A FRAMEWORK FOR PREDICTING URBAN FREEWAY INCIDENT DELAYS IN REAL TIME

by

Scott Mastbrook

Thesis submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in

Civil Engineering

APPROVED:

A. G. Hobeika, Ph.D., Chairman

D. R. Drew, Ph.D.

R. Sivanandan, Ph.D.

January 25, 1996
Blacksburg, Virginia

Key Words: freeway, incident, delay

c.2

A FRAMEWORK FOR PREDICTING URBAN FREEWAY INCIDENT DELAYS IN REAL TIME

by

Scott Mastbrook

A.G. Hobeika, Ph.D., Chairman

Civil Engineering

(ABSTRACT)

Modern traffic management centers often incorporate incident management support systems. One of the major components of these systems is an incident response module. Decisions on incident response are generally made on a prediction of how much impact an incident will have on a freeway in the sense of delays to drivers.

The goal of this research was to build the framework for an algorithm that can more accurately predict delays due to incidents in real-time given a sufficient amount of data. In this research, many existing delay estimation methods were examined. These include heuristic models, queuing theory, shock-wave theory, time-series methods, and simulation. Most of these methods possessed drawbacks that preclude them from being used in a delay prediction algorithm.

The type of model that was found to give the most flexibility while still being fast enough for on-line applications is a macroscopic freeway simulation model. A dynamic extended model for urban freeways was proposed and tested that proved to model urban

freeway traffic phenomena very well. This dynamic model was taken as the basis for a new on-line delay forecasting system.

This research provides an architecture for incorporating a macroscopic simulation with procedures that can be effectively used to provide forecasts of traffic conditions for an urban freeway corridor in real time. The input procedures combine information about freeway characteristics, incidents, and traffic flow conditions into a data package that can be efficiently run through the simulation program. Additional research performed includes a unique survey on the phenomenon of dynamic lane clearance. This phenomenon has a significant impact on delay prediction and was implemented into the architecture. The output from the simulation in terms of traffic flow and speed forecasts is flexible enough be used in many traffic management applications.

When used with a real traffic data set, the simulation has shown to realistically predict future traffic conditions when given the proper input data. An open architecture was used in building the framework, leaving space for additional models that can improve accuracy. Additional research into these areas will further improve the algorithm, increasing both the stability and the accuracy of the forecasts.

ACKNOWLEDGMENTS

The author is indebted to Dr. Antonie Hobieka for his helpful suggestions and comments throughout the research. He would also like to give thanks to Kaan Özbay for his guidance, suggestions, support, and encouragement through all stages of his study. He would like to thank to Dr. Pushkin Kachroo for his helpful suggestions on several topics. He is also very thankful to his fellow students Yunlong Zhang, Wei Wu, and Arvind Narayanan for their helpful comments and suggestions.

He would like to especially thank Karl Petty from the Electrical Engineering Department at the University of California at Berkeley for the F.S.P. data, and his unending technical support and advice. The author also wishes to thank Major Ron Miner from the Virginia Department of Transportation for his extensive aid and cooperation in conducting the dynamic lane clearance survey.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1.Introduction

In the past decade, freeway congestion has become a part of daily life for metropolitan commuters. Congestion causes millions of hours of vehicle-delays each year and in turn millions of dollars in lost productivity. It has been shown that congestion has been increasing in urban areas and will continue to increase in the future. In the long term this may lead to hampered economic development and negative environmental impact.

Incidents are the major cause of freeway congestion today. It is estimated that they account for over 60% of the delays today, and this is likely to increase over the next ten years. Examples of incidents include: stalled vehicles, single vehicle accidents, multiple vehicle accidents, spilled loads, construction operations, extreme weather conditions, or anything that causes a temporary reduction in the capacity of the roadway and an increase in delays. To reduce the impact of these delays, advanced traffic management systems are being developed to better manage both the incident and the roadways affected by the incident. Incident management is the term used by many departments to describe programs that aid in the administration and handling of incident situations. These programs aid in collaborating the various incident clearance agencies together so they may perform as a team, thus expediting the clearance process and reducing delays caused by incidents. Traffic management systems include many surveillance, control, and information dissemination systems that aid in the detection, response, clearance, and handling of incident situations. These too, are aimed at reducing delays. (Cambridge Schematics, 1991)

The objective of this research is to develop procedures that can be used in an advanced traffic management system to predict delays caused by incidents. The research attempts to model incident delays in the terms of queue lengths, queue dissipation times, and increased travel times based on a number of factors such as predicted incident duration, expected traffic flows, and incident characteristics. An accurate prediction of delays caused by an incident can play a major role in decision making in many traffic management systems. Many decisions such as driver diversion, traffic advisories, and signal re-timings are based on the predicted delays of incidents.

The introduction of this thesis is divided into four sections:

1. A discussion on the types and impacts of congestion
2. A discussion on the need for real-time information
3. The scope of the research
4. The outline of the thesis

1.1. Types and Impacts of Freeway Congestion

Traffic congestion on freeways has been categorized into two major groups. These are 1) recurrent traffic congestion, and 2) non-recurrent traffic congestion. Recurrent traffic congestion is distinguished as congestion that occurs at the same location and time every day on a freeway for the same reason. This may be because of poor geometrics, local overloading, lane drops, or at particular freeway entrances. These are major problems, but can be dealt with on an individual basis. For example, in areas where there are high demands at peak periods, ramp metering schemes can be implemented to improve downstream flow.

Nonrecurrent traffic congestion is the result of traffic incidents that cause a temporary reduction in capacity of a particular freeway section and increase delays. As mentioned before, these can include anything that causes a decrease in delays such as stalled vehicles, spilled cargo, accidents, or construction. These can be distinguished from recurring congestion in the fact that they are random and do not occur at exactly the same time and place every day.

There has been some research done on recurrent congestion problems. Most of it has been focused on roadway geometrics and demand management systems such as ramp metering. On the other side of the coin is the nonrecurrent congestion, accounting for better than 60% of the total congestion, that is being addressed by incident management systems. This has been the focus of much of the advanced traffic management research in the forms of traffic diversion strategies, management strategies, signal re-timing plans, and driver information systems. A major decision playing factor in these traffic management tools is the prediction of the delays that are caused by particular incidents. A more accurate prediction of delays can ultimately lead to better implementation of the management strategies.

1.2. The Need for Real-Time Information

Effective incident management systems rely heavily on accurate real-time information on incident situations and traffic conditions. First, to detect incidents many surveillance systems include on-line real-time incident detection algorithms for loop detectors. These algorithms detect small changes in traffic flows, speeds, and densities along a freeway corridor that may indicate an incident. Often, these algorithms can detect an incident has occurred after only a few minutes. This method of incident detection has been proven to speed up the incident management process, and in effect reducing delays.

The second need of real time information is for users of traffic flow and incident information which include the responding agencies and drivers. Responding agencies may include the department of transportation, service patrols, state police, county police, local police, wrecker companies, utility companies, fire & rescue response, hazardous materials cleanup crews, and other environmental agencies. When certain incidents occur, specific agencies must be notified and provide a proper response. Again, real time information about incidents is often distributed by incident management systems to appropriate agencies so a response can be made quickly.

The other potential user of real-time information is the driver. Drivers usually receive information about incidents from either the DOT, through message signs and highway advisory radio, or from private organizations such as Metro Traffic that provides traffic reports to local radio stations. With real-time information, drivers can make decisions to change their trip to avoid the congestion, thus reducing the overall delays. Unfortunately, incident management systems do not always provide accurate real-time information to drivers. Information is often delayed between the point it is obtained at an incident scene to the point that the driver is aware of it. This delay of information dissemination, and inaccuracy leads may drivers to distrust the system. Streamlining the flow and increasing the accuracy of this information is the focus of many incident management systems today. (Wang, 1991)

1.3. Scope of this Research

The goal of this research is to model delays caused by freeway incidents. The perspective taken in this research is in the design and application of integrated real-time incident management systems. There are many facets of incident management system research including: duration estimation, delay estimation, diversion routing, response systems, signal retiming, and information dissemination. In a comprehensive system, these modules should work together to better manage the freeway system under incident conditions. This is illustrated by a conceptual incident management system in figure 1-1. This figure shows the flow of information from incident reports through duration and delay prediction and into the response module. The output of the response module is a combination of many possible actions to alleviate the delay caused by the incidents.

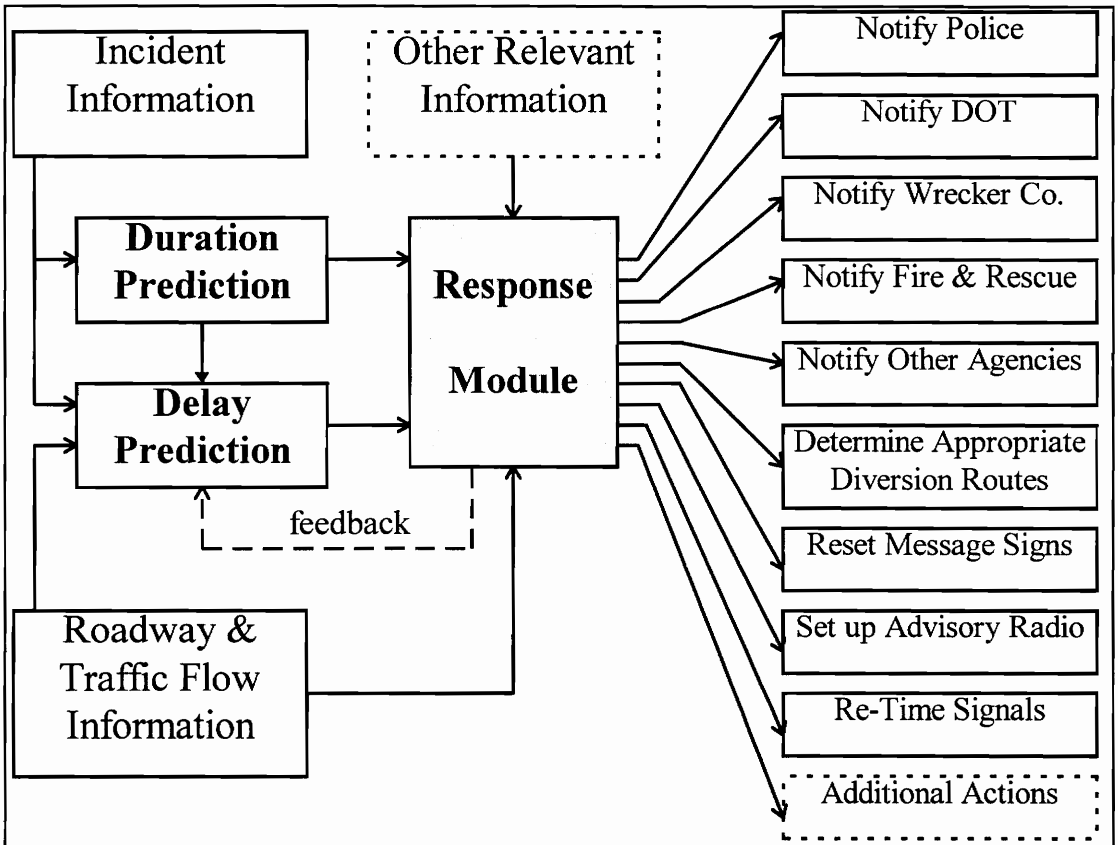


Figure 1-1 A Conceptual Incident Management Response System

Many important decisions regarding incident management and response can be made by a comprehensive incident management system. The response module at the center of the system can take a wealth of information about incidents, traffic flow, roadway geometrics, and others, and compile it into a form where quick decisions can be made about how to best handle the situation. Some important decisions made by an incident management system may be:

- Type and number of response vehicles needed
- Agencies to be notified
- Whether or not to divert traffic
- Where to divert traffic
- Whether or not to re-time signals
- What signal re-timing plan to use
- Messages relayed to drivers
- Extent of information relayed to driver (TMS, HAR, etc.)

These decisions must be made quickly and accurately if a comprehensive system is to succeed in making a significant difference in driver delays. If the response module can make these kind of decisions accurately, then the system would be very powerful.

In order for the response module to make accurate decisions in real time, it must have accurate real time information. This information would include data from a surveillance system on existing traffic conditions, historical traffic conditions, existing incident situations, and roadway geometrics. With the incident information, the response module needs to know details such as expected lengths of incidents and expected delays to be caused by incidents. (Subramaniam et. al., 1994)

The duration estimation module contains algorithms that predict the length of each incident based on the available incident information, while the delay estimation module predicts the extent of delays based on available incident information, the predicted duration, historical traffic flow variables, and real time traffic flow information. It is important to note that while much research has been done into the field of modeling incident duration (Khattak, 1994 and Wang, 1991), there has not been as much research in the field of real time incident delay prediction. There have been numerous papers written about incident duration prediction for real time incident management systems. One study is being conducted by the Center for Transportation Research at Virginia Tech that collected thousands of survey forms from individual incidents in the Northern Virginia area (Subramaniam, et. al., 1994). Results of the study include decision trees that provide incident duration predictions for various combinations of incident types, severity, and other factors. A system such as this can be used to provide a duration estimation to the delay prediction module.

The delay prediction framework proposed in this research will not be applicable as a stand alone unit for an incident management system, but will rather act as a module that works within Wide Area Incident Management Support System (WAIMSS) framework developed at the Center for Transportation Research which is illustrated in figure 1-1. As indicated above, a response module in a system needs to have accurate real-time information about delays that are caused by specific incidents. When an incident occurs on a freeway under guidance of an incident management system, it will be detected either from the surveillance system or from a driver's report. Once it is detected, all the available information about the incident is given to the system. At this point, the duration prediction module predicts how long the incident will take to clear. This prediction, along with real time traffic flow data, historical traffic data, and the incident characteristics, would be passed to the delay forecasting model. This model, as proposed in the research, will give a prediction to how much delay will be produced by the incident. These predictions are then passed to the response module, and then to an operator to make a decision on the response. This whole process as illustrated in figure 1-1, allows for the system to respond in real time, and ultimately leads to less delays because of better management and information.

The research in this paper is focused directly on the delay prediction section of the system. This specific section of the research is detailed in figure 1-2. The research is directed towards creating a procedure that can most accurately forecast future delays based upon incident characteristics, the expected duration, the historical traffic flow profiles, and present traffic flow conditions. The procedure needs to work in real time to provide a quick forecast to the response module.

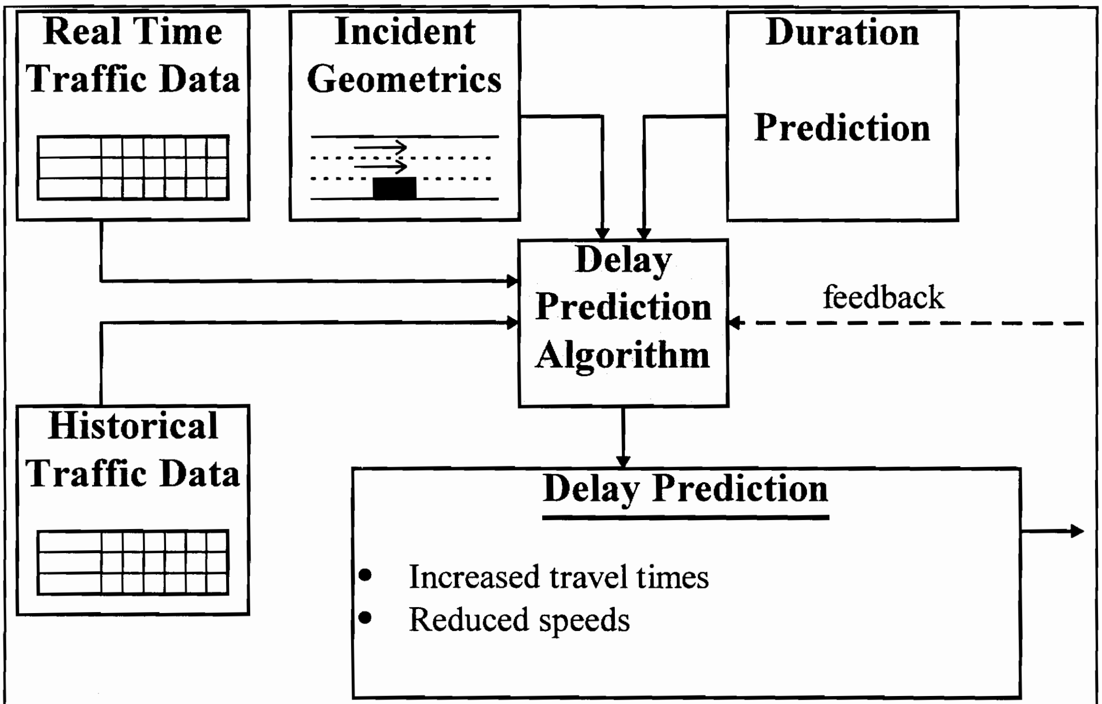


Figure 1-2 Incident Delay Prediction Model Detail

There are different levels of information that may be supplied to the system. For example, a first report of an incident may indicate the number of vehicles involved, but the majority of the other information is not yet available. When an officer arrives at the scene, he may be able to supply the other important information. In order to make best use of the system, there should be some preliminary response based on the limited information, in other words, the framework must be flexible and adapt to a myriad of situations. The goals of the framework developed by this research are summarized as follows:

<u>Goals of the Delay Prediction System</u>

- work in real-time
- work within WAIMSS architecture
- be flexible
    - account for recurring congestion
    - account for multiple incidents
    - account for changing demands
- realistically model changing conditions
- give useful and flexible results
    - effects on drivers (travel times)
    - predict levels of congestion

1.4. The Outline for this Thesis

This thesis concentrates on the design of a delay prediction procedure for on-line incident management use. The second chapter focuses on literature review of several delay estimation methods. It begins with a discussion of the differences between delay estimation and prediction. This chapter describes current methods used to estimate delay, and how they are applicable or not applicable for use in an incident management system. The chapter also analyzes possible traffic time-series forecasting methods, and the inclusion of those in a delay prediction algorithm.

Chapter three begins with a discussion of the available data sets used for analysis in this thesis. It discusses possible uses of the data and applications of the data for evaluating delay prediction strategies. A unique study describing dynamic in-lane incident removal is described and discussed.

Chapters four and five take the recommendations from chapters two and three to provide a basic methodology for constructing and evaluating a delay prediction procedure. Because of the shortcomings found in previous research, a complex framework involving a macroscopic simulation is developed. Data flow within the framework and outputs from the system are described both in text and accompanying flowcharts. Chapter five describes the macroscopic simulation portion of the algorithm in detail. A methodology for calibrating the parameters is described and discussed.

Chapters six and seven conclude the thesis with a discussion of the possible applications and implementation strategies. A procedure for testing, evaluating and calibrating the simulation is presented. Conclusions are given about the algorithm in general along with a discussion of the limitations. Finally, recommendations for further research are given.

2.Literature Review

This chapter begins with a discussion differentiating delay estimation and delay prediction and the problems with applying delay estimation models to delay prediction. The next section reviews methods currently used in delay estimation and their existing applications. The following section reviews models pertaining to time-series traffic forecasting. Finally, traffic simulation models and applications are reviewed and discussed.

2.1. Delay Estimation versus Delay Prediction

First, delay estimation, in the context of incident-related delays, is the process of assessing how much congestion, or traffic delay resulted due to an incident or incidents. This is done through a variety of estimation models. Based on information such as the time of the incident, duration, traffic demand, freeway supply, maximum queue, and travel times, these models give an output that is supposed to represent the actual delay caused by an incident. Studies are often done with incidents in the process of analyzing data to determine level of service on various freeways. The data is collected for many incidents and then run through a delay estimation model such as the shock-wave or queuing analysis models presented below. This is done to measure the effect of freeway improvements such as widening lanes, adding lanes, or implementing freeway service patrols.

The second term, delay prediction, may have similarities to delay estimation, but is notably different. The biggest difference is that estimation is done after all the data about an incident has been collected whereas delay prediction is done with some of the data about the incident missing. The idea is to predict information about an incident before it is over. In this sense, prediction has an entirely different purpose than delay estimation. *The purpose of prediction is to provide an accurate forecast of how traffic conditions will be affected by the incident.* Instead of using the maximum queue length as an input for estimating the delay, for example, the maximum queue length would be an output in a delay prediction algorithm.

The goal of the new procedure developed in this research will be to quickly, and with some degree of accuracy, predict the effects of an incident on traffic flow on a freeway. The inputs to the model will be the average traffic flows, speeds, and occupancies, for the space and time of the freeway it will be used on. Other inputs will be data about the incident such as the type, the lane blockage, the existing traffic variables, and an estimated duration. The estimated duration of the incident will be a major factor in determining the delay. The duration of an incident is predicted through another module in the on-line incident management system and will not be computed within the delay prediction framework.

It should be noted here that many models are used today, and have been used in the past to estimate incident delays on freeways. Because of the new surveillance and communication technologies, the field of incident management has emerged out of

Advanced Traffic Management Systems in ITS. These technologies make traffic data readily available and improved communications bring the advantages of on-line systems to life. An on-line incident management system can help reduce delays in a number of ways. Some of these are:

1. Relay information and traffic forecasts to travelers,
2. Make traffic forecasts for official uses such as setting up traffic diversions, and
3. Speed communications among the various agencies involved in incident management.

In effect, this on-line system will serve to disseminate information to drivers and officials, while speeding up the entire incident management process. It will also allow DOT's to accurately asses the impact of an incident. This will improve the efficiency of incident management and response strategies in real-time. It is hoped that improved information to travelers will discourage drivers from taking congested routes, thus reducing the overall congestion.

2.2. Delay Estimation Models

There have been several models that have been classically used in the past to estimate incident delays. The two major delay estimation models discussed in the following paragraphs include queuing theory and shock-wave analysis.

The first model is based on classical deterministic queuing theory. This model proposed by (Juan Morals, 1986) uses input and output curves on a graph depicting cumulative volume vs. time. This is an analytical procedure that, from the graphs, calculates the cumulative vehicle hours of delay (see figure 2-1). This method takes on a number of assumptions. The first of these being that the demand and capacity are assumed to be constant. In reality of course, these variables are not constant, and the curves are nonlinear. The second assumption is that the demand is initially less than capacity. This is not always the case especially in urban areas where many freeways are plagued by recurring congestion. This method also makes a number of other assumptions for freeway capacities and bottleneck capacities. It was found by (Al-Deek et. al., 1994), that this model is very sensitive to input parameters. There are different methods for calculating the slopes of the demand and bottleneck flow lines. There are also different scenarios to choose from. It was found that changing the input parameters for a single given incident and conditions, yielded significantly different results.

