This is to certify that the

dissertation entitled

Travel Time Predictions Under Dynamic Route Guidance With A Recursive Adaptive Algorithm

presented by

Chronis Stamatiadis

has been accepted towards fulfillment of the requirements for

Doctor of Philosophy degree in Civil Engineering

William C. Taylor
Major professor

Date August 13, 1992

TRAVEL TIME PREDICTIONS UNDER DYNAMIC ROUTE GUIDANCE WITH A RECURSIVE ADAPTIVE ALGORITHM

By

Chronis Stamatiadis

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Civil and Environmental Engineering

1992

ABSTRACT

TRAVEL TIME PREDICTIONS UNDER DYNAMIC ROUTE GUIDANCE WITH A RECURSIVE ADAPTIVE ALGORITHM

By
Chronis Stamatiadis

The key element of a dynamic route guidance system is the knowledge of the travel times on the links in the network, for all time periods in a planning horizon. In real time systems, where route guidance will be performed on-line, the travel times required for the evaluation of alternative routes, have not yet occurred. Therefore the ability to make predictions of such travel times with reasonable accuracy is essential to the route evaluation process.

Many researchers have developed models to estimate future travel times with the usage of traffic assignment models based on some equilibrium principal. The efficiency of such models is based on the assumption that drivers will follow the suggestions given by the guidance system, and that a large percentage of traffic is equipped to receive such information from the system. These conditions may be realistic if guidance systems have been proven efficient enough to attract drivers to actually use them.

Thus, alternative models, that will still operate in real time, have to be devised which will give us the capability of short term travel time predictions. Such models will have to be based on historical data, and they must represent the dynamics of travel times, including variations due to traffic incidents, with sufficient accuracy to be used by the motorists. In this study we examined the

effectiveness of a route guidance system based on the travel time predictions performed by such a model. Predictions are made through the application of a recursive identification algorithm for all the links in the network. The prediction model utilizes a recursive least squares (RLS) algorithm and its input variables include both historical information of the travel time on the link under consideration, and its average travel time in the current time period based on observations from previous days. In addition, travel time information from upstream and downstream links is incorporated so information about evolving traffic waves is considered before such waves reach the link.

The prediction model is applied to a small network with 21 decision nodes and 38 links, and its performance is examined under normal and congested traffic conditions. We consider different structures of the model for improving travel time predictions especially in the case of congestion due to a traffic incident. We also experimented with different frequency rates with which the system is updating predictions on travel times. Obviously, there is a trade off between the number of predictions that we can perform with some reasonable accuracy and the rate with which we update such predictions. Finally different scenarios for market penetration of route guidance systems are considered, for examining how travel time savings change as the percentage of equipped vehicles increases.

To my wife Filio and
to my parents Neofytos and Polymnia

ACKNOWLEDGMENTS

I would like to express my great intellectual debt to my academic advisor Dr. William C. Taylor, for his guidance and advise during the course of my studies and this research.

I also want to express my appreciation to Dr. Francis X. Mckelvey, Dr. Richard W. Lyles, and Dr. Habib Salehi for serving as members on the guidance committee and for their assistance. My gratitude is also extended to Dr. Majid Nayeri for his valuable assistance.

Finally, I want to express my thanks to my wife Filio and my brother Niki, for providing their support and encouragement. My parents have provided continuous encouragement and moral support through the years and my appreciation for them cannot be expressed with words.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

During recent decades, an enormous effort has been made to improve the mobility of individuals, especially in urban areas where demand is concentrated and supply of transportation facilities is restricted. This effort included massive expenditures for providing the means to satisfy such demand, like the construction of the freeway system and the rapid transit system in most of the major metropolitan areas. The hope behind the implementation of such immense projects was that demand will be accommodated by provision of more highways. At the same time operational characteristics of vehicles are greatly improved, regarding safety, emissions, efficiency, and performance. The traffic management systems like traffic signal systems and ramp metering strategies, have also been improved so they respond to variations in traffic demand and conditions.

However, travel demand has been increasing continuously due to socioeconomic factors like the increase of car ownership - and therefore an increase in access to a private automobile by more individuals - as well as the increased need to travel in today's urban areas where the spatial distribution of activities is widely spread. In many metropolitan areas the growth of automobile usage has already outpaced infrastructure investments, and in the near future it is expected that this will also be the case in several other high growth urban areas. Expected levels of delay are measured in billions of vehicle hours, and the increased usage of automobile will result in an increase in the number of traffic accidents (Grenzeback and Woodle 1992). The estimated losses from such delays and accidents, are claimed to be a few hundred billion dollars annually. Traffic congestion is regarded as one of the major problems of urban areas and it is often compared with societal problems like crime, access to health care, or housing.

On the other hand economic and environmental control policies place limits on increasing the physical road capacity. Both social and environmental consequences of such projects may be far more severe than their expected benefits. In addition, congestion relief benefits may not last very long, due to the rate of demand for even more mobility. These two contradicting phenomena - an increase in demand and an inability to increase the capacity of traffic networks to serve this demand - contribute to the creation of traffic congestion, especially on commuting corridors during peak periods.

For these reasons, the attention in transportation planning has turned to studies of how to effectively use the existing capacity, rather than constructing new freeways. The potential of using modern technologies, like advances in communications and computers, as well as a higher level of sophisticated control systems and sensors, are greatly enhancing the possibility of success of such an approach. The usage of such technologies in the transportation system is housed under the term Intelligent Vehicle - Highway Systems (IVHS), and the technologies used are referred to as IVHS technologies.

IVHS includes the systems that use advanced technologies to control and improve mobility on a traffic network. For example, traffic operations can be improved when traffic management is based on processing of real time information regarding the travel conditions on the elements of a network. Without real time information the decisions of individual travelers regarding the time that they will realize their trip, the mode that they will use and the route that will be followed are often based on information that is incomplete and not accurate. When real time information is used, such decisions will approach the optimal ones and the development and application of new and sophisticated control strategies will be possible.

While IVHS as defined above is quite comprehensive, there are four subsystems that can be distinguished and are generally accepted:

(1) The Advanced Driver Information Systems (ADIS);
(2) The Advanced Traffic Management Systems (ATMS);
(3) The Commercial Vehicle Operations; (CVO); and
(4) The Advanced Vehicle Control Systems (AVCS);

Technologies included in the ADIS category are the ones that can navigate the driver through a network, provide information about alternative routes and congestion and reveal the location of service stations, restaurants, or rest areas. Driver information systems can vary based on (Koutsopoulos and Yoblanski 1991):

(1) the ability to provide real time information;
(2) the type of information that they provide;

(3) the ability to address a single vehicle or all the vehicles at a location on the network; and

(4) the communication capabilities of the system linking the vehicles on the road with the traffic control center, ranging from no communication, to one-way or two-way communication.

On the other hand, ATMS includes traffic management systems that operate in real time, so the system will respond to changes in the traffic demand, or even anticipate when and where such changes will occur and apply the appropriate strategy in order to avoid congestion.

Commercial vehicle operations are based on ATMS and ADIS but are concerned mostly with issues like weigh in motion, vehicle tracking, efficient vehicle dispatching, and timely pickups or deliveries. These are operations which can be performed more efficiently through IVHS. On the other, hand advanced vehicle control systems (AVCS) will not make direct use of information about traffic conditions, but they will improve safety and potentially the capacity of a network by assuming partial or complete control of the vehicle. AVCS includes systems that aim to help the driver in the task of driving, like lane keeping systems, enhanced night vision systems, adaptive cruise control, and so forth.

In this study we are concerned with the first subsystem which is closely related with vehicle navigation and route guidance problems of traffic on a network. The key element of a route guidance system, is the knowledge of the "travel cost" on the links in the network, for all time periods in a planning horizon. In real time systems, where route guidance will be performed on-line, these travel costs have not yet occurred. A reasonable assumption is that the aspect that travelers will be willing to optimize in their trip is their travel time. Therefore the ability to make predictions for such travel times with reasonable accuracy is essential to the route evaluation process.

In the following a review of the existing literature relevant to the problem of dynamic route guidance, and short term prediction of travel times is presented. In the next chapter the proposed travel time prediction model is discussed. In chapter 4, a traffic simulation program is presented, which will be used for testing the prediction model. The structure of the prediction model is examined in chapter 5, while the effectiveness of the model incorporated within a route guidance system is presented in chapter 6. Finally chapter 7 presents the conclusions drawn from this research.

Chapter 2

Literature Review

2.1 Route Guidance Systems

Route guidance systems can be defined as those systems that, when the driver provides the destination of his/her trip, are capable of guiding him/her through the most desirable route. Depending on the way that the system represents traffic and the way it performs the routing, we can distinguish three alternative route guidance systems (Chen and Underwood 1991):

(1) *Static guidance systems*, where the costs of the links of the network are assumed to be independent of the time and the amount of traffic that traverse them (constant link travel cost).

(2) *Simple dynamic route guidance systems*, where link costs are allowed to vary with time but traffic is still routed through the network based on a static approach to route selection. This is the case when the guided vehicle receives information only at the beginning of the trip regarding current link costs.

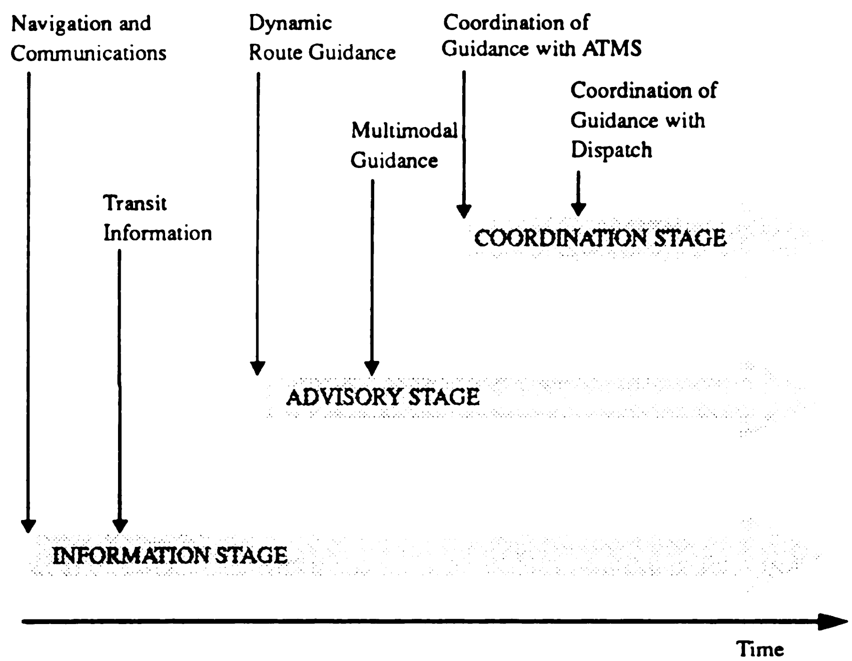
(3) *Dynamic route guidance systems*, where the routing of a vehicle depends not only on the time of departure but on the location of the vehicle as well. In this class of guidance systems link costs vary with time and the most desirable route is reexamined each time the vehicle reaches a

new decision node.

In addition to this classification, depending on the type of information that guidance systems use as a base, route suggestions may be further classified to *responsive*, in the cases that use real time data, and *unresponsive* if they use only historical data obtained from previous days.

For the first two classes of route guidance systems, the determination of the most desirable path can be obtained by using a fastest path algorithm, like the Dijkstra's algorithm, based on Bellman's principal of optimality. In dynamic route guidance systems, such static algorithms will not be sufficient. In this case more complex, and consequently of higher computational effort, algorithms are required, which are capable of finding the minimum path in dynamic networks (Kaufman and Smith 1990).

The most probable scenario of implementing a dynamic route guidance system is depicted in Figure 2.1 (Chen and Undewood 1991), which also follows the market penetration scenario of such systems. Initially ADIS will serve only as a navigation system. Such systems are already successfully in use, like the Advanced Mobile Traffic Information and Communication Systems and the Road Autonomous Communication Systems in Japan or the DEMETER in Europe (French 1990). In this case, information is not provided in real time or the information system performs only "yellow pages" functions. Actual conditions on the network are not known to the driver. Such systems often operate by relating the current position of a vehicle obtained by either the Geographic Position System (GPS) or by dead reckoning, with a digital map. The geographic information of the digital map is stored in a CD-ROM, and navigation advice is often displayed on an in-vehicle monitor. Communications between



(Source: Chen & Underwood 1991)

Figure 2.1: Implementation of ADIS and ATMS.

the vehicle and the traffic control center at this level of systems is not required.

In the next stage, where dynamic route guidance will be possible, information will be broadcast by the system in real time. As is shown in Figure 2.1, at this stage ADIS plays only an advisory role on which is the current optimal route. When information about current and projected traffic conditions is used for guiding traffic, this information has to be updated continually. In this case communication links have to be established, so the information transmitted by the traffic management center (system) will be received by the vehicles on the network which are equipped with some special receivers. The system has to be capable of collecting and processing information about the traffic conditions continuously on the elements of the network and then broadcast the processed information back to the traffic on the network.

In a route guidance system, information received by the equipped vehicles may be either in the form of descriptions about the traffic conditions on the elements of the network, in which case the driver will have to select the optimum path based on his judgement, or in the form of suggestions on which is the current optimum route to follow, or the best time to start the trip. While it could be argued that guidance in the descriptive form is meaningless when with a simple computing device prescriptive information can be obtained, such guidance could be more helpful in the case where the driver wants to avoid a part of the city for security reasons, or a driver wants to stay on the freeway system.

Field tests of such systems are under way in the United States, Europe, and

Japan. For example the Autoguide system in London is the first dynamic route guidance system. Based on real time information, minimum travel time routes are calculated and updated regularly by the traffic control center. This information is transmitted to the network, so equipped vehicles can receive it when they approach "decision nodes". The decision nodes are intersections equipped with electronic signposts which transmit directions to the equipped vehicles, related with the movements of the intersection, i.e. straight ahead, turn right or turn left. The optimum paths are selected based on average conditions on the network for a specific time, day and weather conditions, as well as current information regarding unusual events. This information is updated in rather long time intervals (every 15 minutes), except in the case when an incident has occurred, and in this situation the frequency of updating the transmitted information is increased. The operation of Autoguide is reactive to traffic conditions, meaning that it will divert traffic only when congestion is already evident. A similar system Ali-Scout (or as it is lately renamed to Euro-Scout) is also under testing in the city of West Berlin, and soon it will be tested in Oakland County, Michigan. This system combines the characteristics of Autoguide with a digital map for displaying the vehicle location and the directions to be followed (Boyce 1989, French 1990).

Here it should be noted that the concept of such systems is not new; The idea of alleviation of congestion based on guiding traffic had its origins back in the 1960's. These efforts however were either not successful due to the limited capabilities of the available technologies at that time or not feasible due to the excessive cost of implementing them (Rosen et al., 1970).

In the last stage, dynamic route guidance systems will be coordinated with

ATMS, through the use of traffic assignment models. The route guidance systems described in the advisory stage, in a situation with congestion, will divert all equipped traffic to one single new optimum path. Without more sophisticated coordination, which can be achieved through dynamic traffic assignment models, congestion will be created on the alternative routes. This is because there is no consideration of the effect of the diverted traffic on the travel conditions of the new path. Such strategies will be effective only while the percentage of the equipped vehicles is low. When a large number of vehicles are equipped, everybody will be receiving the same information which will result in the possible transfer of the congestion to another part of the network. Such transfer of congestion is often referred as guidance system induced congestion (Boyce 1989, Kaufman et al. 1990). Although extensive research results are available for the static situation, where traffic demand is considered constant, there are just a handful of theoretical configurations for the dynamic approach.

2.2 Expected Benefits of IVHS

The main benefit expected from the implementation of IVHS is alleviation of congestion that costs society and the environment a great deal. In addition, new technologies are expected to contribute to improvements in safety and productivity of the transportation system. IVHS and mostly AVCS are supposed to contribute to accident prevention in contrast to most accident countermeasures applied currently, that aim to reduce the severity of traffic accidents. Early detection of pending hazards, which is one of the promises of AVCS, will result in providing the driver with the critical time required to react and avoid the accident. Nevertheless potential opposite effects have to be acknowledged. There are some inherent risks of automated highways, and possible distraction of the driver from current in-vehicle information devices may jeopardize his/her safety.

2.2.1 Alleviation of Congestion

In the literature, two types of congestion are conveniently distinguished:

(1) Recurrent congestion that occurs during certain time periods of the day or certain days of the year, due to a high demand that exceeds the capacity of the elements of the network; and

(2) Non recurrent congestion that occurs due to unexpected circumstances like traffic accidents or vehicle breakdowns, which can greatly affect the capacity of the system. Such a reduction in capacity leads to congestion, and it is time independent.

It is claimed that IVHS will reduce delays due to both recurrent and non recurrent congestion. Up to date results suggest that benefits of ADIS systems will be marginal under conditions of recurrent congestion (Al-Deek and Kanafani 1991). However the benefits that will accrue in the case of non recurrent congestion are expected to be much more significant, since the occurrence of such congestion is random, and thus its effects cannot be anticipated by the unequipped driver. On the other hand, some of the effects of recurrent congestion can be anticipated even without ADIS, when drivers have previous experience on the network. Provision of information about the development of congestion to the drivers in real time can be used for diversion of traffic either in space or in time. For example, drivers can be directed to either divert from a congested link of the system to an alternative one with higher level of service, or they can simply choose to postpone their time of departure. The basic difference between the two diversion options lies in the time that they can be implemented. While diversion in time has to be administered before the beginning of the trip, diversion in space can be performed before or during the trip.

Here we consider the case where in the situation of congestion, appropriately equipped vehicles will divert from the originally selected path, en-route, based on "intelligent" suggestions made by the system. However, it should be noted that in large metropolitan areas, congestion is a daily phenomenon not just for a short time period but for several hours. In such areas congestion is spilled even on beltways which were originally designed for through traffic in order to avoid local effects. Before we appraise the benefits of IVHS it is useful to consider if there is any capacity left on the network that is not used, and its usage could improve the situation. If the answer is no, then information on traffic conditions on the network, even in real time, will not contribute anything to the travelers. Under this condition, the application of advanced technologies may help to lead travelers to alternative modes like rapid transit or ride sharing which will possibly qualify them to use HOV facilities.

2.2.2 Findings on Effectiveness of Route Guidance Systems

The expected benefits of IVHS, and more specifically of route guidance systems, will be a reduction in congestion. Several studies have dealt with the effectiveness of route guidance systems, based mostly on simulation models. Depending on the configuration of the route guidance system, reductions in average trip time ranging from 2 to 50 percent were noted (G.A.O. 1991). Although the possible effectiveness of route guidance systems has such a wide range, one aspect that all the studies which examined different levels of market penetration agree is that for high market penetration levels, equipped traffic does not get as much travel time improvement as at low levels.

The issue of reliability of the broadcast information was examined by Chen and Mahmassani (1991), where the supplied travel time was compared to actual travel times experienced in the network after the information was broadcast. Because of the dynamic character of traffic on a network a recommended path may actually be less than optimal as traffic congestion evolves. Furthermore drivers may not switch from their current path unless they know that the time savings from the alternative route are meaningful. With a series of simulations the authors examined various combinations from a set of variables:

(1) different scenarios of information sources including no information at all, information only at the beginning of the trip, only enroute information, or both;

(2) two different behavioral rules for drivers switching routes one where drivers are willing to accept any improvement in their travel, regardless of how small (myopic rule) and one where the new path has to be more than 20% better than the current path and at least one minute less; and

(3) different levels of market penetration.

In their results it was found that reliability of information worsens as market penetration increases when drivers behave myopically and both sources of information are available. For the other cases the reliability of information is better at 10% of market penetration than at 100%, but no trends are obvious for the cases in between. Also at high market penetration, equipped traffic performs better than the rest of the traffic when the behavioral rule is employed, a fact that is attributed to the less "unproductive" path switches due to the indifference band of 20% improvement.

When the reliability of the broadcast information is high, users with access to information experience meaningful travel time improvements. This is true since information is reliable if there have not been many switches that would result in congested routes. Similar results were reported by Hamerslag and van Berkum (1991) where, after simulating a set of different urban networks, it was found that the amount of car-kilometers traveled decreases as the uncertainty of the information available to the drivers decreases.

In an extension of the same study by the authors (Mahmassani and Chen 1991) it was found that when both sources of information are used this has positive results only at low levels of market penetration. These benefits decrease rapidly as market penetration increases and often are less than those obtained with one source of information only. Also home based information only, at high levels of market penetration has negative effects, while benefits due to en-route information appear to be more robust.

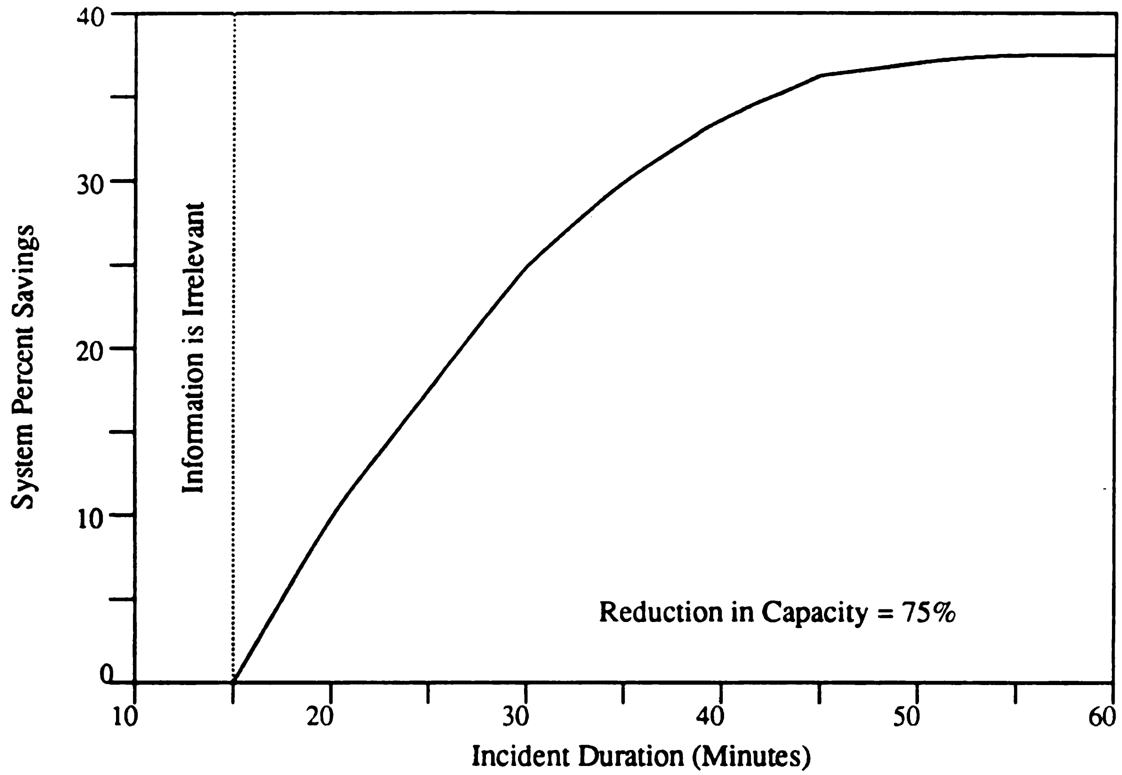
Koutsopoulos and Lotan (1989) investigated the effect of variables like level and amount of information, level of congestion, spatial extent of information and portion of informed drivers, on the system performance under a route guidance system.The results from this study indicated that the total benefits that will accrue from advanced driver information systems will be marginal. Based on simulation results from a small network, the average travel time benefits when perfect information is available was found to be only 4.4%. In the study it was noted that as the level of available information increases, the travel time of the shortest path also increases, since more drivers can make more intelligent decisions; and thus the alternative routes become more crowded. Also as traffic demand increases the benefits per driver (both equipped and unequipped) increase moderately, until the volume over the capacity ratio assumes values in the vicinity of 1.5. After this point, such benefits start to decrease. The benefits experienced by equipped drivers appear to decrease as the congestion level increases. Thus the value of the information provided to the drivers decreases too, since the opportunity to identify better paths is reduced. Also, when perfect information is available for the entire network, travel times decreases by approximately 4% as compared to the case when travel times are known only for the main routes.

Koutsopoulos and Yoblanski (1991) examined design aspects of a route guidance system, with a low market penetration of IVHS. Variables included the location of information nodes (locations where equipped vehicles can update information), frequency of information updates made by the system, and the intelligence of the system. The system was assumed to be of either low or high intelligence. In the case of low intelligence, route suggestions by the system were based on current traffic conditions on the network on the basis that such conditions will not change and in the case of incident occurrences the system would take into account only delays due to queues. In the case of high intelligence, expected delays due to incidents are taken into account based on expected time of arrival at the queue. Also in this scenario, minimum paths were calculated based on projected traffic conditions.

Their results were based on a small network simulated with a microscopic simulation program. From their findings it is indicated that there are no significant reductions in the average travel time of the network that was simulated by the introduction of ADIS, except in the case of a traffic accident. When the density of information nodes increases, the intelligence of the system has little bearing on the average travel times of the equipped vehicles. However the reliability of the trip (the standard deviation of the travel time) is increased when each node is an information node. Also, a frequency of updating information every 3 minutes was almost as good as updating information immediately, while for longer than 3 minutes between information updates, the route guidance system becomes less effective.

Al-Deek and Kanafani (1991) have examined the effect of diverting traffic on the time savings for a small network with two alternative paths to a destination, in the case when an incident occurs on the higher capacity path. Parameters like the capacity of the alternative path, the amount of traffic that arrives at the decision point and the duration, severity and location of the incident were considered. Based on this simple network, the authors found that travel time savings of equipped traffic decreases rapidly once a queue starts forming on the alternative route. When the proportion of diverted traffic exceeds a critical value - which depends on the demand arriving at the decision point and the capacity of the alternative path - benefits to the equipped traffic decrease. When the duration of the incident increases, the benefits from the route guidance system also increase, but not in a linear manner. When incidents are very long (in the example network over 60 minutes) the increase in benefits vanishes, and for very short incidents (less than 15 minutes), there is no benefit of a guidance system (Figure 2.2). The same pattern is true for the severity of the incident, with severity defined as reduction of capacity. It is noted that for severe incidents (with more than 60% reduction in capacity), and of short duration, benefits are more sensitive to the duration than to the severity of the incident. If the alternative route is very long it is never used, while when the travel time of the alternative route has competitive free flow travel time to the free flow travel time of the route with the incident, the benefits from the system are maximized.

Halati and Boyce (1991) have also examined the effectiveness of various in-vehicle navigation systems in the situation of a traffic incident, but on a realistic network (Irvine network, Orange County, California). They simulated different types of drivers, including drivers that are familiar or unfamiliar with the network, drivers that are equipped or unequipped, and drivers that are compliant or non-compliant with the route suggestions supplied by the guidance system. The guidance systems that were considered included two descriptive systems, one with a static map and one with a map system and the added capability of identifying congested segments of the network, and two prescriptive systems where the best path was given to the driver. The two prescriptive systems differ in the aspect that the more advanced one includes a map display. The authors found that while there was no significant differences between the first two systems, there was substantial improvement to the system operations when the dynamic route guidance system was employed. The incident occurred on a freeway section of the network, and the freeway speeds were improved by 0.9%, 11.8% and 53% for 10%, 30% and 50% market penetration of the guidance system respectively. The dynamic route guidance system improved the performance of the surface network too, where the improvements for the entire network were found to be 4.5%, 17.2% and 37.2% for 10%, 30% and 50% market penetration levels (Table 2.1).



(Source Al-Deek and Kanafani 1991)

Figure 2.2: Sensitivity of System Benefits to Incident Duration

Table 2.1: Percent of Speed Improvement with Market Penetration of Prescriptive Route Guidance Systems.

	Market Penetration		
	10%	20%	50%
Freeway Subsystem	0.9	11.8	53.0
Surface Street Subsystem	9.6	23.9	26.6
Entire Network	4.5	17.2	37.2

In the more advanced system, where drivers also have a map display, it was assumed that more drivers divert from the congested path but less drivers follow the alternative route suggested by the system. The result was a slight worsening for low levels of market penetration, while when 50% of the drivers were equipped, the navigation system appears to compensate for the lower compliance rate, and there is an improvement of approximately 4% over the simple dynamic guidance system.

2.3 Prediction of Traffic Conditions

The problem that is recognized in the existing literature is that at least in the last two stages of implementation of route guidance systems where dynamic route guidance will be employed, the required optimum routes are based on traffic conditions that vary with time, and in real time, these conditions have not yet occurred. A more sophisticated handling of such situations will be to guide traffic through a given network with a guidance system which is based not only on the current traffic conditions but also on expected conditions. In other words, the route selection process must be based not only on where traffic is currently, but where traffic will be - or more precisely where it is expected to be - in the short term future. Such estimations of future traffic conditions have to be based on a sophisticated prediction model which will assess the impacts of traffic variations due to both recurrent and non recurrent congestion.

Obviously, two fundamental characteristics of such prediction systems are their operational speed and their responsiveness to actual traffic conditions. Their speed should be at least as fast as the real traffic system, or even faster. This is required, so there will be enough time for the prediction system to perform the predictions, and for the route guidance system to evaluate the alternative routes in real time. Nevertheless, the high speed of the prediction system should not be achieved at the cost of the accuracy of the predictions. In addition the prediction system must be able to respond to both expected and unexpected changes in the traffic conditions, i.e. recurrent or non recurrent congestion. This can be achieved only if the system is capable of collecting the necessary data in real time, and processing such data with the required speed and accuracy, so it will periodically update the information transmitted to the drivers. This way the route guidance system itself will be responsive to such changes in traffic conditions (Chen and Underwood 1991).

There are two strands followed in the development of prediction models, and each may be best applicable at different levels of market penetration of dynamic route guidance systems. The first is based on the traffic assignment models, which are believed to be most efficient when the majority of the traffic will be equipped with such guidance system, while the second is based on adaptive control theory.

2.3.1 Equilibrium Models

Lately, a lot of attention is concentrated on estimating future travel costs with the use of dynamic traffic assignment models, based on some equilibrium principal. The basic concept of the application of such models is that, in the case when all trips are assigned to a network by the system in such a way that some optimization condition is satisfied, the future traffic volumes on each element of the network will be known, and future trips will be assigned in such a manner that the optimization condition will remain valid. A fundamental discrimination of such models can be made based on their objective function. The value that is optimized may be either the delays experienced by each individual driver, in which case optimization is referred to as user optimum solution, or the total delays experienced within the transportation system, in which case optimization is referred to as system optimum. Generally these two optimal conditions do not result in the same solution.

In system optimum modeling, the fundamental goal is to assign traffic so the marginal costs for all alternative routes being used will be equal, under a given demand function. Although it is not known how much the total delays would be reduced in the system optimum case, we can speculate that in very congested networks we may have significant savings, so even the longer routes may be shorter than those obtained by the user optimum, Of course, in this case the incentive for drivers to follow such routes is strong. However, the inherent problem with such an approach is that it may advise drivers to follow extremely long routes in order to minimize total delays. In the case that the route guidance system will be actuated by the driver and it will not be mandatory, drivers whose travel time is worse when using the system will choose not to use it. Since one of the main reasons for developing route guidance systems is to reduce congestion, and therefore reduce the delays that individual drivers are experiencing, improvements in the system performance should not be obtained at the expense of individual drivers.

An alternative objective function can be obtained under the user optimum approach. Based on the principal of optimizing each individuals travel cost, which is often referred to as Wardrop's first principal, traffic is assigned in such a way that the travel cost for all the routes that are used is equal to the minimal route travel cost. Under such a condition, no driver will use a route from his origin to his destination if a shorter route exists. Therefore traffic assigned on the network can no longer improve their cost by switching to alternative routes (Sheffi 1985).

Most of the equilibrium models developed for modeling traffic assignment do not consider the dynamics of the transportation system. For example, traffic demand is assumed to be constant over the analysis period and consideration of changes in the traffic conditions due to changes in traffic demand across time is not present. Vehicles are assigned to a static route and therefore are present on all the links of this route simultaneously. Of course such models are not appropriate for route guidance systems since traffic demand is very dynamic, especially during peak commuting periods. In addition, since guidance systems have to be responsive to actual conditions on the network, they have to be able to change the route while the trip is taking place if non recurrent congestion occurs, a function which cannot be represented with static modeling. Nevertheless the static formulation of the traffic assignment problem provides the framework for developing dynamic traffic assignment models that may be used in dynamic route guidance systems (Boyce 1988).

In their pioneering work, Merchant and Nemhauser (1978), considered a dynamic traffic assignment problem through mathematical programing for optimizing the total delays in a network with many origins and one destination. The model was based on Wardrop's second principal, which requires equal marginal costs for all the used paths, which was generalized in order to be applicable to a dynamic formulation. The planning horizon was divided into equal time intervals, and the demand was assumed to be constant for each time interval. The resulting model was a discrete time, nonlinear, non convex mathematical programming problem, which proved to be a generalization of the static system optimum model. The model does not include explicit capacity restrictions for the links, since in the optimal solution there should not be any large link volumes. The number of vehicles exiting a link during each time interval is a function only of the volume of the link. However, this formulation does not allow transmission of congestion from an already congested link to other links upstream. In the same context, Carey (1987) has extended the Merchant and Nemhauser model for multi origin multi destination networks, and managed to reform it in a nonlinear, convex form.

Recently, the approach of optimal control theory is utilized where the problem of optimal traffic assignment is formulated as an equivalent continuous time optimal control problem, and a performance index has to be optimized (Friesz et al, 1989, Ran et al, 1992, Wie, 1990). For example, Wie (1990) formulated the dynamic user optimum traffic assignment model as an optimal control problem for a single destination network, Wie et al. (1991) extended the same model for multi destination networks. The optimality conditions are derived based on Pontryagin minimum principal, and the optimal solution represents the temporal evolution of traffic flows based on the dynamic generalization of Wardrop's first principal:

> *"If, at each instant in time, for each origin-destination pair, the instantaneous expected unit travel costs for all paths that are being used are identical and equal to the minimum instantaneous expected unit path cost, the corresponding time varying flow pattern is said to be user optimized."*

These models describe the case where the driver has complete knowledge of the current traffic conditions on all the alternative paths from the present position to the destination, and route choices are based only on current conditions. Still congestion is depicted by using exit functions only of the volume of the link, and in the multi destination case, these functions have to be linear. Also the existence and the uniqueness of an optimal solution based on this model is not established.

So far, the relationship between travel cost and link flow in not represented explicitly. For example, in the case where link travel cost is the travel time required to traverse the link, such a representation is necessary so optimal solutions will be realistic. As it is noted by Ran et al (1992), when the objective function is optimized, the speed with which vehicles travel on the links may assume unrealistic large values, with the extreme situation of traffic traveling instantaneously. Therefore in their model formulation, Ran et al (1992), have included constraints regarding the time that vehicles can arrive at any node using the link free flow travel times. Similar to the model developed by Wie, the instantaneous travel time between a decision node and the destination node is found based on the current link travel times, but here the dynamic user optimum is defined as follows:

> *"If for each origin destination pair, at each instant of time, at each decision node, the travel times for all routes that are being used equal the minimal instantaneous route travel time, the dynamic traffic flow over the network is in a dynamic user optimum state."*

which is different from the definition in Wie's model where route travel times are equal at each instant in time while here, travel times are equal to the minimum instantaneous travel time only at each decision node. Their optimal control problem is convex with respect to the control variables, so it provides a unique optimal solution, and the model is implemented on a small network with four links and four nodes and two alternative paths. The model was reformulated as a discrete time non linear problem and the planning period was divided into K=6 equal time intervals, in which the dynamic user optimum was achieved.

Lafortune et al (1991) have approached the problem by modeling the traffic network as a discrete time, integer valued dynamical system. The model describes the state transition function which gives the possible states at time $t+1$ as a function of the state at time t. The state of the system is defined as the number of vehicles on each link with the same destination and the same earliest exit time from the links. The control variables correspond to the assignment of such platoons of vehicles on the same link with the same destination and the same exit time from the link on downstream links at the nodes of the network. Constraints regarding the capacity of the links as well as headways between vehicles are taken into account, so congestion can propagate on the network, and vehicles cannot traverse the links rapidly. However the model formulation is based on a-priori knowledge of the demand levels for all time periods for the entire planning horizon under consideration. As it is noted by the authors, this allows representation of recurrent congestion only, while non recurrent congestion due to unexpected events is not modeled, and thus the model cannot be applied in real time route guidance systems.

Papageorgiou (1990, 1991) has studied the problem of dynamic macroscopic traffic assignment for a multi destination network also via the optimal control approach. As in the models developed by Wie et al (1991), Lafortune et al (1990) and Ran et al (1992) traffic flows on the links consist of subflows with different destination nodes. The behavior of the drivers is represented with the "splitting rates" at each decision node, which are the control variables of the optimal control problem, while a parameter reflecting the compliance rate of drivers to suggested routes by the system is also incorporated. Optimal conditions are derived through feedback regulation, so the model can be implemented in real time route guidance systems. The model does not include any present or future information about demand levels, and perturbations to the system can be controlled via the feedback control. The modeling framework can be applied for either a system optimum or a user optimum solution. It is interesting to note that for the simple network that the model was tested on, the system optimum solution was marginally better than the user optimum one. Also a general formulation of the model is presented, so other control measures such as traffic signals and ramp metering can be included, and optimization can be performed for both splitting rates and control measures simultaneously.

However, the model developed by Papageorgiou (1990) has not been tested on congested networks, where the effectiveness of the feedback rule may not be as good. Route selection, based on the splitting rates, is based on current values of travel costs (travel times) and not future traffic conditions on the link. In addition a feedback control is a good control approach for error actuated systems. In dynamic route guidance systems though, a more sophisticated approach, based on anticipation of the future states of the system would be more advantageous.

As has been noted, the purpose of the dynamic traffic assignment models is to describe with a mathematical model the evolution of a control variable, like link traffic flows, in a time varying system. In the case of high market penetration of route guidance systems with the majority of the drivers who have the system actually using it, the effectiveness of such models will be good. In fact, if all drivers are equipped and follow exactly the suggested routes then such models will be very accurate. However, as long as the percentage of equipped drivers is small to medium, dynamic traffic assignment models seem to be inappropriate, since they fail to describe the behavior of driver 's who do not have knowledge of the optimum route. If drivers do not have perfect knowledge of the traffic conditions on the network, or if all drivers do not have uniform information, then they cannot chose the optimum route, which will result not only in increasing their travel time, but the travel time of other users who share the same route. In this case application of traffic assignment models would not adequately model the evolution of traffic patterns, and their predicting capability would be significantly reduced.

Another issue which must be taken into account is that the output of such models is aggregated flows. Often in the user optimal model formulation and under congested traffic conditions, there is more than one optimum path which can be followed. Since the route guidance system should not direct all drivers to one path, to avoid guidance system induced congestion, the system must be able to allocate drivers to the alternative equal cost paths. If such a mechanism does not exist, then the decision making process in the selection of a route to be followed in a specific trip, from the driver's perspective, is not improved at all.

2.3.2 Real Time Adaptive Forecasting

A different approach for obtaining predictions of traffic characteristics is by treating traffic variables as stochastic processes. Instead of predicting the evolution of traffic based on a deterministic model, traffic characteristics are treated as random variables which change with time, and therefore they constitute stochastic processes. The data needed to describe such variables consist of time series data, since they are repeated observations of the same variable over time. Then, models based on the characteristics of the time series can be developed, with which predictions on the traffic characteristics can be performed.

A difficulty that exists with time series models is that they often are based on the assumption of stationarity of the stochastic process. A stochastic process $\{X_t, t \in T\}$, where T is the time set where observations are made, is said to be stationary if (Brockwell and Davis 1987):

(1) $E|X_t|^2 < \infty \qquad \forall t \in T$

(2) $EX_t = m \qquad \forall t \in T$

(3) $Cov(X_t, X_{t+r}) = \gamma(r) \qquad \forall t, r \in T$

Strictly speaking, traffic variables like traffic flow or travel time are not stationary stochastic processes, since such variables are related with the time of the day and possibly the day of the week. Nevertheless, these variables can be treated as asymptotically stationary with certain conditions. Such conditions could be given time periods, locations and environmental conditions (Lu 1990). Alternatively, the process of a traffic variable can be transformed to a stationary process with appropriate differencing.

Time series models have been used for specific applications such as incident detection and freeway occupancy estimation. For example Ahmed and Cook (1980, 1982) have developed a technique to detect occurrences of traffic incidents automatically by monitoring flows on a network, based on a time series model. Based on observation of flows on different freeways they found that traffic flow time series were best represented by an autoregressive-moving average model with integration (ARIMA) model of low order (0,1,3). With this model formulation they perform predictions for the traffic flows and construct the 95% confidence band for these predictions. When the actual observations (collected in real time) were out of the 95% confidence band the alarm is activated for the existence of an incident. Because of reliability problems of traffic detectors which may create data gaps, Davis and Nihan (1984) developed a time series based model (ARMA) for detecting changes in traffic flow characteristics, despite some missing data point.

Kyte et al (1989) examined the problem of modeling freeway traffic flow with the use of a multivariate transfer function model. The relationship between freeway occupancies of a roadway segment with occupancies at upstream and downstream sections was demonstrated based on occupancy observations on a series of adjusted freeway segments. One of the segments was representing a bottleneck condition, and the resulting congestion effects on upstream traffic were examined. The speed with which traffic waves are transmitted from one segment to another was represented with the cross correlation function of the occupancies of the two segments. For example if the shock wave from the bottleneck requires *n* time steps to propagate to the upstream segment, the cross correlation function was found to have the form:

$$\gamma(k) = \begin{cases} a & \text{for} \quad k = n \\ 0 & \text{for} \quad k \neq n \end{cases} \tag{2.1}$$

In regions without congestion, the model included only terms from upstream traffic and a low order ARIMA model, while for congested regions terms from both upstream and downstream traffic characteristics were included.
However, the main operation of time series models is as historical trend tracking models. The calculated parameters of such models are constant (not time varying) and often computed off line. Therefore, the responsiveness of such time invariant models to changes in the traffic characteristics is significantly restricted, especially when unexpected variations occur. In dynamic route guidance the models that will be used for predicting future traffic conditions must be able to adapt to the dynamics of the transportation system. Methods for allowing the parameters of the model to vary with time are available, and they are included under the more general context of *adaptive prediction models*. An adaptive prediction model can be seen as a system which has structure that is adjustable and it is improved with a learning process through contact with the environment of the system. Such models include, among others, the Kalman filters and the adaptive filters.

A classic example of filtering in traffic science is given by Gazis and Knapp (1971). They developed a method for estimating the number of vehicles on a roadway segment based on real time measurements of speed and flow at the ends of the segment. The model consisted off:

(1) a part where the average travel time of vehicles was estimated periodically based on the time series of speeds;

(2) a "rough estimation" of the number of vehicles based on the travel time estimation; and

(3) a "sequential estimator" of the number of vehicles on the roadway segment.

The sequential estimator was an application of the discrete Kalman filter where the estimations of the number of vehicles was obtained based on the rough estimation from the second part of the model and the error of the previous estimate. The model was tested using data from three sequential links consisting of the three sections of the Lincoln Tunnel (downgrade, level and upgrade sections). Speed measurements were obtained for every 5 and 10 seconds for examining two different sampling rates, and it was found that the 10 second sampling period produced better results. The model was very accurate and the percentage of the time that the errors between the estimated value of number of vehicles and the actual one was less than 10% was greater than 99% for all three sections.

In the above example the model developed by Gazis and Knapp (1971) was used for estimating flows at time t, based on observations of the input variables at the same time interval, and no prediction was performed. Nevertheless the same model could be utilized for performing predictions of traffic flows in the short term future. For example, Okutani and Stephanedes (1984) developed two models based on the discrete time linear Kalman filter, for adaptively predicting traffic volumes on the links of a network. They used traffic data from three previous time intervals, and the input vector to the filter consisted of:

(1) observations of traffic volumes on upstream links; and

(2) observations of traffic volumes on the link under consideration.

In one of their models they used observations of the same day for which the predictions were performed, without any differencing. However, under the assumption that traffic flow fluctuates in a similar fashion from day to day, the parameters $h(d,t)$ that were used in this model for the day (d) and the time interval (t) were updates of the parameters obtained at the same time interval one week ago $h(d\text{-}1,t)$:

$$h(d,t) = h(d-1,t) + e(d-1,t) \tag{2.2}$$

where $e(d\text{-}1,t)$ is some noise variable. In the second model, in order to assure the stationarity of the time series, they difference the observed traffic volumes at the day under consideration with the ones that were observed one week ago at the same time interval. By doing so, variations due to both time of the day and day of the week were eliminated.

The results obtained were compared with those when the UTCS-2 model (Urban Traffic Control System-second generation) was used. Their results were always at least 80% better than those of UTCS-2. The model that incorporated the one week differencing, always performed better than the one without differencing. In the same study it was found that there was no significant improvement of the predictions when observations for more than three previous time intervals were used. The robustness of the models were indicated by the fact that their performance was not significantly affected when predictions for times longer than one time interval ahead were obtained.

In an effort to explain the existing interrelationship between traffic control strategies and the resulting traffic patterns, Stephanedes et al (1990) developed a model for predicting behavioral characteristics of traffic flow in real time. Based on an extended (non linear) Kalman filter their method was used for prediction of traffic demand at the entrance ramps of a freeway and prediction of diversion of this demand. In this study two models were developed, one for predicting arriving traffic at the entrance of a ramp, based on historical information of the same variable, and a second model for predicting the proportion of traffic that enters the ramp. From a previous study conducted by the authors it was found that the factors that affect the decision of a driver to divert from entering a freeway ramp are the rate with which vehicles enter the ramp, the rate with which vehicles enter the freeway, and the number of cars on the ramp. Socioeconomic characteristics of the driver did not appear to be relevant. Therefore, the second model, which calculates the utility attributed by the drivers to enter the ramp, relates the entering proportion of traffic to the ramp with the current freeway entering rate and the number of cars on the ramp from the previous time interval.

The above models were applied on a freeway corridor in Minneapolis, where flows approaching freeway ramps, and diversion of freeway demand was predicted in real time. After implementing the method for a typical weekday it was found that the average error for the predictions of diverted traffic was ranging from 5.4 to 8.8%, while the average error of the predictions of the approaching traffic was ranging from 6.1 to 13.4%. The authors note that the error of the second model would be smaller if in the model for predicting the approaching traffic historical information of the upstream traffic was also included as in Okutani and Stephanedes (1984).

A requirement for implementing such adaptive models is knowledge of the second order statistics of the stochastic process. In the case when no prior statistics are available, but the estimates must relay only on the available data, we need to approach the problem of predicting traffic characteristics with different adaptive prediction models. In such models the statistics of the process are learned by the adaptive processor while the prediction system is in operation (Papoulis 1984). For example, Lu (1990) used such an adaptive prediction model for predicting traffic volumes on a road segment. The prediction model was based on the least mean squares algorithm (LMS), where the estimation of the coefficients of the linear model is made recursively and it is based on the latest observations of the input variable (traffic volume) and the instantaneous error of the last prediction. The model searches for a set of coefficients that will minimize the instantaneous error in the gradient direction by setting to zero the first derivative of the expectation of the squared error, so it can be seen as a *"steepest decent algorithm"*. Although this method is simple to apply, due to the small number of computations, the convergence of the algorithm is not very rapid in the case of a high perturbation (i.e. due to a traffic incident) in the input variable.

Alternatively the Gauss-Newton method can be utilized for obtaining the estimations of the parameters of the model that minimize the sum of the squared errors. The Gauss-Newton method shows faster convergence and as it is shown by Ljung and Söderström (1983) such methods can be more robust in respect to design aspects of the model. In a different type of problem Nihan and Davis (1987, 1989) have compared the Newton-Gauss method for estimating origin destination matrices for freeways segments or signalized intersections from input-output traffic counts. In their approach the parameters of the model represented the portions of traffic leaving a specific origin and destined for a specific exit of the freeway segment or leg of the intersection. The parameters must obviously satisfy the constraints of non negativity and they must total to unity. These constraints are not handled directly by the model, but the estimated proportions are modified in a two stage process (estimation and adjustment) so they will satisfy the required constraints. The results were compared with the ones obtained from off line prediction models based on maximum likelihood estimators, and the ordinary least squares methods. In all cases the results obtained from the Gauss Newton algorithm were much better. In comparing the results from two models based on the Gauss Newton method, one with employing the constraints for the parameters and one without, the models converged in the same proportions, a fact that indicates the robustness of the method.

2.4 Summary of Literature Review

Due to the alarming rate of increase of traffic congestion in most of the major metropolitan areas over the last decade, research is focused on procedures for improving the performance of existing networks with technological means. Such procedures include the development of route guidance systems. Advanced route guidance systems should be able not only to navigate vehicles through a network, but guide them through the most desirable path. A key issue, though, of such a system is that the computation of the most desirable path is based on an estimation of future states of the network.

Several studies have tried to evaluate the benefits of dynamic route guidance systems and the results range from marginal to improvements of travel time up to 50%. The majority of these studies show that route guidance systems will not have a great effect on reducing recurrent congestion, but most of the benefits will accrue during traffic incidents. Factors like traffic demand, capacity of the roadway system, market penetration level and behavioral characteristics of the drivers affect the performance of route guidance systems. Results are based on simulation models where traffic is assigned based on a user equilibrium.

However, none of these studies have included a model for estimating the future states of the network, but they try to maintain the status of the user equilibrium by updating routing suggestions periodically. Models for estimating future states are based mainly on traffic assignment models, which are developed by extending static models to include the time dimension. Such models treat traffic as aggregated volumes and usually they cannot distinguish between equipped and unequipped traffic., and assume that drivers have uniform information, and thus react in a uniform manner to the prevailing traffic conditions. Therefore it seems that such models will be best applicable when the market penetration of route guidance systems will be relatively high, and thus drivers will indeed have perfect information.

Alternatively, prediction models based on the time series of the traffic characteristics can be used. Such models, dynamic by structure and readily applicable in real time, have been used mostly for estimating and/or predicting traffic volumes on roadway segments. Although the potentials of such models are good, they have not been used in route guidance systems. In this study we consider the application of such a model that can adapt its parameters depending on input which consists of the current traffic conditions on the network.

Chapter 3

Prediction Model of Travel Times

3.1 Introduction

The general purpose of this study is to examine the potential of time adaptive control techniques for predicting traffic characteristics in real time. More specifically we will attempt to develop a model formulation that is capable of predicting traffic characteristics under a variety of traffic conditions, i.e. normal traffic conditions or congested conditions. Under the notion that a prediction model is a system itself, adaptive control techniques characterize this system by estimating the system parameters. The key elements of system identification is first the structure of the model, and second the estimation of its parameters.

Because prediction based on the adaptive system will be performed in real time, its operation has to be automatic. Therefore, it is essential to have a good understanding of all aspects of the problem so the structure of the model can be formulated. The structure of the model may not be readily identifiable, either because of our limited knowledge of mechanisms describing the system and the complexity of such mechanisms, or because the system is changing in an unexpected manner, i.e. in the case of a traffic incident. In this case, observed changes in the value of the variables produced by the system can be used to reconstruct the model. In this study we develop a model which is based on the observed values of traffic characteristics.

3.2 Operating Framework

The framework for a model incorporating real time adaptive prediction of traffic conditions consists of three subsystems (Figure 3.1):

(1) The traffic network that is equipped with appropriate devices for collecting information regarding current travel conditions;

(2) The traffic control center that is capable of processing this information, performing predictions for short term future traffic conditions, and broadcasting this information back to the traffic on the network; and

(3) The vehicles that are equipped with some special device that enables them to receive information and evaluate their path towards their destination.

As travel conditions change throughout the duration of the trip, equipped vehicles would use the latest information to reevaluate their routes at the end of each link. Once a vehicle has entered a road segment defining a link of the selected route it cannot divert from that specific link. When it is exiting the link, it is considered to have a new origin and an optimum route for the remainder of the trip is recalculated, based on the latest estimates and predictions of travel conditions. Thus, vehicles can divert from the selected path en-route, if a better path from their current position towards their destination is calculated. Such diversion from the originally selected route could occur either because of poor predictions of travel conditions on the links, or because of the occurrence of an unexpected incident on one of the downstream links after

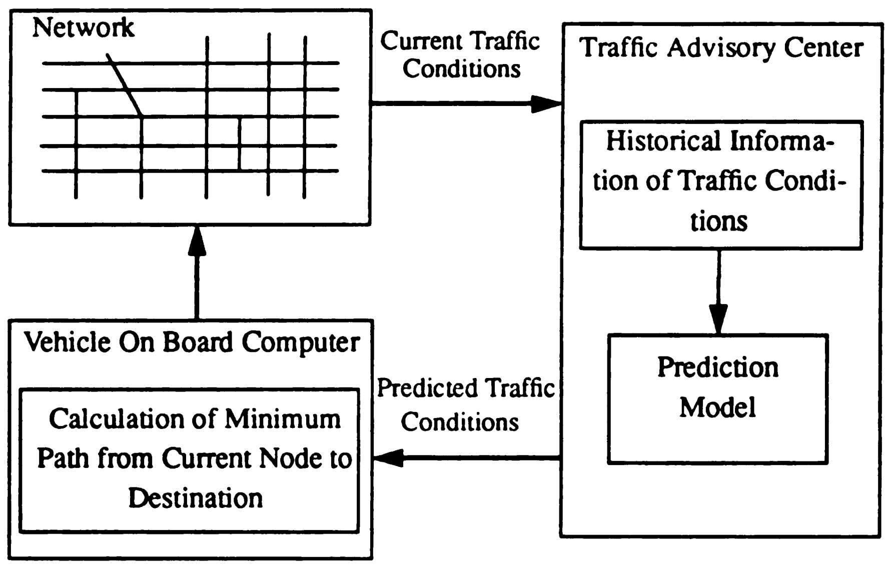


Figure 3.1: Operating Framework for Prediction Model of Travel Conditions on the Network.

departure.

