Georgia Tech Sponsored Research

6A019799

Project	E-20-G98	L416	
Project director	LEONARD	JOHN	L5
Research unit	CIVIL ENGR		
Title	QUANTIFYING SECONDARY CRASHES ON ATLANTA HIGHWAYS		
Project date	7/17/2003		

E-20-G98
#4

Project Final Report

Technical Enhancements to the KWAVES Software

Prepared for:

California Department of Transportation
1120 N Street, Sacramento, CA

by:

John D. Leonard II and Daiheng Ni

School of Civil and Environmental Engineering
Georgia Institute of Technology
Atlanta, GA 30332

April 23, 2004

ABSTRACT

If roadway traffic is viewed as a compressible fluid, kinematic wave models are ideal to help analyze traffic evolution. Among the many variants of the seminal LWR model, the simplified theory of kinematic waves is particularly simple and elegant. It can be summarized in several standard steps and involves only simple mathematical operations such as addition and subtraction, features of which are particularly friendly and efficient for computer simulation. To maximize the advantages of this model, extensions are attempted to generalize this model so that it applies to a general transportation network rather than freeway mainline as intended by the original model. A generalized capacity-based weighted fair queuing (CBWFQ) merge model and a generalized contribution-based weighted splitting (CBWS) diverge model are formulated to deal with merging and diverging traffic. Based on these, a procedure is developed to model a general node where multiple upstream links and multiple downstream links are allowed.

EXECUTIVE SUMMARY

This report documents a study to extend and generalize a macroscopic traffic flow theory discovered by Newell (1993a, 1993b, and 1993c). This study was conducted by a team at the School of Civil and Environmental Engineering, Georgia Institute of Technology. The major goal of this study is to explore the possibility of applying Newell's simplified theory of kinematic waves to more general cases rather than only to freeway mainline intended by the original theory. In addition, this study is expected to help better understand freeway traffic operation and evolution and, thus, provide better modeling and control strategies for freeway traffic.

Newell's theory provides a continuous-time but discrete-space approximation to an earlier traffic model, the L-W-R theory (Lighthill and Whitham, 1955; Richards, 1956), which is continuous in both time and space. The study on extension and generalization of Newell's theory starts with its assumptions. There are four assumptions in the theory: the first, also the most fundamental, assumption defines the relation between flow and density; the second assumption describes on-ramp traffic behavior such that on-ramp traffic has higher priority over mainline traffic; the third assumption gives the off-ramp traffic behavior such that no congestion ever backs up from off-ramp and exiting traffic experiences the same travel time on the mainline as through traffic; the fourth assumption provides the necessary boundary conditions based on which the theory works. However, field observation shows that the second and the third assumptions are not very realistic and this is the starting point of our study.

Relaxation of Newell's second assumption requires a realistic model to describe traffic behavior at a merge where several streams of traffic come into one. Existing merge models (Lebacque 1996, Daganzo 1995, Jin and Zhang 2003, Newell 1993b, Banks 2000) are either lack of realismor calibration-intensive. The proposed Capacity-based Weighted Fair Queuing (CBWFQ) merge model tries to retain the pleasant features of the existing models while staying away their disadvantages.

Relaxation of Newell's third assumption necessitates a diverge model that describes how one stream of traffic splits into several. A number of diverge models (Newell 1993b and 1999, Daganzo 1995 and 1997, Lebaque 1996) are identified. The major discrepancy of the above models pertains to the issue of whether the upstream traffic state should be uniform, i.e., traffic state is the same regardless of destinations. If not, how is traffic state distributed? Based on our fiend observation and traffic behavior analysis, a Contribution-based Weighted Splitting (CBWS) diverge model is proposed to overcome the shortcomings of the existing models.

Based on the proposed CBWFQ and CBWS models, procedures are formulated to model traffic operation at an entrance, an exit, a mainline node, a merge, and a diverge – the basic build blocks for a freeway network. More over, a procedure is also formulated to handle a general node where multiple upstream links and multiple downstream links are allowed. This essentially enables the modeling of a general transportation network.

Testing of the generalized model is performed after model implementation in which Java and XML technologies are employed. Empirical test data comes from the NAVIGATOR system which covers the major highways and freeways in Metro Atlanta area. Tests on three cases, i.e., merge scenario, diverge scenario, and corridor/network scenario, are conducted with each scenario tested on 2 days.

Empirical test results show that the relaxation on the assumptions regarding merging and diverging behavior is valid and the proposed CBWFQ and CBWS models are capable of replicating the reality with sufficient accuracy and precision. Suggested follow-up research is to move the model to on-line application.

TABLE OF CONTENTS

TABLES

FIGURES

CHAPTER 1
INTRODUCTION TO THE PROBLEM

1.1. Overview

Congestion is a key problem that traffic engineers are trying to mitigate. Building more roads, however, is unattractive not only because it is expensive but also it tends to induce more travel and discourage alternative transportation. Therefore, technology combined with smart development of highways make the most sense in terms of money and cost to environment, and it is now generally accepted that Intelligent Transportation Systems (ITS) is the cure of congestion.

At the core of ITS is the better use of traffic data to meet the demand in a timely fashion. Nowadays, most states have installed advanced traffic surveillance systems and traffic information constantly flows in day after day. Unfortunately, the expensive equipment doesn't seem to pay off because the data collected is not fully utilized. We do need something to turn the investment into dollar savings.

This research aims to improve the soundness of engineering judgment with regard to traffic operation and control. It addresses traffic engineers as well as general traveling public. For traffic engineers, this research answers questions such as "Given a regional freeway system and time-varying traffic demand, how well is the freeway facility serving the demand in terms of various measurements of effectiveness (MOEs)?" "Where is the bottleneck?" "How long does it take to dissipate a queue?" "If there is congestion and ramp-meters are used, what's the best ramp-metering strategy?" "In case of traffic accident, what's the best strategy to clean up the accident and resume traffic operation at the minimum cost in terms of predefined MOEs?" For the traveling public, the research answers questions such as "what are the prevailing travel speed and level of service?" "Given the demand and current traffic state, how long does it take if I enter the facility at A and exit at B?" "How long do I have to wait before the accident is cleaned up?"

To run the proposed model of this research, two inputs are generally required: information of roadway facility which is relatively static and information of traffic demand which is relatively dynamic. The requirements of this model are minimal Comparing with more detailed model such as CORSIM. Because of this, this model is light-weighted and possible for online application. On the other hand, the algorithm of the proposed model is very simple - basically some additions and comparisons. All these features enable the simulation to run ahead of real time and this means lots of opportunities for traffic engineers to relieve congestion. For example, just think about "what if we know where the bottleneck will be several minutes before the queue actually builds up?"

1.2. The Broad Picture

To better explain what this research is all about, it is necessary to place it in certain context. This is illustrated by drawing a broad picture and showing where this research fits in.

1.2.1 A Practical View

In a broad sense, this research addresses problems in traffic operation. Nowadays, more and more departments of transportation (DOTs) and local authorities rely on Intelligent Transportation Systems (ITS) to make more out of the existing transportation facilities and mitigate traffic congestion. Among the sub-systems of ITS are Advanced Traffic Management System (ATMS) and Advanced Traveler Information System (ATIS). Both address the general traveling public. To provide useful information to advise traffic control strategies and assist trip generation decisions, it is critical to be able to look forward, either short- or long-term, on traffic demand and predict how the facilities are going to serve the demand. Traffic simulation is an efficient and reliable way to perform this task. Also, it would be very helpful and informative is the above process can be made on-line and automated, the traveling public will benefit a lot from it and make advised decision when they make trips, which, in turn, helps mitigate congestion. The above framework is illustrated in Figure 1-1.

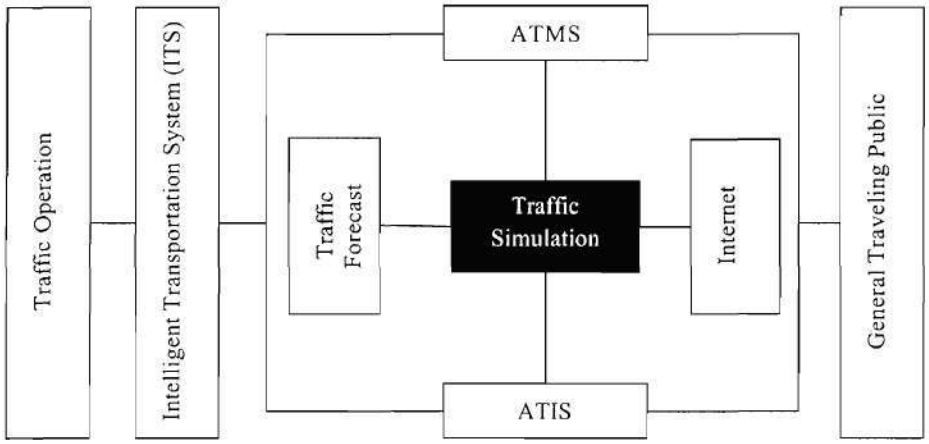


FIGURE 1-1 A Practical View

1.2.2 A Vertical View

In general, simulation is the dynamic representation of part of the world by building a computer model of it and moving it through time. The ultimate goal of simulation is the representation of part of the real world as realistic as possible and the modeling techniques largely depend on how the modeler views the reality. Traffic simulation is an example in case. Suppose one is observing traffic on an airplane in the sky, traffic behaves like fluid whose states (speed, flow, and density, etc.) propagate like waves. Here is the arena of macroscopic approach. If one lowers his/her plane, the sense of wave recedes and a feeling of particle looms. This is the scenario suited for mesoscopic approach. If the plane gets down even more, the scene is dominated by moving particles who interact with each other so as to maintain safe positions in the traffic stream. Here is the typical area where microscopic approach thrives. Now, what if one steps down to the earth and joins the traffic? The picture will be fundamentally different. There is neither wave nor particle, but a complicated nanoscopic system consisting of drivers, vehicles, and environment. They interact with each other, and the mechanism of which, if modeled, would provide insights into the mystery of traffic operation. This is an example of nanoscopic traffic simulation. The research topic (KWaves) that we are addressing here falls in the category of macroscopic traffic simulation. The above categorization is illustrated in Figure 1-2.

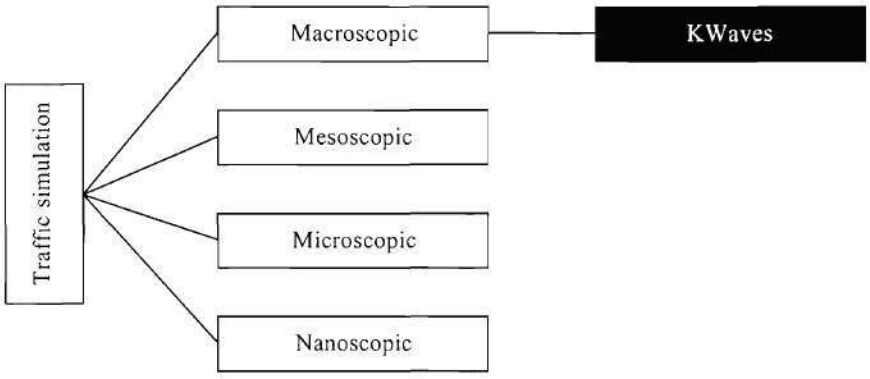


FIGURE 1-2 A Vertical View

1.2.3 A Historical View

Generally, macroscopic traffic simulation borrows ideas from fluid dynamics and treats traffic as continuous compressible one-dimensional fluid, such as gas. A lot of models are available to describe the properties of such fluid, among which are first-order models, which consider its kinematic properties, and higher-order models, which involve factors that resist movements such as viscosity and mass. In the middle 1950's, Lighthill, Whitham, and Richards formulated a seminal model, referenced now as L-W-R theory (Lighthill and Whitham, 1955; Richards, 1956), based on flow conservation and some functional relations between flow and density. This was essentially a first-order model which describes traffic evolution in continuous time-space domain. Nearly 30 years later, Newell (1993a, 1993b, and 1993c) discovered a simple and elegant solution, the Simplified Theory, to numerically approximate the L-W-R model whose solution is typically time-consuming. Slightly later, Daganzo (1994) proposed a sister model called Cell Transmission Model (CTM) that essentially solves the same problem. Newell's theory was proposed without giving empirical tests and this work was performed by Son and Hurdle (Son 1996, Hurdle and Son 2000). Leonard (1997; 1998) coded Newell's theory into software GTWaves, which bridges the theory and its application. Considering that Newell's simplified theory applies only to freeway mainline and some of its simplifications and assumptions, especially those regarding freeway merge and diverge, may not be very true in reality, some extension and generalization are made to enable the simplified theory applicable to freeway networks or even a general transportation network. The relation of the above discussion is summarized in Figure 1-3.

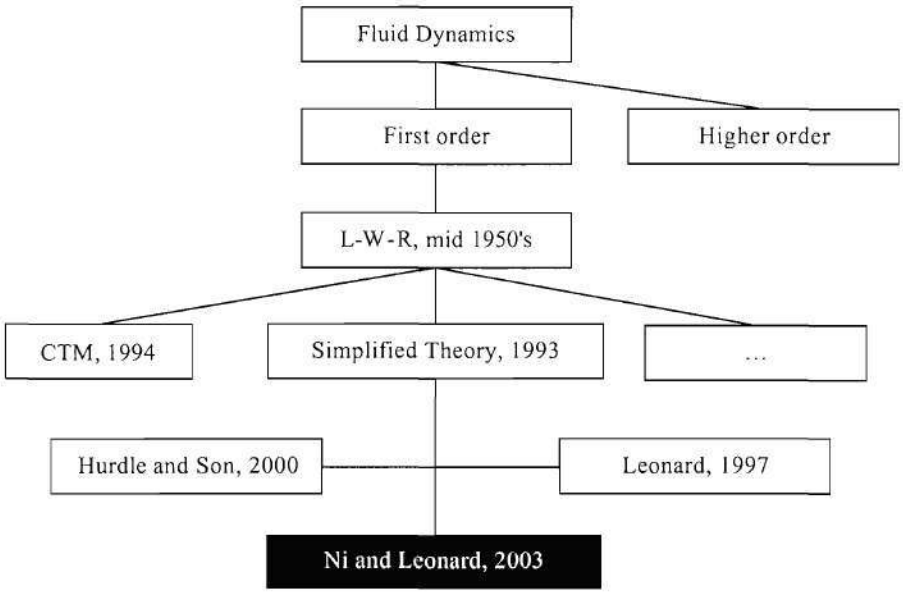


FIGURE 1-3 A Historical View

1.3. Research Motivations

A question to address is that, considering that there are many macroscopic models dealing with the same problem, what's the point of investing yet another model? The reasons are multiple. First, practitioners are more interested in models or tools that are simple and easy to understand, and these properties are exactly what simplified theory features. It is straightforward in that it works on cumulative number of vehicles past selected points along a freeway constrained by upstream demand, downstream queue, and local capacity. Second, simplified theory is computationally efficient, its data requirement is minimal but simulation result is relatively rich, and it works best under heavily congested condition which is the scenario that traffic engineers are most likely to concern. Third, simplified theory also shows the potential of large-scale application due to its simplicity and efficiency. Last but not least, simplified theory is promising for on-line application. This implies that traffic engineers can possibly "preview" how traffic operates, under current demand and condition, before it actually happens. This kind of visibility, even though limited, is very precious to traffic control and management because it not only allows taking

measures in advance to maintain stable flow but also provides better chance to deploy facilities to recover from system breakdown.

Newell's theory, however, deals only with freeway mainline because it assumes that ramp-entering traffic can always bypass a queue, if any, at the merge and experiences no delay and exiting vehicles can always be able to leave the freeway without constraint if they successfully arrive the diverge. These limitations greatly limit the practical appeal of the simplified theory and they have almost been left untouched a decade since the discovery of the simplified theory. This research is intended to explore further along Newell's track and attempt to remove these limitations by generalizing the simplified theory so that the simple and elegant algorithm applies to a freeway corridor, a freeway network, and even a general transportation network.

1.4. Organization of This Report

This report is organized in the following sequence: Section 2 has a literature review on Newell's simplified theory of kinematic waves, based on which the extension / generalization is going to be made. Section 3 presents the key features of the simplified theory, and the simplified theory is summarized in a easy-to-follow 5-step procedure. In section 4, we present a Capacity-based Weighted Fair Queuing (CBWFQ) merge model and a Contribution-based Weighted Splitting (CBWS) diverge model. These two models serve as the basis of the proposed extension / generalization. Section 5 formulates the proposed extension of Newell's simplified theory of kinematic waves. Procedures are proposed for entrance, exit, mainline, merge, and diverge scenarios. With these basic building blocks, it is possible to model traffic in a freeway network. In section 6, we go one step further by generalizing the simplified theory so that it is able to deal with a general transportation network. This is accomplished by proposing a general procedure for a general node which can have multiple upstream links and multiple downstream links simultaneously. This general node procedure is proposed based on a generalized CBWFQ merge model where a merge can have more than 2 merging branches and a generalized CBWS diverge model where a diverge can have more than 2 diverging branches. Section 7 introduces the test sites and collects data for the test sites. The data includes roadway geometry, traffic characteristics, and origin-destination flows. Section 8 presents the test results – the comparison between simulated results and the field observations. Evaluation of the model performance is based on both qualitative and quantitative measures. Section 9 summarizes the work in this research and identifies future directions for this research.