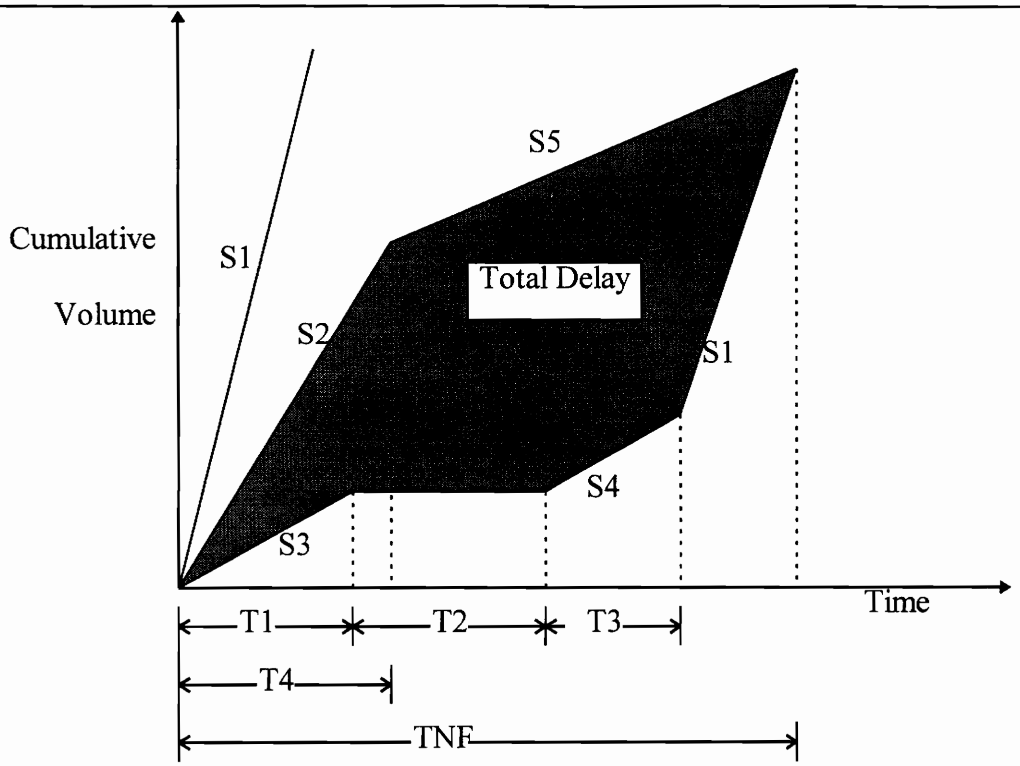


Figure 2-1 General Incident Queuing Diagram

The second model was proposed by (Chow, 1976). He actually proposed and compared two methods for calculating the total incident delay. He used one model with queuing analysis, and another with shock wave analysis. He assumed a unique flow density relationship. Using this assumption, he found that these two methods lead to the same results. He added that if a time-dependent flow-density relationship was used, that the two methods would have yielded different results.

Shock wave analysis was also looked in to detail recently in a new model proposed by (Al-Deek et al., 1994). It is designed to take advantage of detailed loop data for more accurate estimates of delay. For a given incident with a supposed starting and ending time, it uses shock-wave theory with the loop data to determine the time-space domain of the incident. This method specifies a search procedure using the shock wave theory and the loop data to determine the actual start and clear times of the incident. The purpose of the shock wave analysis was not to calculate the delay, but to determine what time-space domain is affected by the incident; namely what is the last detector that should be used in calculating the delay, and at what time should delays stop being calculated. The delays are calculated based on the actual loop data and relative to the historical mean speed profile. He also proposed a method for calculating the delay caused by multiple incidents within the same time-space boundary.

It was found that this method overestimates the incident congestion boundaries in the majority of the cases. It was mentioned that the assumption of linear shock waves was not valid, and that most shock waves are non-linear and dynamic. It can be shown, though that using linear shock waves over-estimates congestion because the actual shock waves will slow down over time due to smaller densities. Thus, the congestion envelope produced by the linear waves includes the actual envelope.

Another problem with this method of estimating delays is that it uses loop measurements taken at a single instance after an incident occurs. A single measurement of loop data may contain noise because often traffic is very unstable just after an incident. Multiple shock-waves formed from slowing vehicles could give an inaccurate representation of flows and speeds at a single instance.

These two models, queuing analysis and shock-wave theory, are the two primary methods currently used for estimating congestion. Both of these models are designed to estimate the delay caused by an incident based on all the collected data. Even with all of the data, it has been shown that neither model gives a precise calculation of delay. They are very sensitive to the input data and do not always give accurate results. This can be a serious problem in the application of the models to a delay prediction algorithm.

These models were not designed to be used on-line with limited information, but they have solid backgrounds, and some attempts have been made to modify them for on-line purposes. The next section discusses the possibility of modifying these traditional models for on-line applications. It discusses what parts of these models are applicable to on-line use, and the problems they present even in a modified format.

2.3. Modification of Delay Estimation Models for Delay Prediction

As shown in the previous discussion, there is still some confusion on how to estimate delay, given all the available data. There are several methods that give significantly different results for the same set of data. Determining the best method for estimating the congestion would be an interesting undertaking but is out of the scope of this thesis. The focus of this thesis will be on the design, formulation, testing, and verification of a procedure for incident delay prediction for use in the practical application of an on-line incident management system. The factor that caused the most confusion in the delay estimation procedure was the total delay (in veh-hrs). This is an important factor in determining levels of service and benefit cost ratios for projects, but is not the most useful output in an on-line incident management system. More useful parameters for an on-line system are items such as the maximum queue length, the reduced speeds, and the increased travel times.

First, the Morales procedure for analytically determining delays will be examined. Two of the outputs from the delay estimation procedure are the TNF, the total time until normal flow is resumed, and Q_{max}, the maximum queue length, could be outputs from the delay prediction algorithm. The TNF depends upon several factors:

- S_1, Capacity flow rate of the freeway, veh/hr
- S_2, Initial demand flow rate, veh/hr
- S_3, Initial bottleneck flow rate, veh/hr
- S_4, Adjusted bottleneck flow rate, veh/hr
- S_5, Revised demand flow rate, veh/hr
- T_1, Incident Duration until first change, min
- T_2, Duration of Total Closure, min

- T_3, Incident Duration under adjusted flow, min
- T_4, Elapsed time under initial demand, min

These factors are illustrated in the queuing diagram in figure 2-1. All of these parameters are usually known at the end of an incident, but many of them are unknown at the start and during the progress of an incident. Those that are unknown are S_4, S_5, T_2, T_3, and T_4. The variables S_4, T_2, and T_3 are related to changes in the lane blockage situation, and would have to be adjusted by another part of the system when that information about an incident becomes available. The other two variables, S_5, and T_4 are related to the change in demand either due to imposed or natural traffic diversions. During an incident with large delays, some of the traffic may avoid the congestion by taking other routes such as parallel arterials. This decrease in demand is difficult to measure, but would have to be added as a calibration factor based on historical and existing demands. This factor would then be calibrated with empirical data. Also, the getaway flow rate (queue discharge rate) in the Morales model is equivalent to the freeway capacity. This is not always the case, as indicated by (Lindley and Tignor, 1979) in their study on Getaway flow rates for Freeway Incidents. This is another factor that would have to be calibrated and adjusted with empirical data.

The second model to be considered for delay prediction will be the shock-wave model that has been used by many researchers including (Chow, 1976) and (Al-Deek et. al., 1994). In shock-wave theory, there are three flow regimes to be concerned with. The first flow regime is the normal condition. This is represented by the traffic conditions including flow (F), density (K), and speed (V) at a time just before the incident occurred. The next regime is the congested condition. This is the condition where capacity is reduced and traffic flow is metered by the incident. At this point, traffic is improved downstream because of the metering, but degraded upstream of the incident. This condition may be measured by the same variables (F, K, and V), as in the first regime. These conditions, however are taken just upstream from the incident, just after the incident occurs. The third flow regime is the recovery condition. Once the incident has been cleared and full capacity restored, traffic flow is restored, but moves slower and at higher densities. This is measured by the conditions just upstream of the incident, just after the incident has been cleared and full capacity restored.

The shock waves themselves are the actual boundaries between the three flow regimes. Although it has been shown that the straight line assumption for these tends to overestimate the congestion, it could be used as an initial approach in creating a shock-wave based delay prediction model. The speeds of the shock waves can be written as follows:

$$W_{12} = (F_1 - F_2)/(K_1 - K_2) \tag{2.1}$$
$$W_{23} = (F_2 - F_3)/(K_2 - K_3) \tag{2.2}$$
$$W_{31} = (F_3 - F_1)/(K_3 - K_1) \tag{2.3}$$

Where
W_{12} = the speed of the shock wave between normal and congested flow
W_{23} = the speed of the shock wave between congested and recovery flow
W_{31} = the speed of the shock wave between recovery and normal flow
F_1 & K_1 = the flow and density of the normal regime
F_2 & K_2 = the flow and density of the congested regime
F_3 & K_3 = the flow and density of the recovery regime

It can be shown that the total delay is:

$$D = T(W_{12})(W_{31}+W_{23})/(W_{31}(W_{23}+W_{12})) \quad (2.4)$$

and the maximum extent of the congestion is

$$X = T(W_{12})(W_{23})/(W_{23}-W_{12}) \quad (2.5)$$

where T is the incident duration

Of course only the absolute magnitudes of the shock-wave speeds are used to calculate D & X (The delay and maximum extent of congestion).

Shock-wave theory is generally considered to be a more realistic modeling technique for traffic flow than queuing theory. This method was used by Al-Deek et. al., for delay estimation purposes in the FSP project (Skabardonis et. al., 1996). It served as a general basis for determining congested regions. At first glance, it may look as though this method could easily be used for prediction purposes. Upon further examination though, there are several weaknesses of the model that make it ineffective for delay prediction. First, when it is used for estimation, data about conditions during all three regimes (normal, congested, and recovery) is known, so making an estimation only entails putting all the collected data into the model. When attempting to use it for on-line prediction however, conditions about all three regimes is not known, therefore these parameters would have to be estimated from point measurements. This increases the likelihood of errors, and in effect can cause a great deal of variance in the prediction process, leading to highly inaccurate results.

In addition to the missing data dilemma, this model with the three flow regimes (normal, congested, and recovery) is extremely over-simplified, and very sensitive to the input variables. It has been shown that this model greatly overestimates the delay because of the linear shock-wave assumption. It also loses accuracy with the assumption of the three flow regimes. In reality, there are many shock waves that form when the demand exceeds capacity. In congested conditions, traffic flow represents a sinusoidal pattern. Speeds are often unsteady, and move up and down through a range within the congested portion, which creates the sinusoidal effect. Each lane may have a different congestion period creating even more inaccuracy. In addition, within the area surrounding the incident many cars are rubbernecking and trying to merge creating additional mini shock waves locally. Therefore, the assumption for the single congested 'regime' is incorrect, and in reality there are many sub-regimes within the congested region. Traffic flow is extremely unstable near the incident and this can cause erroneous measurements.

Attempting to measure the change in flows and densities for a particular region upstream of an incident area will give widely varying results, depending on the point the congested traffic waves are measured at. Differences of flows, speeds, and occupancies may vary greatly between lanes depending on the time and point where they are measured. For these reasons, the simple three regime shock-wave model cannot be used as an on-line delay prediction model.

Delay estimation models are one way of approaching the problem of predicting delays due to incidents. As discussed above, delay estimation has been previously studied in detail to measure differences of delays caused by incidents. As shown through the previous discussion on differences between delay estimation and delay prediction, there is no simple way to directly apply these existing models to prediction of delays on-line and in real time. For this reason, other research methods into delay prediction were looked at.

2.4. Time-Series Traffic Forecasting Models

Some research has shown that time-series forecasting is a way to view the problem of predicting traffic delays. Time-series analysis has traditionally been a powerful tool for analysis of time-based data such as inventory management, production planning, financial planning, staff scheduling, facilities planning, and process control (Montgomery, et. al., 1990). There have been many forecasting models constructed that have shown the capability of predicting traffic flow variables. Most of these forecasting models are based on statistical approaches, heuristic approaches, or mathematical time-series approaches (Chang, 1994).

2.4.1. Heuristic Approaches

The first heuristic approaches were tested and reviewed by (Kreer, 1975). These algorithms are known as the Urban Traffic Control System (UTCS) predictor algorithms. There were two distinct models tested: the UTCS II and the UTCS III. The first model, UTCS II made use of smoothed historical traffic volumes, and current traffic volumes. It made a prediction by weighting the two parameters. If the prediction is close in time to the latest measurement, then the present data received a greater weight. If, however, the prediction was farther in the future, then the smoothed historical data received more weight. The second model, UTCS III, omitted the historical data and used only exponentially smoothed present vehicle flow measurements. The prediction was based on a weighted average between the smoothed and unsmoothed measurements. Kreer tested the algorithms with several data sets and determined that UTCS II performed better than UTCS III. This shows that the historical information is valuable and makes a positive impact on the accuracy of the forecast.

Another heuristic approach was proposed by (Stephanedes et al., 1981). They proposed that their new algorithm had a number of advantages over the UTCS II model. Namely, it did not rely on historical data, and overcame some time-lagging problems. This algorithm predicted volume during a time period t+1 using the current volume, the difference between the current and previous volume, and the average volume during the previous three, four, or five time periods. This algorithm was compared with the UTCS models, and it was found that it performed better than UTCS III in all cases, but did not always perform better than UTCS II. Some disadvantages of the model include the fact that it needs to be updated frequently, and it performed best when used during a particular time of day. For this reason, several equations were needed for each period.

Notwithstanding these problems, it did show an advantage over existing prediction algorithms, and it had the advantage over UTCS II, not requiring historical data.

The third heuristic model reviewed in this research was proposed by (Garbard et al., 1986). This algorithm was used for updating traffic control plans in urban areas. It makes use of both historical and real-time data, to quickly produce updated on-line traffic plans. If the flow was very similar to the historic average, more weight would be put on the historic average. If the flow in the immediate past varied widely from the historical average, then the model weighted the most recent data more. Coefficients are fitted on-line in the system to deal effectively with traffic perturbations.

It should be noted that this algorithm was designed for on-line use in an urban arterial traffic control system. It was coupled with a traffic signal optimizer to change signal plans and reduce delays at intersections in the network. Data gathered by occupancy detectors to detect queues was used by the plan optimizer to change signal timings so queues could discharge faster. It was shown that this model, coupled with the plan optimizer, can have significant results in reducing overall delays on urban arterial street networks.

All of these models discussed present ways for predicting future traffic variables based on combinations of past variables and historical variables. These models are considered because they have been used in the past for predicting traffic flows and speeds. One of the goals of an on-line delay prediction system is to predict similar variables for the use of making delay and travel time predictions. Unfortunately all of these models rely strictly on traffic variable measurements and they cannot incorporate complexities of freeway traffic flow such as geometrics or incidents. Obviously since they cannot incorporate these directly they will not be used as delay prediction algorithms. Where they may play a role though, is accurately predicting future traffic flow variables on simple closed sections such as ramps. These flow predictions for ramps could eventually be used as inputs to a more complex algorithm for the freeway mainline.

2.4.2. Univariate ARIMA models

Another method of using time-series analysis was proposed by (Pankratz, 1983) using Univariate Box-Jenkins Models (UBJ). These models are considered more scientific because they rely on classical probability theory and mathematical statistics, rather than intuitive or heuristic assumptions. The UBJ type of model used for traffic forecasting is the Auto Regressive Integrated Moving Average model (ARIMA). This is not however a single model but a family of models. It has been shown by (Box and Jenkins, 1976) that certain ARIMA models may be more appropriate than others depending on the fluctuations in the data. It has been shown that a properly selected ARIMA model can produce optimal forecasts. The idea behind the ARIMA model is to determine the correlation about Z at the time t base on earlier time periods (Z_{t-1}, Z_{t-2}, etc.).

Figure 2-2 illustrates how the modeling procedure is done. There are 80 time-sequenced observations in the past designated on the left side of the graph. After the 80th observation, an ARIMA model is constructed to determine patterns and correlations in the variable. The two forecasts on the right side 81 and 82, and derived from the model.

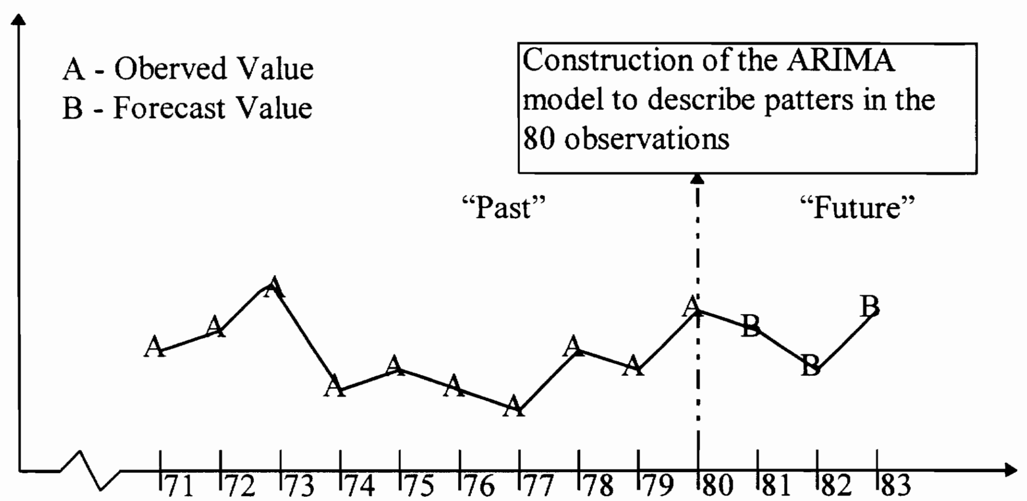


Figure 2-2 The idea of UBJ Forecasting (Box and Jenkins, 1976)

The UBJ family of models can generally handle a wider variety of conditions due to the simplicity of the model structure. The models are built using basic guidelines and comprehensive statistical analysis. This analysis done before the model is built allows the modeller to select a specific model from the ARIMA family that best suits the situation. An important fact to realize about the ARIMA models is that they can usually provide accurate multi-step ahead forecasting that is often needed for traffic management systems. (Pankratz, 1983)

As with the heuristic models, this type of model also has the drawback in that it only considers previous traffic flow variables to predict future variables. It too, cannot account for geometrics and complexities on freeways such as incidents. For these same reasons, it cannot be used as a model in and of itself to predict delays caused by incidents. It may play a significant role however in predicting flow variables on simple sections such as ramps that may be inputs to a larger and more complex model for the freeway mainline. Because it is more scientific than heuristic it may give more accurate forecasts in some situations.

2.5. Simulation

Traffic flow theory has been used for many areas of modeling traffic systems. There are two major types of models: static models that represent steady-state flow conditions, and dynamic models that describe unstable and transitionary areas of traffic flow as well. These models are based on sets of equations derived from established traffic flow theory. Another way to model traffic flow phenomena is through simulation. Although dynamic models are good for modeling specific subsystems of freeway facilities, they can not always account for the details included in complex interconnected networks. Traffic flow over a large freeway network cannot be modeled effectively by a few sets of equations. Traffic flow is often instationary and unstable, so simulation lends itself as a powerful candidate for modeling such systems.

There are two distinct categories of simulation that can be used for modeling freeway traffic flow. Microscopic simulation models individual vehicle's movements and following characteristics. In large microscopic simulations, hundreds of vehicles are individually tracked and modeled through a freeway network. This level of detail may make a model extremely slow and inapplicable to on-line situations. Macroscopic simulation on the other hand, models aggregate traffic variables over the links to define the traffic state.

2.5.1. Macroscopic Simulation

Macroscopic models are essentially dynamic traffic flow equations applied to sub-sections of a freeway network. These equations are interrelated to each other, and are computed in parallel for each time step. A time step is defined as the smallest unit of resolution that a simulation focuses on.

The most applicable model to traffic flow on freeways was propose by (Payne, 1973), and modified by (Cremer, 1976 and 1979). It was shown by (Cremer and Papageorgiou, 1981) that this model reproduces real traffic phenomena well including transitions from free flow to congestion and vice versa. It was also shown that when used in bottleneck situations such as lane drops, or in areas with high ramp volumes that the model behaves unrealistically. Several modifications were made to this model by (Cremer and May, 1986) that address and overcome these shortcomings. It was shown that extensions to this model enable it to model critical situations such as bottlenecks and high-volume ramp counts very well.

The basic model proposed by (Payne, 1973) was designed to model connected sub-sections of a freeway network over time; this is applicable for macroscopic simulation. Each section may have on or off ramps as well as a change in the number of lanes. It is suggested that each section be between 900 and 3000 feet and ramps should be at the ends of the section. An example of this situation is depicted in figure 2-3.

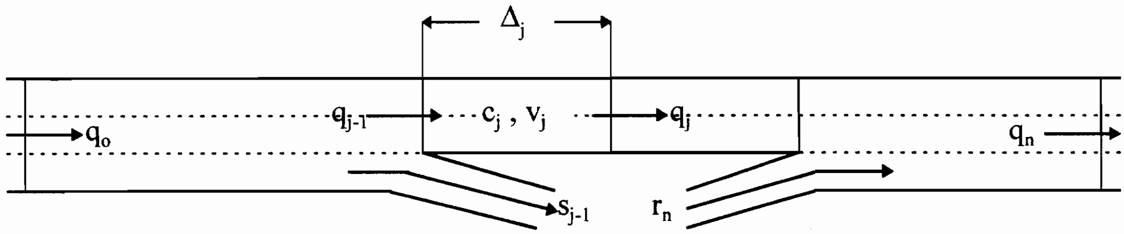


Figure 2-3 Freeway Section (Cremer and May, 1985)

This setup introduces the following variables:

$c_j(k)$	traffic density within subsection j at time k (veh/mi)
$v_j(k)$	mean speed of the vehicles within subsection j at time k (mi/h)
$q_j(k)$	traffic vol. leaving subsection j and entering subsection j+1during time interval $k < t < k+1$(veh/h)
$r_j(k)$	on-ramp volume entering subsection j during time interval $k < t < k+1$ (veh/h)
$s_j(k)$	off-ramp volume leaving subsection j during time interval $k < t < k+1$ (veh/h)
Δ_j	section distance (miles)
T	time step width (h)
l_j	number of lanes in section j

The density relationship can be obtained by balancing the entering and leaving vehicles of each section. The equation is:

$$c_j(k+1) = c_j(k) + \frac{T}{\Delta_j}\left[q_{j-1}(k) - q_j(k) + \left(r_j(k) - s_j(k)\right)\right] \tag{2.8}$$

The mean speed is determined by three major phenomena:

1. the dynamic adaptation of the mean speed to the stationary speed-density characteristic V(c) according to a time constant τ representing the driver's inertia
2. the convection of the speed gradient v_{j-1} - v_j in the downstream direction
3. the driver's anticipation of a density gradient c_{j+1} - c_j as seen in the downstream direction

With these three influences, an empirical equation was built to describe the shift in speed within a subsection j:

$$v_j(k+1) = v_j(k) + \frac{T}{t}\left[V(c_j) - v_j\right]_{(k)} + \frac{T}{\Delta_j}\left[v_j(v_{j-1} - v_j)\right]_{(k)} - \upsilon \cdot \frac{T}{\tau \cdot \Delta_j}\left[\frac{c_{j+1} \cdot \frac{l_j}{l_{j+1}} - c_j}{c_j + \kappa \cdot l_j}\right]_{(k)} \tag{2.9}$$

The density gradient term (the last term in the equation) contains the term l_j that is effective only if the number of lanes between sections j and j+1 is different. The reason for this is inherently because the driver reacts to the density per lane rather than the overall density. The overall sensitivity of this factor is weighted by the calibration factor υ. It should be noted that this is an empirical equation that requires calibration in order to perform accurately. (Cremer & May, 1985)

The stationary speed density characteristic selected for use in the first term of the equation was a general one derived by May and Keller (1967). This equation was originally derived from basic traffic flow theory.

$$V(c) = V_f \left[1 - \left(\frac{c}{c_m} \right)^a \right]^b \tag{2.10}$$

V_f freeflow speed
c_m maximum (jam) density
a, b calibration parameters

This is an empirical equation based on fundamental theory, so all the parameters (V_f, c_m, a, b) need to be calibrated to obtain reasonable results.