3.3 Structure of Prediction Model

Let the traffic network be represented by a graph $G\{N ; L\}$. The graph consists by two subspaces, the subspace N that includes the set of vertexes (or nodes) in the network, and the subspace L that includes the directed arcs of the network, where each arc is represented by an ordered pair of points from the space N ($L \subseteq N \times N$). Nodes may represent either a junction of two roads or highway segments, or points where the physical characteristics associated with each arc change. Arcs represent the physical path (i.e. road or highway segment) from one node to another, and they will be referred to as links. Obviously the property which apply to subspace L is:

$$(i, j) \neq (j, i) \qquad \forall\, i, j \in N$$

If we assume that the traffic control center is collecting data regarding traffic conditions on the links of the network in discrete time t=1,2,3, ... , then such observations comprise time series data. If the observed value of a traffic variable at time t, used in the prediction model is denoted by $x(t)$, then the sequence of such observations up to time t can be arranged in a vector:

$$x^t = \{x(t), x(t-1), \ldots, x(1)\} \tag{3.1}$$

If more than one observable variable is used for the prediction of the state of a link, we denote such variables as $x_1^t, x_2^t, \ldots$ .The model describing the state that a link (i,j) is in at time t can be parametrized in terms of a set of unknown parameters which can be arranged in a vector θ. Then the model will have the general form:

$$y(t) = f(\theta; x_1^t, x_2^t, \ldots) + \varepsilon(t) \tag{3.2}$$

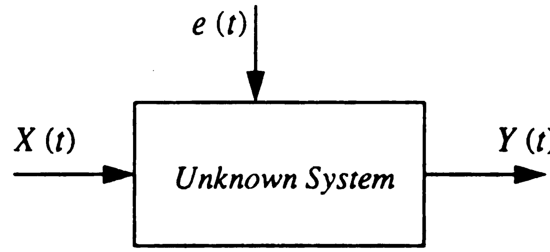


Figure 3.2: Block Diagram of Travel Conditions of a Link

where $y(t)$ is the variable indicating the state of the link at time t, $\varepsilon(t)$ is a noise term of unspecified character, and $f(\cdot)$ is a known function of $x_1^t, x_2^t, \ldots$ and θ. This concept is illustrated in Figure 3.2.

The indicator variable $y(t)$ of the state of the link has to be a variable that can be directly related to the cost associated with traversing this road segment. Under the assumption that the aspect that drivers want to minimize is their travel time, the state of a link (i,j) at time t, will be expressed as the time required to traverse this link at time t.

3.3.1 Input Variables to the Prediction Model

The input variables to the prediction model, $x_1^t, x_2^t, \ldots$, can be observations of either travel times, speeds, or traffic volumes and densities. Measurement of travel times has the inherent drawback of some delay, since estimations can be obtained only after a vehicle exits the link. On the other hand volumes and densities can be estimated as soon as a vehicle enters the link. However, in the case of congestion, simple volume measurements are clearly not a good choice, since their usage can be misleading. As is discussed by Mauro (1991), an incident near the entrance of the link will result in low occupancies but high travel times. In addition, if no traffic is flowing through the bottleneck, no estimation of the travel time of the link will be possible either. Therefore procedures detecting the existence and the length of traffic queues (i.e. image processing devices) are necessary. Then, based on such information, the travel time of the link can be estimated based on some impedance function. In the following it is assumed that the system is capable of estimating current travel times of all the links of the network.

The value of the parameter vector θ of the model for a link (i,j) will be estimated based on the information contained in vectors x_k^t with k=1,2,... . Let $T_{(i,j)}(t)$ be the travel time of link (i,j) at time t to be predicted. The first input variable to the model will be the past history of the travel time of the link under consideration (i,j) until time $(t\text{-}1)$, so $x_1^t = T_{(i,j)}^{t-1}$. This is under the assumption that future states of the link are related to current and past states of the same link.

In addition, there is obviously some relationship with upstream and downstream traffic conditions. Perturbations of traffic patterns either on upstream or downstream links will be transmitted to the link under consideration. For example upstream traffic events will affect downstream traffic patterns with some time lag required by the vehicles on the upstream link to reach the downstream link. Therefore, travel times on upstream links should affect travel times on subsequent links. On the other hand when the capacity on a downstream link is reduced below the prevailing demand as a result of a traffic incident or due to recurrent congestion, a bottleneck is created. Such conditions are known to affect traffic patterns not only on the link with the incident but to the upstream links as well. Figure 3.3 and Figure 3.4 illustrate the rela-

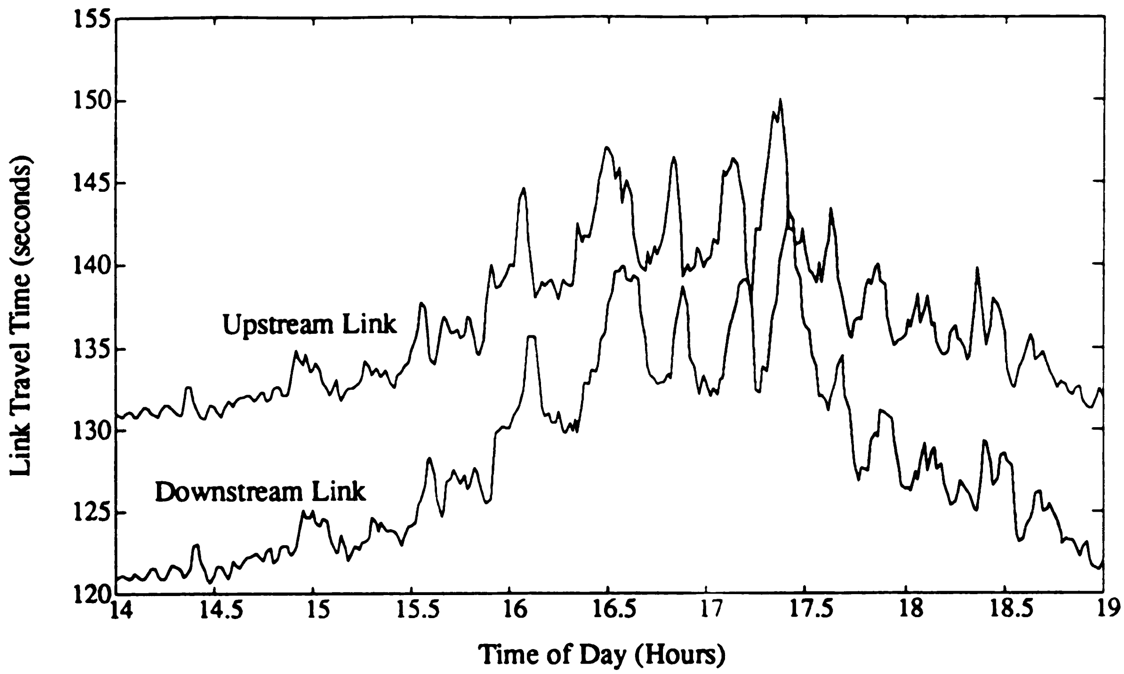


Figure 3.3: Relation Between the Travel Times of Two Links in Sequence (Simulated Observations, Sampling Interval 60 seconds)

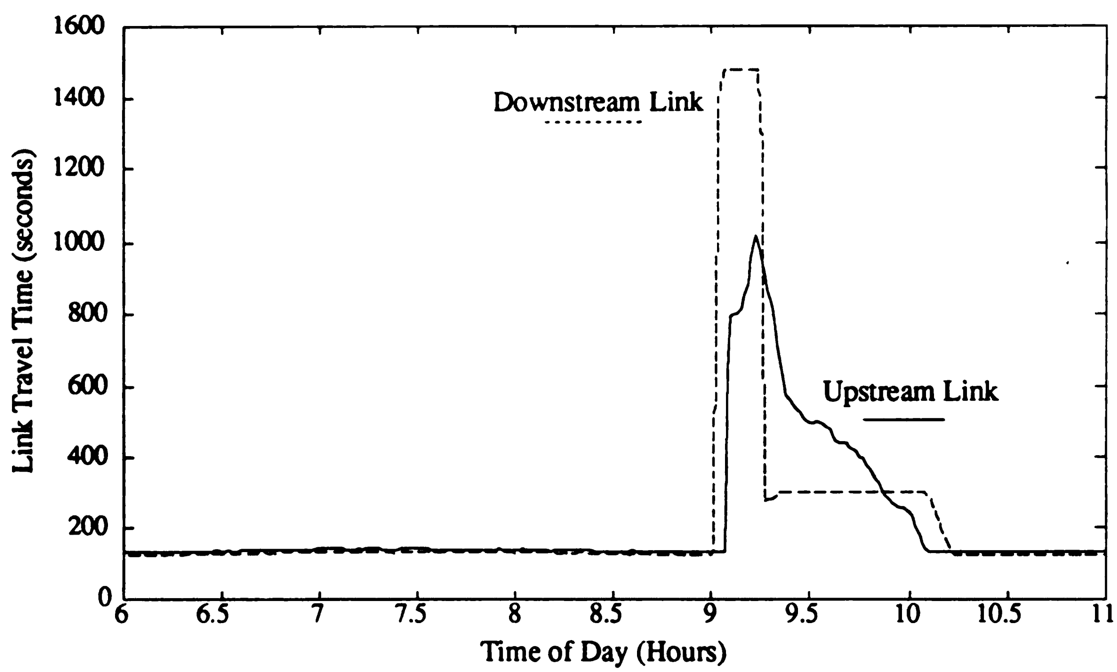


Figure 3.4: Relation Between the Travel Times of Two Links in Sequence with an Accident on the Downstream Link (Simulated Observations, Sampling Interval 60 seconds)

tionship between upstream and downstream traffic. In the first figure where there are no incident occurences travel time on the downstream link depends on travel times of the upstream link. In the second figure an incident occurs on the downstream link at time 09:00:30.2 at which point the relation is reversed until the end of the incident approximately 15 minutes later, and downstream travel times affect travel times on the upstream link.

Finally there is a strong diurnal pattern which has to be taken into account. This pattern is related to the demand level which varies not only by time of day but by day of week as well. Therefore variable $T_{(i,j)}(t)$ is related to some variable that describes this diurnal pattern, for example the expected travel time on link (i,j) at time t, based on observations of $T_{(i,j)}(t)$ from previous days.

3.3.2 Model Formulation

For the sake of simplicity in the remainder of this section we shall denote links with a single character i.e. $l = (i,j)$. Under the assumption that the model for predicting the travel time of link l at time t is linear in its parameters, then it can be written as:

$$A(q^{-1}) \cdot T_l(t) = \sum_{k \in I} B_k(q^{-1}) \cdot T_k(t) + \sum_{p \in O} C_p(q^{-1}) \cdot T_p(t) + d \cdot \tilde{T}_l(t) + \varepsilon(t) \qquad (3.3)$$

where I is the set of links ending at node i, O is the set of links starting at node j, and $\tilde{T}_l(t)$ is the average travel time of link l for the time interval t, obtained from previous days (Figure 3.5). This term is included to express the diurnal pattern of travel times of link l. The noise term $\varepsilon(t)$ is assumed to be a sequence of independent random variables with mean value equal to zero. The terms A, B and C are polynomials of the form:

$$A(q^{-1}) = 1 + a_1 \cdot q^{-1} + a_2 \cdot q^{-2} + \ldots + a_n \cdot q^{-n} \quad (3.4.a)$$

$$B_k(q^{-1}) = b_{k1} \cdot q^{-1} + b_{k2} \cdot q^{-2} + \ldots + b_{km_k} \cdot q^{-m_k} \quad (3.4.b)$$

and

$$C_p(q^{-1}) = c_{p1} \cdot q^{-1} + c_{p2} \cdot q^{-2} + \ldots + c_{pr_p} \cdot q^{-r_p} \quad (3.4.c)$$

and q^{-1} is the delay operator, for example $q^{-1} \cdot T_l(t) = T_l(t-1)$. We will refer to the first term of equation (3.3) which includes the past history of the travel times from link l as the *autoregressive term*, to the term with the history of the upstream links as the *convection term*, to the term with downstream link travel times as the *congestion term*, while the term with the average travel time of link l as the *diurnal term*.

Let $O=\{k_1,k_2,\ldots,k_O\}$ and $E=\{p_1,p_2,\ldots,p_E\}$ be the elements of the link sets O and E respectively. By introducing vector notation equation (3.3) reduces to:

$$T_l(t) = \theta^T \cdot \varphi(t) + \varepsilon(t) \quad (3.5)$$

where θ is the parameter vector:

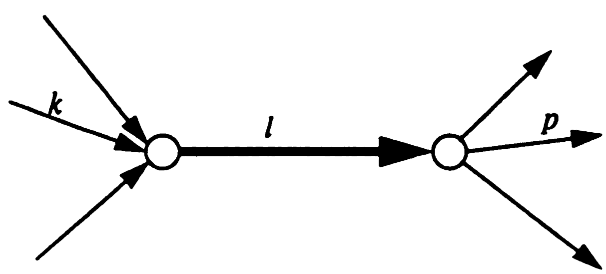


Figure 3.5: Link Layout for Prediction Model of Link l - Equation (3.3)

$$\theta^{T} = \left[a_1, \ldots, a_n, b_{k_1 1}, \ldots, b_{k_1 m_{k_1}}, \ldots, b_{k_O 1}, \ldots, b_{k_O m_{k_O}}, c_{p_1 1}, \ldots, c_{p_1 r_{p_1}}, \ldots \right.$$

$$\left. \ldots, c_{p_E 1}, \ldots, c_{p_E r_{p_E}}, d \right] \tag{3.6}$$

and $\varphi(t)$ is the vector with the observations:

$$\varphi(t) = \left[-T_l(t-1), \ldots, -T_l(t-n), T_{k_1}(t-1), \ldots, T_{k_1}(t-m_{k_1}), \ldots, T_{k_O}(t-1), \ldots \right.$$

$$\left. \ldots, T_{k_O}(t-m_{k_O}), T_{p_1}(t-1), \ldots, T_{p_1}(t-r_{p_1}), \ldots, T_{p_E}(t-1), \ldots, T_{p_E}(t-r_{p_E}), d \right] \tag{3.7}$$

A natural guess of $T_l(t)$, denoted as $\hat{T}_l(t)$, will be:

$$\hat{T}_l(t) = \hat{\theta}^{T} \cdot \varphi(t) \tag{3.8}$$

where, of course, we need some estimation $\hat{\theta}$, of the parameter vector θ.

Equation (3.8) gives the one step ahead prediction for the travel time of link l. The τ-step ahead predictions are obtained by successively using τ one step ahead predictions. Let $\hat{T}_l(t+\tau|\, t-1; \hat{\theta}(t-1))$ be the τ-step ahead prediction of the travel time of link l at time $(t + \tau)$ based on observations up to time $(t - 1)$, with the parameter vector $\hat{\theta}(t-1)$ estimated also based on observations up to time $(t - 1)$. The τ-step ahead prediction will be calculated as:

$$\hat{T}_l(t+\tau|\, t-1; \hat{\theta}(t-1)) = \hat{\theta}(t-1)^{T} \cdot \bar{\varphi}(t+\tau) \tag{3.9}$$

where $\bar{\varphi}(t+\tau)$ will be similar to (3.7) but it will include observed values up to time $(t - 1)$ and expected values after this point. For example consider the simple case where the model (3.3) is given by:

$$A(q^{-1}) \cdot T_l(t) = \varepsilon(t) \tag{3.10}$$

then:

$$\theta^{T} = \left[a_1, \ldots, a_n\right] \qquad \varphi(t) = \left[-T_l(t-1), \ldots, -T_l(t-n)\right] \tag{3.11}$$

If $\tau = 2$ then $\bar{\varphi}(t+\tau)$ will be

$$\bar{\varphi}(t+2) = \left[-\hat{T}_l(t+1|t-1;\hat{\theta}(t-1)), -\hat{T}_l(t|t-1;\hat{\theta}(t-1)), -T_l(t-1), \ldots, -T_l(t-n+2)\right]$$

3.4 Prediction Algorithm

The estimation of travel times will have to be performed on line, so the most recently observed data will be used by the model (3.8) for the most updated predictions. Since travel conditions on a traffic network change continuously, it is necessary to use an algorithm that will update the parameters of the model as new data become available. Because observations of travel times will consist of long data sequences, processing of such data by the prediction model will have to be sequential. Such a procedure is referred to in the literature as *recursive identification* or as *adaptive algorithm.* The significant advantage of such recursive identification techniques is that computational and memory storage demands do not increase with time. Available storage memory can be utilized for the prediction model rather than for storing the data, since we need to store only the latest data while older data can be discarded (Ljung and Söderström 1983).

3.4.1 Time Invariant Parameters

A recursive identification algorithm should be able to identify changes in travel times as such changes take place by responding to the instantaneous error defined as:

$$\hat{\varepsilon}(t) = T_l(t) - \hat{T}_l(t) = T_l(t) - \hat{\theta}^T \cdot \varphi(t) \tag{3.12}$$

The majority of recursive identification algorithms update the estimation of the parameter vector by adding some correction to its previous estimation.

The correction depends on the discrepancy between the observed travel time and the expected travel time for the same time interval, estimated based on the previous estimation of the parameter vector:

$$\hat{\theta}(t) = \hat{\theta}(t-1) + L(t) \cdot [T_l(t) - \hat{\theta}^{T}(t-1) \cdot \varphi(t)] \tag{3.13}$$

where $L(t)$ is the "gain" vector, which regulates the rate with which the prediction error $\hat{\varepsilon}(t)$ contributes to the update of $\hat{\theta}$.

A common technique for estimating the parameter vector is by using the least square principle, where the sum of the squares of the differences between the actually observed and the computed values is a minimum. If we introduce the loss function $V_N(\theta)$:

$$V_N(\theta) = \frac{1}{N} \sum_{t=1}^{N} [T_l(t) - \hat{\theta}^{T} \varphi(t)]^2 \tag{3.14}$$

then $V_N(\theta)$ becomes minimum when:

$$\hat{\theta} = \left[\sum_{t=1}^{N} \varphi(t) \cdot \varphi^{T}(t) \right]^{-1} \cdot \sum_{t=1}^{N} \varphi(t) \cdot T_l(t) \tag{3.15}$$

assuming that the inverse exists. Then with simple calculations we can rewrite (3.14) in a recursive form:

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{1}{t} \cdot R^{-1}(t) \cdot \varphi(t) \cdot [T_l(t) - \hat{\theta}^{T}(t-1) \cdot \varphi(t)] \tag{3.16.a}$$

$$R(t) = R(t-1) + \frac{1}{t} \cdot [\varphi(t) \cdot \varphi^{T}(t) - R(t-1)] \tag{3.16.b}$$

where:

$$R(t) = \sum_{s=1}^{t} \varphi(s) \cdot \varphi^{T}(s) \tag{3.17}$$

given that the inverse of $R(t)$ exists. Equation (3.16.a) has the form of (3.13) for the recursive estimation of the parameter vector $\hat{\theta}$, but in each iteration we will have to find the inverse of matrix $R(t)$. By applying the matrix inversion lemma which states that:

$$[A + B \cdot C \cdot D]^{-1} = A^{-1} - A^{-1} \cdot B \cdot [D \cdot A^{-1} \cdot B + C^{-1}]^{-1} \cdot D \cdot A^{-1} \tag{3.18}$$

and with the transformation:

$$P(t) = \frac{1}{t} \cdot R^{-1}(t) \tag{3.19}$$

we obtain the following equivalent recursive form for (3.16.a)-(3.16.b):

$$\theta(t) = \hat{\theta}(t-1) + L(t) \cdot [T_l(t) - \hat{\theta}^{\mathrm{T}}(t-1) \cdot \varphi(t)] \tag{3.20.a}$$

$$L(t) = \frac{P(t-1) \cdot \varphi(t)}{1 + \varphi^{\mathrm{T}}(t) \cdot P(t-1) \cdot \varphi(t)} \tag{3.20.b}$$

$$P(t) = P(t-1) - \frac{P(t-1) \cdot \varphi(t) \cdot \varphi^{\mathrm{T}}(t) \cdot P(t-1)}{1 + \varphi^{\mathrm{T}}(t) \cdot P(t-1) \cdot \varphi(t)} \tag{3.20.c}$$

(See Ljung and Söderström 1983, pages 17-21). The algorithm given by (3.20.a)-(3.20.c) is the known recursive least squares algorithm (RLS).

3.4.2 Time Varying Parameters

In the previous paragraph the parameter vector was treated as time invariant, with the prediction algorithm trying recursively to estimate the best estimation of $\hat{\theta}(t)$ at time t based on the observations up to time t, according to criterion (3.14). However, the main reason why we need a recursive prediction model for guiding traffic in real time, is because the dynamics of the system are changing with time. Therefore we need to develop a technique to "track" changes in the system, as such changes occur. Under criterion (3.14) the algorithm given by (3.20.a)-(3.20.c) gives the average behavior of the system during the period [0 , t]. In order to obtain an estimate that will be more representative of the current traffic conditions on link l, we can modify criterion (3.14) so older observations of travel times will have a discounted effect on the optimum estimation of the parameter vector $\hat{\theta}(t)$ at time t, while more recent ones will be weighted more heavily. An example of such a criterion would be (Ljung and Söderström 1983):

$$V_t(\theta) = \sum_{s=1}^{t} \beta(s,t)\,[T_l(s) - \hat{\theta}^{T}\varphi(s)]^2 \tag{3.21}$$

where $\beta(s,t)$ is a function that increases in s for given t. To incorporate criterion (3.21) in a recursive algorithm we need a recursive formulation for function $\beta(s,t)$. A typical function that is often used is:

$$\beta(s,t) = \lambda(s) \cdot \beta(s-1,t) = \left[\prod_{j=s+1}^{t} \lambda(j)\right] \cdot \beta(t,t) \tag{3.22}$$

where of course $\beta(t,t)$ is the weight attributed to the latest prediction error, and $\lambda(j) \leq 1$ for all j. In the case that $\lambda(j) = \lambda$ then function (3.22) has an exponential "forgetting pattern":

$$\beta(s,t) = \lambda^{t-s} \cdot \beta(t,t) \tag{3.23}$$

since the weight that is given to older errors of the model is found by (3.23). The parameter λ is chosen so $0 < \lambda \leq 1$, and it is referred to as the *forgetting factor.* Often the weight attributed to the most recent prediction error $\beta(t,t) = \beta$ is set equal to 1. Using the transformation:

$$R(t) = \sum_{w=1}^{t} \beta(t,w) \cdot \varphi(w) \cdot \varphi^{T}(w) \tag{3.24}$$

we get the recursive form of (3.16.a)-(3.16.b):

$$\hat{\theta}(t) = \hat{\theta}(t-1) + R^{-1}(t) \cdot \varphi(t) \cdot \beta \cdot [T_l(t) - \hat{\theta}^{T}(t-1) \cdot \varphi(t)] \tag{3.25.a}$$

$$R(t) = \lambda(t) \cdot R(t-1) + \beta \cdot \varphi(t) \cdot \varphi^{T}(t) \tag{3.25.b}$$

Again with $P(t) = R^{-1}(t)$ and in a similar fashion as for (3.20.a)-(3.20.c) we can get:

$$\theta(t) = \hat{\theta}(t-1) + L(t) \cdot [T_l(t) - \hat{\theta}^{T}(t-1) \cdot \varphi(t)] \tag{3.26.a}$$

$$L(t) = \frac{P(t-1)\cdot\varphi(t)}{\lambda(t)/\beta + \varphi^T(t)\cdot P(t-1)\cdot\varphi(t)} \tag{3.26.b}$$

$$P(t) = \frac{1}{\lambda(t)}\cdot\left[P(t-1) - \frac{P(t-1)\cdot\varphi(t)\cdot\varphi^T(t)\cdot P(t-1)}{\lambda(t)/\beta + \varphi^T(t)\cdot P(t-1)\cdot\varphi(t)}\right] \tag{3.26.c}$$

which for $\lambda(t) = 1$ and $\beta = 1$ coincides with (3.20.a)-(3.20.c). The effect of the forgetting factor such that $0 < \lambda(t) \le 1$ is that the matrix $P(t)$ in (3.26.c) will not tend to zero. The consequence is that the gain vector $L(t)$ will be kept large. This means that the algorithm will always be alert so it will be able to track changes in the dynamics of the travel times of link l. However the compromise for this alertness of the algorithm is that it becomes more sensitive to the random disturbances $\varepsilon(t)$ of the measurements. For values of $\lambda(t)$ close to 0 the algorithm will be very alert to changes in the system but also too sensitive to the prediction errors, while for values of $\lambda(t)$ close to 1 the algorithm will restrain its tracking capability and it will gain in insensitivity to noise.

3.4.3 Initialization of Prediction Algorithm

For the initialization of the recursive algorithm (3.26.a)-(3.26.c) we need some initial values for the parameter vector $\theta(t_0)$ and the matrix $P(t_0)$. Because we were based on the assumption that the invert of matrix $R(t)$ exists, the initial value of $R(t_0)$ corresponding to $P(t_0)$ must also be invertible. Usually we can start the algorithm with $t_0 = 0$ and an invertible matrix $P(0)$.

The importance of the initial values diminishes with time. This can be seen if we set equations (3.26.a)-(3.26.c) in a non-recursive form:

$$\hat{\theta}(t) = \left[P^{-1}(0) + \sum_{w=1}^{t}\beta\varphi(w)\varphi^T(w)\right]^{-1}\left[P^{-1}(0)\hat{\theta}(0) + \sum_{w=1}^{t}\beta\varphi(w)T_l(w)\right] \tag{3.27}$$

where the effect of the values $\hat{\theta}(0)$ and $P(0)$ decays as the sums increase in magnitude. Also as $P^{-1}(0) \rightarrow 0$ the recursive estimation of $\hat{\theta}(t)$ approaches the off-line estimation of the parameter vector. Therefore a common choice for initial values is

$$\hat{\theta}(0) = \mathbf{0} \quad \text{and} \quad P(0) = c \cdot \boldsymbol{I} \tag{3.28}$$

where c is a large constant number.

3.5 Stochastic Interpretation of the RLS Algorithm

If instead of criterion (3.14) we choose $\hat{\theta}(t)$ so the variance of the noise term $\varepsilon(t)$ is minimized at each iteration then the following criterion results:

$$V(\theta) = \frac{1}{2} \mathrm{E} \left[T_l(t) - \hat{\theta}^{\mathrm{T}} \varphi(t) \right]^2 \tag{3.29}$$

where E[·] denotes expectation. The minimum of (3.29) will be:

$$\left[\frac{d}{d\theta} V(\theta) \right]^{\mathrm{T}} = \mathrm{E} \varphi(t) \left[T_l(t) - \hat{\theta}^{\mathrm{T}} \varphi(t) \right] = 0 \tag{3.30}$$

Because the distributions and the moments of $T_l(t)$ and $\varphi(t)$ are not known we cannot solve (3.30) analytically. An approximation can be obtained if we evaluate expected values with sample means which will result in criterion (3.14). Alternatively the Robbins and Monro recursive equation can be used where the solution of the equation $\mathrm{E} Q(x, e(t)) = 0$ is found by:

$$\hat{x}(t) = \hat{x}(t-1) + \gamma(t) \cdot Q(\hat{x}(t-1), e(t)) \tag{3.31}$$

where $e(t)$ is a stochastic process of random variables with unknown distribution, $x(t)$ is a chosen value by the user, $\gamma(t)$ is a sequence of positive numbers tending to zero, and $Q(x,e(t))$ is a known function of $e(t)$ and $x(t)$. Applying

(3.31) to (3.30) with $x(t) = \theta(t)$ we get:

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \gamma(t) \cdot \varphi(t) \cdot [T_l(t) - \hat{\theta}^{T}(t-1)\varphi(t)] \quad (3.32)$$

which for $\gamma(t) = \gamma$ gives the least mean square (LMS) algorithm. This algorithm can be seen as the parallel algorithm of a "steepest descent" algorithm for numerical minimization. Ljung and Söderström (1983) suggest that if we replace the negative gradient direction:

$\left[-\frac{d}{d\theta}V(\theta)\right]^{T}$ with the Newton direction: $-\left[\frac{d^2}{d\theta^2}V(\theta)\right]^{-1} \cdot \left[-\frac{d}{d\theta}V(\theta)\right]^{T}$

the efficiency of (3.29) is increased significantly. Then the natural variation of the Robbins and Monro equation will be:

$$\hat{x}(t) = \hat{x}(t-1) + \gamma(t) \cdot V''(\hat{x}(t-1), e(t))^{-1} \cdot Q(\hat{x}(t-1), e(t)) \quad (3.33)$$

From (3.29) we get that:

$$\frac{d^2}{d\theta^2}V(\theta) = \mathrm{E}\,[\varphi(t) \cdot \varphi^{T}(t)] \quad (3.34)$$

which is independent of θ, and is the autocorrelation matrix of the observation vector. Then from (3.34) the autocorrelation matrix can be determined as the solution R of:

$$\mathrm{E}\,[\varphi(t) \cdot \varphi^{T}(t) - R] = 0 \quad (3.35)$$

From Robbins and Monro equation we find that the estimate of $\frac{d^2}{d\theta^2}V(\theta)$, $R(t)$, at time t will be:

$$R(t) = R(t-1) + \gamma(t) \cdot [\varphi(t) \cdot \varphi^{T}(t) - R(t-1)] \quad (3.36)$$

and from (3.33) we get the stochastic recursive estimation of $\theta(t)$ based on the Newton direction:

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \gamma(t) \cdot R^{-1}(t-1) \cdot \varphi(t) \cdot [T_l(t) - \hat{\theta}^{T}(t-1)\varphi(t)] \quad (3.37)$$

The algorithm (3.36)-(3.37) is the stochastic Newton algorithm. For $\gamma(t) = 1/t$

we get the algorithm (3.16.a)-(3.16.b). If we introduce the transformation:

$$\bar{R}(t) = R(t)/\gamma(t) \tag{3.38}$$

then algorithm (3.36)-(3.37) with simple calculations can be written as:

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \bar{R}^{-1}(t-1)\cdot\varphi(t)\cdot[T_l(t) - \hat{\theta}^{\mathrm{T}}(t-1)\varphi(t)] \tag{3.39.a}$$

$$\bar{R}(t) = \frac{\gamma(t-1)\cdot[1-\gamma(t)]}{\gamma(t)}\cdot\bar{R}(t-1) + \varphi(t)\cdot\varphi^{\mathrm{T}}(t) \tag{3.39.b}$$

which for $\lambda(t) = \dfrac{\gamma(t-1)\cdot[1-\gamma(t)]}{\gamma(t)}$ and $\beta = 1$ coincides with algorithm (3.16.a)-(3.16.b).

The interpretation of the RLS algorithm as a stochastic Newton algorithm reveals some powerful aspects of the method. Convergence is a measure of the speed with which the algorithm converges or "identifies" the true value of the parameter vector θ. It is well recognized that the stochastic Newton algorithm derived here converges much faster than the stochastic gradient algorithm (LMS), and it is generally more insensitive to the choice of the initial values. Haykin (1986) reports that the RLS algorithm converges faster than the LMS algorithm at least by an order of magnitude. By utilizing the inverse of the autocorrelation matrix the RLS algorithm becomes *self-orthogonalizing* since it decorrelates the successive tap inputs. Of course the compromise lays on the considerably higher computational complexity of the algorithm as opposed to the LMS one. The computational complexity of the RLS algorithm increases with the size of the parameter vector M. The number of multiplications required by this method is $3M(3+M)/2$, while for the LMS algorithm it is only $2M+1$.

Chapter 4

Simulation Model

4.1 Introduction

For evaluating the performance of the RLS filter, it was decided to use a simulation model, where vehicular movements on a network would be modeled. This approach was favored for two reasons: First because real world data of travel times on a sequence of links for extended time periods were not available, and second because it was realized that the availability of real time information about travel conditions (in terms of travel times) to a portion of drivers will alter the travel patterns on the network .

For the later reason, it became obvious that testing the performance of the travel time prediction model based on the RLS filter will require the prediction model to be incorporated within the simulation model. Because the usage of an existing simulation package would require access to the source code of the program, in order to make alterations to the path selection routines, as well as to incorporate the prediction model, it was decided to develop the simulation model.

Never the less, models already developed were considered to some extent.

From previous comprehensive studies of previously developed traffic simulation programs (Van Aerde and Yagar 1988a, Underwood 1990), it was found that the majority of the models were developed for a specific subsystem of the transportation system, and none except the INTEGRATION model, were suitable for applications in IVHS technology. From this work it was determined that integrated models, in the sense of incorporating different subsystems in one model, has clear advantages. For example, individual subsystems could be identified for the freeway system, the traffic signal controlled network, or the mass transportation system. This is true since in integrated control, defined as the simultaneous control of subsystems, each subsystem could augment the weaknesses of the others, so the whole system would operate in a higher level than if each subsystem was controlled independently (Yagar 1983). The integration of freeways and signalized arterials constitute the most plausible application of such models, and in the INTEGRATION model these two subsystems are combined with the interactions between them modelled. (Van Aerde and Yagar 1988a,1988b).

In the development of our model, integration would be desirable. However, its complexity, especially when considering semiactuated or fully actuated signals (the case which tends to be the rule in real world traffic networks), made it impossible due to time limitations and this task is reserved for the future. In this simulation model the freeway system, including possibly major arterials, but without consideration of traffic signals, is modelled. This is believed to be adequate for the purpose of developing the model to examine the applicability of adaptive algorithms in travel time predictions.

The option of using a simulation language for a model of this scope (like

SLAM-II, or GPSS) was also rejected basically due to the limited capacity of such packages in their personal computer version (i.e. size of network, number of entities simultaneously on the network, etc.). Therefore the simulation model was developed in FORTRAN computing language, so it can run on a personal computer and includes a real time graphics module. In the following paragraphs the basic concepts for the simulation model are given.

4.2 Modelling Elements

The basic elements of the system that had to be modelled are the network, represented by a directed graph denoted as $G\{N;L\}$ and the entities that traverse the network.

4.2.1 Representation Of Network In The Simulation Model

The graph representation of the network consists by the two subspaces, N , the subspace of the nodes of the graph and L, the set of the links of the graph, as in paragraph 3.3. In addition to the properties of subspace L, introduced at that point:

(1) $(i,j) \neq (j,i) \ \forall \ i,j \in N$

we introduce the following properties for the same subspace:

(2) $(i,j) \in L \Rightarrow i \neq j \ \forall \ i,j \in N$

(3) $(i,j) \in L \Rightarrow$ there is only one pair of $(i,j) : (i,j) \in L$

Property (1) is necessary by the definition of the graph to be directed. Properties (2) and (3) are utilized for simplifying various procedures in the simulation program, and it is believed that they do comply with the reality of traffic networks. In case any of these restrictions has to be removed, this can be achieved by introducing a dummy node.

A sequence of links traversed by an entity defines a path, denoted by the set of links traversed. For example a path with n links will be:

$$\text{Path}\, i_1 \rightarrow i_{n+1} = \{ (i_1, i_2), (i_2, i_3), \ldots (i_n, i_{n+1}) \} \tag{4.1}$$

The traffic network is represented to the simulation model through a matrix graph $G\{N;L\}$ in terms of the adjacency of each vertex to any other vertex of the graph. Such a matrix representation, denoted as A, is called *vertex adjacency list* of a graph $G\{N;L\}$, and is defined as:

(1) the rows of the matrix correspond to the vertices of G, and

(2) a_{ij} = the j^{th} adjacent vertex to the i^{th} vertex if $(i,j) \in L$.

The matrix A is of dimension $n*k$ where n is the number of nodes in the graph, and k is the maximum number of adjacent nodes to any node of the graph. Nodes are labeled with a unique label for each node number, and for simplicity in the programming of the simulation model, the labels associated with each node have to be in sequence starting from label 1.

There are three different types of nodes, generation nodes, destination or termination nodes, and transfer nodes. Vehicles are generated only at generation nodes, and may exit the network only at destination nodes. Any node in the network can be a generation or/and a destination node but the label of each generation and destination node has to be identified to the model. Generation and destination nodes form two new subspaces of N defined as:

(1) $O = \{i{:}i$ is a generation node $\wedge\ i \in N\}$

(2) $D = \{i{:}i$ is a destination node $\wedge$ i $\in N\}$.

Each link of the network (i,j) $i,j \in N$, is described by a set of variables associated with the traffic engineering properties of the link as well as aspects of incident occurrences on the link. Links are numbered in the order that they appear in the vertex adjacency list, for example link 1 is the link from node 1 to node 4, while link 7 is the link from node 5 to node 6.

4.2.2 Representation Of Traffic In The Simulation Model

The entities that traverse the links of the network represent the vehicles on the traffic network. Several macroscopic traffic simulation models (like CONTRAM or FREFLOW) or signal optimization models (like TRANSYT) represent traffic as packets of vehicles that move together as platoons, and instead of tracing individual vehicles they trace the movement of whole platoons. Another class of macroscopic models simulates traffic flow based on the fluid flow analogy. Both of these classes of models examine the status of the network and the development of traffic (platoons or fluid portions) in a sequence of discrete points in time, equally spaced. The microscopic model NETSIM simulates movements of individual vehicles in great detail, but individual vehicles do not have specified destinations, and turning movements at the nodes of the network are prescribed with percentages. The majority of both macroscopic and microscopic models are able to model recurrent congestion and reflect the impact of any bottlenecks and the spill back of congestion.

The modelling of traffic flow in a system operating within the IVHS technology must be combined with a traffic routing component so traffic having a specific destination can be directed into the network realistically. Such a component is present in a few models like INTEGRATION, CONTRAM and

SATURN (Underwood 1990). In addition to this, it must be possible to model traffic holdback due to bottlenecks caused by either excessive demand or traffic incidents, so the interaction of traffic conditions on adjacent links can be included in the model.

The development of this model was made under these constraints. The model traces individual vehicles, and each vehicle maintains its identity throughout the network. By keeping the identity of individual vehicles, it was possible to assign specific attributes to each vehicle identifiable at any point, as well as to collect statistics on individual trips. Such attributes include the destination of the trip, the vehicle type, the generation time, and so forth.

From the point in time that a vehicle enters the network, it is assigned to a destination node and to a vehicle category. The origin destination matrix, denoted *OD*, from all origin nodes to all destination nodes is assumed to be a-priori known, so each vehicle can be assigned to a destination node stochastically, based on the percentages supplied in the *OD* matrix.

In this analysis two different vehicle categories are modelled, depending on the type of information that they have available during their trip on the network. The first category replicates the behavior of traffic that is not equipped with any "smart" device for receiving information regarding the traffic conditions on the network, but they base their decision on selecting their path on experience they have acquired from past utilization of the network. Such vehicles are characterized as "non-smart", as opposed to the second category of vehicles modelled, characterized as "smart". These vehicles base their path selection decision process on present as well as projected travel conditions on the network. This second category of vehicles replicates traffic that is equipped with "smart" devices for receiving and processing such information. The percentage of vehicles created from each category has to be given to the model so each vehicle will be categorized as "smart" or "non-smart" at the time of its entrance into the network.

4.3 Description Of The Simulation Model

The simulation model is discrete and event oriented, meaning that it describes changes in the system that occur at discrete points in time. The isolated point in time where the state of the system changes is the event time, and such changes are dictated by the logic describing each event. Changes in the state of the system have to occur in a time ordered sequence, and events occur at prescribed times during the simulation. Therefore the primary function of a discrete event simulation model is scheduling events in a "calendar" file, so they will be processed in chronological order, and advance the simulation clock at the time of the event in execution.

The general logic of the simulation model is depicted in Figure 4.1. The first step of the model is to read the input data file, containing information about the network, as well as the starting and ending time of the simulation run. The run is initialized by setting the simulation clock to the starting time, which coincides with the time of the very first event in the calendar. Then the first event is called for processing, and after it is executed it is removed from the calendar. The simulation clock is advanced to the time of the next event in the calendar, and this process is repeated till the simulation clock exceeds the end time at which point the simulation run is completed.

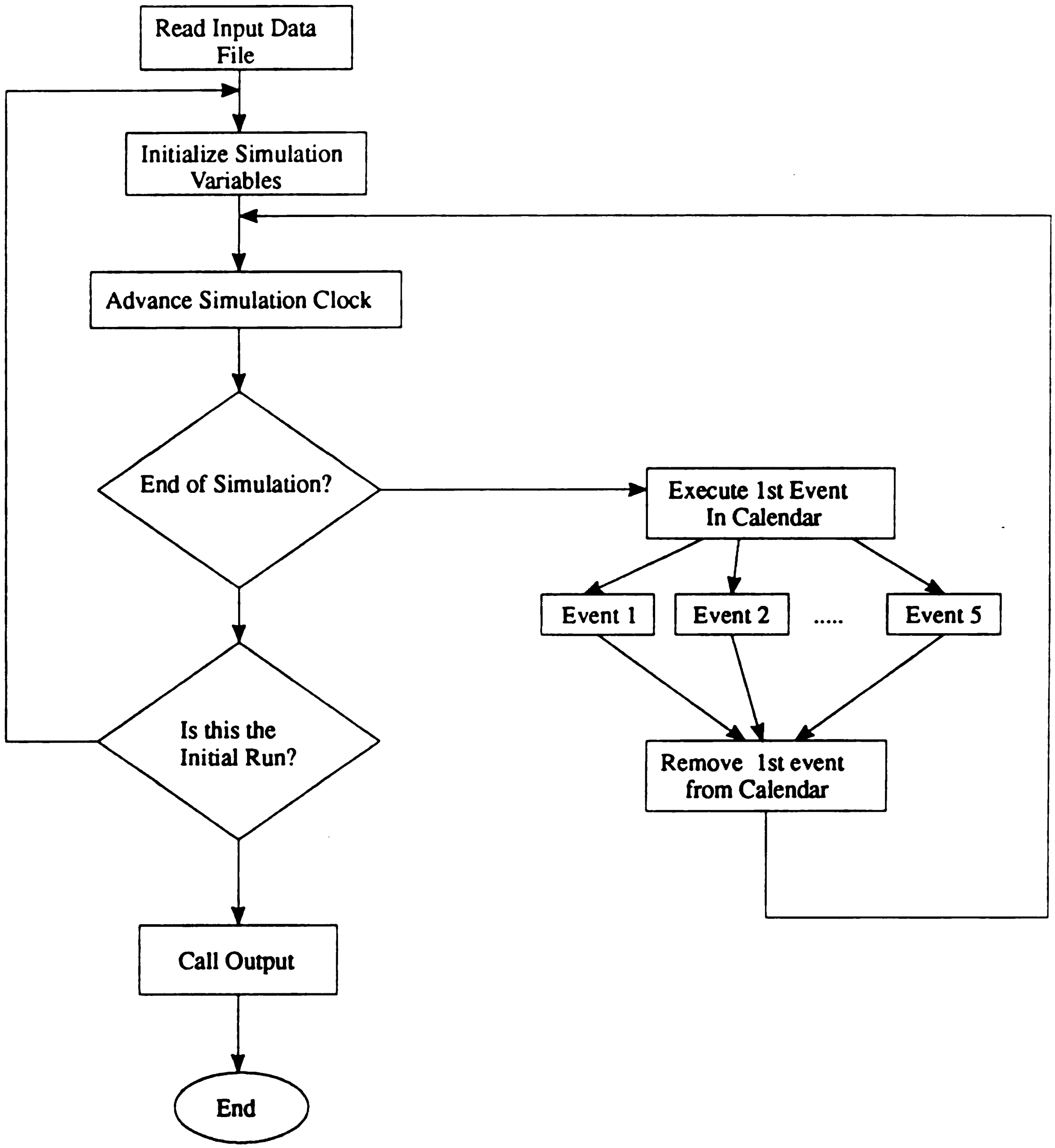


Figure 4.1: Flow Chart of Main Function of Simulation Program

The simulation model is microscopic at the vehicle level. In-link movements of vehicles are not modelled, but in essence vehicles wait on the link until their travel time is expired, and then they are transferred to the subsequent link of their path. In this level of modelling the events that were modelled are:

(1) Updating the minimum path tables of the network,

(2) A vehicle is entering the network,

(3) A vehicle is exiting a link to be transferred to another link, or exit the network if it has reached its destination,

(4) Start a traffic incident,

(5) End a traffic incident.

Both vehicle types have to base their path selection process on some data base containing travel time information for the links of the network categorized in time intervals. While this data base for "smart" vehicles is created and updated in real-time during the simulation run, for "non smart" vehicles such information has to be known a-priori, to emulate the past experience concept. In order to create such a data base, the simulation model is run twice. In the first run, in which the data base is constructed, all vehicles are considered to be "smart" and no traffic incidents are created. This concept is shown in Figure 4.1, where the first run is referred as the initialization run. The second run is the regular run where vehicles can be from either category, and the procedure for creating traffic incidents is reactivated. For convenience, the initial run can be skipped if data about average travel times are available (i.e. from previous runs of the simulation model for the same network) to be used for the travel time data base emulating the past experience information.

4.3.1 Event 1: Updating The Minimum Path Tables Of The Network

As was mentioned, the travel time data base for smart vehicles is created and updated while the regular run is taking place. This process has to be accomplished in regular time intervals, referred to as time steps, at which point specific tasks have to be executed for the proper operation of the prediction models. Therefore, the basic function of this event is that in each time step the RLS filter updates the parameter vectors and then performs a prediction of travel times on the network links, based on the most recent parameter vector.

The first step of this event (Figure 4.2) is to schedule in the calendar the next call of itself, which will be exactly one time step ahead. Following this, the average travel times, $\bar{T}_{(i,j)}(k)$, on all the links of the network *(i,j)* for the time step that has just ended (k) have to be computed, so they can be utilized for the structure of the input vectors to the RLS filter. The travel time of a link is estimated by the system each time the status of the link is changed. The status of the link is considered to be changing whenever:

(1) one vehicle enters the link,

(2) one vehicle exits the link,

(3) a traffic incident starts,

(4) a traffic incident ends and

(5) if there is a queue formation on the link, when the length of the queue is altered.

During one time step k the average travel time of a link *(i,j)*, $\bar{T}_{(i,j)}(k)$ is calculated as the weighted average of the values of the travel time that it assumes due to changes in the link status, $T_{(i,j)}(k-1+d_m)$, where d_m is the elapsed time since the start of the last time step, when a new sample measurement

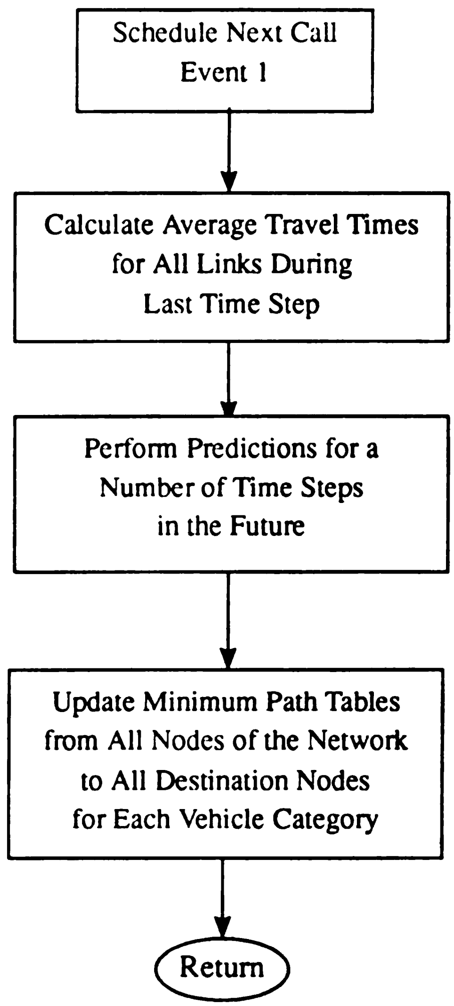


Figure 4.2: Flow Chart of Event 1: Update the Minimum Path Tables

was acquired. The weights used for the weighted average are the time lengths for which each sample measurement of the travel time of the link was valid. If we have M-1 changes in the status of the link, this is expressed in the following equations:

$$\bar{T}_{(i,j)}(k) = \frac{\sum_{m=0}^{M} T_{(i,j)}(k-1+d_m) \cdot (d_{m+1}-d_m)}{\sum_{m=0}^{M} (d_{m+1}-d_m)} \tag{4.2.a}$$

where:

$$\sum_{m=0}^{M} (d_{m+1}-d_m) = 1\,\text{Time Step} \qquad d_0 = 0 \qquad d_M = 1\,\text{Time Step} \tag{4.2.b}$$

and

$$d_{m+1} \geq d_m \qquad d_m \leq 1\,\text{Time Step} \qquad \forall d_m, m \in [0, \ldots, M] \tag{4.2.c}$$

The next step is to call the filter routine so the parameter vector of the prediction model of each link will be calculated, so predictions for a number of time steps in the future can be performed. With this information the simulation model calls the minimum path routine, and constructs the minimum path tables. The minimum path routine finds the minimum path from every node in the network to all destination nodes, by considering only time "costs" and "savings" over alternative paths. This is based on the assumption that drivers are only interested in minimizing their travel time toward their destination.

It is recognized that other factors affect the final choice of the actual path that each individual driver is going to follow - like safety considerations, familiarity with the road system, trip purpose and so forth. While it would be possible to include some of these behavioral aspects in the cost function (instead of just travel time), full account of the drivers decision making process would not be completely modelled, due to the complexity of such a process. In addition to this, and most important task in this analysis, the results of the simulation regarding the performance of the RLS filter in predicting travel times with adequate accuracy, would be clouded with possibly irrelevant information.

The minimum path routine examines all possible paths from a node i_l to a destination node i_{n+l}. The travel time of each path is calculated by considering the cost of each link for the entire downstream path, depending on the expected time of arrival at the entrance of the link (starting node of link). For example, for the path from node i_l to destination node i_{n+l} as given in equation (4.1), the cost of the path is calculated as:

$$C_{(i_1 \rightarrow i_{n+1})} = \sum_{j=1}^{n} \hat{T}_{(i_j, i_{j+1})}(k_j) \tag{4.3.a}$$

where:

$$k'_1 = 0 \qquad k'_{j+1} = k'_j + \hat{T}_{(i_j, i_{j+1})}(k_j) \qquad k_j = \left[k'_j\right] + 1 \tag{4.3.b}$$

and $\hat{T}_{(i_j, i_{j+1})}(k_j)$ is the expected travel time of link (i_j, i_{j+l}) at k_j time steps ahead. The brackets in (4.3.b) denote the integer part of a real number. Equation (4.3.a) takes into account the dynamic nature of the traffic network. For "non smart" vehicles the cost of a path is calculated in the same manner, but instead of using the expected travel time of the links included in the path, the travel times as calculated in the initial run are used:

$$C_{(i_1 \rightarrow i_{n+1})} = \sum_{j=1}^{n} \overline{T}_{(i_j, i_{j+1})}(q_j) \tag{4.4.a}$$

where:

$$q'_1 = \frac{\text{Current Time}}{\text{Time Interval}} \qquad q'_{j+1} = q'_j + \bar{T}_{(i_j, i_{j+1})}(q_j) \qquad q_j = \left[q'_j\right] + 1 \qquad (4.4.b)$$

and $\bar{T}_{(i_j, i_{j+1})}(q_j)$ is the average travel time on link (i_j, i_{j+1}) at the q_j time interval. Here it should be noted that the time interval q has a different length than the time step k used in equation (4.3.a). Thus, the value of q can be large enough, so the classification of travel conditions on the links of the network in the memory of the "non smart" driver would be reflected. Because of the way path costs are calculated (equations (4.3.a) and (4.4.a)), static shortest path algorithms (like Dijkstra's or Dreyfus' algorithms, the label correction method, the label reaching method or the recursive -fixing method) could not be applied. This becomes obvious if we let $Y_{i,j}$ be the minimum time to travel from node i to j and $T_{(k,j)}$ be the cost to travel link (k,j). Note that Y_{ij}, and $T_{(k,j)}$ are not dependent on the time that the vehicle starts the trip or it enters the link, since in this case the network is considered to be static. In a network with M nodes, the basic principal of such algorithms is (Romeijn and Smith 1990, Hall 1986):

$$Y_{ij} = \min_{j \neq i}\{Y_{ik} + T_{(k,j)}\} \qquad \text{for} \qquad k = 1, \ldots, M \qquad (4.5)$$

The underlying meaning of this principal is that the optimum path towards a node - for example node j - is also optimum for the nodes contained in it - for example node k (Kaufman, and Smith 1990). However, in a dynamic network, such as the one that we have to model, the fastest path to reach a node (k) may not be the fastest path if the ultimate destination is a subsequent node (j), since the cost of link (k,j) may be significantly larger if we reach (k) early than if we reach it a few time units later. This is illustrated in Figure 4.3, where we can reach node 4 via paths {(1,2),(2,4)} and {(1,3),(3,2),(2,4)}. Although the shortest path to reach node 2 is {(1,2)} with cost 10 time units, the total cost to

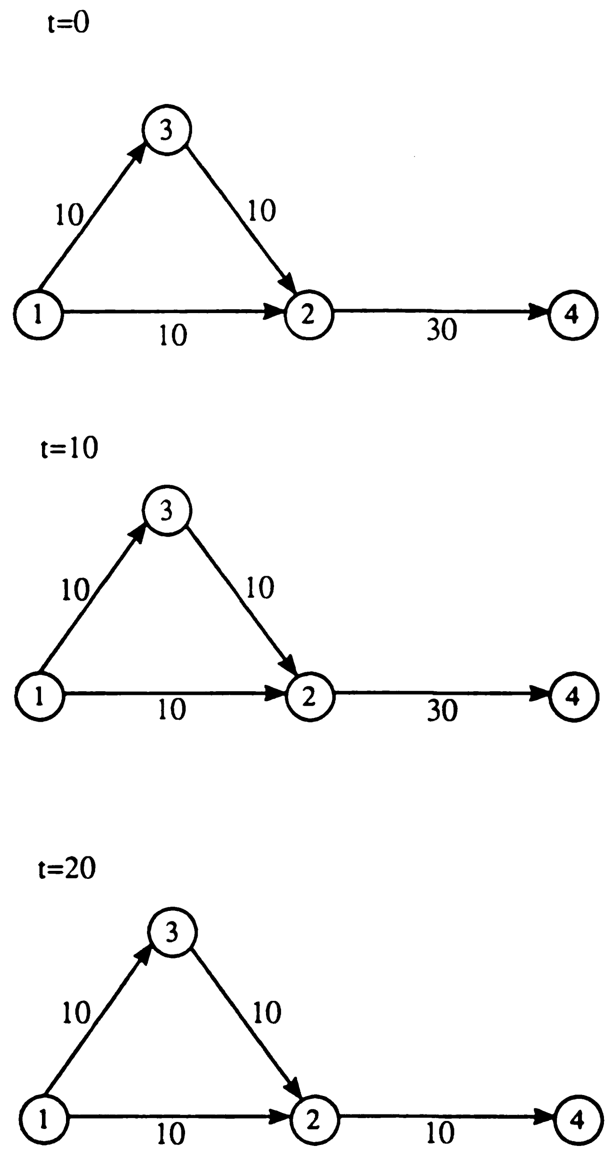


Figure 4.3: Violation of Principle of Optimality in Dynamic Networks

node 4 if we arrive at this point in time 10 is 40 while if we arrive 10 time units later it will drop to 30.

