CWI Tracts

Managing Editors

Editorial Board

CWI
P.O. Box 94079, 1090 GB Amsterdam, The Netherlands
Telephone 31 - 20 592 9333, telex 12571 (mactr nl),
telefax 31 - 20 592 4199

CWI is the nationally funded Dutch institute for research in Mathematics and Computer Science.

Control of freeway traffic flow

S.A. Smulders

1991 Mathematics Subject Classification: 93A30.
ISBN 90 6196 451 2
NUGI-code: 811

Preface

The results described in this monograph were obtained during four years of research at the Centre for Mathematics and Computer Science (CWI) in Amsterdam. I would like to thank Jan van Schuppen, who has largely contributed to the successful completion of the research project. His continuous interest and valuable advice have been very stimulating.

Professor Huibert Kwakernaak has been my guide for the past six years, first during my study at the University of Twente and next during the project described here. I would like to thank him for his assistance all these years.

The research work presented here has been supported by the Technology Foundation. I would like to thank Frans van den Beemt, not only for the very generous support his organisation has given me, but also for his genuine interest in the research topic.

From the 'Gebruikerscommissie' associated with the project I would like to thank mr. A. Brinkman, prof. J.W. Cohen, J.J. Klijnhout and Martin van Maarseveen for their contribution to many a stimulating discussion.

Special thanks go to Jaap van Toorenburg and Henk Stipdonk from the Traffic Engineering Division of Rijkswaterstaat. Their assistance and their insight in traffic theory has been invaluable. Rijkswaterstaat is thanked for providing the necessary traffic data.

I would like to thank Paul de Zeeuw from the CWI for assisting me in solving the optimization problem described in Chapter 6 and for carrying out the required computations.

Contents

Chapter 1

Introduction

In 1896 notary J.P. Backx started the first automobile in the Netherlands and drove from Amsterdam to the Wieringermeer, a distance of approximately 50 km [96]. Although this first tour was a great success, the high maintenance cost and the technological knowledge required for repairs were the reasons that the automobile would be a rare means of transport till after World War I. Moreover, the existing road system was hardly suited for automobile traffic. Only a small part of the inter-urban roads had paved surfaces, and the ones that had were only three to four meters wide. Most of the roads were just sand or clay roads, metalled roads or they were of the macadam type, constructed according to the principles described by the Scottish engineer John L. MacAdam. The macadam roads, constructed from about 1830, at least allowed stage-coach traffic throughout the year. The construction of these roads was initiated by stage-coach proprietors, who were allowed to levy tolls as a compensation for their investments. The macadam principle allowed a fairly cheap improvement of roads. The paved roads were mainly constructed during the reign of King Willem I of the Netherlands. Louis Napoleon already built a paved road, however, to allow for a comfortable ride from Utrecht to the Loo, his hunting lodge.

The era of individual transport by private car that began after World War I necessitated the extension and improvement of the road system. The quality of the existing roads had deteriorated by the end of the nineteenth century as a result of the popularity of railway transport. During the first congress on roads, the 'Nederlandse Wegencongres' in 1920, road and transport problems were discussed by engineers and officials. This led to the levying of a yearly road tax on the use of a motorvehicle, which tax was put into a highway fund. This fund was used for the improvement of highways defined in a primary highway plan or 'Rijkswegenplan', and in provincial plans for secondary or tertiary highways. These plans had to be revised every ten years. The first highway plan was

presented in 1927 and was followed by others in 1937, 1948, 1958, 1968 and 1984. Using improved techniques of road construction, like asphalt surfacing, led to a rapid extension of the network, as Table 1.1 shows (Source: Dutch Ministry of Transport).

year	1940	1950	1960	1970	1975	1982
km	2283	2385	2585	3280	3902	4590

TABLE 1.1. Total length of primary roads in the Netherlands.

The idea of constructing roads especially for motor vehicles was first put into practice in Germany in 1921 [41]. The first road, reserved exclusively for motor vehicles, outside built-up areas, was the Köln-Bonn motor road, completed in 1932. This was the start of the freeway era which, for well-known reasons, provided Germany by 1939 with a transport network of 3500 km reserved for motor vehicles.

In the Netherlands the large scale construction of freeways only began after World War II. The first Dutch freeway, from Amsterdam to The Hague, was already constructed in 1938, however. Freeways were designed to suit fast and safe transport over large, inter-urban distances. They have slight gradients, gentle curves and long sight distances with no visual obstructions. Opposing streams of traffic are completely separated by a continuous central reservation. In order to allow for the smooth and safe flow of traffic at high speeds access to the freeway is widely spaced. It is provided by one of various forms of freeway interchange which give complete grade-separation and feed traffic through an acceleration lane on to the slow lane of the freeway and allow it to leave through a deceleration lane [61]. The rapid increase over the years of the length of the Dutch freeway network is clear from Table 1.2 (Source: Dutch Ministry of Transport).

year	1940	1950	1960	1970	1975	1980	1982	1986
km	87	121	333	981	1525	1798	1894	1960

TABLE 1.2. Total length of Dutch freeways.

Up to the 1970's there seemed to be no limitation to the increase. Then it was realized that further extension of the network could mean unacceptable damage to the natural environment. Also, public opinion and politics tended to favour public transport over individual means of transport, as this would limit the undesirable side-effects of air pollution and traffic noise. This tendency did not stop the further increase in the number of car owners or the increase in traffic intensity on the freeways, however. This is illustrated in Tables 1.3 and 1.4 (Source: Netherlands Central Bureau of Statistics).

year	1910	1920	1930	1939	1946	1950	1960	1970	1987
veh./1000 inh.	0.3	1.6	8.0	11.5	5.1	13.8	46	174	350

TABLE 1.3. Automobile ownership in the Netherlands.

year	1975	1980	1985	1986	1987
total	77	100	106	114	122
freeways	79	100	110	118	126

TABLE 1.4. Intensity indices of traffic outside built-up areas (1980 = 100).

The increase in intensity on freeways amounted to about 50% during the last decade, whereas the length of the freeway network only increased by just over 25% during the same period. As a result congestion more and more frequently occurs, with the annoying features of long travel times, a large number of accidents (mainly during the onset of congestion) and further environmental damage and economical loss.

To overcome the congestion problem a large set of measures have been proposed. Construction of new roads is acceptable only up to a certain extent, other measures such as the promotion of public transport, are important too. A promising way to reduce the oversaturation problem consists of increasing the efficiency of the existing transport network by means of supervisory control.

From the early 1950's the possibilities of control measures based upon mathematical descriptions of traffic flow were recognized by researchers from several disciplines and led to the development of a vehicular traffic science. An overview of a large range of traffic problems that have been considered since then is given by Tabak [89]. A historical overview is given by Gazis and Edie [35]. The earliest investigations were concerned with the determination of road capacity [39]. Attempts to model the behaviour of individual drivers and the stability of a string of vehicles were motivated by problems occurring in the Lincoln and Holland tunnels in New York City. It was observed that at peak periods of traffic demand the traffic throughput of the tunnel could decrease below its capacity and also that severe instabilities like *stop-start waves* could occur. This was thought to be undesirable because of safety and efficiency reasons and therefore research started with the aim of proposing control measures that would overcome the problems. These investigations started in the mid 50's by a General Motors research group and others [35] and led to the control of access to the tunnel by traffic signals. This enabled keeping the traffic density (number of vehicles per length of road) below a critical level and assured a reasonably smooth flow.

Other traffic problems occurred at intersections where heavy flows cross. Simple traffic rules were no longer sufficient to guarantee an efficient flow, and some co-ordinated action at a higher level is asked for. To overcome the problems

traffic signals were usually installed that allowed the different non-conflicting streams to cross the intersection cyclically. It was clear that in case the flows from the different streams are unbalanced, strategies could be designed for the length of the green phases of the traffic signals, that would lead to an improvement of the throughput and a reduction of average delay.

The first open-loop, off-line or time-of-day strategies indeed led to improvement. Next it was realized that further improvement could be obtained by means of closed-loop or demand-responsive strategies, as traffic flow is highly stochastic in nature and large, unpredictable fluctuations may occur. The use of partly demand-responsive strategies did not lead to substantial improvement, however, and truly demand-responsive strategies are required [33]. The problem in designing such strategies is to find an accurate model that produces good predictions of future traffic flows. A promising algorithm seems to be OPAC, developed by Gartner [34].

The traffic problem considered in this thesis is concerned with traffic flow on freeways. From about 1960 it was realized that problems similar to the ones occurring in tunnels could also occur on freeways, when traffic demand approaches capacity. Let us now consider the problem in more detail to explain the main characteristics.

As traffic on a freeway becomes more and more dense, drivers tend to compete for the available space and are forced to drive close to their predecessors [92]. A cautious driver who tries to keep at a safe distance, is forced to slow down frequently because less cautious drivers fill the gap he has left open. This competitive behaviour therefore leads drivers to drive so close and at such a high speed that small disturbances in the speed of one car quickly amplify and may finally lead to a standstill.

The stability of a string of vehicles was first investigated in connection with tunnel traffic [48]. It was found that to assure stability of the string, car spacing or time headway should not decrease below a critical value. Apparently, in dense freeway traffic drivers are forced to drive so close that, although their individual safety is still assured, this stability condition is no longer satisfied. Disturbances may occur for all kinds of reasons and therefore congestion is likely to result when traffic is dense. Again, as in the case of traffic at oversaturated intersections, some coordination at a higher level is required to preserve safety and efficiency.

In macroscopic terms the instability phenomenon may be explained by Figure 1.1, which depicts the relation between traffic *density* (the number of vehicles per km per lane) and traffic *intensity* (the number of passing vehicles per time unit) as it is observed in practice. As it turns out, when traffic is dilute, drivers drive at their preferred speed, without being influenced by other drivers. This leads to an almost linear increase of traffic intensity with increasing density. As traffic becomes more dense, however, drivers tend to slow down out of safety reasons, or because they are temporarily held up by slower drivers. These interactions between different drivers lead to a less steep rise of intensity as density increases. Finally, in even denser traffic the interactions become so frequent that each extra vehicle leads to a decrease of intensity. This argument leads to the parabolic like

shape of Figure 1.1.

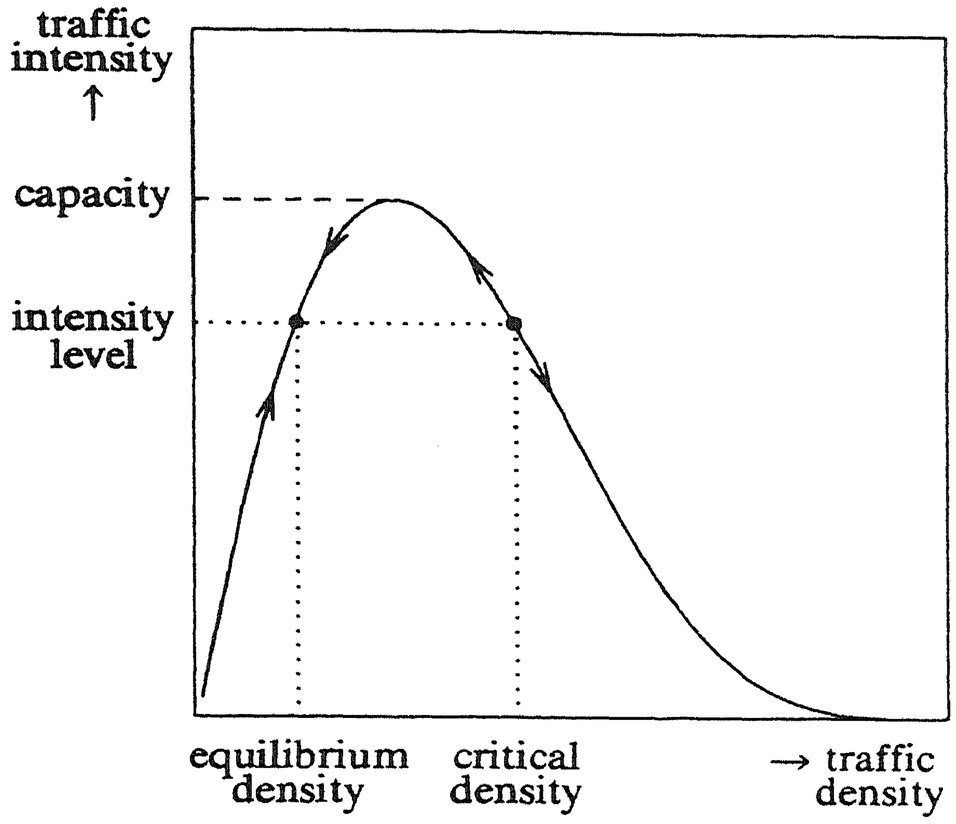

FIGURE 1.1. The fundamental diagram of freeway traffic.

The instability phenomenon may now be explained using the intensity-density relation. With each level of intensity an equilibrium value of traffic density is associated. When intensity is high, this density is high. Next, corresponding to the critical space headway of the microscopic analysis, there exists a critical density. Once density exceeds the critical level, traffic flow collapses with high probability and congestion sets in. The critical density value also depends on the level of intensity. As the latter obtains higher values, the equilibrium and the critical density value approach each other, until they meet at capacity level. If intensity exceeds capacity, due to on-ramp flows for example, congestion is sure to follow. Congestion not only occurs if intensity exceeds capacity, however, but usually sets in earlier. This is caused by the disturbances, the stochastic character of traffic flow, which allows the occurrence of large deviations from the equilibrium density. In general the disturbances die out as the density is continually attracted to the equilibrium value. This is represented by the arrows in Figure 1.1. A fluctuation may lead to exceeding the critical value, however. As said before, congestion will then result. The arrows in Figure 1.1 indicate this. As will be shown in Chapter 6, the probability of congestion is significant from intensity values as far as 15% below capacity.

Once congestion has set in, traffic flow is unstable and stop-start waves may occur, resulting in a decreased throughput and in unsafety. As it turns out the accident rate on a congested freeway is about twice the rate on a freeway with freely flowing traffic [49]. Recovery from the congested state only occurs after a period of low intensity.

It was already argued that exerting some kind of control may help to solve the problem. The most drastic measure consists of a fully automated highway in which each vehicle is controlled to keep at a safe distance from its predecessor. Research devoted to this subject is reported in the literature, but the results are not suitable for application yet. Less drastic measures are applied in practice, the most popular measure being *on-ramp control*. Here access to a freeway is limited by traffic signals on the on-ramps. This allows direct control over the number of merging vehicles and by maintaining the density on the freeway far enough below the critical level, congestion may be avoided. This problem has been investigated in a large number of papers and resulted in algorithms that are applied in practice, some with considerable success [41]. The on-ramp control measure of the Queen Elizabeth Freeway in Canada, for example, is reported to have led to an increase of throughput of 17% [49].

Another control measure consists of displaying *variable (speed) message signs* to the drivers. This is the type of control that will be considered in this thesis.

The problems on Dutch freeways stimulated the Dutch Ministry of Transport in 1975 to start a project to develop a Motorway Control and Signalling System with the following objectives [80]:

—detection of incidents such as an accident or congestion;
—better utilization of existing capacity;
—achieving greater safety;
—enabling road repair works in an easy fashion;
—allow the collection of data for research purposes.

The second objective is the one that may be achieved by exerting some kind of control as on-ramp regulation or control by means of variable message signs. The system was setup as displayed in Figure 1.2. Paired detection loops, embedded in the road surface every 500 meters, allow the registration of passing cars and measuring their speed. The gantries on which matrix type changeable message signs are mounted are spaced approximately 1000 m apart and allow the display of several message signs such as speeds ranging from 0 to 90 km/h, lane change and lane closure signs. In the system a large amount of data is processed locally, at the outstations, and the aggregated information is processed and analysed by the central computer and is available to the operators at the control centre. From there the message signs may be set to achieve the objectives mentioned. At the moment the system depicted in Figure 1.2 is installed on the A2/A12 freeways near Utrecht and on parts of the A13/A16/A19/A20 freeways around Rotterdam [80]. This amounts to a total length of controlled freeway of 80 km.

Currently the system is equipped with an Automatic Incident Detection algorithm (AID), which is installed locally and is meant to detect severe disturbances of the traffic stream, such as accidents or congestion. The algorithm automatically generates suitable signs for the oncoming traffic and turns out to be successful in reducing the number of accidents considerably. The occurrence of congestion cannot be avoided by this measure, however.

Experiments with a control scheme called *homogenization* were carried out in

1983 [93]. Here one and the same speed sign is displayed on a fixed set of gantries for all lanes. The idea is to achieve a traffic stream that is homogeneous with respect to speed, distribution over lanes, etc.

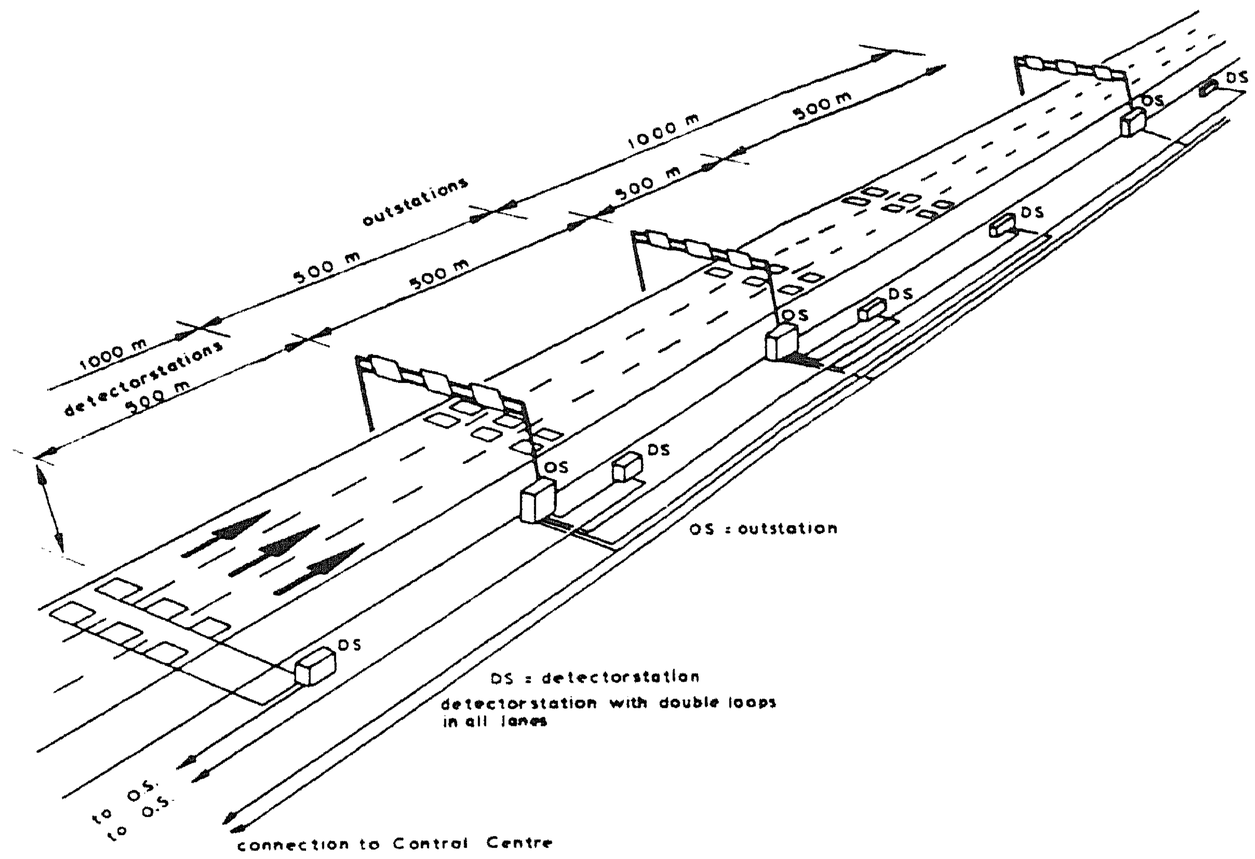

FIGURE 1.2. The Dutch Motorway Control and Signalling System.

In 1983 an evaluation study was completed in which it was concluded that the Dutch Motorway Control and Signalling System is successful in substantially decreasing the number of accidents (-24%) and in increasing the stability of the traffic stream [49]. One of the recommendations was that the homogenization measure should be incorporated in the system in an automatic fashion. In Chapter 6 of this thesis an automatic algorithm is proposed for the homogenization measure.

To be able to study the relative benefits of different control schemes it is necessary to have a detailed model of freeway traffic available. Traffic flow on freeways is a highly irregular process that one can only hope to model approximately. The complexity of the process becomes clear when one tries to describe all the different manoeuvres a driver may take [2]. For single lane freeways or on roads with no overtaking the car-following models, developed for tunnel traffic may be appropriate, or the process may be modelled as a queueing system [68]. For multilane freeway traffic macroscopic models were introduced, representing the traffic stream as a flow in which waves occur [56, 81]. These macroscopic models have the advantage of averaging out individual driver behaviour, which may result in a process that is better to predict.

The flow model consists of a conservation equation and a flow-density relation. This model was extended by Whitham [99] who added a momentum equation for the evolution of the speed of the traffic stream. Payne derived the model of Whitham from car-following models using an aggregation argument, and proposed discretization schemes for its use in practice. This model has become the most popular for describing freeway traffic and still seems to be the most promising one, allowing extensions to account for lane changing, weaving and merging phenomena [64]. Lane changing models have been considered by Michalopoulos and others [67] and merging due to a lane drop was studied by Cremer and May [21].

Notwithstanding the results reported so far, major modelling questions are still unanswered and require further study. This is not the subject of this thesis, however. We are mainly concerned with the development of an algorithm that produces estimates of the state of traffic from the available measurements (filter), and in the development of control policies for variable message signs. These will be based upon the macroscopic model of Payne in a slightly different, stochastic setup, as proposed and developed by Van Maarseveen [62]. Cremer developed a different setup which forces the discretization of time, and leads to a more or less artificial introduction of randomness in the traffic density process [20]. Van Maarseveen describes the theoretical setup of the freeway traffic model and derives approximations to the theoretically optimal filter. The filter was tested against simulated data. An identification procedure for the estimation of model parameters from measurement data was developed and applied as well. In these investigations only count data, the registration of cars passing, were used, speed measurements were omitted.

After presenting Van Maarseveen's model in Chapter 2 and extending it to account for the speed measurements, the model is evaluated for some basic traffic situations by means of stochastic simulation in Chapter 3. In Chapter 4 the measurement data are analysed to obtain estimates of several model parameters. In Chapter 5 the freeway traffic model is used as a basis for an approximate filter, which is then extensively tested against simulated as well as real traffic data. Finally, in Chapter 6 the problem of controlling the variable message signs of the motorway control system is addressed. Chapter 7 gives a summary of the results and suggestions for further research.

Chapter 2

Modelling of Freeway Traffic Flow

In this chapter we present a stochastic model of traffic flow. We start by giving some introductory remarks in Section 2.1. In the next section the stochastic differential equations of the model are derived. The observations or measurements are modelled in Section 2.3.

2.1. Introduction

The numerous freeway traffic models presented in the literature may be classified as micro- or macroscopic. In microscopic models the behaviour of individual drivers is described explicitly. In these models it is often assumed that drivers react to a difference in speed in comparison to the driver right in front by accelerating or decelerating. This is generally referred to as *car following* [36]. The intensity of the reaction may depend on the distance between the two vehicles, on perceptual or on other factors. Computer programs for the simulation of such models have appeared in the literature, [100], for example. A major drawback of this type of model in application to freeway traffic is the amount of computation needed. In dense traffic, when there are 25 veh/km/lane on the freeway on the average, the simulation of one carriageway of a dual two-lane freeway with a length of 10 km would require the solution of a set of 500 differential equations, one for each vehicle. In this case macroscopic models, in which freeway traffic is described in terms of aggregate variables, seem to be more appropriate. For a typical macroscopic model 40 differential equations would suffice in the example mentioned above.

Considering the quality of microscopic and macroscopic models it is clear that in the latter a large amount of detail is lost. Detailed conclusions about traffic characteristics like headways cannot be drawn. However, it would not be correct to state that microscopic models are better beforehand. The problem with these models is in describing the behaviour of individual drivers, which is extremely

difficult. In macroscopic models the individual behaviour of several drivers is aggregated, which may improve predictability.

