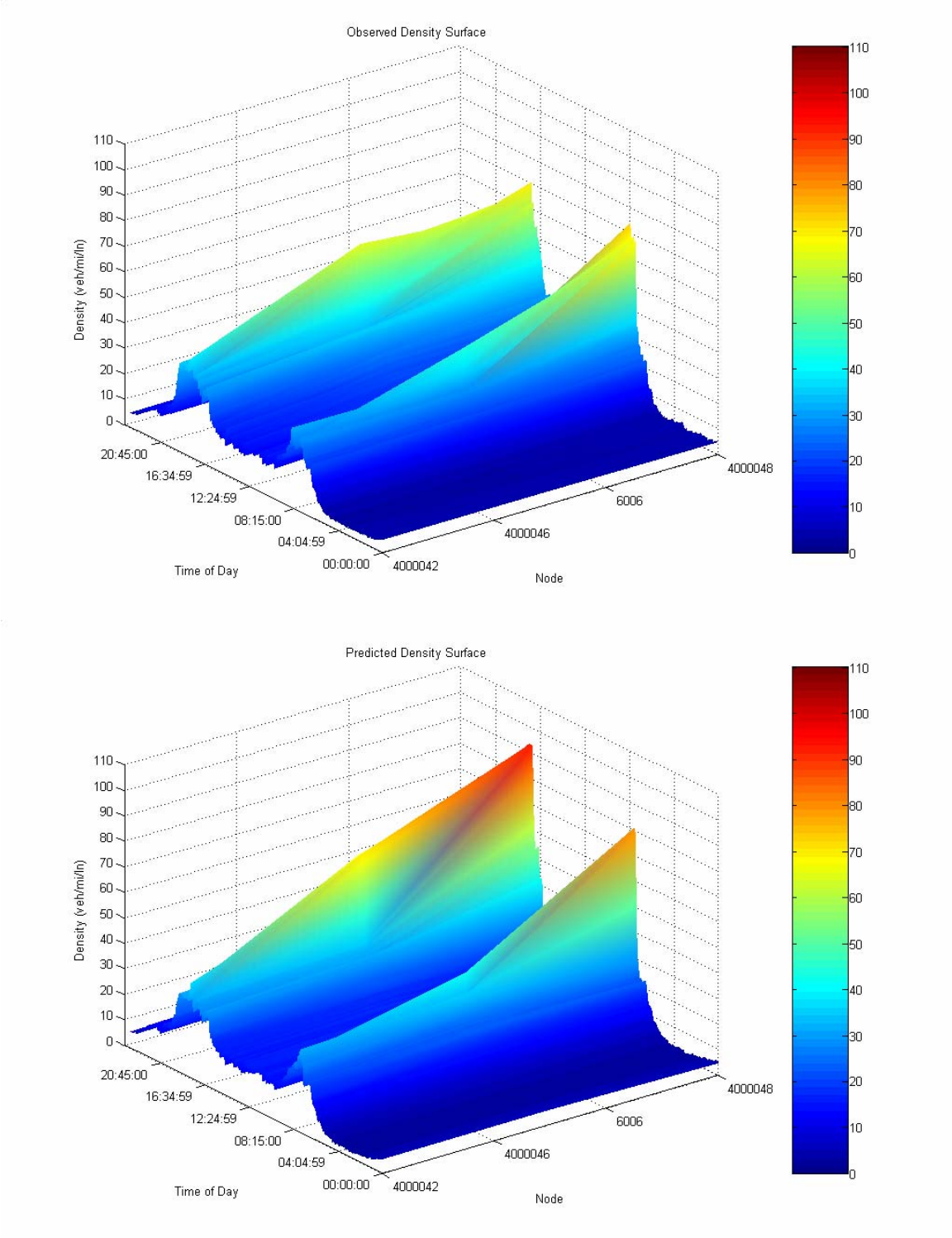


FIGURE 9-23 Comparison of density in 3D space (diverging scenario, day 2)

Plot of prediction against observation

Figure 9-24 has 1144 samples in total, most of which were distributed along the diagonal line where density was under 45 veh/mi/ln. At higher density, more data points are above the diagonal line than under, which confirm the finds above regarding the over-prediction.

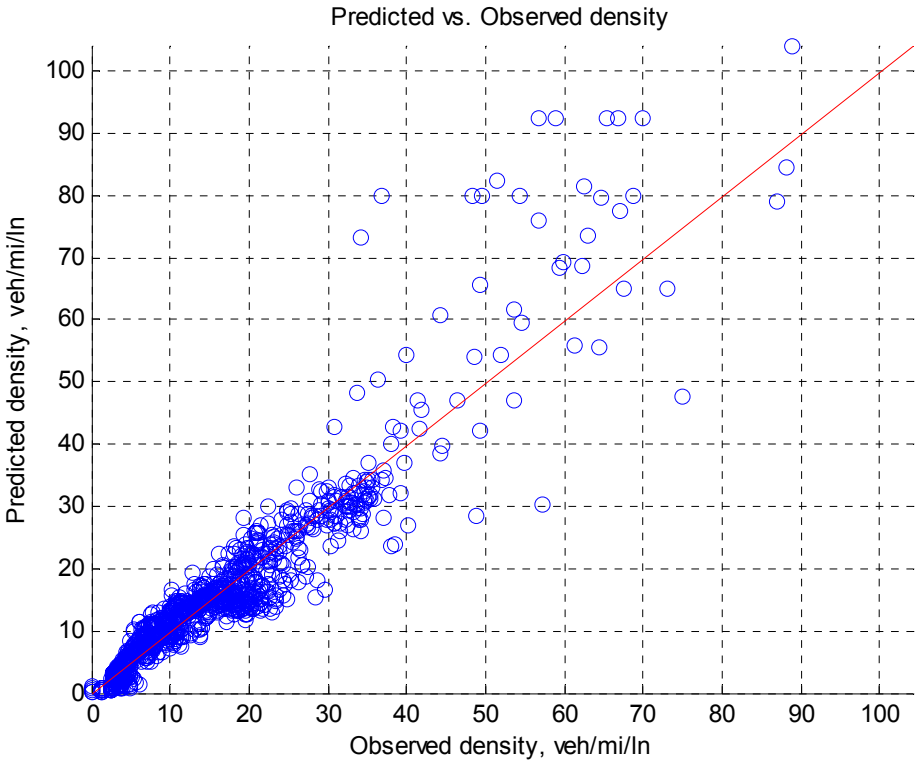


FIGURE 9-24 Plot of prediction against observation (diverging scenario, day 2)

Frequency of modeling error

The highest bar in Figure 9-25 represents 722 samples, and the second highest 335 points, so, roughly, modeling error within ±10 veh/mi/ln accounts for more than 92% of the total samples.

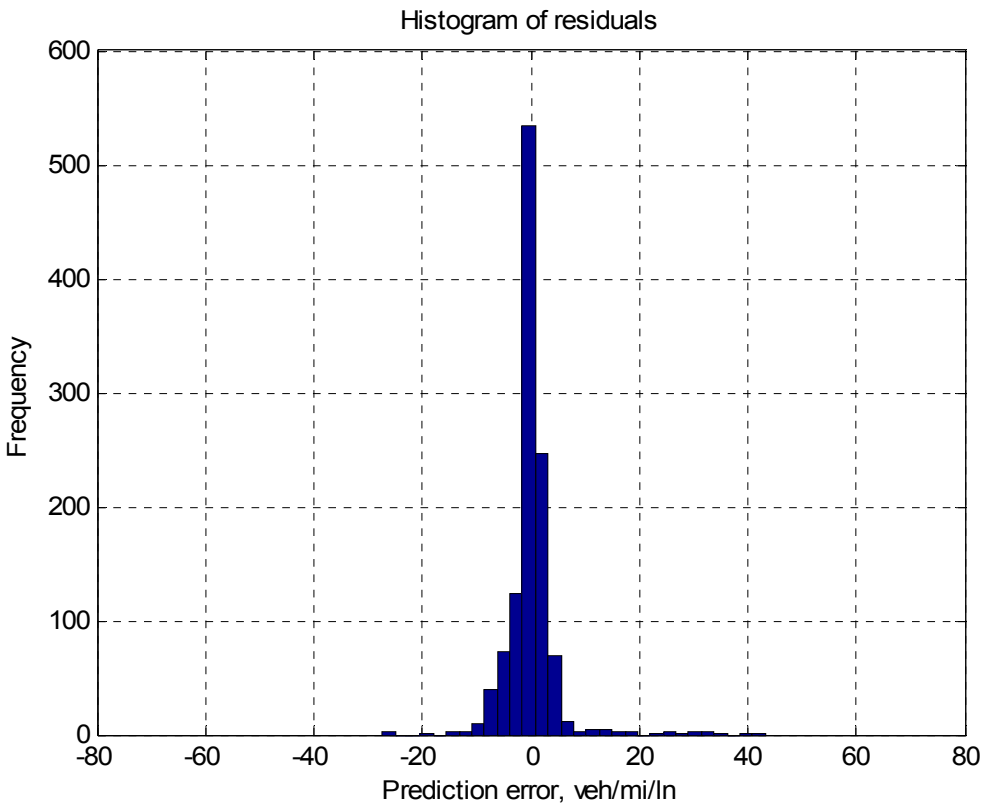


FIGURE 9-25 Frequency of modeling error (diverging scenario, day 2)

Comparison of Predicted and Observed Flow-Density Relationships

Figure 9-26 shows the predicted and the observed flow-density relationships on the same figure.

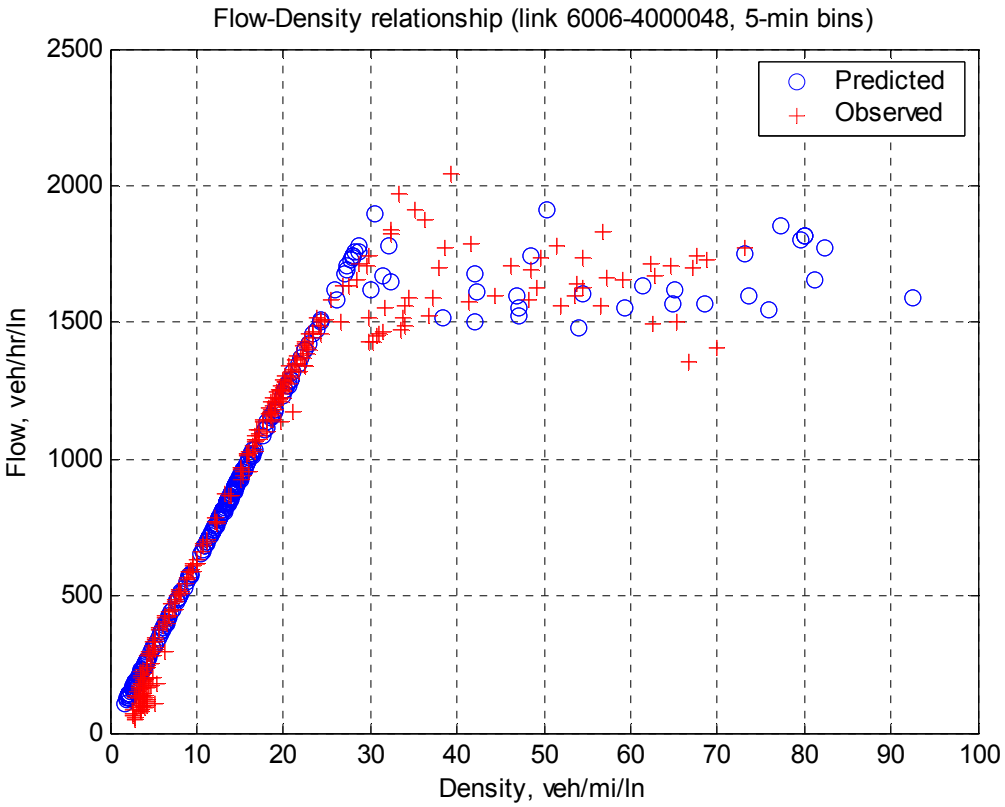


FIGURE 9-26 Comparison of flow-density relationships (diverging scenario, day 2)

9.3.2.2 Quantitative evaluation

The result of simultaneous statistical tests is listed in Table 9-4.