In a more dynamic approach Cooke and Halsey (1966) tried to solve the problem by starting from the destination node and considering the shortest path to reach it progressively towards the starting node. However, they still considered fixed travel time. Hall (1986) has suggested a method for networks with time varying random travel times that are boundable from below, by combining a branch and bound technique with an M shortest paths algorithm (algorithm that finds the M best paths between two nodes). This method would only have advantages (in terms of computational effort) over an exhaustive search if the computation of the expected travel times of paths is difficult. In our model expected values of travel times are produced by the RLS filter anyway.

Kaufman and Smith (1990) suggested a modification on Dreyfus' algorithm so it could be used in dynamic networks. This was achieved under the assumption that each intermediate node of the optimum path has to be reached as early as possible. In this algorithm costs are time variant, but they must satisfy (under the assumption stated above) the following restriction:

$$s + T_{(i,j)}(s) \leq r + T_{(i,j)}(r) \qquad \forall s \leq r \tag{4.6}$$

where $T_{ij}(s)$ is the cost of link (i,j) at time s. This assumption is necessary to avoid the case described in Figure 4.3. It is argued that it is not too restrictive because driver behavior is not changing in such a dynamic pattern.

In order to avoid such assumptions which may not be valid in the case of non recurrent congestion, and to cover cases like the one described in Figure 4.3, the only feasible solution appears to be an exhaustive search of the cost of all possible acyclic paths from a node towards a destination node. The term acyclic path is used to indicate that paths where vehicles circle in a loop waiting for the cost of a subsequent link to drop are not allowed. This is accomplished by using a depth-first algorithm for identifying the possible paths, and as soon as a path is identified its cost is computed with equations (4.3.a) or (4.4.a). This cost is compared with the cost of previously identified paths towards the same destination and if it is found to be smaller it replaces the old value. This process continuous until all possible paths are evaluated at which point the minimum paths in the dynamic network are found (Paull 1988).

After the establishment of a better path the two first nodes of the path, excluding the starting node, are kept in memory. This is done for all the nodes of the network for both vehicle categories, so two sets of minimum path tables are constructed, on for each category. Each set of tables consists of (i) the first node matrix, denoted F_c, with c=1,2 (for each vehicle category) and defined as:

(1) The rows of the matrix correspond to the nodes of the network, and the columns correspond to the destination nodes,

(2) f_{ij} with $i \in N$ and $j \in D$, is the first node of the minimum path to be followed if the trip starts at node i and has destination the node j,

and (ii) the second node matrix, denoted S_c, and defined as:

(1) The rows of the matrix correspond to the nodes of the network, and the columns correspond to the destination nodes,

(2) s_{ij} with $i \in N$ and $j \in D$, is the second node of the minimum path to be followed if the trip starts at node i and has destination the node j.

These tables are used as reference during the following time step, and they are updated in each time step. While the significance of the first node is obvious since it indicates the link to be followed immediately after arriving at the end of a link, the importance of the second node will be discussed later.

4.3.2 Event 2: Entrance Of Vehicle In The Network

The logic describing event 2 is shown in Figure 4.4. As soon as a vehicle enters the network from one of the generation nodes, the next vehicle generation at the same node is scheduled. In this way, the model does not need to keep in memory all the entities to be created until the end of the simulation run, but only those that are on the network.

To capture the dynamic nature of traffic networks, in terms of traffic demand, a dynamic generation rate (rate with which vehicles will be entering the network) was utilized. Such dynamic behavior was reflected in the modelling process by utilizing generation rates that are dependent on the time of the day. Arrivals are created according to a Poisson distribution (exponentially distributed interarrival times), with parameter based on the demand level at the current time of the simulation clock, while the demand curve of each generation node is assumed to be known apriori.

Following the scheduling of the forthcoming vehicle from the current generation node, the entering vehicle was assigned to a link, thus increasing the number of vehicles currently traversing this link. The selection of this link is done by referring to the minimum path tables created in event 1. Based on the vehicle category c, the entering vehicle is assigned to the link defined by the entering node and the node f_{ij} from the first node matrix F_c, where i is the

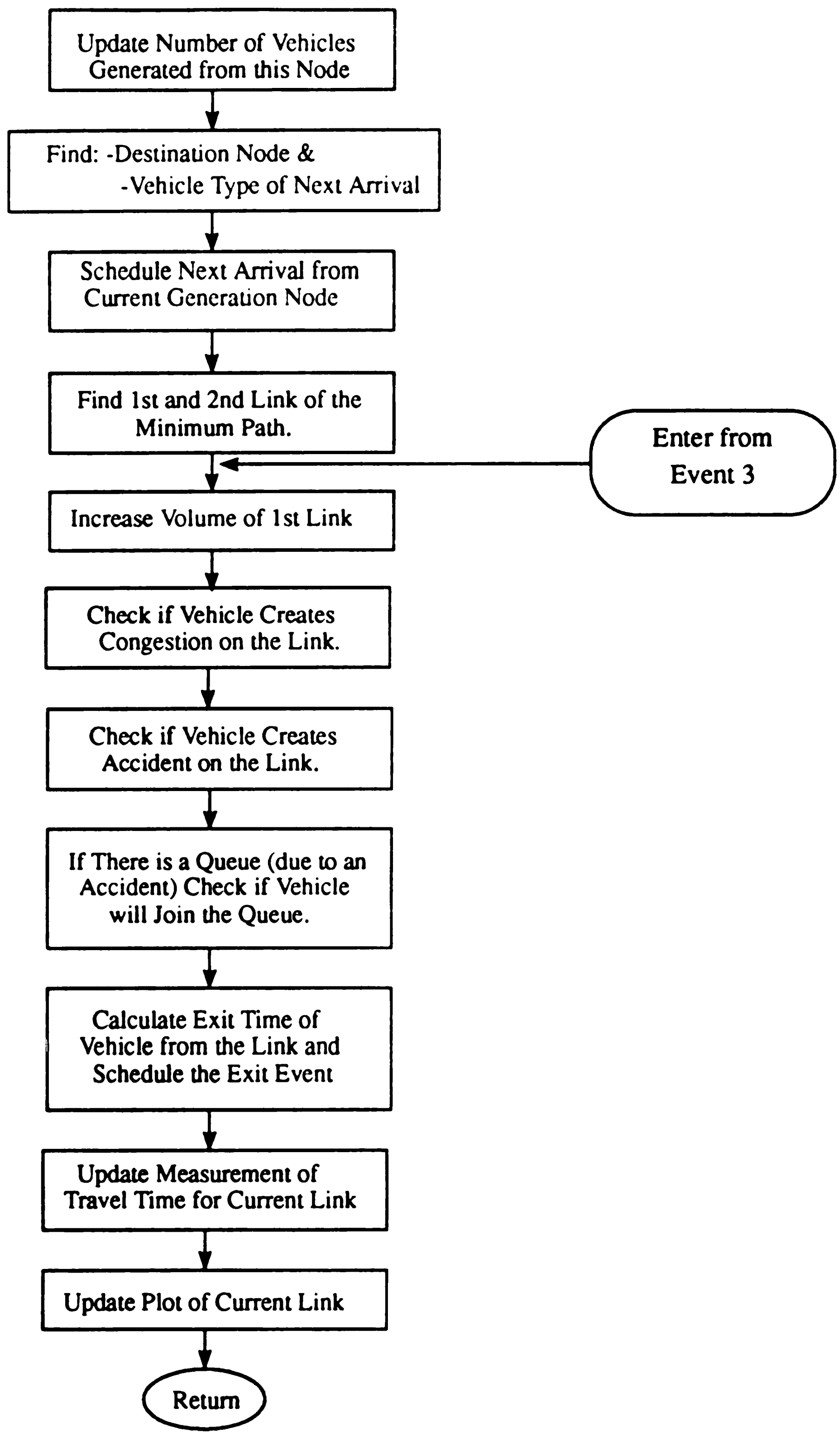


Figure 4.4: Flow Chart of Event 2: Entrance of a Vehicle in the Network

entering node and j is the destination node. In addition to this, the second node of the current optimum path towards the destination j, s_{ij} from the second node matrix, is stored as an attribute of this vehicle.

As was mentioned earlier, the traversing of a link by a vehicle is simulated by storing the vehicle on the link for a period of time equal to its travel time. The remaining functions performed in this event have as a goal the computation of travel time on the link. These functions are executed not only when a vehicle enters the network but also each time it is transferred to a link from another link. The calculation of the travel time is based on the Bureau of Public Roads travel time function, shown in the following equation (TRB SR165):

$$T_{(i,j)} = FFTT_{(i,j)} \cdot \left(1 + 0.15 \cdot \left(\frac{v_{(i,j)}}{c_{(i,j)}}\right)^4\right) \tag{4.7}$$

where $T_{(i,j)}$ is the travel time of the link (i,j), $FFTT_{(i,j)}$ is the Free Flow Travel Time of the link, and v/c is the volume over the capacity ratio for the link. Free flow travel time is defined as the time required by a vehicle to traverse the link if this vehicle is the only one on the link.

To make the above equation usable for our purposes, we introduce two new variables to replace v and c in equation (4.7). First the Occupancy ($OCC_{(i,j)}(t)$) of the link is defined as the number of vehicles traversing the link at a specific point in time. Thus:

$$OCC_{(i,j)}(t) = v_{(i,j)}(t) \cdot dt \tag{4.8}$$

The second variable is the Free Flow Maximum Occupancy ($FFMO_{(i,j)}$) of the link, which is defined as the maximum number of vehicles that can "fit" on the link at any point in time, and have traffic still move freely. Since capacity (c) is defined for a given time period, $FFMO$ can be regarded as the instantaneous capacity or:

$$FFMO_{(i,j)} = c_{(i,j)}(t) \cdot dt \tag{4.9}$$

From (4.8) and (4.9) equation (4.7) can be written as:

$$T_{(i,j)}(t) = FFTT_{(i,j)} \cdot \left(1 + 0.15 \cdot \left(\frac{OCC_{(i,j)}(t)}{FFMO_{(i,j)}}\right)^4\right) \tag{4.10}$$

Equation (4.10) does not take into account delays due to queuing, and it is operational until traffic flows with some minimal speed. As the occupancy on the link increases so does the travel time, and when the link approaches the congestion state small increases in the occupancy result in significant increases in the travel time (Figure 4.5). In order to prevent excessive travel times due to unrealistic occupancies, it was assumed that the maximum occupancy that a link can handle is twice the free flow maximum occupancy. As shown in Figure 4.5, this will result in a travel time at capacity that is 3.40 times greater than the free flow travel time, while when the occupancy is equal to the free flow maximum occupancy the travel time is increased by 1.15 times the free flow travel time. The factor which sets the ceiling in the occupancy of the link, denoted as f_{max}, is a parameter of the model and can be altered by the user.

When a link is in the congestion status, it can exit only when the occupancy on the link drops back to 1.70 times the free flow maximum occupancy. This factor, denoted as f_{free}, is also a parameter that can be altered by the user.

Therefore, the first step after a vehicle is assigned to a link is to determine whether or not this vehicle will congest the link that it just entered. If it does, then the link is assumed to be closed and no other vehicle is allowed to enter

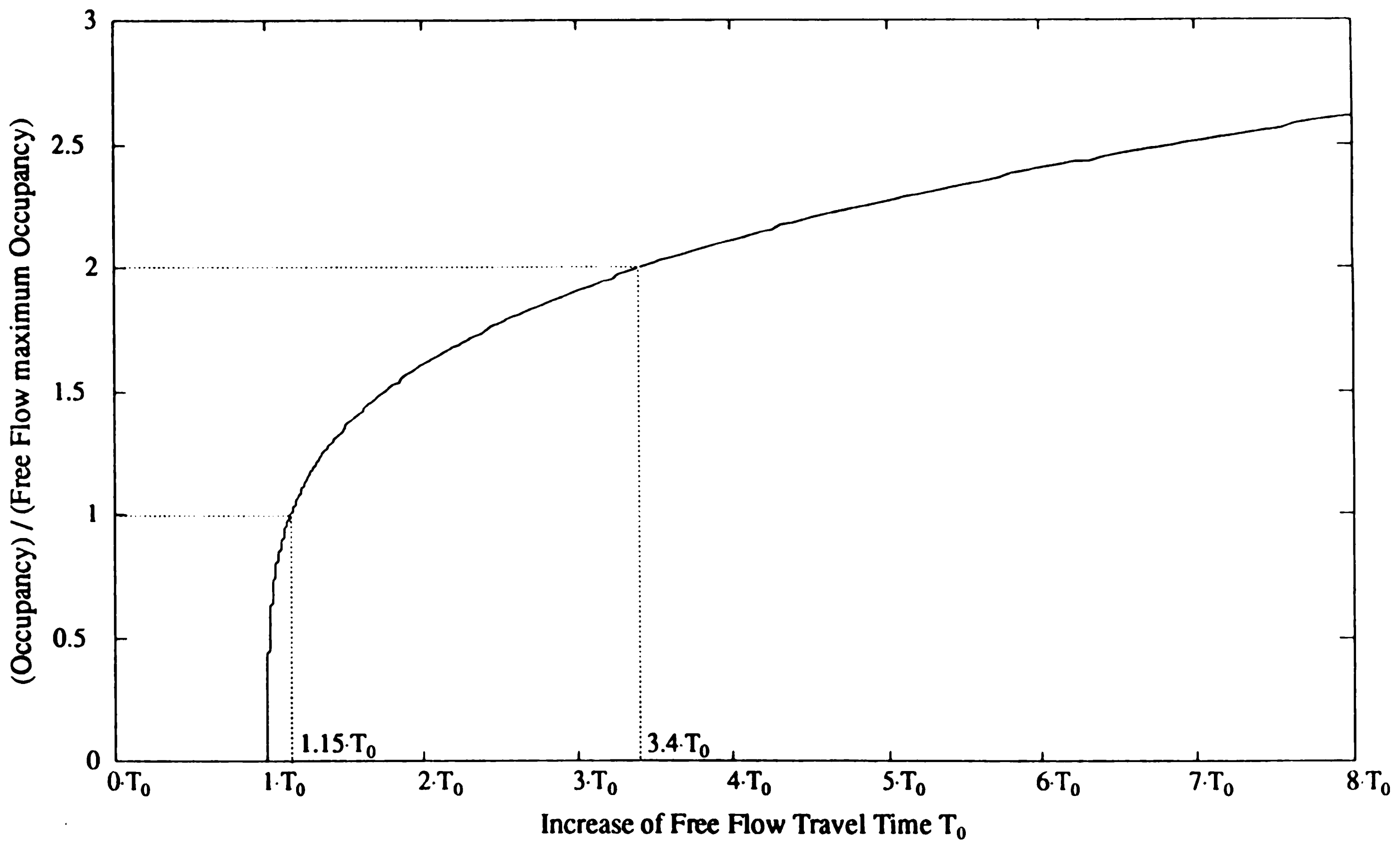


Figure 4.5: Relationship Between Link Travel Time and Link Occupancy Based on the BPR Function

until the link leaves the congestion status.

The next step is to consider if the entering vehicle is about to create a traffic incident. A traffic incident can be caused only by a vehicle, and if there are no entities traversing a link, the probability of having an incident is zero.

Two types of traffic incidents are assumed to occur in this model (although more types could be included): traffic accidents which result in the complete blockage of the link, and vehicle breakdowns which result in partial blockage of one lane of the link. The probability of a vehicle causing a traffic accident P_a, the probability of causing a vehicle break down P_b, or the probability of not causing any incident P_0, along with the average duration of each incident type are assumed to be known for each link of the network. Obviously $P_a+P_b+P_0=1$. In this analysis the probability that a vehicle will create an incident is assumed to be 0.00010/lane-mile, from which 80% of the incidents that occur are vehicle breakdowns and 20% traffic accidents. This ratio of traffic incidents is obtained from the result reported by Guiliano for high volume urban freeways (Guiliano 1989). The probabilities used are higher than real accident rates, but this was done intentionally, so there will be some incident occurrences on the network within a single simulation run.

After the entrance of the vehicle in a link, a multinomial random variable $(P_a;P_b;P_0)$ is constructed to determine if the vehicle will indeed cause an incident, and the type of the incident. If an incident is about to be initialized, the point at which the vehicle causes the incident on the link (*RLOC*) as a percentage of the length of the link is found. This is also a random variable, uniformly distributed with $U[0,1]$. The starting time of the incident depends on the percentage of the length of the link that the vehicle has travelled before the incident. The average duration of each incident type is assumed to be different, as vehicle breakdowns, although less serious in severity and capacity reduction, may last longer due to lower priority for service by emergency units (Koutsopoulos and Yabloski 1991). Due to the great variance in incident duration it was thought reasonable to assume that the distribution of the duration of traffic incidents is exponential, with parameter known. The start and the end of the traffic incident are scheduled as separate events in the calendar of the simulation. Because of the memory requirements, and the complexity of multiple queue formations on the same link, it was assumed that the occurrence of two incidents of the same type on the same link simultaneously is not allowed. This is not too restrictive, since the probabilities of having an incident is already very small.

If the occurrence of a traffic accident has been initialized on the link, the link is assumed to be completely blocked, and a queue starts forming behind the point where the accident occurred. Upon the entrance of a vehicle in the link, in the case that a queue exists on the link, a check is made to determine whether or not the vehicle will join the queue. If the vehicle is the one that causes the accident then it will also be the one that initializes the formation of the queue. After this vehicle, each vehicle that enters the link is checked to determine if it will reach the end of the queue before the formed queue has dissipated.

If the vehicle entered the link at time t then the time required to reach the end of the queue (TRQ) is calculated as a percentage of the free flow travel time of the link, as if all vehicles entering before the one under consideration, have already joined the queue:

$$TRQ(t) = FFTT_{(i,j)} \cdot \left(RLOC_{(i,j)} - \frac{Q_{(i,j)}(t)}{FFMO_{(i,j)} \cdot f_{max}} \right) \tag{4.11}$$

where $Q_{(i,j)}(t)$ is the length of the queue on link (i,j) at time t. This technique produces somewhat shorter travel times for the vehicles to reach the end of the queue, but the differences are considered to be insignificant and it simulates the worse (fastest) case scenario for reaching the end of the queue. Considering the time that the accident will end, the current time of the simulation, and the time to reach the end of the queue we can determine if the queue have started to disperse when the vehicle reaches the last car in the queue. If the queue has started dissipating, we need to examine if, at the point in time that the vehicle reaches the queue, there is any queue left. The time required for the formed queue to be dissipated at time t, is calculated simply as:

$$T_{qd}(t) = \frac{r_d \cdot Q_{(i,j)}(t)}{L_{(i,j)}} \tag{4.12}$$

where r_d is the queue dissipation rate. r_d is a parameter defined by the user and in this analysis is assumed to be equal to 4.8 sec/veh. By comparing the time required to dissipate the queue with the time to reach the end of the queue we can determine if the vehicle will join the queue or not. If the vehicle is going to join the queue the queue length is increased by one.

In the situation of an accident on the link, the number of vehicles that occupy the portion of the link behind the point where the accident occurred, is checked to see whether it has reached the maximum capacity of this portion of the link. If so, the link enters the congestion status and no other vehicle is allowed to enter the link.

After completing these checks, the travel time of the vehicle is calculated. As was mentioned, the travel time calculated by equation (4.10) does not take into account delays due to queueing or delays due to accidents, so these delays have to be added to the travel time calculated by equation (4.10). When a vehicle breakdown occurs, it is assumed that the only effect that it has is to reduce the free flow maximum occupancy of the link (or equivalently the capacity of the link) for the duration of the incident. If at time t there is a vehicle breakdown half a lane will be closed, so the free flow maximum occupancy of the link will be reduced by a factor $a(t)$, where:

$$a(t) = \begin{cases} \dfrac{(L_{ij} - 0.5)}{L_{ij}} & \text{if there is a vehicle break down on link } (i,j) \\ 1 & \text{if there is no vehicle break down on link } (i,j) \end{cases} \tag{4.13}$$

Therefore, equation (4.10) takes the form:

$$T_{(i,j)}(t) = FFTT_{(i,j)} \cdot \left(1 + 0.15 \cdot \left(\frac{OCC_{(i,j)}(t)}{FFMO_{(i,j)} \cdot a(t)}\right)^4\right) + \\ + T_{acc}(t) + T_{qd}(t) \tag{4.14}$$

where $T_{acc}(t)$ is the remaining duration of the accident at time t, if there is an accident occurrence on the link. Note that $T_{acc}(t)$ will be added to the travel time of an entering vehicle even if the accident has not started yet, but a vehicle that has already entered the link is about to cause the accident. In this case the full duration of the accident is added. Further more, in order to control the rate with which vehicles exit the link the exit time form the same link of the previous vehicle is recorded, and if the difference between the scheduled exit of this vehicle minus the scheduled exit time of the previous vehicle is smaller than a threshold value w, the travel time is increased so this difference will be equal to w. Such a threshold value, which represents the headway, is assumed to be fixed at one second per lane. Then the exit of the vehicle is scheduled in the calendar of the simulation based on the current time of the simulation clock and $T_{(i,j)}(t)$:

$$T_{exit} = \begin{cases} T_{(i,j)}(t) + t & \text{if } T_{previous} - T_{(i,j)}(t) - t \geq w \\ T_{previous} + w & \text{if } T_{previous} - T_{(i,j)}(t) - t < w \end{cases} \tag{4.15}$$

where T_{exit} is the exit time of the vehicle from link (i,j), and $T_{previous}$ is the exit time of the previous vehicle from the same link.

To simulate the way that the system control center is measuring the travel time of the link, (so the average of the travel time in each time step utilized by the prediction model can be calculated) equation (4.14) has to be modified. Under the assumption that the network system is equipped with detectors and counters, so current occupancies, length of queues and vehicle breakdowns are detectable immediately, the only variable that will still contain some uncertainty in equation (4.14) is the remaining duration of the accident. The remaining duration of the accident will be calculated as the conditional expectation of the duration of the accident, given that it has not ended yet. Under the assumption that this duration is exponentially distributed the expected remaining time is equal to the mean of the distribution. This is due to the well known *"memoryless property"* of the exponential distribution (Ross 1985). Thus for the calculation made by the system control center, $T_{acc}(t)$ will be replaced by its expected value, denoted with a tilde:

$$\tilde{T}_{acc}(t) = \begin{cases} E[d_{(ij)}] & \text{accident has started and has not ended yet} \\ 0 & \text{otherwise} \end{cases} \tag{4.16}$$

where $E[d_{(i,j)}]$ is the expected duration of an accident for link (i,j). If the accident has not started yet then $\tilde{T}_{acc}$ is equal to zero (no effect before the start of the accident).

4.3.3 Event 3: Exit Of A Vehicle From A Link

After the completion of the travel time of a vehicle on a link, the vehicle is transferred to a subsequent link towards its destination node. When a vehicle is about to exit its current link there are several functions to be executed to check if the vehicle is able to leave this link, and to determine which link it will be transferred to.

As was described in event 2, queues can be formed because of an accident. Besides accidents, recurrent congestion can cause formation of queues on the network, and this aspect has to be included in the model. Generally, when a vehicle tries to exit from its current link to another one, and the new link is blocked, then this vehicle will start the formation of a queue right at the exit point of its link. Such a queue is discriminated from a queue created due to accidents which can start at any point on the link, and thereafter will be referred to as congestion queues.

The procedures followed in this event are depicted in Figure 4.6. Before the vehicle exits the link a check is made to determine if there is a congestion queue on the link, and if the exiting vehicle belongs to this queue. If it does,

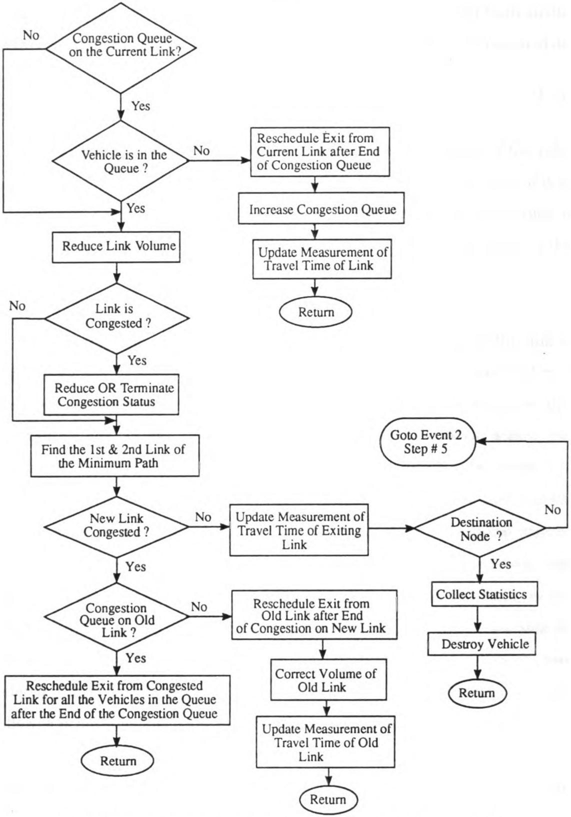


Figure 4.6: Flowchart of Event of Exit of a Vehicle from a Link.

then the exit is rejected and the vehicle joins the end of the congestion queue. In this case the exit time of the vehicle from the current link is reevaluated as:

$$T_{newexit} = T_{previous} + \frac{r_d}{L_{(i,j)}} \tag{4.17}$$

where the second term takes into account delays due to dissipation of the vehicles from the queue, and $T_{previous}$ is the exit time of the vehicle in front of it in the congestion queue. If the vehicle belongs in the congestion queue then it may be allowed to exit, if the new link that it will be assigned to is not in the congestion status.

In the case that the vehicle exits the link, then the occupancy of the link is reduced by one. If the link that the vehicle just left was congested and/or if there was a congestion queue on this link, then these variables are also reduced by one. In the case that the link is in the congestion status the reduced occupancy is checked to see whether it is low enough to terminate the congestion status on the link. Following the exit from the current link a check is made to see if the vehicle has arrived at its destination node. If the current node is not the destination of the vehicle then is assigned to a new link, and the second part of event 2 is executed for this vehicle and the new link. On the other hand, if the vehicle has arrived at its destination, statistics regarding its travel time are collected and the vehicle is removed from the network and deleted from the simulation calendar. Information on the path that the vehicle followed is not collected because of the large requirements for memory.

If the new link that the vehicle is assigned to is already in the congestion status then the vehicle cannot exit its current link, and it starts forming a congestion queue. In such a case there are two possible situations: (1) The old link does not currently have a congestion queue, and (2) The old link already has a congestion queue. In the first case the exiting vehicle will start the formation of a congestion queue and it will reschedule its exit, depending on the time that the link to which it was assigned to, but could not enter, will exit the congestion status. For example, if the vehicle was on link *(i,j)* and it is assigned to link *(j,k)* which is congested, then the new exit time of the vehicle from link *(i,j)* is calculated as:

$$T_{newexit} = T_{(j,k)free} + \frac{r_d}{L_{(i,j)}} \tag{4.18}$$

where $T_{(j,k)free}$ is the time that link *(j,k)* will exit the congestion status. In the second case the exiting vehicle will trigger the rescheduling not only of its own exit from the link, but the exit of all the vehicles from the current link as well. Again the exit time of the first vehicle in the queue will be calculated by equation (4.18), while subsequent vehicles will be scheduled to exit based on equation (4.17).

Each time that a vehicle exits a link the status of the link is changed. Therefore, the current travel time of the link has to be reevaluated so it can be utilized in the calculations of the average travel times. Two additional terms have to be added to equation (4.14), so delays due to congestion queues and delays due to congestion on successor links will be taken into account. Delays due to congestion queues are calculated in a similar fashion as in equation (4.12). For the calculation of delays due to congestion in a subsequent link, the time that the new link will exit the congestion status is utilized, as in equation (4.18). Thus, for the calculation of the current travel time of a link *(i,j)* as the system control center perceives $\tilde{T}_{ij}(t)$, it will be given by the following expression:

$$\tilde{T}_{(i,j)}(t) = FFTT_{(i,j)} \cdot \left(1 + 0.15 \cdot \left(\frac{OCC_{(i,j)}(t)}{FFMO_{(i,j)} \cdot a(t)}\right)^4\right) + \tilde{T}_{acc}(t) + \\ + T_{qd}(t) + T_{cqd}(t) + T_{c\,(j,k)}(t) \tag{4.19}$$

where $T_{cqd}(t)$ is the delay due to congestion queues and $T_{c(j,k)}(t)$ is the delay due to congestion to the successor link (j,k) at time t. These variables have value if there is a congestion queue on the link.

In the case that the vehicle is allowed to exit from its current link (there is no congestion queue on the link or the vehicle is in the top of the queue if such a queue exists) the model has to decide to which successor link the vehicle will be assigned. The selection of the remaining path towards its destination node is made in a time adaptive fashion. This means that, each time the vehicle reaches a new node, new information is acquired, so its path towards its destination is reevaluated. The decision process of selecting the next link of the path is depicted in Figure 4.7.

A first indication of a possible next link is acquired from the first node matrix F_c for the appropriate vehicle category, from the minimum path tables of the network. The link defined by the end node of the current link i and the node f_{ij} of the matrix F_c, where j is the destination node of the vehicle, will be referred to as the new link for simplicity. If the new link has entered the congestion status in the last time interval, then the minimum path from node i to node j is reevaluated and the first and second nodes of the new minimum path replace the corresponding elements of matrices F_c and S_c. This step is applied for both vehicle types, since it is assumed that even "non smart" vehicles are able to detect if the next link of their path is blocked. This way vehicles will

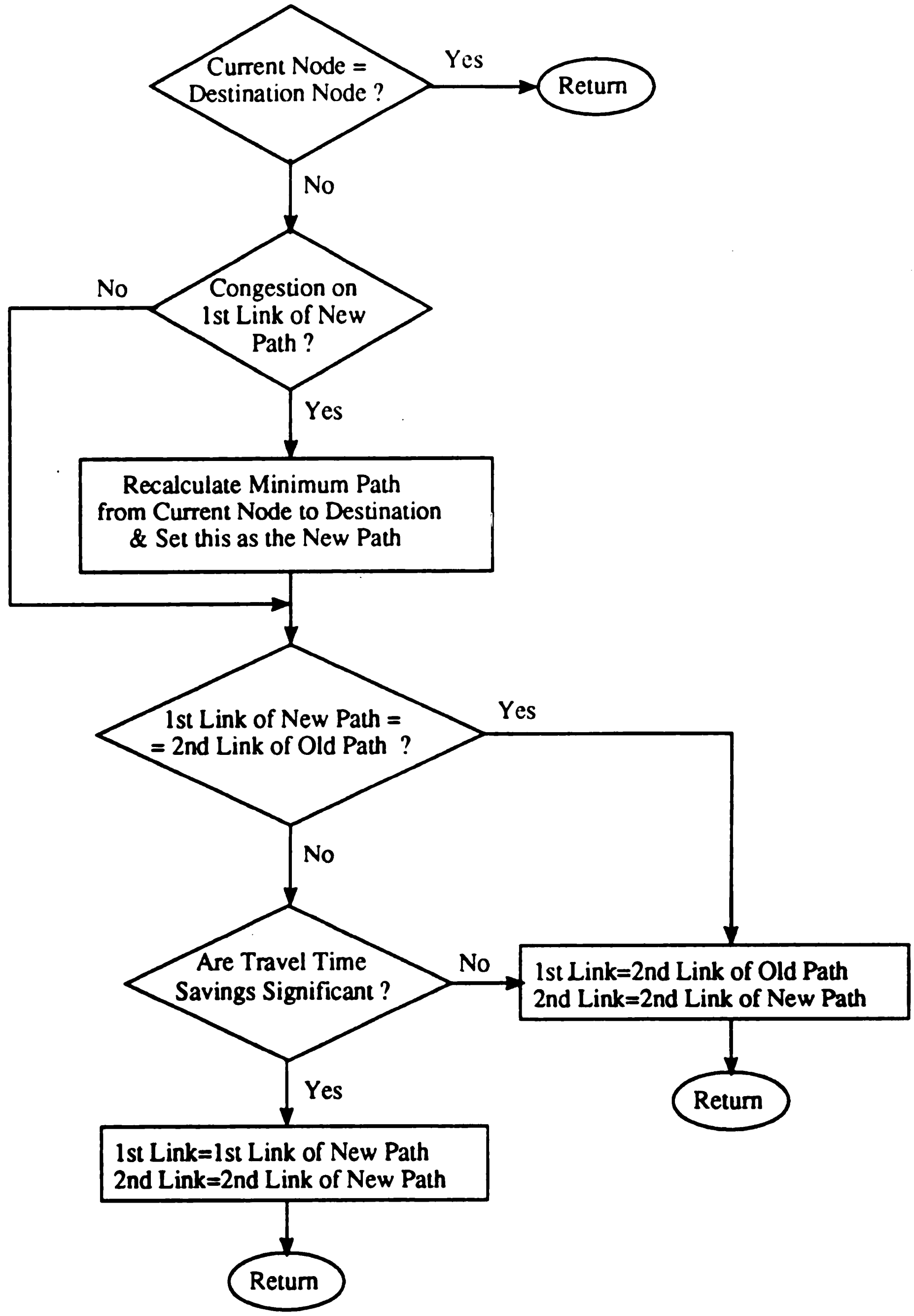


Figure 4.7: Flow Chart of Selection of Next Two Links of Path

not start forming congestion queues if there is an alternative path more attractive than the one with the congested link.

Vehicles should not change paths for very small gains in their travel time. If vehicles always followed their minimum path, then the new link would be the one to which a vehicle would be assigned. But, the assumption that drivers are willing to change paths along their way in pursuit of any gain, no matter how insignificant, is too extreme, and as suggested by Mahmassani and Chen (1991), the driver's switching behavior exhibits a rather bounded rational character anchored in his current path. In this model a relative indifference band similar to the one developed by Mahmassani and Jayakrishan (1990) will be utilized:

$$\delta = \begin{cases} 1 & \text{if} \quad Y_{old} - Y_{new} > max\{\alpha \cdot Y_{old}, \tau\} \\ 0 & \text{otherwise} \end{cases} \tag{4.20}$$

where Y_{old} is the travel time from the old minimum path, Y_{new} is the travel time of the new minimum path, α is the relative indifference band as a percentage of its remaining travel time from the old path to its destination, and τ is the minimum gain over the old path travel time, required to change from the old path to the new. The vehicle will switch to the new path if δ is equal to 1, and it will not if δ is equal to 0. Parameters α and τ can be specified for each vehicle category individually, but for simplicity in the application runs of the model they were kept equal.

The prevailing requirement for rule (4.20) to be applicable is knowledge of the old minimum path. Because of memory restrictions, the entire path selected at each node for each vehicle in the network, could not be kept in memory.

Therefore it was assumed that if the first link of the new path is the same as the second link of the old path, the two paths are equal. In the case where these two links are not the same, it is obvious that the two paths are not the same.

Thus, the new link will be compared with the link defined by the nodes f_{ij} and s_{ij} which the vehicle had kept as attributes when it entered the link that it just traversed. If these two links are the same then it is assumed that there is no change in the path and the vehicle is assigned to the new link. If the two links are different then rule (4.20) is utilized to define if the time saving from the new path is significant.

4.3.4 Events 4 & 5: Start& End Of A Traffic Incident

A traffic incident is defined as any event that causes non recurrent congestion or a reduction in the flow rate. The effects of traffic incidents may vary significantly from complete blockage of a link to vehicle breakdowns blocking only a portion of a lane. The duration of such incidents depends on the severity of the incident, as well as the placement of emergency units. The importance of simulating traffic incidents is evident, if we consider that benefits of ADIS will be greatest during the occurrence of such events.

As was described in event 2, when a vehicle enters a link it is examined to see if it will cause the initialization of a traffic incident. In the case that it does initialize an incident, the incident will not start immediately but after some random portion of the vehicles' travel time has elapsed. At the point in time that the incident starts the parameters of the link on which the incident occurs have to change. Depending on the type of the incident, the effect is reflected in the travel time equation, as reductions in the free flow maximum occupancy (by the factor $\alpha(t)$) in the case of a vehicle break down, or as delays, that are additive to the travel time of the link (delays due to queues $T_{qd}(t)$) in the case of a traffic accident.

When the incident ends, the free flow maximum occupancy assumes its original value in the case of a vehicle break down, or in the case of a traffic accident the process for dissipating the queue formed behind the location of the accident is initialized. In the case that the queue formed behind the location at which the accident had blocked the link, the link is in congestion status. In this case the new length of the queue is examined to see whether it is short enough for the link to exit the congestion status.

The basic modelling protocol of traffic incidents on the link of the network is as random events. To be able to compare the results of two different runs of the simulation model with different model parameters, it was necessary to be able to create traffic incidents at exactly the same points in time in both runs. Thus in addition to the random traffic incident occurrences, the model also has the capability to preschedule traffic incidents from the beginning of the run. In the case that the option of random traffic incidents is selected, the characteristics of the incident (starting and ending time, link on which it occurs, type of incident, and location of incident along the link) are written in an output file. If the prescheduled incidents option is selected the random generation of incidents is disabled, and the start and end of the incidents are scheduled in the model calendar from the beginning of the model.

4.4 Conclusions from the Development of the Simulation Model

The development of the traffic simulation model for testing travel time prediction models was presented. The model is a microscopic event oriented simulation model, where individual vehicles are traced throughout their traverse on the network. The model is developed in modular structure, so the various model components can be accessed independently, and additional extensions of the model will be relatively easy. This is especially true for the travel time prediction component of the model, where different models can be tested by simply supplying the appropriate routine that includes the prediction algorithm.

Traffic enters the network through generation nodes, and is directed to move toward a destination node, prespecified for each vehicle. Along their trip, vehicles reevaluate their path at the nodes of the network, based on the most recent information about traffic incidents, delays due to queues or congested links. This way traffic is assigned on the network based on the most recent minimum path toward their destination. By reevaluating the path of each vehicle at each node based on updated travel time data, a continuous dynamic equilibrium network is emulated. Currently all the nodes of the network are regarded as decision nodes where vehicles can divert from the originally chosen path toward their destination. An option that is reserved for future extensions of the model is to have only specific decision nodes (i.e. signalized intersections where the installations of broadcasting devices for the predicted travel times is more likely). This could be helpful, especially in the first steps of implementation of IVHS technology, when the infrastructure will be more scattered.

Each vehicle may be one of the two vehicle categories, "smart" and "non smart", that are currently modeled. Depending on the category that the vehicle belongs to, it has access to a different data base of travel times on the network. If the vehicle is "smart" then it has access to real time information regarding travel conditions on the network. Otherwise the vehicle has access to information built in an initial run of the simulation model, reflecting the experience that such drivers acquire by using the network. "Non smart" vehicles though are allowed to divert to better routes when links are blocked.

Currently drivers are assigned to paths by utilizing an indifference band for shifting to a better path from their original path. The path that a vehicle follows is defined as a sequence of only two links, the one that is assigned to, and the one right after. Instead of the indifference band, the two or three best paths could be found with an M-shortest path algorithm, and a utility function used to randomly assign vehicles to one of these paths.

Although the model is considered to be adequate for this research effort, a number of modelling aspects remain to be further validated and calibrated. Such aspects include the factors for entering and/or exiting the congestion status on a link and the queue dissipation rate.

Chapter 5

Application of the Prediction Model

5.1 Introduction

As was mentioned previously, the purpose of this study is to examine the potential of the recursive prediction model presented in Chapter 3 when applied in a real time route guidance system. The performance of the prediction algorithm will be measured by the error of the predictions for one as well as many steps ahead, for a variety of traffic conditions. The simulation model presented in Chapter 4 will be utilized where a portion of the simulated traffic will be guided toward its destination based on the on-line predictions performed by the travel time prediction routine. The rest of the traffic chooses their routes based on past experience on travel through the network.

Because of the interrelations between traffic patterns in a traffic network and availability of traffic information regarding future travel times, the prediction algorithm will be tested for a variety of percentages of traffic with access to the predicted travel times. In addition, different time intervals between updates of forecasted travel times will be utilized, to examine the effect of the frequency of updating information on travel times and the reliability of predicted travel times.

5.2 Test Network

The network that was utilized for testing the prediction algorithm is shown in Figure 5.1 with the respective node and link labels. Table 5.1 presents the adjacency list for the network. The network consists of three parallel arterials representing an urban corridor consisting of a freeway and two major arterials, connected with several crossing links.

Traffic arrives at the generation nodes 1, 2, and 3 and it is destined to nodes 19, 20, and 21. Therefore, according to the previous notation $O=\{1,2,3\}$ and $D=\{19,20,21\}$. The crossing links provide a number of alternative routes from the origin nodes and the intermediate nodes to the destination nodes. The arrival pattern of traffic at the generation nodes is shown in Figure 5.2. This pattern was based on examples from the 1985 TRB Highway Capacity Manual (TRB SR209). Vehicles arriving at each generation node may have any destination from set D. The percentage of traffic from each generation node towards each destination node, used for all the experiments conducted in this study is given in Table 5.2.

The only difference among the three parallel arterials of the network is the speed limit. The speed limit is 55 m.p.h. for the freeway facility, and 45 m.p.h. and 35 m.p.h. for the two arterials, while the crossing links are assumed to have a speed limit of 40 m.p.h. The speed limit represents the maximum speed with which vehicles can travel on the facilities, and is utilized for calculating

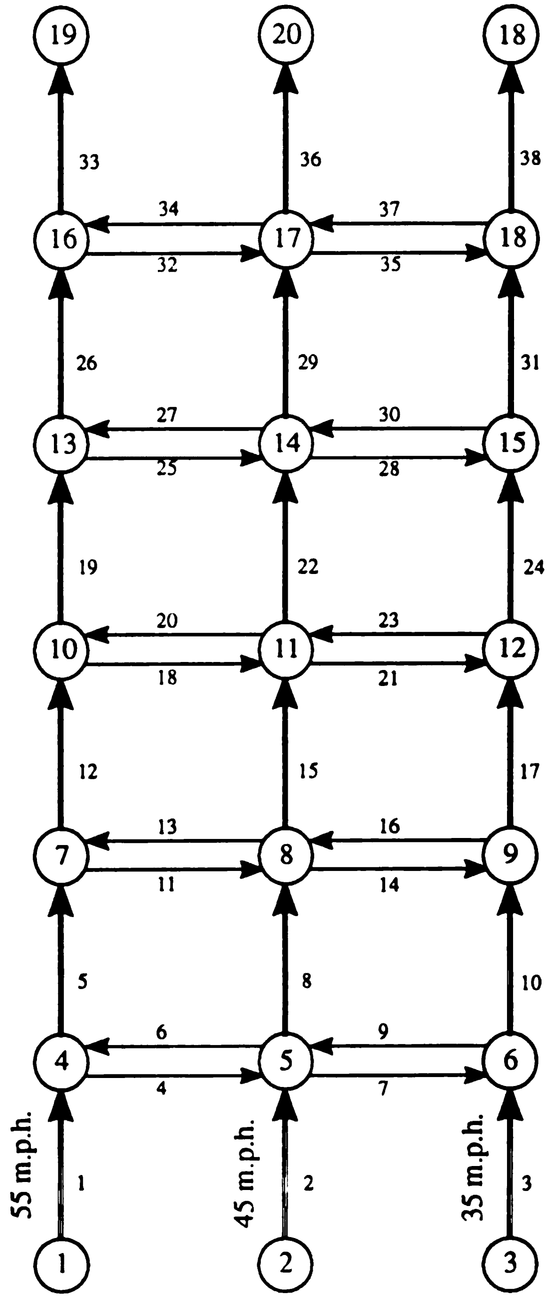


Figure 5.1: Test Network Layout

Table 5.1: Adjacency List of Test Network

From	To Node		
1	4	0	0
2	5	0	0
3	6	0	0
4	5	7	0
5	4	6	8
6	5	9	0
7	8	10	0
8	7	9	11
9	8	12	0
10	11	13	0
11	10	12	14
12	11	15	0
13	14	15	0
14	13	15	17
15	14	18	0
16	17	19	0
17	16	18	20
18	17	21	0
19	0	0	0
20	0	0	0
21	0	0	0

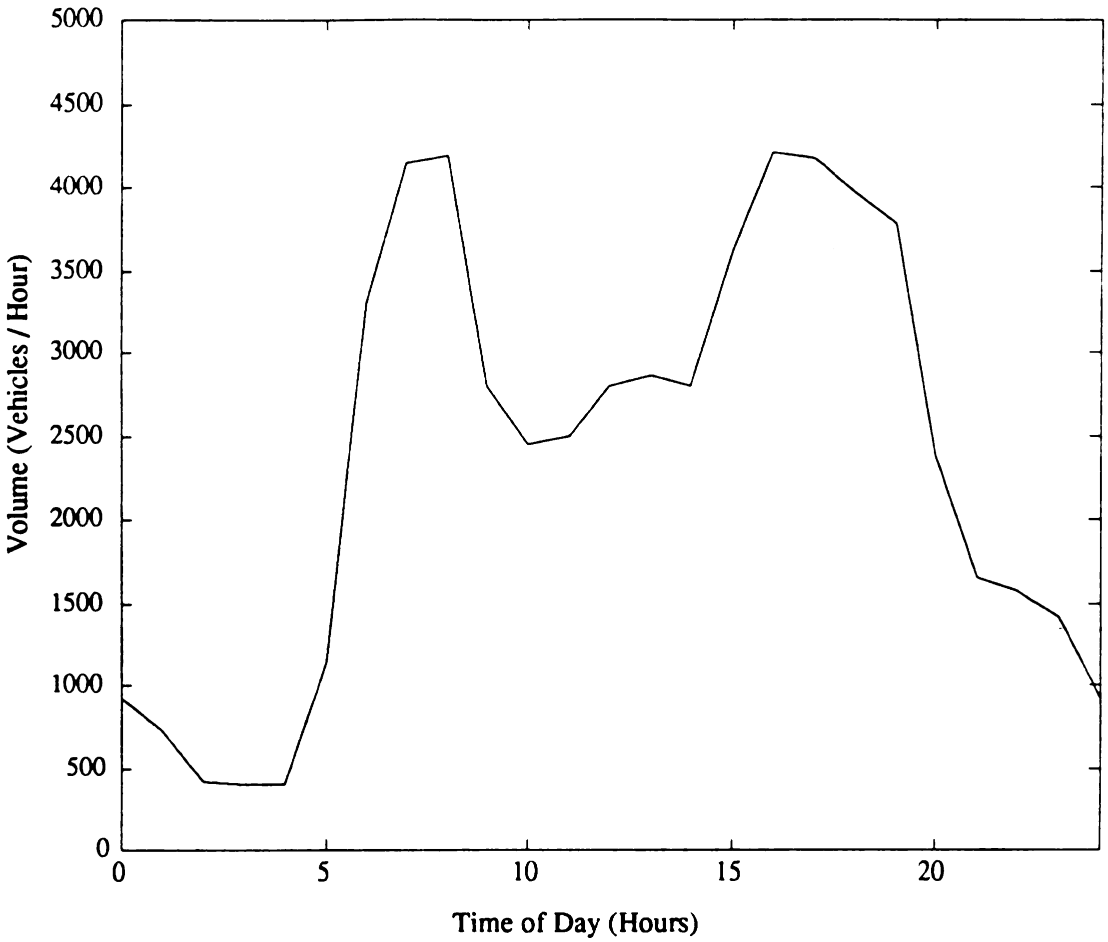


Figure 5.2: Traffic Generation Pattern at Nodes 1, 2, and 3.

Table 5.2: Origin - Destination Matrix

	To 19	To 20	To 21
From 1	50%	40%	10%
From 2	20%	70%	10%
From 3	10%	20%	70%

Table 5.3: Lengths, Free Flow Maximum occupancies and Free Flow travel Times of Links of Network.

Link	Length (miles)	Max. Occup. (vehicles)	Free Flow Travel Time (seconds)		
			Arterial 1 (55 mph)	Arterial 2 (45 mph)	Arterial 3 (35 mph)
1, 2, 3	0.50	66	30	40	50
5, 8, 10	1.50	198	100	120	155
12, 15, 17	2.00	264	130	160	205
19, 22, 24	1.80	244	120	150	185
26, 29, 31	2.00	264	130	160	205
33, 36, 38	0.50	66	30	40	50
Connectors	1.00	66	90		

the free flow travel time. The free flow travel time, along with the length of each link are given in Table 5.3.

All three arterials are two lane facilities, while the crossing links are one lane (for each direction) facilities. Table 5.3 also shows the values of the free flow maximum occupancies of each link utilized in the following experiments. The free flow maximum occupancies were calculated based on the value of 66 vehicles/mile/lane, given in the 1985 HCM for level of service D, which represents the transient phase from free flowing traffic to congested traffic (TRB SR209).Therefore, according to the assumptions made in paragraph 4.3.2, when the occupancy of a link is greater than double the free flow maximum occupancy, thus 132 vehicles/mile/lane, the link is considered congested.

The first step, prior to application of the prediction model to the route guidance system, was to establish the experience data base, for use in routing the non smart vehicles through the network. Because these data are to represent the travel time experience of non equipped drivers, a rough classification of link travel time is required. For this reason, in the initialization run of the simulation model the five-minute mean link travel time was collected, which was then smoothed by averaging each value with the previous and the next five-minute mean value, i.e.:

$$\bar{T}_l(t) = \frac{T_l(t-1) + T_l(t) + T_l(t+1)}{3} \qquad (5.1)$$

where $T_l(t)$ is the observed five-minute mean value of the link l travel time at time t, and $\bar{T}_l(t)$ is the smoothed five-minute mean value for the same time interval. In Figure 5.3 are illustrated the smoothed five-minute mean travel times of links 19 and 22, which are typical for all the links of the three arterials.

5.3 Structure of the Prediction Model

The first step in the application of the model was to find the optimum structure of the prediction model given by equation (3.3). The procedure that was followed was to experiment with different model structures on given sequences of link travel times under two different scenarios of traffic conditions: normal traffic conditions, where there was no congestion induced by a traffic incident, and congested traffic conditions due to a traffic accident where a link is completely blocked. The various model structures differed not only in the number of components (time series) included but also in the order (the number of past values of a time series in the model) of each component.

Because predictions yielded from the RLS algorithm are not bounded, the prediction model was applied in such a way that if the predictions obtained exceeded some reasonable limits then the predictions assumed these limits. Of course the lower limit of the predictions should be the free flow travel time of the links, while the upper limit was set to be fifteen times the free flow travel time of the link.