CHAPTER 2
LITERATURE REVIEW

Macroscopic traffic simulation is based on hydrodynamic theory and was originally formulated in the so-called L-W-R theory (Lighthill and Whitham, 1955; Richards, 1956). At the core of the seminal LWR model is a vehicle conservation law summarized by changes of flow and density in a continuous time-space domain. On the other hand, flow, speed, and density are three macroscopic measures of roadway traffic. Knowing them determines traffic states. This leads to two general approaches in approximating the model: a density-driven approach and a flow-driven approach. The former, such as FREFLO (Payne 1979) and KRONOS (Michalopoulos 1984, and Michalopoulos et al. 1991), works on density and determines speed and density based on some fundamental relations with the density. The latter, such as Newell's simplified theory of kinematic waves (Newell 1993a, 1993b, 1993c) and Daganzo's cell transmission model (CTM) (Daganzo 1994, 1995), works on flow and determines speed and density based on the time and space information associated with the flow. After years of development, the parallel models of the simplified theory have become full-fledged and are capable of dealing with network traffic. However, as what is commonly accepted, the simplified theory, in its original form, applies only to freeway mainline, which greatly limits its practical appeal. There has been a strong belief and curiosity that this simple and elegant algorithm can be extended to deal with network traffic, too. If this is true, it will be beneficial to communities of both theoretical and practical, and this study serves as such an endeavor.

If one takes another perspective, solutions to an L-W-R problem can be continuous or discrete. Continuous solution (Gerlough and Huber, 1975; Stephanopoulos and Michalopoulos, 1979; Stephanopoulos and Michalopoulos, 1981; Wong and Wong, 2002) keeps track of shock paths over the entire time-space diagram, while discrete solution (Payne, 1979; Michalopoulos et al, 1984; Michalopoulos, 1984; Michalopoulos et al, 1987; Newell, 1993a, 1993b, 1993c; Daganzo, 1995a, b; Daganzo, 1999; Lebacque, 1996;) works on traffic states at lattice points in the time-space diagram.

Newell's simplified theory of kinematic waves is actually a continuous time discrete space approach and deals with a special case of the LWR problem by assuming a triangular flow-density relation. The simplified theory combines kinematic wave theory with deterministic queuing theory and keeps track of the cumulative numbers of vehicles past a set of specific points on a freeway. Shock condition is then interpreted as the minimum of cumulative traffic counts when viewed from both sides of the traffic. For some reason, Newell confined his theoretical presentation to a freeway mainline scenario, i.e., no queuing on ramps. This greatly limits its practical appeal and this limitation has almost been left untouched a decade since the discovery of the simplified theory. This paper attempts to remove the limitation by generalizing the simplified theory so that the simple and elegant algorithm applies to a general transportation network.

A number of endeavors furthering Newell's work are identified. 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. Validation of Newell's theory was conducted by Hurdle and Son (Son, 1996; Hurdle and Son, 2000) who 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 bridges the theory and its application.

CHAPTER 3
KEY FEATURES OF NEWELL'S SIMPLIFIED THEORY OF KINEMATIC WAVES

This section summarizes the key assumptions and features of Newell's theory. To be consistent with the original papers, this section follows Newell's notation.

3.1. Assumptions

The following assumptions (or simplifications) are made (or implied) in Newell's simplified theory of kinematic waves:

- Assumption 1: To avoid some unpleasant mathematical complications, Newell postulated a triangular flow-density relationship. Therefore, there are only two constant wave speeds: a forward wave speed in under-saturated flow, which is the free flow speed and independent of flow, and a backward wave speed in congested flow, which is a simple function of free flow speed, capacity, and jam density. To be able to apply this simplification, the freeway mainline is subdivided until each freeway section can be considered homogeneous.
- Assumption 2: When dealing with an on-ramp, Newell assumed that ramp entering flow can always bypass a queue, if any, 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.
- Assumption 3: Travel time of all vehicles in a 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.
- 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 really simply and can be accomplished by some translations.

3.2.1. Forward Wave Propagation Rule

If there is no congestion, ie., traffic is dictated by forward waves, upstream demand will arrive at a downstream location after a free trip time. Therefore, the cumulative number of vehicles versus t curve at location x, $A(x,t)$, can be obtained by horizontally translate the cumulative curve $A(x_0,t)$ versus t at an upstream location x_0 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.

3.2.2. Backward Wave Propagation Rule

If there is congestion, i.e., traffic is dictated by backward waves, a downstream disturbance will be detected at an upstream location after a backward wave travel time and an adjustment of jam storage. Therefore, the cumulative curve $A(x,t)$ versus t at location x in section (x_0, x_1) can be obtained by translating the cumulative curve $A(x_1,t)$ at downstream location x_1 horizontally to the right by a displacement $(x_1-x)/u$ and vertically

upward by a displacement $k(x_1 - x)$, where $x_0 < x < x_1$, u is the backward wave speed, and k is the jam density of the section.

3.3. The 5-Step Modeling Procedure

The algorithm 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. The algorithm works on the cumulative arrival and departure curves at each lattice point iteratively and progressively such that, at each time increment τ, all locations, from the upstream end to the downstream end, are evaluated one by one, and then system time advances one step. The above process is repeated until the whole simulation period is traversed. The following procedure summarizes how the algorithm works at a lattice point (x_i,t) assuming that all previous lattice points have been processed.

A. Departure to the right

The cumulative departure curve past point x_i by time t, $D(x_i^+,t)$, is determined by the following 4 constraints:

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 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 found by tracing $D_{i+1}(x_{i-1}^+,t)$ back until some prior time t' such

that $D_{i+1}(x_{i-1}^{+},t')$ is equal to $D_{i+1}(x_i^{-},t)$. Then $T_i(t)=t-t'$ is the travel time for the vehicle bearing the "number" $D_{i+1}(x_{i-1}^{+},t)$ in this section.

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 $D_j(x_{i-1}^{+},t)$, $(j=i,i+1,i+2,...)$, are known, the 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., $D_j(x_i^{-},t)=D_j(x_{i-1}^{+},t-T_i(t))$.

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.

CHAPTER 4
BASICS OF THE PROPOSED GENERALIZATION

Since the major limitation of the simplified theory is its ignorance of constraints on ramp traffic. This is because the simplified theory makes two assumptions, one for on-ramp and the other for off-ramp, regarding traffic behavior on freeway ramps. Therefore, a good start point to relax the limitation is the formulation of some models to deal with ramp traffic. In this section, a capacity-based weighted fair queuing (CBWFQ) merge model and a contribution-based weighted splitting (CBWS) diverge model are proposed.

4.1 Capacity-based Weighted Fair Queuing (CBWFQ) Merge Model

4.1.1. Review of Existing Work On Merge

To facilitate future discussion, a merge is sketched below which has two upstream links/branches 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-leg intersection if properly defined. Suppose at time t, link 1 wants to send S_1 vehicles, link 2 wants to send S_2 vehicles, and the downstream link can receive R vehicles. Denote d_1 the out flow (i.e., departure count) of link 1, d_2 the out flow 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.

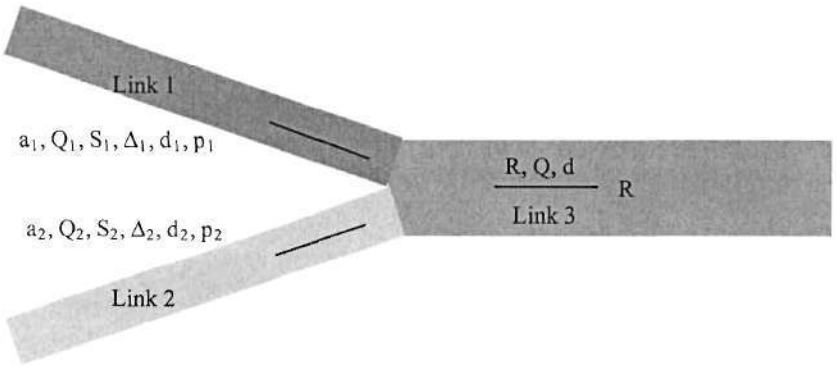


FIGURE 4-1 Sketch of a merge

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 the 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 splitting coefficients. If R is fixed and $S_1 + S_2 \le R$, the solution lies on bold line S_2-1-2-3-4-S_1. If otherwise $S_1 + S_2 > R$, the solution lies 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 out flows can be greater than downstream supply, and something has to be done to make it feasible. Nevertheless, this model is comprehensive and yields the largest solution space.

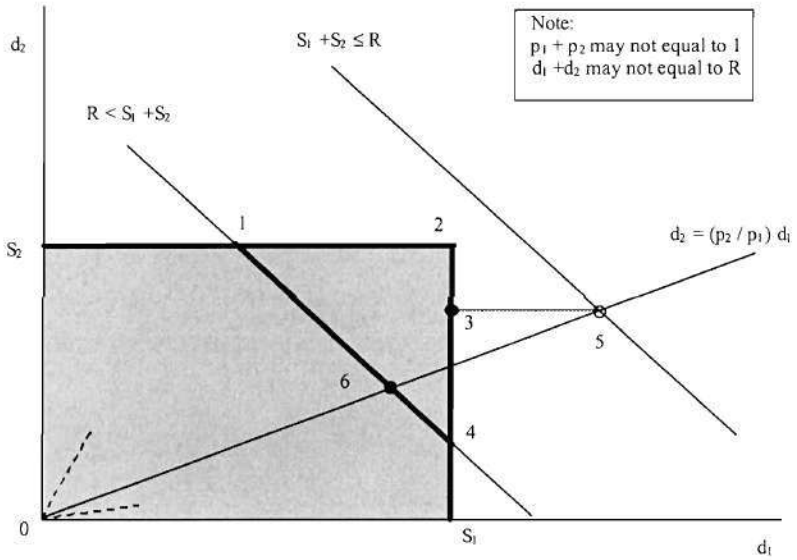


FIGURE 4-2 Lebacque's lane-based merge model

Daganzo's Priority-Based Merge Model

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

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

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

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.

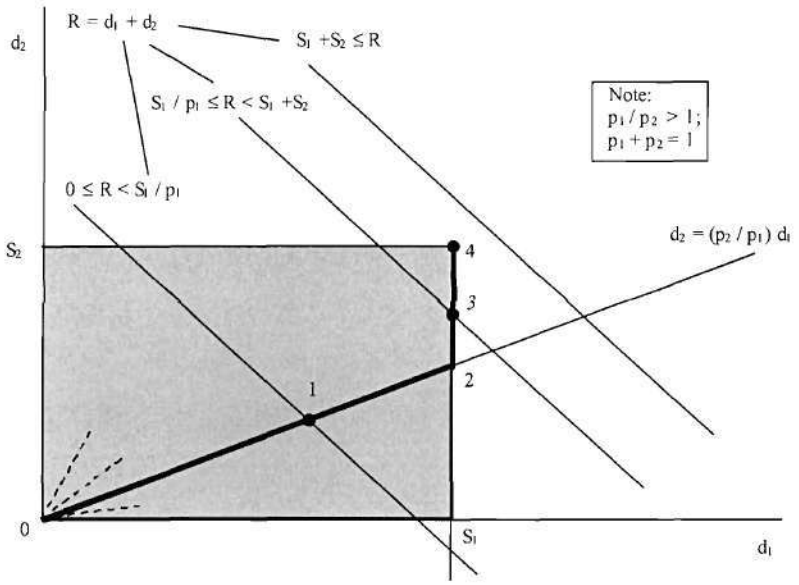


FIGURE 4-3 Daganzo's priority-based merge model

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.,

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

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

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

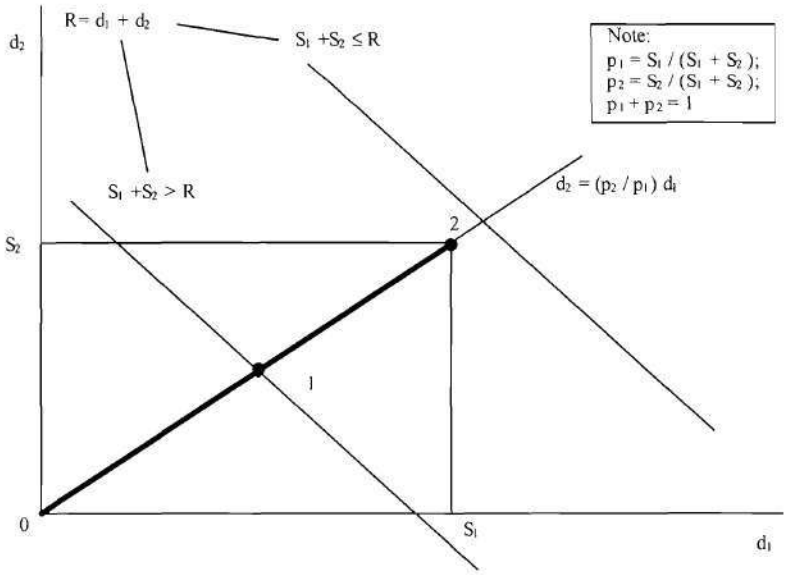


FIGURE 4-4 Jin and Zhang's demand-based merge model

Unfortunately, this model may yield unrealistic results under certain conditions. For example, suppose that 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≈1818 and d_2≈ 182. This implies that the on-ramp can depart only 1 vehicle at the departure of every 10 vehicles at the upstream mainline, a scenario that is very rare in real life.

Newell's Simplified Merge Model

As one of his underlying assumptions of the simplified theory of kinematic waves, 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 is reduced to the bold line indicated in Figure 4-5. The solution is at point 1 if link 1 is dictated by backward waves or at point 2 if link 1 is dictated by forward waves. In either case, link 2 is always dictated by forward waves.

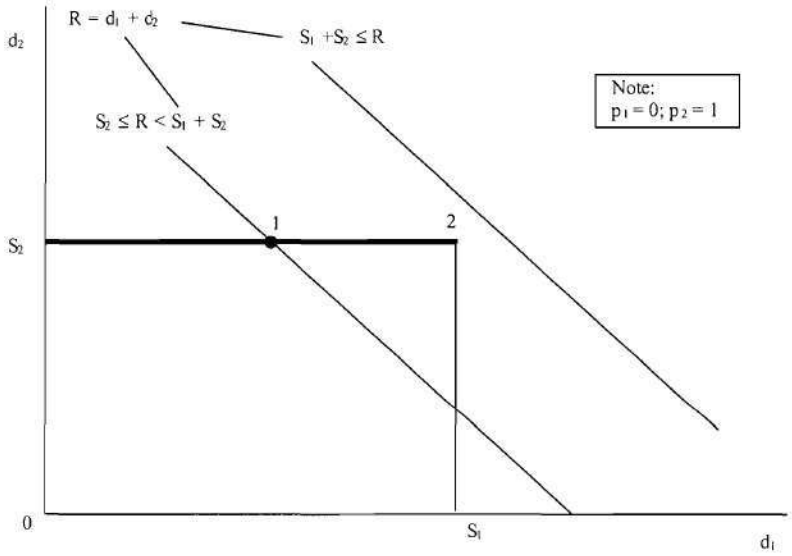


FIGURE 4-5 Newell's simplified merge model

Banks' Ramp-Metering Merge Model

As part of his effort to analyze the effect of ramp-metering in reducing traffic delay, Banks (2000) extends a little bit Newell's assumption on a merge such that ramp out flow 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 basically 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.

Generally, 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, a relationship that is illustrated in Figure 4-6. As a point of interest, the model we are going to proposed can also be viewed as a special case since its priority factor is based on capacity.

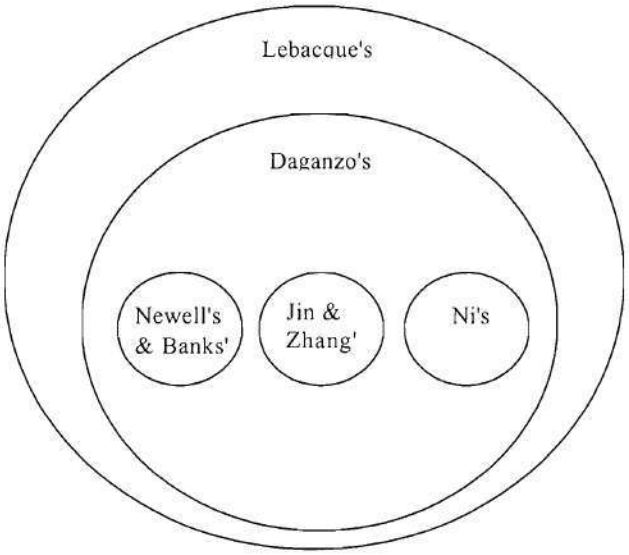


FIGURE 4-6 Relationship of merge models under analysis

As can be seen from above analysis, 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 / splitting coefficients 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 results under certain conditions. Therefore, there exists some room for further improvement regarding modeling queuing at a merge.

This paper is going to propose a merge mode that preserves the advantages of the above models while addressing their unattractive features. The proposed model is expected to achieve the following objectives: (1) It deals with both forward and backward waves, so it can be integrated into a larger kinematic wave problem; (2) It always yields unique solution, so it is well-formulated; (3) The solutions are physically meaningful and highly likely, so the model eliminates lots of unnecessary possibilities that may come at costs; (4) The model is simple and easy to understand and implement, so it makes a lot of sense to traffic engineers; (5) it takes into consideration as many factors as possible, such as demand, supply, road geometry, capacity, ramp -metering strategies, etc. at no extra cost, so it has lots of free stuff. (6) It has a good extensibility, so it has the potential to deal with on-ramps, merging highways, intersections, etc. Moreover, this research is also part of the effort to extend Newell's simplified theory of kinematic wave from mainline application to corridor and network application, which is of great interest both theoretically and practically.

4.1.2. The Proposed Model

As part of the effort to extend Newell's simplified theory of kinematic waves, a model is desired to deal with kinematic waves at a merge. Though it still works under the demand-supply framework, the model proposed in this section is developed independently and exhibits some attractive features that make it different.

To be consistent with the discussion early on and later on, notations are formally redefined as follows:

a_1 : Number of vehicles, waiting somewhere upstream the merge at the side of link 1, that want to use link 3. a_1 can also be viewed as the arrival waiting to be served by link 1. a_1 is referred to as (current) upstream arrival thereafter. Here, we are refrained from using capital A since it is reserved for cumulative arrival count.