The flow transition between sections was described by Payne (1973) as an analogy to mass transport in hydrodynamics. The volumes from one subsection to the next are expressed as a product of speed and density:

$$q_j(k) = \alpha \cdot c_j(k) \cdot v_j(k) + (1-\alpha) \cdot c_{j+1}(k) \cdot v_{j+1}(k) \tag{2.11}$$

The weighting factor, α reflects that the flow exiting section j may be affected by traffic characteristics in both section j and section j+1. Combining the equations yields a modular structure where each sub-section j is defined by the state variables $c_j(k)$, and $v_j(k)$ for each time instance k. Equation (2.8) allows the on and off ramp flows to act as external forces, or inputs, to the system. This basic structure can be used to construct a custom and comprehensive simulation model for any size freeway by linking sections together appropriately.

This model was originally verified and validated with data sets from the Autobahn in Germany. With the data sets, a properly structured model, and an optimization routine, the calibration parameters (V_f, c_m, a, b, α, κ, υ, and τ) were realized. The two lane freeway section used for analysis was 2.7 km long with no bottlenecks or on ramps. The results of this verification proved that the model performed very well, very closely following the real data between transitions from free flow to congestion and back again. (Cremer & May, 1985)

When the model was tested with sections including lane drop bottlenecks, and high ramp entrance volumes, it gave unrealistic results. These problems were also noted by (Hauer and Hurdle, 1979), and (Babcock et al., 1982). Specifically the following major problems were described:

1. At lane-drop bottlenecks, congestion begins one to two kilometers downstream of the bottleneck and proceeds to move upstream over time
2. At the start of the simulation, over capacity flows are observed within the bottleneck region

3. After 30 minutes, once the congestion has moved upstream, the velocity gradient is significantly less that what is observed in the real data.

4. The volumes near the end of the simulation seem to be below the real data

Similar problems were also observed when high on-ramp flows merged with under high-traffic conditions. These problems were overcome by (Babcock et. al., 1982) when they used a scheme to sub-divide the sections when traffic gradients became critical, unfortunately this drastically increased the complexity of the algorithm and thus the computation time. Cremer and May however, devised extensions to the model that do not increase the complexity, but do overcome the shortcomings of the original model. The following extensions were added to the model:

1. limitations of the flows entering a bottleneck section to avert flow rates from going over the capacity: $q_j(k) \leq l_{j+1}*Cap$ where Cap is capacity per lane (1700-2000 veh/h/lane)

2. modification of the density gradient term in the section upstream of a bottleneck:

$$\Delta c = \beta\, c_{j+1}(k) - c_j(k) \qquad \beta > 0 \tag{2.12}$$

3. weakening of the convection term for the first section downstream of a bottleneck:

$$\gamma = \frac{T}{\Delta_j}\left[v_{j+1}(k)\cdot(v_j(k) - v_{j+1}(k))\right. \qquad \gamma < 1 \tag{2.13}$$

4. modification of the density gradient for a section upstream of an on-ramp:

$$\frac{\nu \cdot T}{\Delta_j \cdot \tau} \cdot \frac{\frac{l_j}{l_{j+1}}(c_{j+1}(k) + \delta \cdot \frac{T}{\Delta_j} \cdot r_{j+1}(k)) - c_j(k))}{c_j(k) + l_j \cdot \kappa} \tag{2.14}$$

5. modification of the density gradient for a section with an off-ramp:

$$\frac{\nu \cdot T}{\Delta_j \cdot \tau} \cdot \frac{\frac{l_j}{l_{j+1}}(c_{j+1}(k) - \delta \cdot \frac{T}{\Delta_j} \cdot s_j(k)) - c_j(k))}{c_j(k) + l_j \cdot \kappa} \tag{2.15}$$

Qualitative studies were done with a sample data set, and the new extensions proved to give more reasonable results. In order to get more quantitative results, the new parameters (γ, β, and δ) had to be calibrated together with the original parameters with real observed data. The data selected was taken from a 4 mile section of the eastbound Santa Monica Freeway for a three hour period between 2 p.m. and 5 p.m. from six mainline detector stations and from all the on- and off-ramps (13). The actual calibration and optimization methodology is detailed in (Cremer and May, 1985).

The results showed that the extended, calibrated model reflected most key freeway traffic phenomena: traffic flow transitions into and out of congestion, traffic flow through bottleneck sections, merging and high on-ramp volume sections, and changing lane geometrics. The model also uses a fixed time step width and spatial intervals throughout a simulation period. This keeps the computation time from growing as it does with some other modifications. This makes it more applicable to real-time applications that require quick computation times.

Re-calibration of this model with other data sets still needs to be performed in the future. It has been shown in previous research that this model originally proposed by Payne, and later modified by many other researchers including Cremer and May, is a the most comprehensive macroscopic freeway model and it can model a wide variety of situations. Other, more simplified models have also been proposed, but this model is the most comprehensive, thus the most applicable to this research of an on-line delay prediction system.

2.5.2. Microscopic Simulation

Microscopic simulation of freeway networks is also an area of interest to modeling freeway delays. This type of simulation models individual vehicle's movements and properties as they travel through a computer-coded network. There are a number of simulation tools designed by the Federal Highway Administration for use in microscopic simulation. Among these tools are INTRAS, NETSIM, FRESIM, and CORSIM. FRESIM is the successor to the early INTRAS simulation model and is the only one of these models to be studied in this research. NETSIM is actually designed to simulate urban arterial street networks. CORSIM is designed to simulate both freeway and urban arterial networks by combining both NETSIM and FRESIM. Unfortunately, time constraints allowed only the FRESIM package to be evaluated.

FRESIM is a comprehensive package that allows virtually any freeway geometrics to be simulated. It includes features such as:

- Variations in grade, curvature, and pavement along any stretch of the freeway
- Lane additions or drops at any point
- Incidents, and on-line incident detection
- A comprehensive lane-changing model

- Six different vehicle types, each with its own performance capability
- Ten different driver types ranging from timid to aggressive drivers
- Ramp-metering strategies

Just about any freeway situation can be modeled with this package (FRESIM User Guide, 1994).

Instead of tracking aggregate traffic variables such as flow, density, and speed, this simulation tracks properties of each vehicle such as vehicle type, driver type, speed, acceleration, origin, destination, and location. In theory, this type of modeling should give better results than a macroscopic simulation because it can completely model the situation down to the individual vehicles. Unfortunately, this also requires very detailed input on the exact location of lane drops, ramp entrances, driver demographics, vehicle demographics, roadway geometrics, and origin-destination parameters. This level of detailed input is valuable to have in a simulation package, but is difficult to obtain from real-world situations.

Some preliminary analysis was done with FRESIM. In general, it was found that it requires a great deal of calibration in terms of modeling incidents and congested traffic conditions. Setting up an incident in FRESIM is rather subjective and presents problems with modeling delays. Even after running the simulation many times and adjusting factors for a specific incident, the output of the simulation could not match the real-world data. A general conclusion from this case study is that the simulation can model an incident, but congestion formed by the incident does not clear out as quickly as in the real-world case.

This might be fixed with the adjustment of more of the input factors that go into modeling an incident in FRESIM.

Each incident situation is usually very different. With so many parameters to calibrate in the FRESIM simulation, it makes it very difficult to consider using it for on-line delay prediction purposes. Even if it could be calibrated for certain cases, the computation time is very slow for large networks. For a 3-mile section of a 5-lane freeway simulated for an hour with a 15-minute incident, the computation time took around twelve minutes on a Pentium-90 PC system. For an on-line system, the model used to represent a freeway system must be fast. For complex, interconnected urban networks with miles of freeway, the algorithm must be able to simulate the network for a long period and predict delays in less than one minute. While microscopic simulation may be detailed, it is too complicated and too slow to allow for fast on-line prediction of incident delays.

2.6. Selection of "Candiate Method"

The purpose of the literature review has been twofold. First, it was done to get a better grasp of the problem of delay prediction. Studying the various methods and models opens up the complexities and difficulties of modeling such a detailed phenomenon. This meticulous type of study gives a big picture of the research accomplished in the field as well as some direction for further research. The second purpose of the literature review was to determine what previous research is applicable, if any, to the design of an on-line system for the prediction delays on freeways.

As shown in the previous sections there are many problems that plague previous methods making them inapplicable for use in an on-line system. Some of these problems are: inability to deal with multiple incidents, inability to deal with recurring congestion, over-sensitivity, inaccuracies, slow speeds, and inflexibility of outputs. Only one method that was studied could overcome most of these limitations therefore it is chosen as a "candidate method" for continuing research and application in the new on-line delay prediction framework.

Macroscopic simulation has been chosen as this "candidate method" for an on-line incident delay prediction algorithm. It can incorporate many of the complexities of a freeway such as lane geometrics, changing flow conditions, ramps, and incidents. It can handle a sufficient enough amount of detail to accurately model almost any freeway situation. It is not so detailed though, that the speed of the model is a problem when calculating delays on-line as it does in the microscopic simulation. This method will be the basis for the new system and will work within the framework described in chapter four.

3. Data Collection

In order to properly examine and analyze any model, real world data is needed. For the study and analysis of freeway traffic delays due to incidents, data is needed on both freeway incidents and on traffic flow. For the development and testing of a macroscopic simulation for delay prediction, both kinds of data are needed for the same study area. Information about the incidents on the freeway is needed along with the corresponding traffic flow data for the time-space domain of the incidents. Many studies have been done that collect incident data for purposes of statistical analysis and record keeping. Many other studies have been done that collect traffic flow data for many other purposes. Rarely have there been studies that collect both incident and traffic flow data simultaneously. The first part of this chapter will discuss the combined database of the Freeway Service Patrol (F.S.P.) (Skardonis et. al., 1995).

A second type of data study was also done in this research. New research was performed in the area of dynamic lane clearance. Some previous studies have shown that incidents with multiple lane closures do not have their lanes cleared all at once. For example, a sixty minute, three-lane incident usually does not have all three lanes blocked for the entire duration of the incident. Incident management teams are very attentive to the fact that lanes should be opened to traffic as soon as safety conditions permit in order to reduce delays. The simple matter of opening one lane before the clearance is complete can make a tremendous reduction in the amount of delay experienced in the freeway corridor. No study has been done in the past to determine how multi-lane incidents are cleared dynamically. A unique survey and study was done to determine the behavior of this phenomena. This survey will be the focus of the second section of this chapter.

3.1. The Freeway Service Patrol Database

Before the database is described, the background of the F.S.P. project (Skabardonis et. al., 1995) should be discussed. The freeway service patrol on California freeway system was designed to reduce delays by actively surveilling the freeway with service patrol vehicles. They could inform the traffic management centers of problems before they got out of hand. In addition to that they helped to quickly clear minor incidents by quickly fixing small mechanical problems, adding gas, helping change a tire, or giving a motorist a push over to the shoulder, where less delay would be caused.

The project was created out of the interest of implementing additional freeway service patrols on other California Freeways. Some previous benefit to cost ratios determined from FSP projects has been from 2 to 1 to 36 to 1 (Skabardonis, 1996). The FSP program in California covered 600 miles of freeways and cost an estimated $28 million that was funded through federal, state, and local monies. There was a need to assess the effectiveness of the program in reducing freeway delays and adverse impacts of incidents.

A six mile section of Interstate I-880 between San Jose and Oakland was selected because this section had extensive surveillance in the way of loop stations. Loop stations included a set of double loops for each mainline, and a single loop for each on or off ramp. The loop stations were located approximately every 1/3 mile along this stretch. See figure 3-1 for more details. For the study periods in the morning and afternoon, each set of loops gave updated conditions every 30 seconds. These conditions were converted into one-minute averages of flow, speeds, and occupancies. These extensive datafiles can be used for very precise modeling.

During the study period, four probe vehicles were driven back and forth along the freeway at approximately 8 minute headways. They were equipped with on-board computers that stored information about their progress. The first thing that the computer did was to record key-presses that the driver made. Each time the car passed a fixed point along the freeway, or an incident, the driver pressed the appropriate key. In addition, whenever a driver got close to an incident he radioed the message to the command center.

The incident section of the database has over 50 fields of information. This includes the time, date, location, lanes blocked, clearance time, weather, and a host of other variables affecting the incidents. Although there were over 1000 incidents reported in the database it was found that less than 100 of them blocked lanes for a significant amount of time. Nonetheless, it is still a significant amount of data and is extremely valuable when combined with the detailed loop detector data.

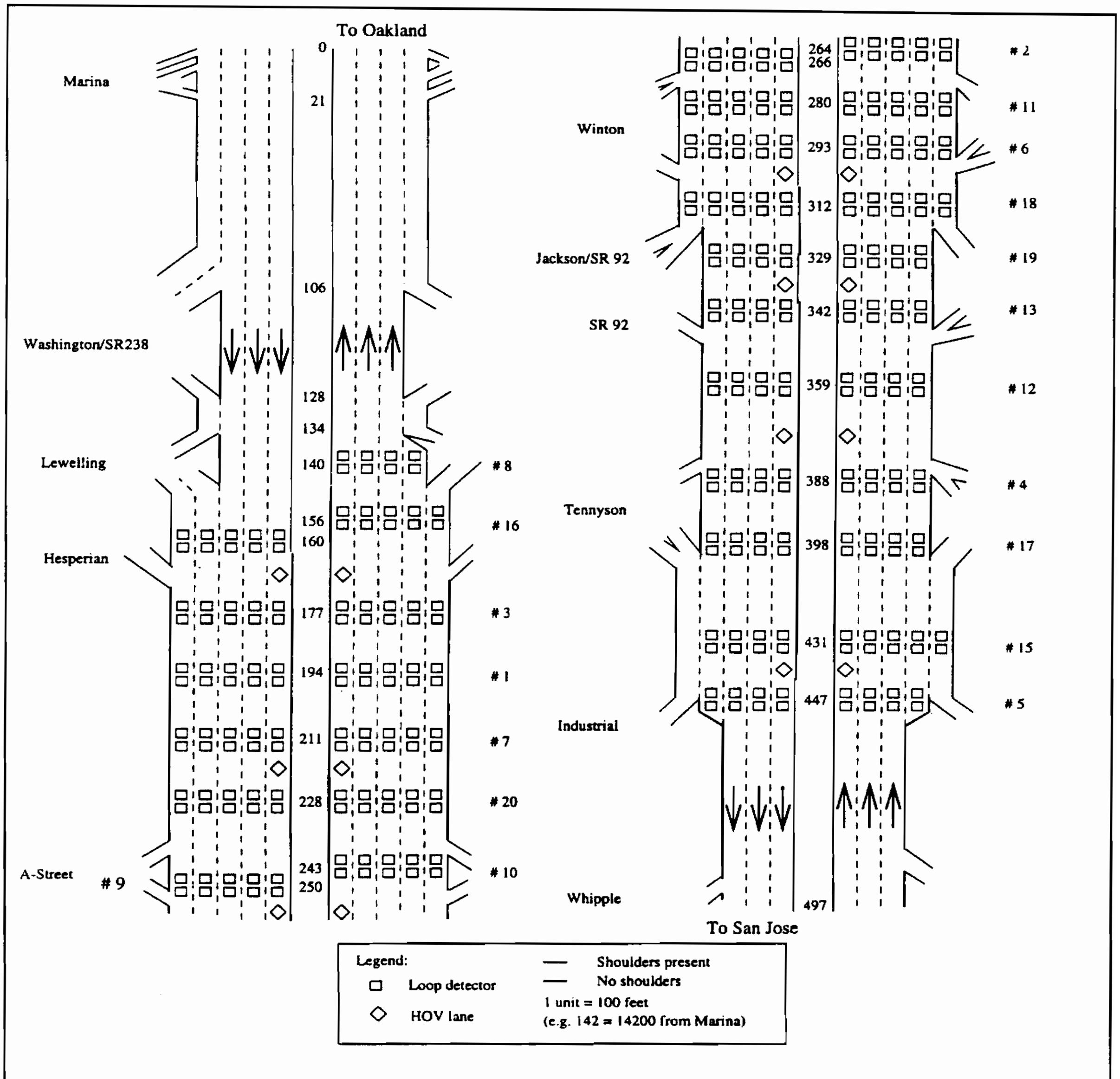


Figure 3-1 Interstate 880 Test Section

3.2. Dynamic Lane Clearance Survey

3.2.1. Description

As mentioned in the introduction to the chapter, incidents that block multiple lanes are not always cleared in one sweep. Lane blocking incidents, as opposed to shoulder incidents, are usually cleared in phases. Many times, vehicles that are blocking lanes can be moved either by their own power or by a push or tow. Modern incident management plans are focused on restoring normal traffic flow as quickly as possible. One directive in some plans is to open as many lanes as possible during the clearance to reduce delays. (Cambridge Schematics, 1990). This removal of vehicles and re-opening of lanes during an incident usually provides a significant decrease in delays. An incident that blocks two out of four lanes for sixty minutes will produce significantly more delay than an incident that blocks two lanes for ten minutes, then just the shoulder for fifty minutes. This phenomena has not been studied in the past, but it will have significant effects on any delay prediction. For this reason, a study was undertaken to determine how in-lane incidents are dynamically cleared.

The best way to study this phenomenon would be to have a time-based incident study that records the exact time it takes to clear each lane for a large sample of various incidents. Unfortunately it was beyond the scope of this project to conduct such a lengthy and detailed study. For this reason, a survey was conducted not of individual incidents themselves, but rather of incident management personnel.

A survey was presented to over 50 people involved in incident management in the Northern Virginia region. A sample portion of this survey is depicted in appendix A. The survey was divided up into several divisions. First, there was a division for each major incident type: property damage, personal injury (minor), personal injury (severe), truck, and cargo spill. Secondly each type was presented for various durations. For each one of these types a situation was presented in the form of: (#lanes blocked/# lanes in section). The situations considered include: (4/4), (3/4), (2/4), (1/4), (3/3), (2/3), (1/3). The situation of one lane blocked was looked at because a car blocking one lane can be pushed to the shoulder frequently, thus reducing delays considerably. Because the majority of the freeways in Northern Virginia are three and four lanes, these were the only ones considered in this survey.

3.2.2. Results of the Survey

The survey was copied and distributed to many incident clearance personnel including 75 state and local police officers. The information from the surveys was compiled and summarized in a spreadsheet. These summaries are illustrated in a set of tables in appendix B. They are first subdivided by duration and incident type. The tables are then divided by number of lanes available and number of lanes initially blocked. The bold number under each lane indicates the average lane clearance time response given. The number just beneath the average is the standard deviation among all the surveys.

It seems that dynamic lane clearance does indeed occur in the minds of Northern Virginia incident management personnel. On average it is the opinion that incidents that initially block lanes only block them for 70% of the total duration. This means on average, an incident that takes a total of 100 minutes to clear will spend only 70 of those minutes blocking travel lanes.

This survey gives a general idea of how much time lane blocking incidents actually spend blocking travel lanes over the course of their duration. The survey, however limited, produced an enormous amount of data, and some detailed statistical analysis could be done with it to make some more general conclusions about dynamic lane clearance. Due to constraints, this kind of rich analysis could not be done in this research. This would be an interesting undertaking in future research.

3.3. Applications of the Data for Algorithm Development

The F.S.P. data has many applications for the validation and verification of traffic flow theory. Specifically it can be used for the testing and evaluating macroscopic simulation program. The macroscopic simulation model will use the data of the freeway corridor input, and the ramp flows to simulate changes in the main line flow segments. Compiled averages of the data can be used as historical data inputs to the model. Output of main line flows and speeds can be compared to those of the actual collected data in the F.S.P. database. Because the simulation model has many adjustable parameters, they can be calibrated to "tune" the simulation output closer to that of the real data. The closeness of the calibrated output to the real data will give a sense of how valid the model is. This kind of analysis was originally done by (Cremer & May, 1985) with data from the Santa Montica Freeway. Their data used coarse resolutions of five and even fifteen minute averages. The FSP dataset on the other hand, uses much finer one minute averages. It is expected that this will increase the accuracy of the survey.

The dynamic lane clearance information can be used in combination with capacity reduction tables to set up incident scenarios in the simulation where capacity may be reduced in certain segments for certain lengths of time. Instead of blocking three lanes in a simulation for the predicted duration of the incident, each lane will have its own clearance time making the simulation more realistic, and hopefully more accurate. This process is discussed in detail later in this research.

4. A New Delay Prediction Framework

The next two chapters will describe the proposed new delay prediction structure in detail. It was determined that real-time simulation was desired to be used in order to achieve the most accurate results. Because a simulation model is so data intensive it needs a complex and detailed support architecture. This chapter will begin with a basic conclusion of the literature review. A description of the general architecture will follow along with detailed descriptions of the various parts. Data flow paths will be illustrated and described along with the input and output procedures. The following chapter will describe the simulation itself along with a method that can be used for calibration of the simulation parameters.

4.1. Methodology

As mentioned in the literature review, there has been no single model that can accurately predict delays caused by freeway incidents. Out of all the models studied, only the macroscopic simulation model proved to give accurate enough information to give traffic flow predictions under many situations. Simulation itself cannot predict the delay caused by an incident, but it can be a powerful tool in traffic forecasting.