The measures of performance used were:

(1) the mean absolute relative error $\bar{e}$ which is a measure of the percent expected error, and is defined as:

$$\bar{e} = \frac{1}{N} \cdot \sum_{t} \frac{|T(t) - \hat{T}(t)|}{T(t)} \tag{5.2}$$

(2) the square root of the mean square relative error $\bar{e}_s$ which gives more

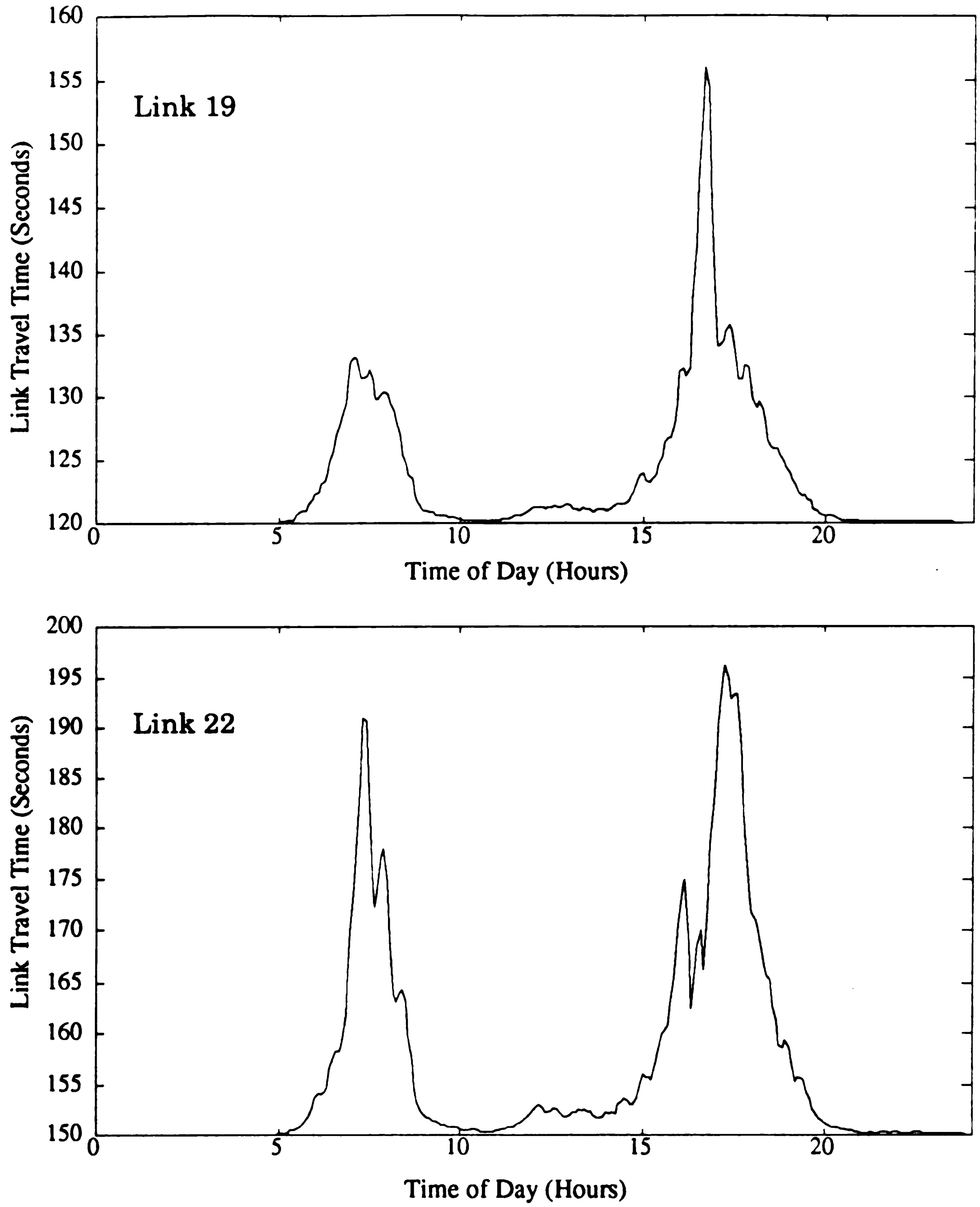


Figure 5.3: Smoothed Five-Minute Average Travel Times of Links 19 and 22

weight to the larger errors, and is defined as:

$$\bar{e}_s = \frac{1}{N} \cdot \sqrt{\sum_t \left[\frac{T(t) - \hat{T}(t)}{T(t)} \right]^2} \tag{5.3}$$

(3) the maximum absolute error e_m defined as:

$$e_m = \max_t \left\{ \frac{|T(t) - \hat{T}(t)|}{T(t)} \right\} \tag{5.4}$$

where $\hat{T}(t)$ is the predicted travel time and $T(t)$ is the observed travel time for the time interval t, and N is the number of observations in the time period that the performance of the prediction model is examined. These error measures were compared with the corresponding error measures for the case where no predictions were performed, and the current travel times were used to obtain the minimum paths.

For this phase of the study the traffic operations on the network were simulated with 30% smart vehicles and the time interval between information updates regarding traffic conditions was set at 60 seconds. The time series consisted of 2880 observations (two days), with the first day free of any traffic incidents. The occurrence of a traffic accident was simulated on link 19 that started at time 08:34:10.2 and was cleared at time 08:50:21.5 in the second day of the simulation run.

Shortly after the accident occurrence, the link is blocked by traffic that had entered the link before the incident, and by traffic that does not know about the traffic conditions of the link (i.e. non smart traffic). Meanwhile, smart vehicles are diverted to link 18. After link 19 is completely blocked, all traffic is diverted to link 18. Because of this diverted traffic link 18 becomes congested and after a few minutes a queue is formed on link 12.

During the accident, the simulation model assumes that the observed travel time of the link with the accident is equal to the expected duration of the accident plus the travel time of the link based on the current length of the queue and the current volume of the link. The expected duration of an accident on link 19 was set at 1000 seconds. After the accident ended, the queue from link 12 feeds link 19 with a steady rate, one vehicle from each lane every 2.8 seconds. The queue formed on link 19 starts dissipating after the end of the accident at the same rate of 2.8 seconds per vehicle per lane. Because, the rate with which the queue on link 19 dissipates is equal to the rate of incoming traffic, the travel time on link 19 is almost constant until the demand is decreased around 10:00 a.m. At that time the queue on link 12 is gradually dissolved completely, and then, the queue on link 19 is also dissolved, and the travel time on the link returns to normal levels.

5.3.1 Autoregressive Models

First the performance of a number of prediction models with just the autoregressive component in equation (3.3) were examined:

$$T_l(t) = \sum_{i=1}^{n} a_i(t-1) \cdot T_l(t-i) + \varepsilon(t) \qquad (5.5)$$

so the relation of future travel times of a link with the past history of the travel time of the same link were studied. The error measures were computed for the one, five and ten step ahead predictions, which represent prediction of the link travel time for one, five and ten minutes ahead. Models with different autoregressive order (value of n) were tested with the order ranging from one to fifty. Initially the forgetting factor λ was set equal to 1.00.

5.3.1.1 Normal Traffic Conditions

For the situation of normal traffic operations (no traffic incident occurrence) the travel time series for link 15 is illustrated. The error measures given by equations (5.2), (5.3), and (5.4) were computed for two time periods of the second simulated day, the morning period from 6:00 to 11:00 and the afternoon period from 14:00 to 19:00. These two time periods were used because for the rest of the time, link travel time, for almost all links, remains very close to the free flow travel time of the link.

The mean relative error, the mean square error and the maximum error produced by a series of autoregressive models for the morning and evening periods are given in Table 5.4 and Table 5.5 respectively. In almost all cases the resulting prediction errors from the models were smaller than the ones resulting if no predictions were performed. Only in the evening period and for the models AR(3) and AR(5) the mean relative error and the mean square error of the five and ten steps ahead are larger than those produced by the no predictions case. As can be seen from the results of these experiments, the prediction error is getting smaller for higher order models (order of 10 and higher) for both the morning and evening time periods.

However, the large number of variables n used in the prediction model has the adverse effect that the resulting parameter vector of the model is probably overfitted to the data. The model, although it is large enough to include the system describing the link travel time, is overparameterized and individual

Table 5.4: Prediction Errors of Link 15 for the Morning Period under Normal Traffic Conditions (300 Observation, 06:00 - 11:00)

	AR order	1 Step Ahead Predictions			5 Step Ahead Predictions			10 Step Ahead Predictions		
		$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m
$\lambda = 1.000$	1	0.00442	0.00040	0.03467	0.01224	0.00129	0.11778	0.01904	0.00203	0.17910
	2	0.00412	0.00037	0.02857	0.01218	0.00124	0.10998	0.01889	0.00201	0.18351
	3	0.00417	0.00037	0.02538	0.01245	0.00123	0.10460	0.01896	0.00202	0.18626
	4	0.00416	0.00037	0.02882	0.01228	0.00129	0.11968	0.01896	0.00201	0.17818
	5	0.00412	0.00036	0.02491	0.01234	0.00124	0.10715	0.01892	0.00203	0.18686
	10	0.00407	0.00036	0.02370	0.01218	0.00118	0.09737	0.01884	0.00192	0.16951
	15	0.00402	0.00035	0.02306	0.01228	0.00116	0.08685	0.01937	0.00186	0.14572
	25	0.00400	0.00035	0.02444	0.01239	0.00116	0.09073	0.01993	0.00188	0.14475
	50	0.00408	0.00036	0.02379	0.01236	0.00119	0.10020	0.01905	0.00193	0.16432

No Predictions	0.00442	0.00040	0.03462	0.01238	0.00129	0.11725	0.01914	0.00208	0.18957

Table 5.5: Prediction Errors of Link 15 for the Evening Period under Normal Traffic Conditions (300 Observation, 14:00-19:00)

	AR	1 Step Ahead Predictions			5 Step Ahead Predictions			10 Step Ahead Predictions		
	order	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m
$\lambda = 1.000$	1	0.00567	0.00045	0.05043	0.01523	0.00131	0.11313	0.02060	0.00186	0.16455
	2	0.00545	0.00044	0.04377	0.01553	0.00130	0.10753	0.02096	0.00188	0.16699
	3	0.00537	0.00043	0.04397	0.01582	0.00131	0.10623	0.02136	0.00190	0.17142
	4	0.00531	0.00043	0.04331	0.01514	0.00129	0.10712	0.02039	0.00183	0.16323
	5	0.00535	0.00043	0.04396	0.01579	0.00131	0.10753	0.02135	0.00190	0.16975
	10	0.00526	0.00042	0.04472	0.01549	0.00125	0.09883	0.01990	0.00176	0.15280
	15	0.00533	0.00042	0.04514	0.01550	0.00123	0.09108	0.02029	0.00171	0.13827
	25	0.00528	0.00042	0.04808	0.01567	0.00123	0.08782	0.02089	0.00171	0.13392
	50	0.00530	0.00041	0.04380	0.01550	0.00121	0.08723	0.01949	0.00165	0.13815

No Predictions	0.00567	0.00045	0.05035	0.01553	0.00131	0.11269	0.02108	0.00189	0.17345

parameters are not affecting the predictions. The effect of old observations that most likely do not correlate as well as more recent ones with the current and the near future travel times is carried for a long time (i.e. for as long as these observations are in the model), and if sudden changes in the travel time occur, the result would be a dramatic increase in the error. For the same reason, the filter is not kept sufficiently alert to follow the most recent changes in the traffic conditions on the link. The low correlation of the older observations to the most recent ones can be seen by studying the parameters of the autoregressive models as n increases. From Table 5.6 we can see that observations older than five time steps do not contribute significantly to the predictions (as compared to the contribution of observations from one to four time steps old).

From the same results, we can observe that as expected, one step ahead predictions have less error than five step ahead predictions which are better than the 10 step predictions. The one step and ten step ahead predictions with an autoregressive model of order three, for the morning period are shown in Fig-

Table 5.6: Parameters for Autoregressive Models at the End of the Evening Period (19:00) of the Second Simulated Day ($\lambda = 1.00$).

AR	a_1	a_2	a_3	a_4	a_5	a_6
1	1.000	--	--	--	--	--
2	1.354	-0.355	--	--	--	--
3	1.340	-0.301	-0.041	--	--	--
4	1.347	-0.249	-0.267	0.169	--	--
5	1.346	-0.249	-0.266	0.165	0.003	--
6	1.347	-0.249	-0.265	0.166	-0.004	0.004

ure 5.4 and Figure 5.5 respectively. Figure 5.6 and Figure 5.7 show the predictions obtained with an autoregressive model of order five for the evening period. In both cases it can be seen that while the one step ahead predictions are very close to the observed data, the ten step ahead predictions almost replicate a shift of the observations ten time steps ahead. This can also be deduced by comparing the error measures of these models with the no prediction case. In Table 5.7 and Table 5.8 the percent difference between the error of each model and the no predictions case is given. Values in parentheses indicate the relative ranking of each error measure for the given prediction, while in the column under the heading 'Total Rank' the sum of the ranks along with the relative rank of these figures is given. From these results we can see that the AR(1) model in both the morning and the evening periods score the lowest, while the AR(2) and AR(4) models are the best. Nevertheless, improvements over the no predictions case are too small to be considered significant.

In the following, the forgetting factor was set to values smaller than 1.00, so the parameters of the models will be allowed to vary with time, and less weight will be given to prediction errors encountered in the distant past. For example, in the case where $\lambda = 0.975$, the weight attributed to the error of the prediction made one time step before is 0.975 while the weight attributed to the error of the prediction 20 time steps before is only 0.402. In Figure 5.8 and Figure 5.9 the parameters of the model AR(3) for $\lambda=1.000$ and $\lambda=0.975$ are shown. As can be seen from these figures, in the case of $\lambda=1.00$ the parameters are converging to constant values, while in the case of $\lambda=0.975$ the parameters of the model are allowed to vary with time.

The prediction errors for forgetting factor equal to 0.990, 0.975 and 0.90 are

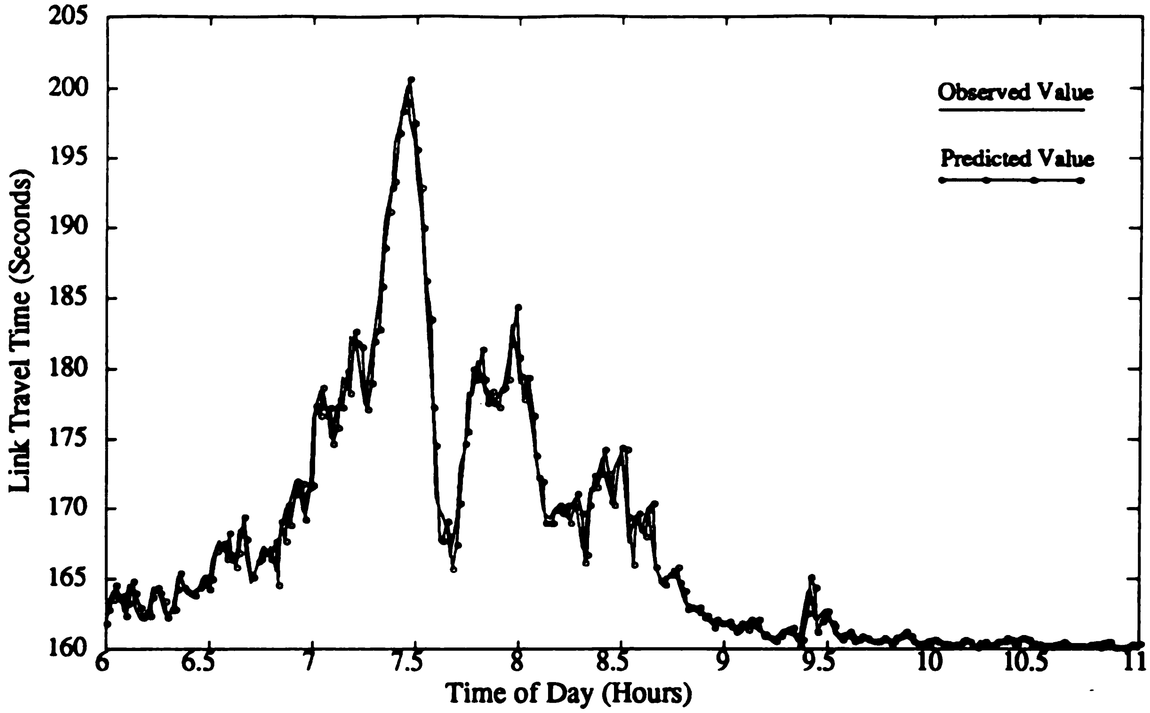


Figure 5.4: 1 Step Predictions of Travel Times of Link 15 with Autoregressive Model of Order 3 - Normal Traffic Operations, Morning Period

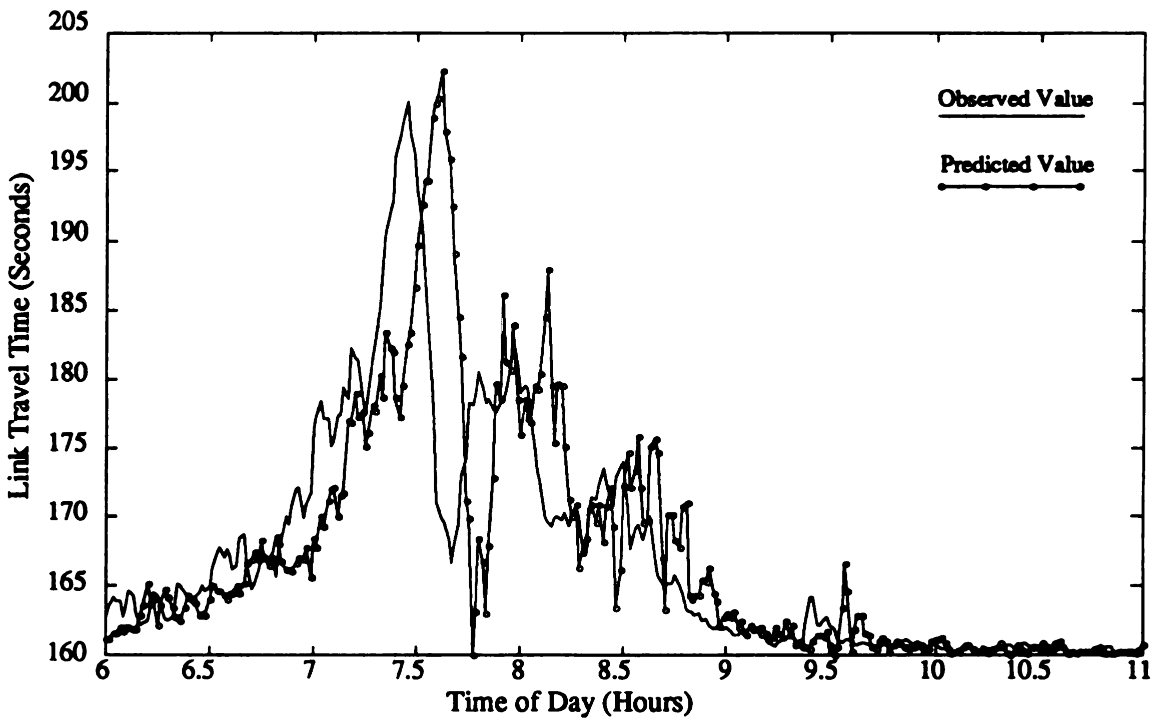


Figure 5.5: 10 Step Predictions of Travel Times of Link 15 with Autoregressive Model of Order 3 - Normal Traffic Operations, Morning Period

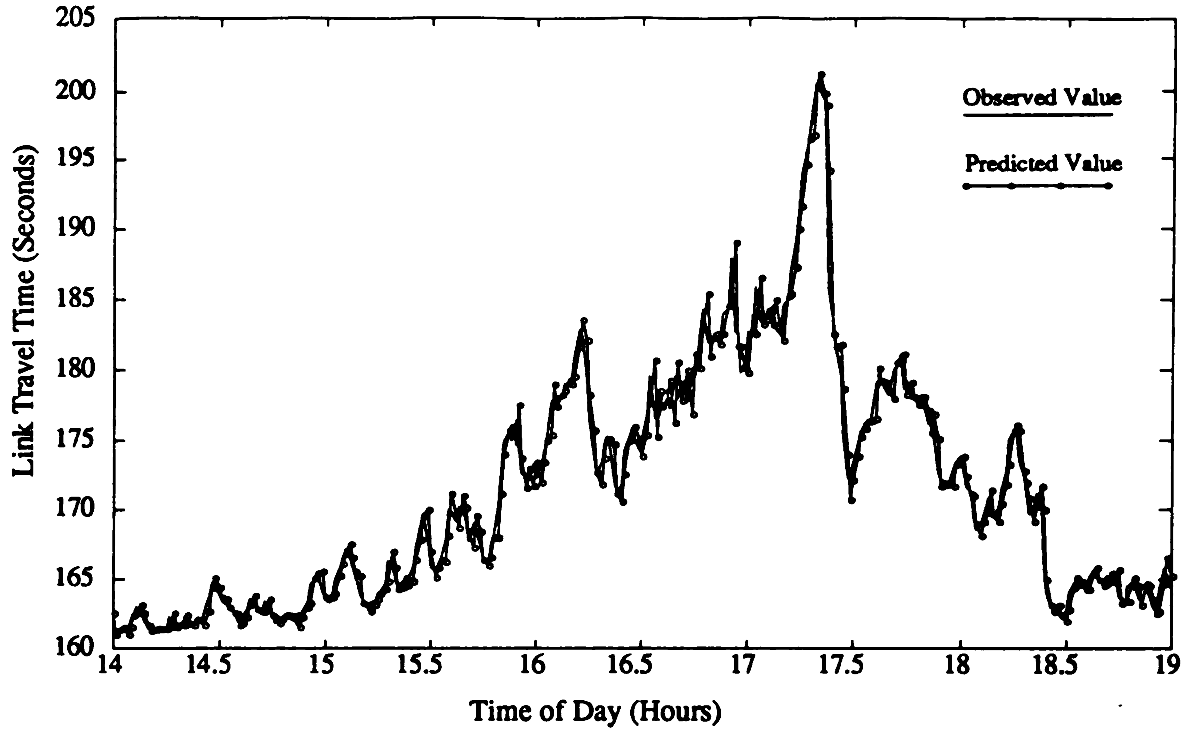


Figure 5.6: 1 Step Predictions of Travel Times of Link 15 with Autoregressive Model of Order 5 - Normal Traffic Operations, Evening Period

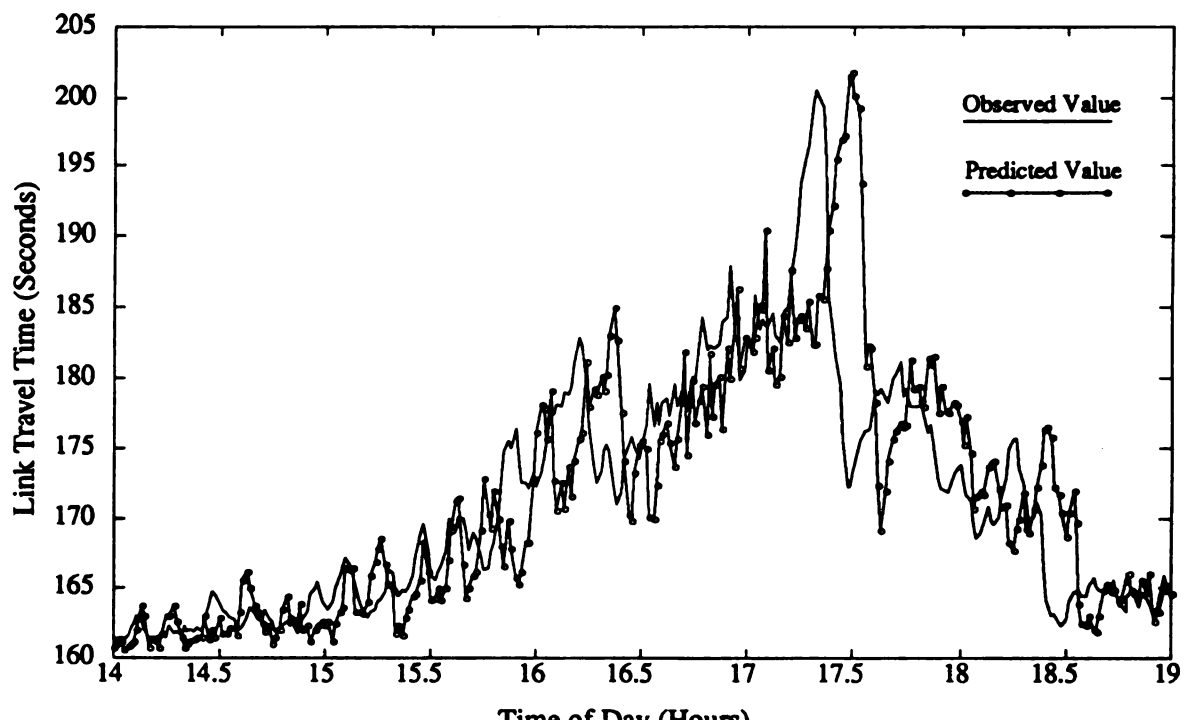


Figure 5.7: 10 Step Predictions of Travel Times of Link 15 with Autoregressive Model of Order 5 - Normal Traffic Operations, Evening Period

Table 5.7: Percent Difference of Prediction Error of Autoregressive Models with No Predictions case for Morning Period (06:00-11:00)

AR	1 Step Ahead[1] Predictions			5 Step Ahead Predictions			10 Step Ahead Predictions			Total Rank[2]
	$\Delta\bar{e}$	$\Delta\bar{e}_s$	Δe_m	$\Delta\bar{e}$	$\Delta\bar{e}_s$	Δe_m	$\Delta\bar{e}$	$\Delta\bar{e}_s$	Δe_m	
1	0.0 (5)	0.0 (5)	0.1 (5)	-1.1 (2)	0.0 (4)	+0.5 (4)	-0.5 (5)	-2.4 (4)	-5.5 (2)	36 (5)
2	-6.8 (1)	-7.5 (2)	-17.5 (3)	-1.6 (1)	-3.9 (2)	-6.2 (3)	-1.3 (2)	-3.4 (1)	-3.2 (3)	18 (1)
3	-5.7 (4)	-7.5 (2)	-26.7 (2)	+0.6 (5)	-4.7 (1)	-10.8 (1)	-0.9 (3)	-2.9 (3)	-1.7 (4)	25 (3)
4	-5.9 (3)	-7.2 (4)	-16.7 (4)	-0.8 (3)	+0.4 (5)	2.1 (5)	-0.9 (3)	-3.3 (2)	-6.0 (1)	30 (4)
5	-6.8 (1)	-10.0 (1)	-28.1 (1)	-0.3 (4)	-3.9 (2)	-8.6 (2)	-1.1 (1)	-2.4 (4)	-1.4 (5)	21 (2)

1. Numbers in parentheses denote relative rank.
2. Cell numbers indicate the sum of ranks, and numbers in parentheses indicate relative rank of the sums.

shown in Table 5.9 and Table 5.10 for the morning and evening periods respectively. As can be seen from these tables, the error measures get worse as the forgetting factor is set to smaller values. The smaller the value of the forgetting factor the better the tracking capability of the model which is trying to find not only the optimum values of the parameters, but its optimum values at each point in time (see Figure 5.8 and Figure 5.9). This is achieved though, at the expense of the sensitivity of the model to the prediction errors. Obviously and as it was discussed in paragraph 3.4.2 there is a trade off which has to be made between the tracking ability of the model and its sensitivity to the prediction errors. This is especially evident for the ten step ahead predictions where the mean absolute relative error is almost double the corresponding

Table 5.8: Percent Difference of Prediction Error of Autoregressive Models with No Predictions case for Evening Period (14:00-19:00)

AR	1 Step Ahead Predictions[1]			5 Step Ahead Predictions			10 Step Ahead Predictions			Total Rank[2]
	$\Delta\bar{e}$	$\Delta\bar{e}_s$	Δe_m	$\Delta\bar{e}$	$\Delta\bar{e}_s$	Δe_m	$\Delta\bar{e}$	$\Delta\bar{e}_s$	Δe_m	
1	0.0 (5)	0.0 (5)	+0.2 (5)	-1.9 (2)	0.0 (3)	0.4 (5)	-2.3 (2)	-1.6 (2)	-5.1 (2)	31 (4)
2	-3.9 (4)	-2.2 (4)	-13.1 (2)	0.0 (3)	-0.8 (2)	-4.6 (3)	-0.6 (3)	-0.5 (3)	-3.7 (3)	27 (2)
3	-5.3 (3)	-4.4 (1)	-12.7 (3)	+1.9 (5)	0.0 (3)	-5.7 (1)	+1.3 (4)	0.5 (4)	-1.2 (5)	31 (4)
4	-6.4 (1)	-4.2 (3)	-14.0 (1)	-2.5 (1)	-1.1 (1)	-4.9 (2)	-3.3 (1)	-2.7 (1)	-5.9 (1)	12 (1)
5	-5.6 (2)	-4.4 (1)	-12.7 (3)	+1.7 (4)	0.0 (3)	-4.6 (3)	+1.3 (4)	+0.5 (4)	-2.1 (4)	28 (3)

1. Numbers in parentheses denote relative rank.
2. Cell numbers indicate the sum of ranks, and numbers in parentheses indicate relative rank of the sums.

error of the no prediction situation (i.e. for the morning period, 0.02288 and 0.01114 respectively). In the case of λ=0.990 and λ=0.975 the one and five step predictions have not deteriorated too much, and the AR(3) model give the best results, which still are only marginally better than the no predictions case (i.e. for the morning period with λ=0.975, 0.01227 and 0.01238 respectively).

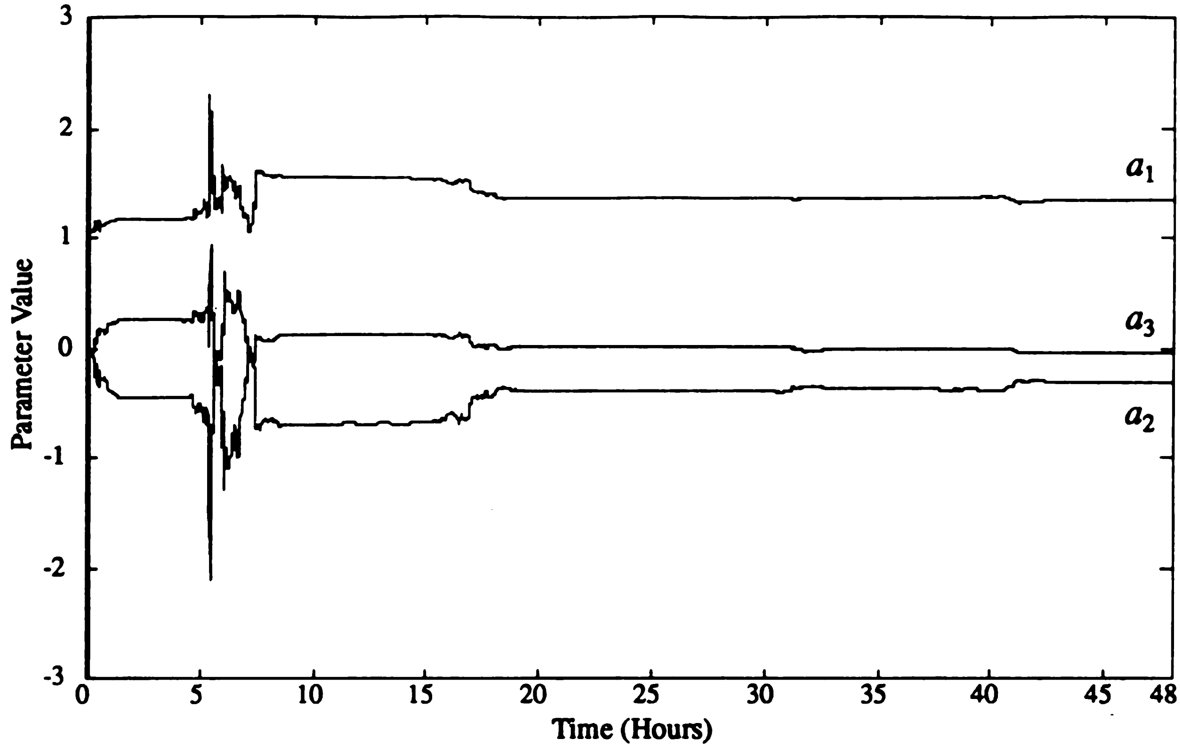


Figure 5.8: Parameters of Autoregressive Model of Order 3 - Forgetting Factor λ=1.000

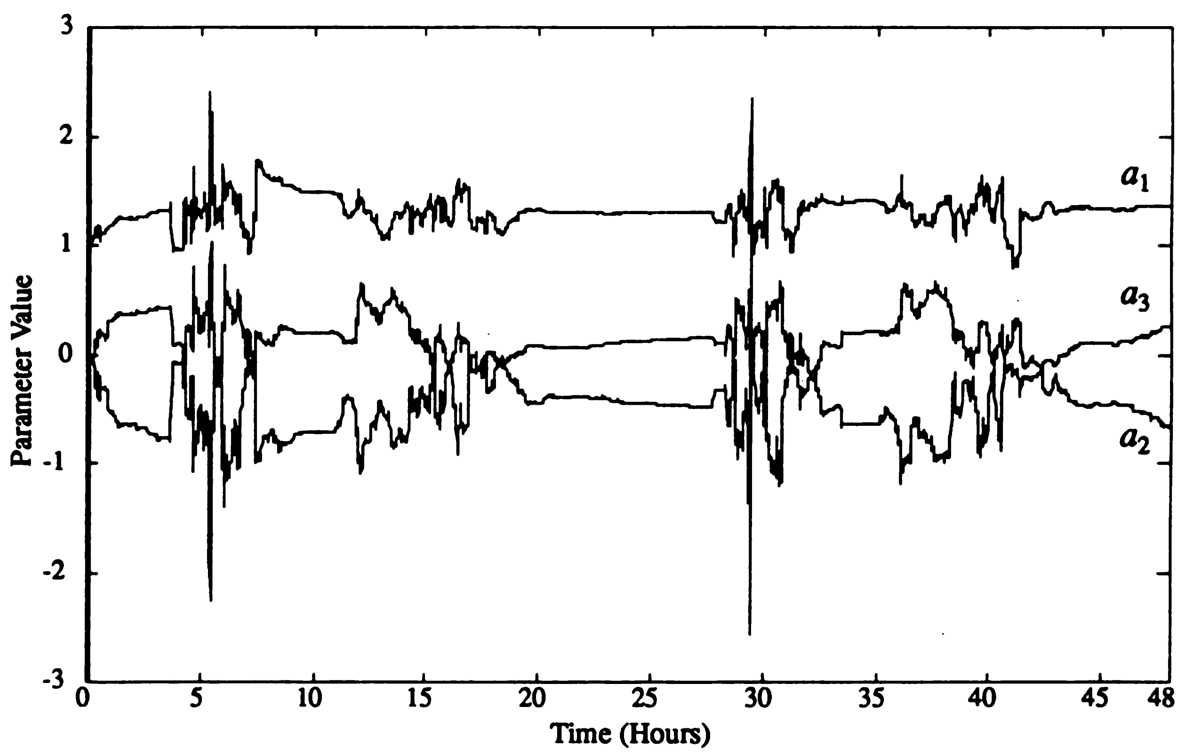


Figure 5.9: Parameters of Autoregressive Model of Order 3 - Forgetting Factor λ=0.975

Table 5.9: Prediction Errors of Link 12 for Morning Period under Normal Traffic Conditions and Different Values of the Forgetting Factor (300 Observations, 06:00-11:00)

	AR	1 Step Ahead Predictions			5 Step Ahead Predictions			10 Step Ahead Predictions		
	order	e	$\bar{e}_s$	e_m	e	$\bar{e}_s$	e_m	e	$\bar{e}_s$	e_m
$\lambda = 0.990$	1	0.00442	0.00040	0.03501	0.01243	0.00132	0.12475	0.01914	0.00214	0.20201
	2	0.00419	0.00037	0.02619	0.01256	0.00127	0.11581	0.01921	0.00214	0.20454
	3	0.00419	0.00037	0.02656	0.01229	0.00125	0.10474	0.01984	0.00219	0.20740
	5	0.00424	0.00037	0.02084	0.01259	0.00131	0.11771	0.02025	0.00225	0.21114
$\lambda = 0.975$	1	0.00441	0.00040	0.03439	0.01273	0.00135	0.11284	0.01956	0.00226	0.22298
	2	0.00420	0.00037	0.02534	0.01283	0.00129	0.11620	0.01985	0.00226	0.22465
	3	0.00419	0.00037	0.02636	0.01227	0.00125	0.09861	0.02049	0.00231	0.22675
	5	0.00432	0.00037	0.02054	0.01268	0.00133	0.12046	0.02106	0.00237	0.22908
$\lambda = 0.900$	1	0.00443	0.00040	0.02822	0.01399	0.00145	0.11861	0.02288	0.00261	0.26360
	2	0.00438	0.00039	0.02552	0.01409	0.00137	0.10400	0.02402	0.00266	0.26400
	3	0.00435	0.00039	0.03096	0.01302	0.00130	0.11034	0.02363	0.00263	0.26298
	5	0.00479	0.00041	0.02926	0.01134	0.00136	0.10478	0.02442	0.00270	0.26529
No Predictions		0.00442	0.00040	0.03462	0.01238	0.00129	0.11725	0.01114	0.00208	0.18957

Table 5.10: Prediction Errors of Link 12 for Evening Period (300 Observations, 14:00-19:00) under Normal Traffic Conditions and Different Values of the Forgetting Factor

	AR	1 Step Ahead Predictions			5 Step Ahead Predictions			10 Step Ahead Predictions		
	order	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m
$\lambda = 0.990$	1	0.00567	0.00046	0.05160	0.01153	0.00134	0.11932	0.02090	0.00195	0.18200
	2	0.00543	0.00044	0.05023	0.01567	0.00135	0.11767	0.02136	0.00196	0.18212
	3	0.00544	0.00044	0.04900	0.01536	0.00132	0.11440	0.02117	0.00197	0.18838
	5	0.00537	0.00044	0.04806	0.01571	0.00141	0.12418	0.02086	0.00191	0.17703
$\lambda = 0.975$	1	0.00571	0.00046	0.05219	0.01575	0.00138	0.11228	0.02197	0.00205	0.19593
	2	0.00549	0.00044	0.05136	0.01623	0.00138	0.12302	0.02255	0.00207	0.19587
	3	0.00555	0.00045	0.04913	0.01592	0.00137	0.11867	0.02279	0.00212	0.20351
	5	0.00542	0.00450	0.04800	0.01606	0.00145	0.12580	0.02193	0.00203	0.19274
$\lambda = 0.900$	1	0.00585	0.00046	0.05210	0.01784	0.00153	0.13303	0.02720	0.00253	0.23730
	2	0.00582	0.00048	0.05667	0.01887	0.00158	0.13038	0.02900	0.00261	0.23722
	3	0.00605	0.00051	0.06374	0.01847	0.00154	0.13336	0.02906	0.00264	0.24769
	5	0.00613	0.00051	0.06141	0.01896	0.00167	0.16923	0.03064	0.00377	0.83479

No Predictions	0.00567	0.00045	0.05035	0.01553	0.00131	0.11269	0.02108	0.00189	0.17345

5.3.1.2 Congested Conditions Due to Traffic Accident

For the situation of congested traffic operations due to a traffic accident the series of travel time observations on link 19 were considered. A traffic accident occurred on link 19 at time 08:34:10.2 and ended at time 08:50:21.5. Autoregressive models of order ranging from one to five were tested and the forgetting factor was set to four different values: 1.00, 0.99, 0.90 and 0.80. The error measures obtained by these models for a time period that includes the incident and its aftermath (starting at 8:00 and ending at 10:30) are shown in Table 5.11.

The performance of the autoregressive model is depicted in Figure 5.10 and Figure 5.11 where the one step and the ten step ahead predictions respectively obtained with an autoregressive model of order one and with λ=1.00 are shown. As can be seen in these figures, like in the normal conditions situation, predictions are almost only a shift of the observations. In Figure 5.12 and Figure 5.13 the same predictions with an autoregressive model of order one and with λ=0.80 are shown. Due to the smaller forgetting factor in the later set of figures, the model is in a more alert condition, and thus response to the change of the travel time is more extreme. Obviously, the one step predictions are worse, at least in the beginning of the accident, but the ten step predictions are better than in the no predictions case. In the case where λ=0.80, the model reacted to the large error at the beginning of the incident, and predictions reached the ceiling of the allowable range of the travel time predictions for one time step. After the end of the incident the model also responded to the abrupt drop of the travel time of the link, and especially close to the end of the aftermath of the incident 10 step predictions return to observation values more rapidly than in the case where λ=1.00.

Table 5.11: Prediction Errors of Link 19 with Autoregressive Models and Different Values of the Forgetting Factor under Congested Conditions due to a Traffic Accident (50 Observations 08:00 -10:30)

	AR	1 Step Ahead Predictions			5 Step Ahead Predictions			10 Step Ahead Predictions		
	order	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m
$\lambda = 1.000$	1	0.02408	0.00709	0.89068	0.12169	0.02149	0.91370	0.26335	0.04102	1.84358
	2	0.02903	0.00877	0.89097	0.12555	0.02184	0.91038	0.26483	0.04095	1.83488
	3	0.03290	0.00969	0.89092	0.13045	0.02242	0.90667	0.26911	0.04109	1.82417
	4	0.03609	0.01038	0.89323	0.13576	0.02317	0.90154	0.27118	0.04101	1.81037
	5	0.04220	0.01207	0.89323	0.13599	0.02310	0.89739	0.28301	0.04207	1.79822
$\lambda = 0.990$	1	0.02649	0.00707	0.89072	0.13132	0.02107	0.89084	0.27748	0.03920	1.68092
	2	0.03207	0.00922	0.89325	0.13644	0.02179	0.89323	0.28024	0.03924	1.66462
	3	0.03768	0.01096	0.89325	0.14129	0.02244	0.89323	0.28245	0.03925	1.64559
	4	0.04135	0.01166	0.89325	0.13985	0.02175	0.89080	0.288014	0.03970	1.62335
	5	0.04705	0.01308	0.89325	0.14454	0.02240	0.89323	0.29007	0.03977	1.61182
No Predictions		0.02404	0.00715	0.89068	0.12450	0.02228	0.95956	0.27445	0.04363	1.98757

Table 5.11 (continued)

	AR	1 Step Ahead Predictions			5 Step Ahead Predictions			10 Step Ahead Predictions		
	order	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m
$\lambda = 0.900$	1	0.02645	0.00797	0.89101	0.11735	0.02007	0.89232	0.24879	0.03774	1.66933
	2	0.03246	0.01005	0.89325	0.11243	0.01936	0.89323	0.22315	0.03421	1.66935
	3	0.03975	0.01185	0.89325	0.11696	0.02021	0.89323	0.22430	0.03441	1.66948
	4	0.04682	0.01278	0.89325	0.12005	0.02030	0.89194	0.24241	0.03694	1.66972
	5	0.05426	0.01436	0.89325	0.17266	0.04173	4.96054	0.32284	0.05784	5.60336
$\lambda = 0.800$	1	0.02425	0.00785	0.89115	0.10276	0.01883	0.89300	0.22369	0.03517	1.67477
	2	0.03133	0.01005	0.89116	0.10347	0.01900	1.06604	0.20067	0.03132	1.67479
	3	0.04015	0.01198	0.89325	0.11271	0.02014	0.91140	0.22517	0.04142	4.09473
	4	0.04886	0.01312	0.89325	0.14357	0.03856	4.65295	0.32047	0.10162	13.84573
	5	0.05723	0.01507	0.89325	0.25979	0.08667	9.17902	0.38524	0.11779	13.80208
No Predictions		0.02404	0.00715	0.89068	0.12450	0.02228	0.95956	0.27445	0.04363	1.98757

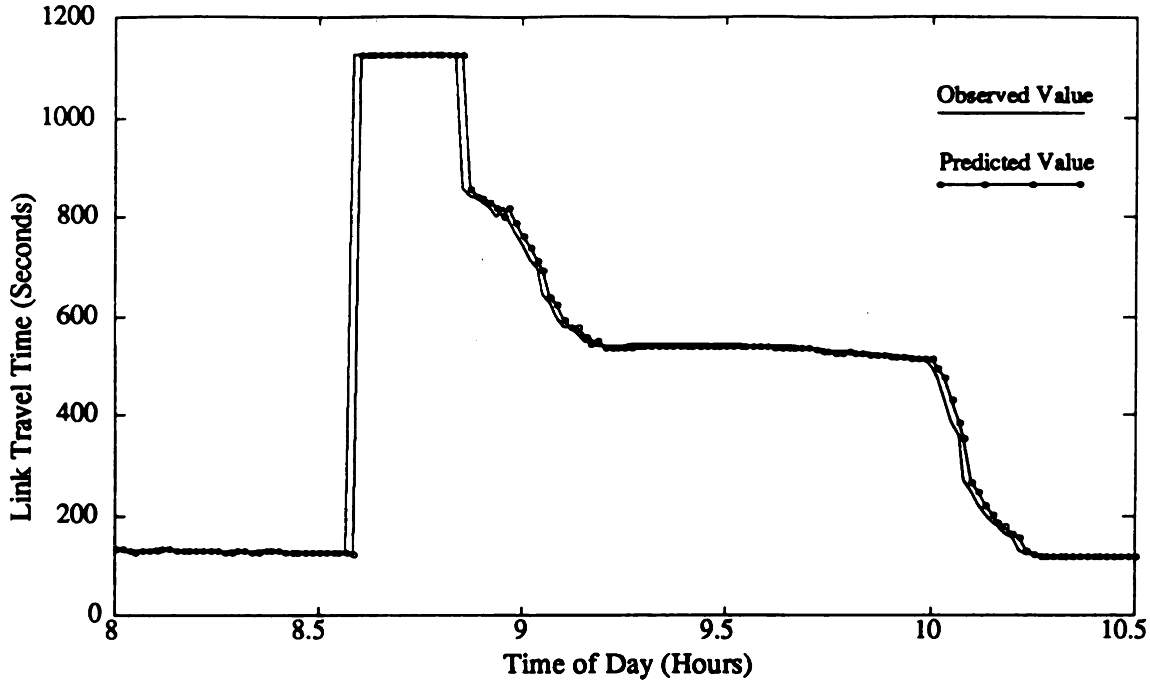


Figure 5.10: 1 Step Ahead Predictions of Travel Times of Link 19 with Autoregressive Model of Order 1 and λ=1.00 - Congested Traffic Conditions Due to Accident

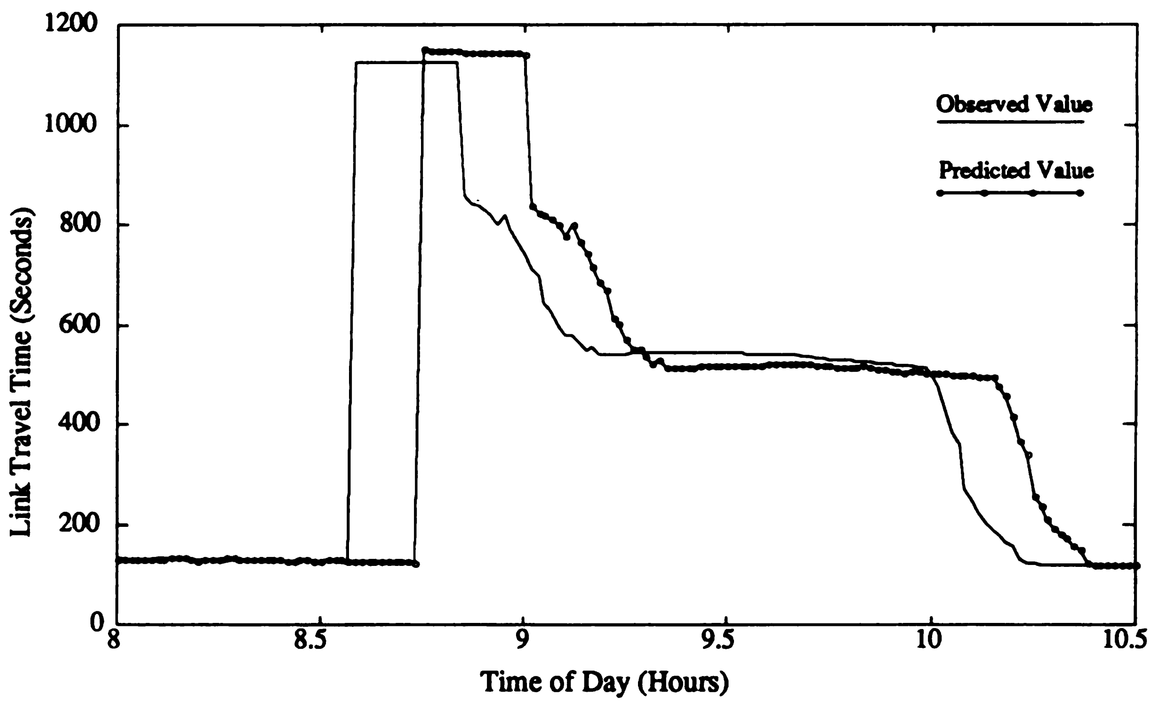


Figure 5.11: 10 Step Ahead Predictions of Travel Times of Link 19 with Autoregressive Model of Order 1 and λ=1.00 - Congested Traffic Conditions Due to Accident

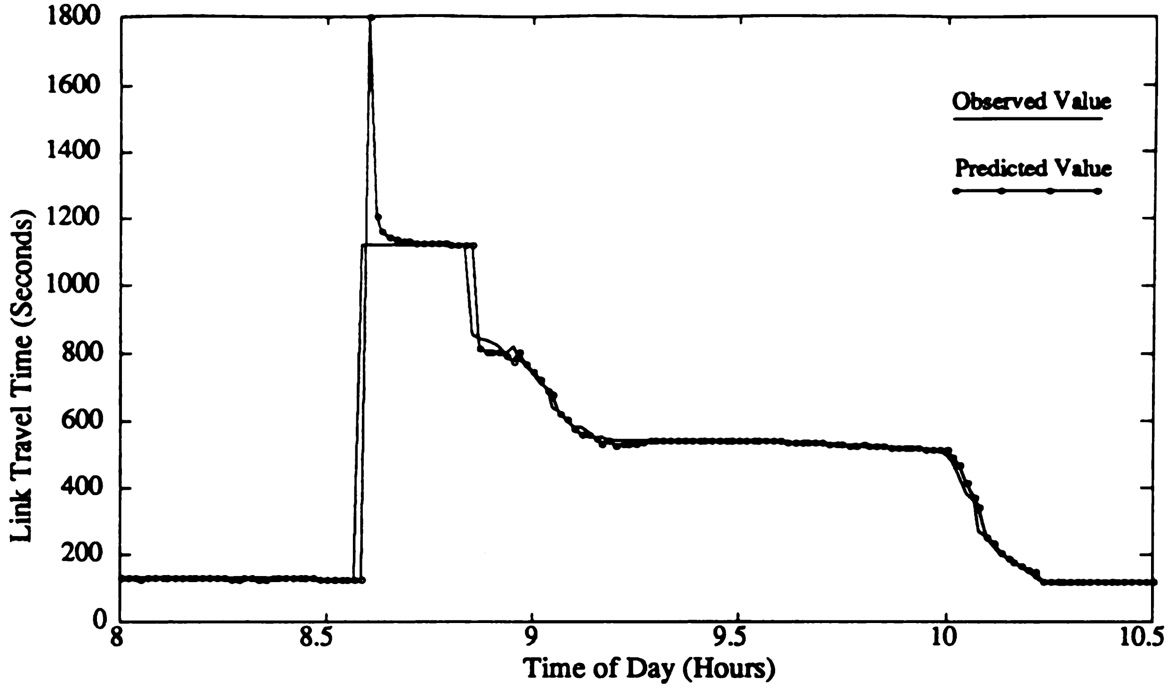


Figure 5.12: 1 Step Ahead Predictions of Travel Times of Link 19 with Autoregressive Model of Order 1 and λ=0.80 - Congested Traffic Conditions Due to Accident

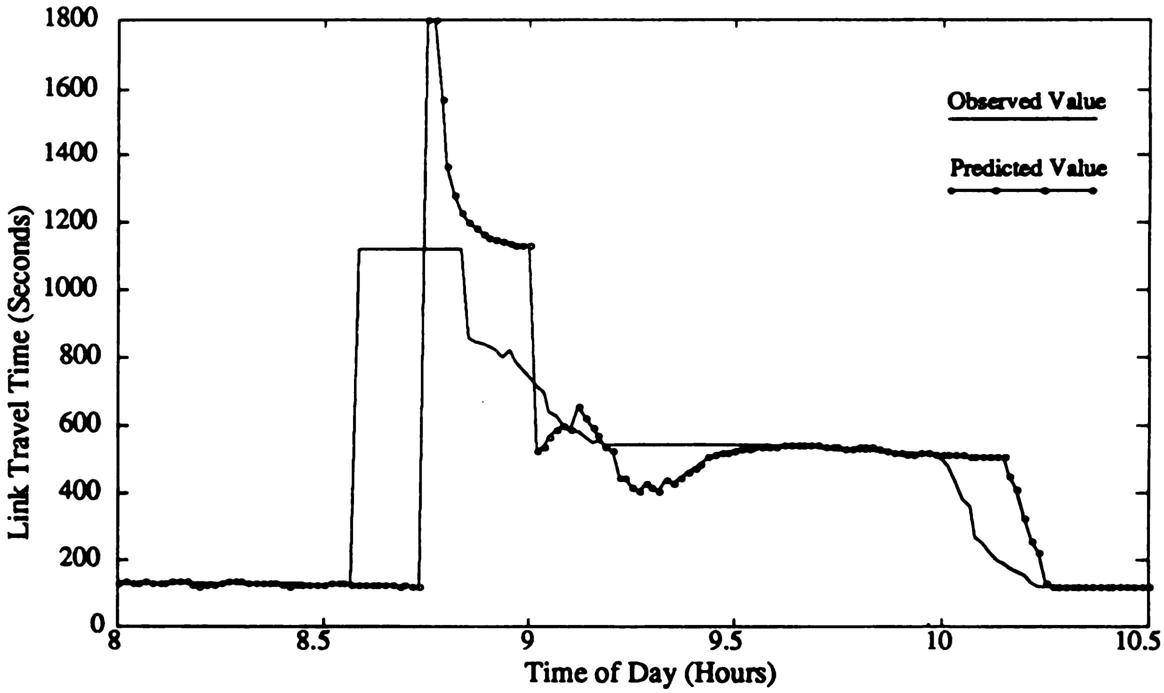


Figure 5.13: 10 Step Ahead Predictions of Travel Times of Link 19 with Autoregressive Model of Order 1 and λ=0.80 - Congested Traffic Conditions Due to Accident

This is shown more clearly in Table 5.11 where the error measures for these traffic conditions and for different values of the forgetting factor are shown. For larger order models and for small values of the forgetting factor the model reacts too spasmodically as it can be deduced by the value of the maximum error, which reaches a high of 13.85 for $\lambda=0.80$ and $n=4$. Generally, for all the values of the forgetting factor, models with smaller autoregressive order n (i.e. $n=1$ or 2) gave better results than models of larger order, and this was especially apparent when fewer past prediction errors of the algorithm were considered in the computation of the parameters of the model with the forgetting factor set at 0.90 and 0.80. For example, for $n=1$ the mean square relative error of the five step ahead predictions is 0.02149 for $\lambda=1.00$ and 0.02107 for $\lambda=0.99$, which represent a 3.5% and 5.4% reduction over the no prediction case. On the other hand, the same error is 0.02007 for $\lambda=0.9$ and 0.01883 for $\lambda=0.80$, which represent a 9.9% and 15.5% reduction respectively.