TABLE 9-4 Simultaneous statistical test result of test day 2 (diverging scenario)

Item	Value
Batch size (m)	142
Batches (b)	8
Total samples (n)	1140
Significance level (α)	0.05
Grand mean ($\overline{W}$)	0.011845
Variance ($\hat{V}_B$)	0.001255
T-test Statistic (t_0)	0.945795
Critical Value ($t_{0.025,7}$)	2.36462
T-test result	Fail to reject H_0, i.e., $E[\overline{W}_i] = 0$.
Small number (ε)	0.00062
Critical χ^2 value ($\chi^2_{0.05,7}$)	14.07
95 % confidence interval for percentage modeling error	(-0.037, 0.064) $\times 100\%$

Interpreting the results is as before. Briefly, the result suggests that the mean of the modeling error is not statistically different than 0 and the 95% confidence interval for percentage modeling error is (-0.037, 0.064) $\times 100\%$.

9.4 EMPIRICAL TEST - NETWORK SCENARIO

Empirical test on network scenario requires O-D flows which may introduce exogenous error, and this differs from empirical tests on the previous two scenarios. Information regarding test site 3 (for testing the network scenario) has been detailed in Section 7.3.4 of Chapter 7 and is not repeated here. Figure 9-27 shows the structure of the test site that is actually used in the tests.

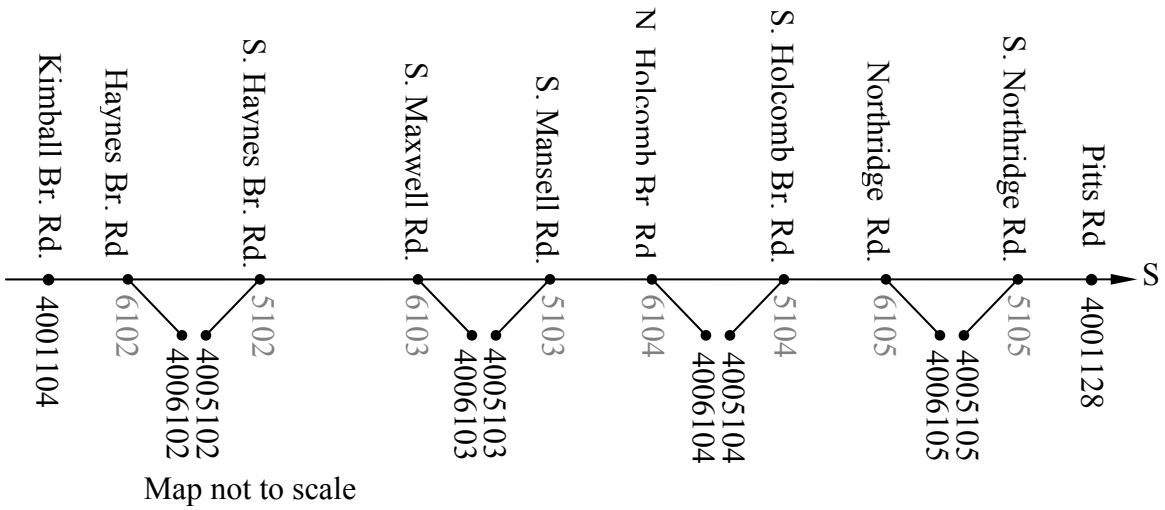


FIGURE 9-27 Model of test site 3 - network scenario (GA 400 southbound)

9.4.1 Test Day 1

Data for test day 1 of network scenario was collected on Friday, Oct. 11, 2002 from 05:50:00 to 19:20:00.

9.4.1.1 Qualitative evaluation

Comparison of prediction and observation in time series

Figure 9-28 shows density vs. time curves of some of the links of the test site. In Part A, the predicted and observed density curves match perfectly except some minor discrepancies. Part B and C follow a similar pattern to Part A but with relatively higher observed density at the beginning and the end. The most noticeable discrepancy in Part B arises from 5:40 to 8:00 where the observed density is around 40 veh/mi/ln, while the predicted is around 25 veh/mi/ln. A close look at the observed data reveals that the

observed flow during this period was unrealistically high. The observed flow level at link 4001104-6102 during this period was approximately 1800 veh/hr/ln. The exit flow during the same period was roughly 200 veh/hr/ln, so the observed flow level at link 6102-5102 should be about 1600 veh/hr/ln. However, the observed flow here remained 1800 veh/hr/ln, and this explains most of the discrepancy of the two density curves. Part C is the extension of the pattern in Part B. Part D shows both the morning peak and the afternoon peak. The origin of the morning peak can be traced to the downstream end in Part I where a queue built up from further downstream and backed up to this link. The queue continued to propagate backwards and diminished at link 6103-5103.

The afternoon peak originated from link 6104-4006104 (Part N) where traffic is somewhat constrained between 17:00-18:00, due primarily to high exit volume during this time and, probably, limited discharging capability at downstream intersection. The constrained flow grew backwards passing the diverging point and onto the upstream mainline, i.e., link 5103-6104. Notice that traffic density at links 5103-6104 and 6104-4006104 stayed, in most of the time, under 45 veh/mi/ln. This means that, though traffic was constrained, a queue might not have been formed. However, this was no longer the case at link 6103-5103 (Part D) where traffic density skyrockets and a queue has certainly formed here. One can also suspect that the queue might have been caused by a sudden inflow from on-ramp link 4005103-5105 (Part M). It should be pointed out that, though there was a local peak of flow at this link during this time, it was not high enough to be a major contributing factor to the queue. This queue eventually diminished somewhere at link 5102-6103 (Part C). This confirms our intuition that a minor disturbance downstream can develop a huge queue upstream. It also confirms our assumption that congestion

(though a queue has not formed yet here) at exit links can grow backwards passing diverging point and further constrain upstream traffic.

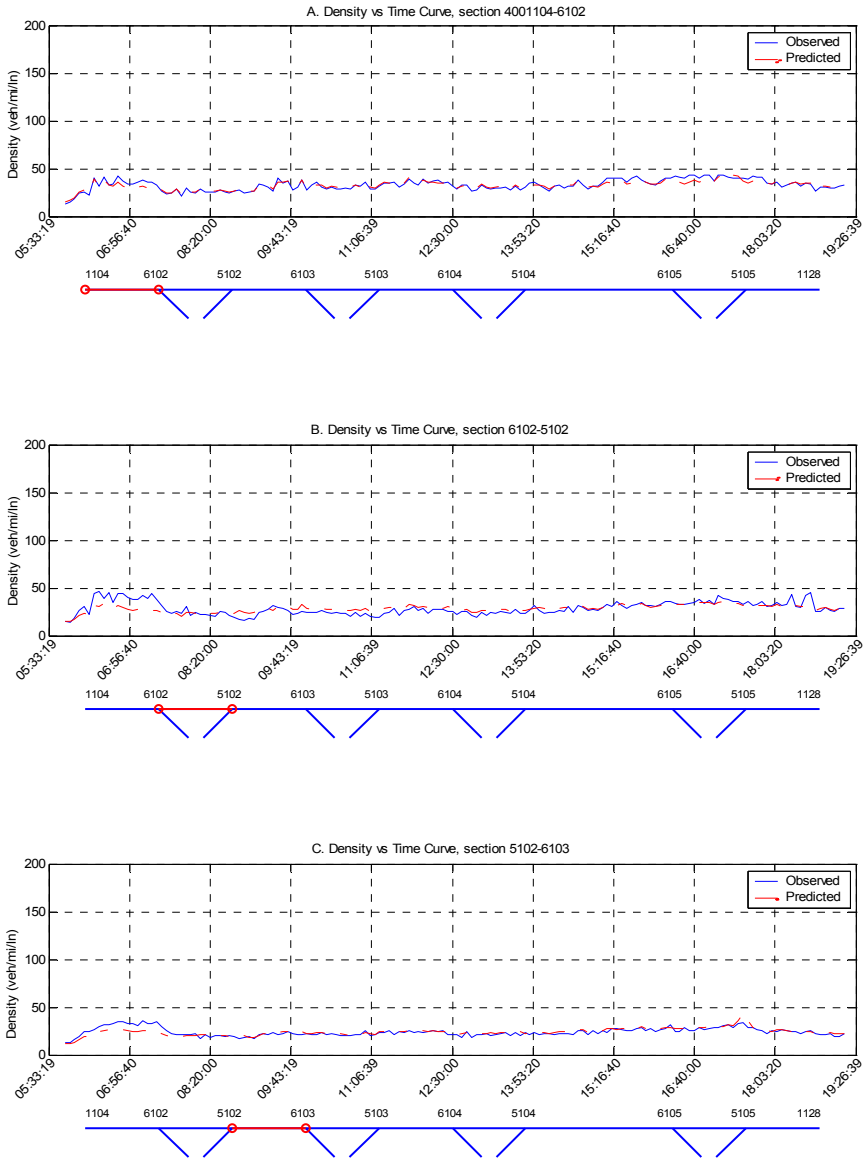


FIGURE 9-28 Time series plots of density (network scenario, day 1)

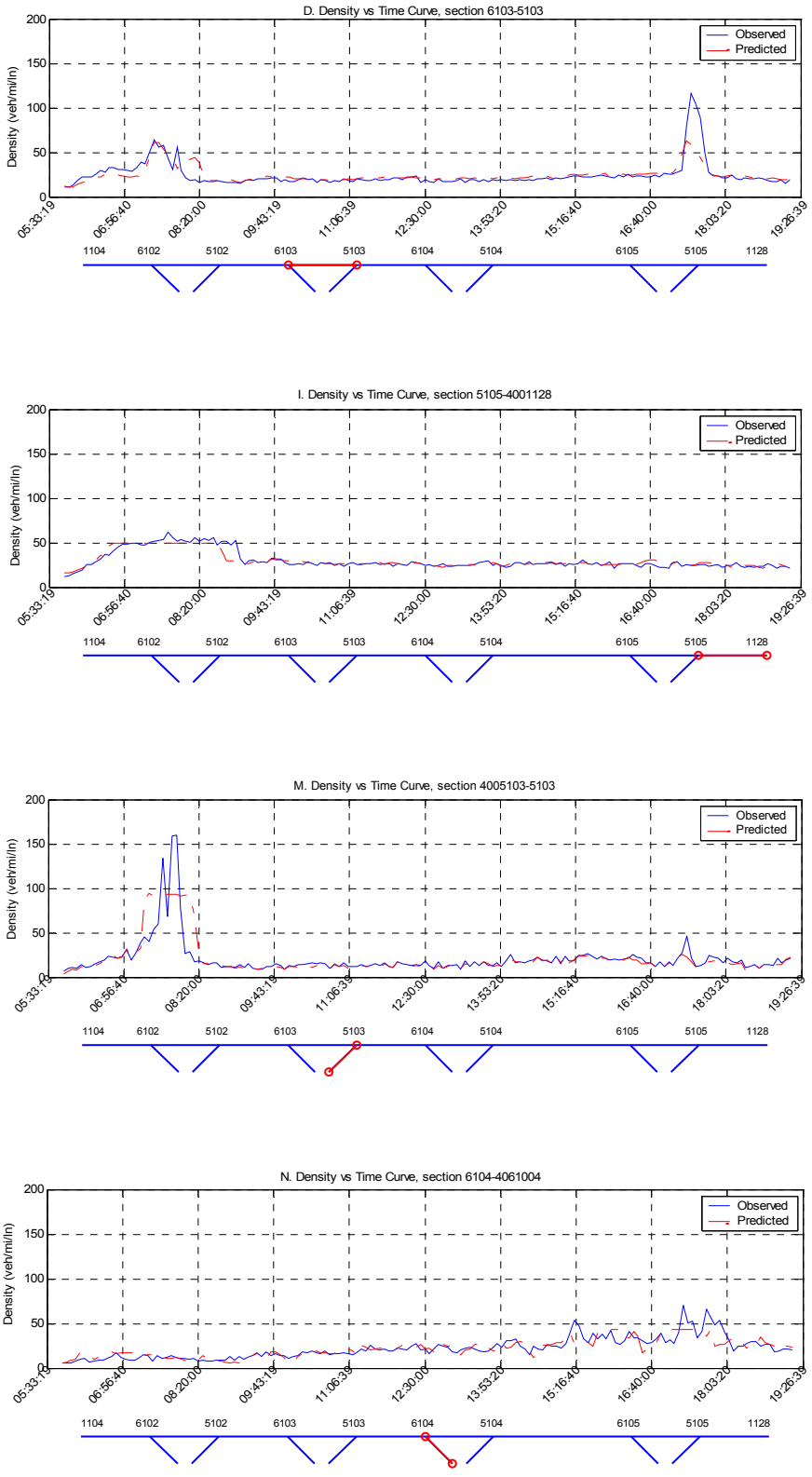


FIGURE 9-28 Continued

Comparison of prediction and observation in terms of congestion

Figure 9-29 shows the morning congestion appeared approximately between 7:00 and 8:30 and the afternoon congestion was much smaller in scale. The figure shows a good agreement between the prediction and the observation.