The macroscopic model as developed by Payne [77] is the most widely used, see for example Cremer [20] , Papageorgiou [75] , Van Maarseveen [62, 63]. An earlier version of the model was investigated by Lighthill and Whitham [56]. Payne derived his model by starting from a general microscopic car following model and applying an aggregation procedure. The model obtained consists of two partial differential equations for the variables density and speed and may be interpreted as a fluid flow model:

$$\begin{cases} \dfrac{\partial \rho}{\partial t} + \dfrac{\partial q}{\partial x} = 0 \\ \dfrac{\partial v}{\partial t} + v\dfrac{\partial v}{\partial x} = -\dfrac{1}{T}\left[v - v^e(\rho)\right] + \dfrac{\dot{v}^e(\rho)}{2\rho T}\dfrac{\partial \rho}{\partial x} \end{cases} \tag{2.1}$$

where

$\rho(x,t)$: density (veh/km/lane)

$v(x,t)$: speed (km/h)

$q(x,t)$: flow (veh/h/lane)

and

$$q(x,t) = \rho(x,t)\cdot v(x,t).$$

Here x is the space variable, the distance from a fixed origin along the freeway, and t is the time variable. The first equation of (2.1) represents the conservation-of-vehicles principle and the second equation describes the evolution of speed in time. The speed is supposed to be influenced by two processes: relaxation (the first term on the right-hand side) and anticipation (the second term on the right-hand side). This is explained in more detail in Section 2.2. Several discretization schemes have been proposed to solve the equations numerically. For our purpose it seems to be natural to discretise with respect to space such that a carriageway of the freeway is split up into sections with measuring loops at the boundaries of each section, see Figure 2.1.
We define

$\rho_i(t)$: the number of vehicles in section i at time t

per length of road per lane (density)

$v_i(t)$: the mean speed of the vehicles in section i at time t

L_i : the length of section i

l_i : the number of lanes of section i.

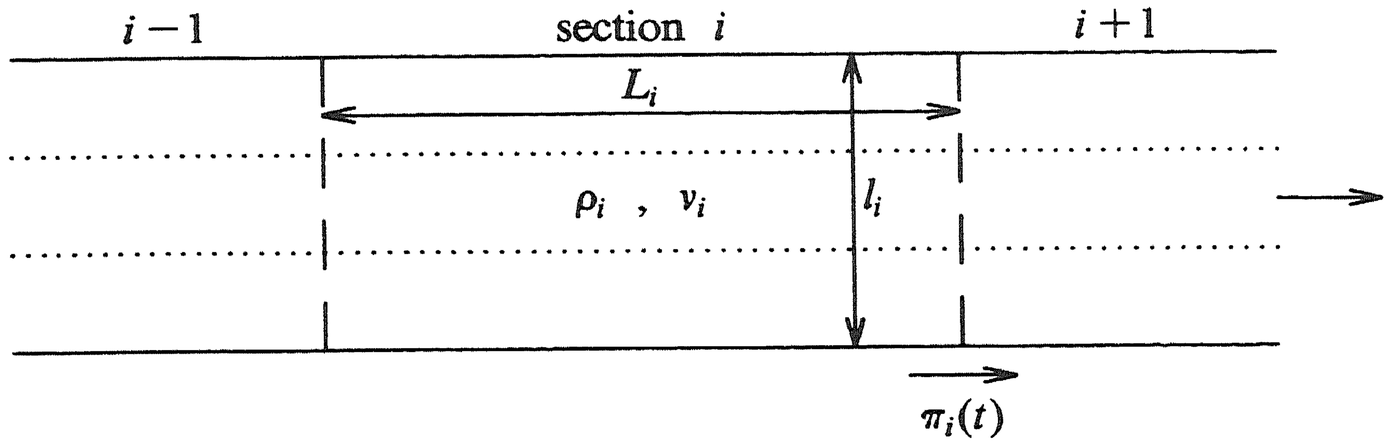

FIGURE 2.1. One carriageway of a freeway.

Our model has the following general form:

$$d\rho_i(t) = f_i(\ \dots\ , \rho_i(t), v_i(t),\ \dots\ )dt + dz_i(t) \tag{2.2}$$

$$dv_i(t) = g_i(\ \dots\ , \rho_i(t), v_i(t),\ \dots\ )dt + dw_i(t) \tag{2.3}$$

where w_i and z_i represent stochastic fluctuations, for $i = 1, \dots, n$ when n is the number of freeway sections we consider, and for $t \geqslant t_0$.

The measurements provided by the signalling system consist of passing times and passing speeds. Now introduce

$\pi_i(t)$: the number of vehicles that left section i since t_0, at time t.

This is a *counting process* for which a large amount of theory has been developed [13, 17]. A short overview of the main definitions and results is given in the Appendix. According to one of these results, the *Doob-Meyer decomposition* of submartingales, we may write

$$d\pi_i(t) = \lambda_i(t)dt + dm_i(t) \tag{2.4}$$

where

λ_i : the intensity of the process π_i

m_i : the martingale associated with π_i.

The decomposition (2.4) holds with respect to the family of σ-algebras $\{\mathcal{F}_t,\ t \geqslant t_0\}$ where

$\mathcal{F}_t$: the smallest σ-algebra generated by all stochastic processes involved.

Intuitively, $\mathcal{F}_t$ represents the past. The variable λ_i may be interpreted as the conditional mean of the number of jumps of π_i per unit of time. The mean is

conditioned on the past $\mathcal{F}_t$. Hence λ_i is the mathematical representation of traffic volume. The decomposition (2.4) shows that $\lambda_i(t)$ is the instantaneous rate of increase of $\pi_i(t)$ at time t, apart from some stochastic fluctuations represented by $m_i(t)$. In the simplest case $\lambda_i(t)$ will be a constant in which case $\pi_i(t)$ is a Poisson process with parameter $\lambda = \lambda_i$. In general λ_i will depend on time or on other stochastic processes, however. In our case it is obvious that λ_i depends on the density and mean speed of the sections i and $i+1$.

In the next section we use the decomposition (2.4) in the derivation of the differential equation for the density.

2.2. State equations

As mentioned in the previous section our model has the general form (2.2), (2.3). In this section we present the exact equations.

Following the definition of ρ_i, π_i and the other variables in Section 2.1 we find that

$$\rho_i(t) = \rho_i(t_0) + \frac{1}{l_i L_i}[\pi_{i-1}(t) - \pi_i(t)] + \frac{1}{l_i L_i}[r_i^{on}(t) - r_i^{off}(t)]$$

for $t \geqslant t_0$. Here we assume that a section contains at most one on- and one off-ramp:

r_i^{on} : the on-ramp counting process

r_i^{off} : the off-ramp counting process.

Also, we assume that the number of lanes is constant within a section. The equation above is just the principle of conservation of vehicles: the number of vehicles in a section only changes when vehicles enter or leave.

In differential form the conservation law looks like

$$d\rho_i(t) = \frac{1}{l_i L_i}[d\pi_{i-1}(t) - d\pi_i(t)] + \frac{1}{l_i L_i}[dr_i^{on}(t) - dr_i^{off}(t)].$$

Using the Doob-Meyer decomposition (2.4) we obtain

$$d\rho_i(t) = \frac{1}{l_i L_i}[\lambda_{i-1}(t) - \lambda_i(t)]dt + \frac{1}{l_i L_i}[dr_i^{on}(t) - dr_i^{off}(t)] + \frac{1}{l_i L_i}[dm_{i-1}(t) - dm_i(t)]. \tag{2.5}$$

As stated in Section 2.1, $\lambda_i(t)$ is the intensity or traffic volume of the counting process $\pi_i(t)$ at time t.

A common way [20] of modelling $\lambda_i(t)$ is by the convex sum approximation

$$\lambda_i(t) = \xi_i[\alpha\rho_i(t) + (1-\alpha)\rho_{i+1}(t)][\alpha v_i(t) + (1-\alpha)v_{i+1}(t)], \tag{2.6}$$

where α is a weighting factor and

$$\alpha \in [0,1].$$

Furthermore, ξ_i measures the width of the freeway at the boundary of section i

and $i+1$. If the number of lanes is equal in both sections, $\xi_i = l_i$, if the number changes one may take $\xi_i = \min(l_i, l_{i+1})$ to emphasize the bottleneck situation. A change of the number of lanes is not considered in the sequel and therefore $\xi_i = l_i$ always holds.

The approximation (2.6) is justified by the fact that in the case of stationary and homogeneous flow the relation is exact. (As ρ_i is measured in veh/km/lane and π_i counts over all lanes the factor l_i appears in (2.6)). In (2.6) the factor $[\alpha\rho_i(t)+(1-\alpha)\rho_{i+1}(t)]$ is an approximation of the density in the vicinity of the common boundary of section i and section $i+1$ and analogously $[\alpha v_i(t)+(1-\alpha)v_{i+1}(t)]$ an approximation of the speed. In general α is taken to be larger than 0.5 to emphasize the fact that only vehicles upstream of the boundary can cross this boundary.

Substituting (2.6) into (2.5) we finally obtain

$$d\rho_i(t) = \frac{1}{l_i L_i}\Big\{ l_{i-1}[\alpha\rho_{i-1}(t)+(1-\alpha)\rho_i(t)][\alpha v_{i-1}(t)+(1-\alpha)v_i(t)] - l_i[\alpha\rho_i(t)+(1-\alpha)\rho_{i+1}(t)][\alpha v_i(t)+(1-\alpha)v_{i+1}(t)] \Big\} dt + \frac{1}{l_i L_i}[dr_i^{on}(t) - dr_i^{off}(t)] + \frac{1}{l_i L_i}[dm_{i-1}(t) - dm_i(t)]. \tag{2.7}$$

The stochastic differential equation (2.7) describes the time evolution of the density in section i. This evolution depends on the density and mean speed in the section itself and in the neighbouring sections and is also susceptible to certain stochastic fluctuations.

This equation is the one used by Van Maarseveen [62]. In comparison with equation (2.38) of Cremer [20] our equation (2.7) differs in the following aspects:

—our equation is continuous in time whereas Cremer takes a time step of approximately ten seconds;
—the accuracy of the approximation in equation (2.32) of Cremer is influenced by the size of the time step. If this step is too small the approximation will be too rough but it cannot be taken too large either if one wants to model short term phenomena like instabilities. This problem does not occur in our approach as in our view λ_i is the (conditional) *mean* number of vehicles that pass per unit of time, where the mean is taken over all realizations;
—in our case the stochastics appear in a natural way: they follow from the character of a counting process. Cremer adds an artificial Gaussian random variable to the density equation.

We now present the second equation of our freeway traffic model. Our equation will be the discrete analogue of the speed equation of Payne [77].

As one can see from the second equation of (2.1) in Section 2.1 three processes are supposed to influence the mean speed of a section:

1. *Relaxation*: In practice there is a tendency of the mean speed to move to an equilibrium value v^e. The value of this equilibrium speed depends on the density: $v^e(\rho)$. The simplest model for this relation was given by Greenshields [39], and consists of a linear decrease of the mean speed with increasing density:

$$v^e(\rho) = v_{free}\left(1 - \frac{\rho}{\rho_{jam}}\right)$$

where

v_{free} : the equilibrium speed when density approaches zero (km/h)

ρ_{jam} : jam density (veh/km/lane).

We then propose to model the influence of the relaxation process on the mean speed as follows:

$$\left\{ dv_i(t) \right\}_{rel} = -\frac{1}{T}\left[v_i(t) - v^e(\rho_i(t)) \right] dt$$

where

T : relaxation time (h);

2. *Anticipation*: It is clear that drivers react to their immediate predecessors and tend to anticipate changing conditions downstream. The effect of this microscopic phenomenon on the evolution of mean speed is represented by the anticipation term of the model. Following Payne's derivation we obtain:

$$\left\{ dv_i(t) \right\}_{ant} = -\frac{\nu}{T(L_i + L_{i+1})}\left[\frac{\rho_{i+1}(t) - \rho_i(t)}{\rho_i(t) + c} \right] dt$$

where

ν : a positive constant (km^2/h)

c : a constant (veh/km/lane), a modification of Payne's model originating from Cremer;

3. *Convection*: The mean speed in a section is influenced by vehicles entering or leaving the section. From simple calculations it follows that

$$\left\{ dv_i(t) \right\}_{con} = \frac{1}{l_i L_i \rho_i(t_-) - 1}\left[v_i(t_-) - v^{out}(\pi_i(t)) \right] d\pi_i(t)$$
$$+ \frac{1}{l_i L_i \rho_i(t_-) + 1}\left[v^{in}(\pi_i(t)) - v_i(t_-) \right] d\pi_{i-1}(t).$$

Here

$v^{out}(k)$: the speed of the k-th vehicle leaving section i since t_0

$v^{in}(k)$: the speed of the k-th vehicle entering section i since t_0

t_- : the time just before a vehicle enters or leaves the section.

We neglect the convection effects of the ramps, assuming that these play a minor role. Now, as in the case of π_i in Section 2.1 we would like to decompose $\{ v_i(t) \}_{con}$ in a martingale part and a part which represents the behaviour in the mean. Unfortunately this *semimartingale decomposition* is very complex. As the convection effect only represents a small adjustment of the mean speed, some approximation seems acceptable. Instead of considering the continuum of all possible values that the passing speed can take just consider an approximate mean passing speed, $\alpha v_i + (1-\alpha)v_{i+1}$, where α is the same weighting factor as in (2.6). Using this in the equation for $\{ dv_i(t) \}_{con}$ and using the decomposition (2.4) for $d\pi_i(t)$ we find

$$\left\{ dv_i(t) \right\}_{con} \approx \frac{l_i[\alpha\rho_i + (1-\alpha)\rho_{i+1}][\alpha v_i + (1-\alpha)v_{i+1}]}{l_i L_i \rho_i(t_-) - 1}(1-\alpha)[v_i - v_{i+1}]dt$$

$$+ \frac{l_{i-1}[\alpha\rho_{i-1} + (1-\alpha)\rho_i][\alpha v_{i-1} + (1-\alpha)v_i]}{l_i L_i \rho_i(t_-) + 1}\alpha[v_{i-1} - v_i]dt$$

$$+ \frac{(1-\alpha)[v_i - v_{i+1}]}{l_i L_i \rho_i(t_-) - 1}dm_i(t) + \frac{\alpha[v_{i-1} - v_i]}{l_i L_i \rho_i(t_-) + 1}dm_{i-1}(t)$$

where the argument of ρ_{i-1}, v_{i-1}, ρ_i, v_i , ρ_{i+1}, v_{i+1} is t_-. To simplify matters choose α equal to 1, neglect the $+1$ and -1 in comparison to $l_i L_i \rho_i(t_-)$, set $\frac{\rho_{i-1}}{\rho_i}$ equal to 1 and neglect the martingales. This results in

$$\left\{ dv_i(t) \right\}_{con} \approx \frac{l_{i-1}}{l_i L_i} v_{i-1}[v_{i-1} - v_i]dt.$$

Other approximations are possible but will not be discussed here, because of the relatively small importance of convection.

Summarizing, the equation for the mean speed looks like:

$$dv_i(t) = \left\{ dv_i(t) \right\}_{rel} + \left\{ dv_i(t) \right\}_{ant} + \left\{ dv_i(t) \right\}_{con} + dw_i(t)$$

$$= -\frac{1}{T}\left[v_i(t) - v^e(\rho_i(t)) \right]dt - \frac{\nu}{T(L_i + L_{i+1})}\left[\frac{\rho_{i+1}(t) - \rho_i(t)}{\rho_i(t) + c} \right]dt$$

$$+ \frac{l_{i-1}}{l_i L_i} v_{i-1}(t)[v_{i-1}(t) - v_i(t)]dt + dw_i(t) \qquad (2.8)$$

where $w_i(t)$ is a Brownian motion process that has been added to represent stochastic fluctuations.

Comparing (2.8) to equation (2.38) of Cremer [20] one notes that the equations are much alike, apart from the fact that we work in continuous instead of discrete time. There is only a small difference in the convection term owing to a different approximation.

Equation (2.7) and (2.8) for $i = 1,...,n$ build up our model of freeway traffic flow and are the same as in the model of Van Maarseveen [62].

In Van Maarseveen's investigations the parameters T, ρ_{jam}, v_{free}, ν, c were identified using real traffic data from the signalling system. Only passing time information was used in the estimation procedure, the speed measurements were neglected. The weighting parameter α appearing in (2.6) was chosen beforehand to be equal to 0.5. The counting errors and acceleration noise were assumed to be zero. The estimation procedure resulted in the parameter values presented in Table 2.1.

parameter	value	unit
T	0.0044	h
ρ_{jam}	116.0	veh/km/lane
v_{free}	106.0	km/h
ν	138.0	km^2/h
c	10.0	veh/km/lane

TABLE 2.1. The model parameters.

In the model the variables ρ_0, v_0, ρ_{n+1}, v_{n+1} are not defined. Suitable boundary conditions at the entrance and at the exit of the freeway stretch have to be chosen to obtain a closed set of equations. As noted by Cremer [20] this is not an easy task. In some traffic situations the values at the boundary are determined by the internal state of the system while in other situations the boundary values determine the state. In stationary traffic flow, for instance, the entrance values ρ_0, v_0 (or to be more clear: the entrance intensity) determine(s) the evolution of the density and mean speed of all the other sections including the imaginary $(n+1)^{th}$ one. When there is a bottleneck downstream, the roles are interchanged, however. In that case the values at the exit determine the values in all other sections.

We now discuss several possible choices of boundary conditions.

1. Prescribed intensity

$$\text{entrance}: \quad \rho_0 = \left[\frac{\lambda_0}{l v_1} - (1-\alpha)\rho_1 \right]/\alpha$$

$$v_0 = v_1$$

$$\text{exit}: \quad \rho_{n+1} = \rho_n$$

$$v_{n+1} = \left[\frac{\lambda_n}{l \rho_n} - \alpha v_n \right]/(1-\alpha)$$

where $\lambda_0(t)$ and $\lambda_n(t)$ are given as functions of time;

2. Stationarity

$$\text{entrance}: \quad \rho_0 = \rho_1, \quad v_0 = v_1$$

$$\text{exit}: \quad \rho_{n+1} = \rho_n, \quad v_{n+1} = v_n;$$

3. Extrapolation

$$\text{entrance}: \quad \rho_0 - \rho_1 = \epsilon(\rho_1 - \rho_2), \quad v_0 - v_1 = \epsilon(v_1 - v_2)$$

$$\text{exit}: \quad \rho_{n+1} - \rho_n = \delta(\rho_n - \rho_{n-1}), \quad v_{n+1} - v_n = \delta(v_n - v_{n-1})$$

where $\epsilon, \delta \in [0,1]$.

We have experimented with the first two possibilities and found that prescribing the entrance intensity and choosing a stationarity condition at the exit gives satisfactory results in general. Choosing a stationarity condition at the entrance leads to unrealistically large fluctuations in the densities. This is caused by the fact that in case of stationarity conditions the model possesses a continuum of equilibrium points. For all density levels an equilibrium point exists and the density may be driven to high values by the martingales in (2.7). Prescribing the entrance intensity results in an isolated equilibrium point, however, to which the density is attracted. Prescribing the exit intensity leads to large fluctuations in the density of the last section, which then influence the whole freeway stretch. We did not experiment with extrapolation conditions.

In the next chapter model simulations are carried out. These lead to modifications and parameter changes. Before presenting these, modelling the observations is discussed in next section.

2.3. Observation equations

As explained in the introductory chapter, the signalling system does not provide the exact values of density and mean speed but only registers the passing of vehicles and their speed at the measuring sites. To be able to control the speed signals of the system in an optimal way we need a good estimate of the state of traffic at all time instants. Our task therefore is to develop a filter that estimates traffic density and mean speed from the measurements. This will only be possible if we have a good model for the relation between the measurements and the traffic state. This modelling is discussed here. We start by modelling the passing times and in the second part we model the passing speeds.

In the previous sections we already introduced the counting processes $\pi_i(t)$ ($i = 1, \ldots, n$), which count the number of vehicles that pass section boundaries from time t_0 on. We also gave a decomposition of these processes

$$d\pi_i(t) = \lambda_i(t)dt + dm_i(t), \tag{2.9}$$

which describes the dynamics of the process as the sum of a mean evolution term and a stochastic term. For the intensity process λ_i we used the approximate relation

$$\lambda_i = l_i[\alpha\rho_i + (1-\alpha)\rho_{i+1}]\cdot[\alpha v_i + (1-\alpha)v_{i+1}]. \tag{2.10}$$

If there would be no measuring errors (2.9) and (2.10) together would satisfy our purposes: there is a clear relation between the counting observations and the traffic state variables. Unfortunately there are measuring errors, mainly due to vehicles changing lanes at the measuring sites: we may *miss* a vehicle passing or we may have a *false count*. In order to model these errors introduce

$n_i(t)$: the number of vehicles that were *observed* leaving section i since t_0, at time t

$e_i^m(t)$: the number of vehicles we missed leaving section i since t_0, at time t

$e_i^f(t)$: the number of false counts at the common boundary of section i and section $i+1$ since t_0, at time t.

Then

$$n_i(t) = \pi_i(t) + e_i^f(t) - e_i^m(t). \tag{2.11}$$

Next we need to make some assumptions about the nature of the error processes: how often do errors occur, do they depend on traffic volume, etc. In absence of any detailed information we assume

$$de_i^f(t) = \epsilon_i^f\lambda_i(t)dt + dr_i^f(t) \tag{2.12}$$

$$de_i^m(t) = \epsilon_i^m\lambda_i(t)dt + dr_i^m(t) \tag{2.13}$$

where

$$\epsilon_i^f, \epsilon_i^m \in [0,1]$$

and $r_i^f(t)$, $r_i^m(t)$ are martingales. Hence a fixed fraction of the vehicles that pass is supposed to be missed and there is a fixed fraction of false counts. These fractions ϵ_i^m, ϵ_i^f are be much smaller than one in general. Together with (2.9) and (2.10) we now have as our observation equation:

$$\begin{aligned} dn_i(t) = {} & (1+\epsilon_i^f-\epsilon_i^m)\, l_i[\alpha\rho_i(t)+(1-\alpha)\rho_{i+1}(t)][\alpha v_i(t)+(1-\alpha)v_{i+1}(t)]dt \\ & + d[m_i + r_i^f - r_i^m](t). \end{aligned} \tag{2.14}$$

A direct way of modelling the passing speeds would be to introduce the *jump* or *mark accumulator* processes

$$y_i(t) = \sum_{k=0}^{\pi_i(t)} v_i(k) \tag{2.15}$$

where

$v_i(k)$: the passing speed of the k-th vehicle at site i

and $v_i(0)=0$. In contrast to the jumps of the process π_i, the jumps of y_i do not have unit size but a size equal to the speed of the passing vehicle. Note that from the process $y_i(t)$ we can deduce both passing times and passing speeds. We neglect measurement errors for the moment.

It turns out that a filter based on the observation processes (2.15) is too complicated to implement: it leads to an integro-differential equation requiring the integration over all possible passing speeds at every time instant. We therefore discretise the domain of passing speeds in m classes:

$$V^1, V^2, \ldots, V^m$$

where

$$V^j = [\, v^{j-1}, v^j \,), \qquad \text{for } j=1, \ldots, m$$

with $v^0 < v^1 < \cdots < v^m$. Now introduce the counting processes

$\pi_i^j(t)$: the number of vehicles that left section i after t_0

with passing speed in class V^j

for i=0, ... ,n and j=1, ... ,m.

Again, the processes $\pi_i^j(t)$ contain the information of passing times:

$$\pi_i(t) = \sum_{j=1}^{m} \pi_i^j(t)$$

Applying the Doob-Meyer decomposition for submartingales as mentioned in Section 2.1 to π_i^j we obtain

$$d\pi_i^j(t) = \lambda_i^j dt + dm_i^j(t) \tag{2.16}$$

where λ_i^j is the intensity and m_i^j is the martingale. As before, λ_i^j depends on the traffic density and mean speed, for

$$\lambda_i = \sum_{j=1}^{m} \lambda_i^j = l_i[\alpha\rho_i + (1-\alpha)\rho_{i+1}][\alpha v_i + (1-\alpha)v_{i+1}].$$

The problem now is to distribute the total intensity λ_i over the speed classes. In studying this problem we came up with two possible solutions.

In the first approach instead of considering $\{\rho_i, v_i\}$ as the traffic state we consider

$\rho_i^j(t)$: the number of vehicles in section i per km per lane

that at time t have a speed that falls in class V^j.

Instead of the general model (2.2), (2.3) we then have

$$d\rho_i^j(t) = f_i^j(\ldots, \rho_i^j(t), \ldots)dt + dz_i^j(t) \tag{2.17}$$

for $i=1, \ldots, n$ and $j=1, \ldots, m$. The modelling of the intensity is easy:

$$\lambda_i^j = l_i[\alpha\rho_i^j + (1-\alpha)\rho_{i+1}^j](v^{j-1}+v^j)/2$$

in analogy with (2.6). The speed $(v^{j-1}+v^j)/2$ is taken to be the representative of class V^j. But how to find a reasonable expression for f_i^j in (2.17)? Note that in fact we are trying to model the dynamic behaviour of an approximation of the speed distribution, because

$$\rho_i^j(t) / \sum_{j=1}^{m} \rho_i^j(t)$$

is the fraction of vehicles driving in class V^j at time t. Attempts to model this behaviour have been made, originating from Prigogine [71]. At the moment there does not seem to be a satisfactory model of this type. The original model by Prigogine has some serious defects and improved versions turn out to be much more complicated [1, 76]. Hardly any practical investigation with these models has been reported [29].

Summarizing we may state that introduction of the variables ρ_i^j complicates matters considerably whereas the benefits are doubtful. We therefore not proceed in this way but opt for another approach which is discussed next.