While the maximum error is decreasing for larger order models, the mean relative error and the mean square error have their best values when $n=1$. For the models with $n=1$, the five step and ten step ahead predictions obtained with models with the forgetting factor equal to 0.90 are better by almost 5% than those obtained by the models with a forgetting factor of 1.00, and the same predictions obtained with models with a forgetting factor of 0.80 have improved the error measures roughly by an additional 10%. However, the error measures for the one step prediction becomes worse as the forgetting factor is set to smaller values.

5.3.2 Autoregressive Models Including the Average Travel Time of the Link

The next set of models that tested were those that include both the autoregressive component and the average travel time of the link for the corresponding time interval. Thus, the general form of these models was:

$$T_l(t) = \sum_{i=1}^{n} a_i(t-1) \cdot T_l(t-i) + d \cdot \tilde{T}_l(t) + \varepsilon(t) \qquad (5.6)$$

Again the error measures are computed for the one, five and ten steps ahead predictions for models of different autoregressive order, and they were compared to the no prediction case. The average travel time of each link used in this set of prediction models were the travel times used as a basis for routing "non-smart" vehicles, defined by equation (5.1).

5.3.2.1 Normal traffic Conditions

For evaluating the performance of the prediction models defined by equation (5.6), the travel time series of link 15 was used. The error measures were computed for the same two time periods as in the previous section, the morning (06:00 to 11:00) and evening (14:00 to 19:00) period. Table 5.12 and Table 5.13 show the values of the error measures obtained when the average travel time of the link for the corresponding time step is included, along with the autoregressive component. The results are for models with order of the autoregressive component ranging from one to ten, and for different values for the forgetting factor: 1.00, 0.99, 0.975 and 0.90.

Similar to the models with just the autoregressive component, when the forgetting factor λ is set to unity, and thus all previous prediction errors encoun-

Table 5.12: Prediction Errors of Link 15 with Autoregressive Models Including the Average Travel time of the Link under Normal Traffic Conditions - Morning Period (300 Observations 06:00-11:00)

	AR	1 Step Ahead Predictions			5 Step Ahead Predictions			10 Step Ahead Predictions		
	order	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m
$\lambda = 1.000$	1	0.00418	0.00039	0.03466	0.01018	0.00106	0.08406	0.01220	0.00133	0.10266
	2	0.00390	0.00036	0.02835	0.01016	0.00103	0.08182	0.01219	0.00129	0.10505
	3	0.00391	0.00035	0.02721	0.01013	0.00103	0.08190	0.01216	0.00129	0.10501
	4	0.00396	0.00036	0.02800	0.01026	0.00107	0.08464	0.01208	0.00131	0.10476
	5	0.00395	0.00035	0.02774	0.01018	0.00107	0.08463	0.01198	0.00130	0.10473
	10	0.00394	0.00036	0.02781	0.01027	0.00109	0.08821	0.01191	0.00131	0.10543
$\lambda = 0.990$	1	0.00429	0.00039	0.03510	0.01134	0.00123	0.10858	0.01522	0.00195	0.22840
	2	0.00402	0.00036	0.02694	0.01111	0.00114	0.09443	0.01465	0.00178	0.21003
	3	0.00411	0.00036	0.02448	0.01039	0.00104	0.07601	0.01355	0.00162	0.19190
	4	0.00413	0.00036	0.01988	0.01059	0.00114	0.09765	0.01392	0.00180	0.22006
	5	0.00410	0.00035	0.01806	0.01059	0.00109	0.08447	0.01420	0.00171	0.21366
	10	0.00426	0.00038	0.02182	0.01150	0.00118	0.09808	0.01542	0.00186	0.23765
No Predictions		0.00442	0.00040	0.03462	0.01238	0.00129	0.11725	0.01914	0.00208	0.18957

Table 5.12 (continued)

	AR	1 Step Ahead Predictions			5 Step Ahead Predictions			10 Step Ahead Predictions		
	order	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m
$\lambda = 0.975$	1	0.00429	0.00039	0.03528	0.01177	0.00129	0.11246	0.01688	0.00229	0.27290
	2	0.00402	0.00036	0.02599	0.01167	0.00120	0.10106	0.01627	0.00210	0.25884
	3	0.00408	0.00036	0.02472	0.01069	0.00108	0.09294	0.01492	0.00190	0.23743
	4	0.00412	0.00036	0.01922	0.01092	0.00118	0.10797	0.01558	0.00216	0.27648
	5	0.00416	0.00036	0.01936	0.01097	0.00114	0.10600	0.01581	0.00205	0.26972
	10	0.00439	0.00039	0.02316	0.01182	0.00123	0.11342	0.01779	0.00223	0.28312
$\lambda = 0.900$	1	0.00440	0.00040	0.03270	0.01346	0.00143	0.11421	0.02284	0.00296	0.29033
	2	0.00423	0.00038	0.02447	0.01279	0.00130	0.11175	0.02092	0.00268	0.28405
	3	0.00426	0.00037	0.02655	0.01231	0.00124	0.10545	0.02014	0.00250	0.26838
	4	0.00445	0.00038	0.02646	0.01181	0.00119	0.11365	0.02032	0.00264	0.33126
	5	0.00467	0.00040	0.02802	0.01194	0.00121	0.11321	0.02109	0.00270	0.33887
	10	0.00520	0.00046	0.03120	0.01422	0.00153	0.15649	0.02584	0.00341	0.51087
No Predictions		0.00442	0.00040	0.03462	0.01238	0.00129	0.11725	0.01914	0.00208	0.18957

Table 5.13: Prediction Errors of Link 15 with Autoregressive Models Including the Average Travel Time of the Link under Normal Traffic Conditions - Evening Period (300 Observations 14:00-19:00)

	AR	1 Step Ahead Predictions			5 Step Ahead Predictions			10 Step Ahead Predictions		
	order	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m
$\lambda = 1.000$	1	0.00555	0.00044	0.04185	0.01278	0.00110	0.07863	0.01388	0.00128	0.10044
	2	0.00529	0.00042	0.03441	0.01256	0.00105	0.07902	0.01353	0.00122	0.10483
	3	0.00527	0.00042	0.03431	0.01253	0.00104	0.07881	0.01345	0.00121	0.10587
	4	0.00517	0.00042	0.03485	0.01248	0.00107	0.08079	0.01355	0.00123	0.10445
	5	0.00517	0.00042	0.03484	0.01255	0.00106	0.08062	0.01354	0.00123	0.10451
	10	0.00510	0.00041	0.03585	0.01238	0.00105	0.08115	0.01341	0.00122	0.10480
$\lambda = 0.990$	1	0.00557	0.00045	0.04977	0.01426	0.00134	0.12068	0.01730	0.00193	0.22040
	2	0.00529	0.00043	0.04736	0.01393	0.00131	0.11937	0.01691	0.00189	0.21995
	3	0.00539	0.00043	0.04361	0.01367	0.00124	0.11039	0.01582	0.00172	0.19675
	4	0.00535	0.00043	0.04442	0.01349	0.00125	0.11539	0.01584	0.00178	0.21106
	5	0.00527	0.00043	0.04731	0.01427	0.00137	0.12808	0.01663	0.00196	0.24760
	10	0.00532	0.00043	0.04722	0.01402	0.00132	0.12438	0.01656	0.00188	0.23397
No Predictions		0.00567	0.00045	0.05035	0.01553	0.00131	0.11269	0.02108	0.00189	0.17345

Table 5.13 (continued)

	AR	1 Step Ahead Predictions			5 Step Ahead Predictions			10 Step Ahead Predictions		
	order	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m
$\lambda = 0.975$	1	0.00563	0.00045	0.05087	0.01503	0.00144	0.14075	0.01985	0.00234	0.28502
	2	0.00536	0.00043	0.04918	0.01455	0.00141	0.14695	0.01942	0.00238	0.29600
	3	0.00547	0.00044	0.04376	0.01453	0.00136	0.14075	0.01828	0.00219	0.27240
	4	0.00548	0.00044	0.04380	0.01439	0.00137	0.14376	0.01832	0.00223	0.28288
	5	0.00537	0.00044	0.04767	0.01505	0.00150	0.16187	0.01909	0.00248	0.32702
	10	0.00557	0.00045	0.04733	0.01544	0.00150	0.15957	0.02012	0.00250	0.32194
$\lambda = 0.900$	1	0.00589	0.00046	0.05031	0.01776	0.00164	0.15388	0.02833	0.00342	0.41967
	2	0.00565	0.00047	0.05479	0.01652	0.00156	0.16097	0.02663	0.00356	0.48698
	3	0.00587	0.00049	0.06364	0.01681	0.00124	0.10545	0.02014	0.00250	0.26838
	4	0.00606	0.00051	0.06106	0.01727	0.00155	0.15285	0.02641	0.00334	0.47000
	5	0.00597	0.00050	0.06100	0.01722	0.00160	0.17353	0.02620	0.00335	0.42282
	10	0.00687	0.00058	0.06395	0.02058	0.00206	0.24755	0.03152	0.00368	0.37447
No Predictions		0.00567	0.00045	0.05035	0.01553	0.00131	0.11269	0.02108	0.00189	0.17345

tered are considered in the calculation of the parameters a_i and d, the resulting errors are smaller than the ones obtained when λ is set to smaller values. This is true for both time periods, for all predictions, the one, five and ten steps ahead, and for all three error measures, the mean absolute relative error, the mean relative square error and the maximum error. As λ is set to smaller values the error measures increase as we move further into the future, i.e. for the ten step ahead predictions. In fact, for the cases that λ is equal to 0.975 and 0.90 the error measures became worse than the no predictions case. For example in the case where n=1 and for the five step ahead predictions, the mean square error for λ=1.00 is 0.00106 while for λ=0.80 the same error of the same model is 0.00143. This happens for the same reason as explained in the case of the simple autoregressive models concerning the trade off between the sensitivity to noise and the tracking ability of the model.

Ten step ahead predictions are approximately 20% worse than five step ahead predictions which are approximately two to three times worse than the one step predictions. The one step and ten step ahead predictions for the morning period obtained with a model containing an autoregressive component of order 3 and the average link travel time are shown in Figure 5.14 and Figure 5.15. The same predictions for the evening period are shown in Figure 5.16 and Figure 5.17. As can be seen in these figures, the one step prediction is very close to the observed travel time, while the ten step ahead prediction follows the general pattern of the observed values, but they are more smoothed, since they are more affected by the average values. For example, this can be seen if we consider that for the model depicted in these figures, the two step ahead prediction at time t, $\hat{T}(t+2)$ is based on the one step ahead prediction which is already based on the average travel time at time t+1, the average travel at

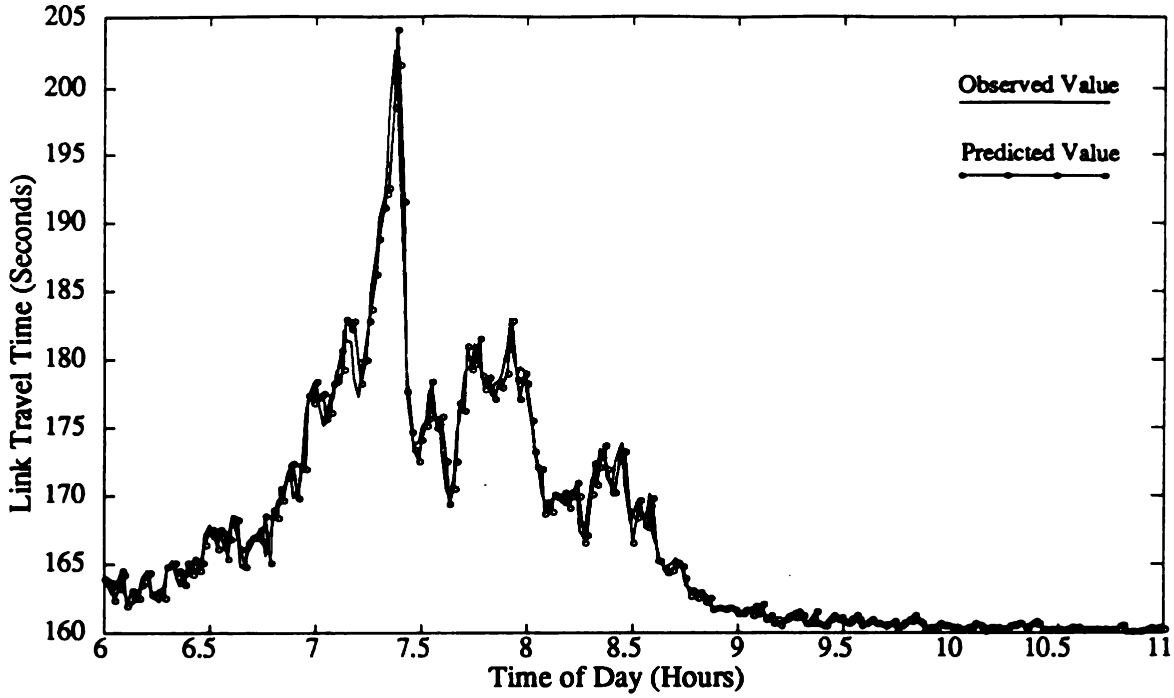


Figure 5.14: 1 Step Ahead Predictions of Travel Times of Link 15 with Autoregressive Model of Order 3, Including the Average Travel Time - λ=1.00 - Normal Operating Conditions - Morning Period

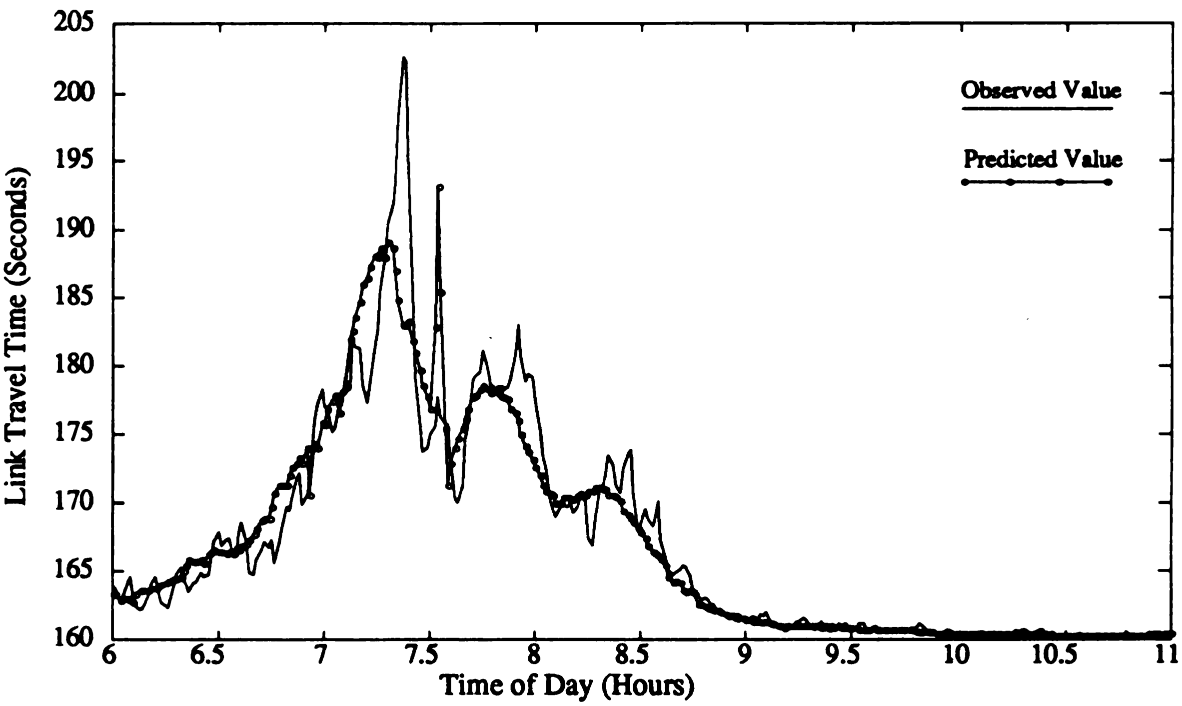


Figure 5.15: 10 Step Ahead Predictions of Travel Times of Link 15 with Autoregressive Model of Order 3, Including the Average Travel Time - λ=1.00 - Normal Operating Condition - Morning Period

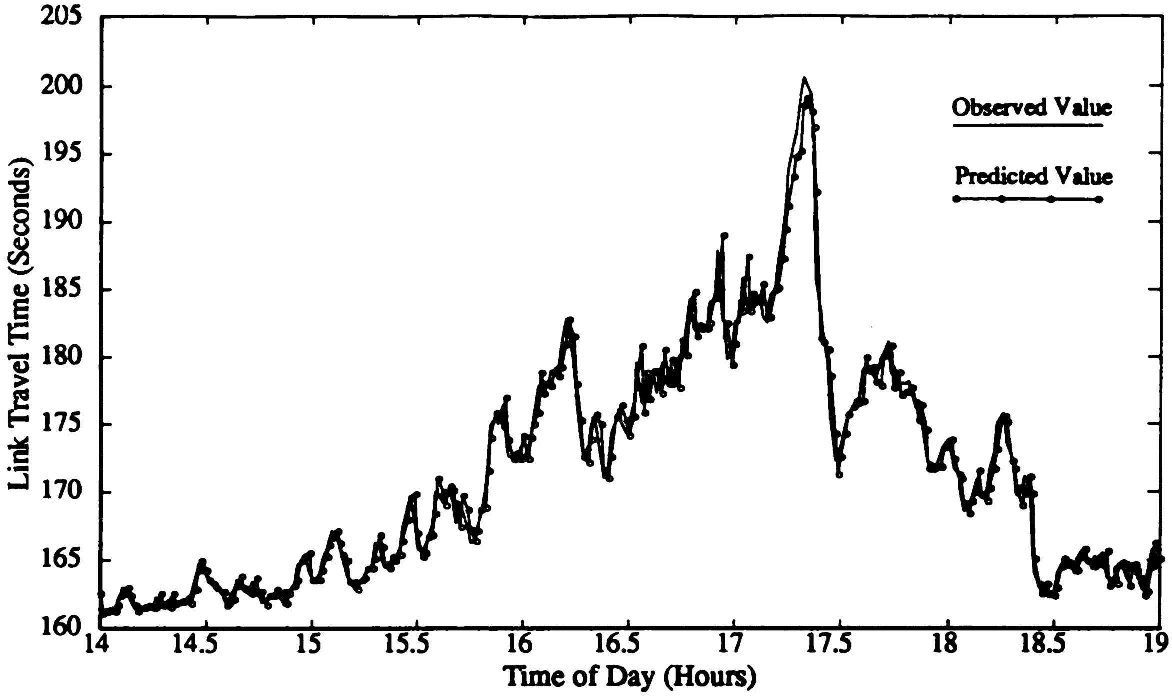


Figure 5.16: 1 Step Ahead Predictions of Travel Times of Link 15 with Autoregressive Model of Order 3, Including the Average Travel Time - λ=1.00 - Normal Operating Conditions - Evening Period

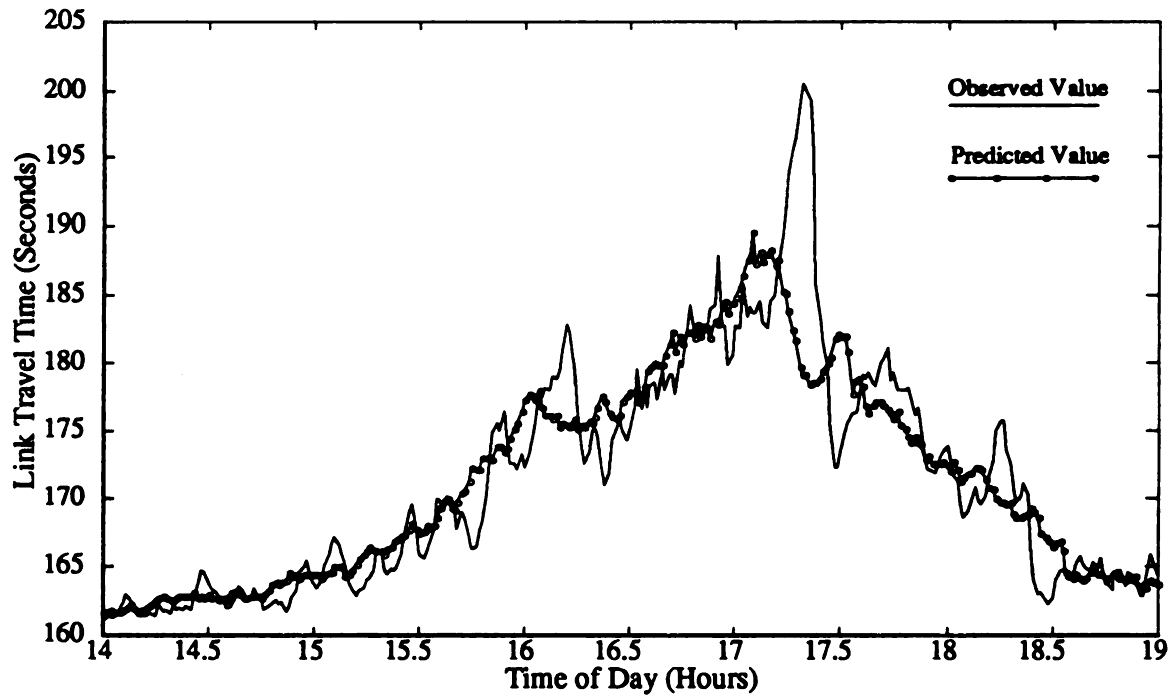


Figure 5.17: 10 Step Ahead Predictions of Travel Times of Link 15 with Autoregressive Model of Order 3, Including the Average Travel Time - λ=1.00 - Normal Operating Conditions - Evening Period

time t+2 and the observed values at times t and t-1:

$$\hat{T}(t+2) = a_1\hat{T}(t+1) + a_2T(t) + a_3T(t-1) + d\tilde{T}(t+2) =$$
$$(a_1^2 + a_2)\,T(t) + \,(a_3 + a_2a_1)\,T(t-1) + a_1a_3T(t-2) + d\tilde{T}(t+2) + da_1\tilde{T}(t+1)$$

Since the average link travel times are given in five minute intervals, the effect of the same value for the average travel time is magnified further into the future predictions.

The relative improvement of these models over the no prediction case when λ is set to 1.00 and 0.99 is shown in Table 5.14, and Table 5.15 for the morning and evening periods respectively. In the case where λ is set to 1.00 the improvement for both time periods are close to 20% and 35% for the mean relative error for the five and ten step predictions respectively, while the maximum error is reduced by approximately 30% and 44%. The one step predictions do not show such dramatic improvements but the errors associated with the one step predictions are already very small. When λ is set to values smaller than one, the maximum error deteriorates. Even when λ is set to 0.99 there are instances, at least for the ten step ahead predictions, with a larger error than in the no prediction counterpart. In the same table, the ranking of the models is shown for values of λ equal to 1.00 and 0.99. The model with an autoregressive component of order three appears to give the best results, for both time periods. The models with autoregression order 1 as well as those of order 10 gave the worst results, which indicates that in the first case the model was too “small” to include the system, while in the later case the model is overparametrized.

Table 5.14: Percent Difference of Prediction Error of AR Models Including the Average Link Travel Time Term from the No Predictions Case - Morning Period (06:00-11:00)

	AR	1 Step Ahead Predictions[1]			5 Step Ahead Predictions			10 Step Ahead Predictions			Rank[2]
		$\Delta\bar{e}$	$\Delta\bar{e}_s$	Δe_m	$\Delta\bar{e}$	$\Delta\bar{e}_s$	Δe_m	$\Delta\bar{e}$	$\Delta\bar{e}_s$	Δe_m	
$\lambda = 1.000$	1	-5.4 (6)	-3.6 (6)	+0.1 (6)	-17.8 (3)	-18.2 (3)	-28.3 (3)	-36.2 (6)	-36.0 (6)	-45.8 (1)	40 (5)
	2	-11.7 (1)	-11.3 (1)	-18.1 (5)	-18.0 (2)	-19.9 (2)	-30.2 (2)	-36.3 (5)	-37.8 (2)	-44.6 (4)	24 (2)
	3	-11.5 (2)	-11.4 (1)	-21.4 (1)	-18.2 (1)	-20.5 (1)	-30.1 (1)	-36.5 (4)	-38.0 (1)	-44.6 (4)	16 (1)
	4	-10.5 (5)	-11.0 (5)	-19.1 (4)	-17.2 (5)	-16.9 (5)	-27.8 (4)	-36.9 (3)	-37.2 (4)	-44.7 (3)	38 (4)
	5	-10.7 (4)	-11.3 (1)	-19.9 (2)	-17.8 (3)	-17.4 (4)	-27.8 (4)	-37.4 (2)	-37.5 (3)	-44.8 (2)	25 (3)
	10	-10.9 (3)	-11.2 (4)	-19.7 (3)	-17.1 (6)	-15.3 (6)	-24.8 (6)	-37.8 (1)	-36.9 (5)	-44.4 (6)	40 (5)
$\lambda = 0.990$	1	-2.9 (6)	-2.3 (6)	+1.4 (6)	-8.4 (5)	-5.0 (6)	-7.4 (6)	-20.5 (5)	-6.4 (6)	+20.5 (5)	51 (6)
	2	-8.9 (1)	-9.9 (2)	-22.2 (5)	-10.3 (4)	-11.9 (3)	-19.5 (3)	-23.5 (4)	-14.5 (3)	+10.8 (2)	27 (3)
	3	-7.1 (2)	-9.7 (4)	-29.3 (4)	-16.1 (1)	-19.5 (1)	-35.2 (1)	-29.2 (1)	-22.1 (1)	+1.2 (1)	16 (1)
	4	-6.6 (4)	-9.8 (3)	-42.6 (2)	-14.5 (2)	-11.3 (4)	-16.7 (4)	-27.3 (2)	-13.3 (4)	+16.1 (4)	29 (4)
	5	-7.3 (3)	-11.3 (1)	-47.8 (1)	-11.5 (3)	-15.4 (2)	-27.9 (2)	-25.8 (3)	-17.9 (2)	+12.7 (3)	20 (2)
	10	-3.6 (5)	-5.4 (5)	-37.0 (3)	-7.1 (6)	-8.4 (5)	-16.4 (5)	-19.5 (6)	-10.4 (5)	+25.4 (6)	46 (5)

1. Numbers in parentheses denote relative rank.
2. Cell numbers indicate the sum of ranks, and numbers in parentheses indicate relative rank of the sums.

Table 5.15: Percent Difference of Prediction Error of AR Models Including the Average Link Travel Time Term from the No Predictions Case - Evening Period (14:00-19:00)

	AR	1 Step Ahead Predictions[1]			5 Step Ahead Predictions			10 Step Ahead Predictions			Rank[2]
		$\Delta\bar{e}$	$\Delta\bar{e}_s$	Δe_m	$\Delta\bar{e}$	$\Delta\bar{e}_s$	Δe_m	$\Delta\bar{e}$	$\Delta\bar{e}_s$	Δe_m	
$\lambda = 1.000$	1	-2.1 (6)	-2.2 (3)	-16.9 (5)	-17.7 (5)	-16.0 (6)	-30.2 (1)	-34.2 (5)	-32.3 (5)	-42.1 (1)	37 (4)
	2	-6.6 (5)	-6.6 (2)	-31.7 (2)	-19.2 (4)	-19.7 (3)	-29.9 (3)	-35.8 (3)	-35.4 (2)	-39.6 (4)	28 (3)
	3	-7.1 (4)	-6.7 (1)	-31.8 (1)	-19.3 (3)	-20.6 (1)	-30.1 (2)	-36.2 (2)	-35.9 (1)	-38.9 (5)	20 (1)
	4	-8.8 (3)	-6.7 (1)	-30.8 (3)	-19.6 (2)	-18.3 (5)	-28.3 (5)	-35.7 (4)	-35.1 (3)	-39.8 (2)	28 (3)
	5	-8.9 (2)	-6.7 (1)	-30.8 (3)	-19.2 (4)	-18.8 (4)	-28.5 (4)	-35.8 (3)	-34.9 (4)	-39.7 (3)	28 (3)
	10	-10.1 (1)	-6.7 (1)	-28.8 (4)	-20.3 (1)	-19.8 (2)	-27.9 (6)	-36.4 (1)	-35.4 (2)	-39.6 (4)	22 (2)
$\lambda = 0.990$	1	-1.8 (6)	0.0 (2)	-1.1 (6)	-8.2 (5)	+2.4 (5)	+7.1 (4)	-17.9 (6)	+2.2 (5)	+27.1 (4)	43 (6)
	2	-6.6 (2)	-4.4 (1)	-5.9 (5)	-10.3 (3)	0.0 (3)	+5.9 (3)	-19.8 (5)	0.0 (4)	+26.8 (3)	29 (3)
	3	-5.0 (5)	-4.4 (1)	-13.4 (1)	-12.0 (2)	-5.3 (1)	-2.0 (1)	-25.0 (1)	-8.9 (1)	+13.4 (1)	14 (1)
	4	-5.7 (4)	-4.4 (1)	-11.8 (2)	-13.1 (1)	-4.3 (2)	+2.4 (2)	-24.9 (2)	-5.9 (2)	+21.7 (2)	18 (2)
	5	-7.0 (1)	-4.4 (1)	-6.0 (4)	-8.1 (6)	+4.5 (6)	+13.7 (6)	-21.1 (4)	+3.7 (6)	+42.8 (6)	40 (5)
	10	-6.1 (3)	-4.4 (1)	-6.2 (3)	-9.7 (4)	+0.8 (4)	+10.4 (5)	-21.4 (3)	-0.4 (3)	+34.9 (5)	31 (4)

1. Numbers in parentheses denote relative rank.
2. Cell numbers indicate the sum of ranks, and numbers in parentheses indicate relative rank of the sums.

5.3.2.2 Congested Conditions Due to a Traffic Accident

In the case where the prediction model was applied when a traffic accident had occurred, predictions demonstrated similar behavior as those obtained with models containing just the autoregressive component. This was expected since the average link travel time does not contain much information related to the traffic conditions governing the link in the case of a traffic accident. The error measures for this set of models for the congested conditions for the time period from 8:30 to 10:00 are shown in Table 5.16.

When λ was set to one, the mean relative and mean square errors produced by this set of models were almost the same as those obtained with the models including the autoregressive component alone. The maximum errors are smaller due to the smoothing effect of the average link travel time term that is included in the model. For smaller values of the forgetting factor, i.e. 0.90 and 0.80, the maximum error is increased up to 13.98 for the ten step predictions when the autoregressive component of the model is of order higher than 2. However, for larger values of the forgetting factor, the mean errors are similar or even worse than those produced by the no predictions case, while for smaller values of λ both mean errors become smaller. When n=1 the mean relative error is decreased by 24% and 33% for the five step predictions and the ten step prediction respectively, and the mean square error by 20% and 31% respectively for the same predictions. This improvement of the mean prediction errors can partially be attributed to the fact that predictions during the first 34 minutes (from 8:00 to 8:34 when the accident started) made with this models are of better quality than those obtained with the autoregressive component alone, as was shown in the previous paragraph.

Table 5.16: Prediction Errors of Link 19 with Autoregressive Models Including the Average Travel Time of the Link under Congested Traffic Conditions Due to Accident - (150 Observations 8:00-10:30)

	AR order	1 Step Ahead Predictions $\bar{e}$	$\bar{e}_s$	e_m	5 Step Ahead Predictions $\bar{e}$	$\bar{e}_s$	e_m	10 Step Ahead Predictions $\bar{e}$	$\bar{e}_s$	e_m
$\lambda = 1.000$	1	0.02676	0.00718	0.89060	0.12894	0.02127	0.89058	0.26633	0.03862	1.69649
	2	0.03169	0.00924	0.89325	0.13292	0.02173	0.89323	0.26782	0.03835	1.67946
	3	0.03564	0.01021	0.89089	0.13620	0.02202	0.89071	0.27244	0.03855	1.65882
	4	0.03884	0.01076	0.89323	0.14083	0.02250	0.89323	0.27550	0.03846	1.63336
	5	0.04486	0.01221	0.89323	0.14323	0.02252	0.89323	0.28528	0.03911	1.60844
$\lambda = 0.990$	1	0.03502	0.00944	0.89325	0.15234	0.02419	0.94775	0.31782	0.04520	2.02017
	2	0.03506	0.00942	0.89325	0.15509	0.02432	0.94273	0.32181	0.04525	2.01934
	3	0.04499	0.01183	0.89325	0.16691	0.02539	0.93581	0.33408	0.04577	2.00920
	4	0.04753	0.01208	0.89325	0.16819	0.02486	0.92634	0.34560	0.04641	1.98993
	5	0.05501	0.01381	0.89325	0.17435	0.02541	0.91750	0.35187	0.04654	1.96846
No Predictions		0.02404	0.00715	0.89068	0.12450	0.02228	0.95956	0.27445	0.04363	1.98757

Table 5.16 (continued)

	AR	1 Step Ahead Predictions			5 Step Ahead Predictions			10 Step Ahead Predictions		
	order	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m
$\lambda = 0.900$	1	0.03027	0.00928	0.89325	0.09983	0.01886	0.89323	0.19867	0.03324	1.71187
	2	0.03699	0.01125	0.89325	0.10589	0.01975	0.89323	0.21160	0.03442	1.71285
	3	0.04293	0.01283	0.89325	0.13509	0.02502	2.06907	0.29611	0.06120	7.29225
	4	0.05066	0.01365	0.89325	0.20119	0.05273	6.68046	0.49741	0.14128	13.97905
	5	0.05623	0.01518	0.89325	0.19797	0.06051	8.18434	0.33184	0.07253	7.45277
$\lambda = 0.800$	1	0.02863	0.00924	0.89325	0.09495	0.01784	0.89323	0.18470	0.02998	1.68386
	2	0.03819	0.01153	0.89325	0.12546	0.03062	3.56942	0.19408	0.03086	1.68382
	3	0.04727	0.01359	0.89325	0.18931	0.06447	8.51631	0.29894	0.07008	8.25381
	4	0.05224	0.01448	0.89325	0.30967	0.10975	12.58208	0.47823	0.14471	13.98564
	5	0.06177	0.01615	0.89325	0.27221	0.09908	12.94149	0.44152	0.12820	13.92341

No Predictions	0.02404	0.00715	0.89068	0.12450	0.02228	0.95956	0.27445	0.04363	1.98757

5.3.3 Models Including the Convection Term

Based on the findings so far, models including an autoregressive component of order up to four, and the average link travel time for the corresponding time interval have given the best results, at least in the normal traffic operations case. In the following models, including the convection term will be examined to determine if further improvements can be made to the quality of the link travel time predictions. According to equations (3.3) and (3.4.b), the general form of these models will be:

$$T_l(t) = \sum_{i=1}^{n} a_i(t-1) \cdot T_l(t-i) + \sum_{k \in I} \sum_{j=1}^{m_k} b_k(t-1) \cdot T_k(t-j) + D \cdot \tilde{T}_l(t) + \varepsilon(t) \qquad (5.7)$$

where I is the set of links k that end at the starting node of link l. Because no cycling is allowed in the network (vehicles looping in a sequence of links), set I will have up to three elements, one link from one of the major arterials of the test network and one or two connector links.

The number of variables in the autoregressive component will be held constant and equal to three, since the combination of this AR component with the average travel time gave the best results in the previous step. Furthermore, for simplicity we will assume that the number of variables m_k used in the model from each connector link k ending at the entrance of the link under consideration l, will be the same and equal to m_c, unless it is otherwise specified. The error measures given by equations (5.2), (5.3) and (5.4) will be computed for a set of models with different values for m_a, where a is the previous major link, and the results will again be compared to the no prediction case.

5.3.3.1 Normal Traffic Conditions

For normal traffic conditions the series with travel time observations from link 15 was used again. In this case, the set I consists of the links {8, 11, 16}. The error measures for the morning and evening time periods, given by this set of models are shown in Table 5.17 and Table 5.18. Since the results that have been found so far indicate the performance of the prediction model deteriorates for values of the forgetting factor smaller than one, only the results for λ=1.00 and λ=0.99 are shown in these tables. Different values for the number of the past values of each link from set I included in the model, ranging from one to five, were tested.

When the number of past link travel times included in the model from the connector links, m_c, was greater than 2 the model became unstable and the error increased rapidly. This happens because links 11 and 16 are empty most of the time, and therefore their travel time remains equal to their free flow travel time for long periods. When light volumes occur this creates a big shock to the system, especially when λ=0.99. The RLS prediction algorithm tries to identify the true values of the respective parameters, which should be close to zero. When more than one such parameter exists in the model, the algorithm may assign values to the parameters such that the total contribution made by the connector links to the predictions of the travel time of link 15 is zero or close to zero. This is illustrated in Figure 5.18, where the sum of the parameters ($b_{11,1}+b_{11,2}$) and ($b_{16,1}+b_{16,2}$) are approximately equal so they cancel each other, since the travel time on both links 11 and 16 is equal to their free flow travel time, 100 seconds. At the end of the second simulated day, and for the case of λ=1.00, all parameters corresponding to the connector links converge to zero. However, when λ is set to values smaller than one, even though the sum of the

Table 5.17: Prediction Errors of Link 15 with Models Including the Convection Term - Normal Traffic Conditions - Morning Period (300 Observations 06:00-11:00)

	m_a	1 Step Ahead Predictions			5 Step Ahead Predictions			10 Step Ahead Predictions		
		$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m
$\lambda = 1.000$	1	0.00362	0.00032	0.02282	0.00719	0.00066	0.04601	0.01129	0.00119	0.09540
	2	0.00315	0.00029	0.03569	0.00717	0.00063	0.04635	0.01106	0.00116	0.09085
	3	0.00317	0.00030	0.03556	0.00717	0.00063	0.04681	0.01107	0.00116	0.09082
	4	0.00303	0.00031	0.04791	0.00781	0.00068	0.05192	0.01124	0.00117	0.09042
	5	0.00309	0.00032	0.04882	0.00770	0.00066	0.04952	0.01118	0.00116	0.09050
$\lambda = 0.990$	1	0.00379	0.00034	0.02120	0.00840	0.00078	0.06726	0.01372	0.00161	0.20249
	2	0.00303	0.00028	0.02793	0.00816	0.00073	0.05936	0.01318	0.00153	0.18218
	3	0.00284	0.00027	0.03422	0.00831	0.00074	0.06697	0.01333	0.00156	0.19258
	4	0.00281	0.00029	0.04027	0.00837	0.00076	0.06749	0.01329	0.00155	0.19143
	5	0.00281	0.00029	0.04017	0.00850	0.00082	0.07277	0.01357	0.00161	0.18568
No Predictions		0.00442	0.00040	0.03462	0.01238	0.00129	0.11725	0.01914	0.00208	0.18957

Table 5.18: Prediction Errors of Link 15 with Models Including the Convection Term - Normal Traffic Conditions - Evening Period (300 Observations 14:00-19:00)

		1 Step Ahead Predictions			5 Step Ahead Predictions			10 Step Ahead Predictions		
	m_a	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m
$\lambda = 1.000$	1	0.00484	0.00038	0.03155	0.00864	0.00064	0.04447	0.01364	0.00118	0.09882
	2	0.00353	0.00029	0.02357	0.00881	0.00065	0.03127	0.01339	0.00116	0.09776
	3	0.00350	0.00029	0.02300	0.00872	0.00064	0.03141	0.01339	0.00116	0.09767
	4	0.00334	0.00028	0.02256	0.00958	0.00070	0.03519	0.01336	0.00117	0.09818
	5	0.00339	0.00028	0.02267	0.00962	0.00070	0.03128	0.01337	0.00117	0.09825
$\lambda = 0.990$	1	0.00488	0.00040	0.04143	0.01005	0.00082	0.07298	0.01538	0.00145	0.12561
	2	0.00349	0.00029	0.02560	0.00871	0.00068	0.04965	0.01508	0.00143	0.12597
	3	0.00329	0.00027	0.02038	0.00891	0.00069	0.04293	0.01531	0.00147	0.13791
	4	0.00306	0.00026	0.02516	0.00889	0.00070	0.05874	0.01528	0.00145	0.13044
	5	0.00298	0.00026	0.02892	0.00886	0.00072	0.07793	0.01544	0.00145	0.12181
No Predictions		0.00567	0.00045	0.05035	0.01553	0.00131	0.11269	0.02108	0.00189	0.17345

parameters $b_{11,j}$ and $b_{16,j}$ still tend to zero, the individual parameters may assume large values, and even very small differences in the travel time of the connector links will introduce large errors to the predictions made by the algorithm. As it is noted by Ljung and Söderström (Ljung and Söderström 1983, page 203) this is a natural result, since those parameters, or linear combination of parameters, which are not affecting the predictions, cannot be estimated by input-output data, and in such instances the correlation matrix $R(t)$ becomes singular. Therefore, in all experiments described in the following the variable m_c will be set at 0.

In Figure 5.19 the parameters $b_{8,j}$, j=1,2,3 are also shown. These parameters correspond to the values of the travel time of link 8 that are used in the prediction model. As can be seen, $b_{8,1}$ and $b_{8,3}$ are close to zero, while $b_{8,2}$ has a larger value. This time lag represents the time required for upstream traffic to reach downstream links. The travel time of link 8 is 120 seconds, and thus the effect of upstream traffic on downstream travel conditions is lagged by two to three time steps. Actually the convection wave propagates to the downstream link somewhat faster than the speed of the vehicles traversing links 8 to enter link 15, which is indeed in accordance with the findings of Kyte et. al. (Kyte et. al. 1989).

As can be seen from the results in Table 5.19 and Table 5.20 the model with m_a equal to two gives the best results for the morning period for both λ=1.00 and λ=0.99. For the evening period the model with m_a equal to three is slightly better than the model with m_a equal to two, while for both time periods including a larger number of variables into the model increases the error. The ranking of the models for the morning and evening time periods is shown in Table 5.19

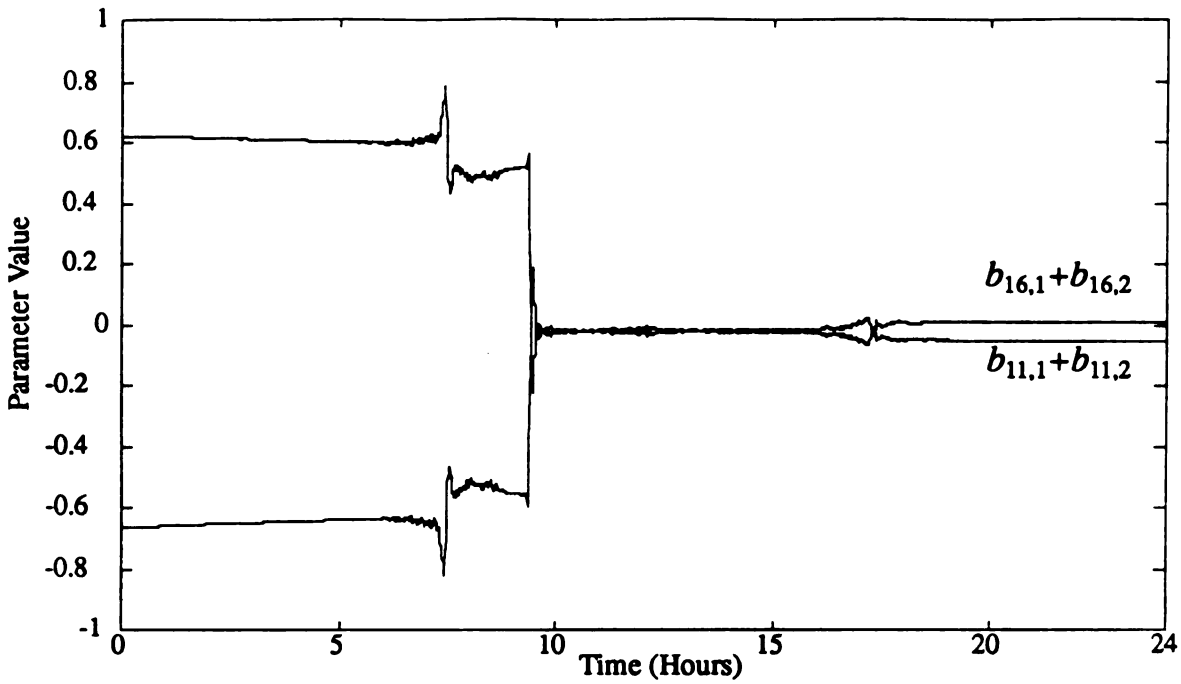


Figure 5.18: Value of Parameters Corresponding to Connector Links (11 and 16) in the Convection Term

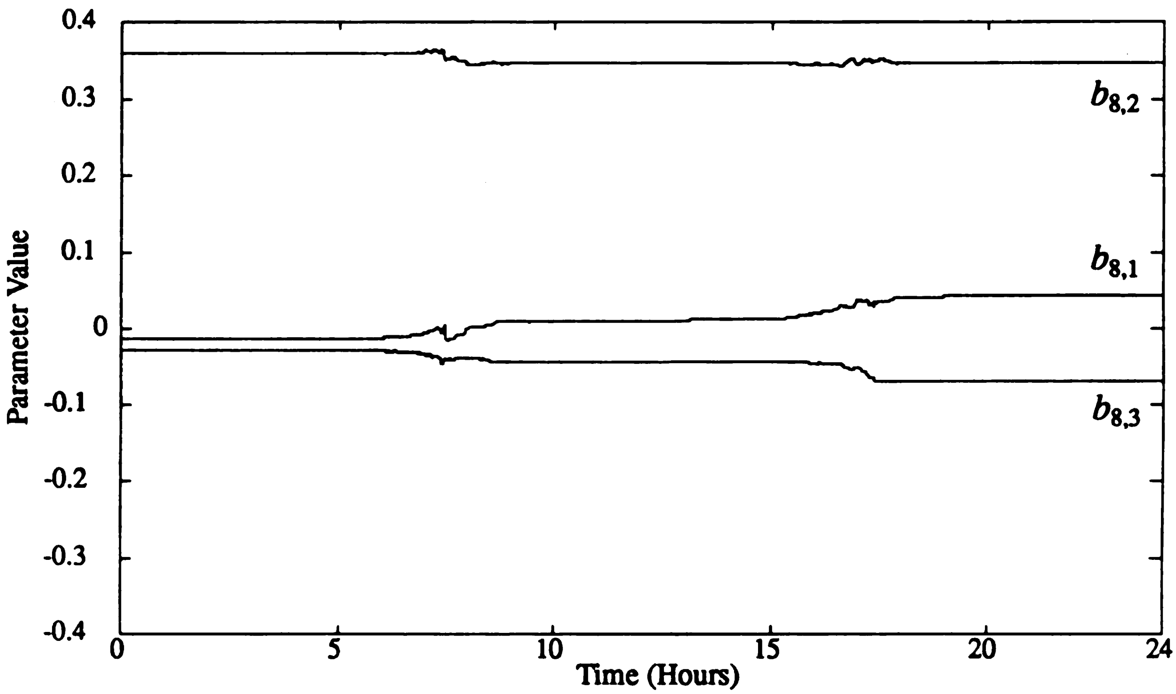


Figure 5.19: Value of Parameters Corresponding to the Major Arterial Link (8) in the Convection Term

and Table 5.20 respectively. In the same table the improvement of these models over the no predictions case is also shown. As can be seen, all models give significant improvements over the no predictions case. In the morning period error reductions reach up to 60% for the maximum error, when λ=1.00, while in the evening period reductions of the same error exceed 70%. While for the one step predictions, the maximum error is getting worse, the mean relative and the mean square errors are up to 30% better. However this is not too serious since prediction errors for the one step ahead predictions are already very small. On the other hand, all the error measures for the five and ten step ahead predictions for both time periods are significantly better than the no predictions case. From all the models that were tried the ones with m_a=1 gave the worst results in both time periods, while the errors produced from the rest of the models were similar. This is explained by the fact that the one step back travel time values of the upstream links do not contain any information for the downstream traffic conditions, since it has not yet started affecting these conditions.

Also in Table 5.21 are shown the relative improvement of the performance of the model including the convection term with m_a=2 over the no prediction case, the simple autoregressive model and the model consisting of only the autoregressive component and the average travel time of the link. As can be seen, the improvement to the one step predictions are of the order of 20% for the morning and 30% for the evening period. The five step ahead predictions also show significant improvements over the models that include the average travel time of the link: approximately 30% for both the morning and the evening time periods for the mean relative absolute error, 36% for the mean square error and above 40% for the maximum error. The ten step ahead pre-

Table 5.19: Percent Difference of Prediction Error of Models Including the Convection Term from the Case of No Predictions - Morning Period (6:00-11:00)

	m_a	1 Step Ahead Predictions[1]			5 Step Ahead Predictions			10 Step Ahead Predictions			Rank[2]
		$\Delta\bar{e}$	$\Delta\bar{e}_s$	Δe_m	$\Delta\bar{e}$	$\Delta\bar{e}_s$	Δe_m	$\Delta\bar{e}$	$\Delta\bar{e}_s$	Δe_m	
$\lambda = 1.000$	1	-18.1 (5)	-20.0 (4)	-34.1 (1)	-41.9 (3)	-48.8 (5)	-60.8 (1)	-41.0 (5)	-57.5 (5)	-49.7 (5)	34 (5)
	2	-28.7 (3)	-27.5 (1)	+3.1 (3)	-42.1 (1)	-51.2 (1)	-60.5 (2)	-42.2 (1)	-58.6 (1)	-52.1 (2)	15 (1)
	3	-28.3 (4)	-25.0 (2)	+2.7 (2)	-42.1 (1)	-51.2 (1)	-60.1 (3)	-42.2 (1)	-58.6 (1)	-52.1 (2)	17 (2)
	4	-31.4 (1)	-22.5 (3)	+38.4 (4)	-36.9 (5)	-47.3 (4)	-55.7 (5)	-41.3 (4)	-58.2 (4)	-52.1 (2)	32 (4)
	5	-30.1 (2)	-20.0 (4)	+41.0 (5)	-37.8 (4)	-48.8 (3)	-57.8 (4)	-41.6 (3)	-58.6 (1)	-52.3 (2)	28 (3)
$\lambda = 0.990$	1	-14.3 (5)	-15.0 (5)	-38.8 (1)	-32.1 (4)	-39.5 (4)	-42.6 (3)	-28.3 (5)	-42.5 (4)	+6.8 (5)	36 (5)
	2	-31.4 (4)	-30.0 (2)	-19.3 (2)	-34.1 (1)	-43.4 (2)	-49.4 (1)	-36.4 (1)	-45.4 (1)	-3.9 (1)	15 (1)
	3	-35.7 (3)	-32.5 (1)	-1.2 (3)	-32.9 (2)	-42.6 (1)	-42.9 (2)	-30.4 (3)	-44.3 (3)	+1.6 (4)	22 (2)
	4	-36.4 (1)	-27.5 (3)	+16.3 (5)	-32.4 (3)	-41.1 (3)	-42.4 (4)	-30.6 (2)	-44.6 (2)	+1.0 (3)	26 (3)
	5	-36.4 (1)	-27.5 (3)	+16.0 (4)	-31.3 (1)	-36.4 (5)	-37.9 (5)	-29.1 (4)	-42.5 (4)	-2.1 (2)	29 (4)

1. Numbers in parentheses denote relative rank.
2. Cell numbers indicate the sum of ranks, and numbers in parentheses indicate relative rank of the sums.