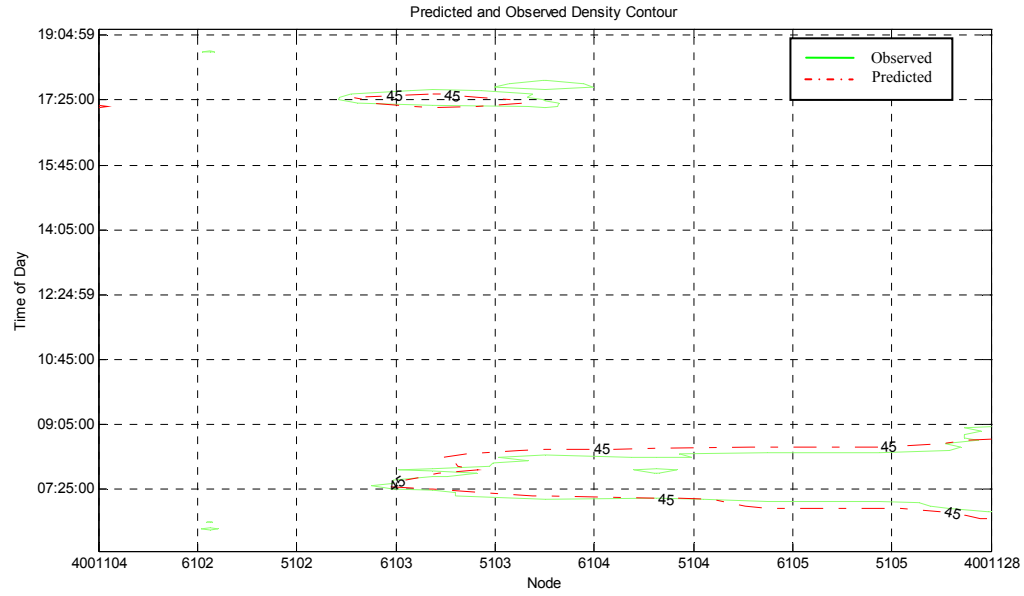
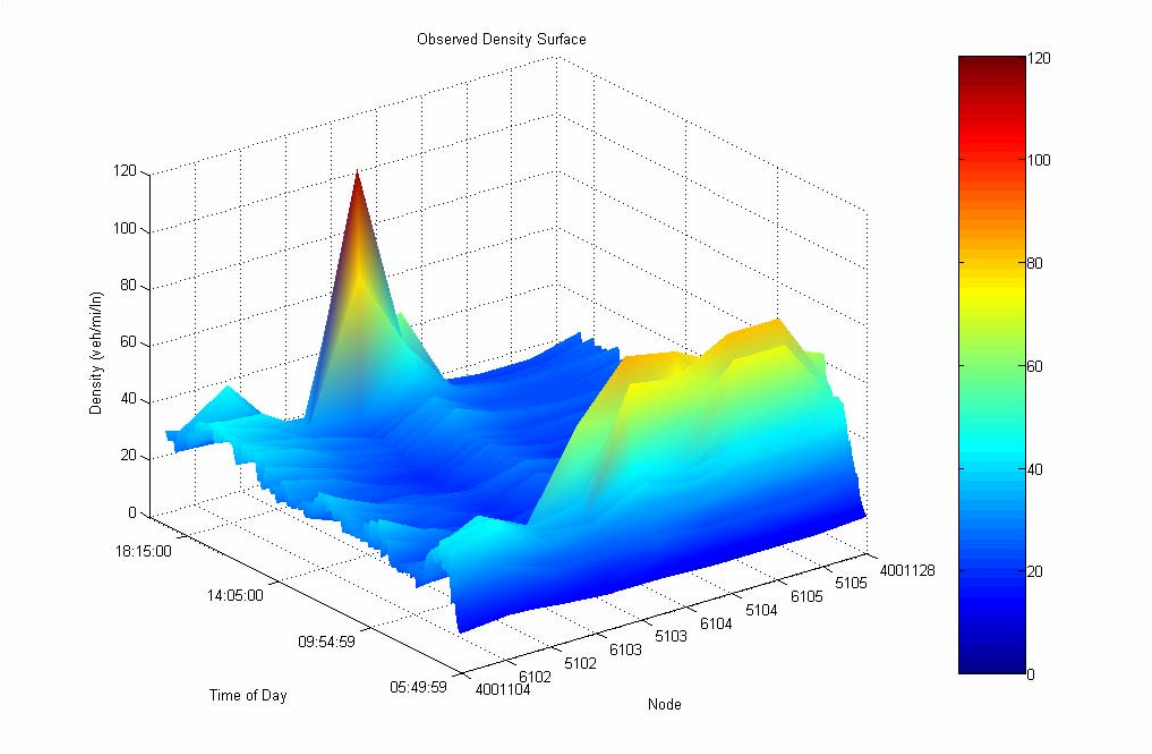


FIGURE 9-29 Comparison of congestion (network scenario, day 1)

Comparison of prediction and observation in 3D space

Figure 9-30 shows a sharp spike in the observation and this is the major difference comparing with the prediction. Fortunately, the model predicts congestions during the same period of time and to the same extent as what are observed, and this is what traffic engineers are most likely to concern.



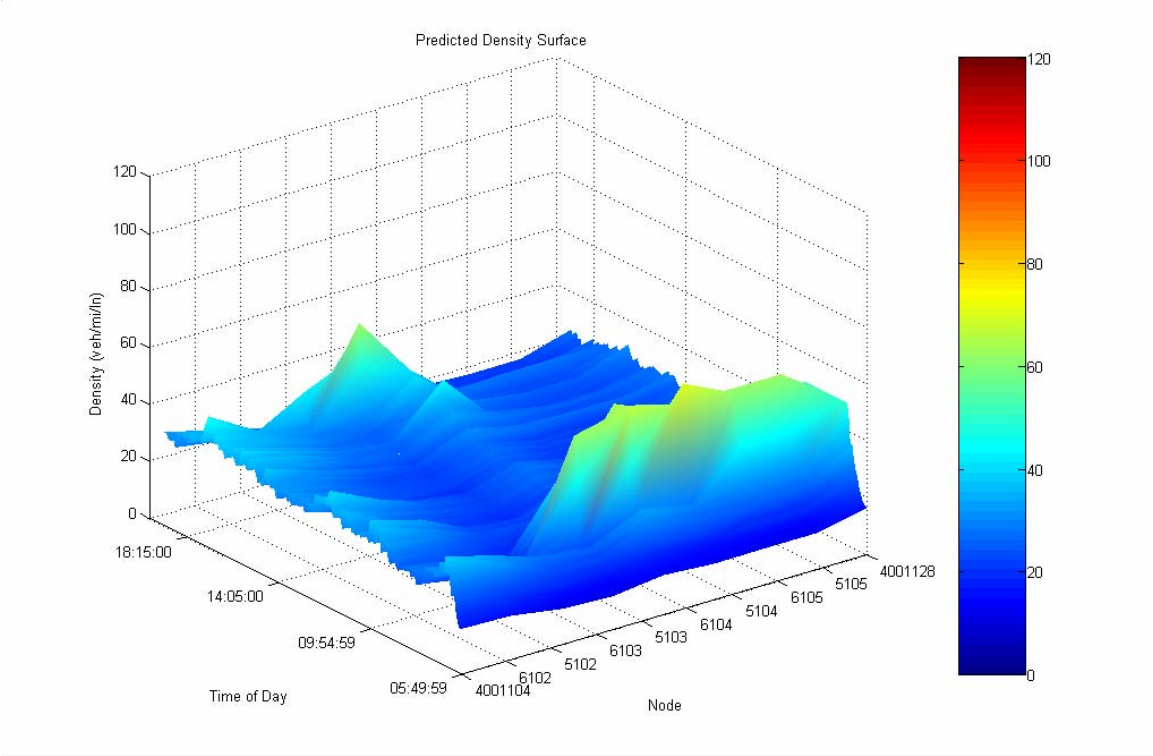


FIGURE 9-30 Comparison of density in 3D space (network scenario, day 1)

Plot of prediction against observation

Figure 9-31 confirms good agreement between the prediction and the observation. Although there are several outliers at high density level, they account for less than 2% of the total samples.

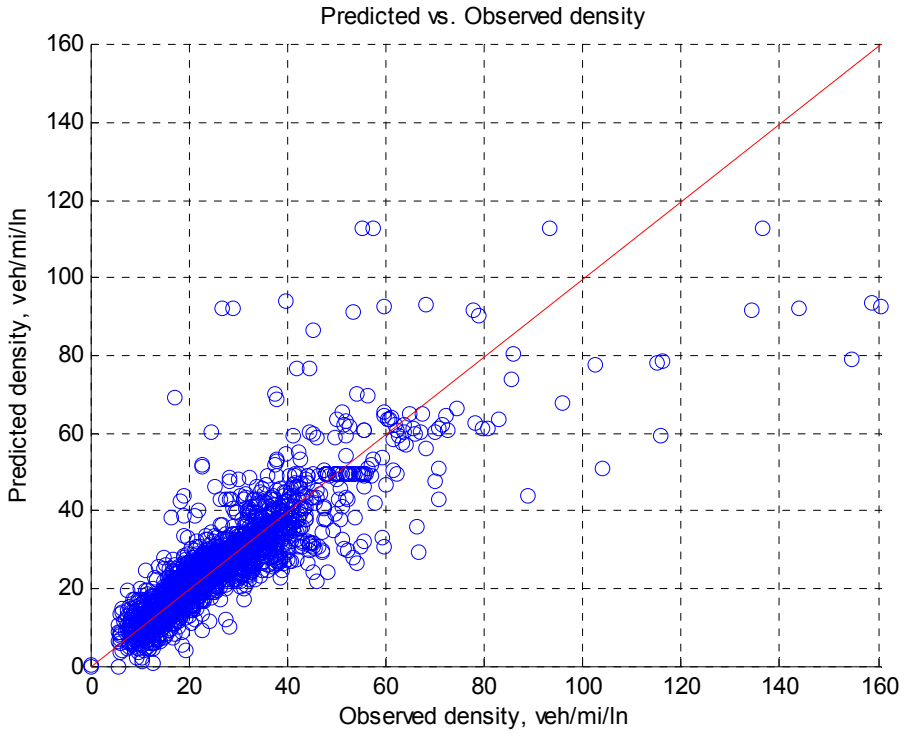


FIGURE 9-31 Plot of prediction against observation (network scenario, day 1)

Frequency of modeling error

In Figure 9-32, the highest bar accounts for 86% of the total samples, the second highest bar makes up 12% of the total samples, and the other bars take the remainder 2%. Notice that high modeling errors account for a negligible portion of the total samples.