One may assume that the passing speed of a vehicle at a fixed location along the freeway is a stochastic variable with an absolutely continuous distribution function

$$F(v) = \int_{-\infty}^{v} f(u)du.$$

Using the speed classes $V^1, V^2, \ldots, V^m$ the *fractions* γ^j in each class may be defined as follows:

$$\gamma^j = \int_{v^{j-1}}^{v^j} f(u)du.$$

γ^j is the expected fraction of cars that pass the location with speed in class V^j. See Figure 2.2 for an example explaining the meaning of the fractions γ^j. Whenever $F(v^0)>0$ or $F(v^m)<1$ a slight correction is required:

$$\gamma^j = \int_{v^{j-1}}^{v^j} f(u)du + \frac{1}{m}[1+F(v^0)-F(v^m)]$$

in order to achieve

$$\sum_{j=1}^{m} \gamma^j = 1.$$

It seems reasonable to assume that the fractions depend on the mean speed and the density of the neighbouring sections only:

$$\gamma_i^j = \gamma_i^j(\rho_i, \rho_{i+1}, v_i, v_{i+1}). \tag{2.18}$$

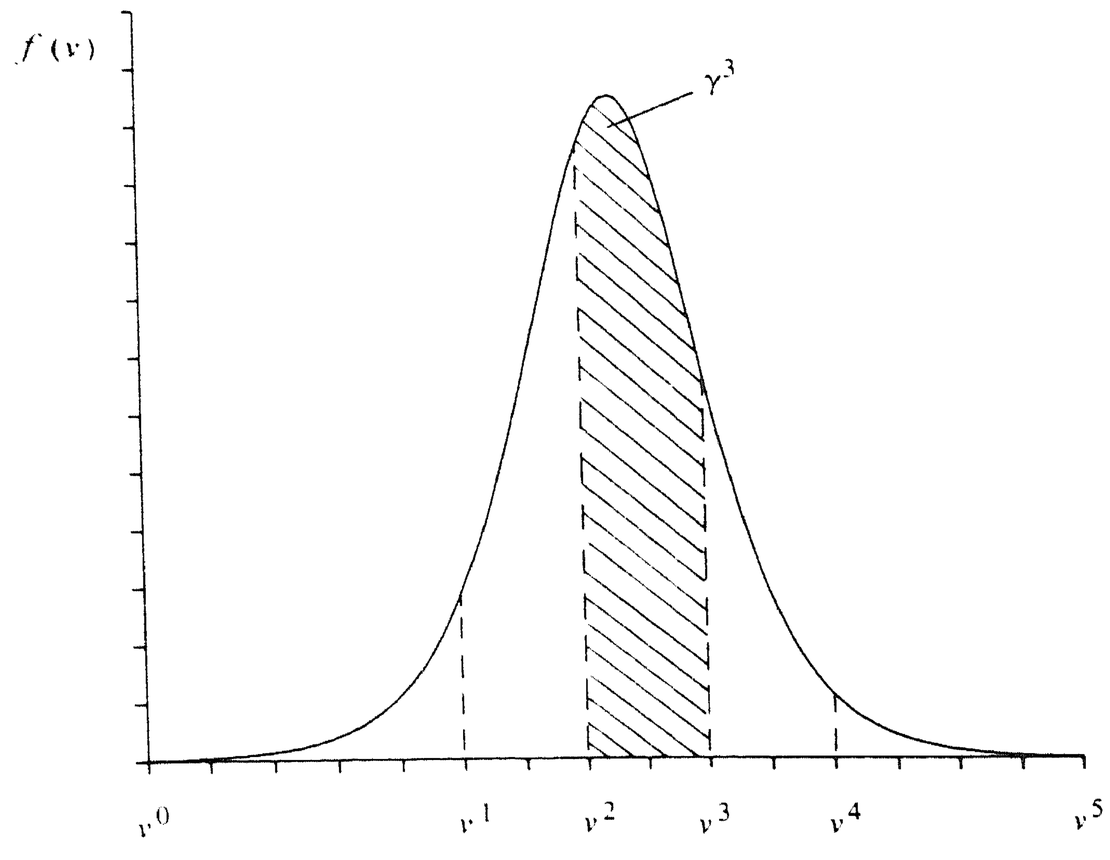

FIGURE 2.2. An example of a fraction γ^j.

Then we may write

$$\lambda_i^j = \gamma_i^j \lambda_i.$$

Note that again we are making assumptions about the behaviour of the speed distribution. Whether the assumption (2.18) holds, even under stationary conditions, has to be checked by analysing real traffic data. This is investigated in Chapter 4. The analysis will give us some insight into the nature of the relation between the form of the distribution and the traffic density and mean speed. We expect the distribution of passing speeds per lane to be approximately normal and the distribution aggregated over all lanes to have several peaks: one for each lane. Furthermore, the variance may depend on the density.

Substituting these results into (2.16) leads to the following expression:

$$d\pi_i^j(t) = \gamma_i^j \cdot l_i[\alpha\rho_i(t)+(1-\alpha)\rho_{i+1}(t)][\alpha v_i(t)+(1-\alpha)v_{i+1}(t)]dt + dm_i^j(t). \quad (2.19)$$

Note that γ_i^j sums to 1 over j and we have

$$\sum_{j=1}^{m} \lambda_i^j = l_i[\alpha\rho_i^j+(1-\alpha)\rho_{i+1}^j][\alpha v_i^j+(1-\alpha)v_{i+1}^j] = \lambda_i.$$

As noted before the processes π_i^j contain the information of passing times and so (2.19) is not an extra observation equation next to (2.14) but (2.19) *replaces* (2.14).

To conclude, a remark concerning measurement errors is required. There are two types of errors: counting errors and speed measurement errors. Analogously to the derivation of (2.14) we find

$$dn_i^j(t) = [1+\epsilon_i^f(j)-\epsilon_i^m(j)]\gamma_i^j\lambda_i(t)dt \; + \; d[m_i^j+r_i^f(j)+r_i^m(j)](t) \qquad (2.20)$$

where

$n_i^j(t)$: the number of vehicles that were *observed* leaving section i with speed in class V^j, at time t

$\epsilon_i^f(j)$: the fraction of false counts in class V^j at site i

$\epsilon_i^m(j)$: the fraction of missed vehicles in class V^j at site i

and

$$\sum_{j=1}^{m}\epsilon_i^f(j) = \epsilon_i^f, \quad \sum_{j=1}^{m}\epsilon_i^m(j) = \epsilon_i^m.$$

As to the speed measurement errors we assume that these are small enough in comparison to the width of the speed classes to justify neglection.

Chapter 3

Simulation of Freeway Traffic Flow

In this chapter a simulation study is undertaken of the freeway traffic model that has been developed in Chapter 2. The aim of simulation is twofold. First, simulations may lead to insight in the dynamic behaviour of the model. Second, this behaviour may be compared to the behaviour of traffic in practice. The simulation results may lead to modifications of the model to improve its performance.

In the introductory section the simulation procedure is described. In the following sections simulations of stationary, unstable and congested traffic are considered. These simulations lead to a modified model, which is summarized in the last section.

3.1. Introduction

The freeway traffic model presented in Chapter 2 consists of a set of stochastic differential equations for the traffic density and mean speed of each section of one carriageway, which are the *state variables* of the model. In addition to these equations there are differential equations for the observations: vehicle counts and speed measurements.

Once initial values have been chosen for density and mean speed and boundary conditions are defined the set of equations can be solved numerically, simulating the stochastics by some random number generation procedure. The behaviour of the model and its validity may thus be studied. If the model does not show a realistic behaviour it may be adjusted. Thus, by means of trial-and-error the model may be refined. The model is further validated in Chapter 5 by means of the filter described there. An additional way to obtain realistic parameter values consists of the optimization of some goodness-of-fit criterion: model identification. Van Maarseveen [62] applied a maximum likelihood identification procedure, the results of which are given in Table 2.1. Model simulations

presented in [62] showed realistic behaviour globally. It is the purpose of this chapter to carry out detailed simulations to assess the ability of the model to represent some basic traffic situations in a realistic fashion. The model will serve as a basis for the filter of Chapter 5 and should at least be able to describe stationary, low density traffic and also be able to represent the instabilities that occur in practice when density is higher.

To simulate the model a computer program was written in CDC FORTRAN 5 (which complies with ANSI FORTRAN 77) and run on a CDC Cyber 750 mainframe. Starting from the initial state $[\rho_1(t_0), v_1(t_0), \cdots, \rho_n(t_0), v_n(t_0)]$ the first jump time of the counting process vector $[\pi_0(t), \cdots, \pi_n(t)]$ is generated. Such a jump corresponds to a vehicle passage over one of the section boundaries. The differential equations (2.7), (2.8) are integrated from t_0 to the first jump time by Euler's method in one or more steps. A maximal step size is to be specified by the user. The use of Euler's method is justified by the fact that in our case the integration error is of second order in h instead of the usual first order (where h is the integration step) [84]. At the jump time the densities corresponding to the freeway sections involved in the vehicle passage are updated and the next jump time is generated.

The counting processes are generated using the so called *thinning* method developed by Lewis and Shedler [55, 73]. This method allows an *exact* generation of realizations whereas other methods are approximate.

The simulation of a homogeneous counting process with constant intensity λ, a Poisson process, is easy: it may be done by independently drawing from an exponential distribution with expectation $1/\lambda$. The numbers drawn are the lengths of the intervals between successive jumps of the counting process.

The counting processes in our model are described by the following equation:

$$d\pi_i(t) = l_i[\alpha\rho_i(t)+(1-\alpha)\rho_{i+1}(t)][\alpha v_i(t)+(1-\alpha)v_{i+1}(t)]dt + dm_i(t),$$

where $\rho_i(t)$, $v_i(t)$ are stochastic processes described by the differential equations (2.7), (2.8). The process $\rho_i(t)$ depends on the past of the counting processes π_{i-1} and π_i: the passage of vehicles over section boundaries. This implies dependence of π_i on its own past. Counting processes with an intensity depending on the past of the process itself are called *self exciting processes*. Counting processes with an intensity depending on other stochastic processes are called *doubly stochastic processes*. When the intensity is a deterministic nonconstant function of time the process is called *nonhomogeneous*.

For the simulation of nonhomogeneous counting processes several procedures have been proposed. For a short overview see Lewis and Shedler [55]. Lewis and Shedler themselves propose a procedure which suits our purposes well. It amounts to the thinning of a realization of a process with higher (and simpler) intensity. The procedure is based on the following theorem.

THEOREM 3.1 [Lewis & Shedler]. *If $t_1, t_2, \ldots, t_{n_T}$ are the jump times in the interval $(t_0, T]$ of the one-dimensional nonhomogeneous counting process $\{n_t,\ t \geq t_0\}$ with intensity $\{\lambda(t),\ t \geq t_0\}$ and every jump time t_i is deleted with probability*

$$1 - \frac{\bar{\lambda}(t_i)}{\lambda(t_i)}$$

for some function $\{\bar{\lambda}(t),\ t \geqslant t_0\}$ *with* $\bar{\lambda}(t) \leqslant \lambda(t)$ *on* $(t_0, T]$ *then the remaining times form a realization on* $(t_0, T]$ *of a nonhomogeneous process* $\{\bar{n}(t),\ t \geqslant t_0\}$ *with intensity* $\{\bar{\lambda}(t),\ t \geqslant t_0\}$.

The theorem may be applied in the simulation of a nonhomogeneous process $\{\bar{n}_t,\ t \geqslant t_0\}$ with intensity $\{\bar{\lambda}_t,\ t \geqslant t_0\}$ as follows:

1. Choose a constant λ, large enough for $\bar{\lambda}(t) \leqslant \lambda$ to hold on the interval to the next jump of $\{n_t,\ t \geqslant t_0\}$, a counting process with intensity λ;
2. Generate the first jump of the homogeneous process $\{n_t,\ t \geqslant t_0\}$;
3. Accept the jump with probability $\bar{\lambda}(t)/\lambda$;
4. Repeat the Steps 1 to 4.

According to the theorem the stochastic process obtained from this procedure is equivalent to $\{\bar{n}_t,\ t \geqslant t_0\}$.

In general, when there is no known upper bound for $\bar{\lambda}(t)$, one cannot be sure that $\bar{\lambda}(t) \leqslant \lambda$ will hold on the interval to the next jump of n_t. The length of this interval depends on λ. So one has to estimate λ. The estimate should not be too large because this would lead to many rejections in Step 3 ($\lambda \gg \bar{\lambda}$), which is inefficient. The estimate should not be too small either for if it turns out that $\bar{\lambda} > \lambda$ somewhere in the interval one has to repeat this step until $\bar{\lambda} \leqslant \lambda$ does hold.

The method may easily be extended for higher-dimensional counting processes.

The procedure described above may be applied to doubly stochastic and self exciting counting processes in an analogous way. The justification for this has been given by Ogata [73] and is not repeated here.

3.2. Stationary traffic

To obtain a first impression of the validity of the model we start by simulating traffic flow at low density: approximately 15 veh/km/lane. This is achieved by taking the entrance intensity constant in time, equal to the equilibrium value of 1383.3 veh/h/lane. The two-lane carriageway consists of 12 sections of 500 m each. The model parameters are chosen according to Table 2.1. In Table 3.1 the initial state is presented along with the results after 2 minutes of simulation (The column indicated by $\nu = 138$).

There are some very large differences in the mean speed of the different sections at time 0.033 h: the speed ranges from 0 to 141 km/h over the freeway stretch. Furthermore, it appears that in the sections where traffic is most dense speed is highest: compare sections 3 and 4 for example. This does not seem to be very realistic. In case of stationary flow one would not expect such large fluctuations in the mean speed, and in denser regions one would expect traffic to flow at a lower speed. It is the anticipation term that forces the speed to behave in the

	$t = 0.0$ h		$t = 0.033$ h					
			$\nu = 138$		$\nu = 40$		$\alpha = 0.85$	
section	ρ	v	ρ	v	ρ	v	ρ	v
1	15.0	92.2	12.0	61.5	13.0	83.9	13.0	93.1
2	15.0	92.2	11.0	109.9	15.0	93.6	12.0	94.1
3	15.0	92.2	1.0	0.0	8.0	79.8	12.0	93.3
4	15.0	92.2	37.0	141.0	23.0	96.3	15.0	94.4
5	15.0	92.2	0.0	1.1	1.0	67.6	12.0	92.0
6	15.0	92.2	23.0	124.2	32.0	93.0	13.0	92.8
7	15.0	92.2	8.0	61.6	10.0	100.0	11.0	88.9
8	15.0	92.2	18.0	102.8	0.0	69.8	19.0	90.2
9	15.0	92.2	15.0	114.7	22.0	80.6	18.0	93.9
10	15.0	92.2	8.0	14.6	25.0	92.5	15.0	97.1
11	15.0	92.2	23.0	115.8	12.0	92.4	5.0	79.5
12	15.0	92.2	17.0	104.2	14.0	91.9	21.0	84.7

TABLE 3.1. Simulation results: stationary traffic.

observed unrealistic manner. Anticipation forces drivers to accelerate when downstream traffic is less dense. Apparently in our model this effect is too strong.

In identifying the parameters of the model Van Maarseveen [62], pp. 177 – 178 used a data set which did not contain enough information to estimate the anticipation factor accurately. Identification of discrete time models by Grewal and Payne [40] and Cremer and Papageorgiou [22] led to much smaller anticipation factors than Van Maarseveen's. We therefore choose a smaller value for this factor. At the moment our choice is a bit arbitrary, we just choose a value that gives reasonable results. We now repeat the previous simulation with $\nu = 40.0$ km^2/h and all other parameter values unchanged. In Table 3.1 the results are presented (The column indicated by $\nu = 40$).

This time the speed is behaving smoothly and the differences between the sections are not as large as before. The density shows very large fluctuations, however. One of the sections is empty for a period of time whereas a neighbouring one contains 32 vehicles. Although large fluctuations are to be expected the ones found appear to be too extreme. This unrealistic behaviour is mainly caused by the fact that when a section is almost empty but its downstream neighbour is not, the intensity at their common boundary is still relatively high. This intensity is modelled as the product of weighted density and mean speed:

$$l_i[\alpha\rho_i(t)+(1-\alpha)\rho_{i+1}(t)][\alpha v_i(t)+(1-\alpha)v_{i+1}(t)]$$

Up until now we took α to be 0.5. Note that this parameter was not estimated by Van Maarseveen [62] but chosen beforehand. By taking a larger value we may increase the influence of the upstream density and thereby decrease the intensity at the boundary when the upstream density is low. We now repeat the previous

simulation with $\alpha=0.85$. See Table 3.1 and Figure 3.1 for the results. The behaviour of density and mean speed has clearly improved.

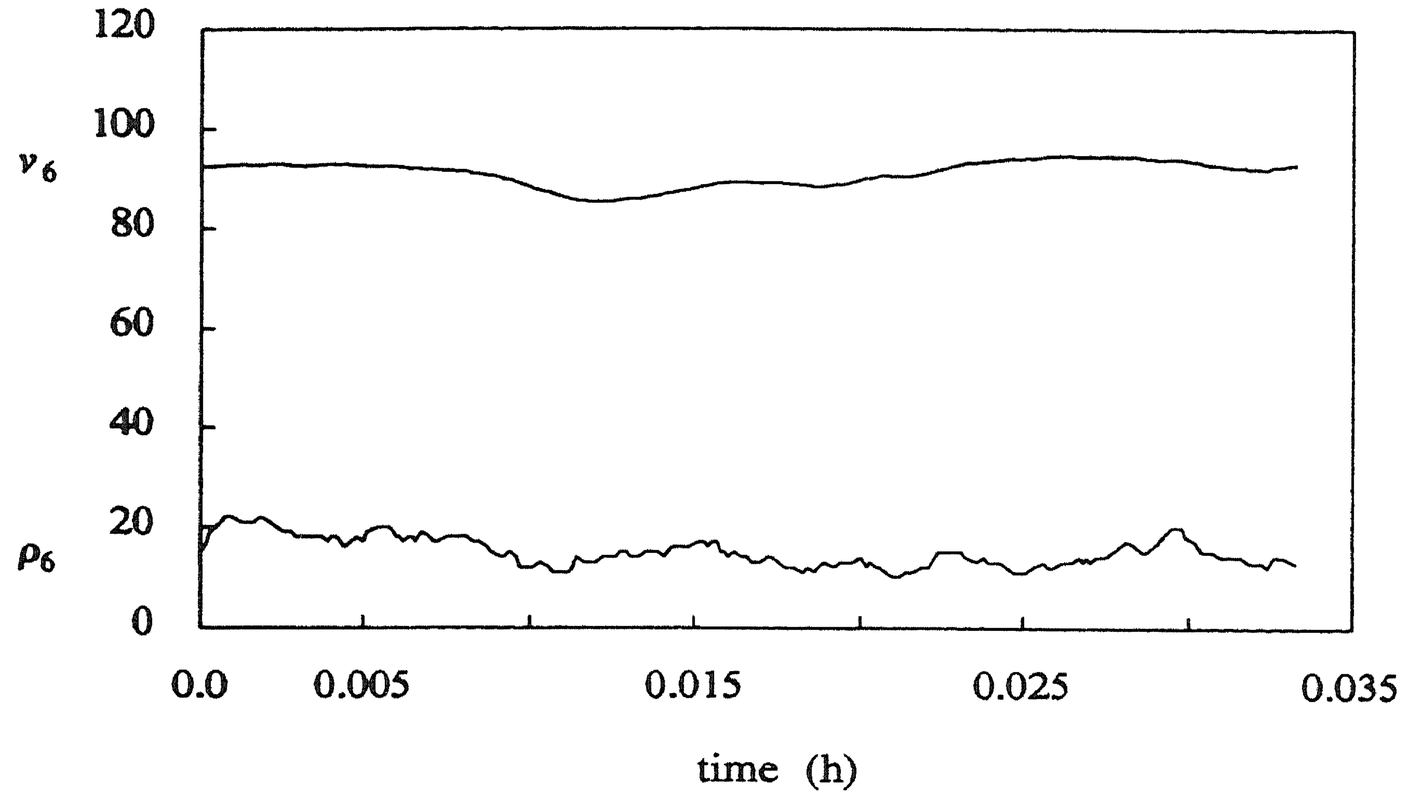

FIGURE 3.1. Simulation realization with $v=40$, $\alpha=0.85$.

The observed stabilizing effect of increasing α is confirmed by a stability analysis of the model. The analysis consists of considering the deterministic version of (2.7), (2.8) (omit the stochastic terms), and a linearization of this model around an equilibrium point. The most interesting equilibrium points are given by $\rho_i=\rho$ and $v_i=v^e(\rho)$ for $i=1,...n$, where ρ is defined as the solution of $l\rho v^e(\rho)=\lambda_0$. For $\lambda_0<\max_\rho l\rho v^e(\rho)$ two equilibrium points are obtained. Other equilibria may exist but are not considered here. The result of the analysis for the model with $\rho=$ 15 veh/km/lane and $v=$ 92.2 km/h is that for values of α of 0.5 or larger the model is stable but for $\alpha=0.1$ for example, the model is unstable.

The simulation experiments of this section have resulted in an estimate of the value of α. Comparison with real data is required to obtain an optimal parameter estimate. This will not be carried out here. We conclude that the model with the given parameter values describes low density traffic in a reasonable way.

3.3. UNSTABLE TRAFFIC

The validity of the model at moderate density values, $20-30$ veh/km/lane, is now tested. Note that in practice the critical density, at which the traffic stream becomes unstable, is approximately 25 veh/km/lane. Again one carriageway of a dual two-lane freeway of 12 sections with a length of 500 m each is considered. Parameters are chosen as in the last simulation of the previous section. We start

with 20 veh/km/lane in all freeway sections and choose the entrance intensity λ_0 to be constant in time and equal to 4000 veh/h. In equilibrium this corresponds to a density of approximately 24 veh/km/lane.

All realizations we produced show stable traffic behaviour analogous to the behaviour at low density. The density nevertheless reached a value of about 30 veh/km/lane for a considerable period of time. In practice traffic flow is highly irregular and unstable in such a situation. One would therefore expect the model to show instabilities in at least some of the realizations. Stability analysis of the deterministic model, obtained from (2.7), (2.8) by deleting the stochastic terms, confirms the simulation results in that it shows the model to be stable for densities up to 92 veh/km/lane. For $\lambda_0 = 4000$ veh/h we obtain an equilibrium point at $\rho = 23.7$ veh/km/lane and one at 92.3 veh/km/lane. Computation of the eigenvalues of the linearized model shows the first point to be stable and the second to be unstable. Determining the exact domain of attraction of the stable equilibrium point is a difficult problem in general, but it is clear from the analysis that, starting at a point near the stable equilibrium, the probability of congestion is small when $\lambda_0 = 4000$ veh/h. Only in case stable and unstable equilibrium points are close instabilities of the traffic stream may be expected. This means that λ_0 will have to be close to the *capacity* of the freeway, defined by $\lambda^{cap} = \max_{\rho} l\rho v^e(\rho)$. In our case:

$$v^e(\rho) = 106.0[1-\rho/116.0], \qquad 0 \leqslant \rho \leqslant 116. \tag{3.1}$$

Hence

$$\lambda^{cap} = 6148 \text{ veh/h}, \quad \text{at } \rho = 58 \text{ veh/km/lane}.$$

In Figure 3.2 the equilibrium relations between speed and density and between intensity and density are depicted.