Q_1 : Capacity of link 1. Capacity is affected by factors such as number of lanes, per lane capacity, and various traffic control strategies such as ramp-metering if applicable.

S_1 : Number of vehicles on link 1 waiting to be served by link 3. S_1 is referred to as upstream demand thereafter.

Δ_1 : Downstream supply shared by link 1 proportionally to its capacity.

d_1 : Number of vehicles that can actually move onto link 3 from link 1, i.e., the out flow or departure count thereafter. Similar to a and A, capital D is reserved for cumulative departure count.

All the above symbols may be functions of time which is omitted to keep the notation simple. The same set of symbols applies to link 2.

R : Number of vehicles that can be received by link 3, i.e., the downstream supply thereafter.

Q : Capacity of link 3.

d : Number of vehicles that can actually move onto link 3.

Since upstream demand is constrained by upstream arrival and capacity (including traffic control strategies such as ramp metering if any), we have:

$S_1 = \min\{ a_1, Q_1 \}$ for link 1.

$S_2 = \min\{ a_2, Q_2 \}$ for link 2.

To formulate the merge model, let's assume that traffic is greedy and wants to pass as many vehicles as possible. We also assume a capacity-based weighted fair queuing (CBWFQ) at a merge, which means that, for those merging branches dictated by backward waves, their demands are satisfied, or downstream supply is distributed, proportionally to their respective capacities.

Suppose the downstream supply is R, let's define:

$\Delta_1 = R \times \frac{Q_1}{Q_1 + Q_2}$ as the fair share of downstream supply for link 1, and

$\Delta_2 = R \times \frac{Q_2}{Q_1 + Q_2}$ as the fair share for link 2.

Suppose the upstream demands are S_1 and S_2. There are 4 possible cases:

Case 1: For each upstream link, its demand is less than or equal to its fair share, i.e.,

$$S_1 \le \Delta_1 \text{ and } S_2 \le \Delta_2$$

In this case, each upstream link is dictated by forward wave and every one gets chance to depart without delay, i.e.,

$$d_1 = S_1 \text{ and } d_2 = S_2$$

Case 2: Demand of link 1 is less than or equal to its fair share, but demand of link 2 is greater than its fair share, i.e.,

$$S_1 \le \Delta_1 \text{ and } S_2 > \Delta_2$$

In this case, traffic on link 1 deserves its chance and can depart without delay, while traffic on link 2 can depart depending on the remainder of the downstream supply, i.e.,

$$d_1 = S_1 \text{ and } d_2 = \min\{ S_2, R - d_1 \}$$

Case 3: the situation of case 2 reverses, i.e.,

$$S_1 > \Delta_1 \text{ and } S_2 \le \Delta_2$$

In this case, traffic on link 2 can depart without delay, while traffic on link 1 make use of the remainder of downstream supply, i.e.,

$$d_1 = \min\{ S_1, R - d_2 \} \text{ and } d_2 = S_2$$

Case 4: For each upstream link, its demand exceeds its fair share, i.e.,

$$S_1 > \Delta_1 \text{ and } S_2 > \Delta_2$$

Actually, this is the case where our assumption of CBWFQ applies, i.e., traffic on each upstream link departs proportionally to its capacity:

$$d_1 = \Delta_1 \text{ and } d_2 = \Delta_2$$

The solution space is illustrated Figure 4-7 by the shaded areas. The working of the diagram is very simple and straightforward. A downsloping line at an angle of 135^0 passing point B_4 denotes supply line where the coordinates of every point on it sum up to downstream supply. Δ_1 and Δ_2 are link 1 and link2's fair shares of downstream supply, respectively. Lines $\Delta_2 B_4$ and $\Delta_1 B_4$ divide the first quadrant into four areas numbered in circles. Let C denotes a demand point whose coordinates are upstream demands (S_1, S_2) and B denotes a solution point whose coordinates are upstream out flows (d_1, d_2). Area 1 corresponds to case 1 where demand point C_1 coincides with solution point B_1. Area 2 corresponds to case 2 and is further divided into two subsections by the supply line. In the lower left (shaded) subsection, demand point C_2 coincides with solution point B_2, while, in the upper right subsection, solution point B_2' is found at the intersection of the supply line and a vertical line through demand point C_2'. Similar situation occurs in area 3 which corresponds to case 3, and the solution can be found as illustrated. Area 4 corresponds to case 4 where the solution is at B_4 regardless of where the demand point C_4 is.

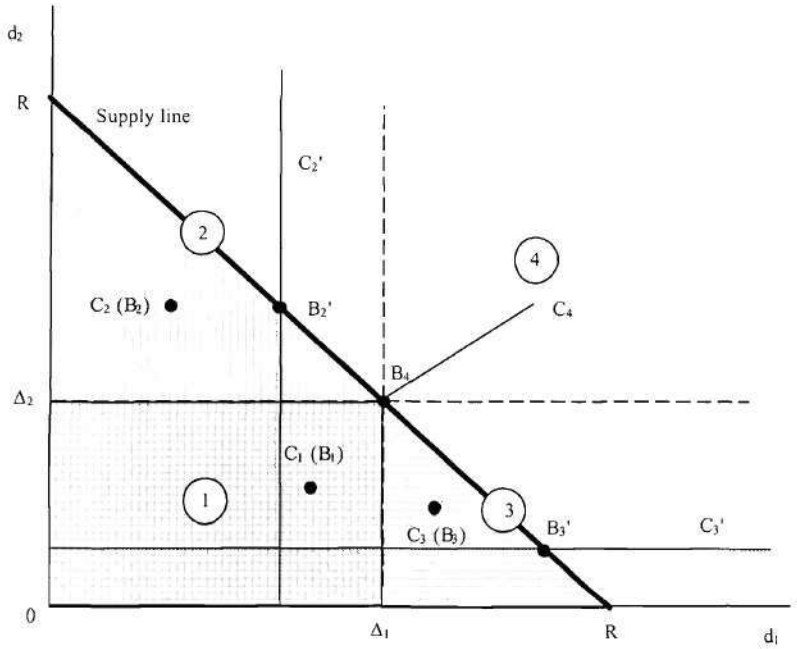


FIGURE 4-7 The proposed CBWFQ merge model

There are several attractive features with the proposed model. First, it separates the concepts of arrival, demand, capacity, and departure/out flow, a feature that makes this model suitable for plugging in the simplified theory. Second, it assumes priority only under certain circumstance, which avoids unrealistic distribution found in some models. Third, the priority is based on capacity, which reflects factors such as number of lanes, per lane capacity, and various traffic control strategies, and, therefore, is highly realistic and physically meaningful. Forth, the proposed model is easy to understand and cheap to implement, a feature that traffic engineers love to see. Last but not least, the model is extensible at virtually no additional cost as the number of merging branches gets big.

4.1.3. The Generalized Model

To facilitate future generalization of the simplified theory, a generalized CBWFQ model is formulated. Figure 4-8A sketches a general merge with α merging branches and one downstream link. Let:

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

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

$a_1, a_2, \ldots, a_\alpha$ denote demands, i.e., arrival counts, of merging branches $1,2,\ldots,\alpha$, respectively.

$Q, Q_1, Q_2, \ldots, Q_\alpha$ denote capacities of the downstream link and merging branches $1,2,\ldots,\alpha$, respectively.

$\Delta_1, \Delta_2, \ldots, \Delta_\alpha$ denote the fair shares of downstream supply by merging branches $1,2,\ldots,\alpha$, respectively.

$d_1, d_2, \ldots, d_\alpha$ denote out flows, i.e., departure counts, of merging branches $1,2,\ldots,\alpha$, respectively.

The generalized a capacity-based weighted fair queuing (CBWFQ) model determines out flows based on the following algorithm.

1. Compute the fair share of the downstream supply for each of the merging branches based on its capacity.

$$\Delta_i = R \times \frac{Q_i}{\sum_{j=1}^{\alpha} Q_j}, \quad i \in B$$

2. For each merging branch, if its demand is less than or equal to its fair share, set its out flow to its demand, 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.:

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

3. Based on 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.
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 its out flow to its fair share.

Figure 4-8B and 4-8C illustrate solutions to two special cases of the generalized CBWFQ model. In the 3-dinensional (3D) case as shown in Figure 4-8C, there are three merging branches whose fair shares determine the fair share cube $O\Delta_2P_{12}\Delta_1\Delta_3P_{23}PP_{13}$. The supply plane $P_1P_2P_3$ consists of points whose coordinates sum up to downstream supply. the plane and the cube intersect at point P. if the demand of each branch is greater than its fair share, the demand point lies in the half-space $PP'_{12}P'_{13}P'_{23}$, and the solution is point P. Otherwise, there must be at least one branch, say branch 3, whose demand is less than or equal to its fair share. By satisfying this demand, subtracting this amount from the supply, and removing this branch from the set, the problem reduces to a 2-dimensional (2D) case which can be viewed as the result of cutting the supply plane with plane $d_3 = a_3 \le \Delta_3$, as shown by the dotted triangle. In the 2D case as shown in Figure 4-8B, the fair share cube reduces to the fair share rectangle $O\Delta_2P\Delta_1$ and the supply plane reduces to the supply line P_1P_2. If the demand of each branch is greater than its fair share, the demand point must lie in the half-space $P\Delta'_1\Delta'_2$, and the solution is P. Otherwise, there must be at least one branch, say branch 2, whose demand is less than or equal to its fair share. By satisfying this demand, subtracting this amount from the supply, and removing this branch from the set, the problem reduces to a 1-dimensional (1D) case which can be viewed as the result of intersecting the supply line with line $d_2 = a_2 \le \Delta_2$. In this 1D case, if the demand is greater than its fair share, the solution is the fair share, otherwise, the solution is the demand, as shown by the dotted line 36'6 in Figure 4-8B.

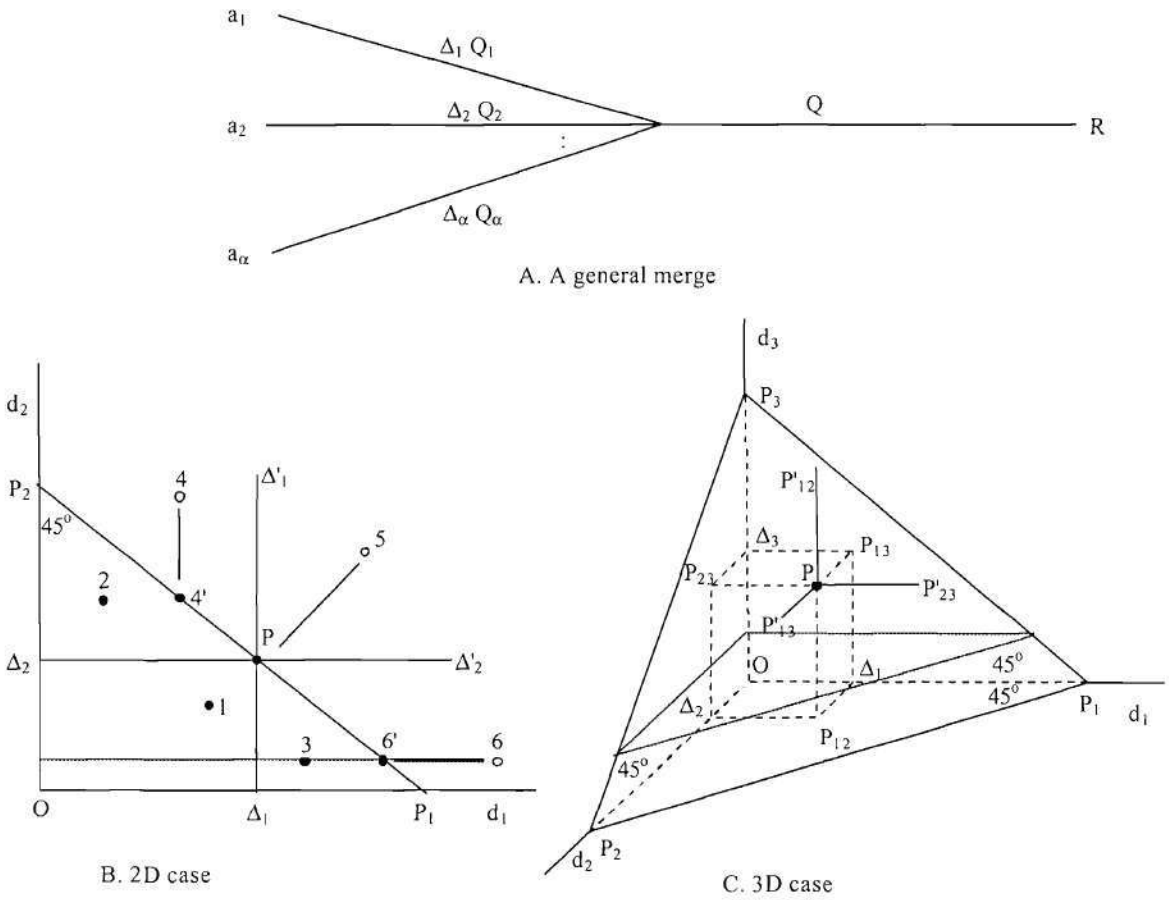


FIGURE 4-8 Generalized CBWFQ merge model

4.2 Contribution-based Weighted Splitting (CBWS) Diverge Model

4.2.1. Review of Existing Work on Diverges

A diverge comes into play because it links two or more downstream routes that are possibly interact at their common upstream 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 of kinematic waves 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 a problem of dealing with 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 (1995) 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 his another paper regarding generalized theory of kinematic waves (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 basically 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 doesn't. 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 destination. This is relatively cheap to model but at some cost of realism. In this paper, we are going to extend the simplified theory of kinematic waves to model queuing at a diverge, which 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

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.

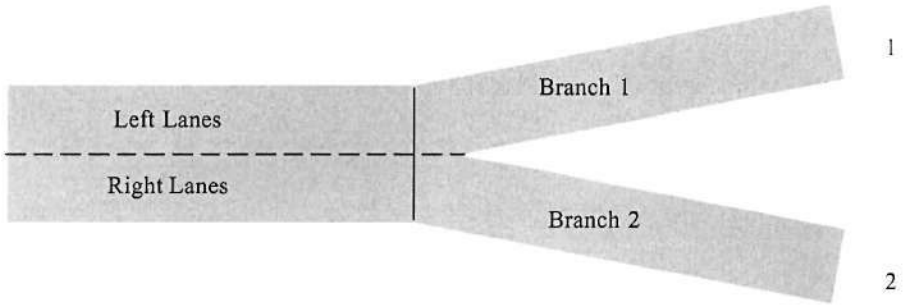


FIGURE 4-9 Sketch of a diverge

Before any queue backs up from downstream of the diverge, let's assume that traffic upstream is operating in the free-flow regime and traffic state is pretty 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 can originate 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 to branch 2. Several pictures are proposed to model queuing at the diverge.

One picture is illustrated in Figure 4-10 where the 2-vehicle queue dictates the overall traffic condition upstream. This is the diverge model assumed in Newell (1993a, 1993b, 1993c) and in Daganzo (1995) 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 cost of some realism. This model also makes more sense if the upstream link has only one lane or the majority of the road users are tourists often characterized as alert, courteous, and curious. However, this model might not be realistic when the upstream link has multiple lanes, especially when the road users are predominated daily commuters, a case that is often seen in urban freeways.

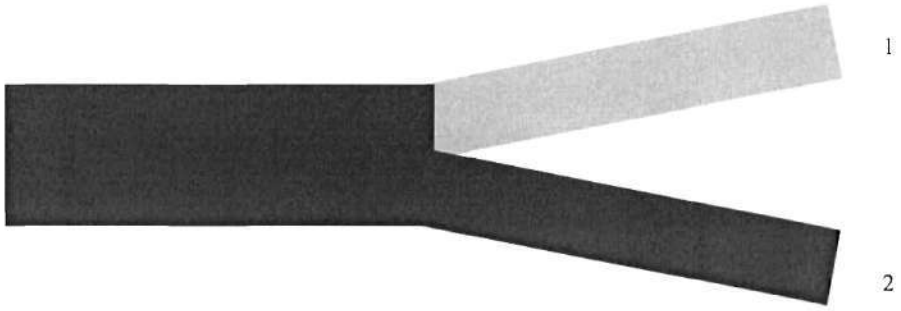


FIGURE 4-10 Diverge model 1 – Newell's 1st model, Daganzo's model, and Lebacque's 1st model

Another picture is proposed by Newell (1996) 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 of kinematic waves to corridor/network application. Despite of this, Newell did imply that queues from different branches need to be treated differently, which is the idea that this paper is in favor of.

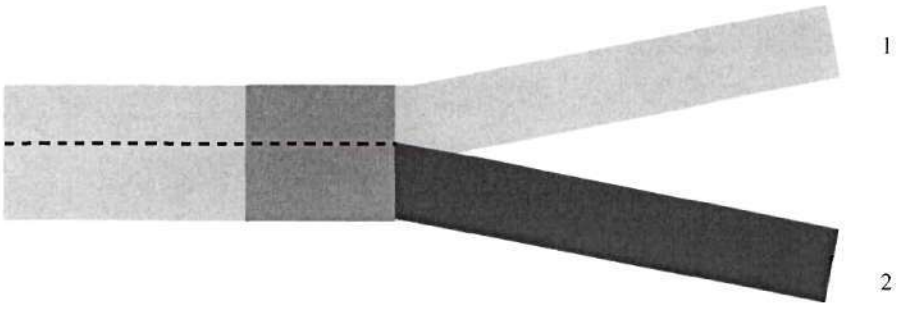


FIGURE 4-11 Diverge model 2 - Newell's 2nd model

Figure 4-12 probably gives a more realistic representation of queuing at a diverge. As 2-vehicle queue backs onto the upstream 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 on the right lanes near the diverge, traffic state is almost the same as that of branch 2 because the former is a nature 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 most macroscopic model does. 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 because they are not intended to exit via branch 2 and because traffic is traveling 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

who are close to the diverge may not be able to change to the left lanes even though they want to divert. However, the difference in traffic states between the left lanes and the right lanes diminishes as one goes further upstream due to a transition from congested to uncongested condition on the right lanes. Traffic density in the transition area is lower then that of the congested area, but higher than the uncongested area. Still in this 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. Though there are some 2-vehicles who bypass the queue and squeeze in at the head of the queue at the expense of others' delay, these vehicles are relatively few and the relatively high speed difference near the diverge may not be favorable for them to sneak in. This and other phenomena such 1-vehicle's squeezing out from right lanes can be modeled by a friction at the interface between the left and right lanes. This friction, in turn, acts as a variable constraint on effective capacity based on current traffic state (e.g., density). 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.