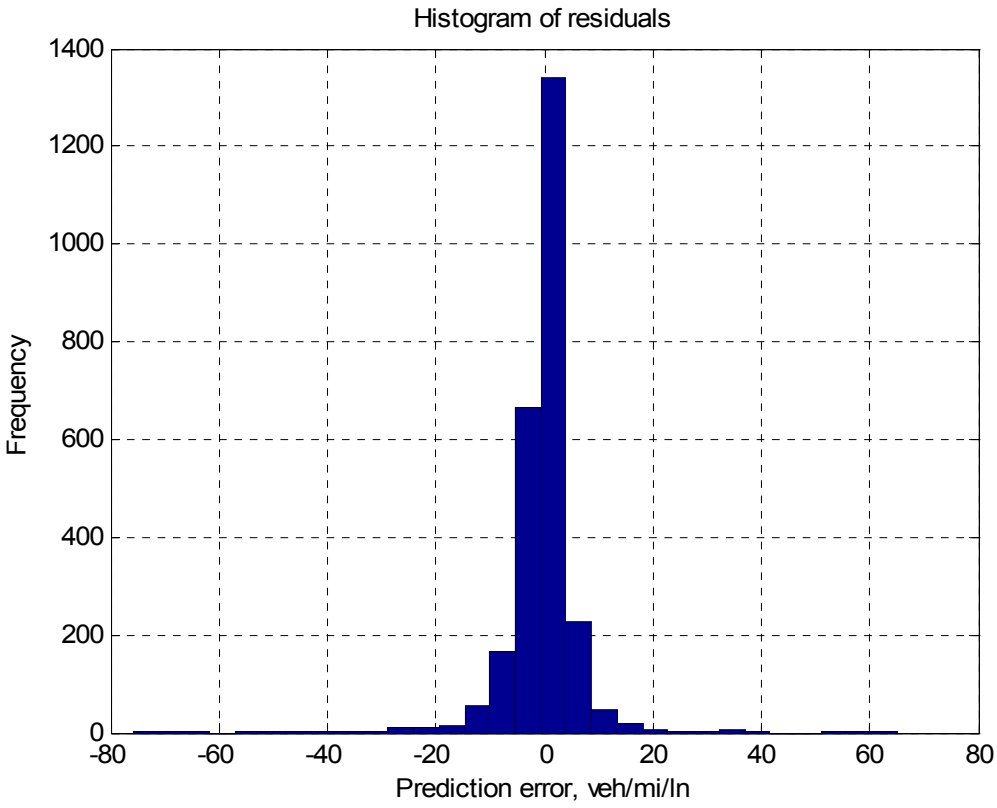


FIGURE 9-32 Frequency of modeling error (network scenario, day 1)

Comparison of Predicted and Observed Flow-Density Relationships

Figure 9-33 shows the predicted and the observed flow-density relationships on the same figure.

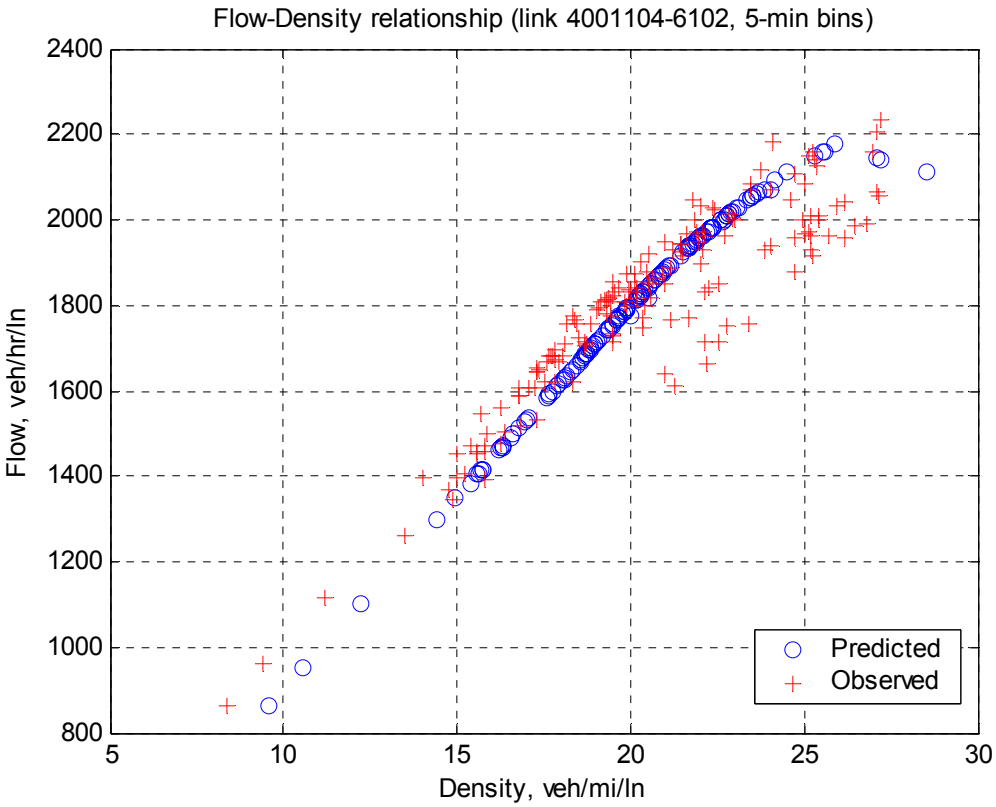


FIGURE 9-33 Comparison of flow-density relationships (network scenario, day 1)

9.4.1.2 Quantitative evaluation

Table 9-5 summarizes the result of the simultaneous statistical test.

TABLE 9-5 Simultaneous statistical test result of test day 1 (network scenario)

Item	Value
Batch size (m)	81
Batches (b)	32
Total samples (n)	2592
Significance level (α)	0.05
Grand mean ($\overline{W}$)	-0.000937
Variance ($\hat{V}_B$)	0.003034
T-test Statistic (t_0)	-0.096263
Critical Value ($t_{0.025,31}$)	2.04227
T-test result	Fail to reject H_0, i.e., $E[\overline{W}_i] = 0$.
Small number (ε)	0.0021
Critical χ^2 value ($\chi^2_{0.05,31}$)	45.0
95 % confidence interval for percentage modeling error	(-0.088, 0.095) $\times 100\%$

Interpreting the results is as before. Briefly, the result suggests that the mean of the modeling error is not statistically different than 0 and the 95% confidence interval for percentage modeling error is (-0.088, 0.095) $\times 100\%$.

Compared with the previous empirical tests, the confidence interval in this test is larger than the previous tests. This is probably because this test involves O-D estimation, a process that is prone to estimation error, while the previous tests do not. Given the existence of O-D estimation error, the above test result is still satisfactory.

9.4.2 Test Day 2

Data for test day 2 of network scenario was collected on Monday, Oct. 14, 2002 from 5:52 to 19:57, a duration of approximately 14 hours..

9.4.2.1 Qualitative evaluation

Comparison of prediction and observation in time series

Figure 9-34 shows part of the links to highlight the two peaks on the day. The figure indicates that the prediction agrees the observation well.

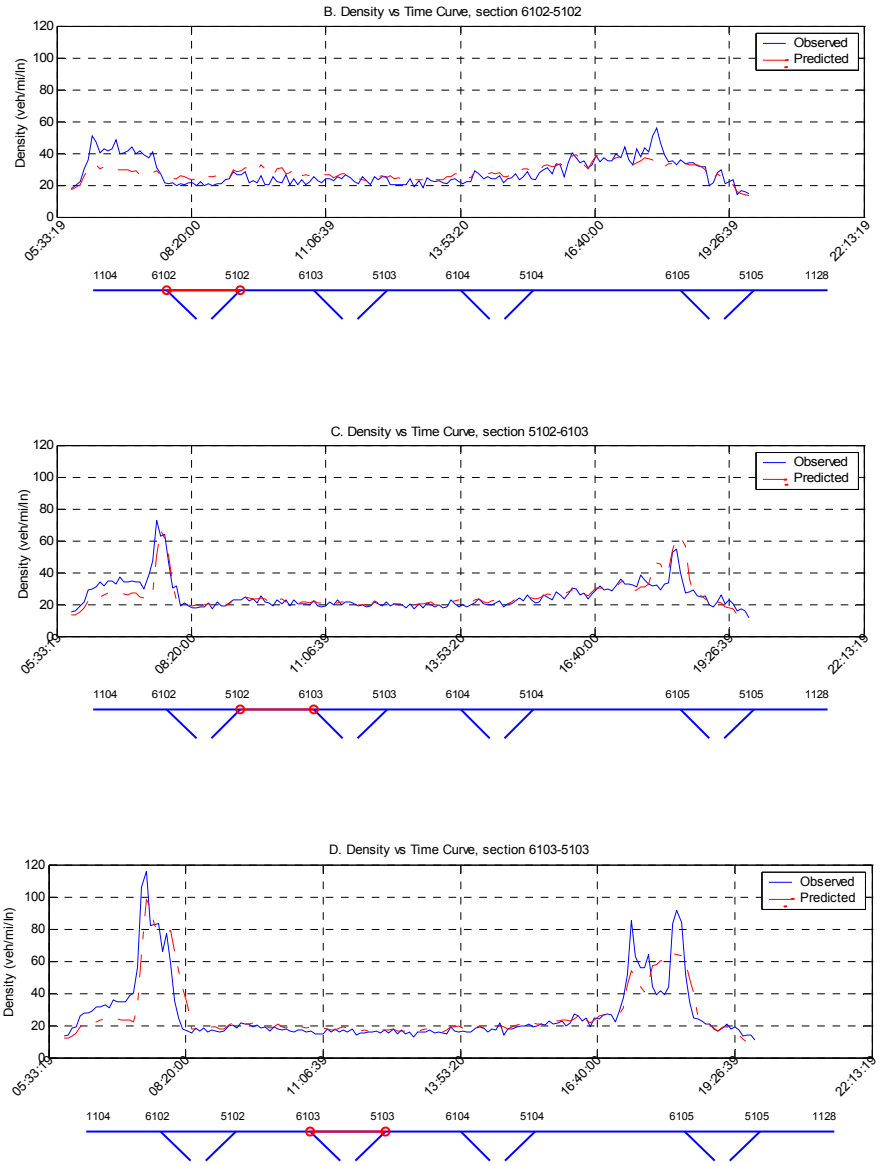


FIGURE 9-34 Time series comparison of density (network scenario, day 2)