We conclude that the equilibrium relation $v^e(\rho)$ has to be modified to achieve unstable behaviour at a realistic density level. In fact, the relation used until now is the simplest one, already introduced in 1934 by Greenshields [39]. Since then a large amount of literature has appeared concerning the equilibrium relationship between density, speed and intensity [20, 46]. It turns out to be difficult to choose between the various relations proposed. We therefore propose a simple modification of (3.1) which suits our purposes:

$$v^e(\rho) = \begin{cases} v_{free}\left[1 - \dfrac{\rho}{\rho_{jam}}\right], & 0 \leqslant \rho \leqslant \rho_{crit} \\ \\ d\left[\dfrac{1}{\rho} - \dfrac{1}{\rho_{jam}}\right], & \rho_{crit} \leqslant \rho \leqslant \rho_{jam} \end{cases} \tag{3.2}$$

where d is chosen in such a way that v^e is continuous at ρ_{crit}. We will take $\rho_{jam} = 110.0$ veh/km/lane, $v_{free} = 110$ km/h and $\rho_{crit} = 27.0$ veh/km/lane, which leads to $d = 2970.0$ 1/h. The above relation is linear in the low density region and decreases fast enough for densities larger than the critical value to prevent the capacity λ^{cap} to become too large. See Figure 3.2. From equation (3.2) it

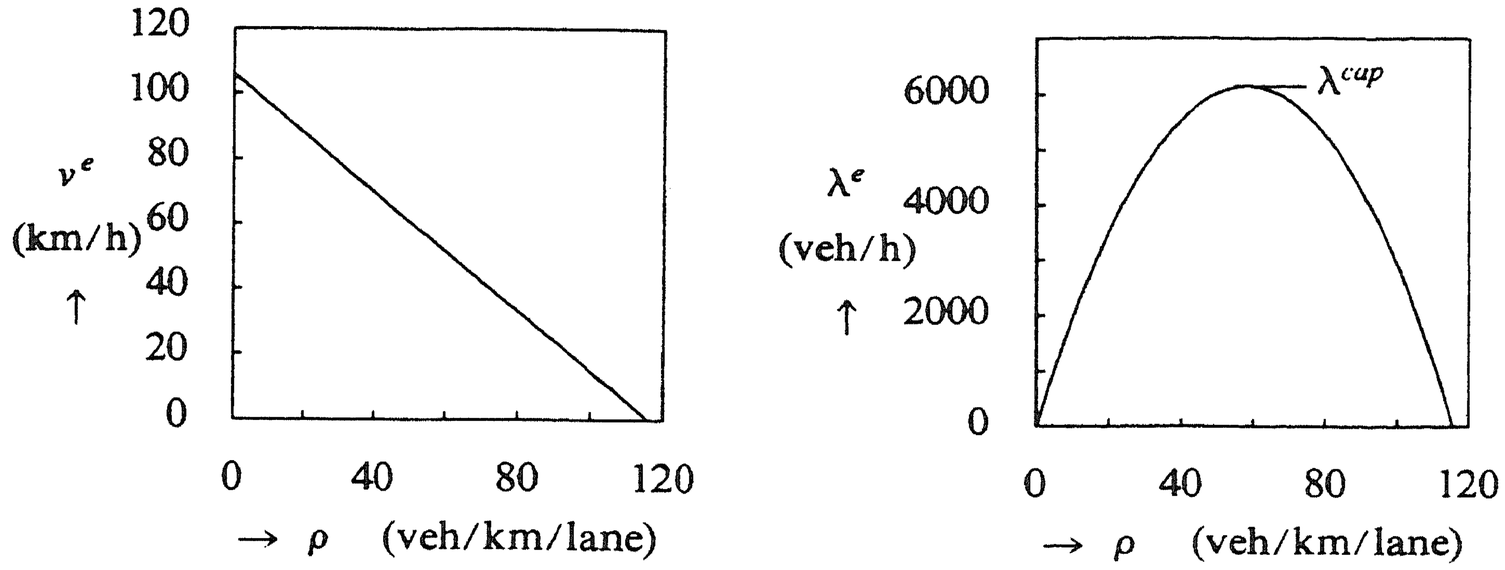

FIGURE 3.2. The equilibrium relations corresponding to (3.1).

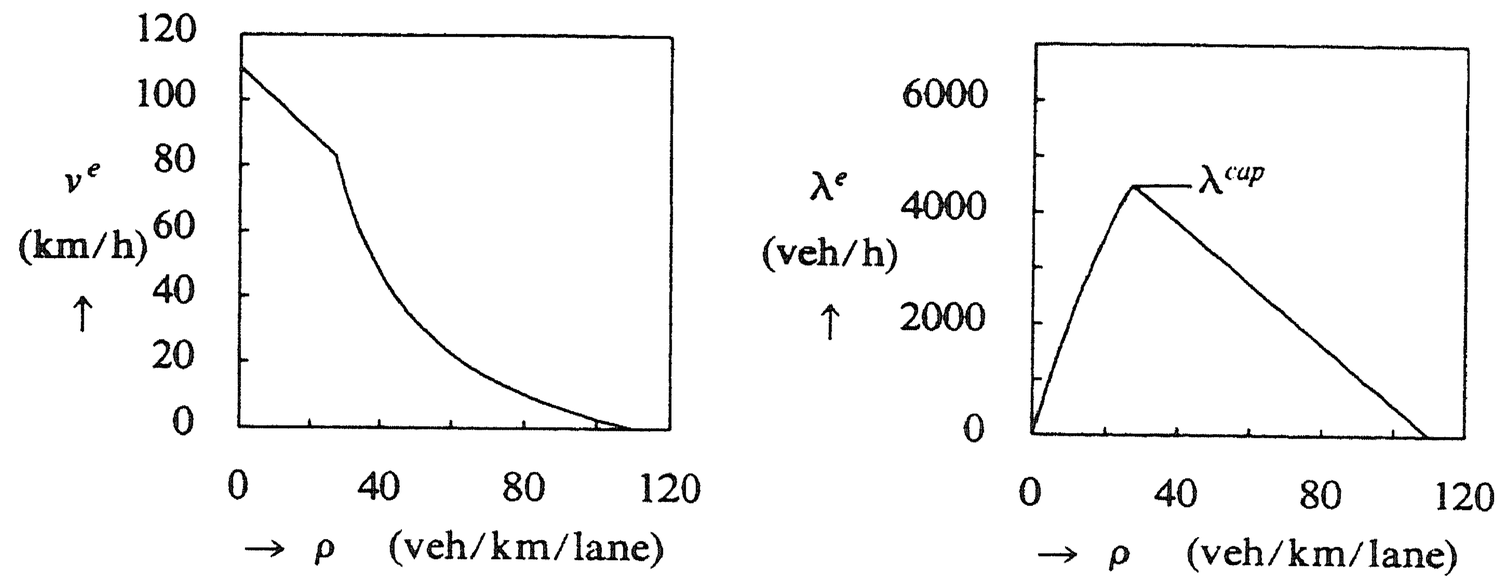

FIGURE 3.3. The equilibrium relations corresponding to (3.2).

follows that

$$\lambda^{cap} = 4482 \text{ veh/h at } \rho = 27 \text{ veh/km/lane.}$$

To motivate the proposed relation further, consider a string of vehicles in one lane and suppose that each driver maintains a distance to his immediate predecessor of

$$le + di + vT_r$$

kilometers, where

le: the average length of a vehicle (km)

di: the minimally acceptable distance between two vehicles standing still (km)

T_r: the reaction time of a driver (h)

v: the driving speed (km/h).

If all vehicles drive at this distance and at speed v then we find that

$$\rho = \frac{1}{le + di + vT_r} \quad \text{veh/km/lane}$$

which leads to

$$v = \frac{1}{T_r}[\frac{1}{\rho} - (le + di)].$$

This relation is therefore expected to be the equilibrium relation for high density values. It equals (3.2) when $d = 1/T_r$ and $\rho_{jam} = 1/(le + di)$. For the proposed values of d and ρ_{jam} we obtain

$$T_r = 1.2 \text{ seconds}$$

$$di + le = 9.1 \text{ meters},$$

which appear to be realistic [36].

Let us now return to the freeway traffic model. Stability analysis shows the stable equilibrium point to be at $\rho = 23.0$ veh/km/lane and the unstable one at 35.9 veh/km/lane. Excursions to values around 35 veh/km/lane occur and therefore instabilities are present in some realizations.

The results of a simulation of the modified model are presented in Figures 3.4, 3.5 and 3.6.

A collapse of the traffic stream occurs at approximately 0.08 h and the density increases to 40 veh/km/lane and speed decreases to 60 km/h. The disturbance starts in section 1 and then moves downstream with a speed of approximately 35 km/h. We conclude that the form of the equilibrium relation plays a major role in the stability properties of the model and that our choice leads to realistic behaviour.

3.4. Congestion

The model as developed until now satisfies our needs, because we are mainly interested in traffic flow for densities below the critical value. It may nevertheless be interesting to look at an extreme traffic situation: almost complete standstill at maximal density. We consider one carriageway of a dual two-lane freeway of 12 sections, 500 m each, and start from a traffic situation where there is a complete standstill at the end of the stretch, while in the first sections traffic is still moving at high speed. The results are displayed in Table 3.2 (The column indicated by (3.3)) and Figure 3.7.

The results are highly unrealistic: vehicles pile up in sections 7, 8 and 9 and density reaches values up to 216 veh/km/lane. This means that the space available for a single vehicle reduces to 4.6 m on the average. Some of the larger cars may get into trouble here. The jam density was modelled at 110 veh/km/lane. In the upstream sections there is no effect at all: drivers keep driving at about

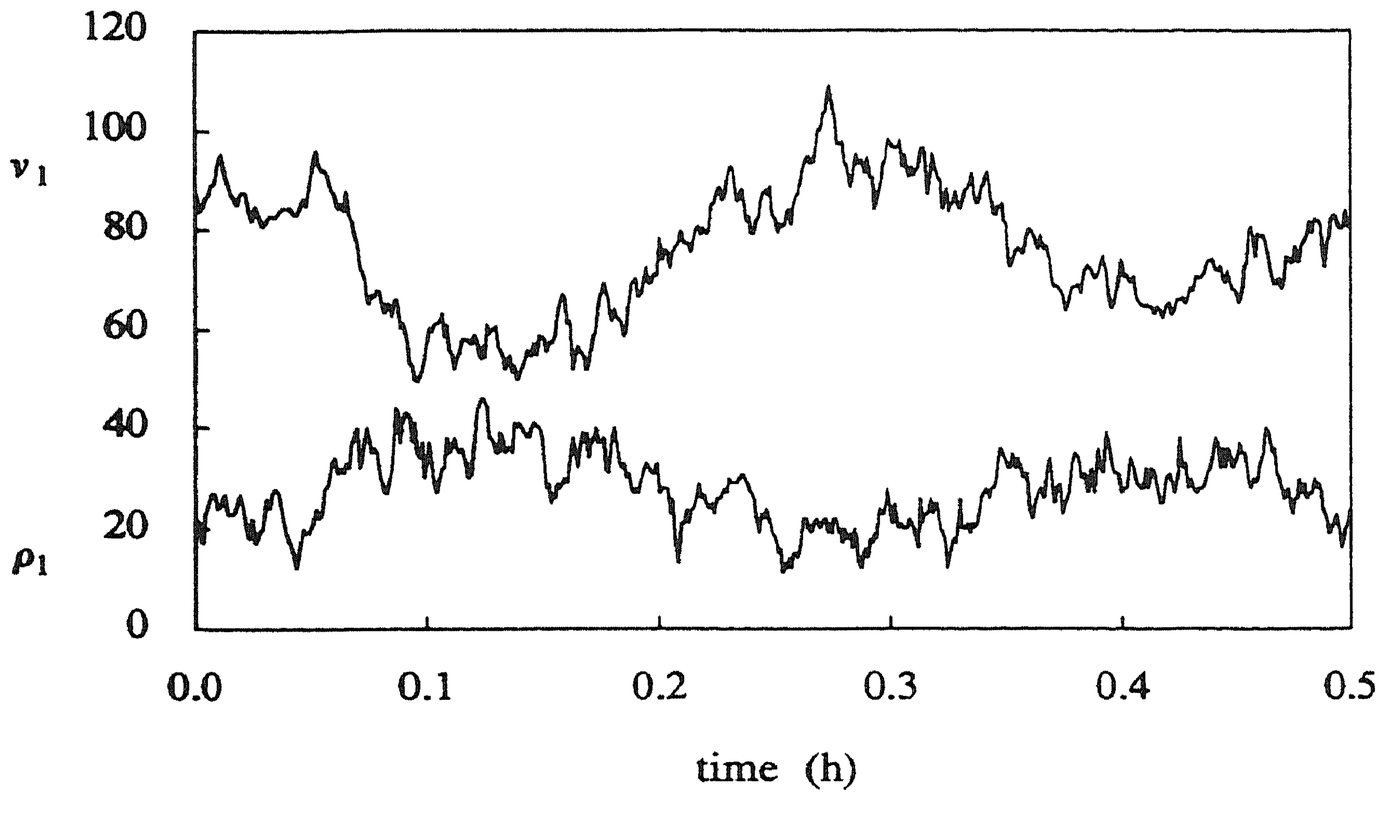

FIGURE 3.4. A collapse of the traffic stream: section 1.

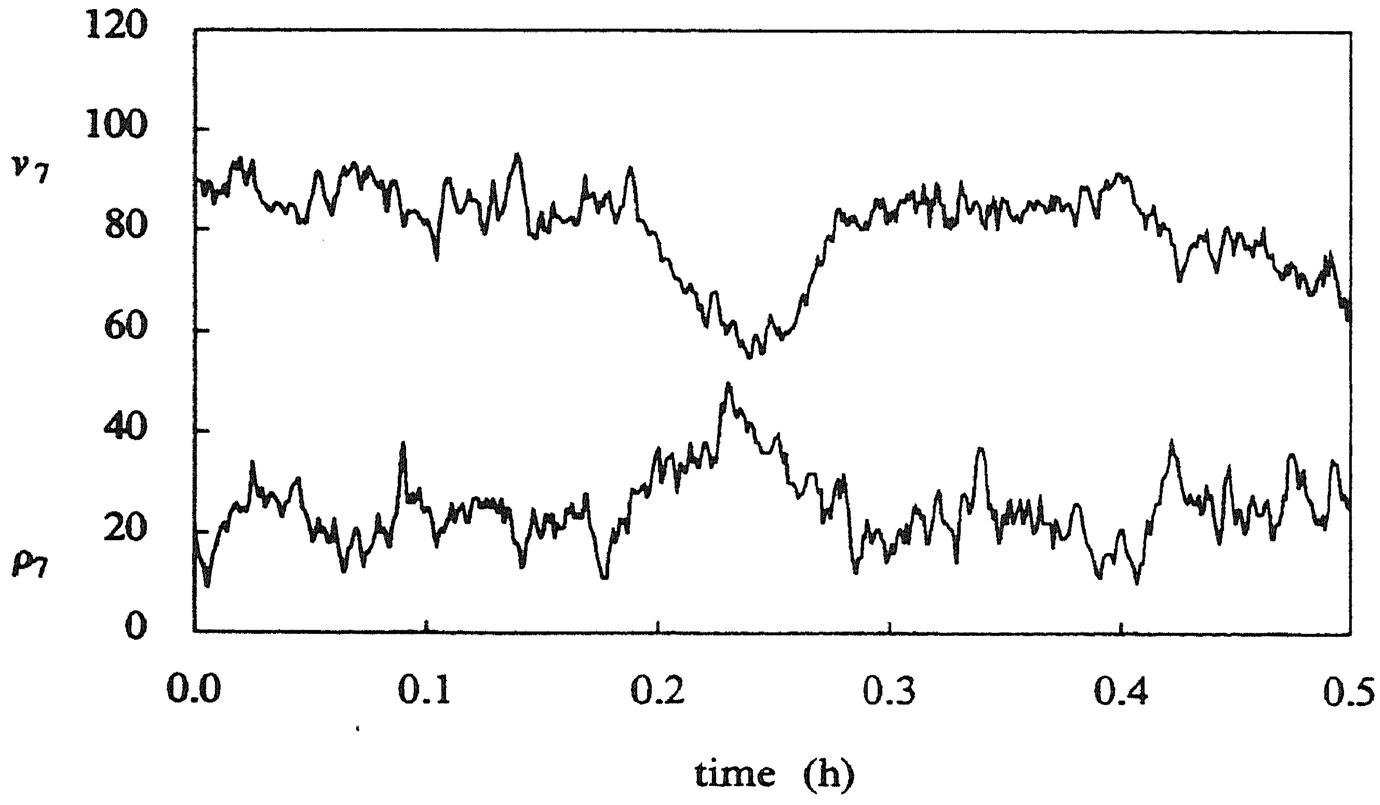

FIGURE 3.5. A collapse of the traffic stream: section 7.

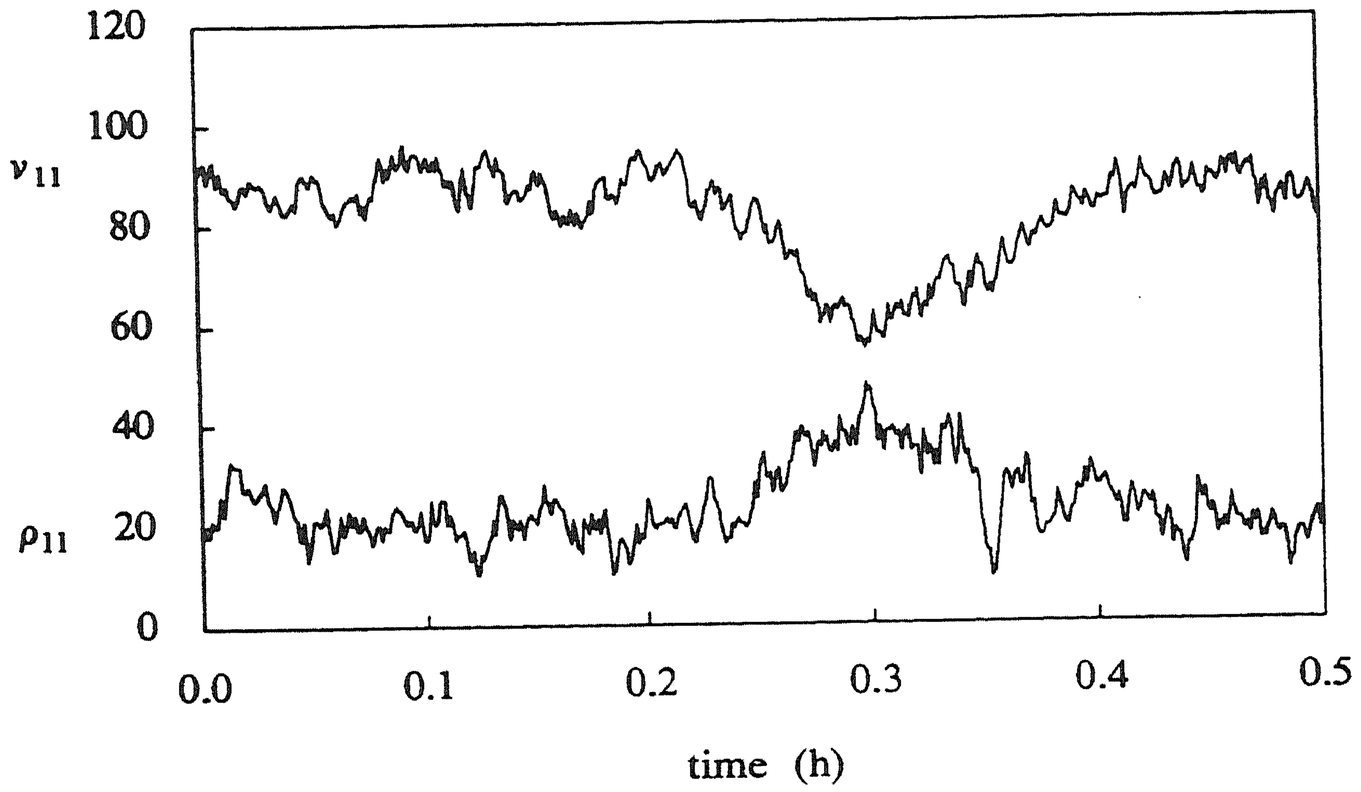

FIGURE 3.6. A collapse of the traffic stream: section 11.

their equilibrium speed.

In practice one would expect the congested region to grow in the upstream direction and density not to exceed the jam value. It is clear that the anticipative effect in the model is too small. Recall that the anticipative effect was weakened in Section 3.2 to avoid unrealistically large speed fluctuations at low densities. It seems that at high densities anticipation should be strong, however.

The above results suggest to model anticipation in such a way that the effect increases with increasing density. The term as it is in our present model:

$$\{ dv_i(t) \}_{ant} = -\frac{\nu}{(L_i + L_{i+1})T} \left[\frac{\rho_{i+1} - \rho_i}{\rho_i + c} \right] dt \qquad (3.3)$$

has just the opposite effect. If the constant c would be zero, which it is in the original model of Payne, the anticipative effect would become infinite as $\rho_i \to 0$. Instead of (3.3) we now propose

$$\{ dv_i(t) \}_{ant} = -\gamma (L_i l_i)^2 [\beta \rho_i + (1-\beta)\rho_{i+1}][\rho_{i+1} - \rho_i] dt \qquad (3.4)$$

where

γ : a constant of dimension km/h^2

β : a weighting factor $\in [0, 1]$.

In (3.4) the factor $[\beta \rho_i + (1-\beta)\rho_{i+1}]$ is an approximation of the prevailing density on the freeway stretch near section i. If $\beta = 1$ the emphasis is on the density

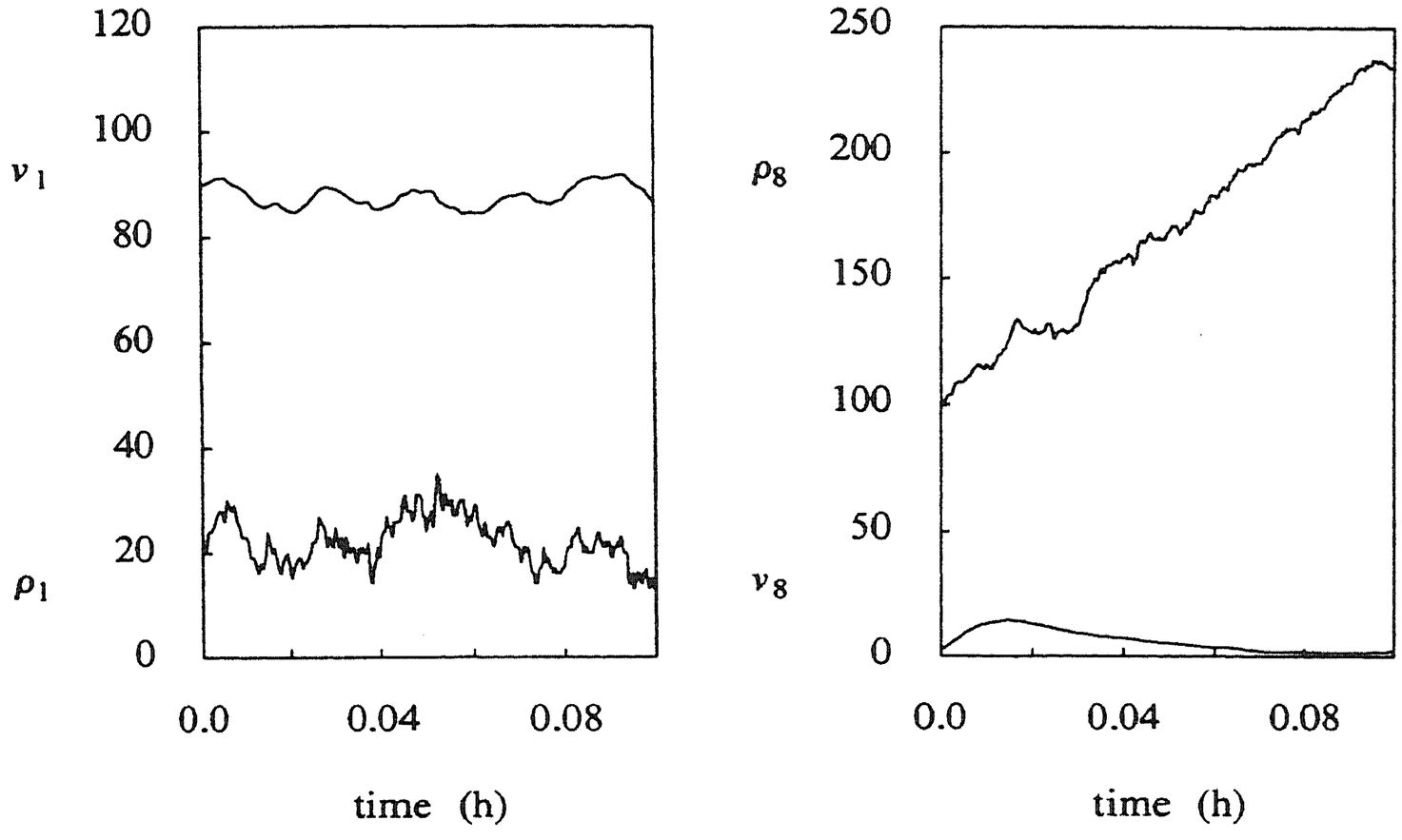

FIGURE 3.7. Simulation results with anticipation term (3.3).

	t = 0.0 h		t = 0.12 h			
			(3.3)		(3.4)	
section	ρ	v	ρ	v	ρ	v
1	20.0	90.0	22.0	89.2	95.0	0.0
2	20.0	90.0	19.0	90.6	100.0	29.8
3	20.0	90.0	14.0	90.5	113.0	25.9
4	20.0	90.0	15.0	88.6	94.0	9.1
5	20.0	90.0	25.0	79.2	97.0	11.9
6	30.0	72.0	41.0	14.1	97.0	4.1
7	60.0	22.5	216.0	0.0	103.0	9.9
8	100.0	2.7	208.0	2.3	100.0	17.5
9	100.0	2.7	132.0	5.4	95.0	0.0
10	100.0	2.7	94.0	8.9	97.0	6.3
11	90.0	6.0	82.0	7.0	98.0	2.6
12	90.0	6.0	88.0	7.0	102.0	4.0

TABLE 3.2. Simulation results: congested traffic.

of the upstream section and there would be no anticipation if ρ_i were 0, even if ρ_{i+1} would be as large as 100. On the other hand there would be a very strong effect if $\rho_i = 100.0$ and $\rho_{i+1} = 0$. If $\beta = 0$ the emphasis is on the downstream section and the conclusions go the other way around. It is clear that it is preferable to choose some intermediate value for β.

Investigations with respect to the so called *hysteresis phenomenon* in traffic flow [95] suggest an asymmetric driver behaviour. This means that a driver anticipating an *increase* of density is likely to react more strongly than when anticipating a *decrease* of density. We may model this effect by taking a value of β smaller than 0.5. For the time being we just choose $\beta = 0.5$. The parameter γ is chosen such that at a density of approximately 20 veh/km/lane the anticipation effect is about as strong as it was in the previous model. This means that the new model shows approximately the same behaviour as the old model for low and moderate densities.