Traditional models (queuing theory, shock wave theory) have focused on the delay caused by a specific incident for a specific, static situation. The shortcomings of these models have shown that they do not represent real world situations because freeways are dynamic, not static. Many factors come into play on freeways including:

recurring congestion, non-recurring congestion (incidents), on-ramp flows, off-ramp flows, and geometrics. There is no way to incorporate all of these into a single, or even set of equations and look-up tables and expect to receive a reliable forecast of delay caused by an incident. This is such a dynamic situation with so many influences that it lends itself towards simulation. Simulation outputs (traffic forecasts) can then be easily used to predict delays and impacts on freeways.

An area of debate in previous research is about the appropriate output of a delay prediction system. It may be interesting to have an estimate of how much delay (in veh-hrs) that will be caused by a specific incident, but it is not practical. In order for an incident management response system to make a reasonable decision on actions to be taken (ramp metering, diversion), it needs to know the effects of an incident on the freeway as a whole. Knowing an incident will produce X vehicle hours of delay is not enough information on which to base a response decision. However, knowing that travel time between segments *A* and *B* will take *Y* minutes more because of an incident will be useful information on which to make a response decision.

Although the original intent of building a delay prediction algorithm was to predict specific incident delays, it is more important to look at the freeway as a whole in this situation. For this reason, the framework will not focus on incident specific delay predictions, but rather, more flexible freeway traffic forecasts incorporating incident and lane blockage information. This also allows the system to make forecasts for other situations. No longer will it be incident-specific, but rather it will be able to model a wide range of situations in effect giving more accurate forecasts. The advantages of this approach are innumerable. This new framework accounts for a wide range of scenarios that would be difficult or impossible with other methods. Some of these are:

- An incident occurs that closes an off-ramp
- An incident in a construction zone
- Incidents in bottleneck areas
- An incident just downstream of a major on-ramp
- Multiple incidents with combined queues
- An incident in a freeway section near a stadium just after a game lets out
- Construction activity blocking lanes during a special event

Outputs that can in-turn be used for response decisions that are not possible with other models are:

- Increased travel times between segments
- Future speeds on segments
- What flow conditions will be in segment X in an hour
- The time at which speeds will rise above 50mph on segment B

These are just some of the advantages of having a traffic flow prediction model as opposed to an incident-specific delay prediction model. The flexibility of this system makes it ideal for use in a traffic management center where quick forecasts are needed for critical responses to freeway emergencies.

4.2. Algorithm Structure

The delay forecasting framework is presented in the form of flow charts and will be run on a fast computer connected with historical traffic data, real-time surveillance equipment, and will work in conjunction with the original system proposed in chapter 1 and illustrated in figure 1-1. Pieces of information are illustrated in the flowcharts as blocks and information flow is indicated with arrows. The algorithm is very detailed, thus it is difficult to illustrate the entire algorithm in one flow chart. It is divided up into many flowcharts, each going into greater detail of sub-modules. Some blocks have a thick grey outline indicating that they are sub-modules. Each of these sub-modules has its own flowchart showing the flow of information in that section. The description of the system will start with the overall system, then go into detail about each of the submodules. Figure 4-1 shows the overall new delay prediction framework.

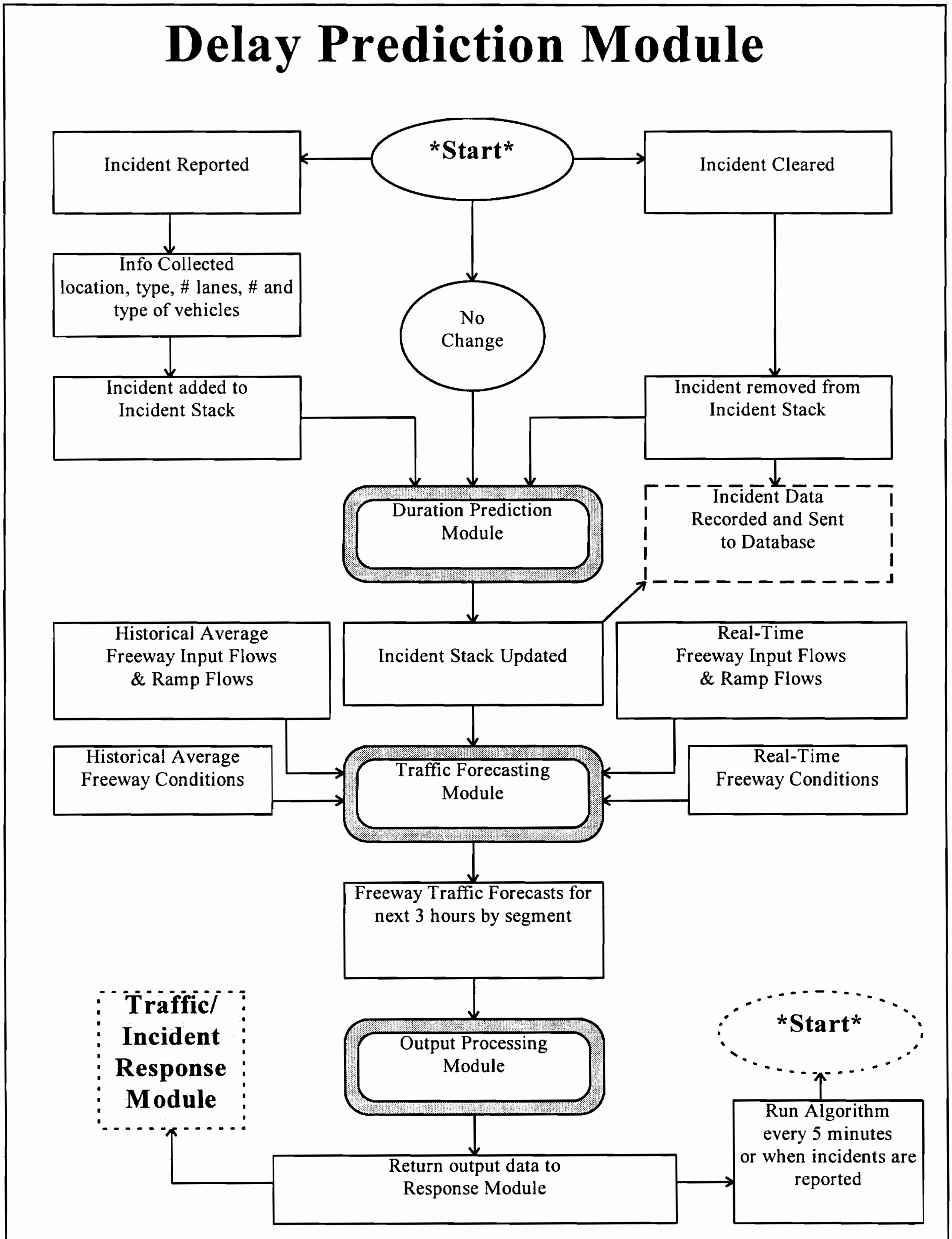


Figure 4-1 Delay Forecasting Framework

While working in the real-time setting of a traffic management center, the freeway response module would be called every time an incident occurred or every five minutes on average to update the forecast. When the algorithm is run the following steps would occur:

Step 1

Information about the incident received from surveillance equipment and entered into the system. This information is put into a database called the incident stack, that holds information about all the incidents active on the freeway at time t+0 (the time when the simulation is run). If an incident has been cleared, the incident is removed from the incident stack and the information about the clearance including the time to clear each lane, and the total clearance time is sent to the cleared incident database. This long-term database can be used to do further research on incident clearance timings.

Step 2

The updated incident stack is then sent to the duration prediction module. This will update the field in the incident stack database pertaining to the duration of each incident.

Step 3

This incident stack along with a plethora of real-time and historic traffic flow information from the freeway corridor is sent to the heart of the algorithm: the Traffic Forecasting Module. This will be discussed in detail in the next section.

Step 4

The output of the traffic forecasting module is the predicted flows and speeds for each segment in the corridor for the next three hours (t+1 to t+180). This information is then sent to the data processing module that uses the data to make travel time predictions between each segments and contour plots to illustrate predicted future congested areas on the freeway. This data processing module will be the last module discussed in this chapter.

4.3. Traffic Forecasting Module

The traffic forecasting module is the heart of the hybrid algorithm discussed in this paper. It takes in many pieces of information including freeway conditions at time t+0 (historical, real-time, and predicted), the incident stack, and freeway corridor inputs and outputs (historical and real-time). All of this information is sent to three sub-modules that generate a total of six arrays that are used as inputs to the macroscopic simulation. The simulation then gives two output arrays: C(s,m), and V(s,m). These are the density and speed respectively for each segment on the freeway from the time t+1 until t+180. Essentially, it is the predicted speed and density for each segment for three hours into the future. This information is returned back to the main algorithm and it is sent to the data processing module. This entire process is illustrated in figure 4-2.

4.3.1. Initial Conditions Module

One of the inputs for the simulation model as discussed in the literature review, is the conditions of each segment at time t+0. In order to do this, historical, predicted, and real-time conditions are weighted for each segment and an estimate of the initial conditions of the freeway is put into an array. This process is depicted in figure 4-3. The weights are based on the type of information that is available. If real time data is available, then more weight will be given to that. If only historical, and previously predicted information is available, then they are weighted appropriately.

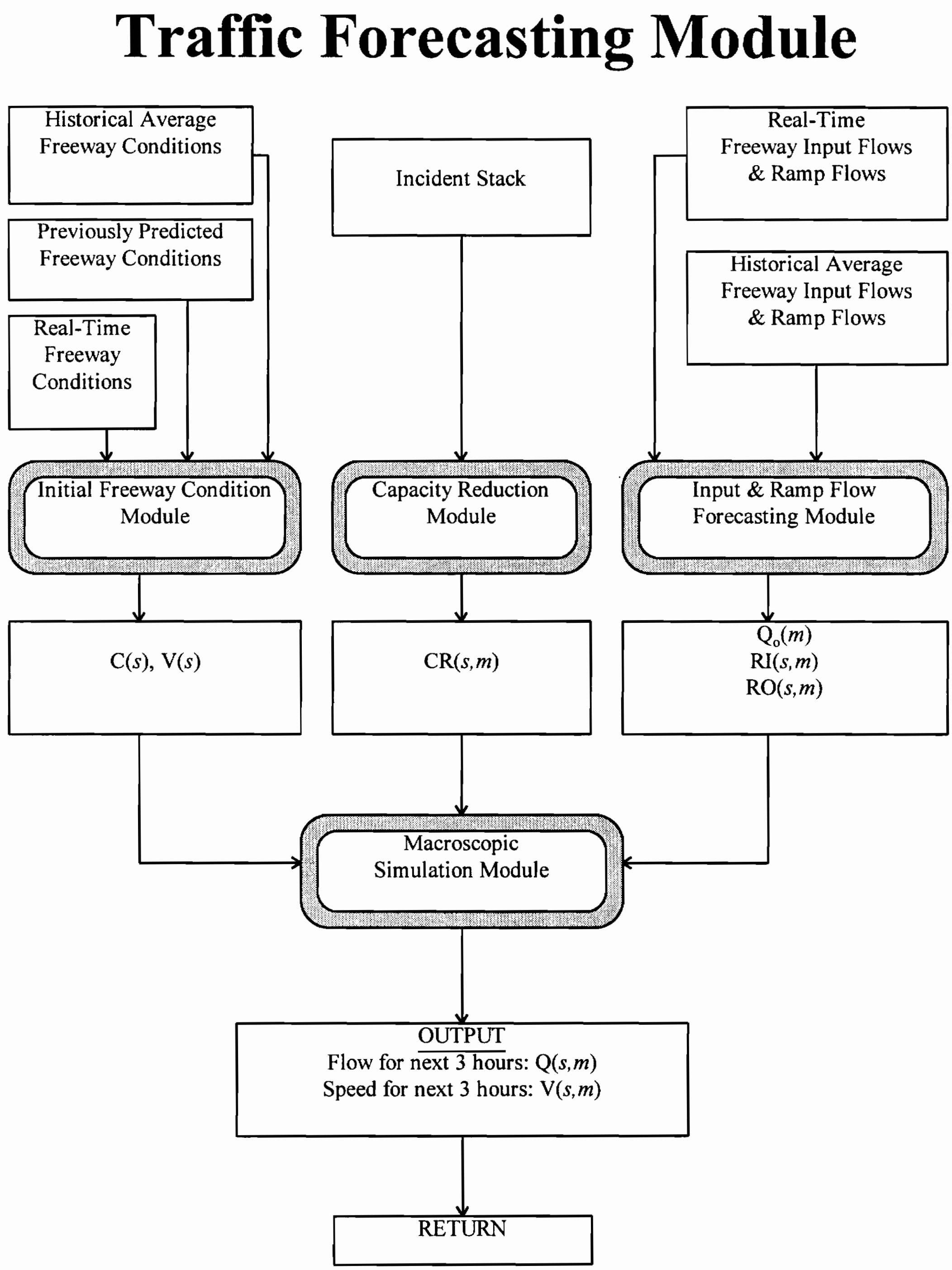


Figure 4-2 Traffic Forecasting Module

4.3.2. Capacity Reduction Module

This module is the portion of the framework that reduces the capacity of each segment based on the incidents that affect them. When lanes on a freeway are blocked, the capacity is severely reduced. This reduction in capacity is the combination of both the physical blockage of lanes and of rubbernecking. This rubbernecking phenomenon occurs when drivers slow down to get a better view of the incident scene. (FRESIM User's Guide, 1994) For this reason, there is a reduction in capacity even for incidents just blocking the shoulder. As additional lanes are blocked and the severity of an incident increases, the delay grows geometrically. This module builds an appropriate scenario that represents the reduction in capacity due to an incident on a freeway over space and time.

The outline of the capacity reduction module is illustrated in figures 4-4 and 4-5. The array created by this module is noted CR(s,m). This is the expected capacity reduction of each segment of the freeway for the time t+1 to t+180. Essentially it is the percent capacity is reduced for each segment for three hours into the future. When this section runs, this array is initialized to all ones. This essentially says that each segment will be running at full capacity for the next three hours.

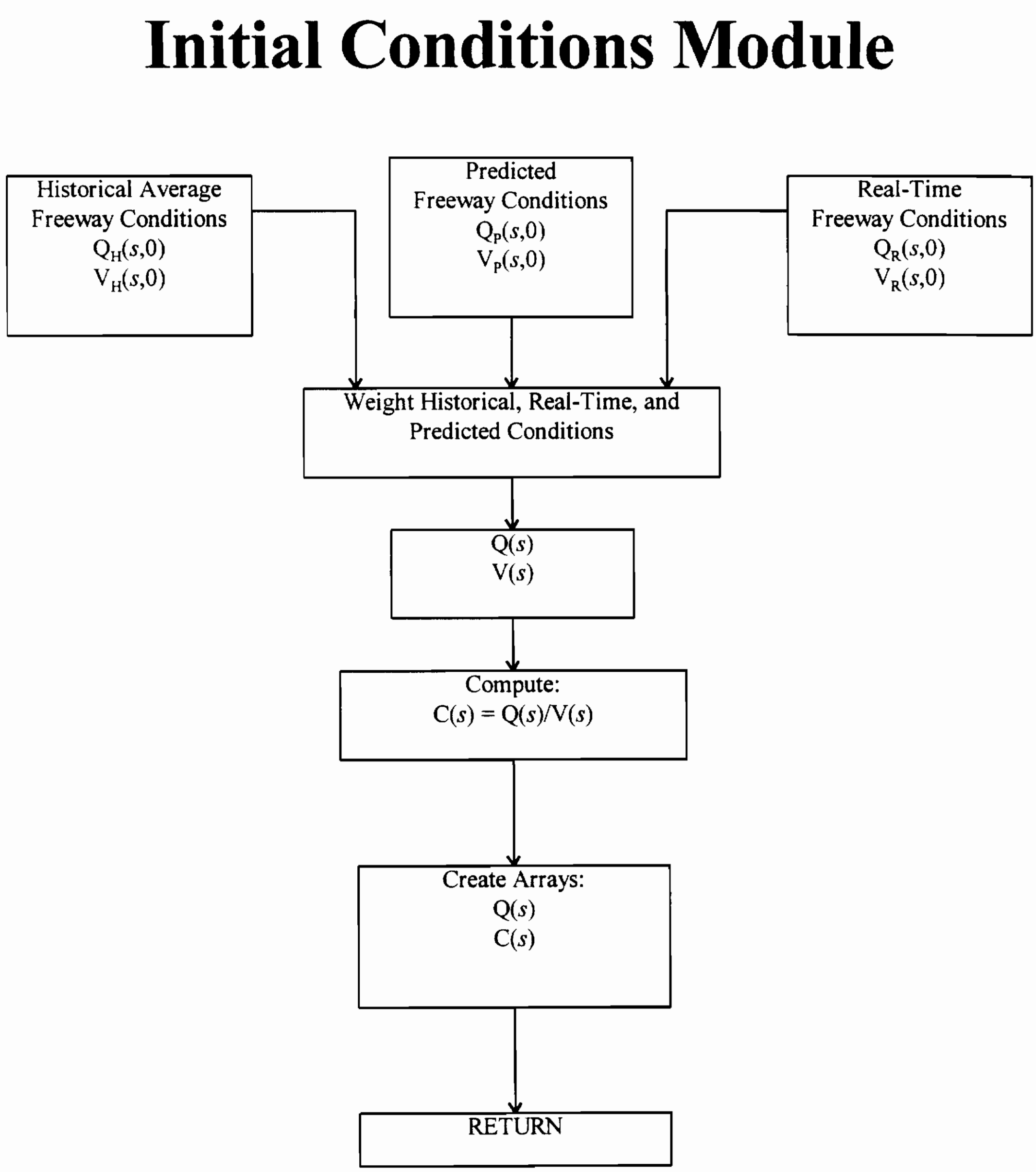


Figure 4-3 Initial Conditions Module

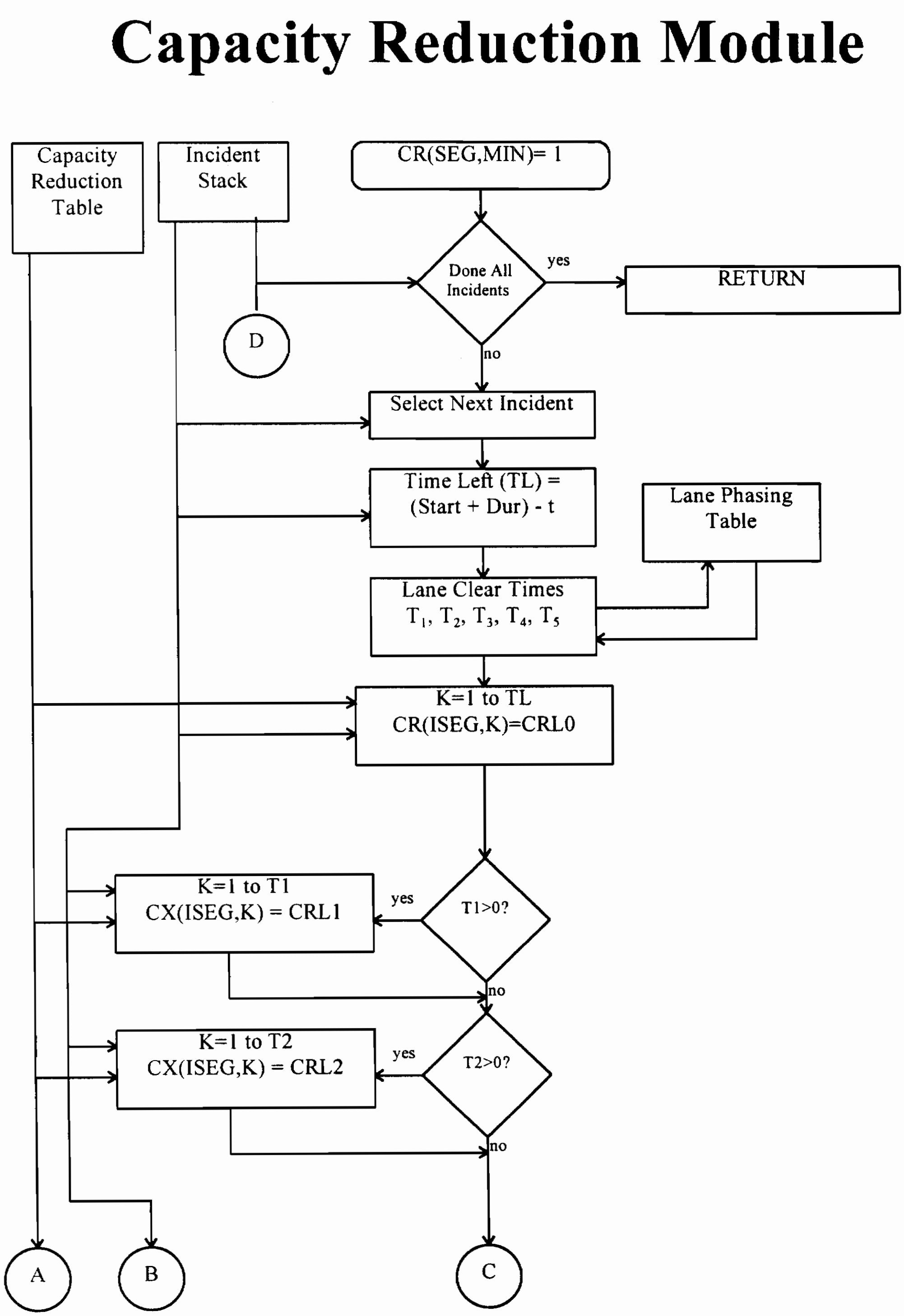


Figure 4-4 Capacity Reduction Module

Capacity Reduction Module, cont.