Technically, it is difficult to model transition of traffic state at a macroscopic level, and this is where a shock wave is introduced to describe the discontinuity of traffic states. Figure 4-12 can be approximated by the Figure 4-13, in which there is an abrupt change of traffic state between branch 2 and the right lanes. This happens to be the second diverge model proposed by Lebacque (1996).

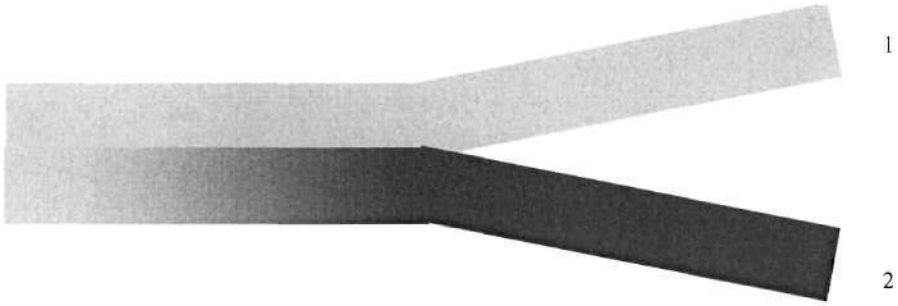


FIGURE 4-12 Diverge model 3 – the realistic model

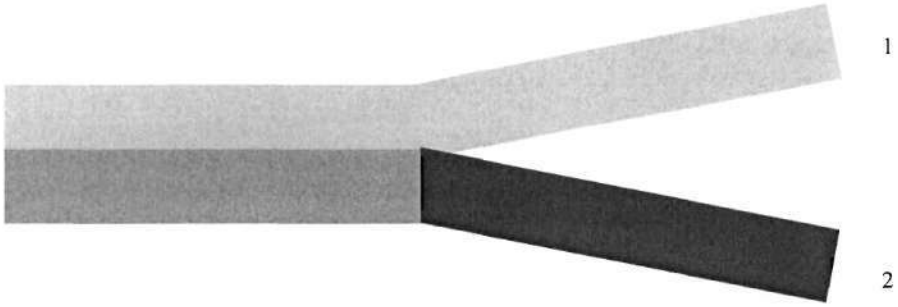


FIGURE 4-13 Diverge model 4 - Lebacque's 2nd model

Considering that queue tail might be located somewhere on the right lanes, the right lanes might contain a congested section and a uncongested section. On the other hand, there might be vehicles exchanging between the left and the right lanes and this acts as a friction at the interface. Given these, a closer approximation of Figure 4-13 can be the one illustrated in Figure 4-14 where there is a discrete changeover from congested to uncongested condition and a traffic state-based variable capacity near the diverge. If there is a concurrent queue from branch 1, the same treatment applies and there is another discrete changeover on the left lanes, and this picture happens to be the second diverge model proposed by Daganzo, i.e., the one with two vehicle types and three traffic states.

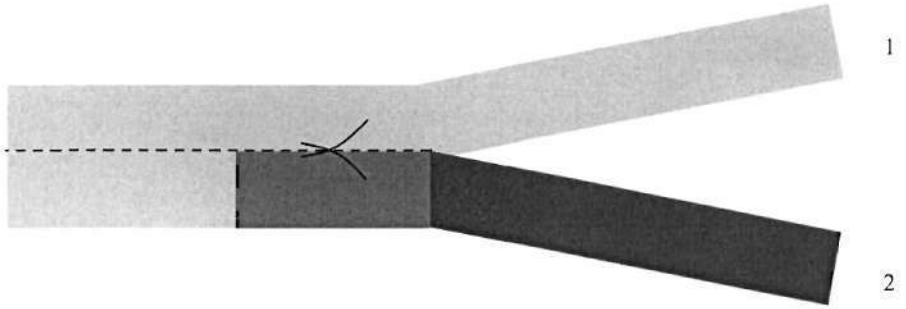


FIGURE 4-14 Diverge model 5 – the contribution-based weighted splitting (CBWS) diverge model

Figure 4-14 depicts the contribution-based weighted splitting (CBWS) diverge model based on which the simplified theory of kinematic waves is going to be extended. As stated before, we are trying to mix the two modeling strategies and taking the best out of them. More specifically, we are going to split flow and move traffic forward based on the strategy of multiple traffic states, while we update link traffic state based on the strategy of single traffic state. The simplified kinematic waves at a diverge works as follows. First, the departure counts of the two diverging branches are evaluated individually. Second, their sum is used as one of the constraints to evaluate the aggregate departure count to the left of the diverge. Also, this is the place where the friction comes into play. A traffic state-based friction factor is applied to the capacity which is also one of the constraints to the aggregate departure count. The resulting aggregate departure count may not exactly be the sum of the departure counts of the two branches determined earlier, so a splitting scheme based on downstream contributions is used to determine the actually downstream departure counts. Queues from different branches are treated separately such that travel times of vehicles in these queues are evaluated individually and the travel times are then used to advance multiple-destination flows.

Of course, one would arguably say that treating different queues 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 lost 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 4-15 sketches a general diverge with β diverging branches and one upstream link. Let:

S denotes the upstream demand, i.e., number of vehicles that want to be sent to the diverging branches.

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

$R_1, R_2, ..., R_\beta$ denote the supplies/contributions of diverging branches $1,2,..., \beta$, respectively.

$d_1, d_2, ... d_\beta$ denote the out flows to diverging branches $1,2,..., \beta$, respectively.

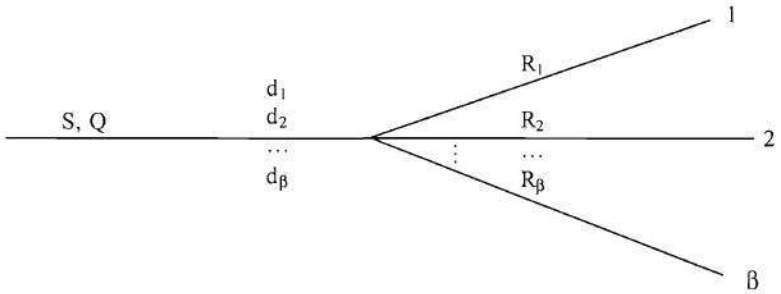


FIGURE 4-15 Generalized CBWS diverge model

The generalized CBWS diverge model determines out flows based on the following algorithm.

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

CHAPTER 5
EXTENSION OF THE SIMPLIFIED THEORY

Now that we are migrating from freeway mainline application to a freeway system, the original one-dimensional (1D) location space has to be extended to a two-dimensional (2D) space. In this section, we assume that a freeway system can be mapped to a spanning tree where a node can have up to 3 legs. Sometimes, a freeway system may involve loops such as I-285 in Atlanta regional freeway system. Loops complicate the problem in two ways: first, it makes O-D estimation difficult because origins and destinations lose their meaning in a loop; second, the nature of the simplified theory requires that processing a node depends on information of all its adjacent and further upstream nodes (to clarify, "upstream" means "adjacent upstream" unless explicitly noted. The same with "downstream"), and a loop turns such a dependent relation into an inter-dependent relation which results in a deadlock. Therefore, something has to be done to break the loop, which is not the main focus of this paper. For now, we just assume a spanning tree configuration.

In the original theory, nodes can be stored in a consecutively indexed a 1D array. An index unambiguously specifies the location of a node on the freeway. However, in the extended theory, nodes have to be stored in a set because a node can have multiple upstream and/or downstream nodes. Nodes in the set have a loose relative relation such that all possible origins of a node bear lower indices and all potential destinations of a node bear higher indices. By this way, the aforementioned dependent relation is maintained. On the other hand, a node remembers its upstream node(s) and potential destination(s).

The focus of this section is to formulate procedures to model some basic building blocks of a freeway system. As most traffic simulation model do, the time-varying O-D flows are assumed known and summarized in some E-E (entrance-exit) tables, i.e., they specify trips from every entrance to every exit of the freeway system. After a simple synthesis, E-D (entrance-destination) tables can be obtained, which contain trips from every entrance to every possible destination. This is the boundary condition of the extended simplified kinematic wave problem as well as the point where the simulation starts. Suppose, after some iterations, we've moved to the current lattice point (x_l, t) and information about all previous lattice points are known. Our task here is to determine the cumulative arrival and departure curves at this lattice point and this can be done by applying the 5-step procedure.

5.1. Simplified Kinematic Waves at an Entrance

In this scenario, we consider a node that is linked with only one downstream node and no upstream node. This scenario 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.

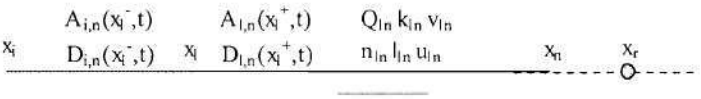


FIGURE 5-1 Entrance scenario

An entrance scenario is sketched in Figure 5-1, 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 5-step procedure for the entrance scenario 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 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 doesn't exist.

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

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

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 given off 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

This name is tricky because there is no link upstream of node x_l, but it is so called to be consistent with other scenarios. Actually the purpose of having some "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 destination. For the same reason, if a vehicle has to waiting some time before entering the system, vehicles for other destinations will experience the same delay. The waiting time is obtained by comparing curve pair $A_{i,n}(x_l^-,t)$ vs. $D_{i,n}(x_l^-,t)$ 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 5-2.

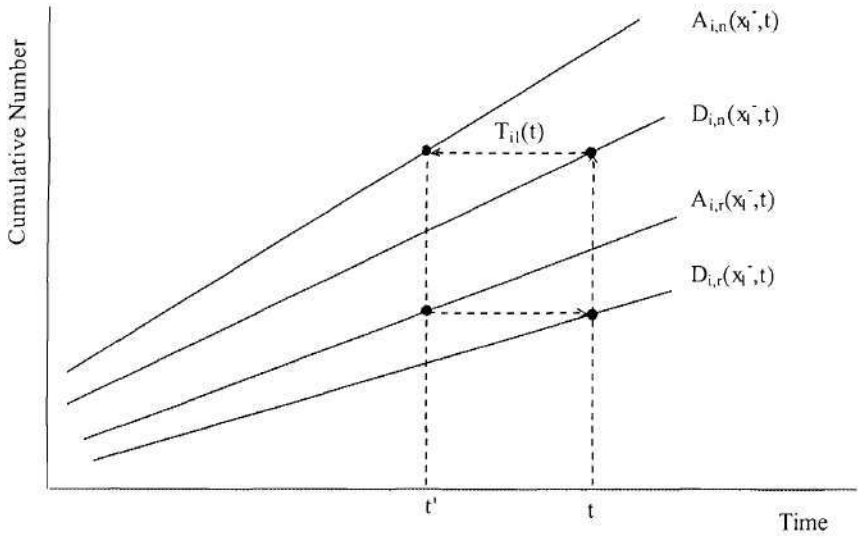


FIGURE 5-2 Illustration of link travel time

D. Departure to the left – multi-destinations

As assumed, the waiting time, $T_{ij}(t)$,determined above is experienced by all vehicle regardless of 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 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 5-step procedure for the entrance scenario, and the algorithm is ready to move on to the next lattice point.

5.2. Simplified Kinematic Waves at an Exit

In this scenario, we consider a node that is linked with only one upstream node and no downstream node. This scenario 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.

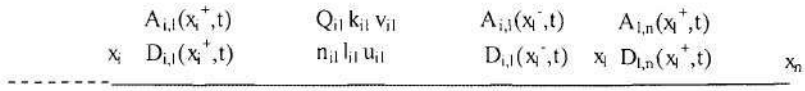


FIGURE 5-3 Exit scenario

An exit scenario is sketched in Figure 5-3, 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 5-step procedure for the exit scenario 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 doesn't exist.

c. Downstream queue

A queue never builds up here because of the infinite capacity of the dummy link, so this constraint doesn't 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 $D_{i,l}(x_i^+, t)$ vs. $D_{i,l}(x_l^-, t)$ 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 in reporting various measures of effectiveness (MOEs).

D. Departure to the left – multi-destinations

None.

E. Departure to the right – multi-destinations

None.

This concludes the 5-step procedure for the exit scenario, and the algorithm is ready to move on to the next lattice point.

5.3. Simplified Kinematic Waves at Freeway Mainline

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

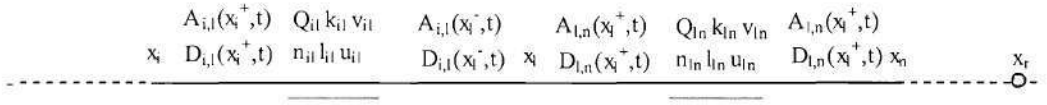


FIGURE 5-4 Mainline scenario

A mainline scenario is sketched in Figure 5-4, 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 5-step procedure for the mainline scenario 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 $D_{i,l}(x_i^+,t)$ vs. $D_{i,l}(x_l^-,t)$ 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 5-step procedure for the mainline scenario, and the algorithm is ready to move on to the next lattice point.

5.4. Simplified Kinematic Waves at a Merge

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

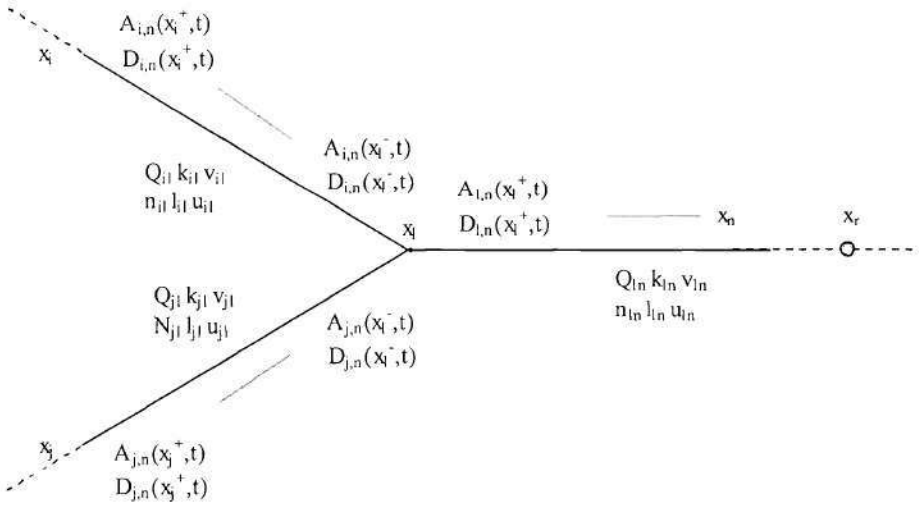


FIGURE 5-5 Merge scenario

A merge scenario is sketched in Figure 5-5, 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 5-step procedure for the merge scenario 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, $A_{j,n}(x_l^-,t)$ can be 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

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

c. Downstream queue

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

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.,

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

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 a capacity-based weighted fair queuing (CBWFQ) merge model. Let:

$a_{i,n}(x_l^-,t)$ be the current arrival to the left of x_l originated from x_i destined for x_n and beyond,

$a_{j,n}(x_l^-,t)$ be the current arrival to the left of x_l originated from x_j destined for x_n and beyond,

$d_{i,n}(x_l^-,t)$ be the current departure to the left of x_l originated from x_i destined for x_n and beyond,

$d_{j,n}(x_l^-,t)$ be the current departure to the left of x_l originated from x_j destined for x_n and beyond,

S_{il}^n be the demand of link (x_i,x_l) waiting to be served by link (x_l,x_n).

S_{jl}^n be the demand of link (x_j,x_l) waiting to be served by link (x_l,x_n).

R_{ln} be the supply of link (x_l,x_n).

Δ_{ln}^i be link (x_i,x_l)'s fair share of downstream supply, and

Δ_{ln}^j be link (x_j,x_l)'s fair share downstream supply.

Obviously, we have:

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

$$\Delta^i_{\text{In}} = R_{\text{In}} \times \frac{Q_{il}}{Q_{il} + Q_{jl}}$$

$$\Delta^j_{\text{In}} = R_{\text{In}} \times \frac{Q_{jl}}{Q_{il} + Q_{jl}}$$

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

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

$$S^n_{il} = \min\left\{a_{i,n}(x_l^-, t), \tau Q_{il}\right\}$$

$$S^n_{jl} = \min\left\{a_{j,n}(x_l^-, t), \tau Q_{jl}\right\}$$

The CBWFQ model determines upstream departure based on the following rules:

Case 1: $S^n_{il} \le \Delta^i_{\text{In}}$ and $S^n_{jl} \le \Delta^j_{\text{In}}$. The solution is:

$$d_{i,n}(x_l^-, t) = S^n_{il} \text{ and } d_{j,n}(x_l^-, t) = S^n_{jl}$$

Case 2: $S^n_{il} \le \Delta^i_{\text{In}}$ and $S^n_{jl} > \Delta^j_{\text{In}}$. The solution is:

$$d_{i,n}(x_l^-, t) = S^n_{il} \text{ and } d_{j,n}(x_l^-, t) = \min\left\{S^n_{jl}, R_{\text{In}} - d_{i,n}(x_l^-, t)\right\}$$

Case 3: $S^n_{il} > \Delta^i_{\text{In}}$ and $S^n_{jl} \le \Delta^j_{\text{In}}$. The solution is:

$$d_{i,n}(x_l^-, t) = \min\left\{S^n_{il}, R_{\text{In}} - d_{j,n}(x_l^-, t)\right\} \text{and } d_{j,n}(x_l^-, t) = S^n_{jl}$$

Case 4: $S^n_{il} > \Delta^i_{\text{In}}$ and $S^n_{jl} > \Delta^j_{\text{In}}$. The solution is:

$$d_{i,n}(x_l^-, t) = \Delta^i_{\text{In}} \text{ and } d_{j,n}(x_l^-, t) = \Delta^j_{\text{In}}$$

Based on the CBWFQ model, 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 determined as follows:

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

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

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_i 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 5-step procedure for merge scenario, and the algorithm is ready to move on to the next lattice point.