We carried out the congestion simulation with the modified model. See Table 3.2 (The column indicated by (3.4)) and Figures 3.8, 3.9 and 3.10 for the results.

We now see the desired behaviour: the congested region extends itself in the upstream direction and density hardly exceeds the jam value. Apart from this we now see a fascinating behaviour in the congested region: *stop-start traffic*. This phenomenon is often observed in practice, especially in tunnel flow [28] or upstream of bottlenecks [46]. In how far the amplitude and period of the stop-start waves are in accordance with reality is not yet clear. For the moment we restrict attention to noting the ability of the new model to describe this type of behaviour at high densities.

Stability analysis confirms the simulation results. In Table 3.3 the eigenvalues of a three section model linearized around $\rho_i = 100$ veh/km/lane and $v_i = v^e(100) = 2.7$ km/h for $i = 1,2,3$ are presented. A stationarity boundary condition is taken at both exit and entrance of the freeway stretch. We observe four eigenvalues with large imaginary parts. Two of these correspond to unstable oscillations with a period of about two minutes. The other two are damped oscillations. The real eigenvalue value -100 corresponds to the relaxation time T of 0.01 h. The zero eigenvalue illustrates the fact that with stationarity at both exit and entry boundary all points $(\rho_1, \rho_2, \rho_3, v_1, v_2, v_3)$ with $\rho_1 = \rho_2 = \rho_3$ and $v_i = v^e(\rho_i)$ are equilibrium points.

A further refinement of the model might consist of replacing the linear anticipation factor $[\beta\rho_i + (1 - \beta\rho_{i+1}]$ by a quadratic one. This would result in smaller anticipative effects when congestion has set in and might lead to stable oscillatory behaviour. This will not be further investigated here.

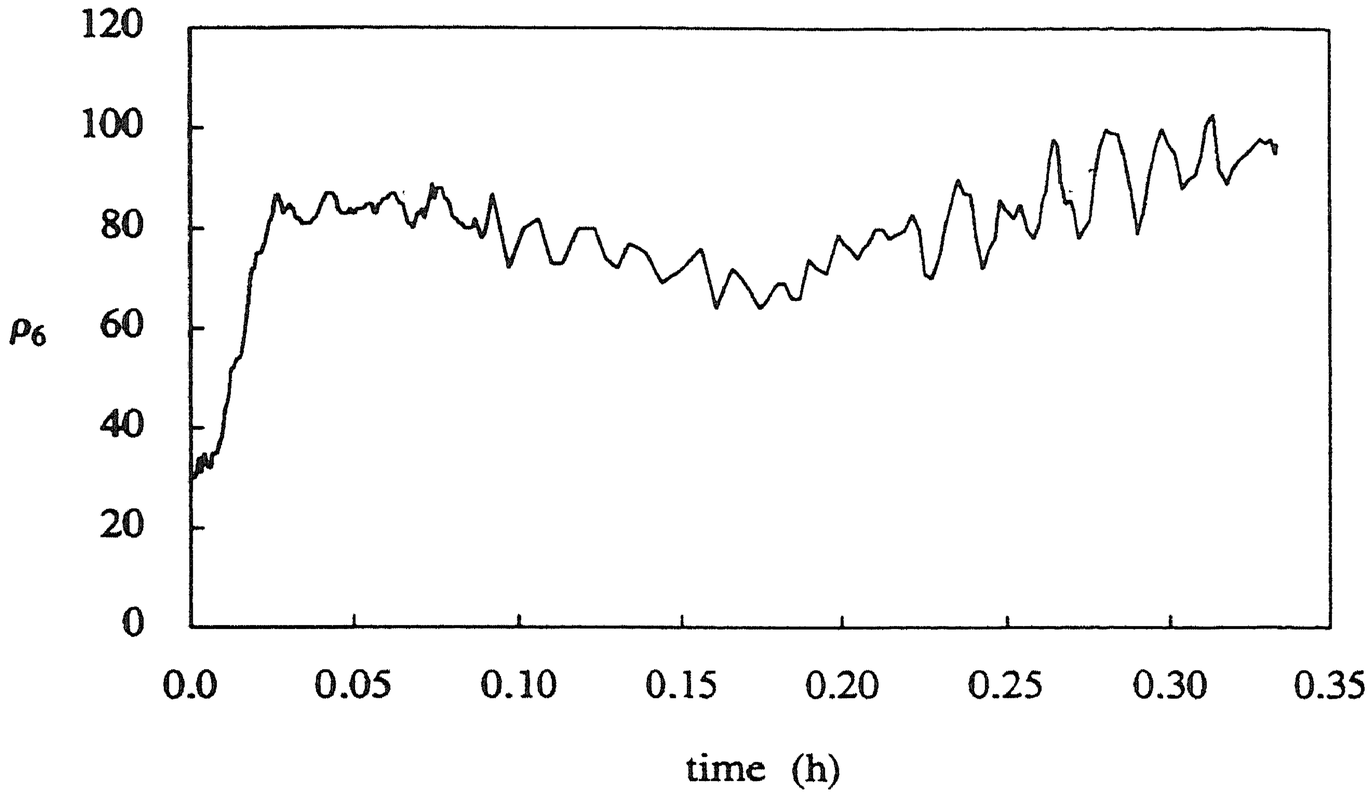

FIGURE 3.8. Extension of congestion upstream: traffic density.

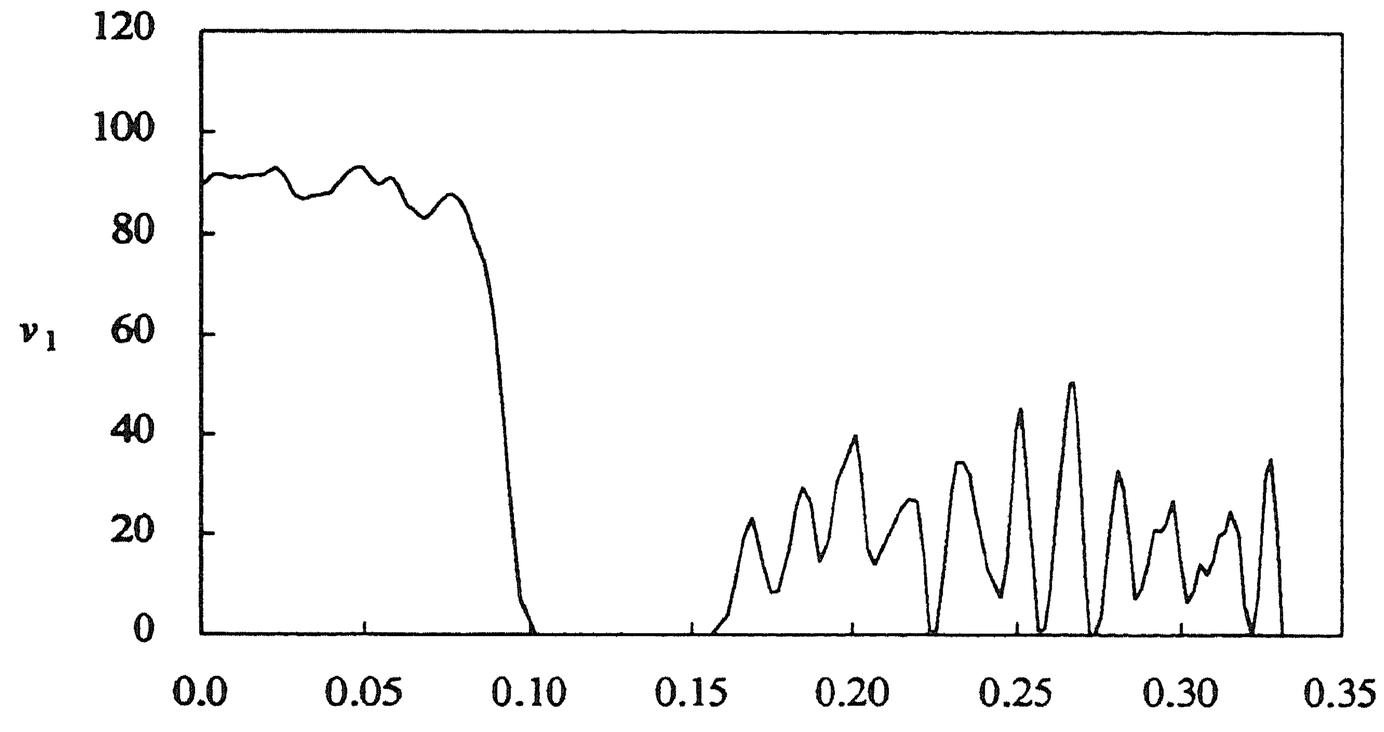

FIGURE 3.9. Extension of congestion upstream: mean speed.

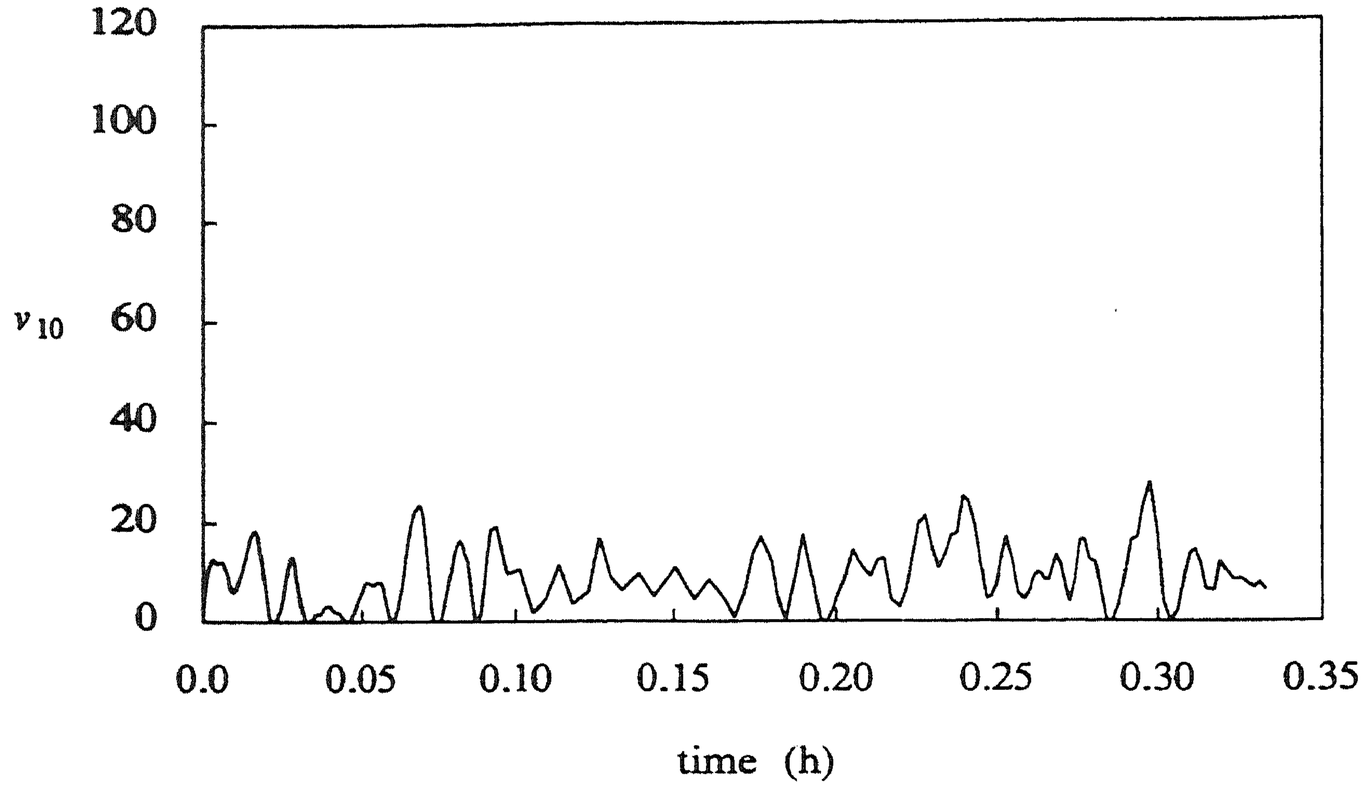

FIGURE 3.10. Stop-start waves in congested traffic.

n=3		
-169.34	±	216.75i
-100.00		
0.00		
62.05	±	212.20i

TABLE 3.3. Eigenvalues of a three-section model.

3.5. SUMMARY

The simulation results of the previous sections show that the original model does not display realistic behaviour in some basic traffic situations, but that modifications do lead to realistic behaviour. The modifications consist of adjustments of the parameter α, the equilibrium speed-density relationship and the anticipation term.

Apparently the extensive identification procedure applied by Van Maarseveen [62] failed to produce satisfactory parameter values. There may be several reasons for this. One reason was mentioned by Van Maarseveen himself: the data set used in the identification process hardly contained high density traffic and the model is therefore not guaranteed to display reasonable behaviour there. The model was unrealistic at low density values as well, however, so this does not provide a complete explanation. Another reason may be that only vehicle count

information was used, the speed measurements were discarded. This leaves a certain amount of freedom to the identification procedure to choose the parameter values in the speed equation and may be interpreted as a lack of identifiability.

The procedure applied by Van Maarseveen was already very involved. Developing an identification procedure that incorporates speed information will be more complicated still. We therefore content ourselves with the tentative values obtained in this chapter. In Chapter 5 it is concluded that the filter based on the traffic model of this chapter is surprisingly insensitive to most model parameters, thereby reducing the necessity for exact values, at least for estimation purposes. The parameters of the equilibrium relation $v^e(\rho)$ may be estimated from the freeway data directly, however. This is investigated in the next chapter.

The modified model consists of the following equations:

$$d\rho_i(t) = \frac{1}{l_i L_i}\Big\{ l_{i-1}[\alpha\rho_{i-1}(t)+(1-\alpha)\rho_i(t)][\alpha v_{i-1}(t)+(1-\alpha)v_i(t)] - l_i[\alpha\rho_i(t)+(1-\alpha)\rho_{i+1}(t)][\alpha v_i(t)+(1-\alpha)v_{i+1}(t)]\Big\}dt + \frac{1}{l_i L_i}[dr_i^{on}(t)-dr_i^{off}(t)] + \frac{1}{l_i L_i}[dm_{i-1}(t)-dm_i(t)] \tag{3.5}$$

$$dv_i(t) = -\frac{1}{T}[v_i(t)-v^e(\rho_i(t))]dt - \gamma(L_i l_i)^2[\beta\rho_i(t)+(1-\beta)\rho_{i+1}(t)][\rho_{i+1}(t)-\rho_i(t)]dt + \frac{l_{i-1}}{l_i L_i}v_{i-1}(t)[v_{i-1}(t)-v_i(t)]dt + dw_i(t) \tag{3.6}$$

where

$$v^e(\rho) = \begin{cases} v_{free}\,[1 - \dfrac{\rho}{\rho_{jam}}]\,, & 0 \leqslant \rho \leqslant \rho_{crit} \\ d\,[\dfrac{1}{\rho} - \dfrac{1}{\rho_{jam}}]\,, & \rho_{crit} \leqslant \rho \leqslant \rho_{jam} \end{cases} \tag{3.7}$$

for $i = 1, \ldots, n$. The parameters are given in Table 3.4.

parameter	value	unit
α	0.85	
T	0.01	h
ρ_{jam}	110.0	veh/km/lane
v_{free}	110.0	km/h
ρ_{crit}	27.0	veh/km/lane
d	2970.0	1/h
γ	6.5	km/h^2
β	0.5	

TABLE 3.4. The model parameters.

Chapter 4

Analysis of Freeway Data

In Chapter 2 an equilibrium relation between mean speed and traffic density was defined, which in Chapter 3 was shown to be of importance for the stability of the freeway traffic model. A functional form for this relation was proposed in the latter chapter, the parameters of which were based upon an identification procedure and simulation results. The vehicle count and speed measurement data available through the Dutch Motorway Control and Signalling System allow us to estimate some of these parameters directly. This will be carried out in Section 3 of this chapter.

The same data may also be used to investigate the dependence of the passing speed distribution defined in Section 2.3 on the traffic variables density and mean speed. Some insight in this dependence is necessary to allow the filter that is to be developed in the next chapter to process the passing speed measurements in a correct manner. The investigation of this dependence will also be pursued in Section 3 of this chapter. Before doing so the subject of probability density estimation will be discussed in Section 2.

Both estimation procedures require traffic data to be available of periods of stationary traffic. The selection of such periods will be discussed in the first section.

4.1. Selection of periods of stationary traffic

For the estimation procedures of the next two sections it is required that data is available of periods of traffic flow during which traffic density and mean speed are stationary processes. Unfortunately these traffic variables are not directly available, as the Signalling System only provides counts and passing speed values at measurement locations. Stationarity of mean speed and traffic density implies stationarity of the increments of the counting process $\{N_t,\ t \geqslant 0\}$, however:

$$dN_t = l\rho_t v_t dt + dm_t$$

at measurement locations. The vehicle counts or time headways at the measurement locations may therefore be used to test for stationarity. The extensive paper of Breiman and Lawrence [9] is devoted to this subject.

In [9] the properties of traffic flow, vehicle counts and time headways are investigated and two procedures for isolation of periods of 'constant' flow from a batch of data are proposed. The main difficulties in using data for statistical tests are concerned with dependence of the observations and with estimation of the moments. The authors show vehicle counts over periods of 18 to 36 seconds to be uncorrelated at all intensity levels. This holds for counts per lane as well as for the total over all lanes. Time headways appear only to be slightly correlated at heavier flows. The conclusion is that it is justified to use either counts or headways in statistical tests. One of the tests developed by Breiman and Lawrence is based upon counts and another is based upon headways [9]. Both are sequential, which poses a problem in that the test statistic gathers a lot of momentum as time proceeds and may react too slowly to an abrupt change in flow. An auxiliary estimation procedure is proposed to overcome this.

It is our goal to select periods during which traffic flow over both lanes is stationary, because the equilibrium relation $v^e(\rho)$ and the passing speed distribution of Chapter 2 are defined for aggregated lanes. This means that time headways are less appropriate for a test procedure. We therefore only use the count data from here on. A problem with the count-based test of Breiman and Lawrence is that an estimate of the variance of the counts is needed: the test is parametric. The authors themselves explain the difficulties in estimating this variance accurately. They suggest a nonparametric test based upon Kendall's tau-test for detection of a trend as an alternative.

We follow the suggestion of Breiman and Lawrence [9] and propose a nonsequential nonparametric test on traffic counts over all lanes over short time intervals, based on Kendall's tau-test. A description of the latter test is given by Gibbons [38]. The tau-statistic measures the *association* between two samples of measurements and may be used to detect a *trend* in one sample by measuring the association with an ordered sequence (1,2, ... for example). For j=0,1, ... define

$$\Delta N^j = N_{(j+1)\Delta t} - N_{j\Delta t},$$

the number of counts at the measurement location under consideration, during the time interval $[j\Delta t, (j+1)\Delta t]$. Now consider the sample

$$S = (\Delta N^0, \Delta N^1, \cdots, \Delta N^{m-1}).$$

The *tau statistic* τ is defined as

$$\tau = O / \binom{m}{2}, \tag{4.1}$$

where O is the number of pairs that occur in the sample S in natural order minus the number of pairs that occur in reverse natural order. Taking pairs has to be carried out systematically, each element being paired with elements to its right

only (otherwise O would always be zero). Note that $-1 \leqslant \tau \leqslant 1$ and that in case $\tau = 1$ the elements in S are in ascending order and if $\tau = -1$ they are in descending order. If $\tau = 0$ no trend exists. Ties, i.e. sample elements of equal rank, have to be taken care of separately.

The statistic τ may now be used in testing the hypothesis

$$H_0 : \text{no trend exists in } S$$

against the alternative

$$A : \text{a trend does exist in } S, \text{ either descending or ascending.}$$

Gibbons produces a table of the distribution of τ under H_0 for different values of m [38]. A critical value τ_{crit} has to be chosen such that the null hypothesis is erroneously rejected only five times out of hundred, or less. Rejection occurs iff $|\tau| \geqslant \tau_{crit}$. The tables given by Gibbons may be used to determine τ_{crit}. We will come back to this in the discussion about ties later on.

Now a nonsequential or moving horizon test is obtained by applying the presented hypothesis test to the samples $S^0, S^1, \cdots$ in succession, where

$$S^i = (\Delta N^i, \cdots, \Delta N^{m-1+i}).$$

That is, the test is applied to a series of m count increments, which moves forward through the data set with step size Δt. Applying the tau-test on a sample we may reject or select it, depending on the value of $|\tau|$. If $|\tau| > \tau_{crit}$ then the sample is rejected, if $|\tau| \leqslant \tau_{crit}$ the sample is selected. To each sample S^i a time interval $[i\Delta t, (i+m)\Delta t]$ corresponds. The procedure therefore ends up with a collection of selected time intervals. Adjacent intervals may be combined to obtain periods of stationary traffic.

ALGORITHM (MOVING HORIZON TEST)

1. Collect m count increments $\Delta N^0, \cdots, \Delta N^{m-1}$ into sample S^0 and set $i = 0$;
2. Compute the τ-statistic of sample S^i;
3. If $|\tau| \leqslant \tau_{crit}$ then the interval $[i\Delta t, (i+m)\Delta t]$ is selected. If $|\tau| > \tau_{crit}$ then the interval $[i\Delta t, (i+m)\Delta t]$ is rejected;
4. Go to Step 5 if the end of the data set is reached. Otherwise, set $i = i+1$, collect ΔN^{m-1+i} and go to Step 2;
5. Combine adjacent selected intervals to obtain periods of stationary traffic.

A disadvantage of the nonsequential test in comparison to a sequential one is that a large number of separate periods results. The data of different periods need to be combined to obtain accurate estimates of traffic variables, and this requires some administration. On the other hand, the nonsequential test in general removes periods with a trend whereas the sequential one distributes these periods over different periods of 'stationary' traffic [9].

The parameters of the test are m, Δt and τ_{crit}. We now discuss the choice of these parameters. As indicated by Breiman and Lawrence [9] Δt should not be taken too small as this introduces correlation. On the other hand, Δt should not be taken too large either as this adds momentum to the test and leads to a slower

detection of trend. The same holds for large m, but taking this parameter too small results in a low resolution of the test. In the extreme case that $m=2$ either $\tau=-1$ or $\tau=1$ and a large number of very short periods will be the result of the selection procedure. A lower bound for Δt appears to be about 15 seconds [9] and this is the value we use. To obtain a reasonable resolution we need m at least as large as 10. We take $m=12$, which means that a single sample S^i covers 3 minutes of data. A trend is now sure to be detected within 3 minutes. The precise detection speed depends on the value of τ_{crit}.

The critical value of τ should be chosen such that a reasonable probability of detection is assured without the introduction of a high false alarm rate. As mentioned before there is a problem in the occurrence of ties, however. For $\Delta t=15$ seconds ΔN_t approximately ranges from 0 to 30, which using $m=12$ gives a reasonable probability of a tie in a sample S^i. The mid-rank procedure of Gibbons is unsatisfactory in that a sample $(10,10,10,\cdots,10,11)$, for instance, leads to rejection of H_0: a trend exists. This is caused by the fact that all ties are essentially combined into one element, so that in the example (10,11) results. Now it is clear that there is an ascending trend! Furthermore, computation of the distribution of the corrected test statistic is complicated and depends not only on the number of ties, but also on their ranks. Gibbons proposes to use an approximation to the exact distribution.

We decided not to use Gibbons's correction procedure but just to compute τ according to (4.1) and use τ only as a measure of trend, without referring to its statistical properties. Some simulation experiments and tests on real data showed that it was necessary to allow a large false alarm probability in order to achieve a high probability of detection. That the false alarm probability is large is observed from the large number of separate periods of stationary traffic that the test ends up with. The value finally chosen for τ_{crit} is 0.25.

The moving horizon test described above was applied to three batches of two-lane freeway data of three hours each, including the morning rush hour. Two measurement locations were taken into consideration, about 4 km apart. The test led to a large number of time intervals of length varying from 15 seconds (Δt) to over 10 minutes. For each interval the intensity was estimated by dividing the number of counts by the length of the interval. This is justified by the stationarity of flow. Only intervals longer than two minutes were taken into account. This is to assure a reasonable accuracy of the estimated intensity. All the intervals of about the same intensity were combined. This led to 26 intensity classes, each of which consists of at least 150 and at most 1000 counts, obtained from 1 to 5 separate stationary periods. See Table 4.1 for the results. Care had to be taken to avoid including data of periods of congested traffic into data of periods of freely flowing traffic. Although these types of traffic have very different characteristics the intensity level may be the same.