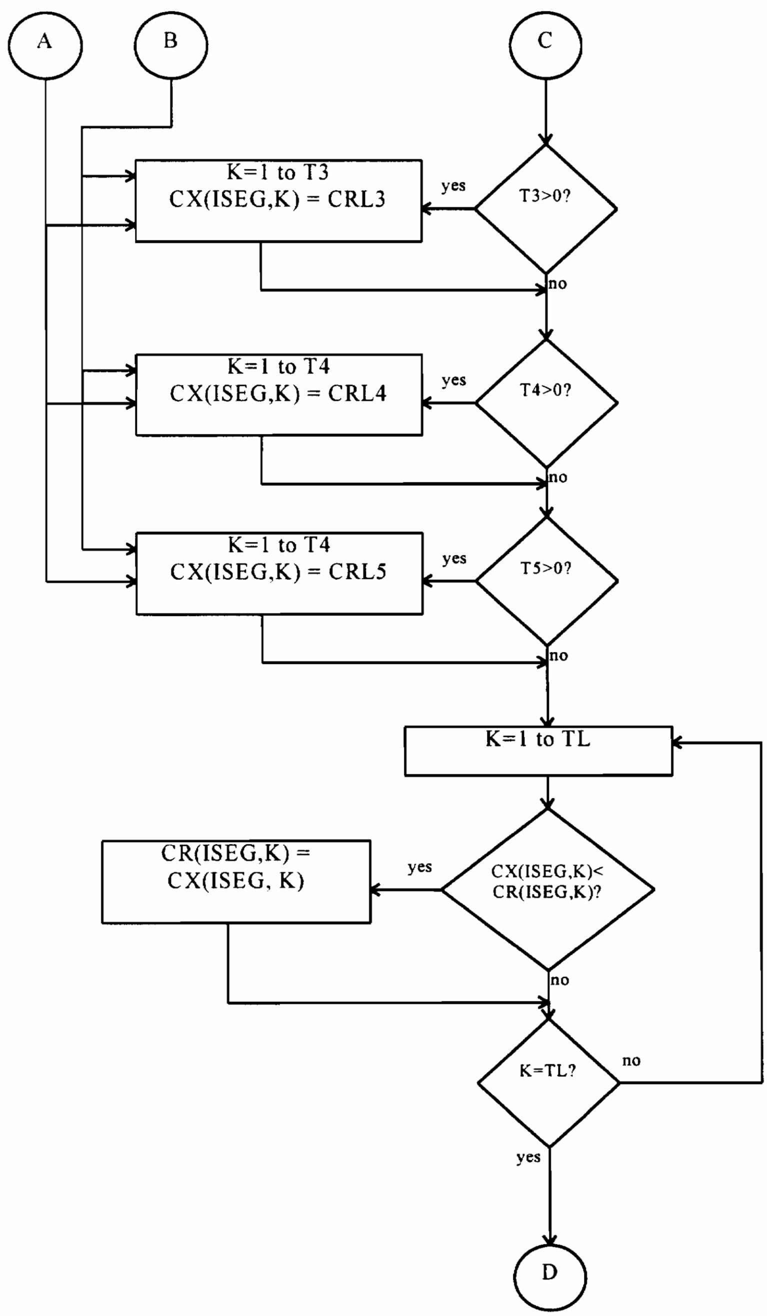


Figure 4-5: Capacity Reduction Module, cont.

The algorithm then begins its loop by selecting the next incident in the incident stack. If there are no incidents left to process then the array returns the capacity reduction array back to the traffic forecasting module. If there is another incident to process, it computes T_l, the time left to completely clear the incident. This is computed by:

T_l=[T_{id}(the predicted incident duration)-T_{is}(the incident start time)]. (4.1)

The next step in the procedure makes use of the data complied about dynamic lane clearance. This module determines how long it will take to clear each lane. It will look in the lane phasing table, and based on the type, the expected duration, the number of lanes in the segment, and the number of lanes blocked, it will estimate the time it takes to clear each lane (T_k: k=1,2,3,4,5). This dynamic lane clearance table is based on the compiled data discussed in chapter three and is illustrated in table 4-1.

The next section of the module reduces the capacity of the freeway segment containing the incident. It does so by looking up the capacity reduction factor for each situation affecting the incident duration. The capacity reduction factors are based on many studies done in the past on freeway lane blockages and traffic flow. For the purposes of this research the factors are taken from (Lindley, 1988). These factors for various incident situations are listed in table 4-2. For example, for a situation where 2 out of 4 lanes are blocked, the capacity is only 25% or 0.25 of what it normally is. The next five sections of this module go through the time each lane is blocked, and apply the appropriate capacity reduction factor.

Incident Type	Expected Duration	Total Lanes	Blocked Initially	4 lanes blocked	3 lanes blocked	2 lanes blocked	1 lane blocked
Prop. Damage	30	4	4	**6**	**10**	**15**	**20**
Prop. Damage	30	4	3		**7**	**12**	**21**
Prop. Damage	30	4	2			**10**	**18**
Prop. Damage	30	4	1				**18**
Prop. Damage	30	3	3		**7**	**12**	**21**
Prop. Damage	30	3	2			**9**	**19**
Prop. Damage	30	3	1				**17**
Prop. Damage	60	4	4	**8**	**12**	**19**	**30**
Prop. Damage	60	4	3		**9**	**16**	**28**
Prop. Damage	60	4	2			**15**	**26**
Prop. Damage	60	4	1				**24**
Prop. Damage	60	3	3		**9**	**17**	**28**
Prop. Damage	60	3	2			**13**	**26**
Prop. Damage	60	3	1				**25**
Prop. Damage	120	4	4	**17**	**25**	**41**	**66**
Prop. Damage	120	4	3		**30**	**39**	**65**
Prop. Damage	120	4	2			**32**	**60**
Prop. Damage	120	4	1				**58**
Prop. Damage	120	3	3		**22**	**50**	**66**
Prop. Damage	120	3	2			**30**	**64**
Prop. Damage	120	3	1				**63**
Minor Injury	30	4	4	**10**	**14**	**19**	**27**
Minor Injury	30	4	3		**11**	**18**	**26**
Minor Injury	30	4	2			**15**	**24**
Minor Injury	30	4	1				**25**
Minor Injury	30	3	3		**10**	**18**	**26**
Minor Injury	30	3	2			**15**	**24**
Minor Injury	30	3	1				**25**
Minor Injury	45	4	4	**11**	**17**	**25**	**36**
Minor Injury	*45*	*4*	*3*		***15***	***23***	***35***
Minor Injury	45	4	2			**20**	**34**
Minor Injury	45	4	1				**34**
Minor Injury	45	3	3		**13**	**22**	**36**
Minor Injury	45	3	2			**19**	**33**
Minor Injury	45	3	1				**34**
Minor Injury	60	4	4	**14**	**22**	**31**	**43**
Minor Injury	60	4	3		**17**	**28**	**43**
Minor Injury	60	4	2			**25**	**42**
Minor Injury	60	4	1				**42**
Minor Injury	60	3	3		**16**	**28**	**44**
etc.	etc.	etc.	etc.	...	...	...	...

Table 4-1: Sample Dynamic Lane Clearance Table

To illustrate this consider an example where an incident with minor injuries that blocks three out of four lanes is predicted to take 45 minutes to clear. Also assume the incident occurred at time *t*-10 (ten minutes before the simulation is run). According to the lane phasing table an incident of this type usually has three lanes blocked only for the first 15 minutes, two lanes blocked for 23 minutes, one lane blocked for 35 minutes and of course the shoulder is blocked for the duration of 45 minutes. The capacity for 3/4 lanes blocked is 0.13. The capacity for 2/4 lanes blocked is 0.25. The capacity for 1/4 lanes blocked is 0.58, and finally, the capacity for just a shoulder blocked in a four lane segment is 0.85. In this situation the capacity reduction table for this segment would look like table 4-3. If dynamic lane clearance were not accounted for, 3 lanes would be considered closed for the duration and the same table would look like table 4-4. This is not as realistic and would result in overpredicting the delay.

	Lanes Blocked				
Number of Lanes	**Shoulder**	**One**	**Two**	**Three**	**Four**
2	**0.81**	**0.35**	**0**	**N/A**	**N/A**
3	**0.83**	**0.49**	**0.17**	**0**	**N/A**
4	**0.85**	**0.58**	**0.25**	**0.13**	**0**

Table 4-2: Fraction of Freeway Capacity Available Under Incident Conditions (Lindley, 1988)

The final section of this module checks to see if the capacity in the segment affected by the incident has already been reduced below that of the presently processed incident. If this incident reduces capacity further, then the capacity reduction values are written to the capacity reduction array. Finally if there are no more incidents to process, the array is returned to the traffic forecasting module.

Time	Capacity Reduction (CR)		Time, cont.	Capacity Reduction (CR)
t+1	**0.13**		t+20	**0.58**
t+2	**0.13**		t+21	**0.58**
t+3	**0.13**		t+22	**0.58**
t+4	**0.13**		t+23	**0.58**
t+5	**0.13**		t+24	**0.58**
t+6	**0.25**		t+25	**0.58**
t+7	**0.25**		t+26	**0.85**
t+8	**0.25**		t+27	**0.85**
t+9	**0.25**		t+28	**0.85**
t+10	**0.25**		t+29	**0.85**
t+11	**0.25**		t+30	**0.85**
t+12	**0.25**		t+31	**0.85**
t+13	**0.25**		t+32	**0.85**
t+14	**0.58**		t+33	**0.85**
t+15	**0.58**		t+34	**0.85**
t+16	**0.58**		t+35	**0.85**
t+17	**0.58**		t+36	**1.00**
t+18	**0.58**		t+37	**1.00**
t+19	**0.58**		t+38	**1.00**

Table 4-3: Sample Capacity Reduction Table

Time	Capacity Reduction (CR)		Time, cont.	Capacity Reduction (CR)
t+1	**0.13**		t+20	**0.13**
t+2	**0.13**		t+21	**0.13**
t+3	**0.13**		t+22	**0.13**
t+4	**0.13**		t+23	**0.13**
t+5	**0.13**		t+24	**0.13**
t+6	**0.13**		t+25	**0.13**
t+7	**0.13**		t+26	**0.13**
t+8	**0.13**		t+27	**0.13**
t+9	**0.13**		t+28	**0.13**
t+10	**0.13**		t+29	**0.13**
t+11	**0.13**		t+30	**0.13**
t+12	**0.13**		t+31	**0.13**
t+13	**0.13**		t+32	**0.13**
t+14	**0.13**		t+33	**0.13**
t+15	**0.13**		t+34	**0.13**
t+16	**0.13**		t+35	**0.13**
t+17	**0.13**		t+36	**1.00**
t+18	**0.13**		t+37	**1.00**
t+19	**0.13**		t+38	**1.00**

Table 4-4: Sample Capacity Reduction Table w/o Dynamic Lane Clearance

4.3.3. Flow Forecasting Module

As discussed in the literature review, the macroscopic simulation model simulates the main line of the freeway as it is affected by the corridor input and ramp flows. These corridor input and ramp flows include: the main corridor input flow, the on ramp flows, and the off ramp flows. These are shown in light grey in figure 4-6. These numbers must be stored in arrays that can be read into the simulation. The form that the arrays take is:

$Q_0(m)$ = The freeway corridor input flow for each minute

$RI(s,m)$ = The on-ramp volume for each segment for each minute

$RO(s,m)$ = The off-ramp volume for each segment for each minute

$m=(t+1, t+2, t+3, ..., t+180)$: forecast for the next 3 hours

$s=(1, 2, 3, ..., sm)$: each segment

The arrays $Q_0(m)$, $RI(s,m)$, and $RO(s,m)$, are termed the input flow arrays because they represent the flows affecting the simulation that act as inputs. The mainline flows for each segment are the output of the simulation, so the array $Q(s,m)$ for $m=(t+1, t+2, t+3, ..., t+180)$ can be termed the output flow array. These input flows are future flows so they must be forecast and put into arrays before the simulation is run.

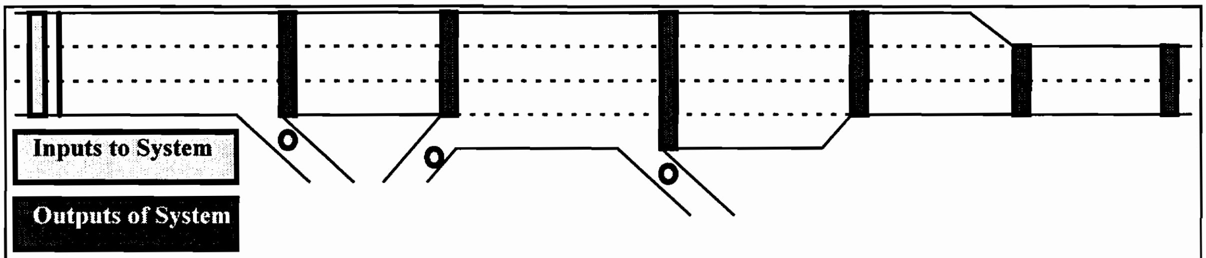


Figure 4-6: Sample Input and Output Data Stations

There are many ways to forecast these flows using both historical and real time information. As mentioned in the literature review, time-series analysis offers many procedures that can be performed to study the combination of historical and real-time flows to make more accurate predictions.Since this process will be forecasting 180 steps into the future it can be considered long-term. For long-term forecasts, (DeRomph, 1992) proved that it is usually advisable to rely only on the historical average.

Where a time-series model can make a difference in the short term. In order to make predictions in the short term better than the historical average, study should be done to determine the most appropriate model for both the corridor input and the ramps on the freeway corridor. Since the majority of the forecast will be for the long-term it is suggested that relying only on the historical average would be a reasonable assumption. A model to improve the input forecasts with real-time data would be implemented as shown in figure 4-7 in the dashed boxes. This area is left for future research and later implementation into the flow forecasting module.

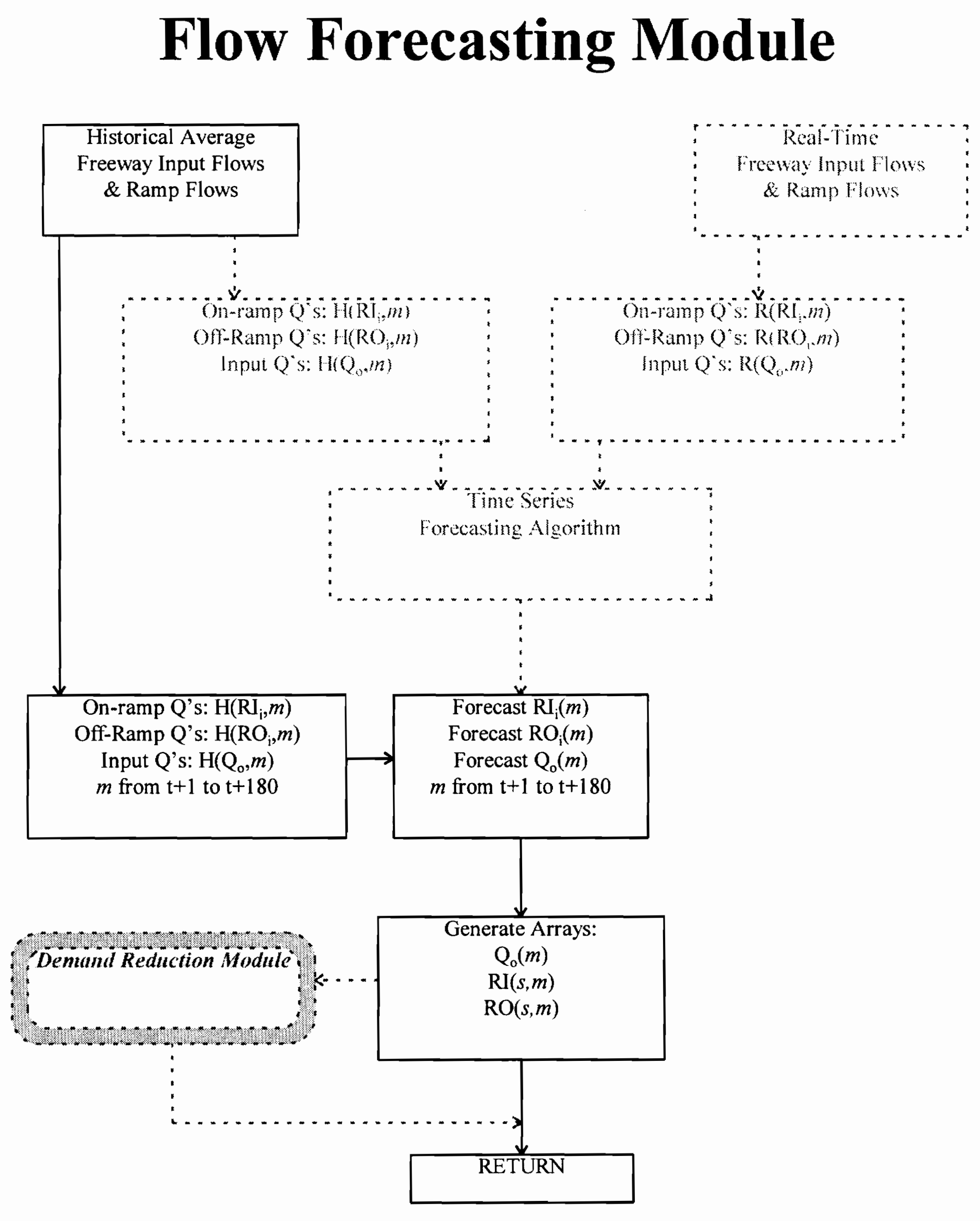


Figure 4-7 Flow Forecasting Module

For the model implemented in this research, the input arrays will be filled with historical information. This information only needs to be read once a day and stored in memory because it will not change much throughout a day's period. Real-time data that is collected can be added to a continuously growing database and the input flows can be averaged once a day and re-read into the arrays. Of course, if a given segment x does not include an on-ramp, then the array RI(x,m)=0 for all m=(t+1, t+2, t+3, ..., t+180). And similarly if a given segment x does not include an off-ramp, then the array RO(x,m)=0 for all m=(t+1, t+2, t+3, ..., t+180). These input arrays are then returned to the traffic forecasting module and sent to the simulation program.

As shown in figure 4-7, a demand reduction module can be implemented but this is left for future research and later implementation. As indicated in figure 1-1, feedback may occur between the response module and the delay prediction module. This feedback may be in terms of ramp metering. If a ramp is specified to meter vehicles at a certain rate in the future, then the on-ramps should not allow more vehicles to enter the freeway than is specified by that metering rate. If a ramp is closed, this will also affect the input arrays, and they would also be adjusted by this module.

If diversion occurs then the exit ramp volumes would be increased. These increases may occur either due to natural or imposed freeway diversions. Drivers naturally divert if they know there will be heavy delays ahead and they know of alternative routes. Determining increases in number or percentage of drivers diverting due to better information an interesting but complicated study, thus it is left for further research. Once more study has been done in this area, an algorithm for increasing on-ramp volumes can be implemented in this delay reduction module, along with other algorithms to adjust ramp flows due to response measures.

4.3.4. Macroscopic Simulation Module

The macroscopic simulation module compiles the various input arrays into a single datafile that is read by the simulation routine at runtime and is illustrated in figure 4-8. This file is referred to as the simulation runfile. It contains all the arrays of data created by the previous module plus fixed data concerning the geometry of the freeway corridor. These data arrays are listed in table 4-4. The label s stands for the total number of segments in the simulation and the label m indicates the total number of minutes to run the simulation for. Typically the simulation would be run for m=180 minutes to provide a 3-hour forecast.

Assuming a 3-hour forecast, there are a total of (545*s)+180 numbers in the simulation runfile. For the sample 6-mile freeway corridor from I-880 used in this research, there are a total of 17 segments. This means a total of 9,445 values in the simulation runfile. This runfile is read all at once and the fed into the appropriate arrays so the simulation runs properly. While the simulation runs it returns the value of speed and flow for each segment for the time t+1 to t+180. These two arrays are written to an output file and returned from the simulation back to the main algorithm. These arrays are each of the size (s x m) so are a total of 360*(s) values returned for a simulation run of 180 minutes. For the sample freeway used in this research, there are a total of 6,120 values returned. An illustration of this process is shown in figure 4-8.

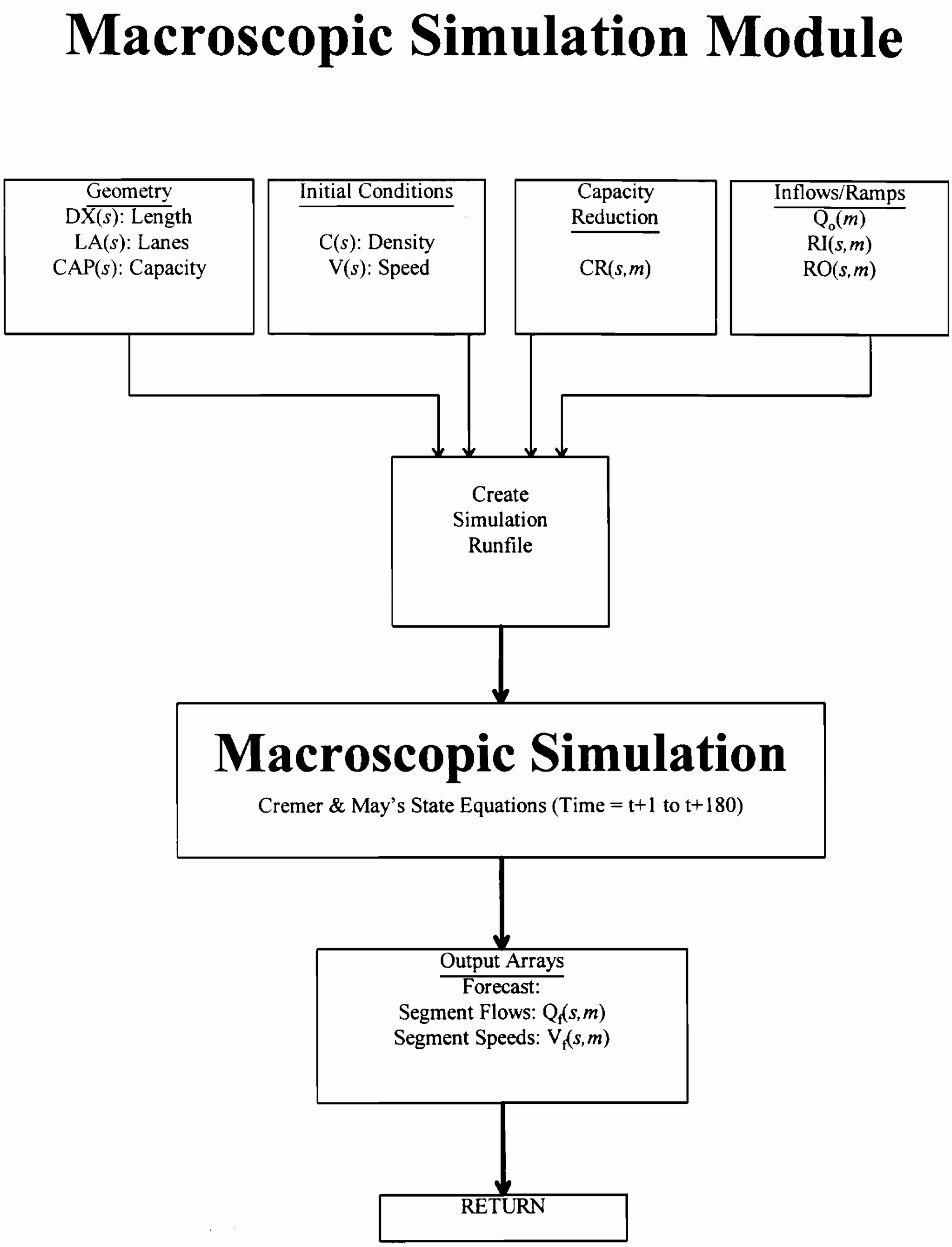


Figure 4-8 Macroscopic Simulation Module

The actual simulation process involves calculating the speeds, flows, and densities for each segment progressively for the 180 minute period in 5 or 10 second time slices. The simulation will calculate output flows and speeds on main links in a freeway corridor. The output stations would be at critical points on the links as shown in dark grey in the sample in figure 4-6. Because this is complicated and is the main focus of the delay prediction system, it is discussed in detail in the next chapter.