5.5. Simplified Kinematic Waves at a Diverge

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

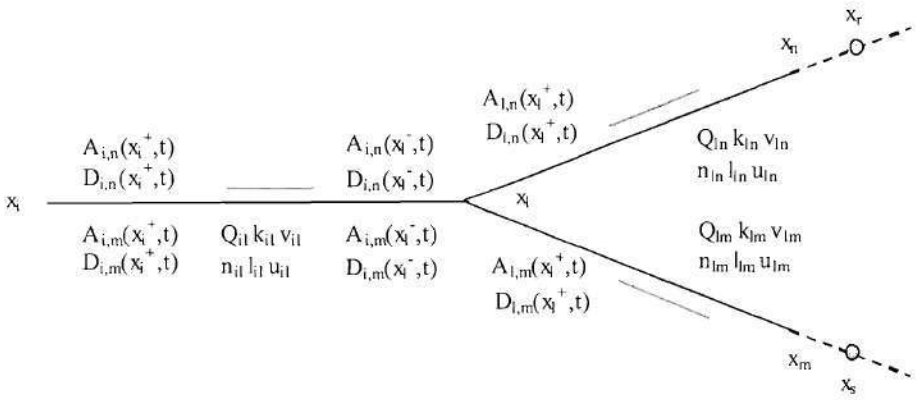
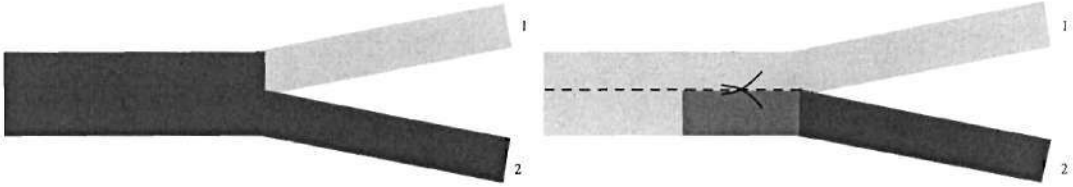


FIGURE 5-6 Diverge scenario

A diverge scenario is sketched in Figure 5-6, 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.



Newell's diverge model The contribution-based weighted splitting (CBWS) model

FIGURE 5-7 Diverge models

Before we go any further, let's spend a little time talking about the contribution-based weighted splitting (CBWS) diverge model based on which the procedure of diverge scenario is proposed. Figure 5-7 shows two models of a diverge with two downstream branches (and hence two vehicle types) labeled as 1 and 2. The left model is assumed by Newell in his original paper which states that, when the 2-vehicle queue backs up, all vehicles upstream will be affected and delay is experienced by everyone regardless of destination. 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 or the majority of the road users are tourists often characterized as alert, courteous, and curious. However, this model might not be realistic when the upstream link has multiple lanes, especially when the road users are predominated daily commuters, a case that is often seen in urban freeways.
The right model is the one we are in favor of and based on which the procedure for the diverge scenario is going to be formulated. It basically says that, when congestion of a diverging branch, say branch 2, backs onto the upstream link, only the corresponding part, say the right lanes, of the upstream traffic will be affected. Therefore, traffic states of the left lanes and the right lanes may not be the same, and there might be some friction between the two due to vehicle exchange. On the other hand, traffic state may not be uniform for the congested lanes, say the right lanes as shown, due to the existence of queue tail, though this feature is not explicitly modeled.
The simplified kinematic waves at a diverge works as follows. First, the departure counts of the two diverging branches are evaluated individually. Second, their sum is used as one of the constraints to evaluate the aggregate departure count to the left of the diverge. Also, this is the place where the friction comes into play. A traffic state-based friction factor is applied to the capacity which is also one of the constraints to the aggregate departure count. The resulting aggregate departure count may not be exactly the sum of the departure counts of the two branches determined earlier, so a splitting scheme based on downstream contributions is used to determine the actually downstream departure counts. Queues from different branches are treated separately such that travel times of vehicles in these queues are evaluated individually and the travel times are then used to advance multiple-destination flows. A 5-step procedure for the diverge scenario 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's 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

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

c. Downstream queue

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

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). As mentioned before, the use of the lanes near the diverge may not be balanced, i.e., 2-vehicles may stay at the right lanes and 1-vehicles may use all lanes though left lanes are usually preferred. Also, there might be vehicle exchange between left and right lanes and this is modeled as a friction which is a function of traffic state (e.g. traffic density). The effect of friction on traffic operation can be reflected by reducing capacity accordingly. Let $f_{il}(\phi,t)$ denotes the friction factor of link (x_i,x_l) at time t and ϕ is the current traffic state. $f_{il}(\phi,t)$ is usually not known and has to be treated as a design parameter. The effective capacity is then $Q_{il}'=Q_{il}(1-f_{il}(\phi,t))$, and the capacity constraint becomes $D_{i,l}(x_l^-,t-\tau)+\tau\times Q_{il}'$. Therefore,

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

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) won't exceed the capacity to the left of x_l.

Now, a new problem 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., $D_{i,m}(x_l^-,t)$? They might be the same as $D_{l,n}(x_l^+,t)$ and $D_{l,m}(x_l^+,t)$, respectively, if $D_{i,l}(x_l^-,t)$ is determined by downstream departures. However, when $D_{i,l}(x_l^-,t)$ is determined by upstream arrival or left capacity, $D_{i,n}(x_l^-,t)$ and $D_{i,m}(x_l^-,t)$ may not be the same as $D_{l,n}(x_l^+,t)$ and $D_{l,m}(x_l^+,t)$, respectively. In either case, $D_{i,l}(x_l^-,t)$ is split between $D_{i,n}(x_l^-,t)$ and $D_{i,m}(x_l^-,t)$ based on their respective downstream contributions. Let:

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

Then

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

If $d_{l,n}(x_l^+,t) + d_{l,m}(x_l^+,t) = 0$, no traffic discharges for either downstream link, i.e., $d_{i,n}(x_l^-,t) = 0$ and $d_{i,m}(x_l^-,t) = 0$.

C. Link travel time

At pervious 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 $D_{i,n}(x_i^+,t)$ vs. $D_{i,n}(x_l^-,t)$ 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 $D_{i,m}(x_i^+,t)$ vs. $D_{i,m}(x_l^-,t)$.

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 scenario, 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 5-step procedure for the diverge scenario, and the algorithm is ready to move on to the next lattice point.

CHAPTER 6
GENERALIZED SIMPLIFIED KINEMATIC WAVES THEORY

Section 5 shows that, by incorporating some models that deal with ramp traffic, the simplified theory can be extended to deal with a freeway corridor or network. The ramp models used in the extension deal with a 3-legged junction where an upstream link connects with two downstream links or a downstream link connects with two upstream links. These scenarios are often seen in a freeway system where a junction involves 4 or more legs are very rare or impossible. However, in a general transportation network, these unlikely scenarios become highly likely. For example, a normal 4-legged intersection can be viewed as an upstream link connecting 3 downstream links or a downstream link connecting with 3 upstream links, and things can be more complicated as the intersection has more legs. Therefore, generalized merge and diverge models wh ich are capable of modeling such complicated cases are critical when generalizing the simplified theory to deal with a general transportation network. Fortunately, the generalized CBWFQ and CBWS have been formulated in section 4. This section is an application of these generalized models based on the 5-step procedure.

6.1 Setting up the Problem

There are several issues to migrate from freeway mainline application to a general transportation network, among which representation of the more complicated location space demands special attention. This section serves as a preparation for subsequent discussion.

6.1.1. K-Wave Space

A K-Wave space, here referred to as the solution space of a kinematic wave problem, consists of a time space and a location space. The original simplified theory essentially deals with a one-dimensional (1D) location space where a freeway is typically represented by an array of nodes that are consecutively indexed. In Particular, consecutively indexed nodes are physically adjacent and a node can have up to two adjacent nodes. On the other hand, the index also guides how the algorithm proceeds as it traverses the lattice in time-location space.
However, as we go from a freeway mainline to a general transportation network, we are migrating to a two-dimensional (2D) location space. Representation of roadway network more complicated because a node may have multiple upstream nodes and/or multiple downstream nodes. In this case, the old representation is not sufficient and a set has to be used to store the nodes. Nodes in the 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, each node remembers its adjacent node and potential destinations. By this way, it is guaranteed that, when processing a lattice point, all the necessary information has been ready.

6.1.2. O-D Flows

O-D flows in Newell's theory are quite simple due to its simp le 1D location space. 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 with the help of 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 our model. The goal of our model is to determine the cumulative arrival and departure curves at every node destined for all its potential destinations.

6.1.3. Notation

Due to the extension to 2D location space, Newell's original notation is insufficient and the revis ed notation is presented below with the aid of a sketch of a general node in Figure 6-1.

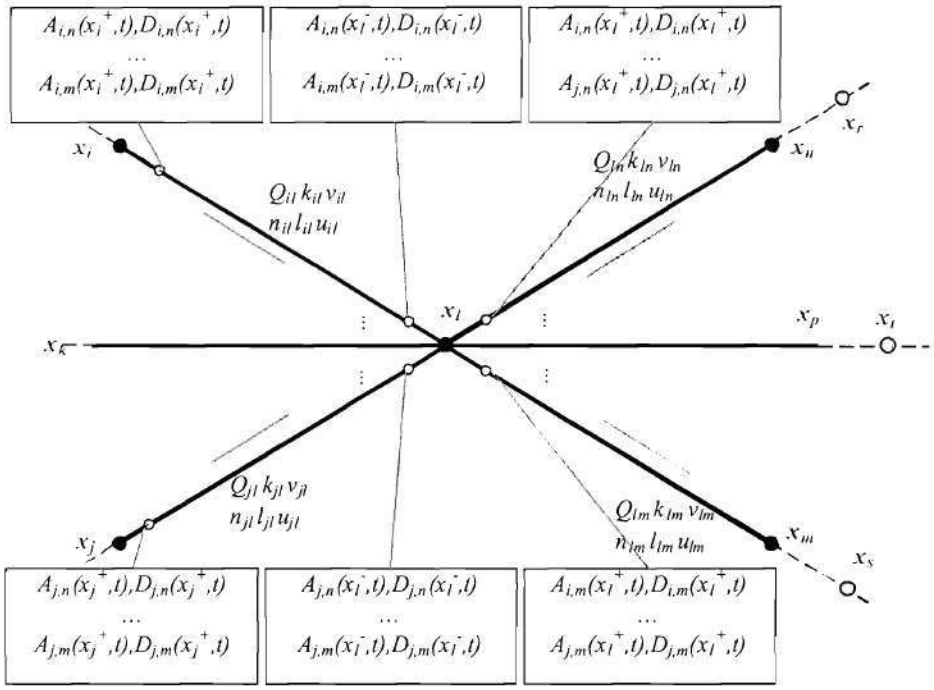


FIGURE 6-1 A General node of a transportation network

- 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 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 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.

6.2. Simplified Theory of Kinematic Waves: General Case

With the sorted set of nodes, we are still able to construct a lattice for the general case problem. Cumulative curves are evaluated progressively at each lattice point in the order mentioned above. At each lattice point, a 5-step procedure formulated below is carried out.

A. Departure to the right
Goal: determining the β downstream cumulative departure curves:

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

Since there are β links to the right of x_l, $\{(x_l, x_n), ..., (x_l, x_m)\}$, cumulative departure curves $D_{l,n}(x_l^+, t)$, ..., $D_{l,m}(x_l^+, t)$ 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:

I. 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.,

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

II. Right capacity

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

III. Downstream queue

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

IV. 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:

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

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:

I. Upstream arrival

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

II. Left capacity

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

III. 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	$\rightarrow$	$R_{\text{In}} = D_{l,n}(x_l^+, t) - D_{l,n}(x_l^+, t-\tau)$
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	$\rightarrow$	$\Delta'_{\text{In}} = R_{\text{In}} \times \dfrac{Q_{il}}{Q_{il} + Q_{jl}}$
	.....	

Δ_α	$\rightarrow$	$\Delta_{\text{In}}^{j} = R_{\text{In}} \times \frac{Q_{jl}}{Q_{il} + Q_{jl}}$
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 $\alpha \times \beta$ matrix whose elements are current departure counts to the right of x_l:

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

...

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

...

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

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:

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

...

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

...

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

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:

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

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

.....

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

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).

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

...

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

...

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

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	$\rightarrow$	$D_{i,l}(x_l^-,t)$ - $D_{i,l}(x_l^-,t-\tau)$
$R_1,\ldots,R_\beta$	$\rightarrow$	$d_{i,n}(x_l^+,t)$,..... $d_{j,n}(x_l^+,t)$, respectively
$d_1,d_2,\ldots d_\beta$	$\rightarrow$	$d_{i,n}(x_l^-,t)$,..... $d_{j,n}(x_l^-,t)$, 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:

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

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

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.,

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

...

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

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

CHAPTER 7
TEST SITE AND TEST DATA

As mentioned in earlier chapters, running the proposed model requires the following information:
Time-varying origin-destination (O-D) flows of the freeway system under study. This is actually the traffic demand at every entrance to every exit, and is required by the model as input information.

- Geometry data of the freeway system, including location of entrances, location of exits, and length and number of lanes of each link of the freeway system. This is also required by the model as input information.
- Traffic characteristic data including capacity, free-flow speed (FFS), and jam density of each link of the freeway system. This data uniquely defines the triangular flow-density curve assumed in the Simplified Theory, and again, is required by the model as input information.
- In addition, observed volume, speed, and density of each link of the freeway system are desirable. In this study, empirical tests are based on the comparison between predicted and observed traffic density. Therefore, observed traffic density in every link is of our primary interest. Though we do not compare predicted and observed volume and speed directly, but observed volume and speed provide complementary information about density and are very useful. On the other hand, they are also the necessary input to estimate the time-varying O-D flows mentioned above.

Data preparation involves selecting study site, determining specific segments of the study site as test sites, collecting the above data for the test sites and estimating origin-destination flows if necessary.

7.1 The study site

ITS data collected on GA 400 by Georgia NAVIGATOR system is used to provide empirical data for testing the proposed model. Currently, the automated traffic surveillance system covers the section of GA 400 between I-285 (to the south) and Old Milton Parkway (to the north), a stretch of road of approximately 12.56 miles and this serves as our study site, as shown in Figure 7-1. Though GA 400 does not provide completely uninterrupted flow along all its way, the section under coverage does. Three test sites are selected from the study site, site 1 is a merge case, site 2 is a diverge case, and site 3 is a corridor/network case.

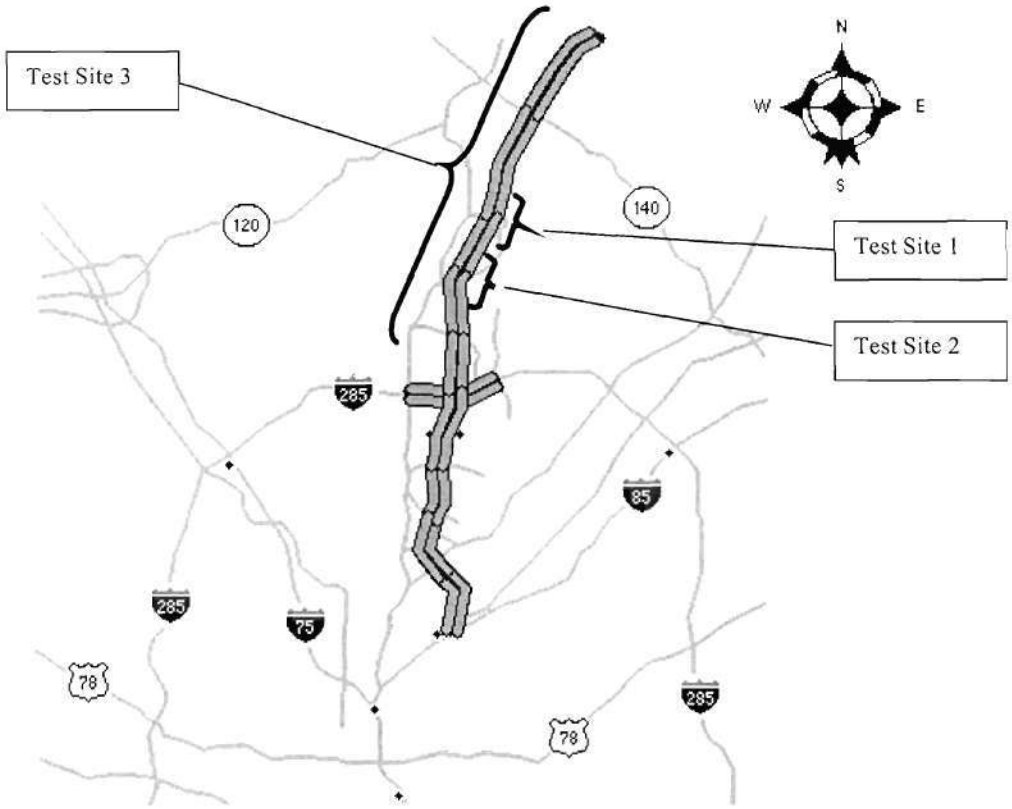


FIGURE 7-1 The study site – GA 400

7.2 Data of the test sites

Figure 7-2 sketches the topology of test site 1 which is a merge. There are 4 links, one entrance link followed by a merge which has two upstream links and one downstream link.