λ (veh/h/lane)	number of data points	λ (veh/h/lane)	number of data points
750 - 900	409	1650 - 1700	558
900 - 1050	524	1700 - 1750	635
1050 - 1100	583	1750 - 1800	380
1100 - 1150	671	1800 - 1850	138
1150 - 1200	996	1850 - 1900	669
1200 - 1250	338	1900 - 1950	490
1250 - 1300	983	2000 - 2050	791
1300 - 1350	488	2050 - 2100	346
1350 - 1400	609	2100 - 2150	158
1450 - 1500	443	2150 - 2200	234
1500 - 1550	654	2200 - 2250	298
1550 - 1600	608	2300 - 2350	156
1600 - 1650	240	2450 - 2500	164

TABLE 4.1. Result of the data selection procedure.

4.2. DENSITY ESTIMATION

The subject of density estimation is a research topic on its own and is therefore discussed separately in this section. Only nonparametric estimators are considered as we know too little about the underlying density to allow for a parametrization. Another reason for considering only nonparametric estimators is that the analysis is for a large part exploratory. In general nonparametric estimators are more robust than parametric ones, at the cost of a lesser performance.

A large number of nonparametric density estimators have been proposed in the literature [25, 90]. The simplest procedure consists of defining a set of disjunct intervals, *bins*, covering the domain of the stochastic variable and computing the fraction of observations per bin. This is the well-known *histogram*. A surprisingly simple improvement of the histogram was found by Rosenblatt and consists of computing the fraction of observations in an interval of fixed length, *window*, which moves over the domain of the stochastic variable [83]. The estimate of the density at a certain point is the fraction of observations in the window centreed around this point. This modified histogram is sometimes called the *moving window* estimator. As shown by Tapia and Thompson [90], the error variance associated with this estimator is of order $n^{-\frac{4}{5}}$ which is a considerable improvement over the order $n^{-\frac{2}{3}}$ of the histogram. A common feature of both estimators is the discontinuity of the resulting estimate. A generalization of the moving window estimator, which allows continuous estimates, is the *kernel estimator*:

$$\hat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K(\frac{x-w_i}{h}) \tag{4.2}$$

where n is the number of data points, h the *window width* or *smoothing factor* and

the w_i are the data points. The *kernel* K is usually taken to be a density function itself as this implies that the estimate $\hat{f}_n$ is a density also (i.e., integrates to one and is nonnegative). The latter is not guaranteed for some other types of estimators and this may be an annoying feature. If the kernel is a continuous function $\hat{f}_n$ is continuous as well. The moving window estimator results from (4.2) by taking a uniform density for K:

$$K(x) = \mathrm{I}_{[-\frac{1}{2},\frac{1}{2}]}.$$

This estimator is therefore also called the *naive* kernel estimator. The idea behind the nonnaive estimator is that data points further away from the centre of the window contain less information about the density function in the centre point and therefore should be weighted less. A common choice for K, leading to a continuous estimate, is the standard normal density.

A further improvement of estimation performance might be expected if the window width h could be optimized. It is intuitively clear that taking h too large *biases* the estimate $\hat{f}_n$ to a function that is too smooth, whereas a small h leads to a large amount of *variation* (peaks). This is illustrated by the following decomposition [90]:

$$E[(\hat{f}_n - f)^2] = \mathrm{var}(\hat{f}_n) + \mathrm{bias}^2(\hat{f}_n)$$

with

$$\mathrm{var}(\hat{f}_n) \sim O(\frac{1}{nh}), \quad \mathrm{bias}^2(\hat{f}_n) \sim O(h^4).$$

Hence, if h is small compared to n then the variance is large, but if h is large then a large bias results. Optimal choices of h may be derived but always depend on the underlying density f itself. An illustration of this is given later on in the discussion of the properties of the kernel estimator. Methods to circumvent this dependence have been proposed and consist of using a guessed approximate form for f (normal for example), iterative techniques in which a parametrized approximating density is used, or data based methods like cross-validation and maximum likelihood estimation. It is recommended to experiment with several values of h always and this is what we will do.

A final refinement of the kernel estimator $\hat{f}_n$ consists of allowing the window with h to vary with x. This is expected to be profitable as for nonuniform densities the data is distributed over the domain in a nonuniform way. A smaller h at the peaks of the distribution would lead to less bias and a larger one at the tails would lead to less variation. Wegman [98] has shown that the estimator in which the window width varies inversely proportional to the density of the data points is the maximum likelihood estimator. An example of a kernel estimator with varying window width is the *nearest neighbour* estimator. Here the window width is chosen to be the distance between the centre of the window and its k-th nearest neighbour in the data points. Instead of h we now have k as a free parameter. The problem with this estimator is its inconsistency at the tails of the density. The variable kernel estimator of Breiman [11] overcomes this problem at the cost

of introducing new parameters that require optimization. The numerical tests of Bean and Tsokos [4] of several variable smoothing factor estimators indicate that these more refined estimators perform worse on small data sets than estimators with a fixed window width. A certain minimum amount of data is needed to pin down the extra degrees of freedom.

We now present some theoretical results concerning the kernel estimator, most of which are taken from Devroye [25].

THEOREM 4.1 (consistency). *Let $\hat{f}_n(x)$ be the kernel estimate (4.2) with $K(x)$ a density-kernel and let h depend on n only. If*

$$\lim_{n\to\infty} h_n = 0\,, \quad \lim_{n\to\infty} nh_n = \infty$$

then

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} |\hat{f}_n(x) - f(x)|\,dx = 0$$

almost surely, for any density f.

The sufficiency conditions on h and n of the theorem are also necessary and illustrate the previous discussion about bias and variation. Apparently h should go to zero to assure that there is no bias in the asymptotic estimate, but h should not go to zero too fast. There should be enough data in each window to prevent a large variation.

Next, one would like to obtain results about the accuracy of the estimate for finite n. A natural question seems to be whether

$$E[\hat{f}_n(x)] = f(x)$$

for all x. The error variance

$$E[(\hat{f}_n(x) - f(x))^2]$$

also is of interest. Tapia and Thompson [90] give expressions for these quantities and show that the kernel estimator is asymptotically unbiased. It is also $\mathfrak{L}_2$-consistent. As Devroye explains, however, pointwise criteria are less important than global ones. The estimated density is only used to compute the measure of specific sets (probabilistic events) and it would therefore be more of interest to investigate an error criterion like

$$E[\sup_A |\int_A \hat{f}_n(x)dx - \int_A f(x)dx\,|],$$

because this measures the maximal error in the estimation of the probability of an event. As it turns out the $\mathfrak{L}_1$-distance has the desired property:

$$E[\int_{-\infty}^{\infty} |\hat{f}_n(x) - f(x)|\,dx] = 2E[\sup_A |\int_A \hat{f}_n(x)dx - \int_A f(x)dx\,|].$$

The proposed $\mathfrak{L}_1$-distance depends on the specific f in general, which is the

unknown. A way out of this problem is to investigate an upper bound for the estimation error:

$$\inf_{f_n} \sup_{f} E\{\int |\hat{f}_n - f|\, dx\}. \tag{4.3}$$

A useful bound is only obtained if we restrict f to be an element of a subclass of the class of all densities. Devroye [25] proves kernel estimators for which the kernel has a certain degree of smoothness to be *minimax optimal* for the density subclasses considered. This means that the kernel estimator achieves a minimax bound up to a multiplicative constant. It is to be expected that for specific f the error will be much smaller than the minimax bound suggests. When we restrict the kernel to be nonnegative, however, (to obtain a estimate which is really a density), the minimax bound is close to the actual bound for a large class of densities f.

THEOREM 4.2 (Devroye). *Let f be a density with a continuously differentiable derivative and $\int x^2 f(x)dx < \infty$. If K is a nonnegative kernel with finite second moment and $\int (1+x^2)K^2(x)dx < \infty$ then*

$$\inf_{h>0} E[\int |\hat{f}_n - f|\, dx] \leqslant (1+o(1))(2^{\frac{1}{5}} + 2^{-\frac{9}{5}})\phi(K)\psi(f)n^{-\frac{2}{5}}$$

where $\hat{f}_n$ is the associated kernel estimator and

$$\phi(K) = \left[\int x^2 K dx (\int K^2 dx)^2\right]^{\frac{1}{5}}$$

$$\psi(f) = \left[\int |f^{(2)}|\, dx (\int \sqrt{f}\, dx)^4\right]^{\frac{1}{5}}.$$

The optimal window width is given by

$$h = n^{-\frac{1}{5}} \frac{\left[\int K^2 dx\right]^{\frac{1}{5}} \left[\int \sqrt{f}\, dx\right]^{\frac{2}{5}}}{\left[2\int |f^{(2)}|\, dx \int x^2 K dx\right]^{\frac{2}{5}}}.$$

It turns out that the kernels satisfying the properties mentioned achieve a fairly uniform estimation accuracy, nearly independent of the specific f. An improved behaviour may only be achieved by restricting the underlying class of densities (i.e., using stronger assumptions) and using tailor-designed or parametric estimators. This will go at the cost of robustness, as illustrated by Devroye [25].

Although the order of convergence is independent of the choice of the kernel (within the assumptions of the theorem) and the form of f, for fixed n optimal choices may be made. It may be shown that the optimal kernel is

$$K(x) = \max\{0, \tfrac{3}{4}(1-x^2)\} \qquad \text{(Bartlett's kernel)},$$

but the Gaussian kernel achieves a value of $\phi(K)$ which is nearly optimal.

The optimal form of f turns out to be an isosceles triangle. A normal density again has a ψ-value that is close to optimal. Bad densities are those with large tails (Cauchy) or discontinuities (exponential).

As for the optimal choice of the window width h we note that this depends on f and n, see Theorem 4.2:

$$h = \alpha(n)\beta(f).$$

Again optimization would require a data dependent estimation procedure. As mentioned at the beginning of our discussion, some of such procedures have been proposed in the literature. In general these require quite some computational effort, however. We restrict attention to experimenting with several values of h.

The theory so far assumed the observations to be independent samples of the underlying distribution. In moderate to heavy flow, passing speeds at a fixed location along the freeway are significantly dependent, however. Investigations in the literature suggest to use the same estimators for correlated data as for independent data, as long as correlation dies out with increasing lag. The only drawback occurs in a slower convergence speed [23]. In our case it may be assumed that [10]

$$\sigma_i = \alpha\beta^{i-1} \quad , i \geqslant 1$$

for some $\alpha > 0$ and $0 < \beta < 1$. Here σ_i is the autocorrelation of lag i. This means that correlation dies out as the lag grows and that using the estimators discussed in this section is justified.

The conclusion of this section is that the kernel estimator is the one most suitable for our purpose. We choose to use a Gaussian kernel to obtain smooth estimates and a good performance. The window width is partly based upon an approximation of the optimal width and partly on trial-and-error.

4.3. Estimation of speed probability density and equilibrium speed

Section 4.1 has made a batch of data of stationary traffic available, corresponding to a series of intensity levels. These may now be used to estimate the probability density of the passage speed of individual vehicles, introduced in the observation equation (2.20) of the traffic model presented in Section 2. Some parameters of the equilibrium speed – density relationship $v^e(\rho)$ may be estimated also.

Recall that the speed domain was discretised in a finite number of speed classes and that we assumed the probability density to depend on mean speed and traffic density of the sections neighbouring the measurement location:

$$\int_{v^{j-1}}^{v^j} f_i(v)dv = \gamma_i^j(\rho_i, \rho_{i+1}, v_i, v_{i+1})$$

where

f_i : the probability density at location i;

v^{j-1}, v^j : the boundaries of speed class V^j;

γ_i^j : the fraction of vehicles that pass location i with speed in class V^j;

$\rho_i, \rho_{i+1}, v_i, v_{i+1}$: the densities and mean speeds of neighbouring sections.

One of the main questions to be asked is whether the probability density f depends on traffic density at all, other than indirectly, via the mean speed. Another question is concerned with the number of modes of the density function: the presence of slow vehicles (trucks) and the fact that there is a large speed difference between different lanes may express itself in a multimodal function.

We will now discuss some manipulations needed before we come to the actual estimation procedure. It is clear that estimating γ_i^j for all quadruples $(\rho_i, \rho_{i+1}, v_i, v_{i+1})$ is infeasible and also unnecessary. We will assume that only the density and mean speed in the neighbourhood of the location are of interest and approximate these by

$$\overline{\rho}_i = [\alpha\rho_i + (1-\alpha)\rho_{i+1}]$$

$$\overline{v}_i = [\alpha v_i + (1-\alpha)v_{i+1}]$$

with $\alpha \in [0, 1]$, so

$$\gamma_i^j = \gamma_i^j(\overline{\rho}_i, \overline{v}_i).$$

The quantities $\overline{\rho}_i$ and $\overline{v}_i$ closely follow an equilibrium curve, a model of which was presented in Chapter 2 and modified in Chapter 3. This means that $\overline{v}_i$ will heavily depend on $\overline{\rho}_i$ and the probability density will only have to be estimated for $\overline{v}_i$ close to $v^e(\overline{\rho}_i)$. Using the approximate relationship

$$\lambda_i \approx l\overline{\rho}_i\overline{v}_i$$

on our stationary traffic data and a direct estimate of $\overline{v}_i$

$$\hat{\overline{v}}_i = \frac{N}{\sum_k \frac{1}{w^k}} \tag{4.4}$$

where w^k is the k-th passage speed and N the number of observations, we obtain the following estimate of the traffic density in the neighbourhood of the measuring location i:

$$\hat{\overline{\rho}}_i \approx \lambda_i / l\,\hat{\overline{v}}_i .$$

The batches of data obtained in the previous section may now be classified according to their density $\overline{\rho}_i$. As it turns out the density varies from about 8 to 27 veh/km/lane: for larger values apparently stationary traffic does not occur. Recall that according to the model of Chapter 2 no stationary traffic occurs for densities above ρ_{crit}. The estimation results therefore lead to the conclusion that

$$\rho_{crit} \approx 27 \text{ veh/km/lane.}$$

Although $\overline{v}_i \approx v^e(\overline{\rho}_i)$ we do not restrict attention to the relation

$$\gamma_i^j = \gamma_i^j(\overline{\rho}_i, v^e(\overline{\rho}_i))$$

but determine a density such that

- the location of the probability density is determined by $\overline{v}_i$

  only: $\int_0^\infty v f_i(v) dv = \overline{v}_i$;

- the other moments are determined by $\overline{\rho}_i$.

Now the density f_i and the fractions γ_i^j are defined for all pairs $(\overline{\rho}_i, \overline{v}_i)$ with $\overline{\rho}_i \in [8, 27]$. The modelling for values of the traffic density above 27 is discussed later. The model for $\overline{\rho}_i < 8$ is a linear extrapolation of the estimated relation.

The Gaussian kernel density estimator introduced in the previous section was applied to the batches of data of Table 4.1. The window width was chosen to be 3 to 5 km/h, as suggested by approximate methods and by experiments. In Table 4.2 the statistical results for each selected class are presented and in Figure 4.1 some of the estimated densities are plotted. In Table 4.2 μ stands for the mean, σ for the standard deviation. MED stands for the median and MAD for Median Absolute Deviation:

$$\mathrm{MAD}(\{x_i\}_1^n) = \mathrm{MED}(\{|x_i - \mathrm{MED}|\}_1^n).$$

It is known that median and MAD are more robust to outliers than mean and standard deviation estimates.

The table displays a clear decrease of the mean speed with increasing density and also a decrease of the standard deviation. This is confirmed by the following relations, obtained by linear regression:

$$\mu = 107 - 0.64\rho, \quad \mathrm{MED} = 109 - 0.72\rho,$$

$$\sigma = 16 - 0.28\rho, \quad \mathrm{MAD} = 12 - 0.24\rho.$$

For the rest the picture is not very clear. A simple unimodal probability density hardly ever occurs, usually several peaks show up, varying in strength and location. Apparently, the speed distribution is not very stable and depends on other factors besides traffic density and mean speed. We do not investigate this further but restrict attention to using the information obtained about the first two central moments.

To obtain a speed probability density parametrized by mean and variance alone, we apply the following procedure. All the data of a selected period corresponding to a specific traffic density value is scaled to achieve a median of zero and a MAD of 1.0. The data of all 26 selected periods is then combined to estimate one single density by means of the Gaussian kernel estimator with window width 0.3 km/h. This involves over 10.000 observations and leads to a lower bound of the $\mathfrak{L}_1$-error of approximately 0.02. Using expressions derived by Tapia and Thompson [90] a lower bound for the bias and the error variance (pointwise) may be computed. For this it was assumed that f is normal and h is

ρ (veh/km/lane)	μ (km/h)	σ (km/h)	MED (km/h)	MAD (km/h)
8.0	102.3	13.2	103.4	9.6
9.8	100.2	13.2	100.0	9.8
10.8	100.2	12.6	100	9.1
11.5	98.2	12.6	97.8	9.3
11.8	99.8	12.7	100.0	9.1
12.5	98.3	12.1	98.4	9.3
13.0	97.1	12.5	97.8	9.3
13.1	100.4	13.2	101.1	10.0
13.8	99.4	13.0	100.0	8.4
14.9	99.7	11.7	98.9	8.2
15.8	97.5	11.5	96.8	7.8
16.0	98.5	11.7	98.4	7.5
16.8	96.0	10.8	96.8	7.9
17.4	96.2	11.7	96.3	8.5
17.8	101.1	11.2	102.3	8.5
18.0	95.9	11.3	95.7	8.3
18.5	96.3	10.7	95.7	7.5
19.6	95.6	11.1	95.7	8.3
21.0	92.2	9.6	91.8	7.1
21.6	93.9	11.1	93.8	8.1
22.2	96.2	11.3	95.7	7.7
23.5	87.9	8.5	88.2	6.5
24.2	90.5	12.5	90.0	8.9
25.0	90.1	8.5	89.1	5.6
25.3	92.0	8.9	91.8	5.0
27.1	91.1	7.1	91.4	5.4

TABLE 4.2. Estimated statistics of the passing speed probability distribution.

chosen optimally, varying with v. As neither of these is the case and the observations are dependent, the results only present lower bounds. An approximate 95% confidence interval around the estimate may be drawn, based on assumed normality of the estimation error. The result is plotted in Figure 4.2. The density function is close to being unimodular. There is a small bump on the left side, however. The density is not symmetric. Nevertheless, a normal density with the same mean and variance approximates the estimated density very well. This is in accordance with results of Breiman [10] where a goodness-of-fit test was developed and applied to speed probability density per lane. Figure 4.3 shows the normal density compared with the real estimate for two different ρ-values.

Note that the normality of the resulting density may only partly be explained by the central limit theorem. Most of the disturbances which lead to a nonnormal density for a finite number of observations indeed average out as the amount of data increases. This may explain the excellent match at the right slope of the

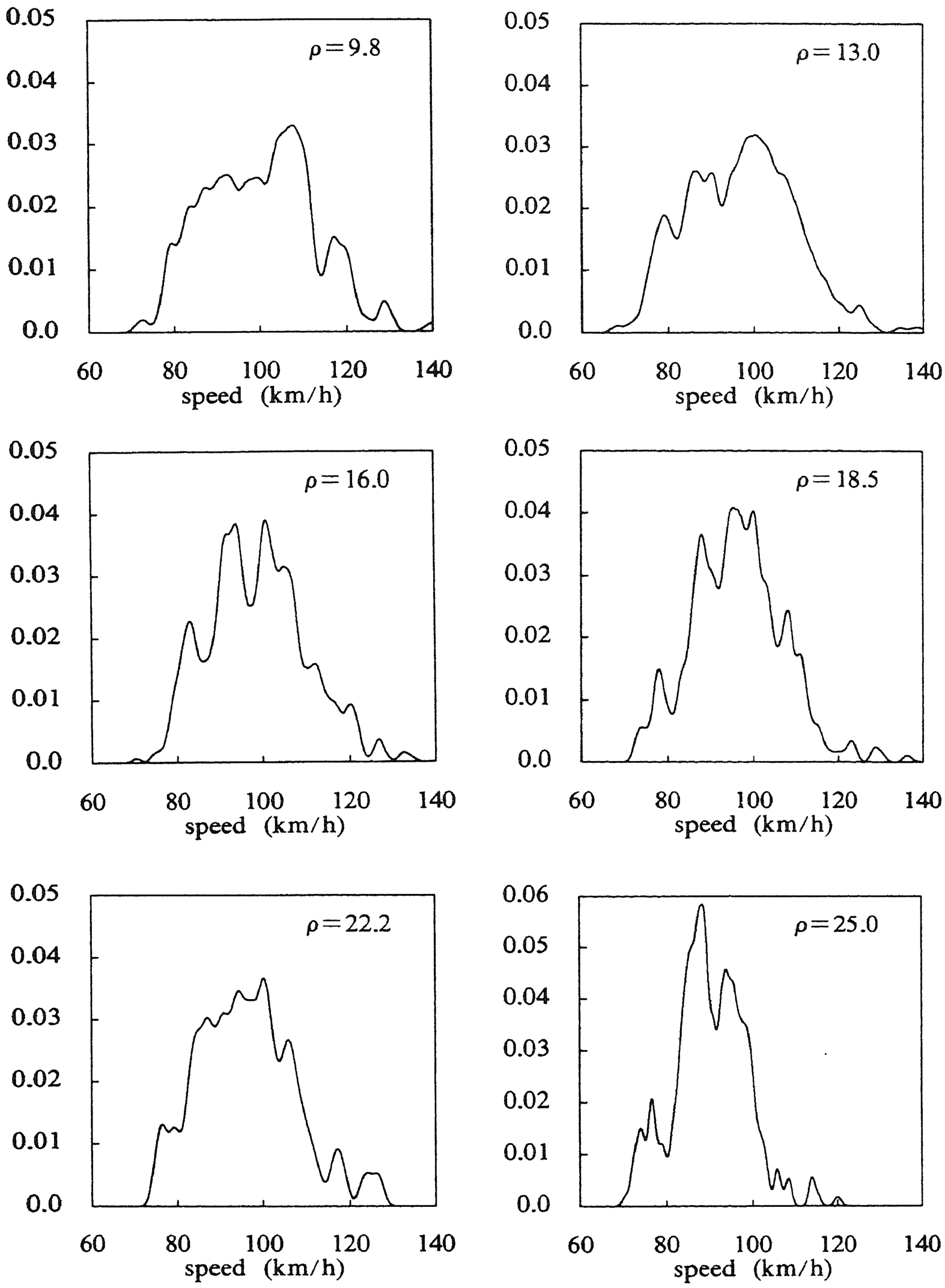

FIGURE 4.1. Estimated passing speed probability densities.

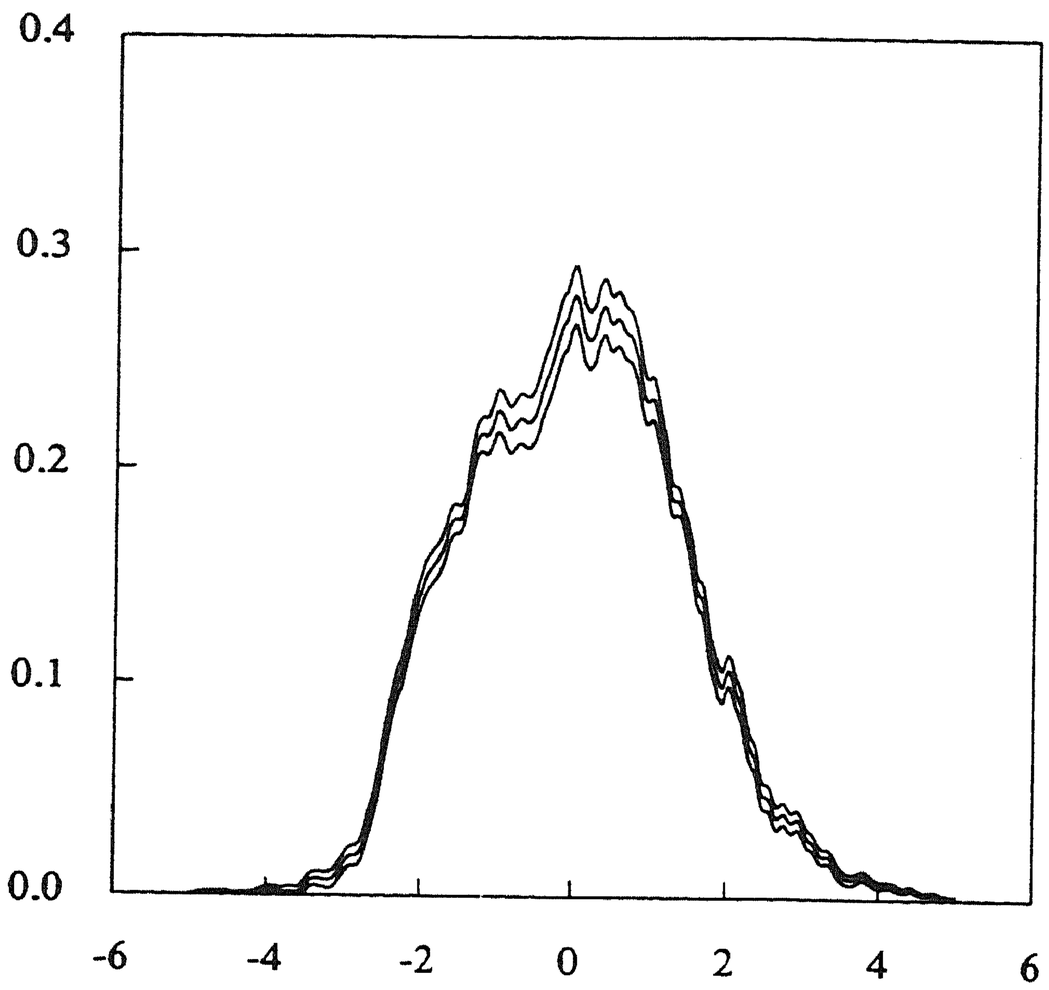

FIGURE 4.2. The scaled probability density and approximate confidence bounds.