Table 5.20: Percent Difference of Prediction Error of Models Including the Convection Term from the Case of No Predictions - Evening Period (14:00-19:00)

	m_a	1 Step Ahead Predictions[1]			5 Step Ahead Predictions			10 Step Ahead Predictions			Rank[2]
		$\Delta\bar{e}$	$\Delta\bar{e}_s$	Δe_m	$\Delta\bar{e}$	$\Delta\bar{e}_s$	Δe_m	$\Delta\bar{e}$	$\Delta\bar{e}_s$	Δe_m	
$\lambda = 1.000$	1	-14.6 (5)	-15.6 (5)	-37.3 (5)	-44.4 (2)	-51.1 (1)	-60.5 (5)	-35.3 (5)	-37.6 (5)	-43.0 (5)	38 (5)
	2	-37.7 (4)	-35.6 (3)	-53.2 (4)	-43.3 (3)	-50.4 (3)	-72.3 (1)	-36.5 (3)	-38.6 (1)	-43.6 (2)	24 (4)
	3	-38.3 (3)	-35.6 (3)	-54.3 (3)	-43.9 (1)	-51.1 (1)	-72.1 (3)	-36.5 (3)	-38.6 (1)	-43.7 (1)	19 (1)
	4	-41.1 (1)	-37.8 (1)	-55.2 (1)	-38.3 (4)	-46.6 (4)	-68.8 (4)	-36.6 (1)	-38.1 (3)	-43.4 (3)	22 (2)
	5	-40.2 (2)	-37.8 (1)	-55.0 (2)	-38.1 (5)	-46.6 (4)	-72.2 (2)	-36.6 (1)	-38.1 (3)	-43.4 (3)	23 (3)
$\lambda = 0.990$	1	-13.9 (5)	-11.1 (5)	-17.7 (5)	-35.3 (5)	-37.4 (5)	-35.2 (4)	-27.0 (4)	-23.3 (2)	-27.6 (2)	37 (5)
	2	-38.4 (4)	-35.6 (4)	-49.2 (3)	-43.9 (1)	-48.1 (1)	-55.9 (2)	-28.5 (1)	-24.3 (1)	-27.4 (3)	20 (1)
	3	-42.0 (3)	-40.0 (3)	-59.5 (1)	-42.6 (5)	-47.3 (2)	-61.9 (1)	-27.4 (3)	-22.2 (5)	-20.5 (5)	28 (4)
	4	-46.0 (2)	-42.2 (1)	-50.0 (2)	-42.8 (4)	-46.6 (3)	-47.9 (3)	-27.5 (2)	-23.3 (2)	-24.8 (4)	23 (2)
	5	-47.4 (1)	-42.2 (1)	-42.6 (4)	-42.9 (3)	-45.0 (4)	-30.8 (5)	-26.8 (5)	-23.3 (2)	-29.8 (1)	26 (3)

1. Numbers in parentheses denote relative rank.
2. Cell numbers indicate the sum of ranks, and numbers in parentheses indicate relative rank of the sums.

Table 5.21: Percent Difference of Prediction Errors of Models Including the Convection Component Over the No Predictions Case, Autoregressive Models and Models Including the Average Travel Time

	Model	1 Step Ahead Predictions			5 Step Ahead Predictions			10 Step Ahead Predictions		
		$\Delta\bar{e}$	$\Delta\bar{e}_s$	Δe_m	$\Delta\bar{e}$	$\Delta\bar{e}_s$	Δe_m	$\Delta\bar{e}$	$\Delta\bar{e}_s$	Δe_m
Morning Period	No Predictions	-28.7	-27.5	+3.1	-42.1	-51.2	-60.5	-42.2	-58.6	-52.1
	AR(2)	-23.5	-21.6	-20.1	-41.0	-46.8	-58.2	-40.2	-40.8	-48.0
	AR(3) & Diurnal Term	-19.2	-19.4	-19.5	-29.2	-35.9	-43.8	-7.4	-7.8	-9.2
Evening Period	No Predictions	-37.7	-35.6	-53.2	-43.3	-50.4	-72.3	-36.5	-38.6	-43.6
	AR(2)	-33.5	-32.6	-45.6	-41.8	-49.6	-70.8	-34.3	-36.6	-40.1
	AR(3) & Diurnal Term	-33.0	-31.0	-31.3	-29.7	-37.5	-60.3	-0.4	-4.1	-7.7

dictions, although better, are not significantly improved. The maximum error shows improvement of approximately 8% for both periods. However the mean relative error, at least for the evening period does not show any significant improvements over the later model. The poorer performance of this set of models for predictions further into the future can be explained because such predictions are based not only on previous predictions of the travel time of the link that is under consideration, but also on predictions of the main arterial link(s) that ends at the entrance node of the link, i.e. link 8. The improvements over the no predictions case and the simple autoregressive models are significant and always greater than 20% for the one step predictions, 40% for

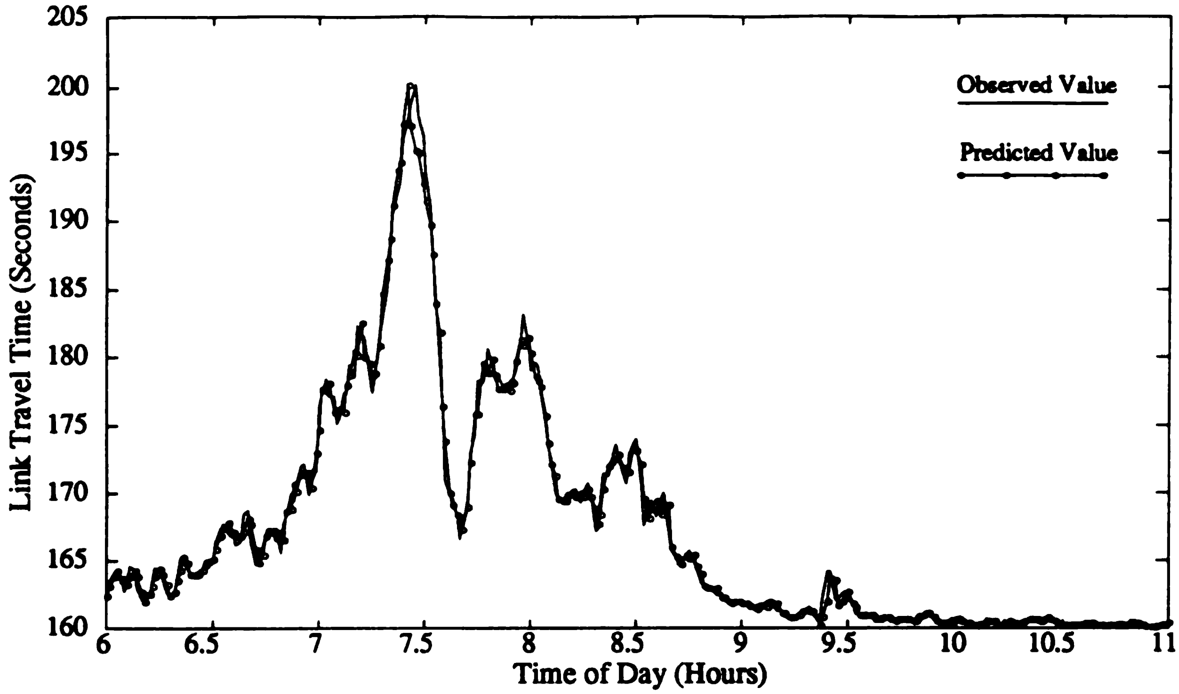


Figure 5.20: 1 Step Predictions and Observed Values of Travel Times of Link 15 - Model Including the Convection Component - Normal Traffic Conditions - Morning Period - λ=1.00, m_a=2

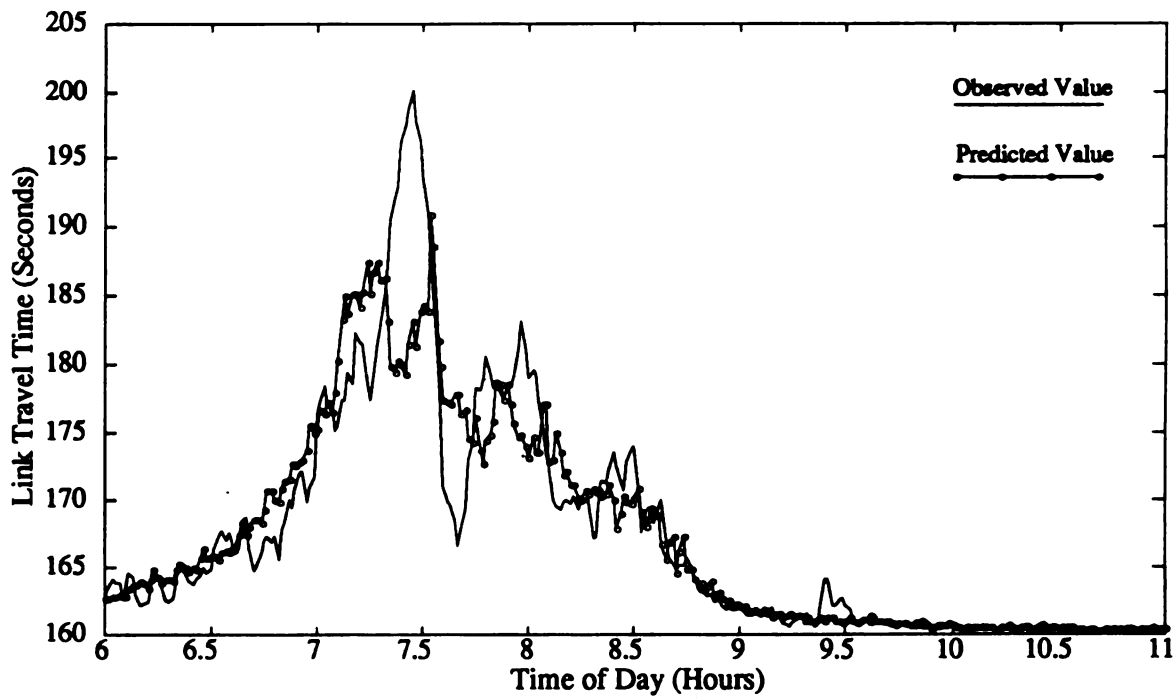


Figure 5.21: 10 Step Predictions and Observed Values of Travel Times of Link 15 - Model Including the Convection Component - Normal Traffic Conditions - Morning Period - λ=1.00, m_a=2

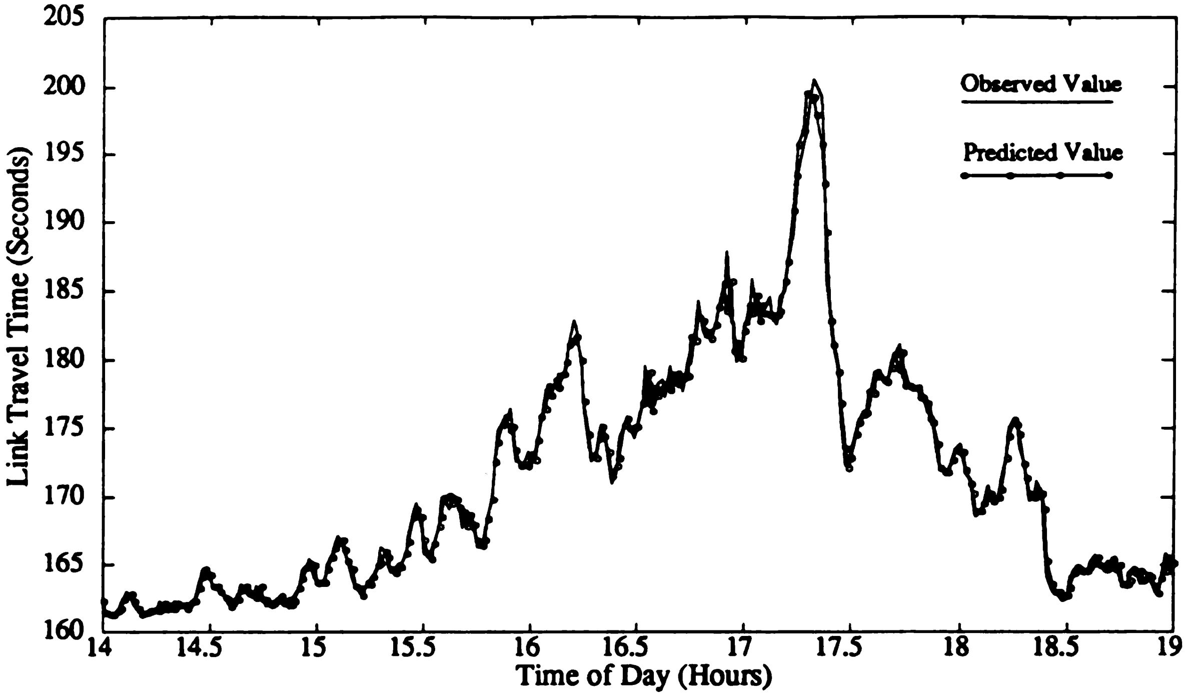


Figure 5.22: 1 Step Predictions and Observed Values of Travel Times of Link 15 - Model Including the Convection Component - Normal Traffic Conditions - Evening Period - λ=1.00, m_a=2

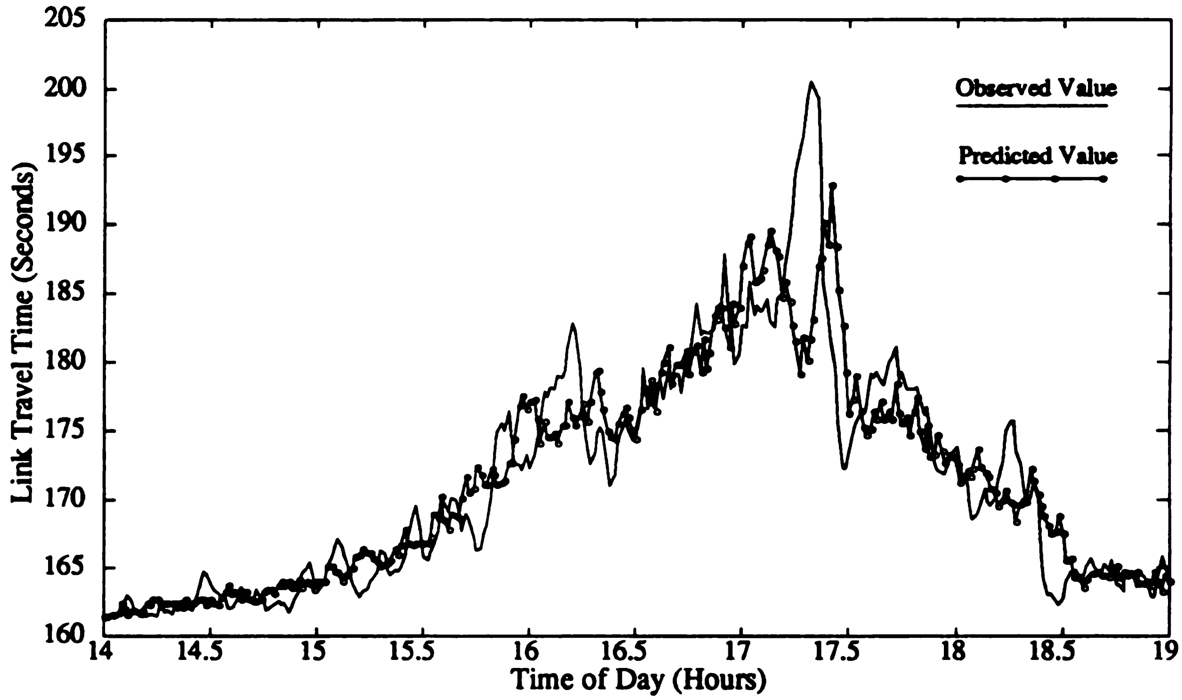


Figure 5.23: 10 Step Predictions and Observed Values of Travel Times of Link 15 - Model Including the Convection Component - Normal Traffic Conditions - Evening Period - λ=1.00, m_a=2

the five step predictions and 35% for the ten step predictions.

The one and ten step ahead predictions for the morning period obtained with these models are shown in Figure 5.20 and Figure 5.21. Also in Figure 5.22 and Figure 5.23 are shown the same predictions obtained for the evening period for the case with λ=1.00. As can be seen in the figures corresponding to the 10 step ahead predictions, the effect of the average travel time is lessened, and predictions follow future travel times more closely than when the model with just the autoregressive component and the average travel time is used.

This is shown more clearly in Figure 5.24 where the five step predictions obtained with a model including the average travel time (o) and those obtained by a model that in addition includes the convection term (+) are plotted. This is explained also by examining the value of the parameter corresponding to the average value of the travel time of the link. While in the former model the parameter stays above 0.08 for the entire second simulated day of the simulation run, for the model that includes the convection term it stays very close to zero for the entire day. At the beginning of the simulated day these two parameters are 0.0929 and 0.0021 respectively, while at the end they become 0.0855 and 0.0072. This means the contribution of the average link travel time is less in the models that contain the convection term, at least by one order of magnitude. However, when the average travel time is excluded from the model, the five step and ten step ahead predictions became worse than when this term is included in the model. The behavior of these two parameters is shown in Figure 5.25 where d_m is the parameter corresponding to the model without the convection term, and d_c to the one with it.

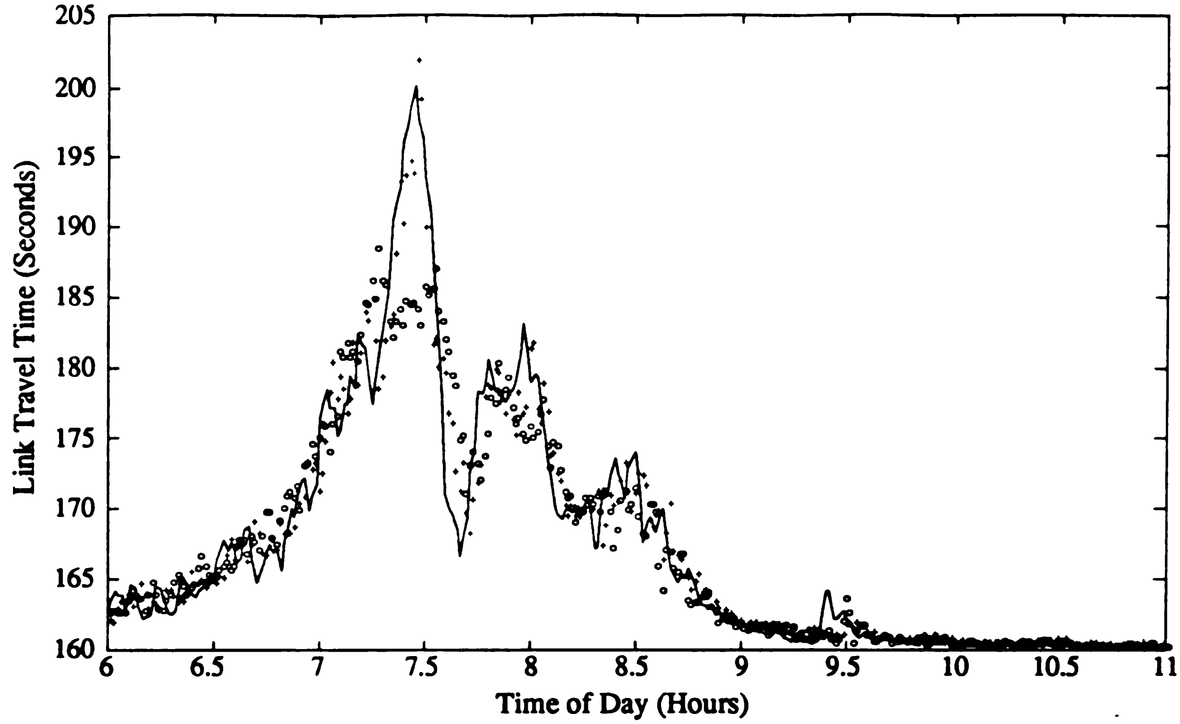


Figure 5.24: 5 Step Predictions with Model Including the Convection Term (+) and with Model Including the Average Travel Time Only (o)

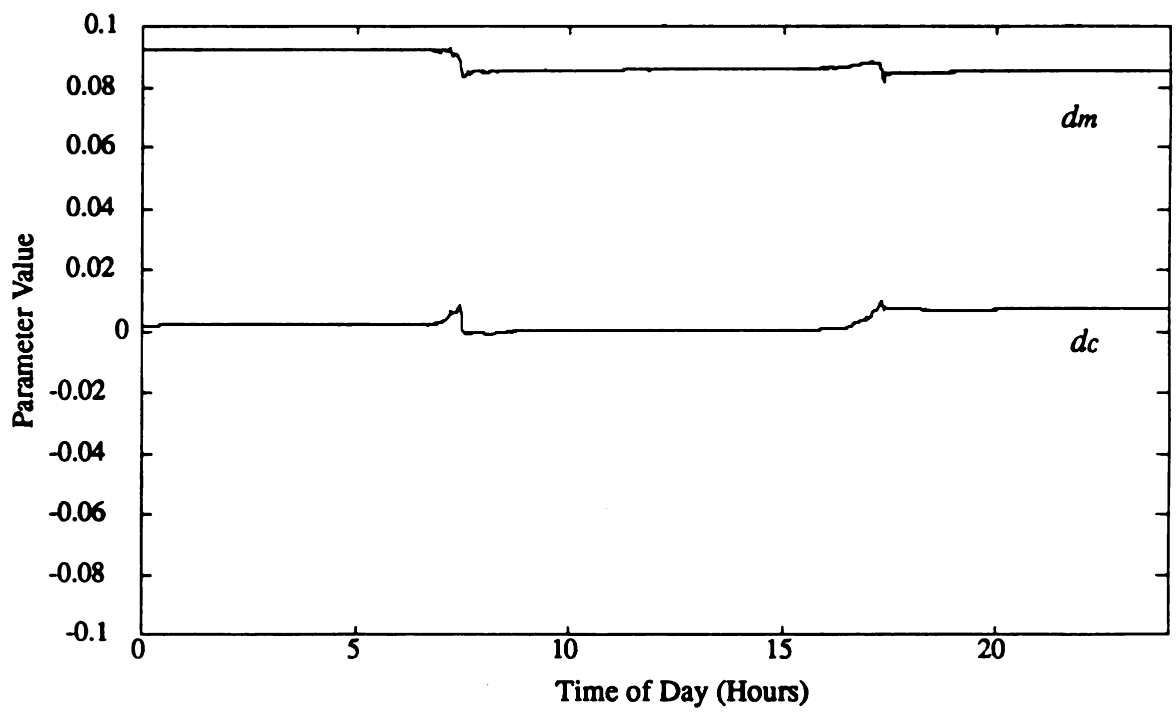


Figure 5.25: Parameter Corresponding to the Average Travel Time Term from Model Including the Convection Term (*dc*) and from Model Including the Average Travel Time Only (*dm*)

5.3.3.2 Congested Conditions Due to Traffic Accident

In the case where the link under consideration is experiencing a traffic accident, demand originating from upstream links will determine the level of congestion in terms of the duration of the aftermath of the accident. The travel time data from link 19 were used in this experiment, and thus links {12,20} comprise set I. The speed with which the queue of link 12 is dissolved will directly affect the speed with which the queue of link 19, which has started dissipating after the end of the accident, will dissolve. Therefore, the travel time of link 19 should be related to the traffic demand on link 12. Again the error measures are calculated for the time period starting at 8:00 and ending at 10:30.

The results obtained in the case when the congested link (link 19) was the one for which the predictions are performed, are shown in Table 5.22. As can be inferred from these figures, there is no improvement in the quality of the predictions made with this model configuration. In all cases, predictions made with the models including just the autoregressive component along with the average link travel time give better quality predictions.

This was expected during the first minutes of the accident, since traffic conditions of upstream traffic do not affect the occurrence of the incident. The one and ten step ahead predictions obtained with these models are shown in Figure 5.26 and Figure 5.27, for the case where λ=1.00 and in Figure 5.28 and Figure 5.29 for the case where λ=0.80. The predictions shown are obtained with a model with m_a=1 including an autoregressive component of order n=2 and the link average travel time.

Table 5.22: Prediction Errors of Link 19 with Autoregressive Models Including the Average Travel Time and the Convection Component of the Link under Congested Traffic Conditions Due to Accident - (150 Observations 08:00-10:30)

		1 Step Ahead Predictions			5 Step Ahead Predictions			10 Step Ahead Predictions		
	m_a	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m
$\lambda = 1.000$	1	0.02934	0.00743	0.89050	0.14259	0.02201	0.89010	0.29661	0.04029	1.73375
	2	0.03145	0.00785	0.89043	0.14582	0.02212	0.89012	0.30358	0.04018	1.70549
	3	0.03269	0.00802	0.88978	0.14492	0.02147	0.89031	0.30038	0.03967	1.69079
	4	0.03597	0.00844	0.88992	0.15166	0.02180	0.89041	0.31106	0.04016	1.67582
	5	0.03737	0.00847	0.88992	0.15549	0.02194	0.89042	0.31905	0.04063	1.66296
$\lambda = 0.990$	1	0.04008	0.01026	0.89325	0.16702	0.02484	0.98547	0.34885	0.04745	2.07952
	2	0.04144	0.01028	0.89325	0.17578	0.02540	0.96799	0.36489	0.04837	2.08513
	3	0.04874	0.01192	0.89325	0.18508	0.02619	0.96081	0.37094	0.04827	2.07889
	4	0.05830	0.01374	0.89325	0.19542	0.02731	1.09205	0.39483	0.04998	2.07180
	5	0.06603	0.01514	0.89325	0.20625	0.02843	1.12914	0.41112	0.05101	2.06354

No Predictions	0.02404	0.00715	0.89068	0.12450	0.02228	0.95956	0.27445	0.04363	1.98757

Table 5.22 (continued)

	m_a	1 Step Ahead Predictions			5 Step Ahead Predictions			10 Step Ahead Predictions		
		$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m
$\lambda = 0.900$	1	0.03559	0.01096	0.89325	0.10394	0.01934	0.89323	0.19734	0.03310	1.68210
	2	0.04086	0.01247	0.89325	0.11229	0.02044	0.89323	0.19917	0.03331	1.71336
	3	0.04994	0.01340	0.89325	0.12233	0.02117	0.89323	0.21028	0.03421	1.70597
	4	0.05514	0.01466	0.89325	0.12311	0.02136	0.89323	0.21414	0.03479	1.70947
	5	0.06065	0.01580	0.89325	0.12924	0.02227	0.89323	0.21268	0.03445	1.72393
$\lambda = 0.800$	1	0.03498	0.01100	0.89325	0.10727	0.01958	0.89323	0.19014	0.03058	1.68636
	2	0.04096	0.01254	0.89325	0.11311	0.02050	0.89323	0.18763	0.03104	1.69460
	3	0.05276	0.01412	0.89325	0.12422	0.02202	0.89323	0.20813	0.03332	1.68264
	4	0.05726	0.01524	0.89325	0.12839	0.02251	0.89323	0.21236	0.03405	1.67721
	5	0.06192	0.01606	0.89325	0.14262	0.02446	0.89323	0.21011	0.03390	1.67447
No Predictions		0.02404	0.00715	0.89068	0.12450	0.02228	0.95956	0.27445	0.04363	1.98757

The erratic behavior of the model during these first minutes is due to the inclusion of the convection term into the model, which creates a delayed response to the rapid increase of the travel time by a few time steps, as compared with the response of the models without the convection component. This occurs because the relationship of the travel times of links ending at the starting node of the link under consideration has been established by the model, and when the incident occurs the parameters associated with the convection term still have large values, thus greatly affecting the predictions. This is also the reason the maximum error, even when m_a is larger than 2 and the forgetting factor is set to small values, does not get such high values as it did in the previous models. During the incident though, this relationship is reduced and the prediction model behaves much like the models without the convection term.

When λ=1.00, and after the incident has ended the model underestimates the predictions for the five and ten steps ahead, because of the strong shock to the system during the accident (travel time increased almost by a factor of 10) and because of the long memory of the model, due to which the parameters associated with the convection term diminishes much slower. In the case of λ=0.80 the shock is overcome fast and predictions during the period of the dissipation of the queue are close to the real values. Also, immediately after the queue is dissipated, ten step predictions return to normal values.

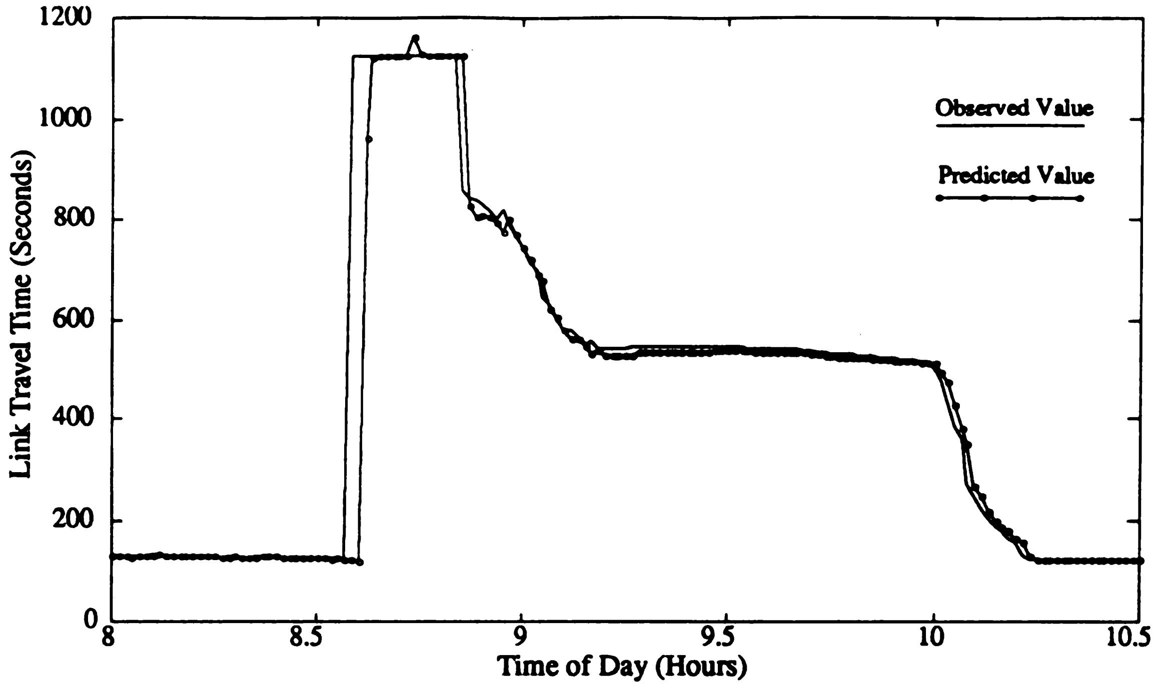


Figure 5.26: 1 Step Predictions of Travel Times of Link 19 with Model Including the Convection Term - Congested Traffic Conditions - λ=1.00

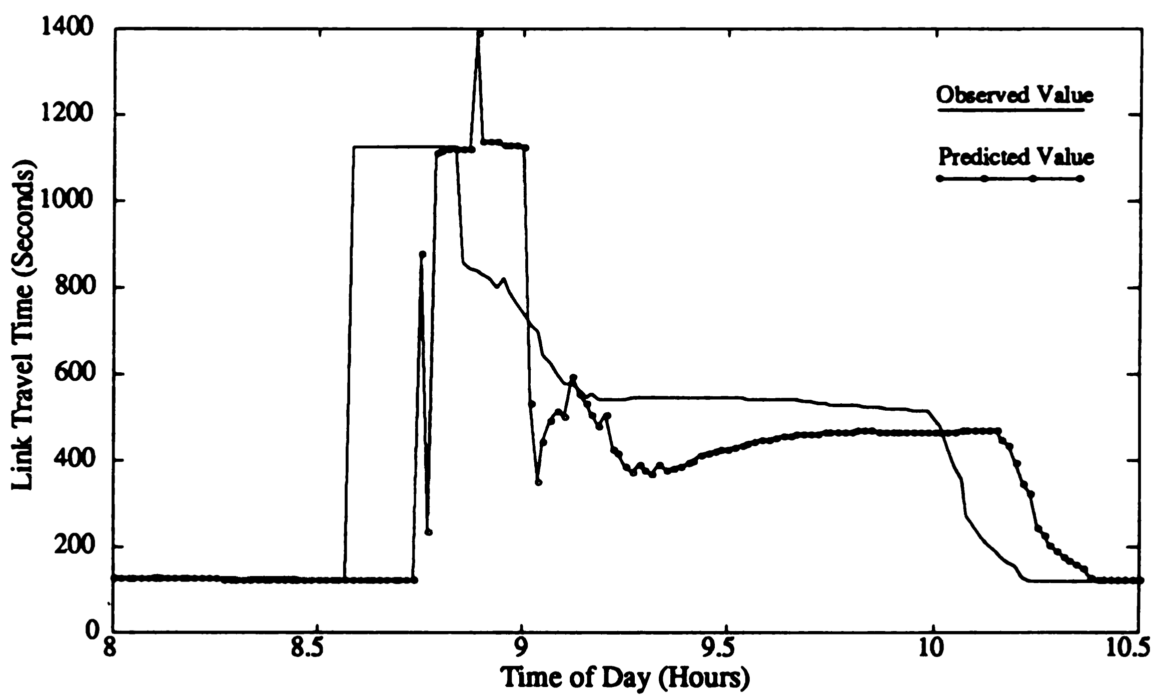


Figure 5.27: 10 Step Predictions of Travel Times of Link 19 with Model Including the Convection Term - Congested Traffic Conditions - λ=1.00

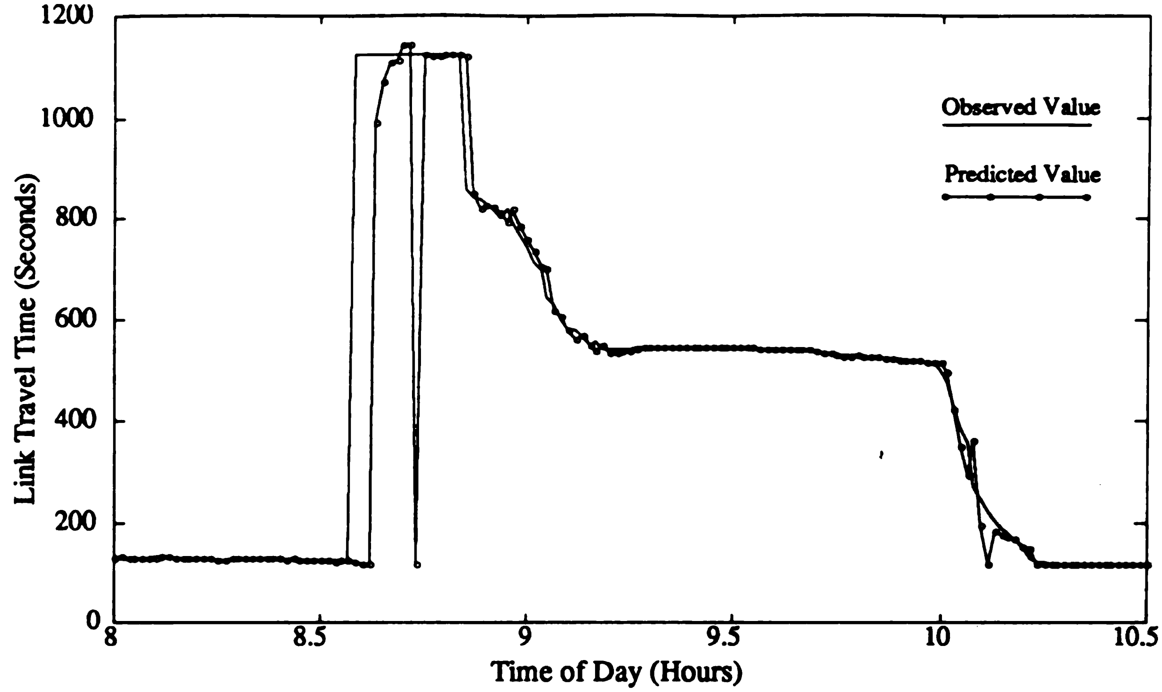


Figure 5.28: 1 Step Predictions of Travel Times of Link 19 with Model Including the Convection Term - Congested Traffic Conditions - λ=0.80

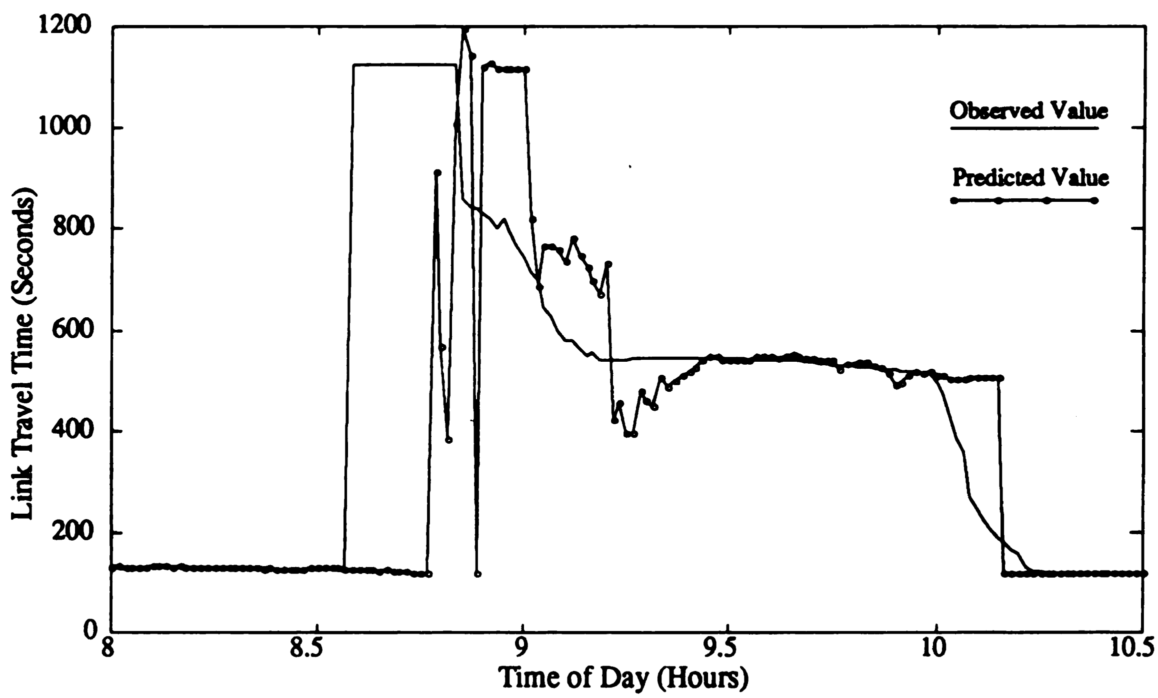


Figure 5.29: 10 Step Predictions of Travel Times of Link 19 with Model Including the Convection Term - Congested Traffic Conditions - λ=0.80

5.3.4 Models Including the Congestion Term

In the last part of this portion of the study we examine the effect on the link travel time predictions of the information from downstream links. For this reason, the *congestion term* was included in the model, which will have the form:

$$T_l(t) = \sum_{i=1}^{n} a_i(t-1) \cdot T_l(t-i) + \sum_{k \in I} \sum_{j=1}^{m_k} b_{kj}(t-1) \cdot T_k(t-j) +$$

$$+ \sum_{p \in O} \sum_{h=1}^{r_p} c_{ph}(t-1) \cdot T_p(t-h) + d \cdot \tilde{T}_l(t) + \varepsilon(t) \qquad (5.8)$$

where O is the set of links p that start at the exit node of link l. Again, due to the geometry of the network that is examined, set O will consist of one or two connector links and one link of one of the major arterials. As was discussed in paragraph 3.3.2, the congestion term is included in the model in order to capture the effect of traffic shock waves travelling in the opposite direction of the traffic. Such shock waves may be produced either due to excessive demand or due to a traffic incident.

In the following experiments, the number of variables from each link in the congestion term that will be used into the prediction model is examined. Again for reasons of simplicity, the number of variables used from the connector links r_c will be equal to the number of variables used from the link belonging to the arterial r_a. The order of the autoregressive term is kept constant and equal to three, and the diurnal term is also included in all models.

5.3.4.1 Normal Traffic Conditions

In the case of normal traffic conditions the travel time series of link 15 was considered, and set O consisted of the links {20, 21, 22}. The mean relative error, the mean square relative error and the maximum relative error for the predictions obtained by these models are shown in Table 5.23 and Table 5.24 for the morning and the evening periods respectively.

The models examined included the convection term with m_a=2 (for link 8), along with the autoregressive and the diurnal terms. In all cases the resulting errors are slightly worse than those obtained with the respective model consisting only by the autoregressive, the convection and the diurnal terms. This was expected since there was no traffic wave traveling backwards (i.e. from link 22 towards link 15). Therefore, current travel times of the link under consideration should not be correlated to the current or past values of the travel times of downstream links. Indeed the parameters associated with the congestion term, in the case of normal traffic operations are converging to zero. This is illustrated in Figure 5.30 where the parameters associated with link 22 are plotted for the second simulated day. After each peak period the parameters are decreased in a stepwise manner. This happens because the rest of the time travel times of both the link for which the predictions are performed and the downstream link are inactive and no change on their relationship can be detected. The contribution of the congestion term:

$$\sum_{p \in O} \sum_{h=1}^{r_p} c_{ph}(t-1) \cdot T_p(t-h)$$

to the one step ahead predicted travel times of link 15 is shown in Figure 5.31 where as it can be seen it converges to zero with the same stepwise manner as the corresponding parameters.

Table 5.23: Prediction Errors of Link 15 with Models Including the Congestion Term - Normal Traffic Conditions, Morning Period (300 Observations 06:00-11:00)

	r_a	1 Step Ahead Predictions			5 Step Ahead Predictions			10 Step Ahead Predictions		
		$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m
$\lambda = 1.000$	1	0.00331	0.00029	0.02550	0.00795	0.00074	0.05947	0.01168	0.00119	0.09440
	2	0.00338	0.00030	0.02489	0.00779	0.00073	0.05752	0.01148	0.00118	0.09501
	3	0.00338	0.00030	0.02487	0.00778	0.00073	0.05745	0.01151	0.00118	0.09490
	4	0.00344	0.00031	0.02425	0.00812	0.00074	0.06014	0.01203	0.00121	0.09499
	5	0.00343	0.00031	0.02365	0.00802	0.00074	0.05960	0.01184	0.001120	0.09494
$\lambda = 0.990$	1	0.00288	0.00029	0.03452	0.00930	0.00098	0.11171	0.01652	0.00200	0.23306
	2	0.00293	0.00029	0.03455	0.00931	0.00100	0.11170	0.01661	0.00201	0.23351
	3	0.00301	0.00030	0.03636	0.00963	0.00106	0.112119	0.01697	0.00205	0.23548
	4	0.00307	0.00030	0.03568	0.00964	0.00106	0.11781	0.01710	0.00210	0.24318
	5	0.00317	0.00031	0.03476	0.00983	0.00106	0.10865	0.01706	0.00206	0.23049

No Predictions	0.00442	0.00040	0.03462	0.01238	0.00129	0.11725	0.01914	0.00208	0.18957

Table 5.24: Prediction Errors of Link 15 with Models Including the Congestion Term - Normal Traffic Conditions, Evening Period (300 Observations 14:00-19:00)

	r_a	1 Step Ahead Predictions			5 Step Ahead Predictions			10 Step Ahead Predictions		
		$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m
$\lambda = 1.000$	1	0.00375	0.00031	0.02559	0.00887	0.00071	0.05670	0.01391	0.00121	0.09921
	2	0.00379	0.00031	0.02661	0.00883	0.00070	0.05616	0.01382	0.00120	0.10018
	3	0.00381	0.00031	0.02660	0.00884	0.00070	0.05622	0.01382	0.00120	0.10017
	4	0.00397	0.00033	0.02429	0.00900	0.00071	0.05295	0.01411	0.00123	0.09830
	5	0.00403	0.00033	0.02462	0.00892	0.00071	0.05430	0.01400	0.00122	0.09903
$\lambda = 0.990$	1	0.00325	0.00027	0.02162	0.00884	0.00069	0.05798	0.01601	0.00146	0.13050
	2	0.00329	0.00027	0.02399	0.00894	0.00070	0.06405	0.01610	0.00147	0.13408
	3	0.00338	0.00028	0.02232	0.00906	0.00069	0.05734	0.01604	0.00147	0.13114
	4	0.00342	0.00028	0.02144	0.00898	0.00069	0.05970	0.01603	0.00147	0.13211
	5	0.00349	0.00028	0.02152	0.00911	0.00070	0.05848	0.01608	0.00147	0.13141
No Predictions		0.00567	0.00045	0.05035	0.01553	0.00131	0.11269	0.02108	0.00189	0.17345

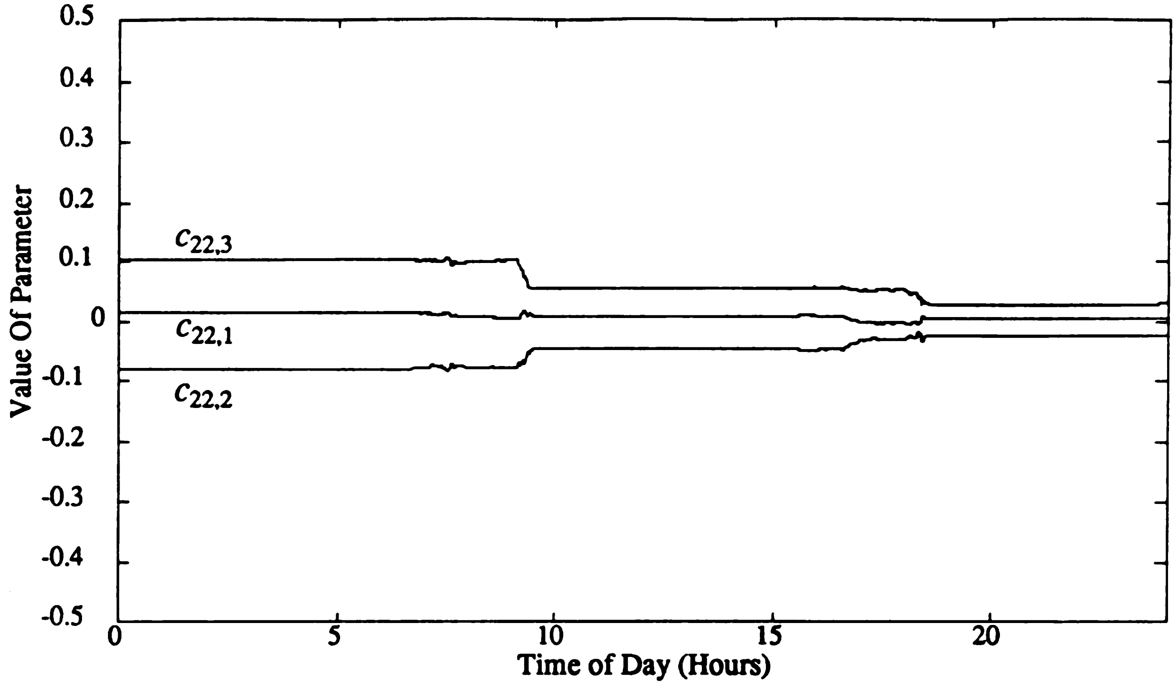


Figure 5.30: Convergence of Parameters of Congestion Term to Zero Under Normal Traffic Conditions.

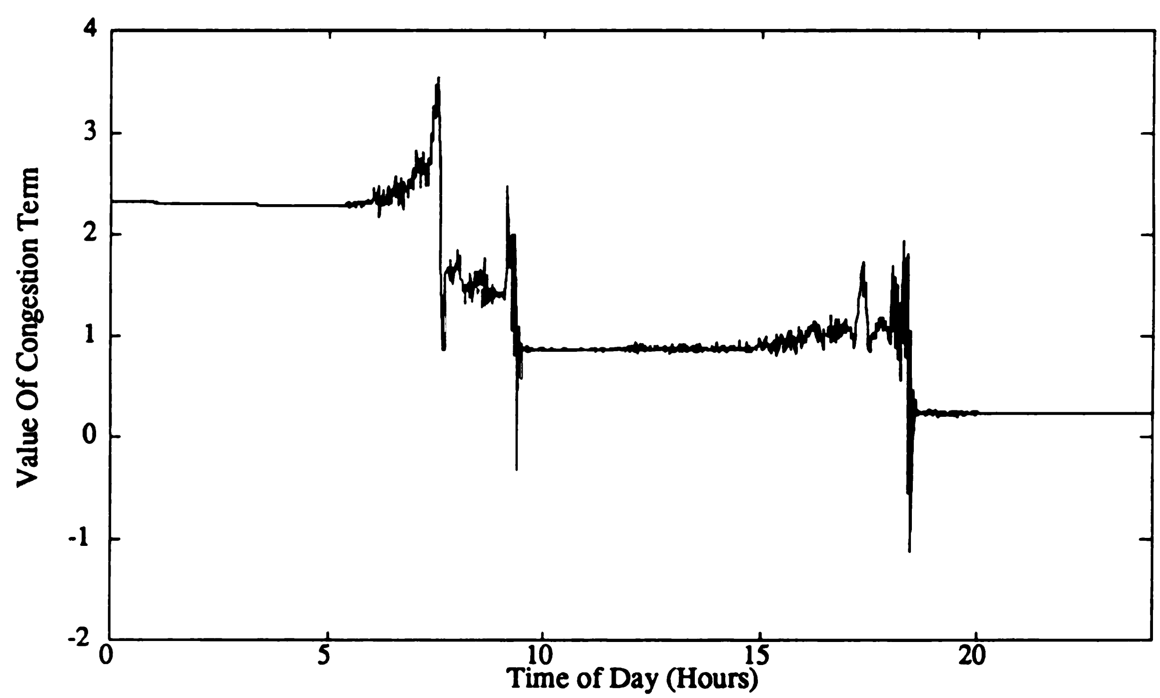


Figure 5.31: Contribution of the Congestion Term to the 1-Step Ahead Predictions Under Normal Traffic Conditions.

5.3.4.2 Congested Conditions Due to Traffic Accident

In the case of congested traffic conditions, the travel time observations of link 12 will be examined. In this case set O consists of the links {18, 19} and the way that the shock wave produced on link 19 due to the accident affects the travel times of link 12 will be investigated.

First the results obtained with a model consisting of the autoregressive component only, the one consisting of the autoregressive term and the diurnal term, and the one including the convection term are shown in Table 5.25. As is shown from these results, analogously to the case of link 19 on which the accident occurred, the simple autoregressive models produces errors similar to the no predictions case when the forgetting factor is equal to 1.00 or 0.99. For smaller values of λ, the model is exited more easily which result in large values of the maximum error. This also results in a large mean square error value although the mean average error is improved slightly. When the diurnal term is added into the model, the results are similar to the ones obtained with only the autoregressive term. When λ was set to 1.00 or 0.99 the resulting errors were worse than the no predictions case while for smaller values of λ the errors are smaller for the ten step ahead predictions but worse for the rest of the predictions. This occurs because at the beginning in the creation of the queue the model reacts too erratically to the sudden increase in the travel time, and it reaches the allowable ceiling value, and then goes back to the allowable floor value once or twice until it stabilizes close to the observed travel times. The result is a few gross errors at the beginning of the creation of the queue which deteriorate the values of the error measures. However, in the ten step predictions this "bang-bang" effect is not so obvious due to the smoothing effect of the average value. The same is also true for the predictions

Table 5.25: Prediction Errors of Link 12 During Traffic Accident on Link 19 (150 Observations 8:00-10:00)

	Model	1 Step Ahead Predictions			5 Step Ahead Predictions			10 Step Ahead Predictions		
		$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m
λ = 1.00	AR(2)	0.02058	0.00541	0.70926	0.09844	0.01559	0.74261	0.20320	0.02806	1.2001
	AR(2) & Avg Travel Time	0.02369	0.00593	0.70926	0.10820	0.01611	0.74390	0.22438	0.02905	1.14618
	AR(2), AvgTT & Conv.	0.02795	0.00793	0.74164	0.11198	0.01689	0.74601	0.22683	0.02946	1.14534
λ = 0.99	AR(2)	0.02593	0.007311	0.74165	0.10710	0.01609	0.74601	0.22242	0.03057	2.04648
	AR(2) & Avg Travel Time	0.03495	0.01056	1.12805	0.15218	0.03176	2.89060	0.25931	0.03546	2.02188
	AR(2), AvgTT & Conv.	0.03845	0.01239	1.30287	0.13453	0.02606	2.89060	0.25231	0.03524	2.02188
	No Predictions	0.02290	0.00558	0.70889	0.10139	0.01597	0.74174	0.20460	0.02872	1.25073

Table 5.25 (continued)

	Model	1 Step Ahead Predictions			5 Step Ahead Predictions			10 Step Ahead Predictions		
		$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m
$\lambda = 0.90$	AR(2)	0.02531	0.00734	0.74165	0.09614	0.01785	1.49473	0.18433	0.03011	2.04648
	AR(2) & Avg Travel Time	0.04681	0.02080	2.78991	0.10388	0.01518	2.89060	0.15428	0.02705	2.02188
	AR(2), AvgTT & Conv.	0.04579	0.02052	2.78991	0.09885	0.02416	2.89060	0.14229	0.02387	1.96884
$\lambda = 0.80$	AR(2)	0.02707	0.00775	0.74165	0.10815	0.02379	2.77755	0.17957	0.02996	2.04647
	AR(2) & Avg Travel Time	0.06066	0.02765	2.87527	0.10179	0.01554	2.89060	0.14636	0.02421	1.96884
	AR(2), AvgTT & Conv.	0.04767	0.02056	2.78991	0.10115	0.01539	2.89060	0.14814	0.02417	2.04647

No Predictions	0.02290	0.00558	0.70889	0.10139	0.01597	0.74174	0.20460	0.02872	1.25073

obtained with the model including the convection term.