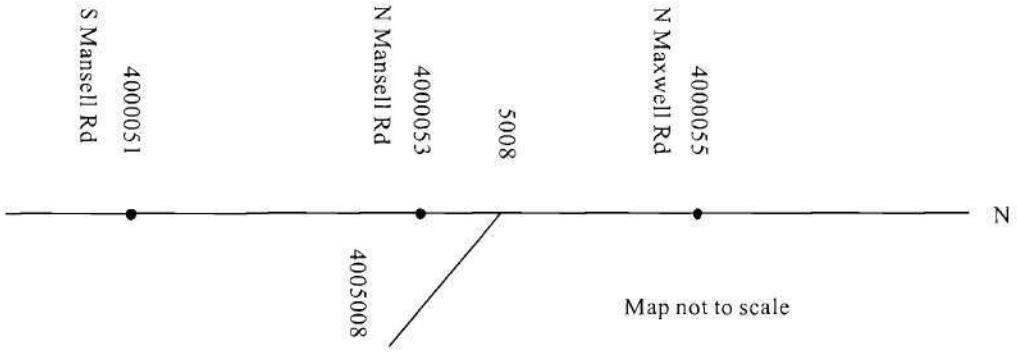


FIGURE 7-2 Test site1

Table 7-1 lists the geometry and traffic characteristics data for test site 1. The merge might be a bottleneck because its downstream capacity is less than the sum of its upstream capacities.

TABLE 7-1 Geometry and traffic characteristics data for test site 1

Link	Up Node	Down Node	Length (mi)	Lanes	Type	FFS (mi/h)	Capacity (veh/h/ln)	Jam Density (veh/mi/ln)
1	4000051	4000053	0.65	3	Mainline	65	2200	180
3	4000053	5008	0.16	3	Mainline	65	2200	180
4	5008	4000055	0.45	3	Mainline	60	2200	180
7	4005008	5008	0.50	1	On-ramp	20	1800	180

Figure 7-3 sketches the topology of test site 2 which is a diverge. This time, an entrance link is followed by a diverge which has one upstream link and two downstream links.

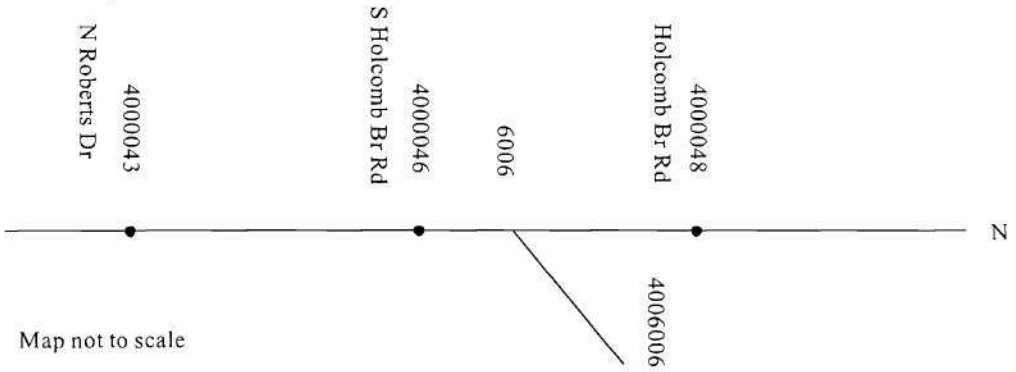


FIGURE 7-3 Test site 2

Table 7-2 lists the geometry and traffic characteristics data for test site 2. The diverge might be a bottleneck because queues may back onto this point from either of the diverging branches.

TABLE 7-2 Geometry and traffic characteristics data for test site 2

Link	Up Node	Down Node	Length (mi)	# of Lanes	Type	FFS (mi/h)	Capacity (veh/h/ln)	Jam Density (veh/mi/ln)
1	4000043	4000046	0.94	4	Mainline	68	2200	180
2	4000046	6006	0.23	4	Mainline	68	2200	180
3	6006	4000048	0.52	4	Mainline	60	2200	180
4	6006	4006006	0.50	1	Off-ramp	60	2000	180

Figure 7-4 sketches the topology of test site 3 in which a freeway has four on-ramps and four off-ramps.

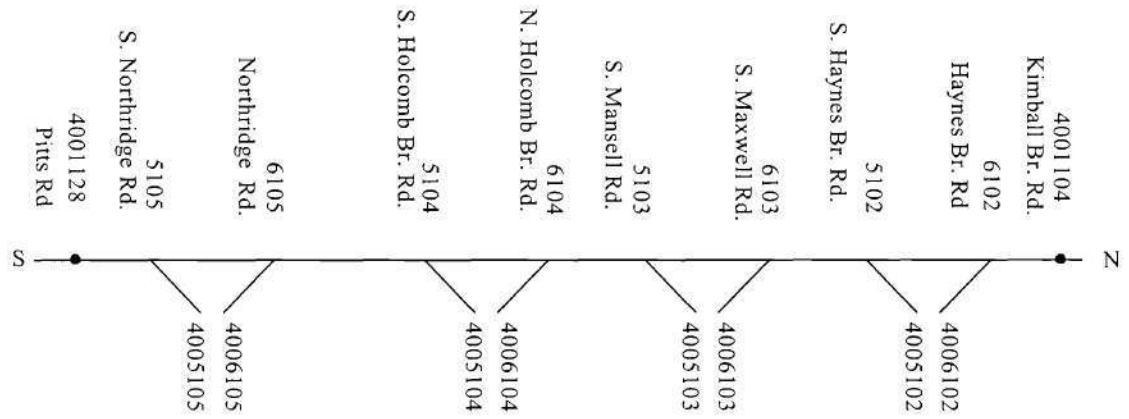


FIGURE 7-4 Test site 3

Table 7-3 lists the geometry and traffic characteristics data for test site 3. There are two potential bottlenecks at this site, one is node 4001128 and the other is node 6104. The first bottleneck is caused by congestion at further downstream that backs onto this node. The second bottleneck is due primarily to high exit volume.

TABLE 7-3 Geometry and traffic characteristics data for test site 3

Link	Up Node	Down Node	Length (mi)	# of Lanes	Type	FFS (mi/h)	Capacity (veh/h/ln)	Jam Density (veh/mi/ln)
1	4001104	6102	1	2	Mainline	56	2200	180
2	6102	5102	0.4	2	Mainline	57	2200	180
3	5102	6103	1.1	3	Mainline	61	2200	180
4	6103	5103	0.9	3	Mainline	65	2100	180
5	5103	6104	0.88	3	Mainline	61	2200	180
6	6104	5104	0.5	3	Mainline	60	1900	180
7	5104	6105	2.9	4	Mainline	64	2100	180
8	6105	5105	0.5	4	Mainline	60	2100	180
9	5105	4001128	0.5	4	Mainline	56	2200	180
10	6102	4006102	0.5	1	Off-ramp	22	1000	180
11	4005102	5102	0.5	1	On-ramp	32	2000	180
12	6103	4006103	0.5	1	Off-ramp	20	1000	180
13	4005103	5103	0.5	1	On-ramp	22	1250	180
14	6104	4006104	0.5	1	Off-ramp	40	1200	180
15	4005104	5104	0.5	2	On-ramp	40	2000	180
16	6105	4006105	0.5	1	Off-ramp	20	1500	180
17	4005105	5105	0.5	1	On-ramp	20	1700	180

7.3 O-D estimation

As stated above, running the model requires as input the time-varying origin-destination (O-D) flows, which is assumed to be known. However, in reality this kind of information is rarely readily available and has to be estimated. O-D estimation is not part of the model but a necessary step to provide input data. Many O-D estimators are available to generate O-D matrices from link traffic counts, and the following considerations drive our selection. First, the estimator needs to have satisfactory accuracy to reproduce O-D pattern from link traffic counts. Second, the estimator needs to be relatively cheap to implement because it is not part of our model but a convenient tool to provide input data. Third, the estimator needs to have the potential of working on-line to facilitate possible online application of the model in the future.

Among the many alternatives, recursive predictions error (RPE) estimator proposed by Nihan and Davis (1987) is selected. The estimator, combining features of recursive method and Kalman filtering method, has a lot of merits and meets the criteria set forth. However, empirical results reveal that the estimation error is still considerable in absolute terms. This error, contained in the input data, will eventually go through our model and mix with modeling error. Considering that it is typically difficult to separate the O-D estimation error from modeling error during validation phase, it is desirable to have some tests that involve no O-D estimation. In this study, test 1 is a merge case and 2 is a diverge case, both require no O-D estimation. Test 3 is a corridor/network case which requires O-D estimation.

CHAPTER 8
EMPIRICAL TEST RESULTS

Any model, to be creditable and useful in practice, has to undergo validation, a formal procedure to assess the performance of the model. Empirical tests generally compare model prediction against field observation. As we know that modeling error is inevitable due to model simplification and abstraction. If, to run a model, the result of some other model, which is also subject to modeling error, has to be used, the effect of these errors can be mixed and propagated. What makes things further complicated is that it is usually very hard to clearly separate errors of different sources, so it is unfair to evaluate a model with the presence of error of other sources. For example, origin-destination (O-D) estimation is such a case to our model which implicitly assumes as part of its input accurate O-D flows. Since this kind of information is rarely readily available and typically has to be estimated from observed link flows. O-D estimation error is introduced and mixed with our model. With this in mind, the empirical tests are carried out with and without the need of O-D flows. More specifically, we are going to test a merge scenario and a diverge scenario which don't require O-D estimation and a corridor/network scenario which requires O-D estimation. The empirical tests are based on comparing observed and predicted traffic densities. Both qualitative and quantitative measures are employed to asses the performance of the proposed model.

8.1. Empirical Test 1 – Merge Scenario

Two tests, with one day for each, are performed based on test site 1, a merge scenario.

8.1.1. Test Day 1

The data for test 1 was collected on Monday, October 14, 2002 from 5:52 to 19:57. Figure 8-1 shows density contour for the traffic condition of the day based on 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 reasons. First, HCM 2000 uses a density value of 45 pc/mi/ln as the upper bound of level of service (LOS) F, which 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 at the middle of this range. Third, the contour lines of 40 and 50 veh/mi/ln are pretty close to each other in most of the cases. Therefore, to keep the contour clean, it is appropriate to use 45 veh/mi/ln to delineate the boundary of congested and uncongested regions.

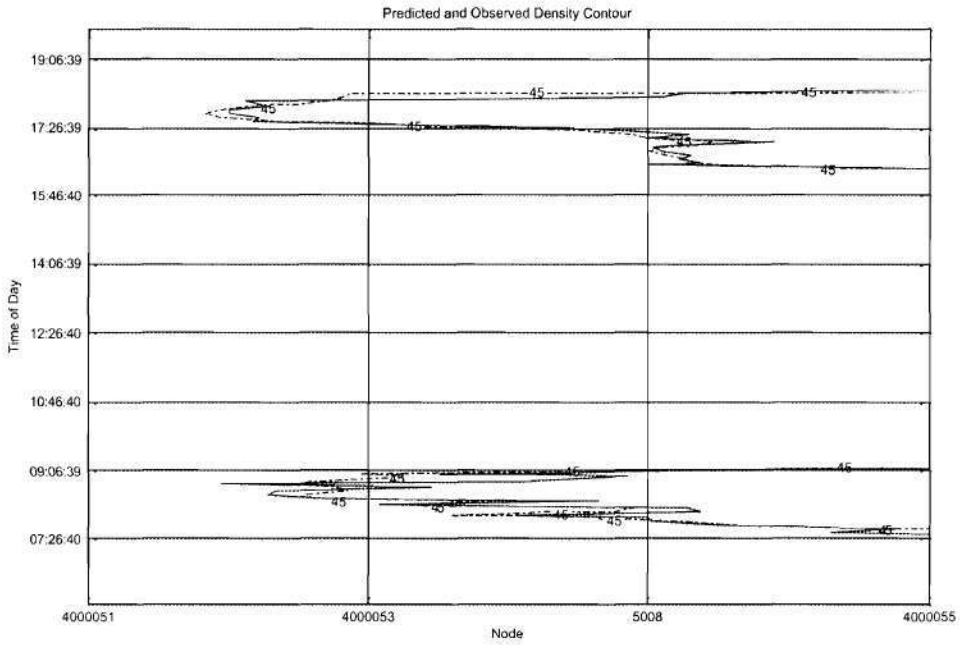


FIGURE 8-1 Density contours of test 1 (Merge Scenario)

The figure shows the contour of the congestion in time-space domain. 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. The figure shows a good fit between the prediction and the observation.

The simulation starts at 5:52 and ends at 19:57. There are two peaks on this day, one in the morning and the other in the afternoon. Both are originated from the downstream of node 4000055, probably due to insufficient capacity there. Since the real bottleneck is located outside of our test site, a dummy link is added at the downstream of node 4000055 and a time varying capacity is applied at the dummy link to simulate the effect of the real bottleneck which is located somewhere further downstream. There is no peak at the on-ramp link 4005008-5008 because its fair share of downstream supply always suffices its demand, which is exactly what the merge scenario predicts.

Generally, it is preferable to have the congested regions contained by the bounding box of 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 and longer period where congestion is unlikely to reach the boundaries. This is particularly difficult for merging and diverge scenarios since these scenarios deals only with a ramp junction. Fortunately, the simplified theory doesn't prohibit congestion backing up from downstream end as long as it doesn't exceed upstream end. Otherwise, the flow at the upstream end doesn't represent the true demand.

Figure 8-2 illustrates the frequency of prediction error. As expected, the error is densely concentrated around zero with the rest balanced at both sides – a bell-shaped distribution.

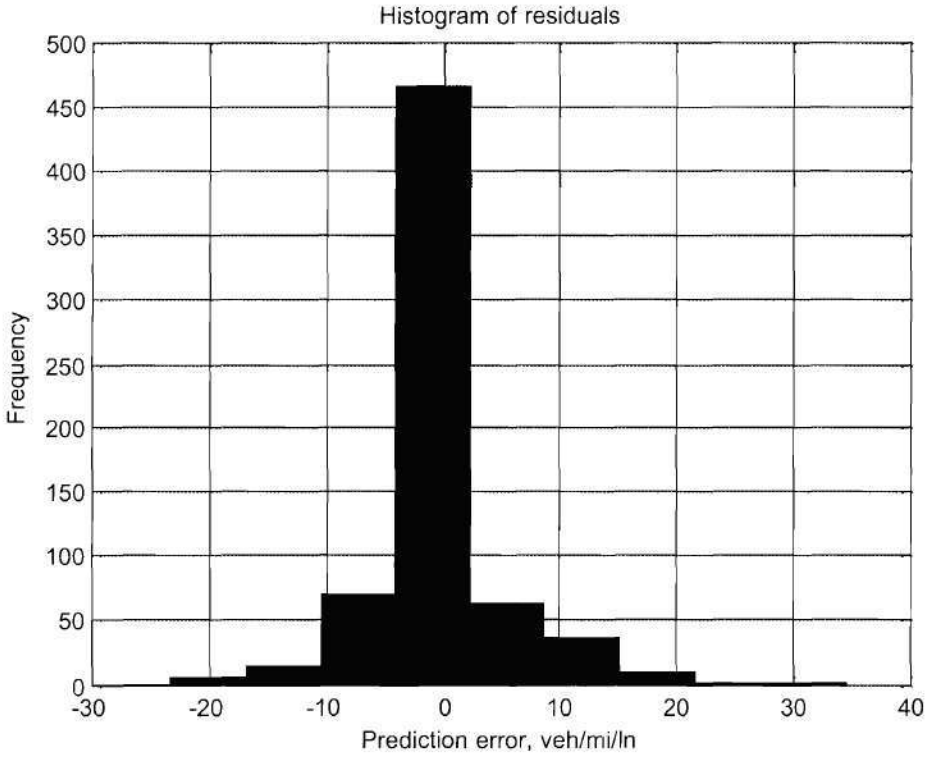


FIGURE 8-2 Histogram of Prediction Error of Test 1 (Merging, 676 Samples)

So far the qualitative measures have shown encouraging results, but we are not sure how much confidence we have before quantitative assessment is made. There are three issues to address when running statistical tests: (a) test whether the model is unbiased in prediction; (b) test whether the model has sufficient precision in prediction; (c) deal with time series and correlation which are typical in simulation outputs. With these in mind, we are going run a simultaneous statistical test based on the following two hypotheses:

Hypothesis 1: the prediction is unbiased, i.e., the mean of prediction error is not statically different than 0. This is intended to address issue (a) and a t-test based on prediction error suffices the need.

Hypothesis 2: the prediction has sufficient precision, i.e., the variance of prediction error is reasonably small. This is intended to address issue (b) and a Chi-square test is employed.

To address issue (c), we are going to run the above tests based on batch means technique (Goldsman and Tokol 2000) which is particularly designed to deal with correlation in simulation outputs. For more details on this technique, please refer to the original paper.

Statistical test results are encouraging, test on hypothesis 1 confirms that the mean of prediction error is not statistically different than 0 at the 95% confidence level. On the other hand, the Chi-square test suggests a 95% confidence interval of (-0.083738, 0.037963) $\times$ 100% for percentage prediction error.