Array	Dim	Size	Description
DX	1	s	Length of each segment (ft)
LA	1	s	Number of lanes in each segment (dim)
CAP	1	s	Capacity of each segment (veh/hr/ln)
C	1	s	Initial Density for each segment (veh/mile)
V	1	s	Initial Speed for each segment (mile/hr)
CR	2	$s \times m$	Capacity reduction factors (dim)
Q_0	1	m	Corridor inflow (veh/hr)
RI	2	$s \times m$	On-ramp flows (veh/hr)
RO	2	$s \times m$	Off-ramp flows (veh/hr)

Table 4-5: Simulation Input Arrays

4.4. Data Processing Module

The raw output of the simulation is extremely detailed, and the framework is flexible. The framework for the output processing module is illustrated in figure 4-9. There are many ways the data can be used by the response module including the update of variable message signs and update of highway advisory messages. Indirectly, the response module may use the data to make decisions on traffic diversion or ramp metering. The data returned is of sufficient resolution to be applicable for these applications if it is processed in the proper manner.

The output can be summarized into many useful output formats including speed vs. time graphs, flow vs. time graphs, speed contour plots, flow contour plots, and travel time matrices. All of these applications could be studied, but it is beyond the scope of this research to delve into all the possibilities of the traffic forecasts. The travel time matrix has the most useful application in giving messages to drivers on future traffic conditions, therefore it will be the method described in this research. Other output formats can be similarly created in parallel with the travel time matrix.

Output Processing Module

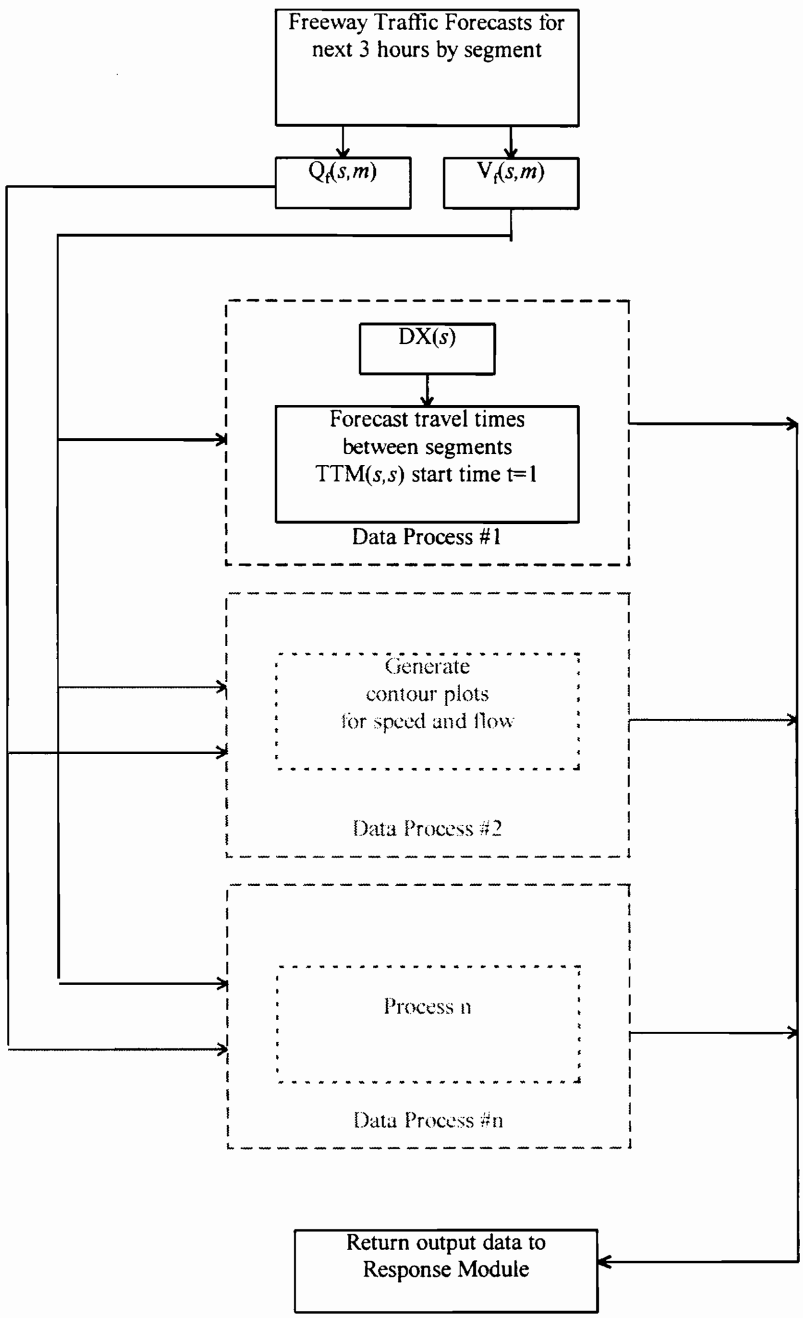


Figure 4-9 Output Processing Module

A travel time matrix giving the travel time from each segment to each downstream segment in the corridor can be easily created from the speed forecasts for each segment. Many sources of error may creep into the travel time predictions so it is advised that they are not directly used. Instead, the travel time forecast matrix can be used in conjunction with a non-congested travel time matrix to determine levels of congestion. For instance, suppose the non-congested travel time between segments A and B is 12 minutes and a variable message sign is located just upstream of segment A. A rule in the response module might give messages for various levels of congestion:

- Travel time <= 13 minutes: "HAVE A NICE DAY"
- 14 minutes ≤ TT ≤ 18 minutes: "MINOR DELAYS AHEAD"
- 19 minutes ≤ TT ≤ 24 minutes: "MAJOR DELAYS AHEAD"
- 25 minutes ≤ TT: "SEVERE CONGESTION *X* MILES AHEAD"

These messages would be established on the base travel time which is 12 minutes between segments A and B. For heavily congested situations, it may be desirable to give drivers a distance that they will go before they encounter the congestion. This information can help in making a decision on if and where to divert. This distance measurement *X* can be done manually by an operator at the traffic control center with the aid of the traffic forecast in the form of a speed contour plot. A speed contour plot program could automatically highlight areas with speeds that are forecasted to drop below a given threshold.

Other methods listed including contour plots would also be done in the data processing module. This module does not have as defined a structure as the other modules because it is more open-ended. Response systems may differ in needs from location to location. Some incident management systems may be able to make use of detailed reports, while others can't. Even still, some response modules may have additional capabilities that require special queries of the data that need to be custom designed for the specific setup. Almost any query imaginable can be done based on the stored geometrics, flow forecasts, and speed forecasts for each segment.

5. Simulation and Calibration

The macroscopic simulation module was noted as the integral part to the delay prediction framework developed in this thesis. The simulation model can be thought of as the "brain" or "engine" to the whole algorithm. The previous chapter discussed the general framework. This discussion included the architecture in which the algorithm runs showing the data inputs, data flow trees, and data outputs. A great deal of research has gone into insuring accurate inputs to the simulation module, for without accurate input, the output cannot be expected to be accurate. Accuracy has been the biggest problem with traditional methods, and the primary goal of this research was to overcome this problem through the incorporation of more complete information and more detailed models. The majority of the architecture is devoted to assuring that the most accurate and complete information is fed into the simulation, and that the output is correctly handled and used efficiently by the incident response system at a traffic management center. While the previous chapter was devoted to the algorithm inputs, outputs, flow, and structure, this chapter will focus on the single key element in the algorithm: the macroscopic freeway corridor simulation. The first part of the chapter will describe the simulation used and the modifications made to incorporate it into the hybrid delay prediction architecture. The latter part of the chapter will describe a method used to calibrate the simulation parameters.

5.1. The Macroscopic Simulation

The code created for this simulation was written to conform to the macroscopic dynamic freeway model that was used by (Cremer and May, 1986) and fully described in detail in section 2.5.1. The method used by this research can be looked at as an extension of Cremer and May's research. They indicated that future work should include testing this extended model with more data sets from different traffic sections on other freeways. They (Cremer & May, 1985) also indicated that a better calibration procedure incorporating data with a higher resolution is also highly desirable. In their research, the data used was based on five or fifteen minute averages, while the data used in this research uses one minute averages. While they used much slower computers a decade ago with reasonable speed, computers today can perform the simulation in a matter of seconds. It should be noted that this research is not a duplication of Cremer & May's research, but rather a modification, extension, and practical application of it. It incorporates the model in a dynamic framework that can be used for traffic and delay prediction on freeways.

The important modifications and extensions to the model proposed by Cremer and May include:

- the program reads in a single data 'file' because of large data requirements
- the capacity of each section for each minute is modified by the CR factors
- volume is always limited to capacity of the segment

- any segment j is treated as a bottleneck if $CAP_j < CAP_{j-1}$

5.2. Calibration of the Parameters

As indicated in the discussion of equations 2.9 through 2.15 there are eleven different parameters that need to be calibrated. A table summarizing these parameters is illustrated in table 5-1. In order to properly calibrate the simulation, the model needs to be looked at as a causal system with inputs and outputs. The inputs flows including Q_0 and all the ramp flows are taken as inputs to the causal system. The system reacts to these inputs through the transitions in the state variables. The response of the system to these changes can be seen in the outputs of the system. These causal inputs and outputs are shown in figure 4-6. In this case the outputs are the main line speeds and flows for each minute for each segment.

speed-density-characteristic				basic model				extensions		
v_f	C_m	a	b	α	κ	υ	τ	β	γ	δ
free velocity	maximum density	exponent	exponent	volume averaging factor	denominator parameter of density gradient	sensitivity of density gradient	time constant	density gradient modification factor	convection term weighting factor	ramp flow weighting factor

Table 5-1 Calibration Parameters

More specifically, the goal of calibration is to select a set of parameters that will minimize the difference between the output variables and the real main line data when real data is given as input to the system. This can be summarized by declaring a performance index, PI:

$$PI = \sum_{s=1}^{SEG} \sum_{m=1}^{MIN} \left\{ \lambda \left(Q_r(s,m) - Q_{sim}(s,m) \right)^2 + \mu \left(V_r(s,m) - V_{sim}(s,m) \right)^2 \right. \qquad (5.1)$$

where Q_r and V_r are real reference values and Q_{sim}, and V_{sim} are corresponding simulation outputs for each segment for each minute. The terms λ and μ are included to balance the error between the flows and speeds. They are chosen at 10^{-6} and 10^{-2} respectively. The goal is to select appropriate values for each of the parameters such that they minimize the performance index PI. (Cremer & May, 1986)

In order to achieve this, an optimization routine was created to select parameters and evaluate respective performance indexes in an iterative manner to search for an optimal solution that minimizes the performance index. Each time a new parameter set is chosen, the simulation must be run to determine the performance index. This process is illustrated in figure 5-1.

Most optimization routines depend on differentiable objective equations. Unfortunately there is no closed form differentiable objective function in this case. The objective function to be minimized is the performance index function, but in order to compute it, a simulation run has to be performed. The only real way to optimize this type of problem then is through a search method. (Hooke and Jeeves, 1961) proposed a method for solving this type of problem through the use of a pattern search. The method has proved successful for similar situations and the method has been coded into a computer program. For more details on the structure of this direct search method the reader is referred to Hooke and Jeeves (1961).

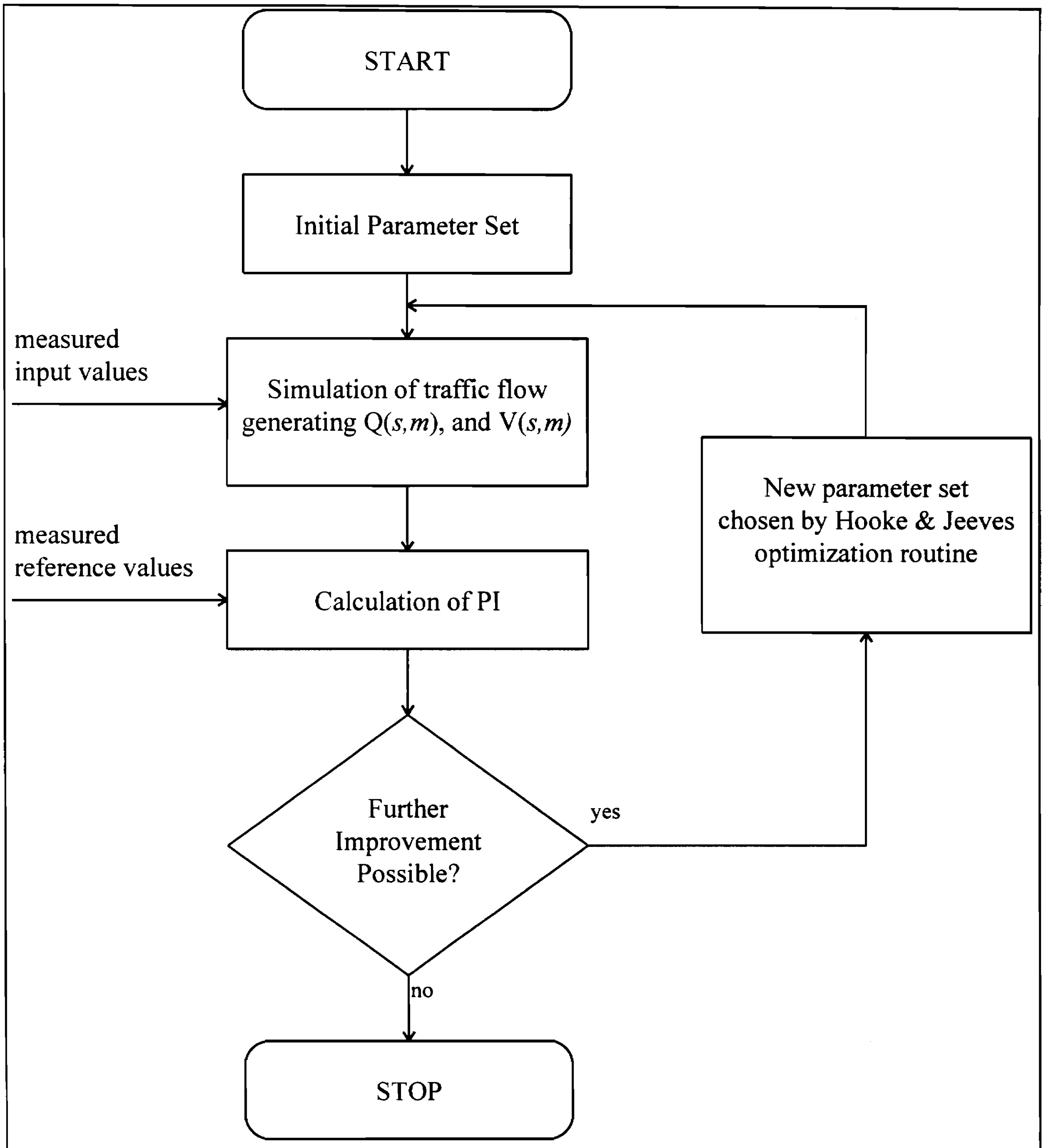


Figure 5-1: Recursive Parameter Optimization (Cremer & May, 1985)

A table showing the initial parameters is illustrated in table 5-2. The I-880 corridor is similar to the Santa Montica Freeway segment that was used in the study by Cremer and May, so it was proposed that the parameters would also be similar.

speed-density characteristic				basic model				extensions		
v_f	c_m	a	b	α	κ	υ	τ	β	γ	δ
93.1 km/h	110 veh/km/lane	1.86	4.05	0.95	9.5 veh/km/lane	23.9 km^2/h	20.4 sec	0.79	0.70	1.35

Table 5-2: Initial Parameter Set (Cremer & May, 1985)

5.3. Simulation Limitations

The extended model for traffic flow on urban freeways proposed by (Cremer and May, 1985) appears to model traffic phenomena very well. It accurately represents changing traffic conditions in the domain of reduced capacity or bottleneck sections and in areas with changing geometry and a myriad of ramps. It performs well even when the freeway changes geometric conditions between segments and is non-uniform, and for this reason does not require calibration of individual speed-density characteristics for each segment.

The speed-density characteristic described by equation 2.10 is highly general, thus it has four parameters to be calibrated. It was assumed in the development that this equation would hold true for each segment when it is calibrated with the freeway.

Unfortunately, changes in freeway geometrical or environmental conditions may negate this assumption. If conditions are non-uniform enough between segments then this assumption would not hold and the model must be expanded to deal with different speed-density characteristics for each segment. This expansion would greatly increase the number of parameters making calibration much more difficult and time-consuming. Therefore this model as it is presented is limited to freeway sections that are reasonably uniform and do not require individual speed-density characteristics for each segment.

The other limitation has to with the boundaries of the freeway section. Incidents or bottlenecks beyond the boundaries of the section studied can have effects on the model that can not be accounted for. Traffic is normally allowed to leave the last segment at up to the capacity flow rate. If an incident or bottleneck occurs downstream of the section then traffic may queue up into the freeway section causing congestion that could not be accounted for. For some of the real data in the F.S.P. database speeds near the downstream boundary decrease significantly in some instances. This is obviously due to a bottleneck or incident downstream. This cannot be accounted for in the model, thus calibration can be cumbersome. It is suggested that for calibration purposes, real data should not be used that has unexplainable congestion near the downstream boundary.

Upstream conditions will also play a role in the accuracy of the prediction. If the traffic entering from the upstream section is extremely variable on a day to day basis, then the accuracy of the simulation will suffer. One of the main inputs to the simulation Q_0, the freeway corridor input, is based on the historical average. If there is not much variability, then the historical average should hold well. On the other hand, if the corridor input is very erratic on a day to day basis due to incidents or other conditions, then the historical average would not be a good predictor of future flows and the accuracy of the simulation would decrease. For this reason, it is suggested that the upstream segment for study or implementation have a fairly uniform flow pattern from day to day that closely follows the historical average.

6. Proposed Testing and Operating Procedure

The thesis began with a basic overview of traffic management systems and the basic architecture of an incident management information system. The concept of delay prediction was discussed and examined. Many historical methods were discussed with their strengths and weaknesses pointed out. The third chapter discussed data requirements for the development of an appropriate model. The data collection and manipulation efforts were also described and summarized. Chapter four developed the new delay prediction framework, describing each of its components in detail. The input, output, and data flow structures were illustrated and described. The heart of the algorithm, the macroscopic freeway simulation model, was described in chapter 5. Calibration of parameters was discussed along with limitations and possible implementation problems that may arise. This chapter will propose a method for testing, calibrating, and implementing the delay prediction framework for use in a wide area incident management support system.

6.1. *Testing and Calibration*

Before the framework can be hard coded and implemented into an incident management system, it must first be tested and calibrated. The Freeway Service Patrol data is ideal for testing the system. There are many different real-world scenarios depicted in the data. The framework should be able to accurately represent those various scenarios before it is implemented for real-time use.

For testing and calibration purposes it is recommended that the data input/output procedures, the simulation, and the calibration routine be combined so the optimization can proceed smoothly. Each time the Hooke and Jeeves algorithm perturbs the parameters, the simulation should automatically be run to determine the performance index for that set of new parameters. When the performance index has been minimized then the parameter set is said to be "optimal" for that particular data set.

For the initial test, it is suggested that the afternoon, northbound shift on March 19, 1993 be used. This particular data set was chosen because only one specific lane blocking incident occurred during that shift. Speeds just upstream of the incident area before the incident occurred were in the 55-60 mph range. This indicated that there was no recurring congestion in the domain of the incident. This simple isolated incident case is simpler than other sets that included multiple incidents and recurring congestion. Although the framework is designed to handle these complex situations, it is suggested that the first case tried should be simple enough so programming errors could be isolated.

The incident from this shift is indicated as #1456 in the incident database. It occurred at 5:04 pm in segment 9 which is a five lane segment. Two lanes were blocked at the start of the incident, but all lanes were open by 5:21 pm. The incident was finally cleared from the shoulder around 5:50 pm. (Al-Deek, et. al. 1994). Just before the incident occurred, speeds were recorded under 30 mph two segments downstream of the incident location. This would indicate that a queue from a bottleneck further downstream reached that segment. After the incident occurred, traffic was metered in segment 9 which in turn improved the traffic downstream. This is typical of many freeway incidents and should be easily handled by the simulation model.

To model it within the delay prediction framework, the input data needs to be created. The geometrics of the I-880 including number of segments, segment lengths, and number of lanes can be realized from (Al-Deek, et. al., 1994). The initial speeds and densities can be taken from the loop data files at the beginning of the incident. The freeway corridor input flows and ramp flows should be taken from the "averaged" loop files in the dataset. Finally, the capacity reduction matrix should be built using the rules described in chapter four. Using the actual incident duration (~45 minutes), incident type (property damage), the number of lanes existing (4), and the number of lanes blocked (2), an appropriate capacity reduction matrix can be realized. All this information should be compiled into a data file and read into the simulation program. Initial values for the parameters can be taken from table 5-2.

When the performance index has been minimized by the optimization procedure, the final parameters should be stored in a database with the incident data. Following runs of the system should be done with additional incidents and loop data from the database. After many runs with many different incident scenarios there will be a compiled set of scenarios and optimal parameter sets. At this point some patterns may be realized. It is hypothesized that similar incident scenarios will also have similar optimal parameter sets.

If a large enough database can be compiled of incident scenarios and optimal parameter sets, then a specific parameter set can be chosen to make optimal forecasts based on some incident variables.

6.2. Implementation

Because of the data-hungry nature of the framework, it is only practical to implement it in areas where there exists detailed real-time data on-line for immediate use. This type of surveillance is just being brought about in many urban areas around the country. If this is available then the open architecture described in this thesis can easily be incorporated into the exiting management systems.

For real-time implementation, the calibration procedure described in the previous section will not be used. Instead, an appropriate workstation program that is compatible with the existing system, will be created based on the framework depicted in figure 4-1. With continuous live data about the freeway conditions available, a comprehensive "historical" average can be made for the freeway corridor inputs and the ramp flows. These can be fed into the system along with actual conditions at time t+0. All of the existing incidents in the incident stack can be quickly modeled with the data lookup tables similar to those depicted in tables 4-1 and 4-2. The reduced capacity matrix along with the other freeway geometry would also be put into the input file to be read by the simulation program. If a database of parameters exists, a specific parameter set would be chosen, otherwise the optimal one determined through procedures described in section 6.1 would be used. The simulation would be run, and output in terms of actual travel times between exits and speed-time-space contour plots would be instantly available to the operator at the traffic management center. The entire procedure described in this paragraph should be accomplished in less than a minute with a minimum of keystrokes required by the user.