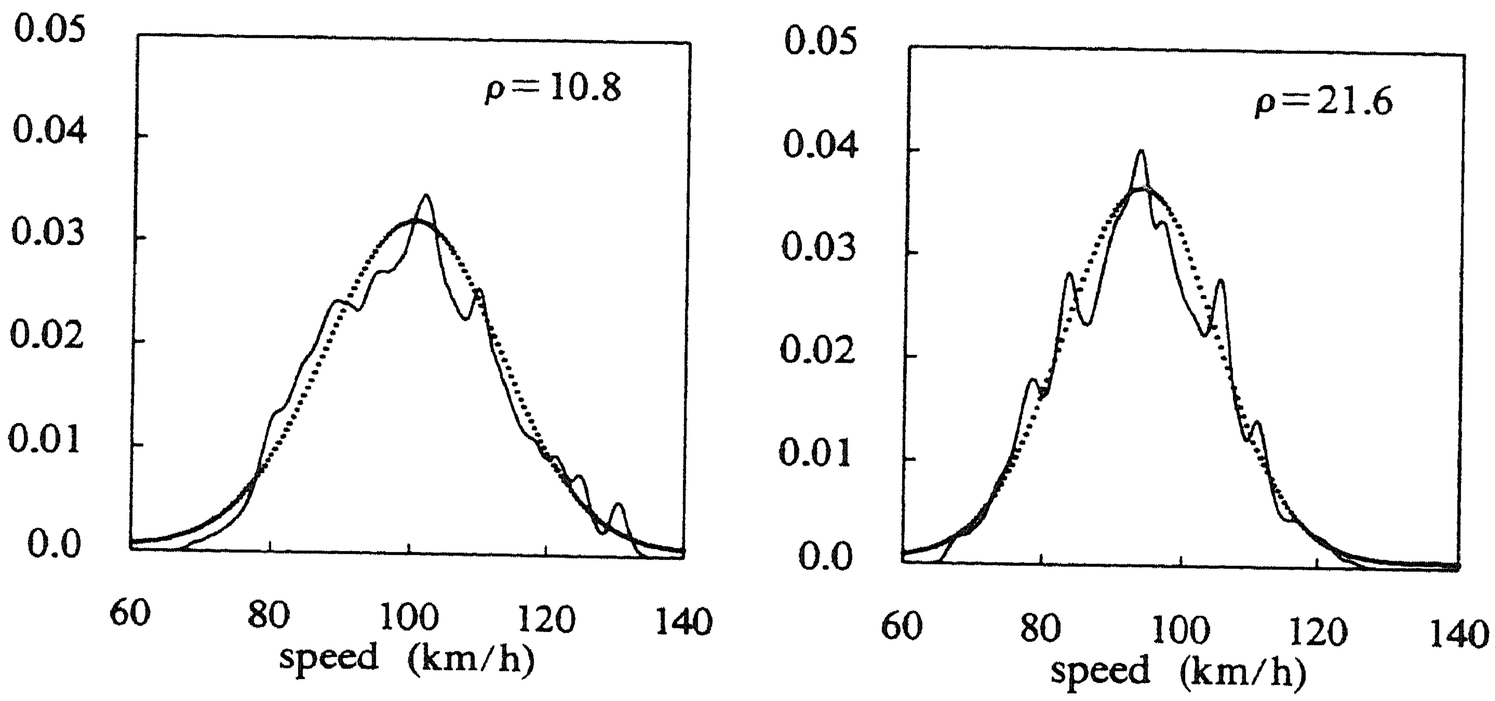

FIGURE 4.3. Comparison of estimated and normal densities.

density. The central limit theorem does not necessarily imply that the resulting density is normal, however. This is immediately clear if we assume the density to be a convex combination of two densities, each for one lane of the freeway. If these are both normal but have different mean the sum cannot be normal, and the estimate does not approximate a normal density, no matter how large the number of observations.

To conclude, the passing speed probability density function at a location i is modelled as a normal density with mean and standard deviation as follows:

$$\mu = \alpha v_i + (1-\alpha) v_{i+1}$$

$$\sigma = 16 - 0.28[\alpha \rho_i + (1-\alpha)\rho_{i+1}]$$

where ρ_i, ρ_{i+1} are the traffic densities of the neighbouring sections and v_i, v_{i+1} the mean speeds.

The relation above only holds for stationary traffic, when $\rho_{crit} \leqslant 27$ veh/km/lane. Investigation of a very limited set of data during which density was high led to the hypothesis that σ does not become smaller than 6 km/h. According to the estimated relation above this value is reached at $\rho \approx 35$ veh/km/lane. We therefore now assume:

$$\sigma = \begin{cases} 16.0 - 0.28\rho\,, & \rho \leqslant 35 \\ 6.0\,, & \rho > 35. \end{cases} \tag{4.5}$$

As the modelled speed probability density only is used in the filtering of measurements in the next chapter and we are mainly interested in traffic up to the critical value, we may assume this approximation not to be too restrictive.

The estimation of the parameters of the equilibrium relation $v^e(\rho)$ is relatively straightforward. We have already noted that $\rho_{crit} = 27$ veh/km/lane and therefore our model becomes:

$$v^e(\rho) = v_{free} - a\rho\,, \quad \rho \leqslant 27. \tag{4.6}$$

The exact values of ρ are not available. We use the estimated $\bar{\rho}_i$ – values of Table 4.1. The corresponding speeds are estimated according to (4.4). Linear regression then leads to

$$v_{free} = 105 \text{ km/h}$$

$$a = 0.58 \text{ km}^2 \cdot \text{lane/h}. \tag{4.7}$$

The other parameters of $v^e(\rho)$ are not estimated from the data. We assume $\rho_{jam} = 110$ veh/km/lane, comparable to the value obtained by the identification procedure of Van Maarseveen [62]. Furthermore, we take

$$d = 3197 \text{ 1/h}$$

to obtain continuity of $v^e(\rho)$ at $\rho = \rho_{crit}$. Whether this is a reasonable assumption

is subject to some doubt. Edie [27] was the first to propose a discontinuity in the equilibrium relation, which was also mentioned by [46]. Recently, Hall [42] and others have argued that the discontinuity that appears in the observed data is in fact caused by the measurement procedure and is not characteristic of traffic flow. Results of the next chapter with the filter also point in this direction.

Chapter 5

Filtering of Freeway Traffic Flow

The results of the three previous chapters now allow us to take the next step in solving the traffic control problem which is the subject of this thesis.

Based upon the freeway traffic model developed until now, and using the measurements available through the Dutch Motorway Control and Signalling System, an estimation algorithm will be derived. This *filter* processes the measurement data and produces an estimate of the state of traffic which is optimal in a minimum variance sense. The filter that results from this chapter is a finite dimensional approximation to the theoretically optimal filter, and is recursive. This means that a limited computational effort and memory is needed at every time step to process the new measurements and to produce the updated estimates.

In the first section the approximate filter is derived. In Section 2 important properties of the filter are investigated and the implications for the asymptotic error variance are discussed. The effect of speed information on the accuracy of the estimates is investigated also. In the third section the filter is applied to simulated data to analyse bias, error variance and robustness. Next the filter is applied to real data to assess the performance in practice. Finally, in the last section the results are summarized and discussed.

5.1. Derivation of an approximate filter

In this section an algorithm is developed for the estimation of the section densities and mean speeds from the measured passing times and vehicle speeds.

To simplify notation introduce the *state vector*

$$X_t = [\rho_1(t), \rho_2(t), \ldots, v_{n-1}(t), v_n(t)]^T,$$

the *measurement vector*

$$N_t = [n_0^1(t), n_0^2(t), \ldots, n_n^m(t)]^T$$

and the vector of intensities:

$$\Lambda(X_t) = [\lambda_0^1(t), \lambda_0^2(t), \ldots, \lambda_n^m(t)]^T.$$

The model equations (3.5), (3.6) may then be summarized as

$$dX_t = F(X_t)dt + dZ_t \tag{5.1}$$

and the measurement equations (2.20) as

$$dN_t = H(X_t)dt + dM_t \tag{5.2}$$

where $F(.)$ and $H(.)$ and Z and M follow from the equations referred to. For later use we write

$$F(X_t) = \begin{bmatrix} F_1(X_t) \\ F_2(X_t) \end{bmatrix} = \begin{bmatrix} A\,\Lambda(X_t) \\ F_2(X_t) \end{bmatrix},$$

where the constant matrix A follows from equation (3.5) and only contains zeroes and entries of the type $\pm 1/L_i l_i$.

The model parameters are summarized in Table 5.1.

parameter	value	unit
α	0.85	
T	0.01	h
ρ_{jam}	110.0	veh/km/lane
v_{free}	105.0	km/h
ρ_{crit}	27.0	veh/km/lane
a	0.58	km^2/h
b	3197.0	1/h
γ	6.5	km/h^2
β	0.5	

TABLE 5.1. The parameters of the freeway traffic model.

For the estimation of the state X_t from the measurements $\{N_s, s \leqslant t\}$ techniques have been developed in the area of system and control theory. Well-known is the Kalman filter for the case where F and H are linear and Z and M are Brownian motion processes. For the case of counting type measurements theory may be found in [13, 62]. We only give a brief account here.

As can be easily shown, the estimator of X_t that minimizes the estimation error variance is given by

$$\hat{X}_t = E[X_t \mid \mathcal{F}_t^N].$$

This is the mean of X_t conditioned upon the measurements. $\mathcal{F}_t^N$ is the σ-algebra generated by $\{N_s, s \leqslant t\}$ and represents the information contained in the measurements up to and including time t. For $\hat{X}_t$ the following differential

representation may be derived:

$$d\hat{X}_t = E[F(X_t) \mid \mathcal{F}_t^N]dt + \Phi_t(dN_t - E[H(X_t) \mid \mathcal{F}_t^N]dt) \tag{5.3}$$

where Φ_t is the *gain*, defined by

$$\Phi_t = \left\{ (E[\tilde{X}_t \tilde{H}(X_t)^T \mid \mathcal{F}_t^N] + E[\frac{d}{dt} <Z,M>_t \mid \mathcal{F}_t^N]) E[\frac{d}{dt} <M,M>_t \mid \mathcal{F}_t^N]^{-1} \right\}_t \tag{5.4}$$

Furthermore,

$$\tilde{X}_t = X_t - \hat{X}_t,$$
$$\tilde{H}(X_t) = H(X_t) - E[H(X_t) \mid \mathcal{F}_t^N]$$

and $<Z,M>_t$, $<M,M>_t$ are predictable (co)variation processes. For a precise definition see the Appendix or [24, Part II, Chapter 7, Theorem 37 and 39]. If we assume that the counting processes in N_t do not have common jumps it may be shown that

$$\frac{d}{dt} <M,M>_t = \text{diag}[H(X_t)] .$$

Note that a diagonal matrix which has the components of a vector X as its diagonal elements is denoted by diag[X]. It may furthermore be shown that in our case

$$\frac{d}{dt} <Z,M>_t = \begin{bmatrix} AM_t \\ O \end{bmatrix}.$$

For a derivation of these results, see [62], Chapter 4, Section 3.

In (5.3) the first term on the right-hand side shows that the filter estimates are based upon a prediction from the traffic model. The second term is a correction term. If the estimated state is in conflict with the observations, this is partly attributed to an estimation error and partly to the measurement noise.

Equation (5.3) gives the exact evolution of $\hat{X}_t$ in time. Unfortunately, $\hat{X}_t$ cannot be computed from (5.3) in general. For the evaluation of $E[F(X_t) \mid \mathcal{F}_t^N]$ the computation of the entire conditional probability distribution of X_t is needed, which is infeasible. The same holds for other terms in the representations for $\hat{X}_t$ and Φ_t. We therefore have to resort to approximations.

The usual approach is to develop Taylor series of the nonlinear functions about the estimated state $\hat{X}_t$, and neglect higher order terms. Depending on whether one only takes first or second order terms into account, a first or a second order filter results. The *truncated second order filter* is given by the following equations:

$$d\hat{X}_t = \hat{F}(\hat{X}_t)dt + \Phi_t[\, dN_t - \hat{H}(\hat{X}_t)dt \,] \tag{5.5}$$

$$\Phi_t = \left\{ P_t^x H'(\hat{X}_t)^T \text{diag}^{-1}[\hat{H}(\hat{X}_t)] + \begin{bmatrix} A \\ O \end{bmatrix} \right\}_t \tag{5.6}$$

$$dP_t^x = \left\{ P_t^x F(\hat{X}_t)^T + F(\hat{X}_t)P_t^x + \begin{bmatrix} A\,\mathrm{diag}[\hat{\Lambda}(\hat{X}_t)]A^T & O \\ O & \Sigma \end{bmatrix} \right.$$

$$\left. - \Phi_t \mathrm{diag}[\hat{H}(\hat{X}_t)]\Phi_t^T \right\} dt. \qquad (5.7)$$

Here $\hat{F}$ is defined as follows:

$$\hat{F}_i(.) = F_i(.) + \frac{1}{2}\sum_j\sum_k \frac{\partial^2 F_i(.)}{\partial x_j \partial x_k}(P_t^x)_{jk}.$$

$\hat{H}$ and $\hat{\Lambda}$ are given by analogous expressions. H' and F' are the Jacobians of H and F respectively.

The filter is called truncated because a martingale term in the P_t^x-equation is neglected. P_t^x is an approximation of the matrix of conditional error covariances

$$E[(X_t - \hat{X}_t)(X_t - \hat{X}_t)^T \mid \mathcal{F}_t^N].$$

In order to compute first or second order approximations to the optimal estimator of X_t it turns out to be necessary to compute (an approximation of) the conditional error covariance matrix as well. This is why equation (5.7) appears. Σ is the covariance matrix of the Brownian motion processes w_i.

To obtain the truncated first order filter one only has to set $\hat{F}_i$ equal to F_i, $\hat{H}_i$ to H_i and $\hat{\Lambda}_i$ to Λ_i, and apply the same equations, (5.5) to (5.7), as for the second order filter.

The dimensions of the various vectors and matrices are as follows: $\hat{X} \in \mathbf{R}^{2n \times 1}$, $N \in \mathbf{R}^{(n+1)m \times 1}$, $\Phi_t \in \mathbf{R}^{2n \times (n+1)m}$, $P_t^x \in \mathbf{R}^{2n \times 2n}$, $A \in \mathbf{R}^{n \times (n+1)m}$, $\Sigma \in \mathbf{R}^{n \times n}$, $\hat{F} \in \mathbf{R}^{2n \times 1}$, $F' \in \mathbf{R}^{2n \times 2n}$, $\hat{H} \in \mathbf{R}^{(n+1)m \times 1}$, $H' \in \mathbf{R}^{(n+1)m \times 2n}$, $\hat{\Lambda} \in \mathbf{R}^{(n+1)m \times 1}$.
Computing $\hat{X}$ therefore involves the integration of $2n+(2n+1)n = 2n^2+3n$ differential equations (P_t^x is symmetric).

The second order filter requires the extra computation of the Hessian matrices of F_i, H_i and Λ_i ($i=1,...,2n$) in comparison to the first order filter. This requires the computation of $6n^2(2n+1)$ extra elements. Although in practice many zeros appear in the Hessians and a well-chosen order of computation may save much computational effort, it is clear that the second order filter is relatively expensive computationally. It is therefore of interest to investigate whether the extra effort is justified by a considerable gain in estimator precision. This question is addressed in Section 5.3.3.

It is well-known that numerical instability may occur in the integration of (5.7) [97]. To reduce the sensitivity to rounding errors so called *square root* algorithms have been developed [66]. Most of these algorithms consider the discrete time case. In the square root algorithm of Morf, Lévy and Kailath [70] for the continuous time filter, the matrix P_t^x is replaced by its Cholesky factor S_t:

$$P_t^x = S_t S_t^T,$$

where S_t is lower triangular. Using this factorization (5.7) may be written as

$$\dot{S}_t S_t^T + S_t \dot{S}_t^T = S_t S_t^T F'(\hat{X}_t)^T + F'(\hat{X}_t) S_t S_t^T$$

$$+ \begin{bmatrix} A\,\mathrm{diag}[\Lambda(X_t)]A^T & O \\ O & \Sigma \end{bmatrix} - \Phi_t \mathrm{diag}[\hat{H}(\hat{X}_t)]\Phi_t^T, \quad (5.8)$$

where a dot denotes differentiation with respect to time. Assuming that $P_t^x > 0$ for all t the inverse of S_t exists. Then pre-multiplication of (5.8) by S_t^{-1} and post-multiplication by S_t^{-T} lead to

$$S_t^{-1}\dot{S}_t + (S_t^{-1}\dot{S}_t)^T = (F'(\hat{X}_t)S_t)^T S_t^{-T} + S_t^{-1}(F'(\hat{X}_t)\, S_t)$$

$$+ S_t^{-1} \begin{bmatrix} A\,\mathrm{diag}[\hat{\Lambda}(\hat{X}_t)]A^T & O \\ O & \Sigma \end{bmatrix} S_t^{-T} - S_t^{-1}\Phi_t \mathrm{diag}[\hat{H}(\hat{X}_t)]\Phi_t^T S_t^{-T}. \quad (5.9)$$

As $S_t^{-1}\dot{S}_t$ is lower triangular we only need to compute the lower triangle of the right-hand side and divide the diagonal by 2. For this we use the notation

$$S_t^{-1}\dot{S}_t = \left[(F'(\hat{X}_t)S_t)^T S_t^{-T} + \cdots \right]_{-/2}$$

and hence

$$\dot{S}_t = S_t\, [\text{ right-hand side of (5.9) }]_{-/2}\,. \quad (5.10)$$

The advantage in integrating (5.10) over integrating (5.7) is that S_t may be interpreted as the square root of P_t^x: the elements are closer together (the difference between two elements, measured in the number of digits, is halved). This reduces the effect of rounding errors, which is especially important when P_t^x is almost singular. The gain in precision is at the cost of increased computational effort. Suppose that the computation of F', $\hat{\Lambda}$, $\hat{H}$, Φ_t requires Δ operations. Then, computing P_t^x according to the original equation (5.7) costs

$$\Delta + 10n^3 + 8mn^2(n+1) \text{ operations}$$

and the square root approach requires

$$\Delta + 18n^3 + 12mn^2(n+1) \text{ operations.}$$

Note that in computing $\hat{\Lambda}$ and $\hat{H}$ and also in computing Φ_t we need P_t^x explicitly and therefore have to perform the multiplication of S_t with its transpose. If the other computations would require relatively little effort (Δ small) the increase due to the square root approach would be over 50 percent. In our case Δ is quite large and the relative effect is smaller, but it is clear that the square root method should only be used if there is a clear indication of possible numerical problems. In our case preliminary investigations showed that P_t^x may become ill-conditioned. This is confirmed by results in Section 5.2.2. We therefore decided to implement the square root algorithm described above.

The square root algorithm proposed by Bar-Itzhack and Oshman [74] is quite

different from the one we just described. Here the eigenvalues and eigenvectors of P_t^x are propagated. The square root differs from the one described above. It is the product of a matrix that has as its columns the eigenvectors of P_t^x, with a diagonal matrix. The diagonal elements of the latter matrix are the square roots of the eigenvalues corresponding to the eigenvectors in the former matrix. Unfortunately, this alternative method is not applicable to the second order filter, because $\hat{\Lambda}$ and $\hat{H}$ require P_t^x explicitly: (5.7) is not of the Riccati type. Moreover, this eigenfactor solution algorithm requires even more computations:

$$\Delta + 22n^3 + 12mn^2(n+1) \text{ operations.}$$

We therefore use the method of Morf, Lévy and Kailath.

To solve (5.5),(5.6) and (5.10) a FORTRAN computer program was written and implemented on a CDC Cyber 750 mainframe. The integration is simply done by Euler's method, where the integration step is dictated by the observed sample path of N_t. If a stretch of four sections of one carriageway of a dual freeway is considered, as in the sequel, and traffic intensity is reasonably large, the resulting integration step is small enough to obtain an accurate solution. More sophisticated methods could be used, but lead to a considerable increase in computational effort. Equations (5.5) and (5.10) are integrated in one step from one jump of a component of N_t to the next jump of one of the components of N_t. Then X_t and Φ_t are updated for the jump in N_t and the next integration step is carried out. In case there is a small number of counting processes, or there are not many jumps (low traffic intensity), the integration step may become too large. To avoid this an upper bound of 0.0001 hour for this step was chosen. This value was tested: the filter turned out to be insensitive to a reduction of the upper bound.

Equation (5.6) requires the inversion of a diagonal matrix. The diagonal consists of the components of $\hat{H}(\hat{X}_t)$, recall equation (2.20):

$$\left[\hat{H}(\hat{X}_t)\right]_{i\times(n-1)+j} = [1+\epsilon_i^f(j)-\epsilon_i^m(j)]\gamma_i^j\hat{\lambda}_i.$$

We may expect numerical problems whenever a counting process intensity $\hat{\lambda}_i^j=\gamma_i^j\hat{\lambda}_i$ becomes small. This may happen when the mean passage speed at a location is far off the center of a speed class. This class then hardly contains any probability mass. Problems are also to be expected when section density or mean speed becomes small. The problems just mentioned have occurred in practice. As a remedy we chose to take a lower bound of 10 veh/h for each of the intensities $\hat{\lambda}_i^j$.

The proposed filter had to undergo another small adjustment: whenever $\hat{\rho}_i<0$ or $\hat{v}_i<0$ occurs during the integration or updating step, these variables are reset to 0. Especially the former may occur because of initial uncertainty about the real density. The filter may estimate the density too low and it may become zero, while in reality there is still a vehicle in the section. This vehicle may leave the section, necessitating a reduction of the estimated density, which would then become negative.

As a final remark we note that instead of modelling the passing speed

distribution as a normal one, as proposed in Chapter 4, a *logistics* distribution will be used [44]. This distribution is known to approximate the normal distribution well, and enjoys an analytical expression:

$$F(v) = \frac{1}{1 + e^{\frac{-\pi(v-\mu)}{\sigma\sqrt{3}}}}.$$

Here μ is the mean and σ the standard deviation. Use of the logistics distribution facilitates the computation of the γ_i^j's. In the future, for the implementation of the filter in practice, a table of a normal distribution may be used.

5.2. Analysis of the filter

Before testing the filter given by (5.5) to (5.7) on data, it may be worthwhile to investigate some of its properties theoretically. Of special interest is the behaviour of the error covariance matrix P_t^x over time. It will turn out that to study this theoretically the detectability and stabilizability properties of associated matrices are of interest. The detectability property is a necessary condition, without which the state cannot be estimated with finite error variance. The stabilizability condition is of importance for the question whether asymptotically exact estimation is possible. The equation for P_t^x, (5.7), and both conditions are discussed in Section 5.2.1. In the second section conclusions about the asymptotic behaviour of the covariance matrix P_t^x are drawn. In 5.2.3 the effect of using speed information is investigated by computing the asymptotic value of the covariance matrix P_t^x.

5.2.1. Detectability and stabilizability.

We are interested in studying the evolution of P_t^x over time. Stated this generally the problem is too difficult to handle. We therefore apply several approximations before coming to the actual analysis. These approximations are discussed next.