The results of the prediction models including the congestion term are shown in Table 5.26. When the forgetting factor is set at 1.00 or 0.99 the resulting errors for all the models that were tested were greater than the errors resulting from the no predictions situation. However, the errors of the model with r_a=10 gave the smallest errors for both λ=1.00 and λ=0.99 and for all the predictions, the one, five and ten step ahead. In the case when λ is set to the smaller values of 0.90 and 0.80 the prediction errors of all models was worse except the model with r_a=10, for which the error is smaller. In the case of λ=0.90 the resulting error from this model was smaller than that obtained without any predictions performed or those obtained by any other model.

This was expected because there is a delay of approximately 8 time steps (8 minutes) between the time the incident occurs on link 19 and the start of the queue on link 12. Therefore there is a time lag before travel time on link 19 starts affecting the travel time of link 12. This time lag represents the time required for the shock wave to be transferred from link 19 to link 12. Thus models with r_a smaller than 8, i.e. 1,2 and 5, fail to capture the relationship between the travel times of links 19 and 12, and the contribution of the congestion term to the predictions is relatively small most of the time (Figure 5.32). On the other hand, when r_a is equal to 10 the contribution of the congestion term to the prediction of the travel time of link 12 is more significant, as is depicted in Figure 5.33. When r_a=10 the total contribution of the congestion term gains not when the accident starts, but when the queue on link 12 starts building up, and drops to a lower value at the end of the incident and decays as the queue generated by the incident is dissolved.

Table 5.26: Prediction Errors of Link 12 with Models Including the Congestion Term - Congested Traffic Conditions (150 Observations 8:00-10:00)

	AR	1 Step Ahead Predictions			5 Step Ahead Predictions			10 Step Ahead Predictions		
	order	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m
$\lambda = 1.00$	1	0.03063	0.00681	0.70865	0.13771	0.02081	0.83003	0.26035	0.03672	1.54726
	2	0.03334	0.00701	0.70870	0.14524	0.021714	0.88632	0.27017	0.038149	1.59704
	5	0.03878	0.00712	0.70943	0.16411	0.02395	1.01303	0.30004	0.04099	1.67073
	10	0.02912	0.00636	0.71421	0.13329	0.01945	0.74733	0.24817	0.03379	1.46591
$\lambda = 0.99$	1	0.03451	0.00883	0.74237	0.14861	0.02807	2.89060	0.26853	0.04031	2.04647
	2	0.03726	0.00941	0.74237	0.15177	0.02713	2.57550	0.267076	0.03929	1.96884
	5	0.03825	0.00855	0.74237	0.16482	0.02757	2.39546	0.28665	0.04007	1.68084
	10	0.02779	0.00629	0.71421	0.12053	0.01767	0.74734	0.23531	0.03249	1.46776

No Predictions	0.02290	0.00558	0.70889	0.10139	0.01597	0.74174	0.20460	0.02872	1.25073

Table 5.26 (continued)

	r_a	1 Step Ahead Predictions			5 Step Ahead Predictions			10 Step Ahead Predictions		
		$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m
$\lambda = 0.90$	1	0.04244	0.01514	1.85706	0.13174	0.02654	2.89061	0.22145	0.03807	2.63712
	2	0.05310	0.01891	1.94214	0.11989	0.02542	2.89060	0.16516	0.02617	1.96884
	5	0.03371	0.01004	0.74733	0.09394	0.01559	0.74816	0.13719	0.02026	0.79854
	10	0.02231	0.00548	0.70421	0.09296	0.01417	0.73727	0.13447	0.01988	0.79751
$\lambda = 0.80$	1	0.03959	0.01036	0.74733	0.15498	0.04587	6.27328	0.28824	0.09954	13.97638
	2	0.06354	0.02658	3.67137	0.12868	0.02330	1.86674	0.19900	0.03674	3.19307
	5	0.03816	0.01026	0.74733	0.16039	0.04930	5.64326	0.24240	0.07855	8.98686
	10	0.02442	0.00586	0.70622	0.11294	0.01905	1.12132	0.15207	0.02512	2.20378

No Predictions	0.02290	0.00558	0.70889	0.10139	0.01597	0.74174	0.20460	0.02872	1.25073

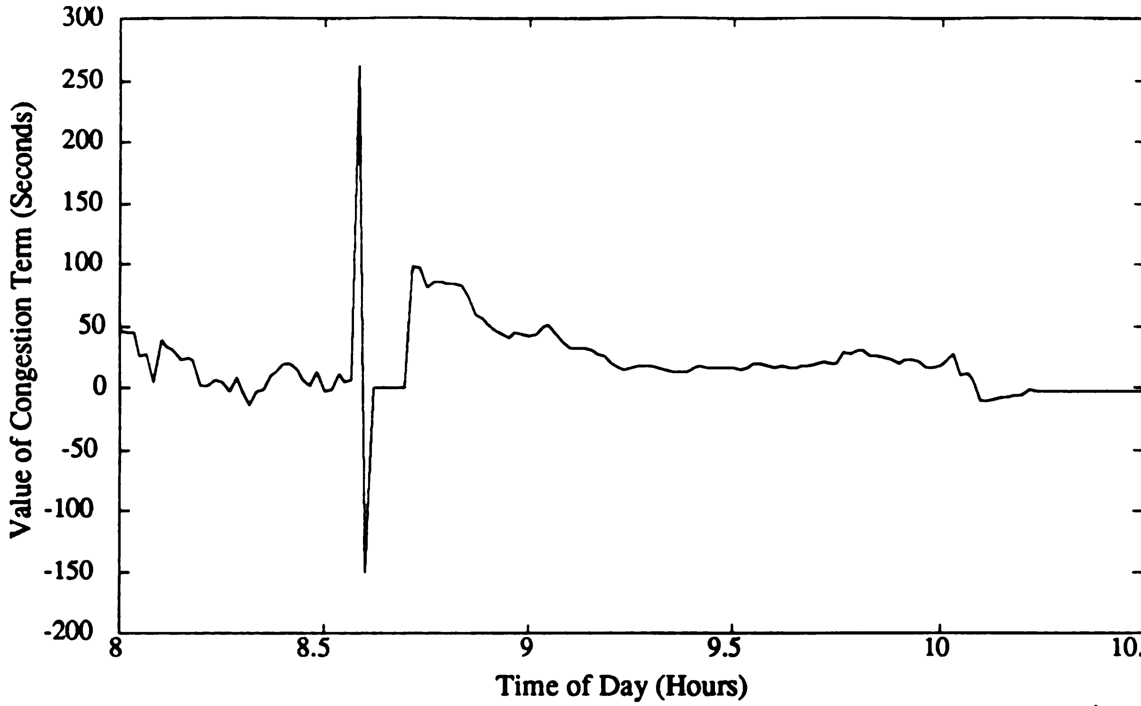


Figure 5.32: Value of the Congestion Term from Link 19 for the 1 Step Travel Time Predictions of Link 12 - r_a=2

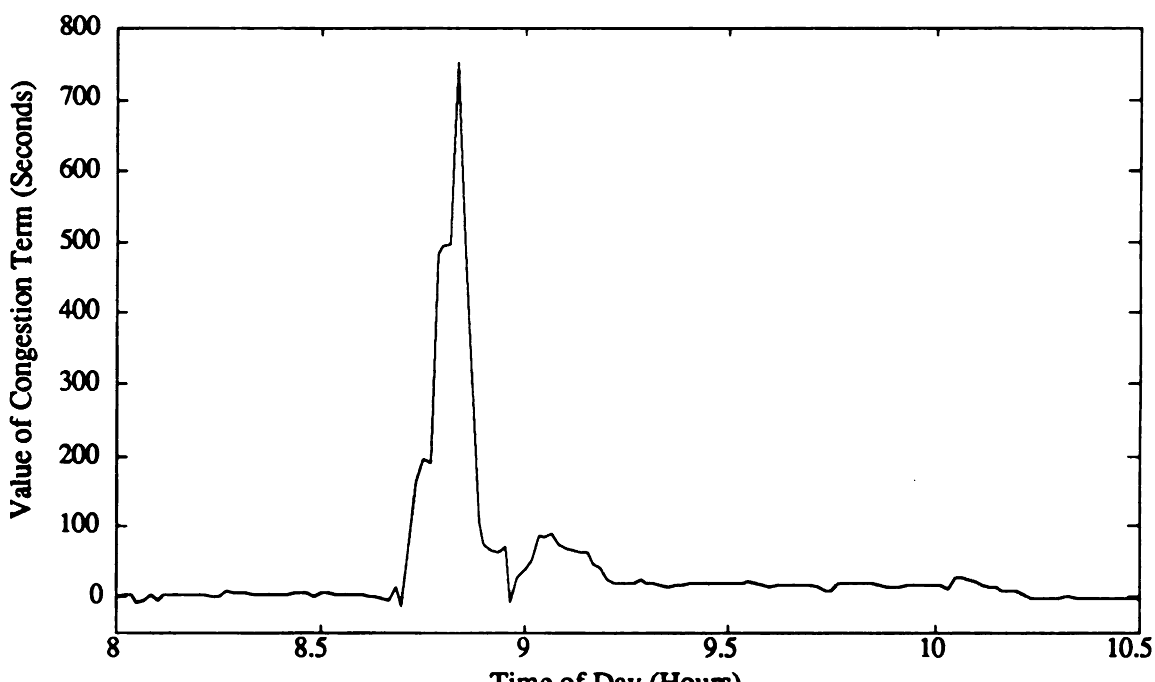


Figure 5.33: Value of the Congestion Term from Link 19 for the 1 Step Travel Time Predictions of Link 12 - r_a=10

The one and ten step ahead predictions obtained with the model including the congestion term with r_a=10 and for λ=1.00 are shown in Figure 5.34 and Figure 5.35. As can be seen in these figures, the effect of the long travel time of link 19 due to the queue that is formed on it is carried for a long time on the ten step ahead predictions, thus increasing the error. On the other hand when λ=0.90 (Figure 5.36 and Figure 5.37) the predicted values are not affected too much by the queue on link 19, and remain closer to the observed travel times.

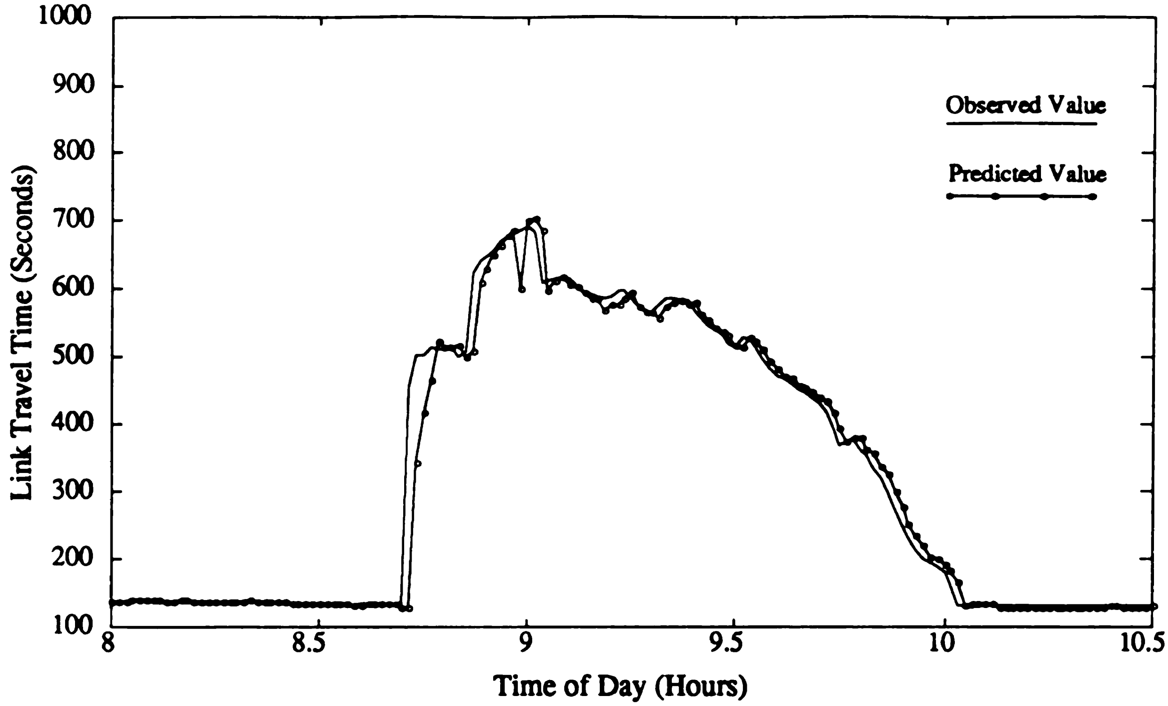


Figure 5.34: 1 Step Predictions of Travel Times of Link 12 with Models Including the Congestion Term - r_a=10, λ=1.00

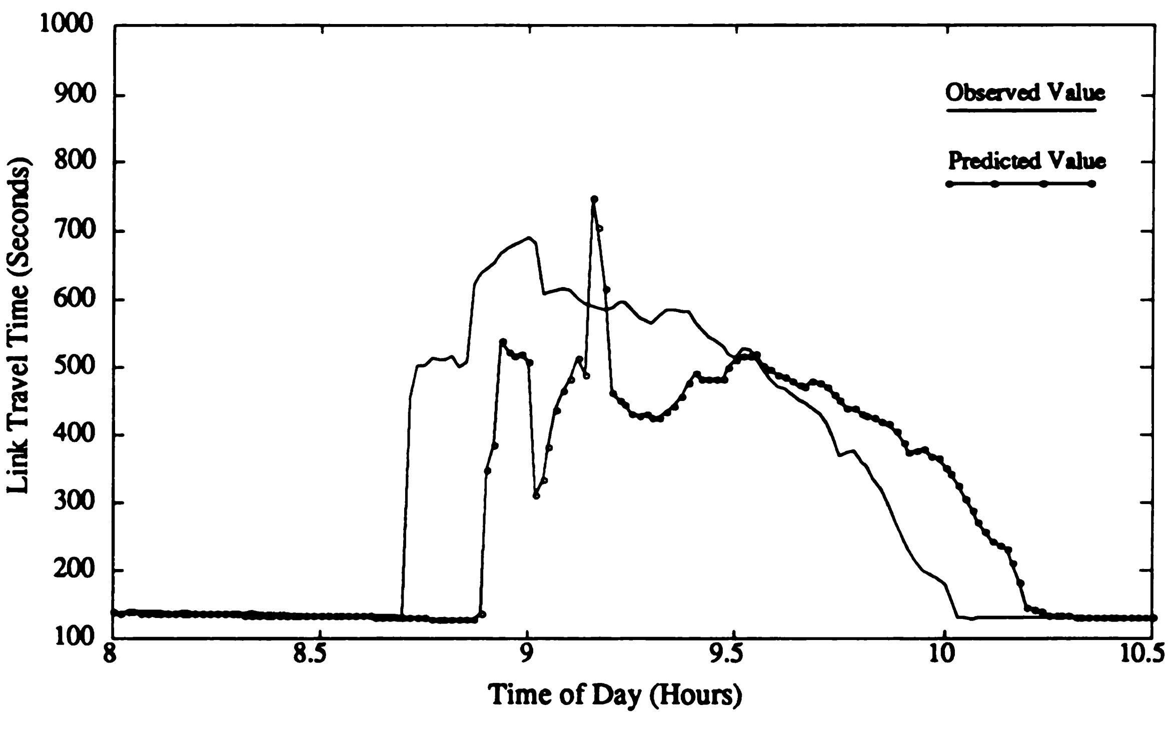


Figure 5.35: 10 Step Predictions of Travel Times of Link 12 with Models Including the Congestion Term - r_a=10, λ=1.00

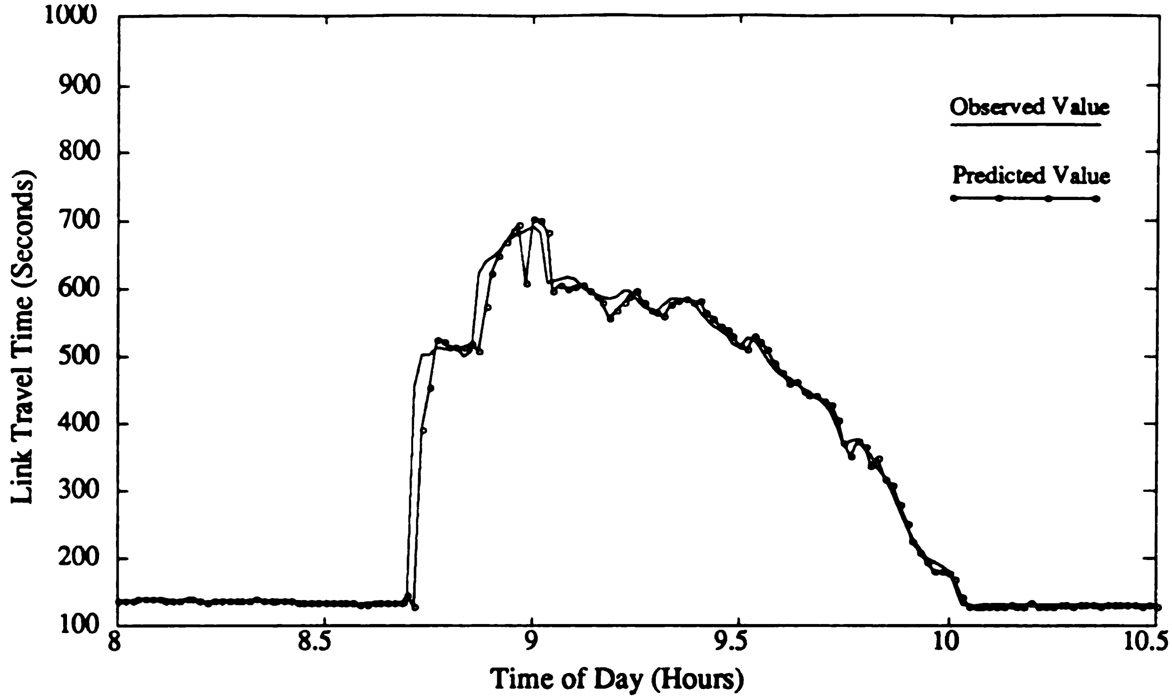


Figure 5.36: 1 Step Predictions of Travel Times of Link 12 with Models Including the Congestion Term - r_a=10, λ=0.90

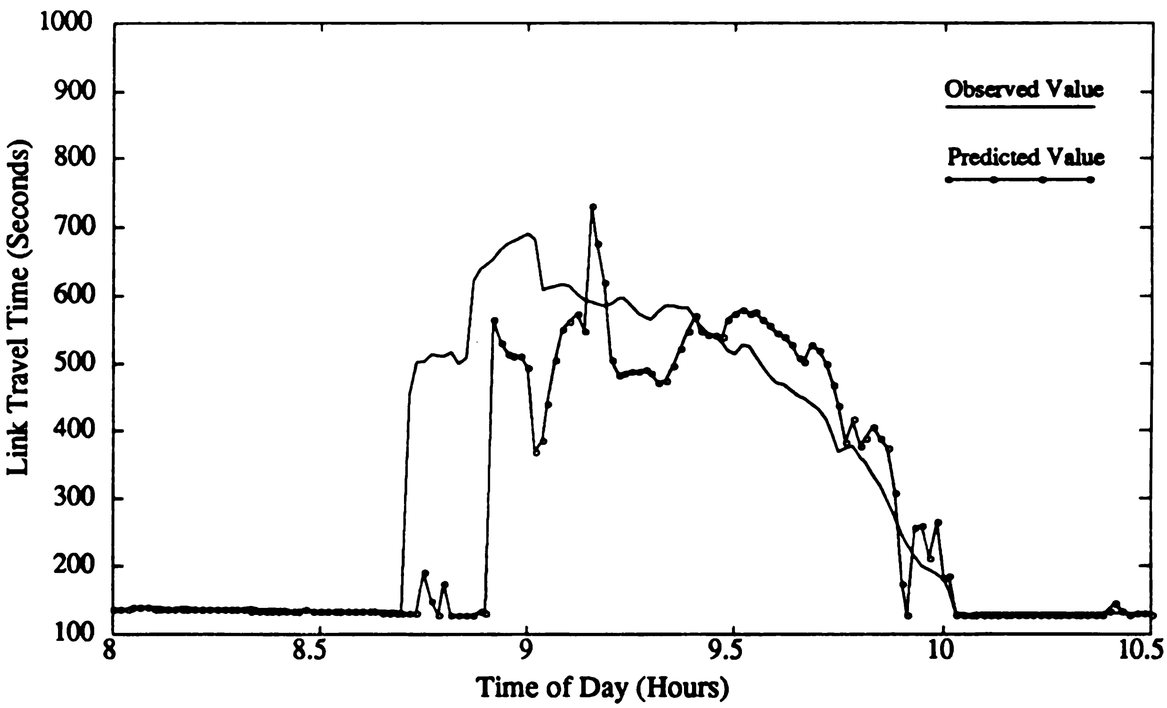


Figure 5.37: 10 Step Predictions of Travel Times of Link 12 with Models Including the Congestion Term - r_a=10, λ=0.90

5.3.5 Discussion of the Results and Further Developments

From the previous analysis it can be concluded that the models including the autoregressive, the diurnal and the convection terms are the most appropriate for the prediction of the link travel times. Models including the convection term did not give acceptable results. A probable reason for this is due to the low influence for long time periods of the input data from the downstream link on the travel times of the link under consideration. For the same reason model that included information from connector links in the convection term produced larger errors when a small change occured on the travel time of the connector link. Therefore, in the following analysis the models that will be used will include the autoregressive, the diurnal and the convection terms with order m_c=0 and m_a depending on the length of the upstream link.

It is obvious that the prediction model works best under normal traffic operation with the forgetting factor set at one. The fact that the results are better in the case of the normal operation than in the case of the traffic accident are not a surprise, since when the accident occurs there is a strong shock to the system and the RLS algorithm operates efficiently only while the parameters of the model are not changing too fast.

Because the environment in which the filter operates is not stationary, even during normal operations - since the stochastic process of the link travel times which is followed by the filter has variations which are time dependent - it was expected that the model would give better results when the forgetting factor, λ was set to values less than one. When such values are assigned to the forgetting factor, the adaptation of the model to the prevailing traffic conditions is fast, but at the expense of an increase in the mean relative error. This is obviously a trade off that has to be made between the tracking ability of the model and its sensitivity to the prediction errors. When the forgetting factor is set to values smaller but very close or equal to 1.00, the result is smaller errors due to the large amount of data that are used for the computation of the parameter vector, since old and recent prediction errors contribute equally to the loss function (3.21).

An important feature of the prediction model that includes the diurnal term, can be used for the justification of the usage of a large value for the forgetting factor under normal traffic condition. The inclusion of the average travel time into the model has as an effect that the link travel time process approximates a stationary one. Of course, the difference $[T_l(t) - \tilde{T}_l(t)]$ is not really stationary, since during non peak periods the value approaches zero since both variables $T_l(t)$ and $\tilde{T}_l(t)$ are very close to the free flow travel time, while during peak periods where there is much higher variation in $T_l(t)$ the difference $[T_l(t) - \tilde{T}_l(t)]$ oscillates around zero. Note that this would be true even if we would apply differencing of one day on $T_l(t)$, i.e. if the process that was modeled was

$$z(t) = T_l^{(d)}(t) - T_l^{(d-1)}(t) \tag{5.9}$$

where $T_l^{(d)}(t)$ is the observation at time t of the day d. Without attempting to prove this here, we can assume that for given time periods, smaller than a peak period, the above difference becomes asymptotically stationary. For this reason and when the diurnal term is included in the model it is suitable to use a forgetting factor λ=1.00.

However, when traffic conditions suddenly change from normal to congested, then the model needs to employ its tracking characteristic. The results obtained with small values of λ such as 0.90 and 0.80 under such conditions demonstrate this need. Of course, such small values for the forgetting factor would have an adverse effect on the prediction errors for the periods when the system is not excited as much due to the increased sensitivity to even small errors. Furthermore, the best results in the case of congested conditions due to the traffic accident were obtained with a model of different structure than the one which gave the best results for the normal operations situation. Since the prediction model will have to be implemented so it will operate automatically, a trade off will have to be made, so the model will operate satisfactory while either normal or congested conditions are in effect.

Therefore, we need to device a mechanism that would employ small values to the forgetting factor only when appropriate, while for the rest of the time the model should operate in a close to stationary environment. Until now we have assumed a constant value for λ. However, there is nothing to prohibit a time varying value. Such a time varying function for λ, $\lambda(t)$ should assume small values when there is a strong excitation to the system, and return to values close to one when conditions are reinstated to normal. An appropriate trigger to induce the reduction of the value of $\lambda(t)$ would be the prediction error $\hat{\varepsilon}(t)$, defined in equation (3.12). Such a function could be defined as:

$$\lambda(t) = \begin{cases} \lambda_0 & \text{if} \quad |\hat{\varepsilon}(t)| \geq \xi, \lambda(t-1) < \lambda \\ \lambda_0 + \sum_{i=1}^{t-s} \frac{1-\lambda(t-1)}{i} & \text{if} \quad t > s, \lambda(t-1) < \lambda \\ \lambda & \text{otherwise} \end{cases} \tag{5.10}$$

where s is the time at which a large prediction error is observed, ξ is the prediction error threshold, λ_0 is a small initial value for the forgetting factor such as 0.80. The function defined in (5.10) allows the usage of a general value for the forgetting factor ($\lambda \leq 1.00$),while when prediction errors occur that exceed ξ, $\lambda(t)$ is set to a small initial value λ_0, and afterwards it approaches 1.00 asymptotically, until it reaches the value of the forgetting factor, λ. With proper selection of the threshold ξ, equation (5.10) will not affect the predictions performed by the RLS algorithm in the case of normal traffic conditions. Such a value is selected to be double of free flow travel time of the link, $(\xi = 2 \times FFTT_l)$. In the case of the congested traffic conditions the modified function for $\lambda(t)$ was examined on link 19.

The model that was used consisted of the autoregressive term with n=3, the diurnal term and the convection term with m_a=2, since this is the model that gave the best results for the normal traffic conditions. The initial value λ_0 was set at 0.80. Smaller values were examined but they gave worse results. The error measures obtained with this prediction model and with the forgetting factor as defined in equation (5.10) are shown in Table 5.27, while in Figure 5.38 and Figure 5.39 the one and the ten step ahead predictions obtained with the same model are shown.

In the same table the percent differences of these errors from the no predictions case, and the errors obtained from the same model in the case that the forgetting factor, is constant and equal to 1.00 and 0.80 are shown. The errors of the one step predictions of the model with the varying forgetting factor are still worse than the no predictions, but better than those obtained by the model with λ=0.80. On the other hand, five and ten step ahead predictions are of better quality than the no predictions case and the case with λ=1.00. The

Table 5.27: Prediction Errors of Model with Time Varying Forgetting Factor for Link 19 - Congested Traffic Conditions Due to Accident - 150 Observations 08:00-10:30

Model: AR(2), Avg. TT & CV(3)	1 Step Ahead Predictions			5 Step Ahead Predictions			10 Step Ahead Predictions		
	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m	$\bar{e}$	$\bar{e}_s$	e_m
$\lambda(t)$	0.03101	0.00959	0.89091	0.10443	0.01897	0.89078	0.21440	0.03393	1.66004
No Prediction	-29.0%	-34.1%	0.0%	16.1%	14.9%	7.2%	21.9%	22.2%	16.5%
λ=1.00	-5.7%	-29.1%	0.0%	26.8%	13.8%	-0.1%	27.7%	15.8%	4.3%
λ=0.80	11.3%	12.8%	0.3%	2.6%	3.1%	0.3%	-12.8%	-11.1%	1.6%

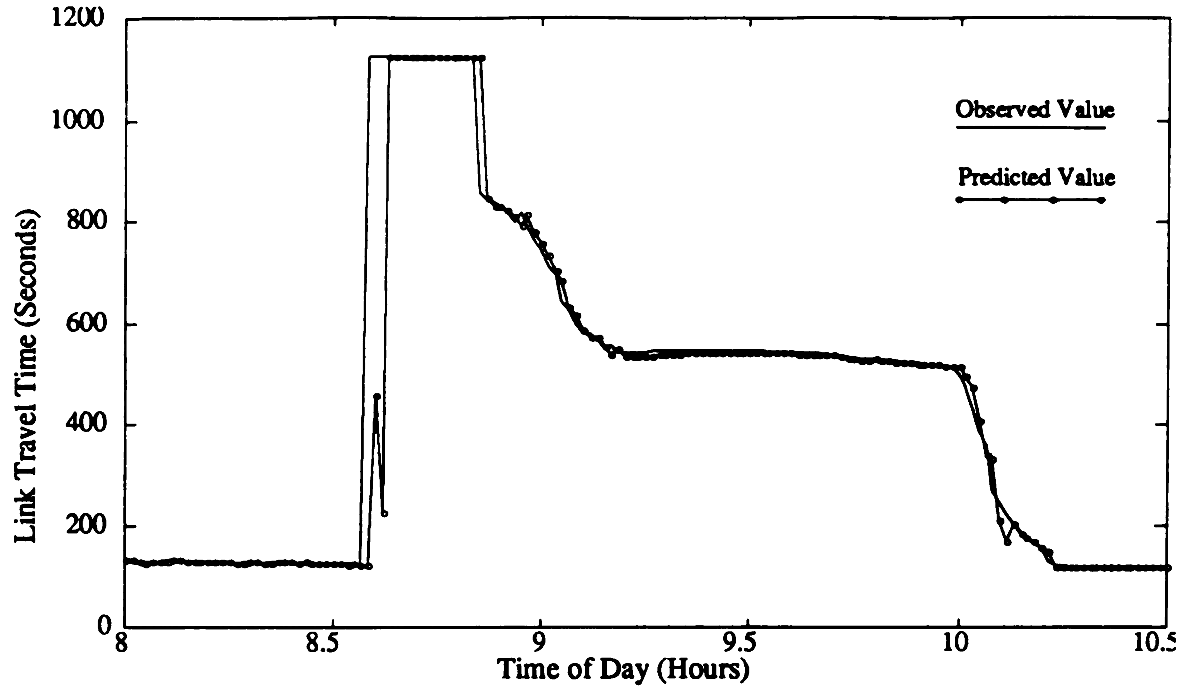


Figure 5.38: 1 Step Ahead Predictions of Travel Time of Link 19 with Time Varying Forgetting Factor

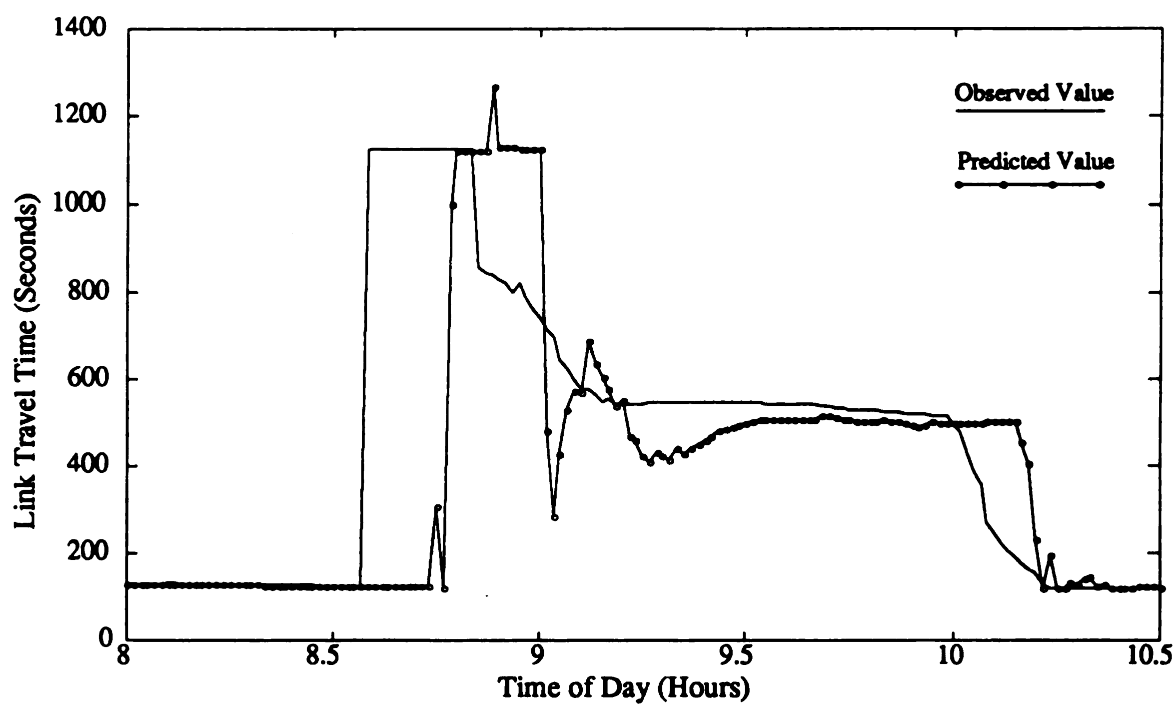


Figure 5.39: 10 Step Ahead Predictions of Travel Time of Link 19 with Time Varying Forgetting Factor

mean square error for the five step predictions is 14.9% and 13.8% smaller than same errors obtained from the no predictions case and the case with $\lambda=1.00$. Compared to the model with $\lambda=0.80$ five step predictions are almost the same (slightly better with the mean square error just 3.1% lower) while ten step predictions are worse by as much as 12.8% for the mean relative error. It should be noted though, that the most crucial predictions are the one step ahead, since errors to further predictions will affect vehicles which are not at the entrance of the congested link, and thus, such vehicles may still have the opportunity to divert to alternative paths.

5.4 Effect of Time Step Length on Predictions

In the following we will examine the effect of the length of the time interval between sampling the link travel times of the network. To that end, the models that suggested good performance under normal and/or congested traffic conditions will be utilized for examining the different time intervals. From the previous paragraph we have seen that the RLS algorithm works efficiently when traffic conditions are not disturbed by incidents which produce large and abrupt anomalies to the system. In the situation where the system is disturbed by a traffic accident, the algorithm gives acceptable results when the forgetting factor is set to small values, so the large errors which occur at the start and at the end of the incident will not affect the travel time predictions of the links for a long time.

Four different sampling intervals will be examined: sampling every 1, 2, 5, and 10 minutes. For evaluating the effect of the length of the time interval, the simulation program was utilized in four separate runs where the time step between updates of the link travel times and the evaluation of the minimum paths in each run was set respectively to 1, 2, 5 and 10 minutes. Of course, predictions of s time steps ahead are not comparable among the different runs, since s represents a different length of time in each run. For this reason, and in addition to the average travel time of the link in every time interval, the one minute average travel time of each link was recorded so the results of the prediction model can be compared with the "real" travel time which will be approximated by the one minute discrete observations. The performance of the prediction model under the different time intervals h, will be based on the error of the estimation of the travel time of the link under consideration k minutes into the future where k=1,2,...15, which is defined as:

$$e_k(t) = T_1(t+k) - \hat{T}_h(s,r) \tag{5.11}$$

where $T_1(t+k)$ is the one minute sample of the travel time of the link at time $(t + k)$ and $\hat{T}_h(s, r)$ is the r-step ahead prediction of the travel time of the link, obtained at time $(s{\cdot}h)$, and with a model utilizing a sampling time step h, where s is given as $s = \left[\frac{t}{h}\right]$, and r is given as $r = \left[\frac{k}{h}\right]$, where the brackets denote the integer part of the fraction. It must be brought to the attention that the results of each different run cannot be compared with the results obtained by the run with the one minute time interval, because the routing of the traffic is different due to the more sparse updates of the traffic conditions on the network, as the time step h increases.

Based on the errors defined by (5.11), and parallel to the measures defined in paragraph 5.3, the measures of comparison among the models with different time steps will be:

(1) the mean absolute error given as:

$$\bar{e}^{(k)} = \frac{1}{N} \cdot \sum_t |e_k(t)| \tag{5.12}$$

(2) the mean square error:

$$\bar{e}_s^{(k)} = \frac{1}{N} \cdot \sqrt{\sum_t [e_k(t)]^2} \tag{5.13}$$

and

(3) the maximum error:

$$e_m^{(k)} = \max_t \{|e_k(t)|\} \tag{5.14}$$

where N is the number of one minute observations in the period that the above errors are evaluated. The effect of the different time intervals will be tested under the two different traffic conditions: the normal traffic operations and the congested ones due to the traffic accident. All simulation runs were made with 10% smart vehicles in the network and the simulation program was run for two simulated days.

5.4.1 Effect of Time Interval under Normal Traffic Conditions

The effect of the length of the time between sampling the travel time of the link on the predictions performed with the RLS algorithm was tested with the results from link 15. The model that was used in all cases consists of the autoregressive term with n=3, the diurnal term and the convection term with m_a depending on the length of the upstream link and m_c=0, and with the forgetting factor set to 1.00. The errors defined by equations (5.12) to (5.14) are calculated for the time period starting at 06:00 and ending at 19:00 of the second simulated day.

The mean absolute errors and the mean square errors calculated for this time period are shown in Figure 5.40 and Figure 5.41 respectively. Also their values for k minutes into the future with k=1,2,...,15 are shown in Table 5.28. As can be seen from Figure 5.40 and Figure 5.41 the algorithm gives consistently better results when sampling on the travel times of the link is most frequent. For the predictions of the travel times for one to four minutes ahead the shorter the time interval the better they appear to be in terms of the mean square error. Only in the case where the time step h was five minutes, predictions of three minutes or further ahead appear to be worse than those obtained in the case of h=10 minutes. Nevertheless, predictions for one and two minutes ahead get consistently worse as h increases.

The quality of the predictions obtained with small time steps deteriorates faster than those obtained with large time steps, at least for the first 8 to 10 minutes, while after this point the quality of the predictions is tempered. This is due to the increasing effect of the diurnal term on predictions k time steps

Table 5.28: Prediction Errors of Link 15 for Different Time Steps h

Time (min)	h=1min			h=2min			h=5min			h=10min		
(k)	$\bar{e}^{(k)}$	$\bar{e}^{(k)}_s$	$e^{(k)}_m$	$\bar{e}^{(k)}$	$\bar{e}^{(k)}_s$	$e^{(k)}_m$	$\bar{e}^{(k)}$	$\bar{e}^{(k)}_s$	$e^{(k)}_m$	$\bar{e}^{(k)}$	$\bar{e}^{(k)}_s$	$e^{(k)}_m$
1	0.496	0.029	7.800	0.767	0.053	22.775	1.076	0.070	20.054	1.342	0.079	14.789
2	0.619	0.038	14.038	1.038	0.067	22.775	1.237	0.078	22.986	1.340	0.078	14.789
3	0.748	0.045	17.441	0.975	0.067	27.231	1.385	0.086	22.986	1.369	0.080	14.789
4	0.905	0.054	17.314	1.236	0.077	27.231	1.523	0.093	22.986	1.392	0.081	14.789
5	1.187	0.067	17.454	1.471	0.089	27.348	1.635	0.100	22.986	1.427	0.083	14.789
6	1.406	0.080	17.608	1.587	0.095	27.348	1.652	0.099	23.089	1.457	0.084	14.789
7	1.513	0.086	17.962	1.736	0.104	27.266	1.665	0.101	23.397	1.500	0.086	14.789
8	1.598	0.092	17.568	1.737	0.106	27.266	1.699	0.104	23.397	1.559	0.089	16.032
9	1.639	0.097	17.536	1.802	0.112	27.707	1.722	0.106	23.397	1.615	0.092	16.032
10	1.642	0.099	17.892	1.765	0.112	27.707	1.777	0.109	23.397	1.701	0.097	16.032
11	1.673	0.102	17.786	1.862	0.118	27.431	1.871	0.123	24.757	2.069	0.122	23.718
12	1.696	0.104	17.414	1.816	0.116	27.431	1.816	0.120	24.757	2.006	0.120	23.718
13	1.709	0.105	17.685	1.918	0.121	27.139	1.769	0.117	23.314	1.954	0.118	23.718
14	1.723	0.106	18.180	1.863	0.118	27.139	1.718	0.114	23.314	1.898	0.116	23.718
15	1.721	0.107	18.584	1.924	0.122	27.516	1.694	0.112	23.314	1.865	0.114	23.718

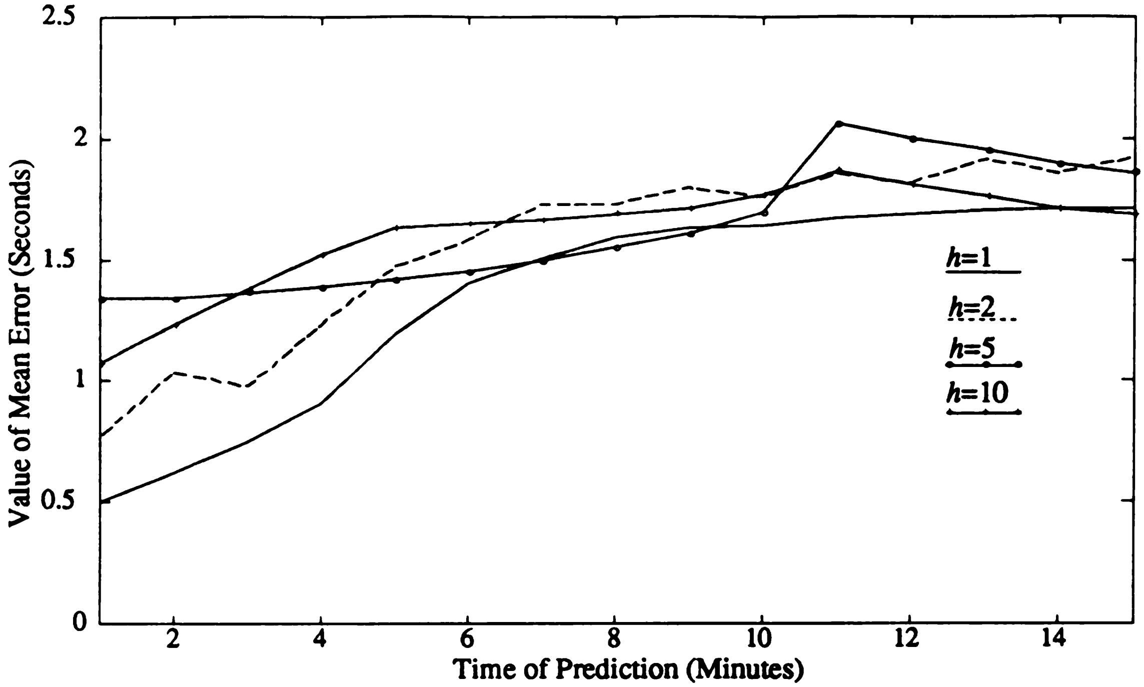


Figure 5.40: Behavior of Mean Absolute Error for Different Time Steps under Normal Traffic Conditions

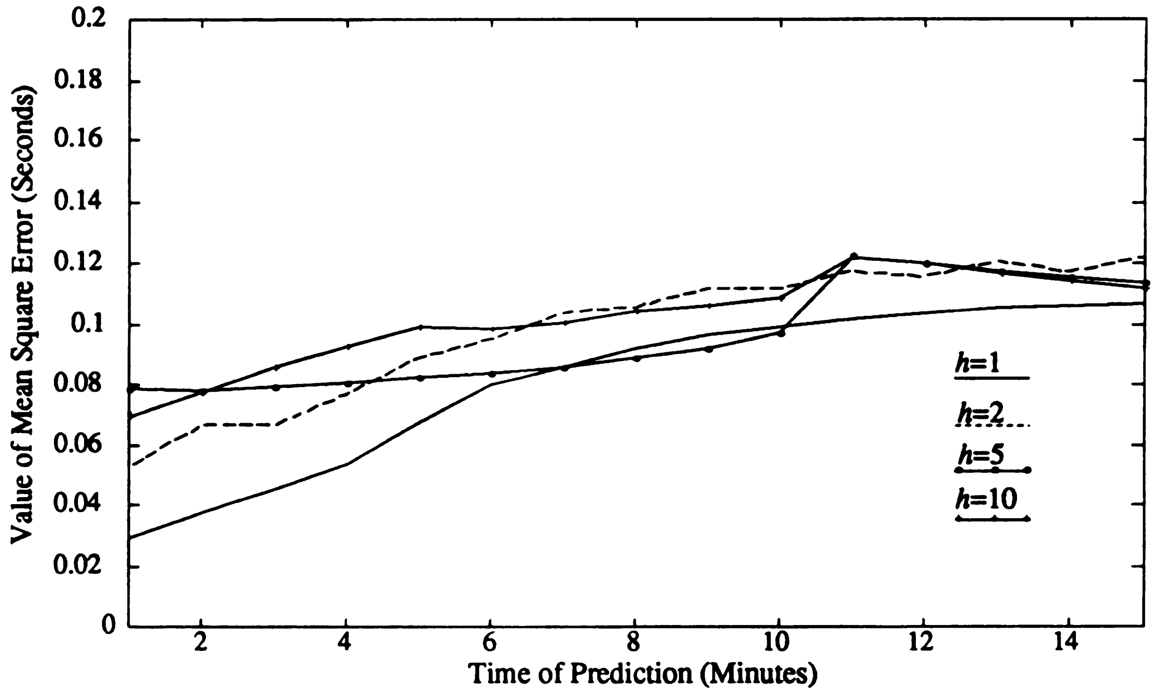


Figure 5.41: Behavior of Mean Square Error for Different Time Steps under Normal Traffic Conditions

ahead as k increases, as was explained previously. On the other hand, the mean error of the predictions obtained in the case of h=10 minutes is almost constant for the first ten minutes, and increases after this point. This consistency in the quality of the predictions when large time steps are used is due to the fact that such predictions are obtained with one or two time steps ahead predictions, while for the case of h=1 this would require ten successive predictions, with the obvious results of the propagation of the errors.

The errors produced in the case of h=10 and h=5 are remarkably small, especially for predictions around the ten minutes ahead (one or two step ahead predictions). In fact, predictions for eight to ten minutes ahead obtained with h=5 appear to be slightly better than those obtained in the case of h=1, while they are clearly better than the case of h=2. In addition, predictions beyond ten minutes appear to have some reduction in their error, instead of an increase. Figure 5.42 illustrates the quality of the two step ahead predictions obtained in the case of h=10, and as it can be seen, predicted values are close to the one minute observations.

This behavior is understood when we examine the parameters of the model when such large time steps are utilized. The total gains of each term in the case of h=10 are shown in Figure 5.43. The importance of the autoregressive term is minimized, while the importance of the other two components of the prediction model is magnified. This is especially true for the diurnal term which throughout the simulated day retains a value close to 0.5, and at the end of the simulated day assumes a value of 0.4981, thus affecting greatly the one and even more the two step ahead (10 and 20 minutes ahead) predictions. This occurs due to the limited sampling of the travel time of the link itself

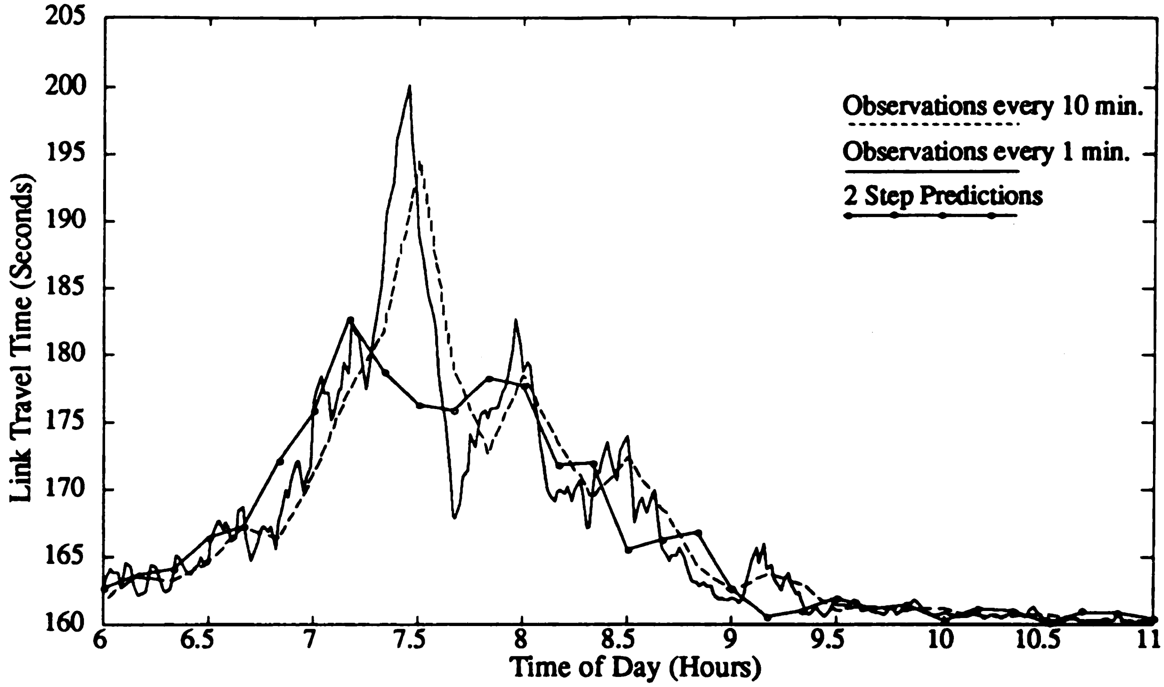


Figure 5.42: 2 Step Predictions of Travel Times of Link 15 with Time Step *h*=10 minutes

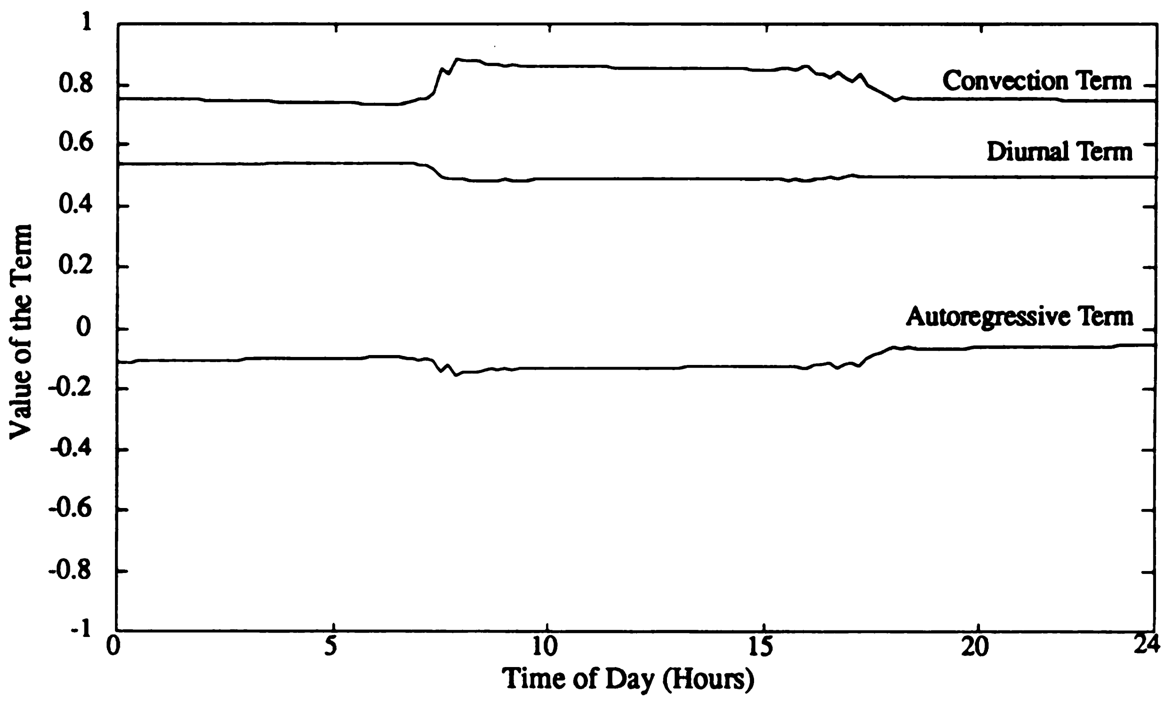


Figure 5.43: Value of Components of Prediction Model for Link 15 with Time Step *h*=10 minutes

which changes values more rapidly than in the case of h=1, and thus the RLS algorithm cannot efficiently track the states of the autoregressive component. Instead, it identifies the relationship of the link travel time with the diurnal term, and therefore, it attributes large values to parameter d. The same general trends of the mean error are observed for the mean square error as well, indicating that many of the prediction errors are smaller than one second.