8.1.2. Test Day 2

The data for test 1 was collected on Friday, September 6, 2002 from 00:01 to 23:51, almost a whole day. Comparison of model simulation and field observation are made based on traffic density. Figure 8-3 shows density contour of the day based on density level of 45 veh/mi/ln. The X axis represents location (nodes along the freeway mainline) and the Y axis represents time of day. The solid line is the observed density contour and the dashed line is the simulated contour. There are two major peaks, one in the morning and the other in the afternoon. Both are originated from the downstream of node 4000055. The morning peak is relatively small in scale, while the afternoon peak is larger and virtually split into two. The splitting effect can be caused by a sudden but brief drop in upstream demand which issues a forward moving shock wave, i.e., the queue tail in this case, heading towards the

downstream bottleneck. When the queue tail barely passes node 4000055, the upstream demand increases again, and the queue begins to grow until next drop in demand. In general, the simulation matches the observation pretty well in queue formation and dissipation, but less satisfactory in transition area between queued and unqueued traffic.

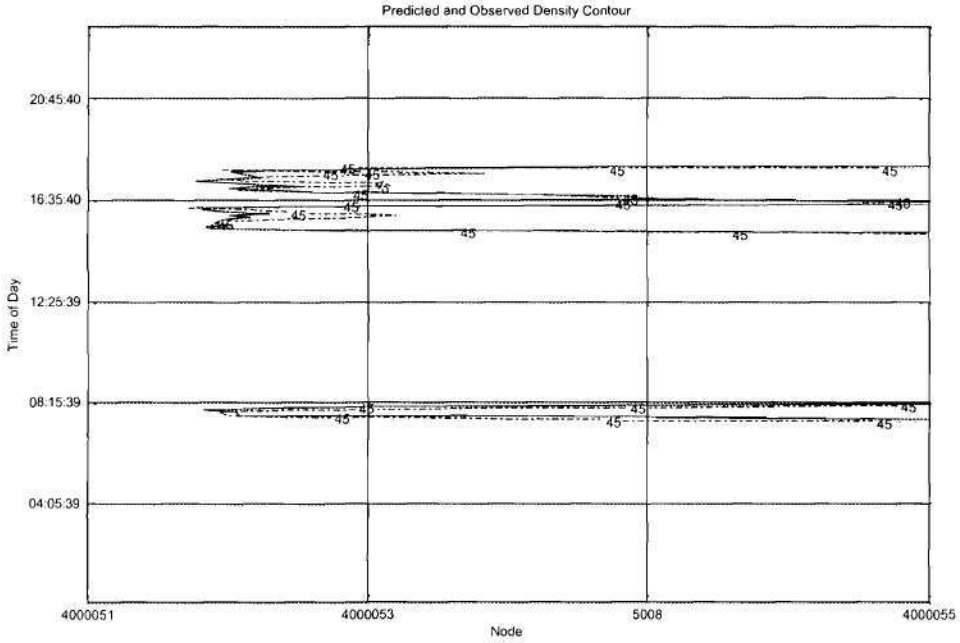


FIGURE 8-3 Density Contours of test 1

Figure 8-4 shows the diagonal plot of simulated vs. observed densities. Density values are log-transformed to account for the fact that simulation error tends to increase as density goes high. A perfect fit is an up-sloping diagonal line which is approximated pretty well by the data points.

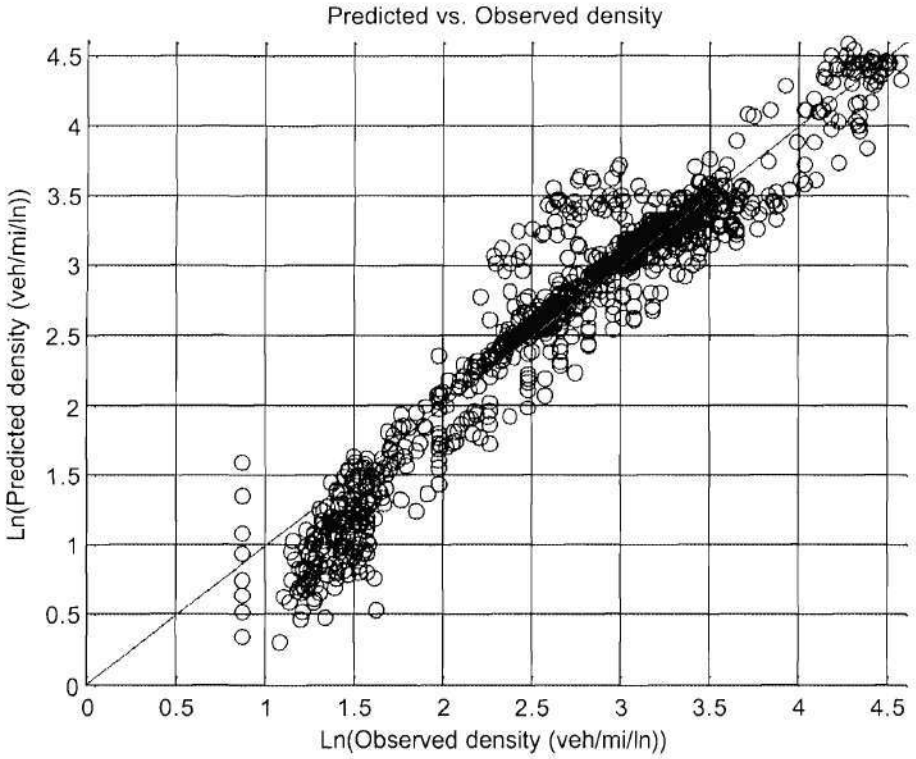


FIGURE 8-4 Diagonal Plot of test 1 (1144 Samples)

Statistical analysis on the simulation outputs indicates that the mean of simulation error is not statistically different than 0 at the 95% confidence level, and the 95% confidence interval of percentage error is *(-0.035499, 0.059814)* $\times 100\%$.

8.2. Empirical Test 2 – Diverge Scenario

Two tests, with one day for each, are performed based on test site 2, a diverge scenario.

8.2.1. Test Day 1

Data for test 2 was collected on Thursday, September 12, 2002. The test runs from 00:01 to 23:51, almost a whole day. Figure 8-5 shows density contours of this test. There are two peaks during the day. The morning peak originates from the downstream of node 4000048 probably caused by insufficient capacity, while the afternoon peak backs up from the downstream of node 4006006 due primarily to high exit volume.

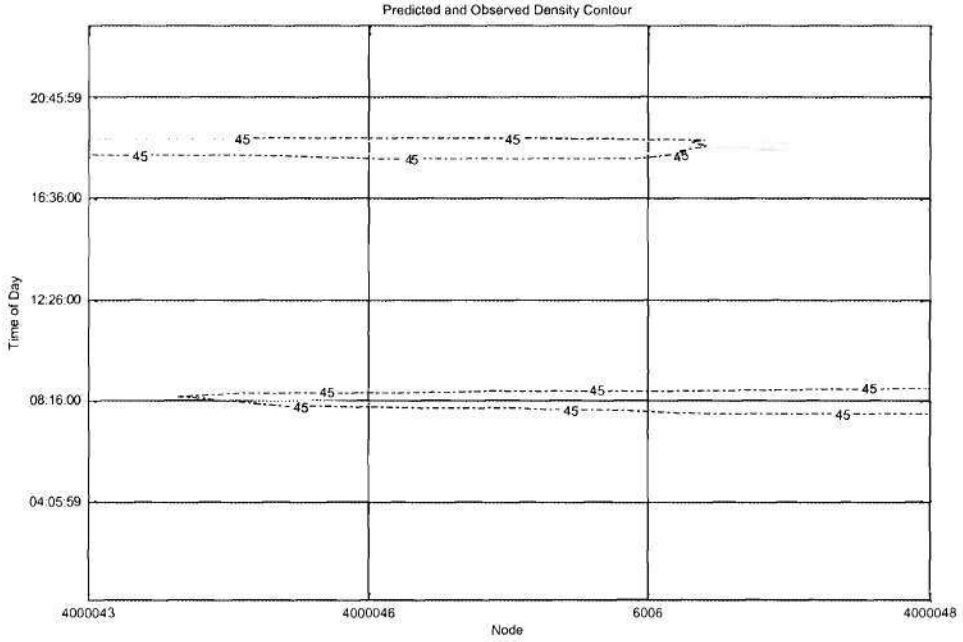


FIGURE 8-5 Density contours of test 2 (Diverge Scenario)

As we discussed in previous section, it is preferable to have congestion contained in the bounding box of the time - space diagram. However, congestion backing up from downstream is also acceptable. If congestion backs up past the upstream end, 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 contained in the new bounding box.

Unfortunately, the afternoon peak in our case passes the upstream end, but we are unable to find an uncongested node further upstream because, otherwise, we would have to incorporate another freeway junction which contradicts the premise of a diverge scenario. Therefore, limited confidence should be given to the portion where the congestion exceeds the upstream boundary.

Nevertheless, this test is a good example to demonstrate the validity of the proposed diverge model, which allows congestion from any downstream branch, and, if there is a queue, upstream traffic will be delayed accordingly. In general, the figure shows a good agreement between the prediction and the observation.

Figure 8-6 shows the histogram of prediction error. Again, a bell-shaped distribution is resulted.
Simultaneous statistical test shows that the mean of prediction errors is not statistically different than 0 at 95% confidence level, and the 95% confidence interval for percentage errors is (-0.034187, 0.045355) 100%.

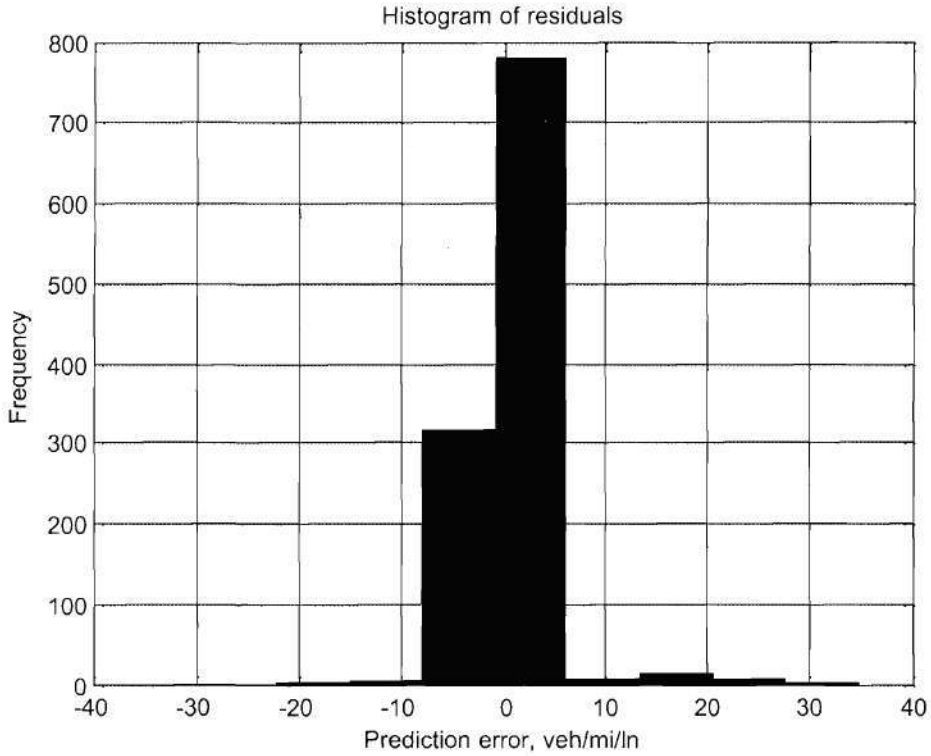


FIGURE 8-6 Histogram of prediction error of test 2 (Diverging, 1144 Samples)

8.2.2. Test Day 2

Data for empirical test 2 was collected on Monday, December 9, 2002 and the density contour of the day is shown in Figure 8-7. There are two peaks during this day, a morning peak and an afternoon peak. The morning peak, originated from the downstream of node 4000048, is caused by insufficient downstream capacity.

What makes this example interesting is that the afternoon peak is caused by congestion at both downstream links (6006-4000048 and 6006-4006006). At approximately 16:40, a queue backs onto link 6006-4000048 and continues to move backward. At approximately 17:51, another queue pops in at link 6006-4006006 and also keeps moving backwards. Both of the queues pass the diverge and back onto link 4000046-6006, jointly constraining upstream traffic. The off-ramp queue disappears rough 24 minutes later at 18:15, and the mainline queue dissipates at approximately 18:00. The figure also shows a good agreement between the simulation and the observation.

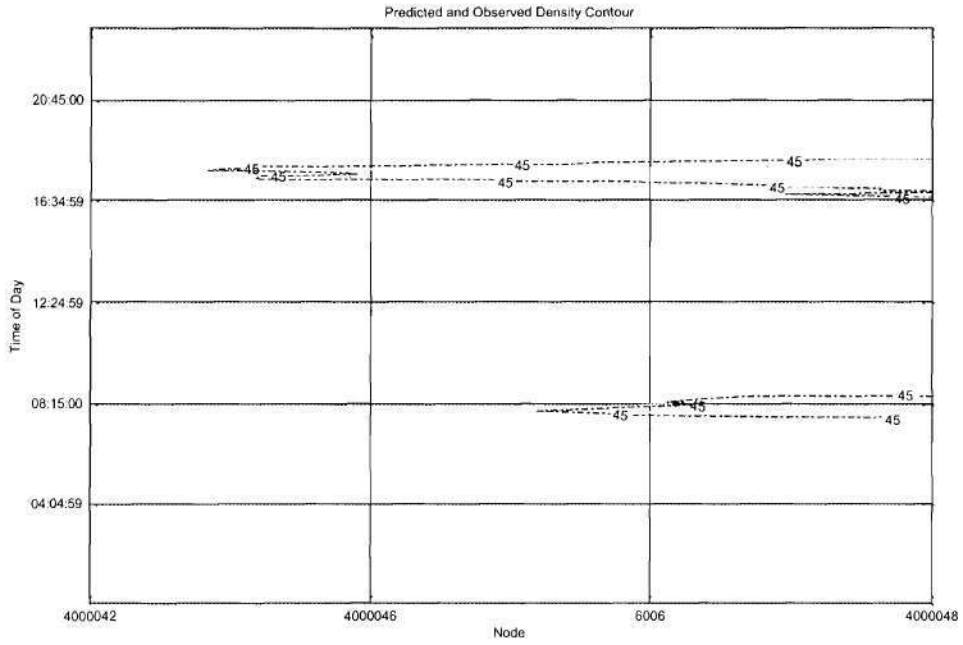


FIGURE 8-7 Density Contours of test 2

Figure 8-8 shows the diagonal plot of the simulated and the observed densities after log-transformation. Again, data points are clustered around 45 degree line, indicating a good fit.

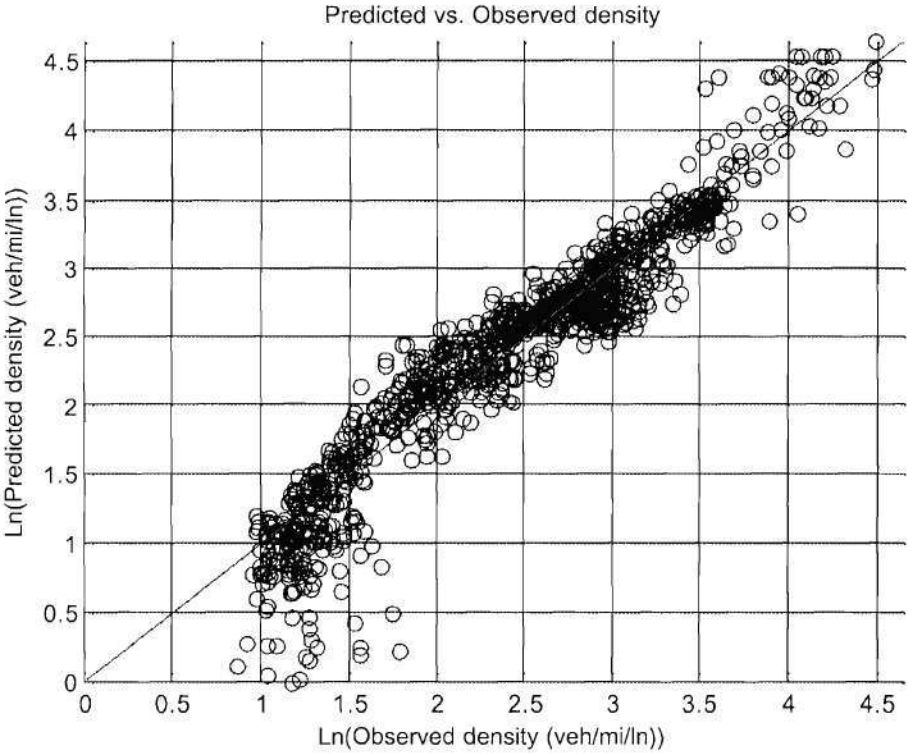


FIGURE 8-8 Diagonal Plot of test 2 (1144 Samples)

Statistical analysis on the simulation outputs indicates that the mean of simulation error is not statistically different than 0 at the 95% confidence level, and the 95% confidence interval of percentage error is (-0.037407, 0.063765) $\times$ 100%.

8.3. Empirical Test 3 – Corridor/Network Scenario

Two tests, with one day for each, are performed based on test site 3, a corridor / network scenario.

8.3.1. Test Day 1

Figure 8-9 shows density contours of the test which has two peaks, one in the morning and the other in the afternoon. The morning queue appears approximately between 7:00 and 8:30 and extends from the downstream end backwards up to node 6103. At about 6:50, the morning queue starts to build up passing node 4001128. About 40 minutes later, the queue reaches node 6103. The queue then starts to dissipate and reaches somewhere at the downstream of node 5103 at 8:15. In the next 15 minutes, the queue swiftly shrinks to node 4001128 and disappears before 9:05.