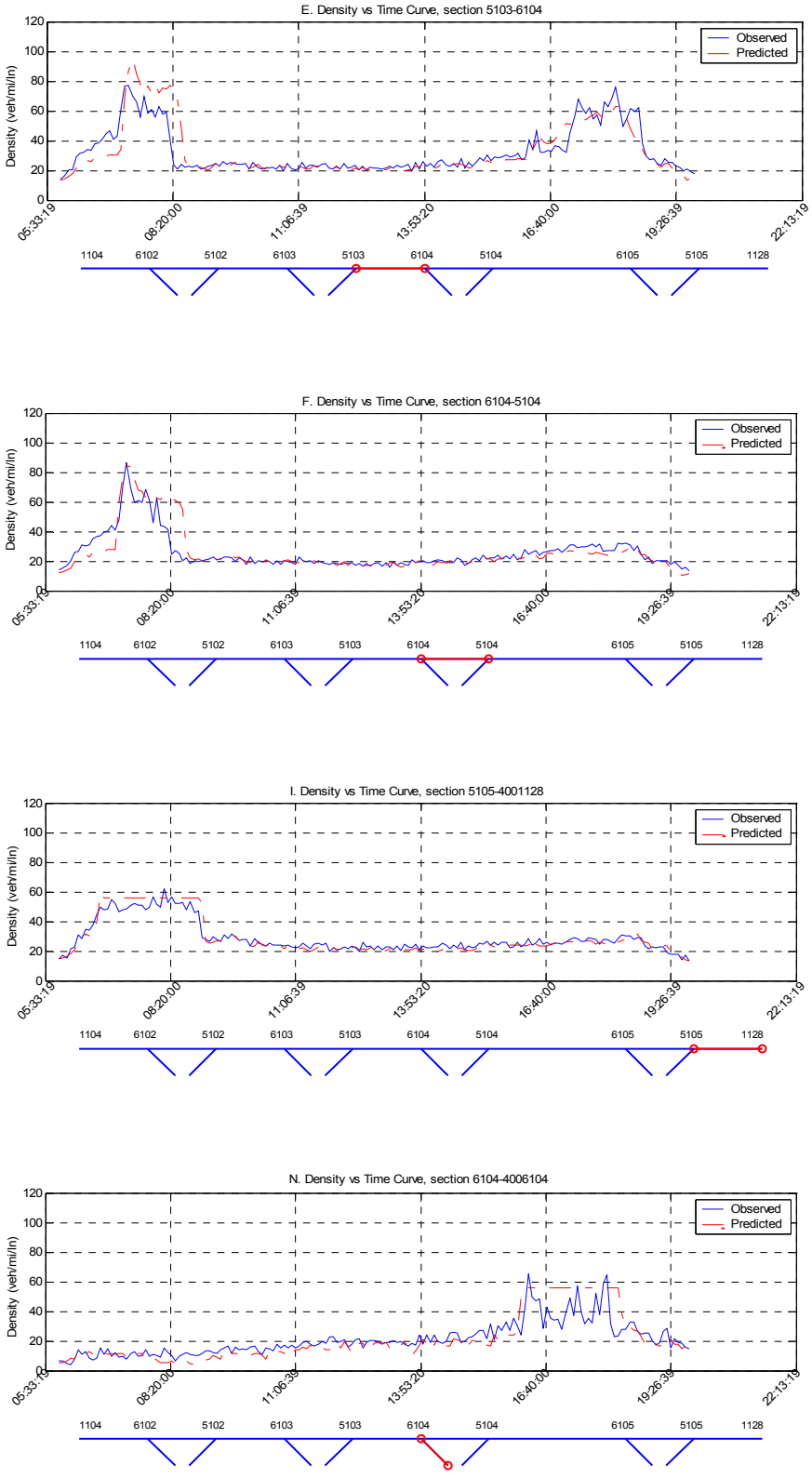


FIGURE 9-34 Continued

Comparison of prediction and observation in terms of congestion

Figure 9-35 confirms the findings above.

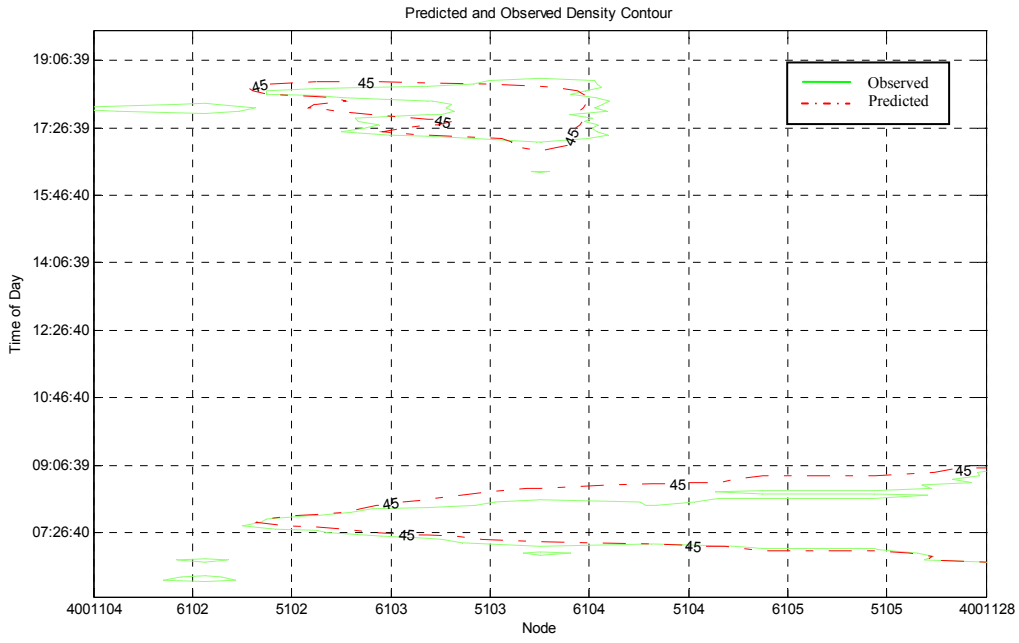


FIGURE 9-35 Comparison of congestion (network scenario, day 2)

Comparison of prediction and observation in 3D space

Figure 9-36 shows that the morning peak grew from node 4001128 all the way back to node 5102, while the afternoon peak was confined in a limited area.

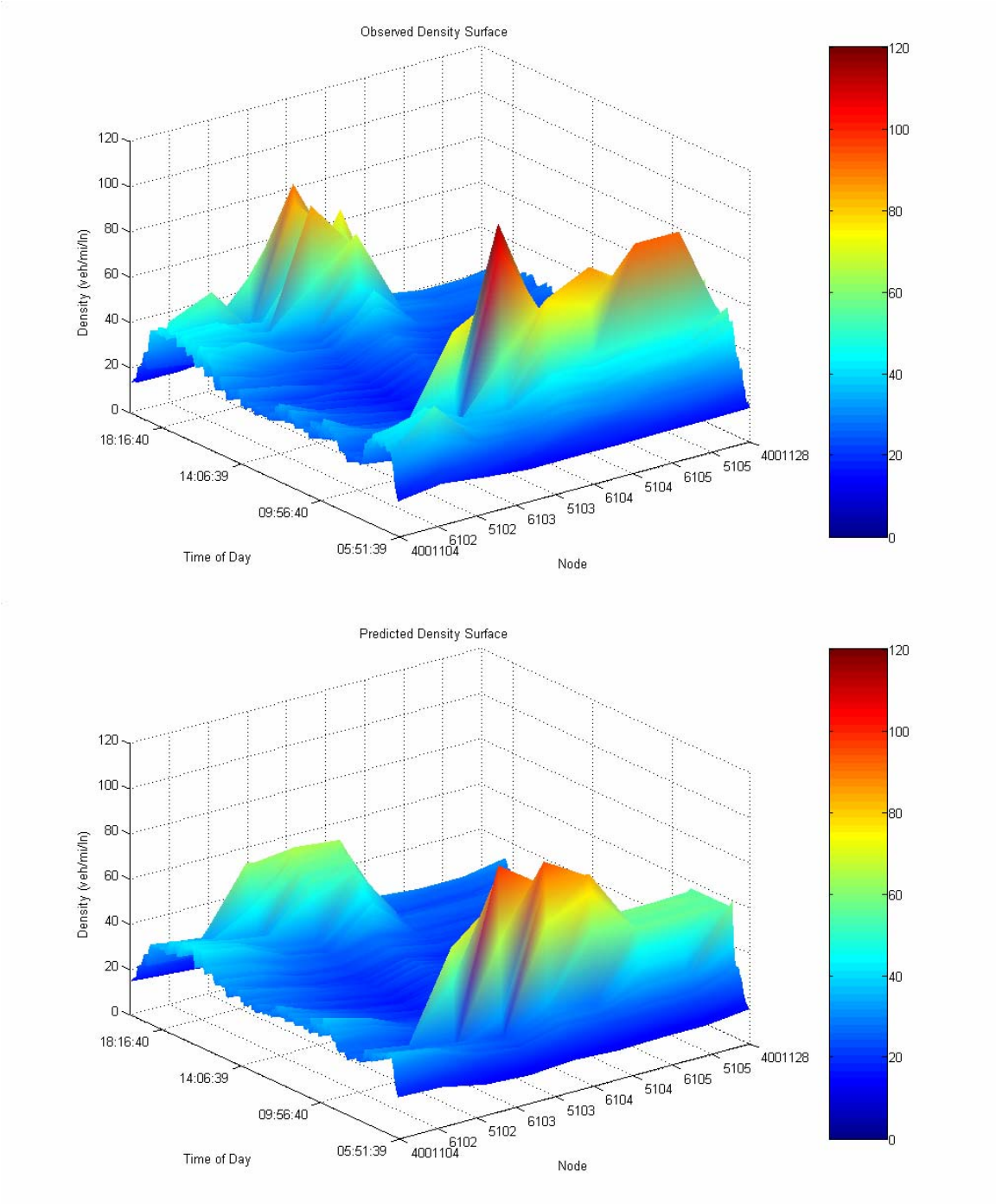


FIGURE 9-36 Comparison of density in 3D space (network scenario, day 2)

Plot of prediction against observation

Figure 9-37 plots the predicted density against the observed density.

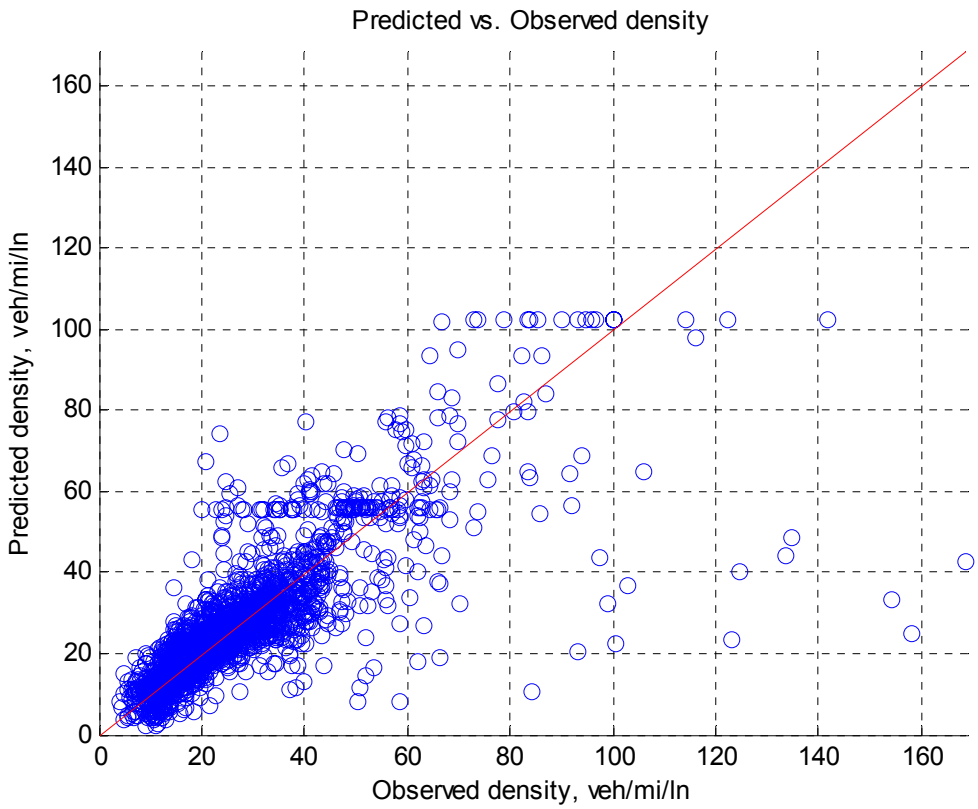


FIGURE 9-37 Plot of prediction against observation (network scenario, day 2)

Frequency of modeling error

Figure 9-38 shows that 81.4% of the total residuals are distributed within interval (-4, 14) and 14.1% within interval (-22, -4). Notice that the figure is a little bit skewed with a longer but skinny tail in the left which suggests that the model might have more occasions of under-prediction than those of over-prediction.