Overall, from the above results it can be concluded that short term predictions in the time range of one to five minutes ahead are of very good quality when they are obtained with smaller time steps. For predictions further into the future, all time steps that were examined appear to give comparable results, although the smaller time step tested was consistently better. However, longer time steps are translated to lower operational cost, since information will have to be predicted and broadcasted fewer times than in the case of small time steps. When the results of the cases h=10 and h=5 are compared with those for h=2 it is more obvious that for longer term predictions, long time steps give better results. First because they depend more heavily on the average travel time of the link and second because they are obtained with fewer step ahead predictions.

5.4.2 Effect of the Time Interval under Congested Traffic Conditions

For the situation of the congested traffic conditions the performance of the RLS algorithm under different time steps was evaluated based on the results of link 19, on which the accident occurred. The model that was used for the prediction of the travel times of this link consisted of the autoregressive term with order n=1 and the diurnal term and the forgetting factor that was set at λ=0.90. The errors were computed for the period of time starting at 8:00 and ending at 12:00.

The results are summarized in Figure 5.44 and Figure 5.45 where the sensitivity of the prediction algorithm to the length of the time step in terms of the mean absolute error and the mean square error is shown. As was expected, the shorter the time step, the smaller the resulting error, since the estimation of the travel time of the link at the beginning and at the end of the incident which contain large errors, were in effect for longer times. This was true for predictions of the travel time of the link for all times that were tested. In contrast to the case of normal operations where errors of the predicted values were tempered after the first few minutes, here it was clear that predictions further into the future contain more error, with almost a linear relationship. As is revealed from Figure 5.44, the mean error of the predictions are consistently worse as the time step increases by almost a constant amount.

When a large time step is used the adaptation of the model to the sudden increase of the travel time is hampered even more due to the limited number of observations, and thus the resulting errors are worse. On the contrary, in the case of small time steps, changes of travel time are perceived by the algorithm as being more gradual and thus adaptation is more successful, and predictions even further into the future are of better quality than those obtained in the case of longer time steps.

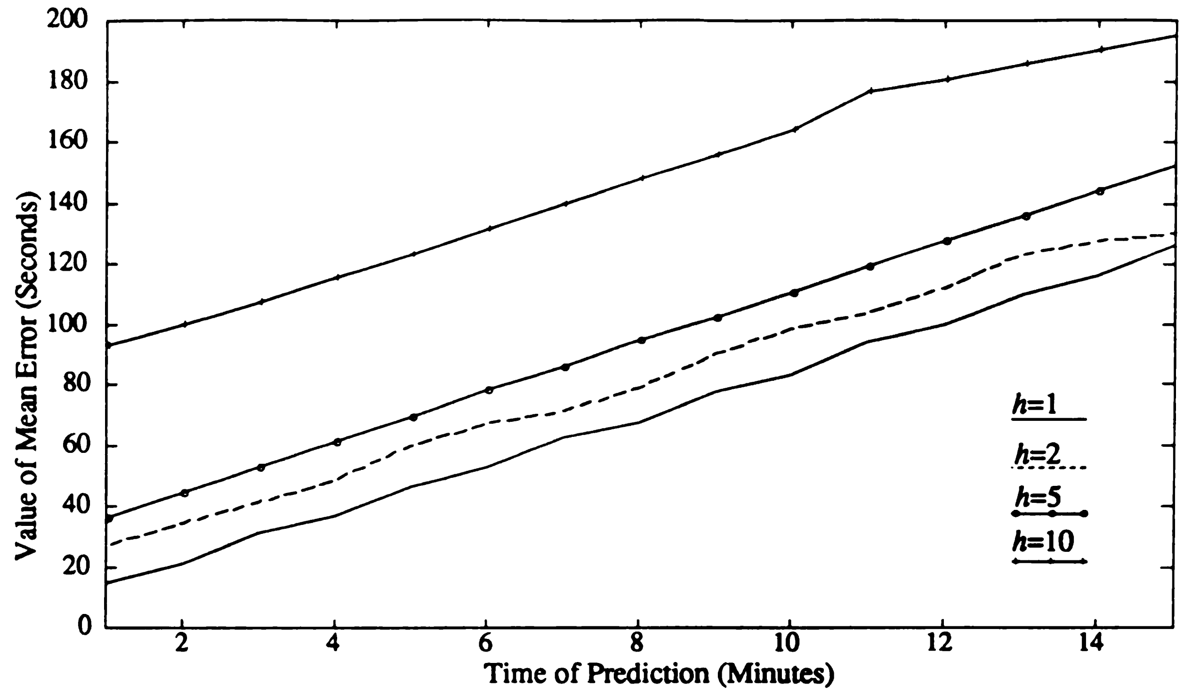


Figure 5.44: Behavior of Mean Absolute Error for Different Time Steps under Congested Traffic Conditions

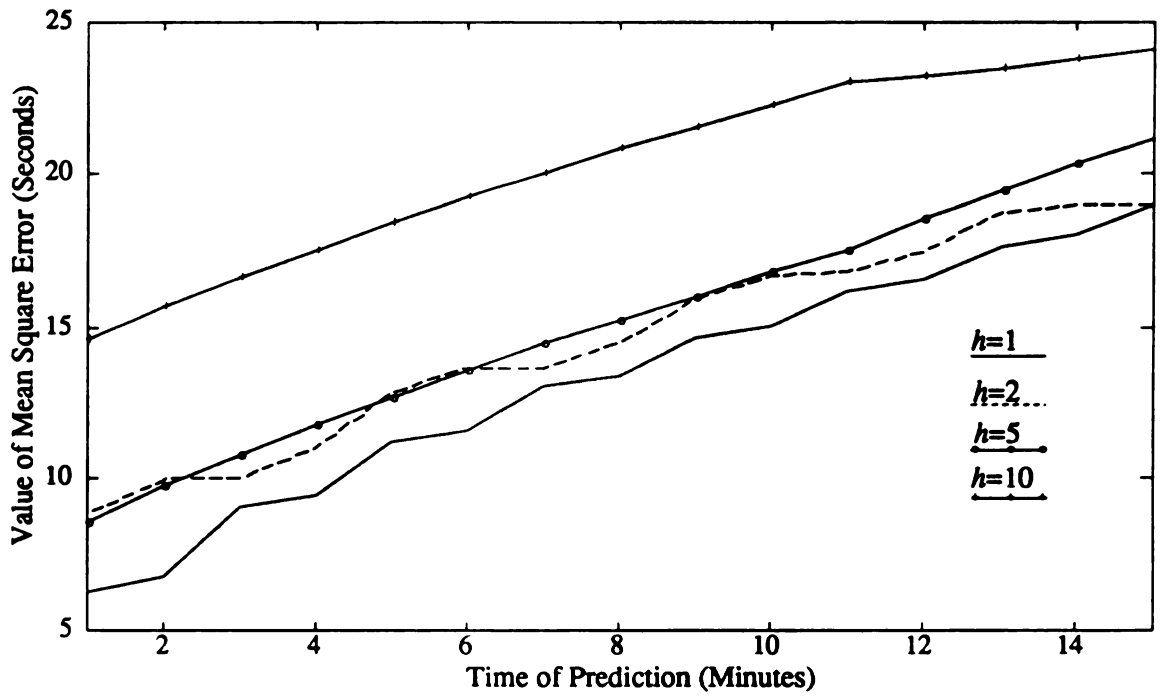


Figure 5.45: Behavior of Mean Square Error for Different Time Steps under Congested Traffic Conditions

Effectiveness of Route Guidance System

6.1 Introduction

In this section of the study we assess the impact that the route guidance system utilizing the prediction algorithm has on the throughput of the network. Of course, poor predictions may have a negative impact on the performance of smart vehicles, with possibly negative effects on the rest of the traffic as well. On the other hand, good predictions should improve the travel times of smart vehicles more than if smart vehicles had information only on current traffic conditions and not on predicted ones. The performance of the system was examined at different levels of market penetration of smart vehicles. In all instances of market penetration and route guidance schemes, the same traffic demand patterns were used. Also in all runs an accident was simulated on link 19 starting at time 08:34:10.2 and lasted till time 08:50:21.5.

6.2 Effectiveness of Route Guidance System at Different Levels of Market Penetration

The hourly average travel time for smart and non smart vehicles for the route guidance schemes utilizing the prediction model and at five different levels of market penetration are illustrated in Figure 6.1 through Figure 6.5. As can be seen from these figures, for time periods other than the one where operations are affected by the traffic accident, travel times of smart vehicles and non smart vehicles are almost identical. They differ only by a few seconds, and smart vehicles do not always have shorter average travel time than non smart vehicles.

The above observation is explained if we consider the definition of the non

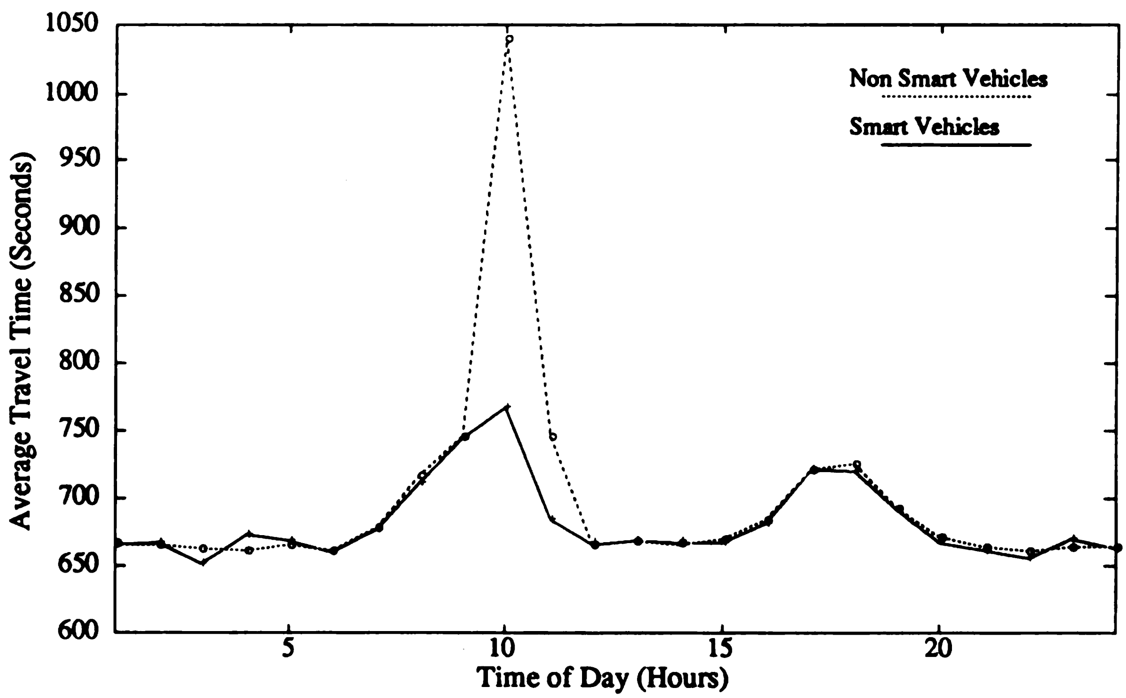


Figure 6.1: Hourly Average Travel Time - Market Penetration 10%

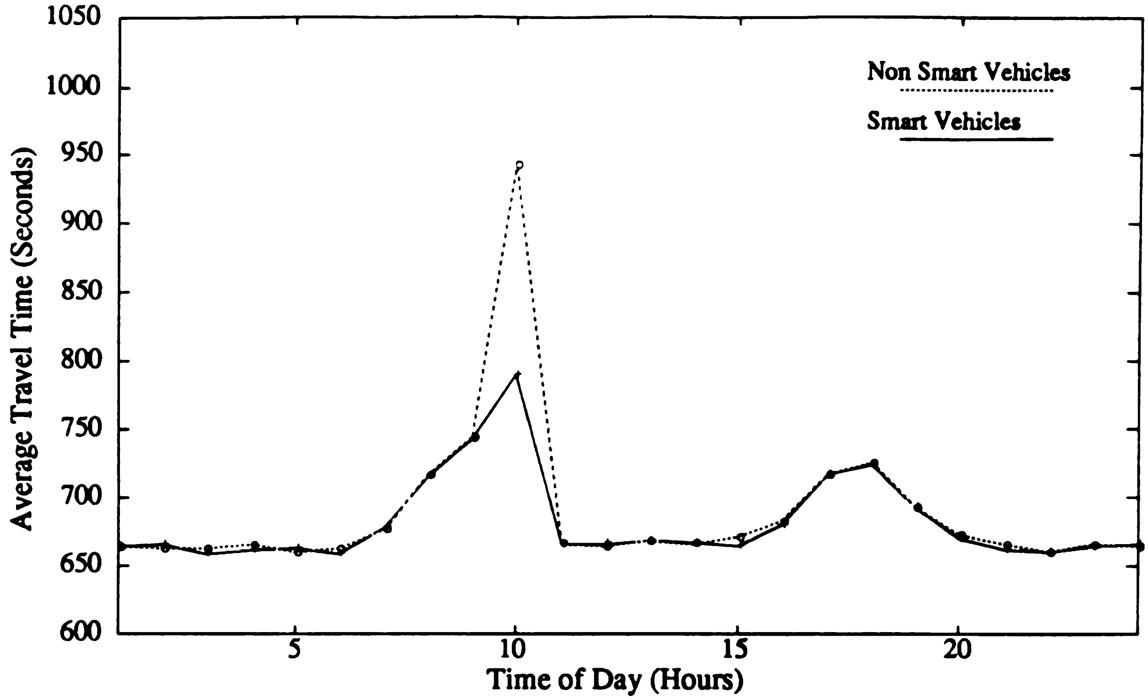


Figure 6.2: Hourly Average Travel Time - Market Penetration 30%

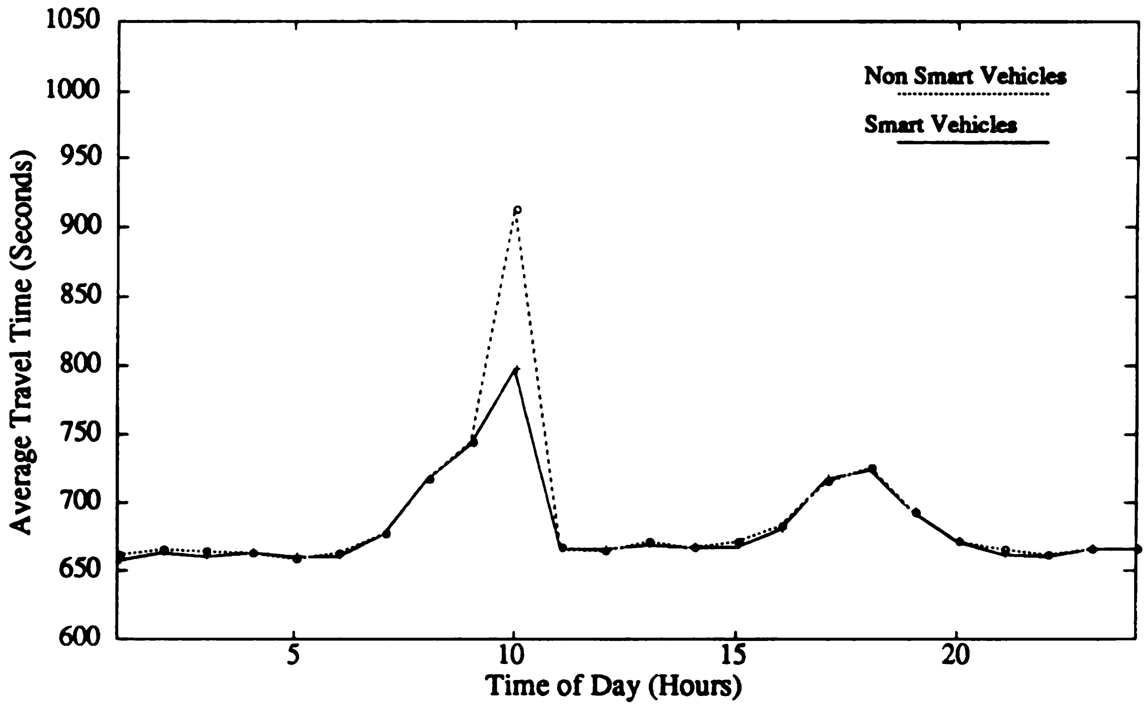


Figure 6.3: Hourly Average Travel Time - Market Penetration 50%

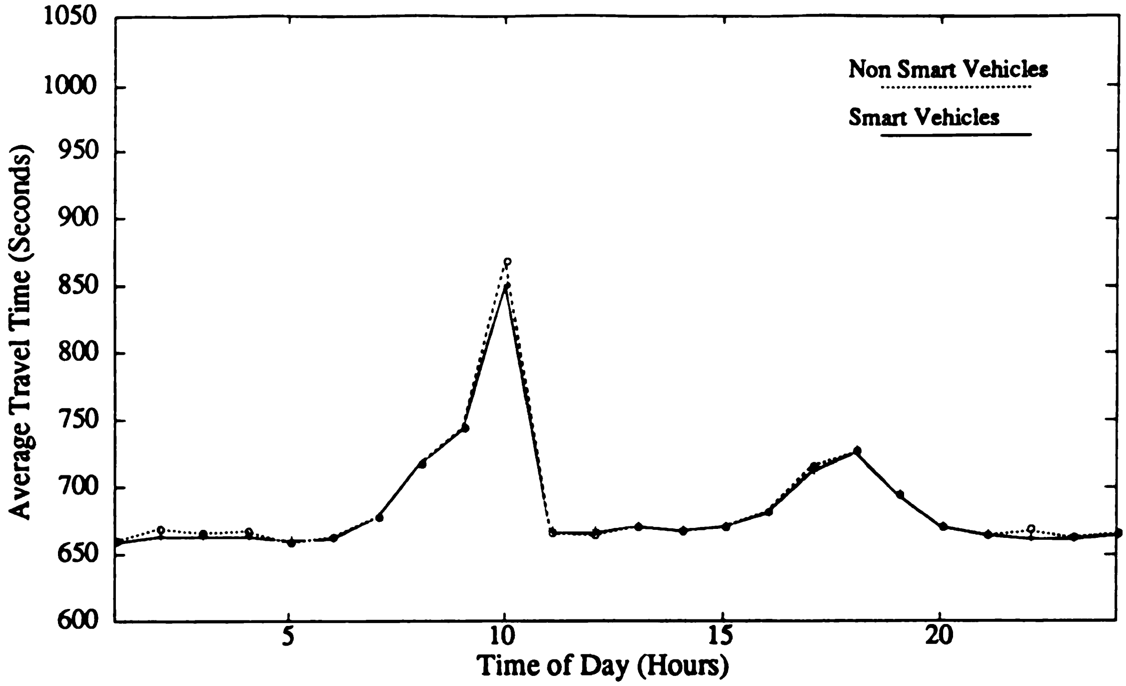


Figure 6.4: Hourly Average Travel Time - Market Penetration 70%

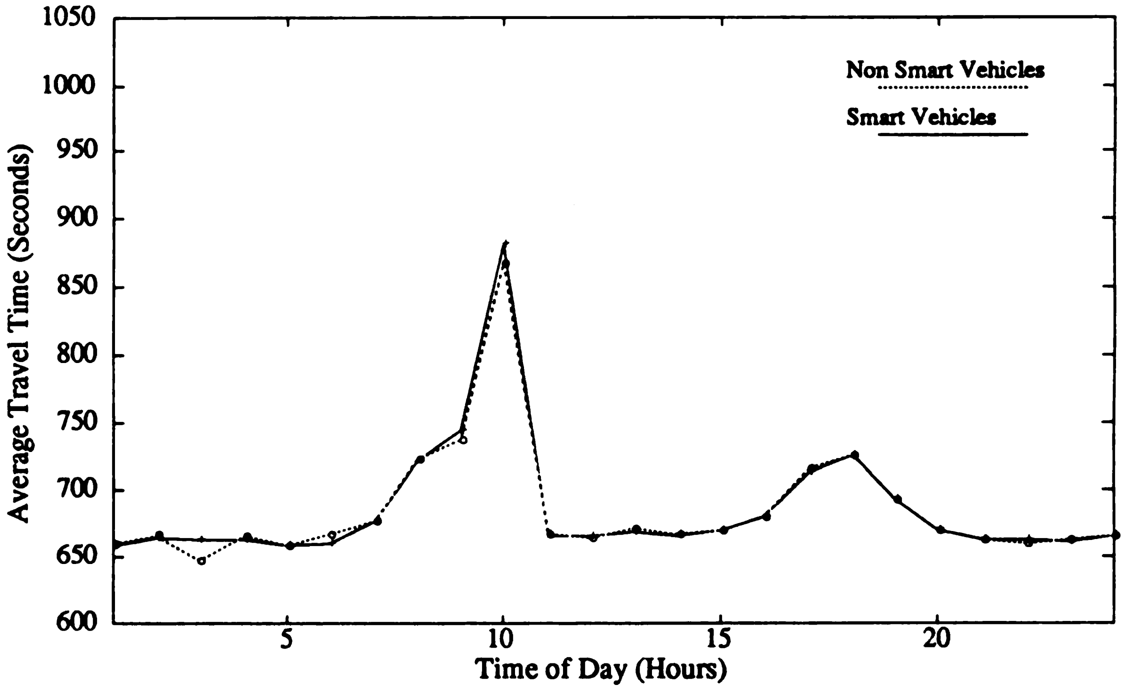


Figure 6.5: Hourly Average Travel Time - Market Penetration 90%

smart vehicles. Actually, in the case of normal operations, non smart vehicles do not differ from smart vehicles. This is true for two reasons: first, the information on which they base their route choice, the smoothed average travel time of the link for the current five minute interval, is a relatively good approximation of the actual travel time. Second, the actual travel time of any given path (under normal traffic conditions) is not that much worse than non smart drivers perceive it to be, so no alternative route would be suggested to smart vehicles. This is further reinforced by the fact that vehicles do not change routes for very small gains in their travel time, but only if the gains of the new path exceed a minimum threshold value. Therefore, when there is no congestion due to an unexpected event, smart and non smart vehicles chose the same route toward their destination, even if the perception of the travel time is more accurate for smart vehicles. In this case any differences in the average travel time between the two vehicle categories is random and not due to any intelligent routing strategy. Of course this behavior of smart and non smart vehicles is very much dependent on the layout of the network, and the threshold (in terms of time) applied to vehicles shifting from one route to an other. For example, if drivers could alter their route with very small penalty in the travel time, which would results from shifting from the original path, it is believed that the savings of the smart vehicles would be evident.

On the other hand in the case of the traffic accident smart vehicles know of the occurrence of the incident when the incident occurs. In addition they have information regarding the expected time of duration of the incident, and as soon as the incident ends they receive this information as well. Due to this information, smart vehicles are guided to alternative paths and thus are able to improve their travel times, when such alternative routes exist. Therefore, the actual time savings (in the case of the given network) for smart vehicle accrue from their route choices during the period that operations on the network are affected by the accident on link 19. However, even in the case of the incident, non smart vehicles employ a quite intelligent routing strategy: they enter the link till the link is filled, and after this point they reevaluate their route toward their destination based on the fact that link 19 is congested.

A prominent difference though is observed among the average travel times of both smart and non smart vehicles for the different market penetration levels during the incident. Figure 6.6 illustrates the profiles of the average travel time of smart and non smart vehicles during the incident for the different penetration levels. For low penetration levels the average travel time of the smart vehicles is much lower than the one for non smart vehicles (from 26.3% to 16.1% lower average travel times for smart vehicles at levels of market penetration of 10% and 30% respectively - Table 6.1). However this difference is diminished for higher percentages of market penetration and for a penetration level as high as 90% it is even inverted in favor of the non smart vehicles.

The total throughput of the network is also shown in Figure 6.6. As can be seen in this figure, for low levels of market penetration the overall performance of the network improves until the 50% level, while after this point, it starts a slow decline. Thus, the route guidance system appears to reduce the effect of the incident on the overall performance of the network at all levels of market penetration.

The accident on link 19 is triggered by a vehicle which has traveled almost 75% of the link before it creates the incident. Therefore, a large portion of the

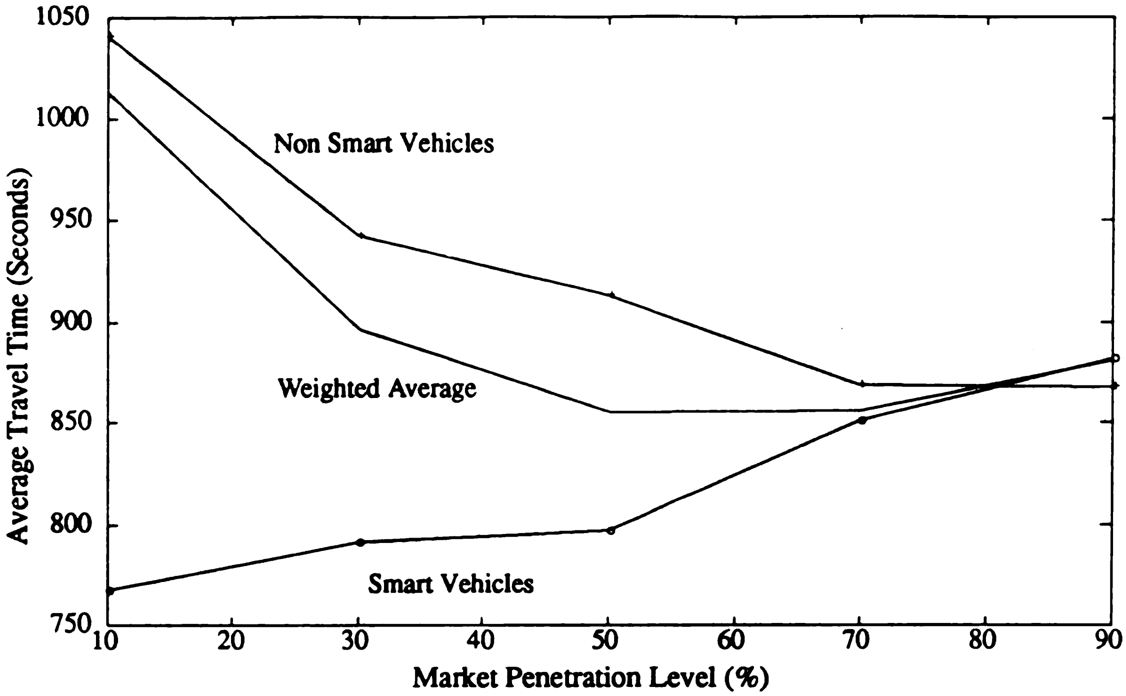


Figure 6.6: Travel Time Profile for Different Levels of Market Penetration

Table 6.1: Travel Times and Percent Differences of Smart and Non Smart Vehicles for Different Market Penetration Levels.

Market Penetration	Smart (seconds)	Non Smart (seconds)	Difference %	Weighted Average
10%	767.09	1041.26	26.3	1013.9
30%	791.23	943.01	16.1	897.5
50%	797.28	913.13	12.7	855.2
70%	851.22	868.38	1.9	856.4
90%	881.76	867.52	-1.6	880.4

high travel time of both smart and non smart traffic, is due to the vehicles from each category that are traversing link 19 before the start of the incident and have entered the link after the entrance of the vehicle that will cause the incident (such vehicles that join the queue in link 19 will have to wait until the end of the incident and thus their travel time is increased by approximately 950 seconds, the duration of the incident). Since smart vehicles will stop entering link 19 right after the start of the incident, while non smart vehicles will enter until the link is completely congested, their average travel time never reaches the high levels of the travel time of the non smart vehicles in the cases of market penetration of 10% and 30%.

However, at high levels of market penetration, and due to the long duration of the incident, smart traffic which is aware of the congestion, in order to bypass the congested link, starts diverting to alternative routes which soon become congested because of the large number of vehicles that divert to these paths. Thus, equipped vehicles which still have the opportunity start diverting to new alternative paths from further upstream. On the other hand, non smart vehicles improve their average travel time as the market penetration increases, since at higher levels of market penetration the formed queues are shorter. This results in faster dissipation of the formed queues on the links that become congested, with the subsequent result of shorter waiting times in such queues.

In the case where there was no incident in the network the average travel time of smart vehicles at the same time period (from 9:00 to 10:00) was almost the same for all levels of market penetration ranging form 665.6 to 667.4 seconds. The same was true for non smart vehicles with average travel times for the same period ranging from 665.8 to 667.5 seconds. Smart vehicles have higher travel times as compared with the travel times at the no incident case for two reasons. First traffic that is diverted to alternative routes is forced to follow longer paths than what it would follow if the incident had not occurred. Second, some smart vehicles are trapped in the queues created due to the incident as explained above.

Due to the time lag between the inspection of the congestion, and the time this information is broadcast to the traffic, equipped vehicles will start diverting to alternative routes only after the system updates the information regarding the current travel conditions on the links. Until then, even smart vehicles will follow the directions of the system from the last time step. However, such directions may lead them to routes which already are becoming congested during the last time step. Of course, the shorter the time between updates of information regarding link traffic conditions during congestion periods, the less severe the effects of the congestion.

This is demonstrated in Figure 6.7 where the average travel time for smart and non smart vehicles are plotted in the case that the link travel time information is updated and broadcast every 30 seconds. From these results it becomes obvious that higher frequency of information update improves the performance of smart vehicles significantly for all levels of market penetration. Of course this is also associated with the fact that predictions of link travel times are also of better quality for smaller time steps, as was discussed in paragraph 5.4.2. The percent gains of the average travel times for the period after the accident of each vehicle category, as well as of the throughput of the network are listed in Table 6.2. The effectiveness of the shorter time

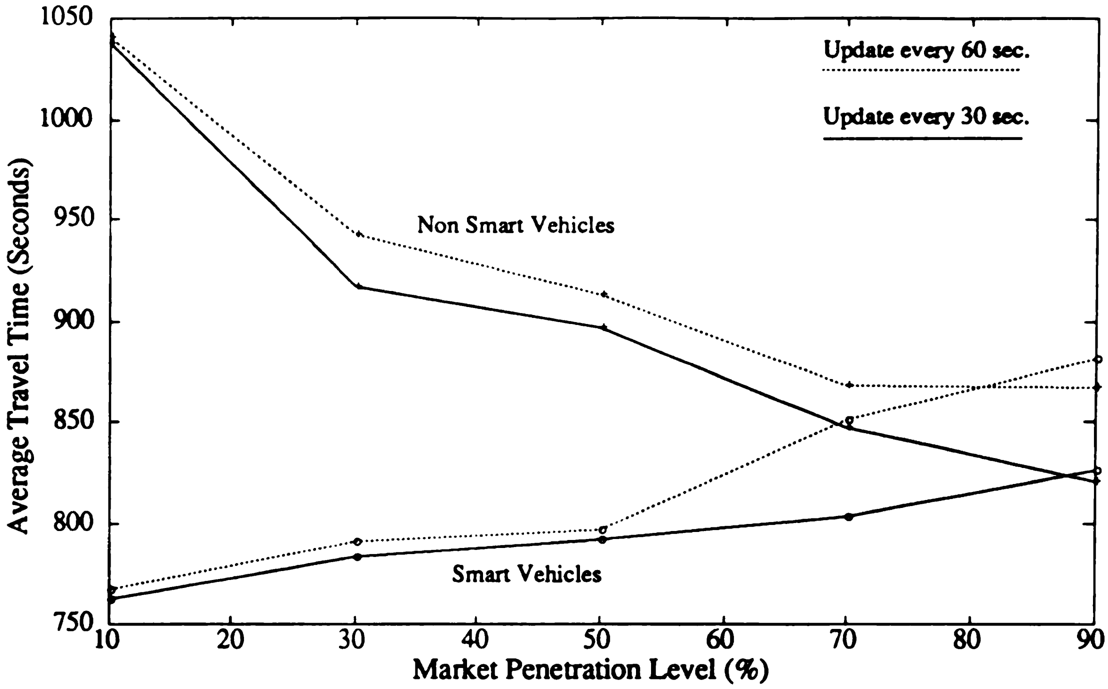


Figure 6.7: Comparative Travel Time Profiles for Information Updates Every 60 and Every 30 Seconds

Table 6.2: Travel Times of Smart and Non Smart Vehicles for Different Market Penetration Levels when Information is Updated Every 30 Seconds

Market Penetration	Smart (seconds)	Non Smart (seconds)	Weighted Average	% Reduction of Delay [1]		
				Smart	Non Smart	Overall
10%	762.54	1038.55	1010.9	4.6	0.7	0.8
30%	783.04	917.22	877.0	6.6	9.3	8.9
50%	792.12	896.68	844.4	4.0	6.7	5.7
70%	803.34	847.36	816.5	26.0	10.5	21.0
90%	825.99	820.83	825.5	26.0	23.3	25.7

1. Delays are calculated as the difference of the average travel times between the case of prediction update every 30 seconds and every 60 seconds reduced by 667 seconds which is approximately the travel time of both smart and non smart traffic in the case of no incident for all market penetration levels.

interval increases with the level of market penetration. This indicates that the importance of having the correct information as soon as possible, is more crucial when a large percentage of the traffic has access to this information, so congestion to the alternative paths which are suggested by the system (or calculated by the on board computers based on the information provided by the system) will be avoided. At penetration levels of 90% smart vehicles still have longer average travel time than the non smart vehicles. However, the difference between their travel time is reduced almost to one third the value in the case of 60 second updates, while at 70% penetration, this difference is much larger, from 17.2 to 44.0 seconds.

When the penetration level is only 10% frequent update of information about link conditions does not appear to have a strong impact on the average travel time of the non smart vehicles and the overall performance of the network, while for smart vehicles there are moderate improvements. On the other hand, when broadcast information is affecting a large percentage of traffic, this traffic moves more efficiently with more frequent information updates. In addition the overall performance of the network is greatly improved (Table 6.2).

Summary and Conclusions

From the literature review it become apparent that models designed to estimate future travel conditions based on current ones all have limitations. However, models of this type are needed in the development and application of IVHS technologies, and more specifically in ADIS. A significant level of research is concentrated on extending equilibrium models for this purpose, but such models are only applicable when the market penetration is rather high.

In this study we examined the use of a recursive identification model for predicting travel time on the links of a traffic network in real time, when the input signal to the model is comprised of current and past information regarding traffic condition on the network. The main reason for using a recursive algorithm is to provide the adaptive characteristics required, to respond to evolving traffic conditions. The proposed prediction algorithm is not new, but is an existing methodology applied to the specific problem of predicting travel times on the links of a network. Similar algorithms have been successfully used in the past, mostly for predicting traffic volumes. However, the use of these models in route guidance systems, where the availability of real time information to drivers may affect the traffic patterns on the network, has not been tested until now.

Due to the lack of real world data from a system where drivers (or a portion of drivers) have access to real time information regarding traffic conditions, it was decided that the prediction model would be tested on simulated data. Furthermore, due to lack of access to the source code of simulation programs that could be modified for simulating such a network-driver system it was neccesary to construct such a traffic simulation program. The simulation program is microscopic at the vehicle level, and was developed so any shape of network could be simulated. A certain percentage of the traffic has access to real and predicted information, and routing of traffic is performed under the properties of dynamic networks.

The prediction algorithm was tested under normal traffic conditions where traffic operations were not affected by any unexpected event such as a traffic accident, and under congested traffic conditions where the operating conditions were greatly influenced by the occurrence of a traffic accident on a link of the network. The input data to the link travel time prediction model considered in this study included the time series of the travel times of the link that is under consideration, as well as information from those links ending at the entrance node of this link and the ones emanating from the exit node of the link. In addition the time dependent average travel time on the link observed during previous days, was included as well. The measures of effectiveness used were based on the errors produced by each model. The magnitude of the error was compared with the errors produced when the only information available to equipped drivers was the current travel time on the link.

When the model includes only the autoregressive term, the predictions were only marginally better than the no predictions case. When a large order model was used, the performance of the algorithm improved, but it is recognized that such models overparametrize the system, with the result that many of the parameters do not affect the predictions. When the model was modified to utilize different weights for the prediction errors, by means of a forgetting factor λ, to discount errors in the distant past, the model performance was expected to improve. However, when the forgetting factor was set to values smaller than 1.00, so the tracking capability of the model would be used, the prediction errors increased.

When a diurnal term was included in the model, predictions were better than those obtained with the autoregressive term alone. This improvement increases for predictions further in the future, mainly because such predictions are smoothed by the increasing effect of the diurnal component. For example one, five and ten step ahead predictions have approximately 10% 20% and 35% less error than the no prediction case. By comparison, when only the autoregressive term was used, the one step predictions were approximately 7% better and five and ten step ahead predictions were only marginally better.

The inclusion of information from upstream links in the model enhanced the performance of the model even more. The expected errors of these prediction models were reduced up 31% and 37% for the one and five step ahead predictions respectively, while ten step ahead predictions gained only an additional

8%. In this model, the effect of a forgetting factor set at 0.99 was positive, and one step ahead predictions were better when $\lambda=1.00$, and five step ahead predictions where almost equal. These results were obtained when information from links that were rarely used was excluded. This was necessary since such information does not affect the predictions, and very small changes in the travel time of such links resulted in large prediction errors.

Using the same model, it was detrmined that the number of past observations in the convection term represents the time lag required for the traffic wave to travel downstream. For example, in the case examined two parameters ($b_{8,1}$ and $b_{8,2}$) were included when using a 60 seconds time step since traffic waves needed approximately 120 seconds to reach from the upstream link to the link under consideration. The parameter $b_{8,1}$ was approaching zero, and exclusion of it from the model may have improved the results.

While these results were obtained for the normal traffic conditions, when a traffic accident occurs , the model performed worse than the case of no prediction when the forgetting factor λ was set to one. However, when the parameters of the model were not assumed to be fixed, and the forgetting factor was set to relatively small values such as 0.80, the model gave predictions that were better for five and ten steps ahead. The one step ahead predictions still contained more error than the no prediction case. When information from downstream links was included in the model, predictions under normal traffic conditions deteriorated, while for the link that is upstream from the one where the accident occurred, predictions were improved when the congestion term with a large order was included. It was expected that such input would improve the predictions for this link since it contains information about the shock wave which is traveling backwards. However, due to the alternative paths available, this relationship was not manifested as early as the shock wave reached the exit of the link with the accident, and thus the large order congestion term had to be used.

Since information regarding the time and location of a traffic accident is not predictable, since traffic accidents are random events with very small probability, it is not possible to operate the prediction model with the structure and the forgetting factor that would give good predictions only under congested conditions. Therefore, to obtain good results it was neccesary to develop a time varying weighting function to replace the forgetting factor. While further development is required for such a function, a function was designed for this study which would assume small values when large prediction errors occur, while the rest of the time λ would remain close to one. The results obtained were not as good as those obtained during the congestion period by the models with a small forgetting factor, but they were better than those obtained with the forgetting factor set at 1.00 or 0.99.

With the model containing the autoregressive the diurnal and the convection terms, the sensitivity of the prediction algorithm to the size of the interval between recursions of the algorithm was examined. It appears that the smaller the interval, the better the predictions of the travel time on the link in the immediate future. For predictions further in the future it was observed that the algorithm is relatively insensitive to the length of the interval, and predictions obtained with large intervals were surprisingly good. This result means that in the selection of the time step, the length of the links in the network should be taken into account. If very short term (i.e. 1 minute) predictions are not needed, a larger time step may be used, which would result in less operational cost. However, in the case of a traffic accident, it was clear that the larger the time step the greater the error in the predictions.

The effectiveness of the route guidance system with the prediction model was examined at different levels of market penetration. From the result of this part of the study, it is obvious that benefits of such systems will accrue mostly during unexpected congestion (such as congestion induced by traffic incidents) when non equipped vehicles are not aware of the downstream congestion. Benefits during recurrent congestion are greatly dependent on the network layout, as well as the behavior of non equipped vehicles. Nevertheless during accident induced congestion, equipped vehicles do significantly better than non equipped vehicles, at least at market penetration levels lower than 50%. At higher levels of market penetration equipped traffic appears to congest the alternative routes, and at levels as high as 90% their travel time is even higher than the travel time of non equipped vehicles. This is tempered when the time step is decreased from 60 to 30 seconds.

Along with the performance of the equipped traffic the performance of the entire system also appears to improve as market penetration increases until it reaches 50% for the 60 second time step, and until it reaches 70% for the 30 second time step. At lower levels of market penetration the performance of the model appears to be insensitive to the length of the time step while for higher levels the benefits are significant. This result indicates that during congested periods accurate information regarding traffic conditions is essential when such information affects the routing decisions of a large portion of the traffic.

Although the effectiveness results are dependent on the network used, the principles underlying the prediction model will be applicable to more general networks as well. While further research is necessary to improve the prediction model, particularly during incident induced congestion, the proposed model appears to perform very well especially during normal traffic conditions. Improvement of the performance of the model may be achieved by filtering the input data and/or the predictions, so gross errors that may occur during the first time steps of an incident, which are also the most critical, are minimized.

BIBLIOGRAPHY

Bibliography

1. Al-Deek Haitham and Kanafani Adib, 1991. *Incident Management with Advanced Traveller Information System.* Paper presented at the VNIS'91 Conference, Dearborn, Michigan.

2. Ahmed Mohamed and Cook Allen, 1982. *Analysis of Freeway Traffic Time Series Data by Using Box-Jenkins Techniques.* Transportation Research Record 722, pp 1-9, TRB, National Research Council Washington, D.C.

3. Ahmed Mohamed and Cook Allen, 1980. *Time Series Models for Freeway Incident Detection.* Transportation Engineering Journal of ASCE, Vol 106, No TE6, American Society of Civil Engineers, New York, New York.

4. Åstrom Karl Johan and Wittenmark Björn, 1989. *Adaptive Control.* Addison Wesley Publishing Company.

5. Boyce David, 1984. *Route Guidance Systems for Improving Urban Travel and Location Choices.* Transportation Research A, Vol 22A, No 4, pp 275-288.

6. Boyce David, 1989. *Route Guidance Systems for Managing Urban Transportation Networks: Review and Prospects.* Paper presented at 10th Italian Regional Conference, Rome, Italy.

7. Brockwell Peter and Davis Richard, 1987. *Time Series: Theory and Methods.* Springer Verlag Book Company, New York.

8. Carey M., 1987. *Optimal Time Varying Flows on Congested Networks.* Operations Research, Vol 35, No 1, pp 58-69, Operation Research Society of America.

9. Chen Kan and Underwood Steven, 1991. *Research on Anticipatory Route Guidance.* Paper presented at the VNIS'91 Conference, Dearborn, Michi-

gan.

10. Chen Peter She-Te and Mahmassani Hani,1991. *Reliability of Real-Time Information Systems for Route Choice Decisions in a Congested Traffic Network: Some Simulation Experiments.* Paper presented at the VNIS'91 Conference, Dearborn, Michigan.

11. Cook, K. L. and Hasley E., 1966. *The Shortest Route Through a Network with Time Dependent Intermodal Transit Times.* Journal of Mathematical Analysis and Applications, Vol 14, pp 493-498.

12. Davis Gary and Nihan Nancy, 1984. *Using Time Series Designs to Estimate Changes in Freeway Level of Service, Despite Missing Data.* Transportation Research A, Vol 18A, No 5/6, pp 431-438.

13. Ervin Robert and Chen Kan, 1990. *Providing Radical Functionality to Serve Highway Transportation: A 20 Year Vision for IVHS.* University of Michigan Transportation Research Institute, Ann Arbor, Michigan.

14. Euler Gary, 1990. *Intelligent Vehicle/Highway Systems: Definitions and Applications.* ITE Journal, Vol 60, No 11, pp 23-31, Institute of Transportation Engineers, Washington, D.C.

15. French Robert, 1990. *Intelligent Vehicle/Highway Systems in Action.* ITE Journal, Vol 60, No 11, pp 23-31, Institute of Transportation Engineers, Washington, D.C.

16. Friesz Terry, Luque F., Tobin Roger and Wie Byung-Wook, 1989. *Dynamic Network Traffic Assignment Considered as a Continuous Time Optimal Control Problem.* Operations Research, Vol 37, No 6, pp 893-901, Operation Research Society of America

17. Gazis Denos and Knapp Charles, 1971. *On Line Estimation of Traffic Densities from Time Series of Traffic and Speed Data.* Transportation Science, Vol. 5, No 3.

18. Grenzeback Lance and Woodle Clyde, 1992. *The True Costs of Highway Congestion.* ITE Journal, Vol 62, No 3, pp 16-20, Institute of Transportation Engineers, Washington, D.C.

19. Guiliano Genevieve., 1989. *Incident Characteristics, Frequency and Duration on a High Volume Urban Freeway.* Transportation Research A, Vol 23A, No 5, pp 387-396.

20. Hamerslag Rudi and van Berkum Eric, 1991. *Effectiveness of Information Systems in Networks With and Without Congestion.* Paper presented at the 70th TRB meeting, Washington, D.C.

21. Halati Abolhassan and Boyce David, 1991. *Effectiveness of In-Vehicle Navigation Systems in Alleviating Non-Recurring Congestion.* Paper presented at the VNIS'91 Conference, Dearborn, Michigan.

22. Hall Randolph, 1986. *The fastest Path through a Network with Random Time Dependent Travel times.* Transportation Science Vol 20, No 3.

23. Haykin Simon, 1984. *Introduction to Adaptive Filters.* Macmillan Publishing Company, New York.

24. Kaufman David, Lee James and Smith Robert, 1990. *Anticipatory Traffic Modeling and Route Guidance in Intelligent Vehicle Highway Systems.* IVHS Technical Report 90-2, University of Michigan Transportation Research Institute, Ann Arbor, Michigan.

25. Kaufman David and Smith Robert, 1990. *Fastest Paths in Dynamic Networks with Application to Intelligent Vehicle Highway Systems.* IVHS Technical Report 90-4, University of Michigan Transportation Research Institute, Ann Arbor, Michigan.

26. Kaufman David, Smith Robert and Wunderlich Karl, 1991. *An Alternative Routing/Assignment Method for Anticipatory Real Time Route Guidance.* Paper presented at the VNIS'91 Conference, Dearborn, Michigan.

27. Koutsopoulos Haris and Lotan Tsippy, 1990. *Motorist Information Systems and Recurrent Traffic Congestion: A Sensitivity Analysis of Expected Benefits.* Paper presented at the 69th TRB meeting, Washington, D.C.

28. Koutsopoulos Haris and Yablonski Ann, 1991. *Design Parameters of Advanced Driver Information Systems: The case of Incident Congestion and Small Market Penetration.* Paper presented at the VNIS'91 Conference, Dearborn, Michigan.

29. Kyte Michael, Marek Joseph, Frith Douglas 1989. *Using Multivariate Time Series Analysis to Model Freeway Traffic Flow.* Paper presented at the 68th TRB meeting, Washington, D.C.

30. Lafortune Stephane, 1990. *Anticipatory Traffic Modeling and Route Guidance: A General Mathematical Formulation.* IVHS Technical Report 90-6,

University of Michigan Transportation Research Institute, Ann Arbor, Michigan.

31. Lafortune Stephane, Sengupta Raja, Kaufman David and Smith Robert 1991. *A Dynamic System Model for Traffic Assignment in Networks*. Paper presented at the VNIS'91 Conference, Dearborn, Michigan.

32. Ljung Lennart and Söderström Torsten 1983. *Theory and Practice of Recursive Identification*. MIT Press, Cambridge, Massachusetts.

33. Lu Jian, 1990. *Prediction of Traffic Flow by an Adaptive Prediction System*. Transportation Research Record 1287, pp 54-61, TRB, National Research Council, Washington, D.C.

34. Mahmassani Hani and Chen Peter Shen-Te, 1991. *Comparative Assessment of Origin-Based and En-Route Real-Time Information Under Alternative User Behavioral Rules*. Paper presented at the 70th TRB meeting, Washington, D.C.

35. Mahmassani Hani and R. Jayakrishan, 1990. *System Performance and User Response under Real Time Information in a Congested Traffic Corridor.* Paper presented at the 69th TRB meeting, Washington, D.C.

36. Mauro Vito, 1991. *Evaluation of Dynamic Network Control: Simulation Results Using NEMIS Urban Microsimulator.* Paper presented at the 70th TRB meeting, Washington, D.C.

37. Merchant Deepak and Nemhauser George, 1978a. *A Model and an Algorithm for the Dynamic Traffic Assignment Problems*. Transportation Science, Vol 12, No 3, pp 183-199.

38. Merchant Deepak and Nemhauser George, 1978b. *Optimality Conditions for a Dynamic Traffic Assignment Model.* Transportation Science, Vol 12, No 3, pp 200-207.

39. Mulgrew Bernard and Cowan Colin, 1988. *Adaptive Filters and Equalizers*. Kluwer Academic Publishers, Boston, Massachusetts.

40. Newell Gordon, 1980. *Traffic Flow on Transportation Networks*. MIT Press, Cambridge, Massachusetts.

41. Nihan Nancy and Davis Gary, 1989. *Application of Prediction-Error Minimization and Maximum Likelihood to Estimate Intersection O-D matrices*

from Traffic Counts. Transportation Science, Vol 23, No 2, pp 77-90.

42. Nihan Nancy and Davis Gary, 1987. *Recursive Estimation of Origin Destination Matrices from Input/Output Counts.* Transportation Research B, Vol 21B, No 2, pp 149-163.

43. Okutani Iwao and Stephanedes Yorgos, 1984. *Dynamic Prediction of Traffic Volume through Kalman Filtering Theory.* Transportation Research B, Vol 18B, No 1, pp 1-11.

44. Papageorgiou Markos, 1983. *Applications of Automatic Control Concepts to Traffic Flow Modelling and Control.* Springer-Verlang Book Company, Lecture Notes in Control and Information Sciences, No 50, Berlin, Germany.

45. Papageorgiou Markos, 1990. *Dynamic Modeling, Assignment, and Route Guidance in Traffic Networks.* Transportation Research B, Vol 24B, No 6.

46. Papageorgiou Markos and Messmer Albert, 1991. *Dynamic Network traffic Assignment and Route Guidance via Feedback Regulation.* Paper presented at the 70th TRB meeting, Washington, D.C.

47. Papoulis Athanasios, 1984. *Probability, Random Variables and Stochastic Processes.* McGraw Hill Book Company, New York.

48. Paull Marvin, 1988. *Algorithm Design: A Recursion Transformation Framework.* Wiley, New York.

49. Pritsker Alan, 1986. *Introduction to Simulation and SLAM II.* Halsted Press, Wiley, New York.

50. Ran Bin, Boyce David and LeBlanc Larry, 1992. *Some Basic Constraints for a Dynamic Traffic Assignment Model.* Paper presented at the 71st TRB meeting, Washington, D.C.

51. Ross Sheldon, 1985. *Introduction to Probability Models.* Academic Press, New York.

52. Romeijin Edwin, Smith Robert, 1990. *Notes on Parallel Algorithms and Aggregation for Solving Shortest Path Problems.* Class notes, Ann Arbor, Michigan.

53. Rosen Dan, Mammano Frank and Favout Rinaldo, 1970. *An Electronic Route Guidance System for Highway Vehicles.* IEEE Transactions on Vehic-

ular Technology, Vol VT-19, No 1.

54. Sen Ashish, 1991. *Short Term Forecasting of Link Travel Times: A Preliminary Proposal.* Class notes, Ann Arbor, Michigan.

55. Sheffi Yossef, 1985. *Urban transportation Networks: Equilibrium Analysis with Mathematical Programming Methods.* Prenctice Hall, New York.

56. Stephanedes Yorgos, Kwon Eil, Michalopoulos Panos, 1990. *On Line Diversion Prediction for Dynamic Control and Vehicle Guidance in Freeway Corridors.* Transportation Research Record 1287, pp 11-19, TRB, National Research Council, Washington, D.C.

57. TRB, 1975. *Traffic Flow Theory.* Special Report 165, Transportation Research Board, National Research Council, Washington, D.C.

58. TRB, 1985. *Highway Capacity Manual.* Special Report 209, Transportation Research Board, National Research Council, Washington, D.C.

59. Underwood Steven, 1990. *Review of Traffic Simulations for Intelligent Vehicle-Highway System Evaluation.* IVHS Technical Report 90-10, University of Michigan Transportation Research Institute, Ann Arbor, Michigan.

60. U.S. General Accounting Office, 1991. *Smart Highways: An Assessment of Their Potential to Improve Travel.*

61. Van Aerde Michel and Yagar Sam, 1988a. *Dynamic Integrated Freeway / Traffic Signal Networks: Problems and Proposed Solutions.* Transportation Research A, Vol 22A, No 6, pp 435-443.

62. Van Aerde Michel and Yagar Sam, 1988b. *Dynamic Integrated Freeway / Traffic Signal Networks: A Routing Based Modelling Approach.* Transportation Research A, Vol 22A, No 6, pp 445-453.

63. Wie Byung-Wook, 1991. *Dynamic Analysis of User Optimized Network Flows with Elastic Travel Demand.* Paper presented at the 70th TRB meeting, Washington, D.C.

64. Wie Byung-Wook, Friesz Terry and Tobin Roger, 1990. *Dynamic User Optimal Traffic Assignment on Congested Multi Destination Networks.* Transportation Research B, Vol 24B, No 6, pp 431-442.