7. Conclusions and Recommendations

7.1. Conclusions

Predicting delay due to incidents on urban freeways is a common problem for many traffic management systems. Many systems today could benefit from an accurate prediction of delays caused by an incident. Traffic management center operators typically base their response actions on past experience and set policies. Although this may be expert judgment, it is not always based on the best information.

An incident may have far-reaching implications on a freeway network depending on the time of day, the location of the incident, the type, the location of on-ramps, and a host of other factors. Any one of these could easily be left out of an important response decision. The delay caused by an incident often is a result of a specific combination of a number of these factors. Usually these factors are considered separately and combined to make a prediction of delays. The combination of various factors as a whole has not been considered in the past, so that was the basis of this research: create a new framework that can account for a plethora of factors regarding an incident, not just individually, but as a combination.

The type of model that encompasses the widest range of variables while still being fast enough for on-line applications is the macroscopic freeway model. A dynamic extended model for urban freeways was proposed and tested by (Cremer and May, 1985) that proved to model urban freeway traffic phenomena very well. This dynamic model was taken as the basis for the new delay prediction architecture.

A simulation cannot predict delays on a freeway without other procedures. The specific procedures a simulation needs to give useful output is an input processing procedure and an output processing procedure. The input for a freeway simulation contains many components including geometric information, real time freeway conditions, historic freeway conditions, traffic control measures, incident information, and ramp flows. These all need to be gathered from various sources and put in a readable format for the simulation program. Once the simulation has run, the output also needs to be processed to give useful results back to the traffic control center to make response decisions.

This thesis provides an architecture for incorporating a macroscopic simulation with input and output procedures that can be effectively used to provide accurate and timely forecasts of traffic conditions for an urban freeway corridor. Due to the complex nature of urban freeway systems, the architecture was left open with many areas for future research.

An historical problem with using simulation in research is the large amount of time it takes to create the program and the lengthy input files. Often there has not been enough data of detailed enough resolution to even perform a simulation. The rich and detailed data that is in the F.S.P. database begs to be used for simulation purposes. This may have been avoided in the past because of the difficulty of applying the data to a simulation program or due to slow speeds of microcomputers. To overcome this, a program was developed to build runfiles directly from the F.S.P. data. This enormously cuts down the time it takes to create satisfactory simulation inputs. This data file program takes input from the user including the date, shift, and direction and makes an appropriate runfile for the simulation calibration program. This runfile can be automatically read into the simulation calibration program. This utility greatly speeds up the research process automating the time-consuming task of setting up the proper simulation model. In this fashion, more time can be spent doing more scientific research such as developing better capacity reduction strategies.

7.2. Recommendations for Future Research

Of course the process described in chapter six to calibrate and test the simulation within the new delay prediction framework still needs to be carried out. It will take additional work to test the different scenarios and calibrate the simulation with the many sets of data from the F.S.P. database. This additional research may also bring up other areas of investigation particularly in the simulation equations. Such detailed simulation analysis with real world data has not been done in the past, so this additional research may bring about more additions and extensions similar to the work that Cremer and May performed. The framework described in this thesis will most definitely facilitate this future research.

Throughout the thesis there were also many parts of the framework that were not fully discussed and/or noted as future research topics, in particular, the parts of the flow figures printed in **grey**. This final section of the thesis lists and summarizes the possible research topics. The topics briefly mentioned in the thesis but left for further study include:

- The capacity reduction module (4.3.2)
    - further research into dynamic lane clearance
    - further research into reduced flow rates through reduced capacity sections
- The flow forecasting module (4.3.3)
    - incorporation of time-series forecasting in lieu of strict historical averages
    - development of a demand reduction module
- The output processing module (4.4)
    - creation of code to create speed and flow contour plots
    - design of a user query module for use in traffic management centers
- Incorporating new algorithms to better calibrate the simulation parameters (5.2)
- Increasing the depth of the simulation itself (2.5.1)

Dynamic lane clearance is a complicated and real phenomenon that directly impacts freeway traffic forecasts that include incidents. A brief study was done in this research surveying 75 incident clearance personnel on their opinions of various freeway incident situations. Although this may provide a general basis for defining dynamic lane clearance in the capacity reduction module, much study still has to be done. A survey of actual incidents involving dynamic lane clearance on freeways the system is to be used on would be an invaluable addition to this research and certainly increase the accuracy of the delay predictions.

Reduced flow rates in reduced capacity sections were taken from previous research done in this field. This research has been limited and insufficient. Further research needs to be completed with more detailed data sets similar to the F.S.P. database in order to gain a better understanding both quantitatively and qualitatively of what reduced flow rates are actually are. A large data collection effort would be very valuable in further refining these reduced capacities for various blockage situations.

Time-series forecasting was discussed in both the literature review (2.4) and in the algorithm development (4.3.3). These methods have convincing potential for creating better short term flow forecasts for the simulation inputs. Currently, the algorithm relies wholly on historical data to forecast future input variables. Extensive study needs to be done to determine what models can best improve the forecasts in the short term. While it has been shown that consistently using the historical average forecasts is adequate, surely a better forecast can be made incorporating real-time conditions.

The demand reduction module also discussed in section 4.3.3 was left for further research. Incorporating ramp metering strategies would be a complicated but interesting study that would make the freeway inputs more accurate. Mainline demand reductions due to drivers diverting is an important circumstance that is not covered by the algorithm. This diversion was also discussed in section 4.3.3 and may have a significant impact on the accuracy of delay predictions. Studying this phenomenon would require a thorough and extensive data collection including information about incidents, traffic reports, other driver information, dynamic lane clearance, and detailed traffic flow information. This type of study would be expensive but is needed to accurately model this phenomenon.

The output data processing module was left open for many additional procedures to process the simulation output. One procedure indicated but not developed was speed and flow contour plots. This type of plot gives a simple one-page graphical summary in three dimensions, giving the operator at a traffic management center a quick overview of the forecast. In addition, there are many pieces of information that an operator at a traffic management center could use that could be generated from the simulation output. A study of operator needs and desires could be done and additional query modules could be easily implemented into this open architecture.

To calibrate the simulation parameters, a program incorporating Hooke & Jeeves pattern search algorithm was used. Other, more complex, algorithms have been used in the past optimize simulations. Some of these algorithms have been shown to perform better than the simple Hooke & Jeeves pattern search (Yunker, 1993). Implementing different optimization algorithms into the program is an important task as it could provide more accurate simulation parameters.

Thus far, the simulation has been used for single, isolated freeway corridors. In most real-world urban applications there exists an interconnected network of freeways. The simulation needs to be expanded to deal with freeway to freeway connections and the intricate details the accompany them. Combining a network of freeways into one simulation program will be more efficient then connecting several isolated simulations together.

This additional research will certainly have an impact on the accuracy and application of the framework presented in this thesis. This research is very young and additional research will help get this system on-line faster, and improve incident management procedures through better delay predictions.

8.REFERENCES

Ahmed, M. and Cook, A., (1979), "Analysis of Freeway Traffic Time Series Data Using Box and Jenkins Techniques", *Transportation Research Record*, 722, 1-8.

Akahane, H. and Kosi, M., (1986), "Short-Term Prediction of Inflow Volumes of Urban Freeways", *Second International Conference on Road Traffic Control*, April, Institution of Electrical Engineers, 35-38.

Al-Deek, H., and Garib, A., (1994), "Methods for Estimating Freeway Incident Congestion Part I: Conventional Methods", Paper presented at 1995 TRB Annual Meeting.

Al-Deek, H., and Garib, A., (1994), "Methods for Estimating Freeway Incident Congestion Part II: New Method", Paper presented at 1995 TRB Annual Meeting, 950705.

Babcock, P.S., A.D. May, D.M. Auslander, and M. Tomizuka, (1982), "Freeway Simulation and Control", *Institution of Transportation Studies*, University of California, Berkeley, Research Report UCB-ITS-RR-82-13.

Banks, James H., (1991), "The Two-Capacity Phenomenon: Some Theoretical Issues", *Transportation Research Record*, 1320, 234-241.

Banks, James H., (1990), "Flow Processes at a Freeway Bottlenecks", *Transportation Research Record*, 1287, 20-28.

Benekohal, Rahim F., (1991), "Procedure for Validation of Microscopic Traffic Flow Simulation Models", *Transportation Research Record*, 1320, 190-202.

Box, G.E.P. and Jenkins, G.M., (1976), *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day.

Cambridge Systematics, Inc., (1990), "Incident Management", Prepared for the Trucking Research Institute Foundation, Inc.

Chow, We-Min, (1976), "A Study of Traffic Performance Models Under an Incident Condition", *Transportation Research Record*, 567, 31-36.

Cremer, M., and May, A.D., (1986), "An Extended Traffic Flow Model for Inner Urban Freeways", *IFAC Proceedings Series No. 6, International Federation of Automatic Control*, Pergamon Press, 239-244.

Cremer, M. and May, A.D., (1985), "An Extended Traffic Model for Freeway Control", *Research Report UCB-ITS-RR-85-7*, University of California, Berkeley, May.

Cremer, M., and Papageorgiou, M., (1981), "Parameter Identification for a Traffic Flow Model", *Automatica*, Vol.17, No.6, 837-843.

Davis, G.A., and Nihan, N.L., (1991), "Nonparametric Regression and Short-Term Freeway Traffic Forecasting", *Journal of Transportation Engineering*, Vol. 17, No. 2, March/April, 178-188.

Davis, G.A., Nihan, N.L., Hamad, M.M., and Jacobson, L.N., (1990), "Adaptive Forecasting of Freeway Traffic Congestion", *Transportation Research Record*, 1287, 29-33.

De Romph, Erik, (1994), "A Dynamic Traffic Assignment Model, Theory & Applications", Distributed by the Delft University of Technology, Department of Infrastructure, Transport Planning and Traffic Engineering Section, The Netherlands.

Derr, B.R., (1987), "Delay Analysis for Freeway Corridor Surveillance, Communications, and Control Systems", *Transportation Research Record*, 1132, 77-81.

FRESIM User Guide, (1994), Version 4.5, Federal Highway Administration, Turner-Fairbank Highway Research Center, McLean, VA.

Hauer, U. and Hurdle, V.F., (1979), "Discussion of FREFLO", *Transportation Research Record*, 722, 75-76.

Highway Capacity Manual, (1994), Special Report 209, TRB, National Research Council, Washington D.C.

Hooke, R., and Jeeves, T.A., (1960), ""Direct Search" Solution of Numerical and Statistical Problems", *Journal of the Association for Computing Machinery*, April, 232-229.

Khattak, Asad, Schofer, J.L., and Wang, Mu-Han, (1994), "A Simple Procedure for Predicting Freeway Incident Duration", *TRB Paper, No. 940379.*

Khattak, Asad, Schofer, J.L., and Keppelman, F.S., (1991), "Factors Influencing Commuters' En Route Diversion Behavior in Response to Delay", *Transportation Research Record*, 1318, 125-136.

Kim, C., (1994), "Development and Evaluation of Traffic Prediction Systems", Ph.D. Dissertation, *Department of Civil Engineering*, Virginia Tech, April.

Lindley, Jeffrey A., (1988), "A Methodology for Quantifying Urban Freeway Congestion"" *Transportation Research Record*, 1132, 1-7.

Lindley, Jeffrey A., and Tignor, S.C., (1979), "Getaway Flow Rates for Freeway Incident and Geometric Bottlenecks", *Public Roads*, Vol. 43, No. 1, pp. 1-7.

May, A.D., (1990), *Traffic Flow Fundamentals*, Prentice-Hall, Englewood Cliffs, N.J.

May, A.D., and Keller, H., (1967), "Non-Integer Car-Following Models", *Highway Research Record*, 199, 19-32.

Messmer, M., and Papageorgiou, M., (1990), "Metanet: A Macroscopic Simulation Program for Motorway Networks", *Traffic Engineering and Control*, September, 465-469.

Michalopoulos, P.G., Kwon, E., and Kang, Jeong-Gyu, (1991), "Enhancement and Field Testing of a Dynamic Freeway Simulation Program", *Transportation Research Record*, 1320, 203-215.

Michalopoulos, P.G., (1984), "Dynamic Freeway Simulation Program for Personal Computers", *Transportation Research Record*, 971, 68-79.

Montgomery, D.C., Johnson, L.A., and Gardiner, J.S., (1990), *Forecasting and Time Series Analysis*, McGraw-Hill, Inc.

Morals, J.M., (1986), "Analytical Procedures for Estimating Freeway Traffic Congestion", *Public Roads*, September, Vol. 50, No. 2, 55-61.

Hobeika A.G., Özbay K., Subramaniam S., Krishnaswamy V., (1994), *A Heuristic Network Generator for Traffic Diversion during Non-Recurrent Congestion.* Presented at 1994 TRB Conference, Washington D.C.

Pankratz, Alan, (1983), *Forecasting with Univariate Box-Jenkins Models*, John Wiley & Sons.

Payne, H.J., (1978), "FREFLO: A Macroscopic Simulation Model of Freeway Traffic", *Transportation Research Record*, 722, 68-76.

Payne, H.J., (1973), "Freeway Traffic Control and Surveillance Model", *Transportation Engineering Journal*, November, Vol. 99, No. TE4, 767-783.

Petty, Karl, (1995), *FSP 1.1: The Analysis Software of the FSP Project*, Department of Electrical Engineering, University of California, Berkeley.

Shirazi, E., Anderson, S., and Stesney, J., (1988), "Commuters' Attitudes Toward Traffic Information Systems and Route Diversion", *Transportation Research Record*, 1168, 9-15.

Skabardonis, A., Petty, K., Noeimi, H., Rydzewski, D., Varaiya, P.P., (1996), *The I-880 Field Experiment: Database Development and Incident Delay Estimation Procedures*, Paper No. 960961, Paper presented at the 75th annual TRB Meeting, Washington D.C.

Stephanedes, Y.J., Kwon, E., and Michalopoulos, P., (1990), "On-line Diversion Prediction for Dynamic Control and Vehicle Guidance in Freeway Corridors", *Transportation Research Record*, 1287, 11-19.

Subramaniam S., Özbay K., (1995), "A Hybrid Expert-GIS Architecture for Wide-Area Incident Management." *ITS Quarterly*, Summer Issue, Volume 3, No. 1, pp.48-57.

Wang, Mu-Han, (1991), *Modeling Freeway Incident Clearance Time*, Masters Thesis, Northwestern University, Evanston, Illinois.

Yunker, J.M., (1993), *Optimization of Simulation Models by Genetic Algorithms: A Comparative Study*, Ph.D. Dissertation, Department of Industrial and Systems Engineering, Virginia Tech, December.

9.APPENDIX

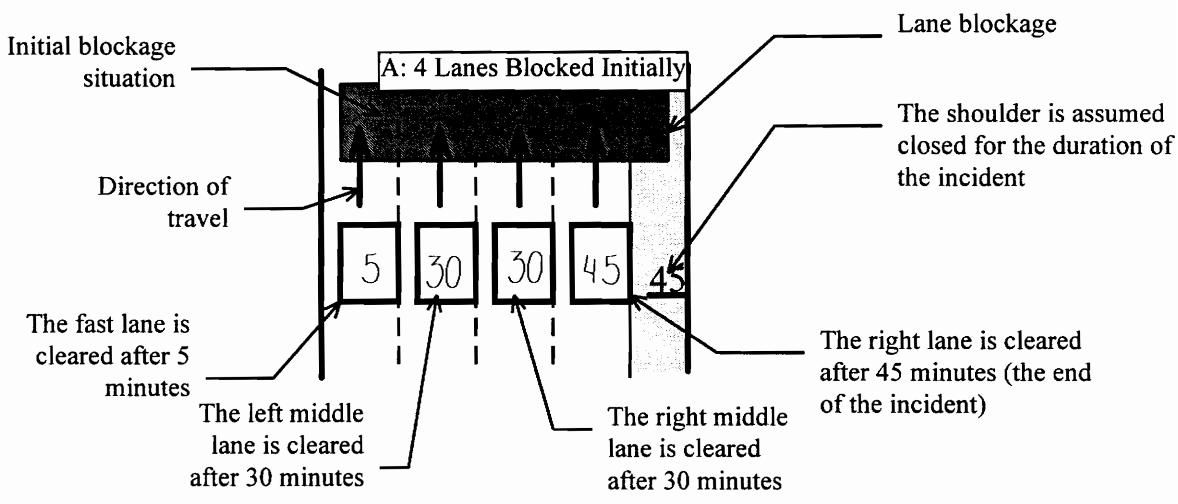
9.1. SAMPLE SURVEY PORTION

Incident Lane Clearance Survey for In-Lane Incidents

This survey is being conducted by the Virginia Tech Center for Transportation Research for incident management system development. The purpose of this survey is to compile knowledge of incident clearance timings from experts in the field of incident management. This knowledge will be used to help develop algorithms that predict incident delays.

Lane blocking incidents, as opposed to shoulder incidents, are usually cleared in phases. Many times, vehicles that are blocking lanes can be moved either by their own power or by a push or tow. Modern incident management plans are focused on restoring normal traffic flow as quickly as possible. One directive in some plans is to open as many lanes as possible during the clearance to reduce delays. This removal of vehicles and re-opening of lanes during an incident usually provides a significant decrease in delays. Part of the incident management expert system being developed at Virginia Tech is focused on forecasting delays caused by incidents. While data is available regarding incident clearance times, there is a significant lack of data on the phasing of incident removal. The goal of this survey is to collect this data in an attempt to gain a better understanding of incident clearance phasing.

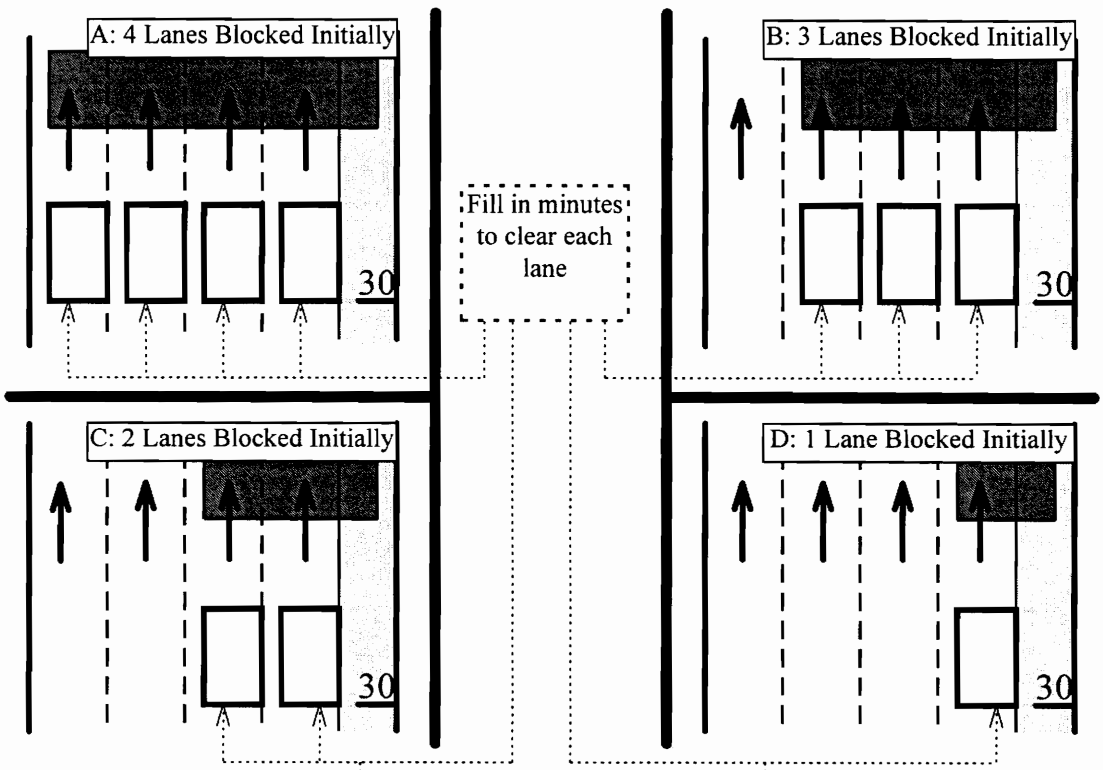
All incidents situations are unique , and there is no way to account for each one. The goal of this survey is to get a <u>general understanding</u> of how basic incident types of various durations are cleared. The incidents presented are on 4-lane and 3-lane sections of freeway with a right shoulder. It is assumed in this survey that lanes are cleared from left to right. It is also assumed that the right shoulder is blocked for the duration of the incident for the staging of incident management vehicles. For each diagram there is a box under each blocked lane. For each box, give your <u>best estimate of the average time</u> it would take to clear that lane given the blockage situation, total incident duration, and the incident type. An example is illustrated below.



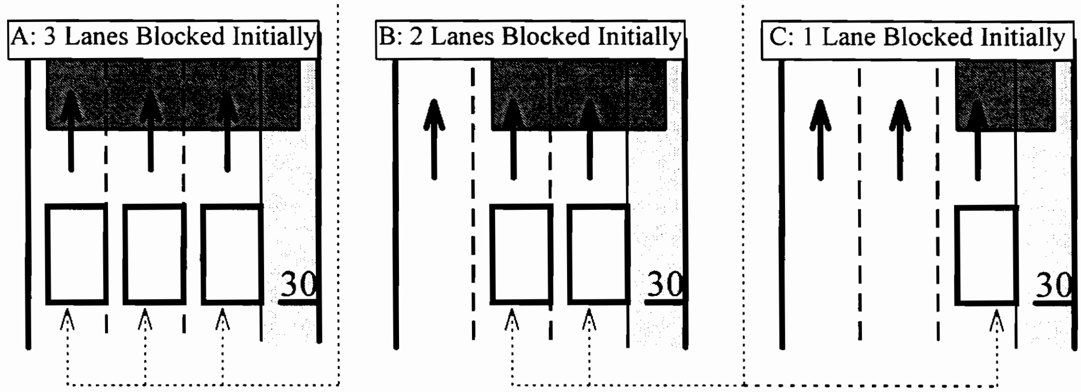
In this example, the opinion of the survey filler indicates that one lane is often opened immediately in this situation. For this incident type, the survey filler feels that two additional lanes are usually opened two thirds of the way through the incident, and that the right lane would normally be closed through the duration of the incident. Of course all incidents of this type will not exactly follow this pattern. This information just indicates that this phasing pattern is fairly common for incidents of this type.