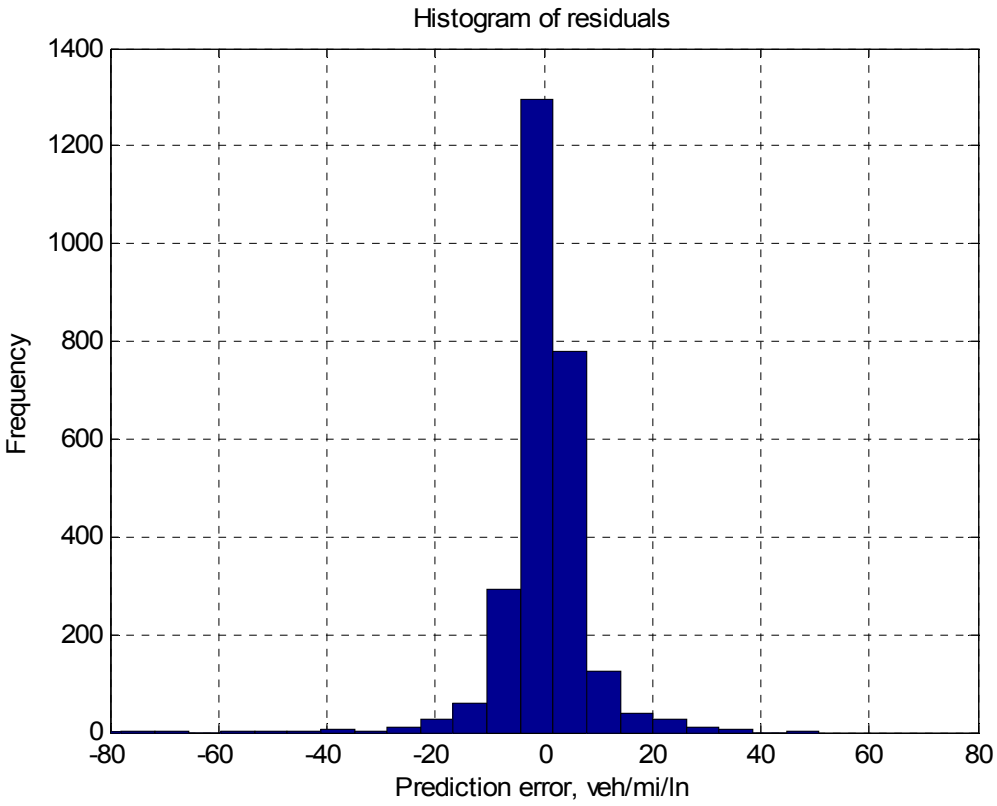


FIGURE 9-38 Frequency of modeling error (network scenario, day 2)

Comparison of Predicted and Observed Flow-Density Relationships

Figure 9-39 shows the predicted and the observed flow-density relationships on the same figure.

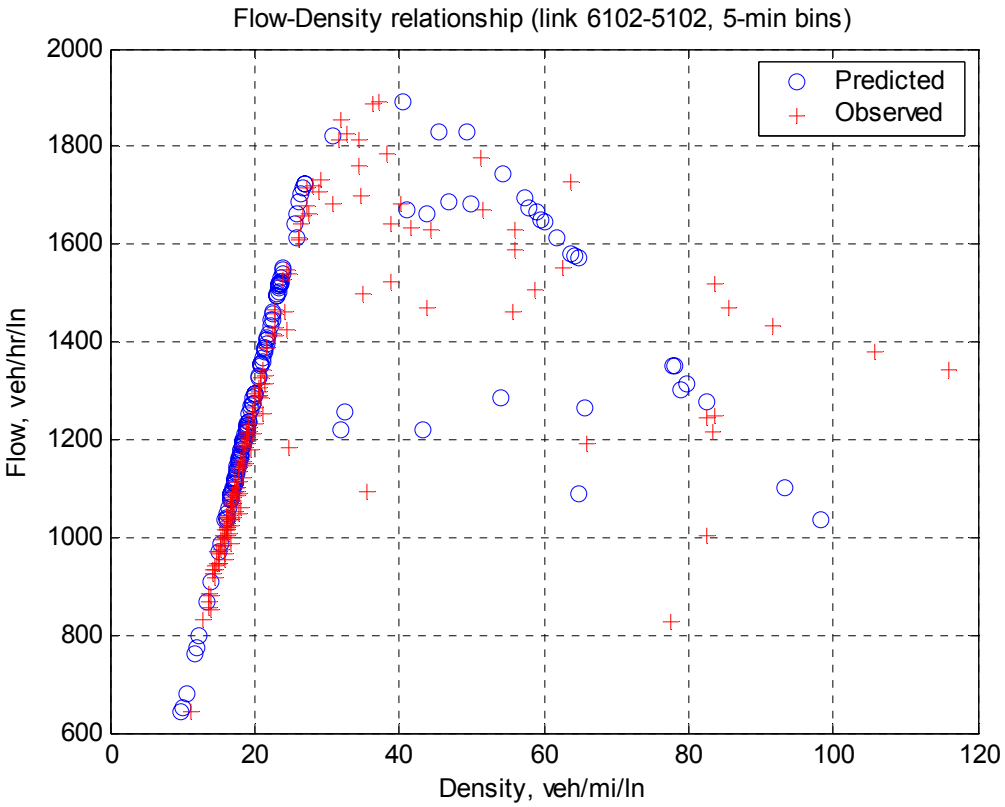


FIGURE 9-39 Comparison of flow-density relationships (network scenario, day 1)

9.4.2.2 Quantitative evaluation

Table 9-6 summarizes the result of simultaneous statistical test for test day 2.

TABLE 9-6 Simultaneous statistical test result of test day 2 (network scenario)

Item	Value
Batch size (m)	84
Batches (b)	32
Total samples (n)	2704
Significance level (α)	0.05
Grand mean ($\overline{W}$)	0.000231
Variance ($\hat{V}_B$)	0.004821
T-test Statistic (t_0)	0.018796
Critical t value ($t_{0.025,31}$)	2.04227
T-test result	Fail to reject H_0, i.e., $E[\overline{W}_i] = 0$.
Small number (ε)	0.0033
Critical χ^2 value ($\chi^2_{0.05,31}$)	45.0
95 % confidence interval for percentage modeling error	(-0.109, 0.122) $\times 100\%$

Interpreting the results is as before. Briefly, the result suggests that the mean of the modeling error is not statistically different than 0 and the 95% confidence interval for percentage modeling error is (-0.109, 0.122) $\times 100\%$. Similar to test day 1 of network scenario, the result of this test is also satisfactory given the existence of O-D estimation error.

9.5 SUMMARY

This chapter presents the results of empirical tests on the merging scenario, the diverging scenario, and the network scenario. The former two scenarios do not require O-D estimation while the last one does. O-D estimation may introduce extra error into model validation and the tests on these scenarios enable us to gain some insights in the model performance with and without O-D estimation.

We conducted the empirical tests by means of comparing the model output against field observations and we made the comparison based on the validation scheme proposed in Chapter 6.

In this chapter, we presented the results of six empirical tests with two tests for each scenario. Test results are promising – the model performs well in every test. Visual comparison generally shows a good agreement between the prediction and the observation, and statistical tests shows that the model is able to predict with reasonable accuracy and precision. The mean of modeling error is not statistically different than 0 and the percentage modeling error falls within the range of approximately ±8.4% at 95% confidence level when O-D estimation is absent and within the range of approximately ±12.3% at 95% confidence level with the presence of O-D estimation.

As assumed in the model, merging vehicles do compete for downstream capacity and queues may back up from any diverging branch and block traffic on the upstream mainline. The good agreement between the prediction and the observation in the empirical tests supports the validity of the proposed model.

CHAPTER 10

SUMMARY, CONCLUSION, AND FUTURE DIRECTIONS

This chapter summarizes this research, draws conclusions, and suggests future research directions.

10.1 SUMMARY

This research proposed an extension and a generalization to Newell's simplified theory of kinematic waves which was intended to help evaluate queue development and dissipation on freeway mainlines. The extension applies the simplified theory to a freeway system where multiple freeways and their on- and off-ramps are allowed and the generalization applies the simplified theory to a general transportation network where a link in the network can have multiple upstream and/or downstream branches.

We began the research with a literature review with a special interest in macroscopic / continuum flow models and their solutions. We constructed a picture of research in this area at a historical perspective and position our research in this context.

We summarized the simplified theory by reviewing its underlying assumptions, the wave propagation rules, and the computation procedure of modeling traffic evolution on a freeway mainline.

We made the proposed extension and generalization by relaxing some of the restrictive assumptions in the simplified theory. We proposed a CBWFQ merge model and a CBWS diverge model, together with their generalized forms, to replace the restrictive assumptions regarding ramp traffic behavior in the simplified theory. The proposed CBWFQ merge model, as well as its generalized form, allows queuing on any of the merging branches and vehicles on these merging branches compete for downstream supply. The basic idea of this model is that every merging branch is associated with a fair share of the downstream supply. We then satisfy merging branches with less-demanding traffic first. After this, we recalculate the fair shares of the remaining merging branches and redistribute the remaining downstream supply among these branches accordingly. If a merging branch depletes its fair share, the branch has to wait until other less-demanding merging branches have been satisfied and then make use of whatever downstream supply leftover. The proposed CBWS diverge model, as well as its generalized form, allows queuing on any of the diverging branches and the queue from any diverging branch can back onto the upstream mainline, further constraining traffic there. When splitting the outflow of the upstream mainline among the diverging branches, we distribute the outflow proportionally to the amounts that these branches can receive.

Next, we discussed some general issues such as the expansion of the K-Wave domain as well as the revised notation to accommodate more complicated roadway structure. With these, we formulated the proposed extension to the simplified theory based on 5 basic building blocks, i.e., an entrance, an exit, a mainline, a merge, and a diverge, and we assumed that a freeway system can be represented by certain

combination of these basic building blocks. We formulated the generalization of the simplified theory based on a generic building block where multiple upstream and downstream links are allowed, and we assumed that a general transportation network can be represented by certain combination of the basic building blocks which are special cases of the generic building block.

This research proposed an extension, which applies to freeway networks, and a generalization, which applies to general transportation networks, to the simplified theory. We only validated the extension to the simplified theory in the research. We discussed the methodology of model validation focusing on issues of data preparation and model validation scheme. In data preparation, we identified the data needs of the model validation and discuss sources of data to meet the needs. We also discussed how to prepare model input data which include three pieces, i.e., geometry data, traffic characteristics data, and time-varying O-D flows. In model validation scheme, we proposed using both qualitative and quantitative means to evaluate model performance. For the latter, we developed a simultaneous statistical inference technique which evaluates both the accuracy and precision of model performance. Considering the impact of O-D estimation error on the model validation, we proposed testing on three scenarios of the model, i.e., a merging scenario, a diverging scenario, and a network scenario, with the first two do not require O-D estimation and the last does. By this way, we obtained some insights in how our model performs with and without O-D estimation.

The field observations used in the model validation came from GA 400. Automated traffic surveillance systems cover the portion that serves as our study site and provide automated data over time and space. We discussed the study site and from it we choose

three smaller portions – the test sites – with each of which corresponds to a scenario we plan to test. Geometry data and traffic characteristics data of these sites were prepared.

We also prepared O-D flows of these sites. The test sites corresponding to the merging and the diverging scenarios do not require O-D estimation. O-D flows of these test sites can be synthesized directly from observed entrance or exit link flows. The test site corresponding to the network scenario requires O-D estimation. We selected an O-D estimator from the existing ones based on our needs and prepare the time-varying O-D flows of the test site using the selected estimator.

Based on the test sites and the data prepared, we conducted model validation by empirically comparing model output against field observations following the model validation methodology developed before. We found that traffic from merging branches do compete for downstream capacity and queues can back up from any diverging branch and block traffic on the upstream mainline. Both qualitative and quantitative means of evaluation showed that the model output agrees well with the field observations. The comparison results are satisfactory even with the existence of O-D estimation error. The good agreement between the prediction and the observation in the empirical tests supports the validity of our model.