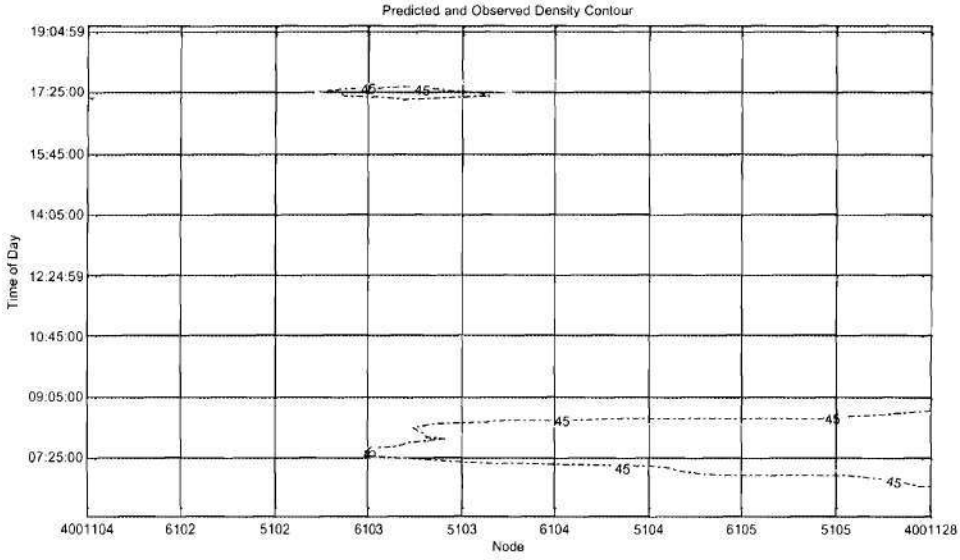


FIGURE 8-9 Density Contours of Test 3 (Corridor/Network Scenario)

The afternoon peak is much smaller in scale. It lasts approximately from 17:15 to 17:55 and spans 3 links between nodes 5102 and 6104. The observed peak consists of two congested areas, a major one and minor one. This is probably due to recurrent congestion, i.e., the congestion first backs up past this point, it then shrinks to somewhere downstream of the point, and back up past this point again before long.

There are, of course, some slight discrepancies in the figure. For example, the predicted morning congestion builds up a little bit earlier and shrinks a little bit later than the observed one. On the other hand, the predicted afternoon congestion fails to capture the recurrent effect. Given these, the density contour shows a good agreement between predicted and observed congestion regions in general.

Figure 8-10 shows the histogram of prediction error. Again, prediction error is densely clustered around 0.

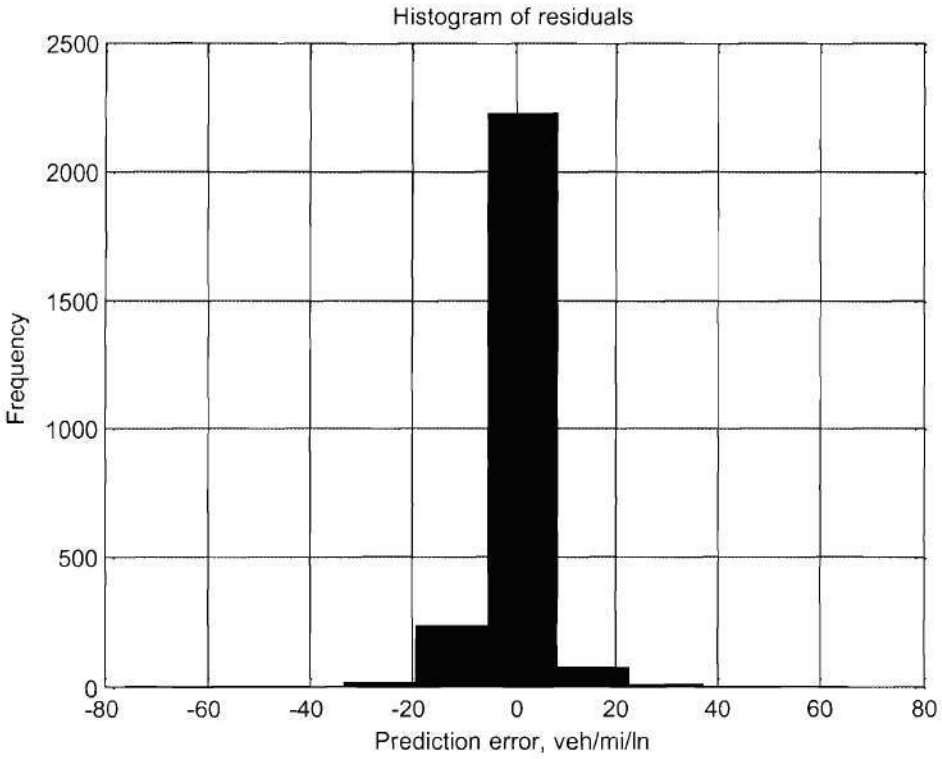


FIGURE 8-10 Histogram of prediction error of test 3 (Corridor/Network, 2592 Samples)

Simultaneous statistical test shows a 0 mean in prediction error at 95% confidence level and a 95% confidence interval of (-0.088238, 0.094723) × 100% for percentage prediction error.

8.3.2. Test Day 2

The data for test 3 was collected on Monday, Oct. 14, 2002 and Figure 8-11 shows the density contour of the day. Again, a morning peak and an afternoon peak show up.

The morning peak, roughly from 6:40 to 9:00, originates from downstream of node 4001128 and propagates backwards at nearly a constant speed. At 7:36, the queue tail reaches somewhere between nodes 6102 and 5102. Then the queue begins to recede at almost the same speed, and moves out of the diagram at approximately 9:00. Notice that the model predicts the queue formation, i.e., the lower boundary of the morning peak, very well, but there is a lag in simulated queue dissipation, i.e., the upper boundary of the morning peak. Nevertheless, the simulation matches the observation pretty well in terms of impact area, the timing of congestion, and the location of queue tail.

The afternoon peak, which occurs approximately from 17:00 to 18:40, is caused by high exit volume at link 6104-4006104. A queue starts to build up there and backs onto the mainline at link 5103-6104. A close look at the data reveals that there is no congestion at link 6104-5104, which means the contribution of this link to the congestion is secondary. This confirms our intuition that a queue can build up from off-ramps and hence supports the validity of our relaxation of Newell's assumption on off-ramps. It is interesting to note that the observed contour exhibits a fork-liked shape, suggesting that the congestion initially reaches somewhere upstream of node 6103, it then shrinks a little bit, and later the queue tail returns and moves further upstream past node 5102. Eventually, the queue dissipates at approximately 18:42. Again, the simulation matches the observation quite closely except that it fails to capture the "fork".

In addition to the two major peaks, there are other 5 little peaks in the contour, with the biggest being the one appearing around 18:00 at the upstream of the afternoon peak. Notice that all of them are tiny in scale and disconnected from anyone else. These suggest that they are probably caused by some spontaneous and random disturbance and result in little impact on traffic operation. The inability of our model to capture these details reveals the generally limited power of macroscopic approach which is typically good at capturing sustained and recurrent queues.

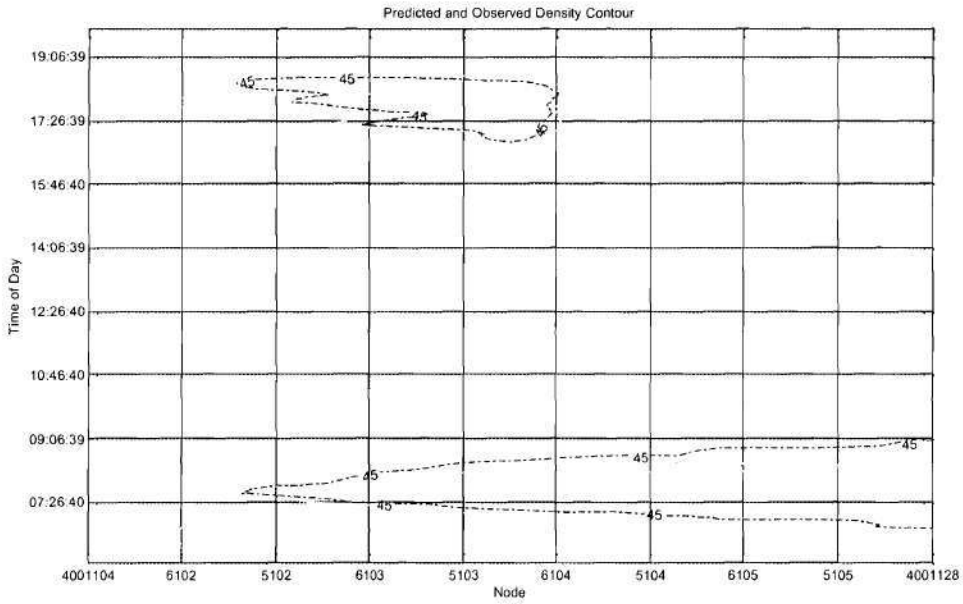


FIGURE 8-11 Density Contours of test 3

Figure 8-12 shows a fatter shape than previous two tests, but the result is quite satisfactory considering that the simple size is more than doubled and the data points are still densely clustered around the diagonal line.

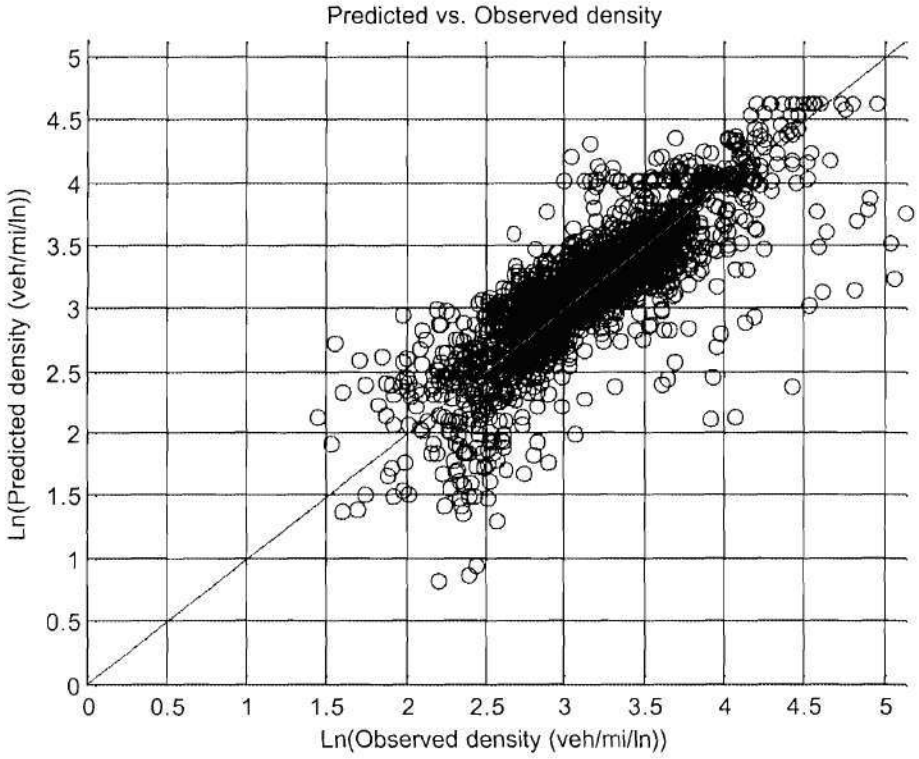


FIGURE 8-12 Diagonal Plot of test 3 (2704 Samples)

Statistical analysis on the simulation outputs indicates that the mean of simulation error is not statistically different than 0 at the 95% confidence level, and the 95% confidence interval of percentage error is (-0.10866, 0.12243) $\times$ 100%.

CHAPTER 9
SUMMARY AND FUTURE DIRECTIONS

This report summarizes Newell's simplified theory of kinematic waves in a 5-step procedure and extends the theory to a freeway system by proposing procedures to deal with a set of basic building blocks. Among the important extensions are merge and diverge scenarios which are made by relaxing some of Newell's assumptions.

The proposed merge scenario relaxes the original assumption that on-ramp traffic always has the priority and can bypass queues, if any, at the merge. As a replacement, a capacity-based weighted fair queuing (CBWFQ) merge model is used, which enables merging traffic to compete downstream supply and applies a capacity-weighted distribution only when both merging branches are dictated by backward waves.

The proposed diverge scenario relaxes the original assumption that exiting traffic can always exit without delay and downstream congestion affects upstream traffic as a whole. With the proposed contribution-based weighted splitting (CBWS) diverge model, this is no longer true because queues from a diverging branch affect only the corresponding part of upstream traffic, and upstream departure is based on contributions of downstream branches.

To enable the simplified theory to deal with a general transportation network, the above two models are even generalized so that multiple merging or diverging branches are allowed. The generalized capacity-based weighted fair queuing (CBWFQ) merge model and the generalized contribution-based weighted splitting (CBWS) diverge model are formulated. With the aid of the two generalized models, Newell's simplified theory is relaxed to deal with a general node with multiple legs, and the modeling of a general transportation network is a comprehensive exercise of the proposed procedure over the K-Wave space.

Three empirical tests are run to evaluate the model. The results show that the model performs consistently well even with the presence of errors of other sources, such as O-D estimation error. Visual comparison shows a good agreement between the prediction and the observation. The model does a good job predicting the formation and dissipation of queues. The majority of the prediction error is concentrated around 0 with the remainder balanced at both sides, which conforms our expectation of a normal-liked distribution. Quantitative assessment confirms the model's accuracy: the mean of prediction error is not statistically different than 0 and the variance of prediction error is reasonably small.

The proposed extensions and generalization have been implemented and is basically ready to use. Currently, it works as a stand-alone command-line simulation software, and some input and output functionalities are going to be developed to make the software more user-friendly. Future directions of the research involve the following: First, finalize the software so that it is easy to use, is able to validate input data, and is capable of generate user-friendly reports. XML technology provides an excellent tool to accomplish input data validation and output data representation. Second, push the software to online application. Since the algorithm is simple and efficient, online application is highly likely given the wide deployment of automated traffic surveillance systems. The benefit of this involves more ability for traffic management centers (TMCs) to make quicker response to congestion and incidents, optimize resources deployment, publish more accurate traveler information, and develop better-supported strategies.

REFERENCES

Banks, J.H. (2000) Are Minimization of Delay and Minimization of Freeway Congestion Compatible Ramp Metering Objectives? Transportation Research Record 1727, 112-119.

Daganzo, C. F. (1994) The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory, Transportation Research, Vol28B, No. 4, 269-287.

Daganzo, C. F. (1995). The Cell Transmission Mode, Part II: Network Traffic. Transportation Research, Vol. 29B, No. 2, 79-93.

Daganzo, C. F. (1995a) A Finite Difference Approximation of the Kinematic Wave Model of Traffic Flow. Transportation Research, Vol. 29B, No.4, 261-276.

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. (1999) The lagged cell-transmission model. Transportation and Traffic Theory, 14th ISTTT Symposium, 81-104.

Gerlough, D. L. and Huber, M. J. (1975) Traffic Flow Theory: A Monograph. Transportation Research Board, S.R.165.

Goldsman, D. and Tokol G. (2000) Output analysis: output analysis procedures for computer simulations. Winter Simulation Conference 2000, 39-45

Hurdle V.F., Son B. (2000) Road Test of a Freeway Model. Transportation Research, Vol. 34A, No. 7, 537-564.

JIN, W. L. AND ZHANG, H. M. 2003. On the Distribution Schemes for Determining Flows Through a Merge. Transportation Research, Vol. 37B, No. 6, 521-540.

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, 647-677.

Leonard II, 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., II. (2001) A tool for evaluating freeway congestion. Online paper located at http://traffic.ce.gatech.edu/gtwaves. Accessed June 2001.

Lighthill, M. and Whitham, G. (1955) On kinematic waves II. A theory of traffic flow on long crowded roads, Proc. Royal Society of London, Part A, Vol. 229, No. 1178, 317-345.

Michalopoulos, P. G. (1984) Dynamic freeway simulation program for personal computers, Transportation Research Record 971, 68-79.

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., Kwon, E., and Kang, J-G. (1991) Enhancement and field testing of dynamic freeway simulation program. Transportation Research Record 1320, 203-215.

Michalopoulos, P.G., Beskos, D.E., Lin, J.K. (1984) Analysis of Interrupted Traffic Flow by Finite Difference Methods. Transportation Research, Vol18B, 409-421.

Newell, G. F. (1993a) A simplified theory on kinematic waves in highway traffic, Part I: general theory. Transportation Research, Vol. 27B, No. 4, 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, 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, 305-313.

Newell, G. F. (1999) Delays Caused by a Queue at a Freeway Exit Ramp. Transportation Research, Vol. 33B, 337-350.

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, 149-163

Payne, H. J. (1979) FREFLO: A macroscopic simulation model for freeway traffic. Transportation Research Record 722, 68-77.

Richards, P. I. (1956) Shock waves on the highway. Operations Research 4, 42-51.

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.

Stephanopoulos, G. and Michalopoulos, P. G. (1981) An Application of Shock Wave Theory to Traffic Signal Control. Transportation Research Journal, Vol. 5B, 35-51

Stephanopoulos, G. and Michalopoulos, P. G. (1979) Modelling and Analysis of Traffic Queue Dynamics at Signalized Intersections. Transportation Research, Vol. 13A, 295-307

Wong, S. C. and Wong, G. C. K. (2002) An Analytical Shock-Ftting AlgorithmFor LWR Kinematic Wave Model Embedded With Linear Speed-Density Relationship. Transportation Research, Vol. 36B, No. 8, 671-754.