30 Minute Property Damage Incidents(No Injury)

4-Lane Cases

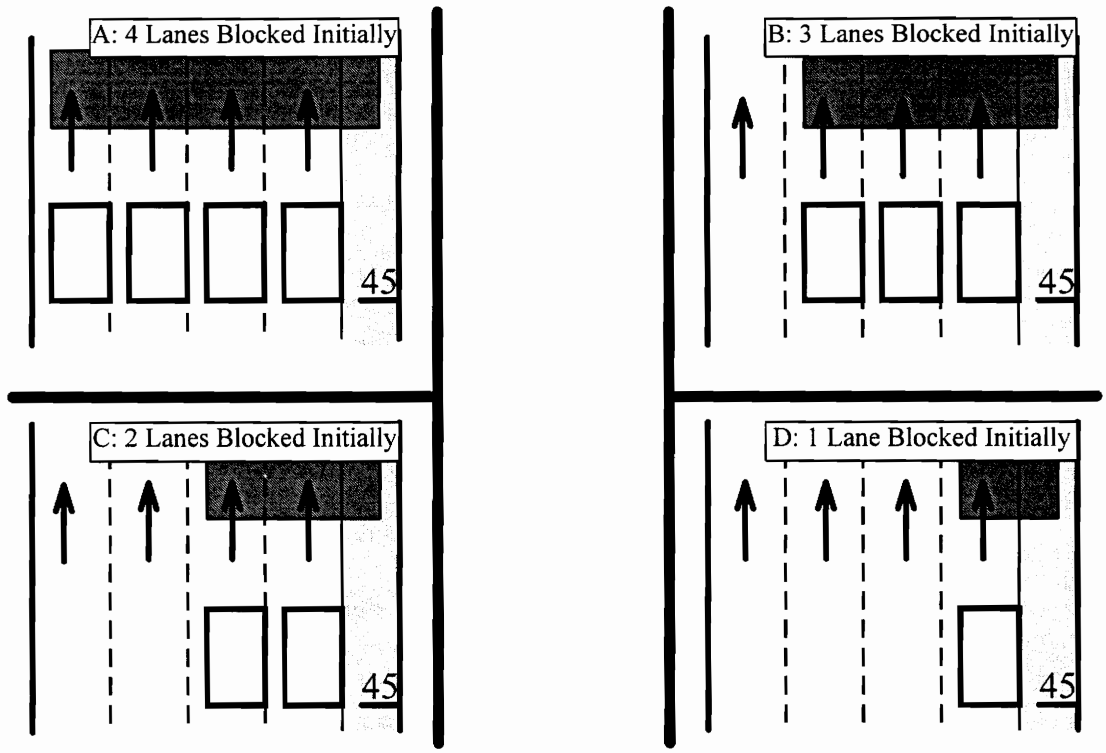


3-Lane Cases

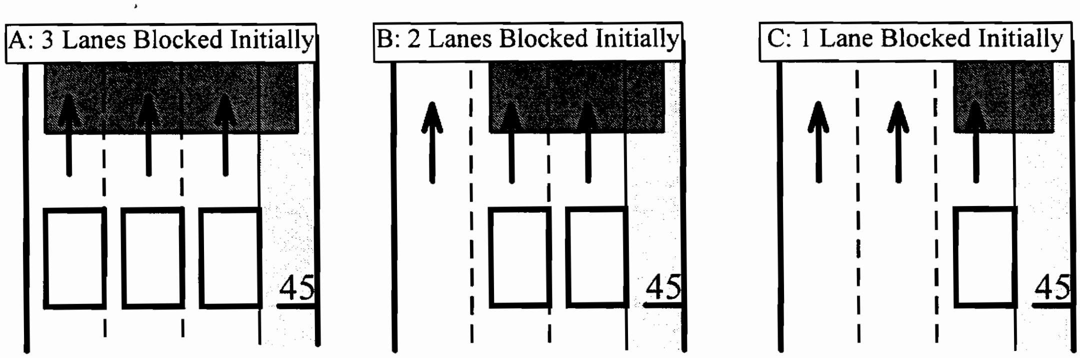


45 Minute Minor Personal Injury Incidents

4-Lane Cases

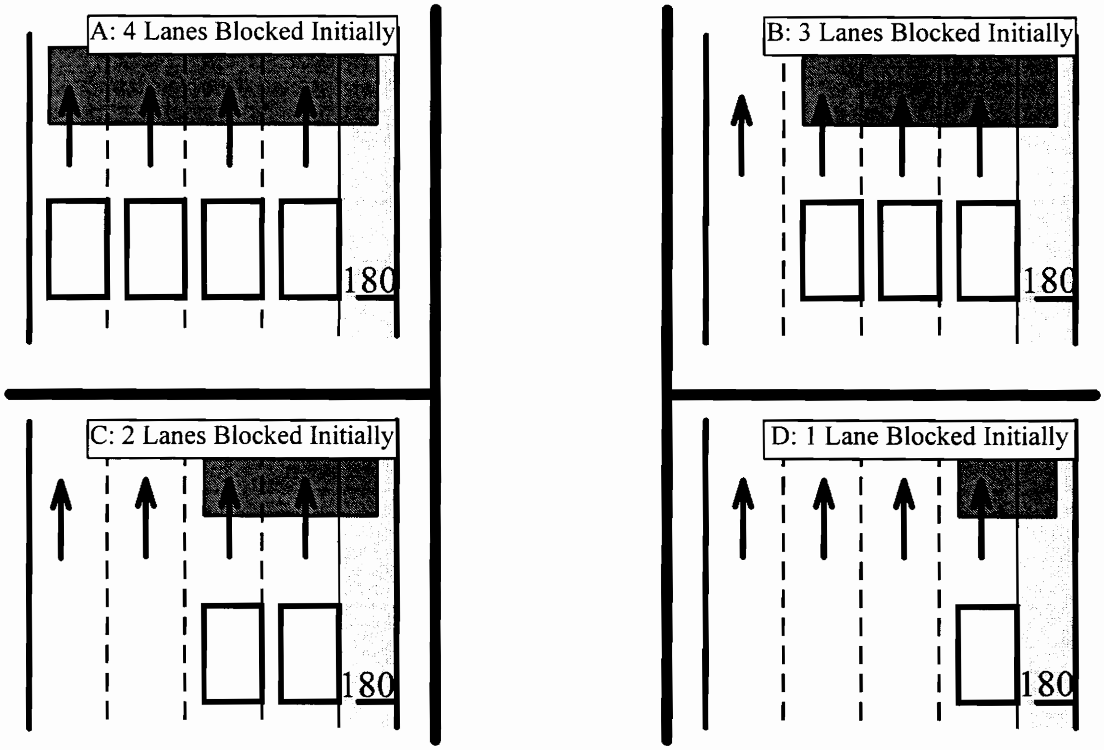


3-Lane Cases

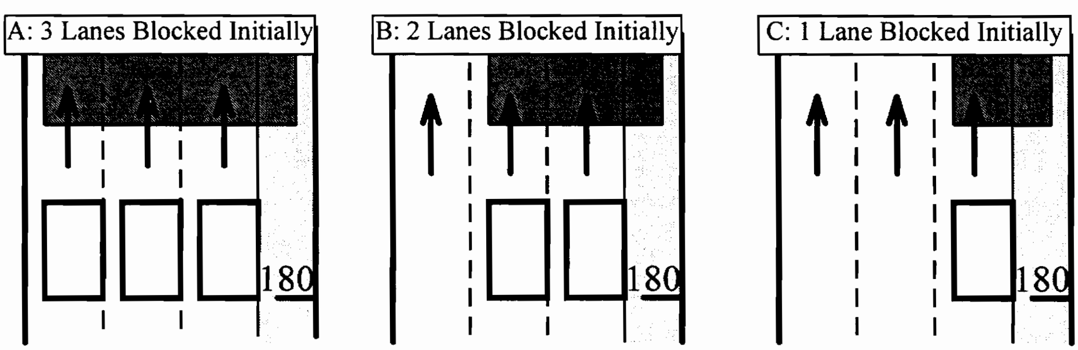


180 Minute Cargo Spill / Hazmat Incidents

4-Lane Cases



3-Lane Cases



9.2. SURVEY RESULTS SUMMARY

30-Minute Property Damage Incidents

4 - Lane Cases					**3 - Lane Cases**		
4 lanes blocked initially					3 lanes blocked initially		
Lane 1	Lane 2	Lane 3	Lane 4		Lane 1	Lane 2	Lane 3
6.135135	**9.945946**	**14.55405**	**20.85135**		**7.094595**	**12.45946**	**20.55405**
3.102471	5.391246	6.664446	9.098671		4.357858	6.020943	9.070714
3 lanes blocked initially					2 lanes blocked initially		
Lane 1	Lane 2	Lane 3	Lane 4		Lane 1	Lane 2	Lane 3
	6.648649	**11.87838**	**20.51351**			**9.351351**	**18.95946**
	3.739579	5.875309	9.328471			5.638499	9.607986
2 lanes blocked initially					1 lane blocked initially		
Lane 1	Lane 2	Lane 3	Lane 4		Lane 1	Lane 2	Lane 3
		9.890411	**18.45205**				**16.7973**
		5.579132	10.17355				10.55057
1 lane blocked intially							
Lane 1	Lane 2	Lane 3	Lane 4				
			17.75342				
			10.82407				

45-Minute Property Damage Incidents

4 - Lane Cases					**3 - Lane Cases**		
4 lanes blocked initially					3 lanes blocked initially		
Lane 1	Lane 2	Lane 3	Lane 4		Lane 1	Lane 2	Lane 3
7.821918	**12.49315**	**19.36986**	**29.82192**		**9.30137**	**17.06849**	**28.12329**
5.064425	7.220963	9.595292	13.79245		6.250966	8.745362	13.75979
3 lanes blocked initially					2 lanes blocked initially		
Lane 1	Lane 2	Lane 3	Lane 4		Lane 1	Lane 2	Lane 3
	9.083333	**16.44444**	**27.68056**			**13.93151**	**26.49315**
	6.124874	9.045875	13.99278			8.748538	14.39051
2 lanes blocked initially					1 lane blocked initially		
Lane 1	Lane 2	Lane 3	Lane 4		Lane 1	Lane 2	Lane 3
		14.76712	**26.0411**				**25.08219**
		10.01044	15.05493				15.3322
1 lane blocked intially							
Lane 1	Lane 2	Lane 3	Lane 4				
			24.27397				
			15.65271				

60-Minute Property Damage Incidents

4 - Lane Cases				3 - Lane Cases		
4 lanes blocked initially				3 lanes blocked initially		
Lane 1	Lane 2	Lane 3	Lane 4	Lane 1	Lane 2	Lane 3
10.60274	**17.15068**	**25.34247**	**37.73973**	**11.54795**	**21.73973**	**35.86301**
7.809787	10.38919	13.02585	17.59974	8.79716	12.2384	18.04879
3 lanes blocked initially				2 lanes blocked initially		
Lane 1	Lane 2	Lane 3	Lane 4	Lane 1	Lane 2	Lane 3
	12.53425	**22.50685**	**36.69863**		**18.90411**	**34.45205**
	9.933225	12.92728	18.31415		13.69323	18.82156
2 lanes blocked initially				1 lane blocked initially		
Lane 1	Lane 2	Lane 3	Lane 4	Lane 1	Lane 2	Lane 3
		20.11111	**34.43056**			**33.50685**
		13.87221	19.26525			19.72669
1 lane blocked intially						
Lane 1	Lane 2	Lane 3	Lane 4			
			32.97183			
			20.04065			

120-Minute Property Damage Incidents

4 - Lane Cases				3 - Lane Cases		
4 lanes blocked initially				3 lanes blocked initially		
Lane 1	Lane 2	Lane 3	Lane 4	Lane 1	Lane 2	Lane 3
16.64384	**25.0274**	**41.24658**	**66.09589**	**21.84932**	**50.05479**	**66.23288**
16.16011	19.74629	27.29254	40.46815	19.95667	104.1485	39.81538
3 lanes blocked initially				2 lanes blocked initially		
Lane 1	Lane 2	Lane 3	Lane 4	Lane 1	Lane 2	Lane 3
	30	**39.33333**	**65.27778**		**30.13699**	**63.49315**
	77.25484	26.27911	39.93735		23.97876	39.80582
2 lanes blocked initially				1 lane blocked initially		
Lane 1	Lane 2	Lane 3	Lane 4	Lane 1	Lane 2	Lane 3
		32.08333	**59.68056**			**63.21918**
		25.00352	40.4241			41.86096
1 lane blocked intially						
Lane 1	Lane 2	Lane 3	Lane 4			
			57.70833			
			39.97303			

30-Minute Minor Injury Incidents

4 - Lane Cases			
4 lanes blocked initially			
Lane 1	Lane 2	Lane 3	Lane 4
9.575342	**14.23288**	**19.89041**	**27.0411**
5.927614	6.588122	8.566021	7.68013
3 lanes blocked initially			
Lane 1	Lane 2	Lane 3	Lane 4
	11.0137	**18.41096**	**25.82192**
	6.115765	8.452211	8.935482
2 lanes blocked initially			
Lane 1	Lane 2	Lane 3	Lane 4
		14.83562	**24.13699**
		7.538814	7.874564
1 lane blocked intially			
Lane 1	Lane 2	Lane 3	Lane 4
			25.27778
			13.98915

3 - Lane Cases		
3 lanes blocked initially		
Lane 1	Lane 2	Lane 3
10.34722	**17.79167**	**26.04167**
6.295896	7.749489	7.695687
2 lanes blocked initially		
Lane 1	Lane 2	Lane 3
	14.5274	**24.36986**
	7.070032	8.240326
1 lane blocked initially		
Lane 1	Lane 2	Lane 3
		24.93151
		8.994333

45-Minute Minor Injury Incidents

4 - Lane Cases			
4 lanes blocked initially			
Lane 1	Lane 2	Lane 3	Lane 4
11.16667	**17.125**	**24.63889**	**35.61111**
7.238395	8.353017	10.27688	11.27075
3 lanes blocked initially			
Lane 1	Lane 2	Lane 3	Lane 4
	14.5	**22.61111**	**34.88889**
	8.607236	10.12813	11.70236
2 lanes blocked initially			
Lane 1	Lane 2	Lane 3	Lane 4
		20	**34.26389**
		10.13987	12.38467
1 lane blocked intially			
Lane 1	Lane 2	Lane 3	Lane 4
			34.44444
			12.63116

3 - Lane Cases		
3 lanes blocked initially		
Lane 1	Lane 2	Lane 3
13.125	**22.48611**	**35.83333**
8.533179	10.85369	11.75333
2 lanes blocked initially		
Lane 1	Lane 2	Lane 3
	19.375	**33.41667**
	10.47927	12.47505
1 lane blocked initially		
Lane 1	Lane 2	Lane 3
		33.61111
		13.1144

60-Minute Minor Injury Incidents

4 - Lane Cases

4 lanes blocked initially			
Lane 1	Lane 2	Lane 3	Lane 4
13.83784	**21.90541**	**31**	**43.17568**
10.155	11.32722	13.68631	16.37686
3 lanes blocked initially			
Lane 1	Lane 2	Lane 3	Lane 4
	17.24658	**28.23288**	**43.45205**
	11.15519	14.64252	15.92695
2 lanes blocked initially			
Lane 1	Lane 2	Lane 3	Lane 4
		25.13889	**41.875**
		14.53437	16.87647
1 lane blocked intially			
Lane 1	Lane 2	Lane 3	Lane 4
			42.12329
			17.31886

3 - Lane Cases

3 lanes blocked initially		
Lane 1	Lane 2	Lane 3
16.32432	**27.55405**	**43.98649**
11.90845	13.84666	16.46374
2 lanes blocked initially		
Lane 1	Lane 2	Lane 3
	24.75342	**43.58108**
	13.66746	16.50073
1 lane blocked initially		
Lane 1	Lane 2	Lane 3
		43.37838
		17.75225

120-Minute Minor Injury Incidents

4 - Lane Cases

4 lanes blocked initially			
Lane 1	Lane 2	Lane 3	Lane 4
17.43243	**28.83784**	**46.47297**	**71.14865**
15.62057	21.93267	30.32754	38.18258
3 lanes blocked initially			
Lane 1	Lane 2	Lane 3	Lane 4
	24.82432	**43.06757**	**72.22973**
	20.17958	29.30208	38.88111
2 lanes blocked initially			
Lane 1	Lane 2	Lane 3	Lane 4
		39.31507	**71.84932**
		27.992	39.15343
1 lane blocked intially			
Lane 1	Lane 2	Lane 3	Lane 4
			67.90411
			40.63187

3 - Lane Cases

3 lanes blocked initially		
Lane 1	Lane 2	Lane 3
21.95946	**41.59459**	**70.74324**
19.61649	26.60823	37.75083
2 lanes blocked initially		
Lane 1	Lane 2	Lane 3
	36.95946	**69.93243**
	25.65335	38.5836
1 lane blocked initially		
Lane 1	Lane 2	Lane 3
		72.16216
		40.08622

45-Minute Severe Injury Incidents

4 - Lane Cases			
4 lanes blocked initially			
Lane 1	Lane 2	Lane 3	Lane 4
16.89189	**25.09459**	**34.27027**	**41.08108**
10.74941	11.99791	10.91097	13.3773
3 lanes blocked initially			
Lane 1	Lane 2	Lane 3	Lane 4
	22.43243	**31.37838**	**40.39189**
	12.5579	11.80516	13.24151
2 lanes blocked initially			
Lane 1	Lane 2	Lane 3	Lane 4
		27.87671	**45.06849**
		13.09468	51.02732
1 lane blocked intially			
Lane 1	Lane 2	Lane 3	Lane 4
			38.63014
			14.19752

3 - Lane Cases		
3 lanes blocked initially		
Lane 1	Lane 2	Lane 3
18.65753	**29.9863**	**40.05479**
10.5951	11.21939	13.83724
2 lanes blocked initially		
Lane 1	Lane 2	Lane 3
	26.78082	**39.0274**
	12.1441	14.07517
1 lane blocked initially		
Lane 1	Lane 2	Lane 3
		39.65753
		13.67603

60-Minute Severe Injury Incidents

4 - Lane Cases			
4 lanes blocked initially			
Lane 1	Lane 2	Lane 3	Lane 4
18.87838	**29.5**	**41.04054**	**49.45946**
11.73095	14.1888	15.65855	14.63172
3 lanes blocked initially			
Lane 1	Lane 2	Lane 3	Lane 4
	25.09589	**37.41096**	**48.34247**
	14.22455	16.72144	14.20702
2 lanes blocked initially			
Lane 1	Lane 2	Lane 3	Lane 4
		43.74324	**47.44595**
		71.20029	15.13333
1 lane blocked intially			
Lane 1	Lane 2	Lane 3	Lane 4
			48.2027
			15.47422

3 - Lane Cases		
3 lanes blocked initially		
Lane 1	Lane 2	Lane 3
20.36486	**35.01351**	**48.83784**
12.62155	15.86803	14.14604
2 lanes blocked initially		
Lane 1	Lane 2	Lane 3
	33.25676	**47.64865**
	16.62649	15.62263
1 lane blocked initially		
Lane 1	Lane 2	Lane 3
		48.40541
		14.94549

120-Minute Severe Injury Incidents

4 - Lane Cases					**3 - Lane Cases**		
4 lanes blocked initially					3 lanes blocked initially		
Lane 1	Lane 2	Lane 3	Lane 4		Lane 1	Lane 2	Lane 3
27.77027	**45.67568**	**67.25676**	**86.35135**		**33.31081**	**60.27027**	**84.86486**
22.2554	28.56529	37.31248	37.56544		26.0305	34.46153	36.5833
3 lanes blocked initially					2 lanes blocked initially		
Lane 1	Lane 2	Lane 3	Lane 4		Lane 1	Lane 2	Lane 3
	39.24658	**60.82192**	**84.12162**			**57.02703**	**83.31081**
	27.03741	34.47033	37.71158			36.09272	37.32748
2 lanes blocked initially					1 lane blocked initially		
Lane 1	Lane 2	Lane 3	Lane 4		Lane 1	Lane 2	Lane 3
		58.78378	**85.2027**				**85.06757**
		36.03471	37.80301				38.16415
1 lane blocked intially							
Lane 1	Lane 2	Lane 3	Lane 4				
			85				
			37.91664				

120-Minute Cargo Spill / Hazmat Incidents

4 - Lane Cases					**3 - Lane Cases**		
4 lanes blocked initially					3 lanes blocked initially		
Lane 1	Lane 2	Lane 3	Lane 4		Lane 1	Lane 2	Lane 3
55.27397	**68.21918**	**89.17808**	**101.8493**		**59.86301**	**87.46575**	**97.87671**
38.87203	39.35566	35.45355	30.63571		37.23026	64.56114	34.86997
3 lanes blocked initially					2 lanes blocked initially		
Lane 1	Lane 2	Lane 3	Lane 4		Lane 1	Lane 2	Lane 3
	68.42466	**86.36986**	**99.72603**			**90.13699**	**101.0274**
	38.612	35.99055	31.98188			101.9258	31.06846
2 lanes blocked initially					1 lane blocked initially		
Lane 1	Lane 2	Lane 3	Lane 4		Lane 1	Lane 2	Lane 3
		81.91781	**99.65753**				**101.3699**
		36.06158	32.87528				31.74618
1 lane blocked intially							
Lane 1	Lane 2	Lane 3	Lane 4				
			100.6164				
			31.92814				

180-Minute Cargo Spill / Hazmat Incidents

4 - Lane Cases			
4 lanes blocked initially			
Lane 1	Lane 2	Lane 3	Lane 4
73.21918	**98.49315**	**129.863**	**158.6575**
54.20748	54.98538	54.86076	70.47778
3 lanes blocked initially			
Lane 1	Lane 2	Lane 3	Lane 4
	89.31507	**122.8082**	**148.9726**
	54.97042	55.47764	51.01715
2 lanes blocked initially			
Lane 1	Lane 2	Lane 3	Lane 4
		119.1667	**148.125**
		55.9678	52.79136
1 lane blocked intially			
Lane 1	Lane 2	Lane 3	Lane 4
			147.2917
			53.72031

3 - Lane Cases		
3 lanes blocked initially		
Lane 1	Lane 2	Lane 3
79.38356	**115.8219**	**149.7945**
55.01859	53.84529	50.03775
2 lanes blocked initially		
Lane 1	Lane 2	Lane 3
	110.2055	**148.9041**
	55.16666	50.33393
1 lane blocked initially		
Lane 1	Lane 2	Lane 3
		147.3288
		52.0422

VITA

Scott A. Mastbrook was born on June 19, 1972 in Silver Spring, Maryland. He attended Virginia Tech in Blacksburg, Virginia from 1990 through 1994 and received his B.S. in Civil Engineering in May, 1994. He attended Virginia Tech for graduate studies from 1994 through December of 1995. He joined Kaman Sciences Corporation in February of 1996 as a transportation engineer. His professional interests include traffic management systems, traffic surveillance, communications, driver information systems, and Internet applications for the transportation industry.