10.2 CONCLUSIONS

Our experiences gained from empirical studies at GA 400 suggest the following:

The proposed model works properly under a wide range, including congested and uncongested conditions, and replicates the real system with reasonable accuracy. However, care must be taken to the conditions with extremely light traffic (e.g., density is

less than 10 veh/mi/ln) and extremely dense traffic (e.g., density is greater than 100 veh/mi/ln) because these conditions might be where the proposed model fails to work properly, as generally suggested by the empirical results in Chapter 9.

O-D estimation is a challenge to applications of the proposed model, especially when the network under study becomes complicated. This is because the O-D flows generated from O-D estimators might not be satisfactory enough to feed in subsequent applications. This research proposed a post-processing strategy to calibrate the initial O-D estimation results so as to minimize the impact of O-D estimation error on model validation. Since this calibration strategy is not part of the selected O-D estimator and a calibration strategy may vary from study to study, we are unable to make recommendations on the suitability of the proposed strategy to other studies at this point.

Though we proposed both the extension and the generalization of the simplified theory, only the extension (i.e., the portion of the model that deals with a freeway system) has been empirically tested. The generalization provides the theoretical potential of modeling general transportation networks. However, formal validation is recommended before use.

Because the proposed model relies on the conditions of upstream nodes to determine the condition at a downstream node, users are cautioned again including loops in their applications without any special treatment because a loop may create an inter-dependent situation where, when processing a node, one needs to know another node which, in turn, depends on the former.

Conceptually, the proposed model merits some advantages that are better than the CTM model. For example, the less restrictive requirement on link length (corresponding

to cell size in CTM) and the use of full O-D paths (corresponding to turn percentages in CTM) to move vehicles through a traffic network.

10.3 FUTURE DIRECTIONS

In conclusion of this research, we identify several future research directions:

Real-time Application: We have implemented the proposed model using JAVA and XML technologies and the resulted software can be used as a stand-alone program to perform various traffic studies. However, the more attractive side of the model is its ability to work real-time and such an application should be investigated as soon as possible.

Validation of the generalization: In this study, we developed the extension and the generalization to the simplified theory. We tested the extension only, and the validation of the generalization is expected as soon as possible.

Triangular (or trapezoidal) flow-density relationship: The simplified theory consists of four underlying assumptions. In this study, we relaxed two of them, i.e., the merge model, the diverge model, and extended the O-D flows from 1D to 2D domain. It would be interesting to go one step further by relaxing the triangular (or trapezoidal) flow-density relationship and see what happens.

Improving O-D estimation: In this study, we noted the existence of O-D estimation error and we developed a strategy to minimize its impact. However, better O-D estimation results make more sense and deserve further investigation.

Improving field observations: In this research, we also recognize the existence of observation error and we either try to avoid bad data or use it as if it were accurate. However, we may have more choices if we can improve the quality of the field

observations by means of data imputation which is a method that fills the gaps and correct erroneous data with statistically sound values.

Comparison of this research and CTM: This research and CTM share something in common and exhibit some differences. More importantly, both models address the same type of problem. It is quite natural to ask "how do they compare to each other?" Though we have shown some conceptual comparison results, empirical comparison is more desirable and such a study is expected to be carried out as soon as possible.

BIBLIOGRAPHY

Ashok, K. and Ben-Akiva, M. E. (2002) Estimation and Prediction of Time-Dependent Origin-Destination Flows with a Stochastic Mapping to Path Flows and Link Flows. Transportation Science, Volume 36 Issue 2, pp. 184-198.

Babcock, P. S., et al. (1984) The Role of Adaptive Discretization in a Freeway Simulation Model. In Transportation Research Record, 971, TRB, National Research Council, Washington, D.C., 1984, pp. 80-92.

Banks, J.H. (2000) Are Minimization of Delay and Minimization of Freeway Congestion Compatible Ramp Metering Objectives? Transportation Research Record 1727, pp. 112-119.

Bhattacharjee, D., Sinha, K. C., Krogmeier, J. V. (1998a) Real-time Freeway O-D Prediction Algorithm under ATIS Environment. Proceedings of the Fifth International Conference of Applications of Advanced Technologies in Transportation. pp. 147-159.

Bhattacharjee, D., Sinha, K. C., Krogmeier, J. V. (1998b) Modeling the Effects of ITS on Route Switching Behavior in a Freeway and its Effects on the Performance of O-D Prediction Algorithm. Preprint CD-ROM of the 77th TRB Annual Meeting. Washington, DC.

Bick, J. H. and Newell, G. F. (1960). A continuum model for two-directional traffic flow. Quarterly of Applied Mathematics. 18(2) pp. 191-204.

Brenninger-Gothe, M., Jornsten, K. O., and Lundgren, J. T. (1989) Estimation of Origin-destination Matrices from Traffic Counts Using Multiobjective Programming Formulations. Transportation Research. 23B, no. 4. pp. 257-269.

Chin, S., H. Hwang and T.Pei (1994) Using Neural Networks to Synthesize Origin-Destination Flows in a Traffic Circle, Transportation Research Record, 1457. pp.134-142.

Cremer, M. and Keller, H. (1987) A new Classes of Dynamic Methods for the Identification of the Origin-Destination Flows. Transportation Research 21B, no. 2, pp.117-132.

Daganzo, C. F. (1994) The Cell-Transmission Model: A Dynamic Representation of Highway Traffic Consistent with the Hydrodynamic Theory. Transportation Research, 28B, pp. 269-288.

Daganzo, C. F. (1995a) A Finite Difference Approximation of the Kinematic Wave Model of Traffic Flow. Transportation Research, Vol. 29B, No.4, pp. 261-276.

Daganzo, C. F. (1995b) The Cell Transmission Model, Part II: Network Traffic. Transportation Research 29B(2), pp. 79-93.

Daganzo, C. F. (1995c) Requiem For Second-order Fluid Approximations of Traffic Flow, Transportation Research - B 29(4), pp. 277-286.

Daganzo, C. F. (1997) A Continuum Theory of Traffic Dynamics for Freeways with Special Lanes. Transportation Research, Vol. 21B, No. 2, pp. 83-102.

Daganzo, C. F. (1997) Fundamentals of Transportation and Traffic Operations, Pergamon.

Daganzo, C. F. (1999) The Lagged Cell-transmission Model. Proc. 14th Int. Symp. on Transportation and Traffic Theory. (Ceder, A., editor) pp. 81-106, Pergamon, New York, N.Y.

Davis, G. A. and Yu, P. (1994) Estimating Freeway Origin-Destination Patterns Using Automation Traffic Counts. Transportation Research Record, 1457, pp. 127-133.

Davis, G. A., and Nihan, N. L. (1991) Stochastic Process Approach to the Estimation of Origin-destination Parameters from Time Series of Traffic Counts. Transportation Research Record 1328, pp. 36-42.

Dawdy, D. R., and Matalas, N. C. (1964) Statistical and Probability Analysis of Hydrologic Data, part III: Analysis of Variance, Covariance and Time Series, in Ven Te Chow, ed., Handbook of applied hydrology, a compendium of water-resources technology: New York, McGraw-Hill Book Company, p. 8.68-8.90.

del Castillo, J. M., Pintado, P., and Benitez, F. G. (1993) A Formulation for the Reaction Time of Traffic Flow Models. In Transportation and Traffic Theory . Proc 12th International Symposium on Transportation and Traffic Theory (C. Daganzo, ed.). Berkeley, CA. pp. 387-405

Dixon, M. (2000) Incorporation of Automatic Vehicle Identification Data into the Synthetic OD Estimation Process, Ph.D. Thesis. Texas A&M University, College Station, TX.

Edie, L. C. and Baverez, E. (1965) Generation and Propagation of Stop-start Traffic Waves. Proceedings of the Third International Symposium on the Theory of Traffic Flow. June 1965.

Edie, L. C. and Foote, R. S. (1958) Traffic Flow in Tunnels. Proc. Highway Research Board, 37, pp. 334-344.

Florian, M. and Chen, Y. (1991) A Bilevel Programming Approach To Estimating O-D Matrix By Traffic Counts. Université de Montréal, Centre de recherche sur les transports.

Fusco, G., and Recchia, R. (1997) O-D matrix estimation and incident detection in urban areas using artificial neural networks. Transportation Systems 1997 (ed. Papageorgiou, M., Pouliezos, A.), Chania, Greece. Volume 2, pp. 907-912.

Gerlough, D. L. and Huber, M. J. (1975) Traffic Flow Theory: A Monograph. Transportation Research Board, S.R.165.

Geva, I., Hauer, E., and Landau, U. (1982) Estimation of Cross-cordon Origin-destination Flows from Cordon Studies. Transportation Research Record 891, pp. 5-10.

Geva, I., Hauer, E., and Landau, U. (1982) Estimation of Cross-Cordon Origin-Destination Flows from Cordon Studies. Transportation Research Record, 891, pp. 5-10.

Goldsman, D. and Tokol G. (2000) Output analysis: output analysis procedures for computer simulations. Proceedings of Winter Simulation Conference 2000, pp. 39-45.

Hai. Y. (1995) Heuristic algorithms for the bilevel origin-destination matrix estimation problem. Transportation Research, 29B, no. 4, pp. 231-242.

Hazelton, M. L. (2000) Estimation of Origin-destination Matrices from Link Flows on Uncongested Networks. Transportation Research. 34B, no. 7. pp. 549-566.

Hu, S. R. (1996) An Adaptive Kalman Filtering Algorithm for the Dynamic Estimation and Prediction of Freeway Origin-destination Matrices. Ph. D. Dissertation, Purdue University.

Hurdle, V. F., Son, B. (2000) Road Test of a Freeway Model. Transportation Research, Vol. 34A, No. 7, pp. 537-564.

Ikenoue, K., Takeuchi, H., and Yasui, K. (1995) Use of AVI Information Linked up with Detector Output in Travel Time Prediction and O-D Flow Estimation. Steps Forward. Intelligent Transport Systems World Congress. Yokohama, Japan. Volume 1, pp. 94.

Jin, W. L. and Zhang, H. M. (2003) On the Distribution Schemes for Determining Flows Through a Merge. Transportation Research, Vol. 37B, No. 6, pp. 521-540.

Kang, Y. (1999) Estimation and Prediction of Dynamic Origin-destination (O-D) Demand and System Consistency Control for Real-time Dynamic Traffic Assignment Operation. PhD dissertation, The University of Texas at Austin.

Kerner, B. S. and Konh?user, P. (1993) Cluster Effect in Initially homogeneous Traffic Flow, Physical Review, E48(4), pp. 2335-2338.

Kühne, R. D. (1984) Macroscopic freeway model for dense traffic - stop-start waves and incident detection. Ninth International Symposium on Transportation and Traffic Theory, pp. 20-42.

Kühne, R. D. (1989) Freeway control and incident detection using a stochastic continuum theory of traffic flow. Proceedings 1st International Conference on Applied Advanced Technology in Transportation Engineering, San Diego, CA, pp. 287-292.

Kwon, E., and Michalopoulos, P. (1995) Macroscopic Simulation of Traffic Flows in Complex Freeway Segments on a Personal Computer. In 1995 Vehicle Navigation & Information Systems Conference Proceedings, pp. 342-345.

Lebacque, J. P. (1996) The Godunov Scheme and What it Means for First Order Traffic Flow Models. Proceedings of the 13th International Symposium on Transportation and Traffic Theory. Lesort, J.B. (Ed.), Elsevier, Lyon, France, pp. 647-677.

Leo, C. J., and Pretty, R. L. (1992) Numerical simulation of macroscopic continuum traffic models. Transportation Research, 268, pp 207-220.

Leonard, J. D. (1997) Computer Implementation of a "Simplified Theory of Kinematic Waves." California Department of Transportation Report FHWA/CA/TO-98-01, Sacramento, CA.

Leonard, J. D. (2001) A tool for evaluating freeway congestion. Online paper located at http://traffic.ce.gatech.edu/gtwaves. Accessed June 2001.

LeVeque, R. J. (1992) Numerical Methods for Conservation Laws, second edition, Birkhauser Verlag, Basel, Switzerland. Lectures in Mathematics ETH.

Li, B. and Moor, B. D. (1999) Recursive Estimation Based on the Equality-constrained Optimization for Intersection Origin-destination Matrices. Transportation Research. 33B, no. 3. pp. 203-214.

Lighthill, M. J. and Whitham, J. B. (1955) On Kinematic Waves II: A Theoryof Traffic Flow on Long Crowded Roads, Proceedings of the Royal Society, Vol. A229, pp. 317-345.

Lo, H. P., Zhang, N., and Lam, W. H. K. (1996) Estimation of an Origin-destination Matrix with Random Link Choice Proportions: A Statistical Approach. Transportation Research. 30B, no. 4. pp. 309-324.

Madanat, S., Krogmeier, J. and Hu, S. R. (1995) An Enhanced Kalman Filtering Algorithm for Dynamic Freeway OD Matrix Estimation and Prediction. Proceedings of the Fourth International Conference on Applications of Advanced Technologies in Transportation Engineering. Capri, Italy. pp. 423-428.

Maher, M. J., Zhang, X., and Vlietb, D. V. (2001) A Bi-level Programming Approach for Trip Matrix Estimation and Traffic Control Problems with Stochastic User Equilibrium Link Flows. Transportation Research. 35B, no. 1. pp. 23-40.

May, A. D., (1998) FREQ User Manual, California Department of Transportation, Berkeley.

May, A. D., et al. (1991) Integrated system of freeway corridor simulation models. Transportation Research Record No. 1320. pp. 177-189

Michalopoulos, P. G. (1984) Dynamic freeway simulation program for personal computers, Transportation Research Record 971, pp. 68-79.

Michalopoulos, P. G. (1984) Dynamic Freeway Simulation Program for Personal Computers. In Transportation Research Record 971, TRB, National Research Council, Washington, D.C.,

Michalopoulos, P. G., and Lin, J. (1986) Integrated Modeling of Freeway Flow and Application to Microcomputers. In Transportation Research Record 1091, TRB, National Research Council, Washington, D.C., pp. 25-28.

Michalopoulos, P. G., Beskos, D. E., Lin, J. K. (1984) Analysis of Interrupted Traffic Flow by Finite Difference Methods. Transportation Research, Vol18B, pp. 409-421.

Michalopoulos, P. G., Beskos, D. E., Yamauchi, Y. (1984) Multilane traffic flow dynamics: some macroscopic considerations. Transportation Research 18B, pp. 377-395.

Michalopoulos, P. G., Lin, J. K., and Beskos, D. E. (1987) Integrated Modelling and Numerical Treatment of Freeway Flow. Appl. Mathem. Model, Vol. 11, No. 401, pp. 447-458.

Michalopoulos, P. G., Yi, P. and Lyrintzis, A. S. (1993) Continuum Modelling of Traffic Dynamics for Congested Freeways, Transportation Research B 27(4), pp. 315-332.

Michalopoulos, P. G., Yi, P., Lyrintzis, A. S. (1993) Continuum modelling of traffic dynamics for congested freeways. Transportation Research 27B, pp. 315-332.

Michalopoulos, P., Kwon, E., and Kang, J-G. (1991) Enhancement and Field Testing of a Dynamic Freeway Simulation Program. In Transportation Research Record 1320, TRB, National Research Council, Washington, D.C., pp. 203-215.

Munjal, P. K. and Pipes, L. A. (1971). Propagation of on-ramp density perturbations on unidirectional two-and three-lane freeways. Transportation Research, 5, pp. 241-255.

Newell, G. F. (1962) Theories of Instability in Dense Highway Traffic. Journal of Operations Research Soc. Japan 5, pp. 9-54.

Newell, G. F. (1993a) A simplified theory on kinematic waves in highway traffic, Part I: general theory. Transportation Research, Vol. 27B, No. 4, pp. 281-287.

Newell, G. F. (1993b) A simplified theory on kinematic waves in highway traffic, Part II: queueing at freeway bottlenecks. Transportation Research, Vol. 27B, No. 4, pp. 289-303.

Newell, G. F. (1993c) A simplified theory on kinematic waves in highway traffic, Part III: multi-destination flows. Transportation Research, Vol. 27B, No. 4, pp. 305-313.

Newell, G. F. (1999) Delays Caused by a Queue at a Freeway Exit Ramp. Transportation Research, Vol. 33B, pp. 337-350.

Ni, D. and Leonard, J. D. (2004a) A Kinematic Wave Model For Merge Queuing. Proceedings of the International Conference on Computing, Communications and Control Technologies (CCCT'04, ed. Hsing-Wei Chu, et al), Austin, Texas, USA, August pp. 14-17.

Ni, D. and Leonard, J. D. (2004b) Simplified Kinematic Waves at a Diverge. Proceedings of the 8th World Multiconference on Systemics, Cybernetics and Informatics (SCI 2004, ed. Callaos, N., Horimoto, K., Chen, J., and Chan A. K. S.), International Institute of Informatics and Systemics. Vol VII, pp. 425-430.

Ni, D. and Leonard, J. D. (2004c) Simulation of Freeway Merging and Diverging Behavior. Proceedings of the 2003 Winter Simulation Conference (ed. S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, eds), Institute of Electrical and Electronics Engineers, Piscataway, NJ.

Ni, D., Leonard, J. D., Guin, A., and Williams, B. M. (2004) A Systematic Approach for Validating Traffic Simulation Models. Transportation Research Record, TRB, National Research Council, Washington, D.C., accepted for publication February, 2004.

Nihan, N. L. and Davis, G. A. (1987) Recursive Estimation of Origin-Destination Matrices from Input/Output Counts. Transportation Research B Vol. 21, No. 2, pp. 149-163

Nihan, N. L. and Davis, G. A. (1989) Application of Prediction Error Minimization and Maximum Likelihood to Estimation Intersection O-D Matrices from Traffic counts, Transportation Science, 23 (2), pp. 77-90.

Payne, H. J. (1971) Models of freeway traffic and control. Simulation Council Proceedings 1, pp. 51-61.

Payne, H. J. (1979) FREFLO: A macroscopic simulation model for freeway traffic. Transportation Research Record 722, pp. 68-77.

Phillips, W. F. (1979) A Kinetic Model for Traffic Flow with Continuum Implications, Transportation Planning and Technology 5, pp. 131-138.

Prevedouros, P. D., Li, H. (2000) Comparison of Freeway Simulation with Integration, Kronos, and Kwaves. Proceedings of Fourth International Symposium on Highway Capacity. Maui, Hawaii. pp. 96-107.

Prigogine, I. (1961) A Boltzmann-Like Approach to the Statistical Theory of Traffic Flow. In Theory of traffic Flow (ed. R. Herman), Elsevier, Amsterdam, p. 158.

Rathi, A. K., Lieberman, E. B., and Yedlin, M. (1987) Enhanced FREFLO: Modeling of Congested Environments. In Transportation Research Record 1112, TRB, National Research Council, Washington, D.C., 1987, pp. 61-71.

Richards, P. I. (1956) Shock waves on the highway. Operations Research 4, pp. 42-51.

Richards, P. I. (1956) Shockwaves on the Highway. Operations research 4, pp. 42-51.

Ross, P. (1988) Traffic dynamics. Transportation Research 22B, pp. 421-435.

Sherali, H. D. and Park T. (2001) Estimation of dynamic origin-destination trip tables for a general network. Transportation Research. 35B, no. 3, pp. 217-235.

Sherali, H. D., Arora, N., and Hobeika, A. G. (1997) Parameter optimization methods for estimating dynamic origin-destination trip-tables. Transportation research. 31B, no. 2. pp. 141-157.

Sherali, H. D., Sivanandan, R., and Hobeika, A. G. (1994) A Linear Programming Approach for Synthesizing Origin-destination Trip Tables from Link Traffic Volumes. Transportation research. 28B, no. 3. pp. 213-233.

Sivanandan, R., Nanda, D., and Arnold, E. D., Jr (1999) Method to Enhance Performance of Synthetic Origin-destination Trip Table Estimation Models. Transportation Research Record 1676, pp. 124-135.

Son, B. (1996) A Study of G.F. Newell's Simplified Theory of Kinematic waves in Highway Traffic. Ph.D. Thesis, Department of Civil Engineering, University of Toronto, Canada.

Sun, C., and Porwal, H. (2000) Dynamic Origin/Destination Estimation Using True Section Densities. PATH research report, UCB-ITS-PRR-2000-5.

Treiber, M., Hennecke, A. and Helbing, D. (1999) Derivation, Properties and Simulation of A Gas-kinetic-based, Nonlocal Traffic Model, Physical Review E 59(1), pp. 239-253.

Turnquist, M. and Gur, Y. (1979) Estimation of Trip Tables from Observed Link Volumes. Transportation Research Record, 730. pp. 1-6.

Van Aerde, M., Yagar, S., Ugge, A., and Case, E. R. (1987) A Review of Candidate Freeway-Arterial Corridor Traffic Models. In Transportation Research Record 1132, TRB, National Research Council, Washington, D.C., pp. 53-65.

Van Der Zijpp, N. J. (1997) Dynamic Origin-destination Matrix Estimation from Traffic Counts and Automated Vehicle Identification Data. Transportation Research Record 1607, pp. 87-94.

Van Der Zijpp, N. J., and Hamerslag, R. (1993) The Real Time Estimation of Origin-destination Matrices for Freeway Cooridors. 26th International Symposium on Automotive Technology & Automation. Aachen, Germany. pp. 297-305.

Van Der Zijpp, N. J., and Hamerslag, R. (1994) Improved Kalman Filtering Approach for Estimating Origin-destination Matrices for Freeway Corridors. Transportation Research Record 1443, pp. 54-64.

Whitham, G. B. (1974) Linear and Nonlinear Waves, John Wileyand Sons Inc, New York, NY.

Yager, S. (1975) CORQ - A Model for Predicting Flows and Queues in a Road Corridor. In Transportation Research Record 533, TRB, National Research Council, Washinton, D.C., pp. 77-87.

Yang, H., Akiyama, T. and Sasaki. T. (1998) Estimation of time-varying origin-destination flows from traffic counts: A neural network approach. Mathematical and Computer Modeling 27 (9-11), pp. 323-334.

Yi, J., Lin, H., Alvarez, L. and Horowitz, R. (2003) Stability of macroscopic traffic flow modeling through wavefront expansion. Transportation Research B, 37(7), pp. 661-679.

Zhang, H. M. (1998) A Theory of Nonequilibrium Traffic Flow, Transportation Research B 32(7), pp. 485-498.

Zhang, X. and Maher, M. J. (1998) The Evaluation and Application of a Fully Disaggregate Method for Trip Matrix Estimation with Platoon Dispersion. Transportation Research. 32B, no. 4. pp. 261-276.

VITA

Daiheng Ni was born in China. He loves vehicles and he has been devoted to research in vehicles for a long time. Vehicle engineering and transportation engineering are the two keywords that heavily influenced his life.

Among others, he is proud of the five degrees he earned:

- Ph.D. in Civil Engineering from Georgia Institute of Technology
- M.Sc. in Industrial Engineering from Georgia Institute of Technology
- M.Sc. of Civil Engineering from Georgia Institute of Technology
- M.Sc. in Mechanical Engineering from Beijing Agricultural Eng. University
- B.Sc. in Mechanical Engineering from Jilin University of Technology

His major research interests include Intelligent Transportation System (ITS), transportation modeling and simulation, transportation system operation and control, traffic flow theory, transportation information systems and technology, transportation safety, transportation logistics systems, and transportation system optimization.