The main difficulty with equation (5.7) is that the right-hand side depends on the estimated state X_t. This means that this equation is coupled with (5.5), in which the observations N_t appear. Assumptions about these have to be made. An approximation which does not appear to be too restrictive and which is suggested by practical applications in traffic theory [9] is the following. The traffic intensity is assumed to be piecewise constant over time and over a stretch of freeway of a certain length. During such a period the traffic densities ρ_i and mean speeds v_i of the sections concerned are constant. Furthermore, v_i is approximately equal to the equilibrium value $v^e(\rho)$. The following step then leads us to an equation which can be analyzed. We assume that at the start of a period of constant flow the filter estimates $\hat{\rho}_i$ and $\hat{v}_i$ are close to the real values, such that $d\hat{X}_t \approx 0$, and $\hat{X}_t$ is approximately constant. This makes (5.7) independent of (5.5). To summarize:

ASSUMPTION. *The estimated section density and mean speed are assumed to be nonzero, constant over time and equal in all sections, and the mean speed is assumed to be equal to the equilibrium speed:*

$$\forall\, i \in \{1,..,n\},\ \forall\, t \geqslant t_0 : \hat{\rho}_i(t) = \hat{\rho} \neq 0\ ,\ \hat{v}_i(t) = v^e(\hat{\rho}) \neq 0.$$

A small problem that occurs is that the matrix P_t^x appears in $\hat{\Lambda}$ and $\hat{H}$ on the right-hand side of (5.7). We therefore only consider the first order filter in the following. With these approximations we have reduced (5.7) to a standard *Riccati equation*:

$$\dot{P}_t = P_t K^T + K P_t - P_t G^T R^{-1} G P_t + Q\ , \tag{5.11}$$
$$P_{t_0} = P_0 \geqslant 0\ ,$$

where

$$K = \begin{bmatrix} A\,\mathrm{diag}(\epsilon)\hat{\lambda}' \\ F_2'(\hat{X}) \end{bmatrix}$$
$$G = \hat{\lambda}'$$
$$R = \mathrm{diag}(1-\epsilon)^{-1}\mathrm{diag}(\hat{\lambda}) \tag{5.12}$$
$$Q = \begin{bmatrix} A\,\mathrm{diag}(\epsilon)\mathrm{diag}(\hat{\lambda})A^T & O \\ O & \Sigma \end{bmatrix}$$

and

$$\epsilon = (\bar{\epsilon}, \bar{\epsilon}, \cdots)^T$$
$$\hat{\lambda} = (l\hat{\rho}\hat{v}, l\hat{\rho}\hat{v}, \cdots)^T.$$

Note that $K \in \mathbb{R}^{2n\times 2n}$, $G \in \mathbb{R}^{(n+1)m\times 2n}$, $R \in \mathbb{R}^{(n+1)m\times(n+1)m}$, $Q \in \mathbb{R}^{2n\times 2n}$, $\epsilon \in \mathbb{R}^{(n+1)m}$, $\hat{\lambda} \in \mathbb{R}^{(n+1)m}$. $F_2' \in \mathbb{R}^{n\times 2n}$ and $\hat{\lambda}' \in \mathbb{R}^{(n+1)m\times 2n}$ are the Jacobian matrices of F_2 and $\hat{\lambda}$ respectively. To simplify computations we have assumed that all sections have an equal number of lanes, l, are equally long, counting error fractions are equal to $\bar{\epsilon}$ at all locations and will further neglect the effect of boundary conditions. By $\epsilon > 0$ we denote that all components of ϵ are strictly positive.

We will now investigate the detectability and stabilizability properties of the pairs of matrices $[G,K]$ and $[K,L]$ respectively.

DEFINITION 5.1. *A pair of matrices [G,K] with $G \in \mathbb{R}^{p\times r}$, $K \in \mathbb{R}^{r\times r}$ is called* observable *if*

$$\mathrm{rank}([\ G^T \mid K^T G^T \mid \ldots \mid (K^{r-1})^T G^T\ ]) = r.$$

DEFINITION 5.2. *A pair of matrices* $[G,K]$ *with* $G \in \mathbb{R}^{p \times r}$, $K \in \mathbb{R}^{r \times r}$ *is called* detectable *if*

$$\text{kernel}\left(\begin{bmatrix} G \\ GK \\ GK^2 \\ \vdots \\ GK^{r-1} \end{bmatrix}\right) \subset \chi^-(K), \tag{5.13}$$

where $\chi^-(K)$ denotes the stable subspace of K. In words: the unobservable subspace corresponding to $[G,K]$ has to be contained in the stable subspace of K. As can easily be seen, observability implies detectability. The matrix on the left-hand side of (5.13) is called the *observability matrix.*

PROPOSITION 5.3. *The pair* $[G,K]$, *with* G *and* K *defined in (5.12), is observable.*
PROOF.
Three cases will be considered: *i*) m>1 ; *ii*) m=1, n>1 ; *iii*) m=1, n=1.

i) In case m>1 matrix G may be shown to consist of 2n independent columns with the exception of the case where $(-0.28\frac{\partial \gamma^j}{\partial \sigma}\hat{\lambda} + \hat{v}\hat{\gamma}^j)/(\frac{\partial \gamma^j}{\partial \mu}\hat{\lambda} + \hat{\rho}\hat{\gamma}^j)$ is constant, independent of j. This means that generically rank(G) equals 2n and so the observability condition is satisfied, for any K.

ii) In case m=1 and n>1, $G \in \mathbb{R}^{(n+1) \times 2n}$ so rank(G)<2n and we have to consider $K^T G^T$ and possibly $(K^2)^T G^T$ etc. It turns out that rank($[G^T \mid K^T G^T]$)=2n and observability of $[G,K]$ is guaranteed.

iii) When n=1 and m=1 rank(G) turns out to be smaller than 2n, but considering $K^T G^T$ again leads to observability. □

Note that the absence of speed information (m=1) does not prohibit the possibility of estimating the traffic state. The count data alone apparently contain enough information.

We now come to the question of stabilizability.

DEFINITION 5.4. *A pair of matrices* $[K,L]$ *with* $K \in \mathbb{R}^{r \times r}$, $L \in \mathbb{R}^{r \times r}$ *is called* controllable *if*

$$\text{rank}([\, L \mid KL \mid \ldots \mid K^{r-1}L \,]) = r.$$

DEFINITION 5.5. *A pair of matrices* $[K,L]$ *with* $K \in \mathbb{R}^{r \times r}$, $L \in \mathbb{R}^{r \times r}$ *is called* stabilizable *if*

$$\text{range}([\, L \mid KL \mid \ldots \mid K^{r-1}L \,]) \supset \chi^+(K), \tag{5.14}$$

where $\chi^+(K)$ denotes the unstable subspace of K. Note that controllability implies stabilizability. The matrix on the left-hand side of (5.14) is called the *controllability matrix.*

PROPOSITION 5.6. *For the pair* $[K,L]$, *with* K *defined in (5.12) and a suitable* L, *such that* $LL^T = Q$, Q *defined in (5.12), the following holds:*

– *if* $\epsilon>0$ *and* $\Sigma>0$ *then the pair* $[K,L]$ *is controllable;*

– *if* $\epsilon=0$ *then the pair* $[K,L]$ *is not stabilizable. There are uncontrollable modes corresponding to the eigenvalue 0 of* K*;*

– *if* $\epsilon>0$ *and* $\Sigma=0$ *then the pair* $[K,L]$ *is controllable, possibly with the exception of the case where* $\gamma=0$ *and* $\frac{\dot{v}^e(\hat{\rho})}{T}=0$.

PROOF.
Three cases are considered: *i)* $\epsilon>0$, $\Sigma>0$; *ii)* $\epsilon=0$; *iii)* $\epsilon>0$, $\Sigma=0$.

i) If $\epsilon>0$ and $\Sigma>0$ then Q has a unique positive definite root $Q^{\frac{1}{2}}$. Take $L=Q^{\frac{1}{2}}$, then rank$(L)=2$n, the maximal value. Controllability of $[K,L]$ follows immediately.

ii) In case $\epsilon=0$ it is clear that $L=\begin{bmatrix} O & O \\ O & \Sigma^{\frac{1}{2}} \end{bmatrix}$ and controllability does not follow from the structure of L. The matrices KL etc. have to be taken into consideration. Notice that

$$K=\begin{bmatrix} O \\ F_2'(\hat{X}_t) \end{bmatrix}=\begin{bmatrix} O & O \\ K_1 & K_2 \end{bmatrix}$$

which leads to

$$KL=\begin{bmatrix} O & O \\ O & K_2\Sigma^{\frac{1}{2}} \end{bmatrix}, \quad K^2L=\begin{bmatrix} O & O \\ O & (K_2)^2\Sigma^{\frac{1}{2}} \end{bmatrix} \quad \text{etc.}$$

This means that there are at most n independent rows in $[\, L \,|\, KL \,|\, \cdots \,]$ and so the pair $[K,L]$ is not controllable. Let us now try to establish the weaker property of stabilizability. Instead of investigating the stabilizability of $[K,L]$, however, it turns out to be handier to consider the equivalent property of detectability of $[L^T,K^T]$.

Clearly

$$\begin{bmatrix} L^T \\ L^TK^T \\ \vdots \end{bmatrix}=\begin{bmatrix} O & O \\ O & \Sigma^{\frac{1}{2}} \\ O & O \\ O & \Sigma^{\frac{1}{2}}K_2^T \\ \vdots & \vdots \end{bmatrix} \quad \text{and} \quad K^T=\begin{bmatrix} O & K_1^T \\ O & K_2^T \end{bmatrix}$$

so K^T has a set of eigenvectors corresponding to the unstable eigenvalue 0, which are in the kernel of the observability matrix. This excludes detectability of $[L^T,K^T]$ and therefore $[K,L]$ is not stabilizable.

iii) In case $\epsilon>0$ and $\Sigma=0$ we take

$$L=\begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \quad \text{with} \quad D=D^T>0 \quad \in \mathbb{R}^{n\times n}.$$

The controllability matrix then has the following structure

$$\begin{bmatrix} D & 0 & * & 0 & * & 0 & .. & .. \\ 0 & 0 & K_1 D & 0 & K_2 K_1 D & 0 & .. & .. \end{bmatrix}$$

which means that whenever rank$(K_1 D)$=n controllability is assured. Now it turns out that whenever $\frac{\dot{v}^e(\hat{\rho})}{T} + \gamma (lL)^2 \hat{\rho} \neq 0$ matrix K_1 is invertible and so will $K_1 D$, implying that rank$(K_1 D)$=n. So, in general controllability is assured. There are realistic parameter values and values for $\hat{\rho}$ for which the condition just mentioned is not satisfied, however. A more detailed investigation shows that in case $\frac{\dot{v}^e(\hat{\rho})}{T} + \gamma (lL)^2 \hat{\rho} = 0$ controllability is still assured as long as $\gamma \neq 0$. The unrealistic case $\gamma = 0$, $\frac{\dot{v}^e(\hat{\rho})}{T} = 0$ requires further analysis that will not be carried out here. □

We conclude that whether the matrix pairs concerned are controllable, stabilizable, observable or detectable depends on the model parameters. This reflects itself in the behaviour of the error covariance matrix P_t^x. Its asymptotic behaviour is discussed in the next section.

5.2.2. Asymptotic error covariance.

The results of the previous section may be used to draw conclusions about the asymptotic behaviour of the solution P_t of (5.11), which is an approximation of the covariance matrix P_t^x of (5.7).

In case $\lim_{t\to\infty} P_t$ exists it satisfies the *Algebraic Riccati Equation* (ARE)

$$0 = PK^T + KP - PG^T R^{-1} GP + Q\,. \tag{5.15}$$

It is therefore of interest to investigate whether a solution of the ARE exist, whether it is unique, positive definite, etc. We are only interested in real, symmetric, nonnegative definite solutions. As a by-product of a solution $\bar{P}$ of the ARE the eigenvalues of the matrix

$$\bar{K} = K - \bar{P} G^T R^{-1} G \tag{5.16}$$

are of interest. Writing down the exact equation for the error process $\hat{X}_t - X_t$ and applying some approximations leads to the result

$$\frac{d}{dt}(\hat{X}_t - X_t) \approx (K - \bar{P} G^T R^{-1} G)(\hat{X}_t - X_t) + \text{ noise terms,}$$

which means that stability of $\bar{K}$ guarantees exponential reduction of estimation errors. It is clear that solutions for which $\bar{K}$ is antistable are of no value.

DEFINITION 5.7. *A solution* $\bar{P}$ *of the ARE (5.15) is called* strong *if all the eigenvalues of* $\bar{K}$ *defined by (5.16) lie in the closed left half of the complex plane, and is called* stabilizing *if they all lie in the open left half of the complex plane.*

The classical result about the solution of the ARE is given by the following theorem:

THEOREM 5.8. *If* $[G,K]$ *is detectable and* $[K,L]$ *is stabilizable, where* $LL^T = Q$, $L \in \mathbf{R}^{2n \times 2n}$, *then the ARE (5.15) has a unique nonnegative definite solution* $\bar{P}$. *Furthermore,* $\bar{P}$ *is stabilizing and if* $[K,L]$ *is controllable* $\bar{P}$ *is positive definite.*

PROOF.

See Kuçera [51], Theorem 5 and Wonham [102], Theorem 4.1.

This theorem completely satisfies our needs in case $\epsilon > 0$. When $\epsilon = 0$, however, $[K,L]$ is not stabilizable and a stronger result is needed. We are able to state the following theorem:

THEOREM 5.9. *Assume* $[G,K]$ *to be detectable. Then the following holds:*

i) *The ARE (5.15) has a unique strong solution* $\bar{P}_{strong}$;

ii) $\bar{P}_{strong}$ *is stabilizing if and only if there are no* $[K,L]$*-uncontrollable modes corresponding to purely imaginary eigenvalues of* K;

iii) $\bar{P}_{strong}$ *is positive definite if and only if* $[-K,L]$ *is stabilizable.*

iv) $\bar{P}_{strong}$ *is the only nonnegative solution of the ARE (5.15) if and only if there are no* $[K,L]$*-uncontrollable modes corresponding to eigenvalues of* K *in the open right half plane.*

PROOF.

i) See Poubelle et al. [79].

ii) According to Theorem 3 of Kuçera [51] a stabilizing solution of the ARE exists if and only if $[G,K]$ is detectable and the associated Hamiltonian matrix has no purely imaginary eigenvalues. But this Hamiltonian has such an eigenvalue if and only if there exists a corresponding $[G,K]-$unobservable and/or $[K,L]-$uncontrollable eigenvalue (Lemma 8 in [50]). The detectability of $[G,K]$ excludes the first option, which means that Kuçera's theorem may be reformulated as "A stabilizing solution exists if and only if $[G,K]$ is detectable and there are no $[K,L]-$uncontrollable modes on the imaginary axis". As there is only one strong solution according to *i*), a stabilizing solution has to be equal to P_{strong}.

iii) $\leftarrow$) Theorem 2 of Richardson and Kwong [82] states that in case the Hamiltonian has no purely imaginary eigenvalues, a positive definite solution exists if and only if $[G,K]$ is detectable and $[-K,L]$ is stabilizable. Our statement immediately follows from the fact that $[-K,L]$ stabilizability and $[G,K]$ detectability imply that the Hamiltonian has no purely imaginary eigenvalues (Kuçera

[50], Lemma 8).

→) Consider Theorem 3 of Richardson and Kwong [82]. This theorem states that the class of matrices K, G, L for which here exists a positive definite solution to the ARE consists precisely of those which, with respect to an appropriate basis, have the form

$$K = \begin{bmatrix} K_1 & O \\ O & K_2 \end{bmatrix}, \quad G = [\, G_1 \; O \,], \quad L = \begin{bmatrix} L_1 \\ O \end{bmatrix}$$

where $[G_1, K_1]$ is detectable, $[-K_1, L_1]$ is stabilizable and K_2 is diagonalizable and has only purely imaginary eigenvalues. $[G,K]$ detectability and positive definiteness of a solution of the ARE now imply that $\dim(K_2)=0$ and stabilizability of $[-K_1, L_1]=[-K, L]$ follows.

iv) ←) Apply the construction used in the necessity part of the proof of Theorem 4 of Kuçera [51].

→) First, from *i*) we know that $\bar{P}_{strong}$ is unique. Next one may show that assuming the existence of a nonnegative solution that is not strong implies the existence of an uncontrollable mode corresponding to an eigenvalue of K in the open right half plane: a contradiction. This may be shown analogously to the sufficiency part of the proof of Theorem 4 in [51]. □

REMARK 1. This theorem is the continuous time equivalent of Theorem 3.2 of De Souza et al. [88].

REMARK 2. An independent proof of *i*) and *iii*) is given by Kano [45].

From Theorem 5.9 one may conclude that detectability alone is sufficient in the estimation of the state from observations: the state is estimated with finite error variance. As it turns out, stabilizability is not a necessary prerequisite. The detectability condition assures that at least all unstable system modes are observable.

In the filtering context the stabilizability property of the pair $[K, L]$ has to do with the question whether modes of the system are perturbed by noise or not. If $[-K, L]$ is a stabilizable pair, all modes corresponding to eigenvalues in the closed left half of the complex plane are perturbed by noise. This implies that none of the states is asymptotically estimated exactly, and therefore a strictly positive definite error covariance matrix results from the estimation procedure.

If oscillatory modes occur, corresponding to purely imaginary eigenvalues, and these are not perturbed by noise, they are estimated exactly asymptotically and show up as modes of the filter. This implies that the asymptotic error covariance matrix is not stabilizing, according to Definition 5.7.

Theorem 5.9 leads to the conclusion that a unique strong solution exists in the case $\epsilon = 0$, but the solution is neither stabilizing nor positive definite. In fact, the strong solution may easily be shown to be

$$\bar{P}_{strong} = \begin{bmatrix} O & O \\ O & P_3 \end{bmatrix},$$

where P_3 is the unique positive definite solution of

$$O = P_3 K_2{}^T + K_2 P_3 - P_3 G_2{}^T R^{-1} G_2 P_3 + \Sigma$$

in case $\Sigma > 0$, and $P_3 = 0$ if $\Sigma = 0$. Here $K = \begin{bmatrix} O & O \\ K_1 & K_2 \end{bmatrix}$ and $G = [G_1\ G_2]$. The singularity of P_{strong} is caused by the fact that the section densities (and possibly the mean speeds) are estimated exactly asymptotically. In this case

$$\bar{K} = \begin{bmatrix} O \\ F_2' \end{bmatrix} - \begin{bmatrix} O \\ P_3 G_2^T R^{-1} G_1 + P_3 G_2^T R^{-1} G_2 \end{bmatrix},$$

which implies that

$$d\hat{\rho}_i = 0 \quad \text{for } i = 1, \ldots, n.$$

This is logical once $\hat{\rho}_i = \rho_i$. Therefore, the non-stabilizability and singularity should not bother us.

Let us now turn to the convergence properties of the solution P_t of the Riccati differential equation (5.11). In general nonnegative solutions of the ARE (5.15), different from $\bar{P}_{strong}$ exist. This may lead P_t to converge to a "wrong" asymptotic value, if the initial value P_{t_0} is not chosen carefully. As part *iv*) of Theorem 5.9 shows, however, absence of antistable $[K, L]$-uncontrollable modes guarantees uniqueness of a nonnegative solution. In our problem $[K, L]$ is controllable if $\epsilon > 0$ and one may show by inspection that the uncontrollable modes that occur when $\epsilon = 0$ are all located in the closed left half plane. P_{strong} is the only nonnegative solution of (5.15) therefore. Unfortunately, to prove convergence of P_t to this unique nonnegative solution is not easy in general [37]. For the case $\epsilon > 0$ the following classical result is available:

THEOREM 5.10. *If* $[K,L]$ *is stabilizable and* $[G,K]$ *is detectable, where* $LL^T = Q$, $L \in \mathbf{R}^{2n \times 2n}$, *then* $\lim_{t \to \infty} P_t = \bar{P}_{strong}$ *for any* $P_{t_0} \geqslant 0$.

PROOF.
See Kuçera [51], Theorem 17. □

In case $\epsilon = 0$ the following result is of interest:

THEOREM 5.11. *If* $[G,K]$ *is detectable and* $P_{t_0} \geqslant \bar{P}_{strong}$ *then* $\lim_{t \to \infty} P_t = \bar{P}_{strong}$.

PROOF.
See Theorem 3 of Poubelle et al. [79]. □

Taking P_{t_0} large enough thus assures convergence to the right solution.

Finally, the following monotonicity result is of interest for the filtering experiments in the next sections:

THEOREM 5.12. *Assume that* $[G,K]$ *is detectable and let* $\bar{P}$ *be the unique strong solution of (5.15). Consider a corresponding Algebraic Riccati Equation where* Σ *in* Q *is*

replaced by $\Sigma_1 < \Sigma$. *Let* P_1 *be the unique strong solution of this equation. Then* $P_1 < \bar{P}$.

PROOF.

For the proof of $P_1 \leqslant \bar{P}$ see Wimmer [101], Theorem 1. Now suppose that $P_1 = P$, then from subtraction of the corresponding Riccati equations it follows that $\Sigma_1 = \Sigma$, which is in contradiction with the assumption. □

To summarize:

—if $\epsilon > 0$ the solution of the Riccati differential equation (5.11) with initial condition $P_{t_0} \geqslant 0$ converges to the unique stabilizing and positive definite solution of the algebraic Riccati equation (5.15);
—if $\epsilon = 0$ the solution of the Riccati differential equation (5.11) converges to the unique nonnegative and strong solution $\bar{P}_{strong}$ of the algebraic Riccati equation (5.15) if $P_{t_0} \geqslant \bar{P}_{strong}$. $\bar{P}_{strong}$ is neither stabilizing nor positive definite.

In our filter applications the components of ϵ will generically be larger than 0, but small (in the order of 0.02). Numerical problems in the computation of P_t^x may therefore be expected. It is for this reason that a square root algorithm as mentioned in Section 5.1 is implemented.

Some idea about the magnitude of the asymptotic estimation errors is given in the next section.

5.2.3. Speed information and estimator accuracy.

The results of the previous section may now be used to investigate the effect of speed information on the accuracy of the estimates.

We may assume $\Sigma > 0$ and $\epsilon > 0$ to hold in practice, which means that the asymptotic error variances are larger than 0. The magnitude of the errors depends on the amount of information available in the estimation procedure: the more detailed the information, the more accurate the estimates are.

The amount of information clearly depends on the number of speed classes we use: m. If m = 1 no speed information is contained in the measurements and in the limiting case where $m \to \infty$ the entire passing speed distribution is available.

The solution $\bar{P}$ of the algebraic Riccati equation (5.15) has been computed using the interactive package MATLAB-SC [69]. Parameters are as in Table 5.1 except that because of filtering results to be presented in the next section α was set to 0.5 and γ to 1.0. Furthermore, we took n = 4, $l = 2$, $L = 0.5$, $\Sigma = \text{diag}(10000)$ and $\bar{\epsilon} = 0.015/\text{m}$ in accordance with values used in the filter tests of Section 5.3.

Now m is varied from 1 to 3 and $\hat{\rho}$ is equal to 20.0, 30.0 and 40.0 veh/km/lane respectively. If m = 2 the boundary between the two speed classes is chosen to be equal to $v^e(\hat{\rho})$, which implies that both classes contain an equal amount of probability mass. If m = 3 the boundaries are also chosen such that each class contains equal probability mass. The results of the computations are given in Table 5.2 and Figures 5.1 and 5.2. In Figure 5.1 the error standard deviation in $\hat{\rho}$ for

several values of m and ρ is plotted. In Figure 5.2 the same is done for $\hat{v}$. In Table 5.2 the diagonal elements of the covariance matrix are listed, these are approximations of error variances in the state estimates. In Figures 5.1 and 5.2 lines are drawn to obtain a clear plot, but note that m can only be integer valued. From the results the following conclusions may be drawn:

—increasing the number of speed classes reduces the error variance, especially when density is high;
—the reduction in error variance is considerable when going from 1 to 2 speed classes, and negligible when going from 2 to 3 classes;
—the minimal value of the standard deviation of the error in $\hat{\rho}$ is about 2 veh/km/lane and in $\hat{v}$ about 4 km/h.

In case of an ideal homogeneous and stationary traffic stream, taking two speed classes apparently is necessary and sufficient to produce accurate estimates. In practice, where the intensity varies considerably, more classes may be necessary. This is investigated in the next section.

ρ (veh/km/lane)	20.0			30.0			40.0		
m	1	2	3	1	2	3	1	2	3
$\widehat{\text{var}}(\rho_1 - \hat{\rho}_1)$	6.5	4.5	3.4	16.7	3.1	2.4	25.8	5.2	4.1
$\widehat{\text{var}}(\rho_2 - \hat{\rho}_2)$	7.7	5.3	4.1	18.6	3.8	3.1	30.7	6.0	4.7
$\widehat{\text{var}}(\rho_3 - \hat{\rho}_3)$	7.7	5.3	4.1	21.3	3.9	3.9	58.6	5.9	4.7
$\widehat{\text{var}}(\rho_4 - \hat{\rho}_4)$	6.0	4.2	3.2	19.8	3.4	4.1	83.7	5.0	4.0
$\widehat{\text{var}}(v_1 - \hat{v}_1)$	17.4	13.7	13.3	44.1	13.8	13.0	38.0	13.2	12.4
$\widehat{\text{var}}(v_2 - \hat{v}_2)$	21.0	14.2	13.7	51.8	14.1	13.4	47.7	13.6	12.8
$\widehat{\text{var}}(v_3 - \hat{v}_3)$	22.0	14.2	13.7	67.4	14.1	13.9	92.1	13.5	12.8
$\widehat{\text{var}}(v_4 - \hat{v}_4)$	22.5	14.5	14.0	82.4	13.7	16.3	137.6	13.0	12.2

TABLE 5.2. Approximate asymptotic error variances.

We may already gain some insight in the number of classes needed by repeating the previous computations for m=2, moving the boundary between the two speed classes, v^1, away from $v^e(\hat{\rho})$. In the limiting cases where either $v^1 = 0$ or $v^1 \to \infty$ we are in fact dealing with just one class, expressing itself in a large error variance. Hence, if we move v^1 away from $v^e(\hat{\rho})$ we expect a gradual deterioration of the estimates.

The results for $\hat{\rho}$=30 veh/km/lane and v^1 = 57.5, 67.5, 77.5 (= v^e), 87.5 and 97.5 km/h are shown in Table 5.3. Note that in case v^1=57.5 and 97.5 km/h only 2 percent of the probability mass of the passing speed distribution is in one class and 98 percent is in the other.

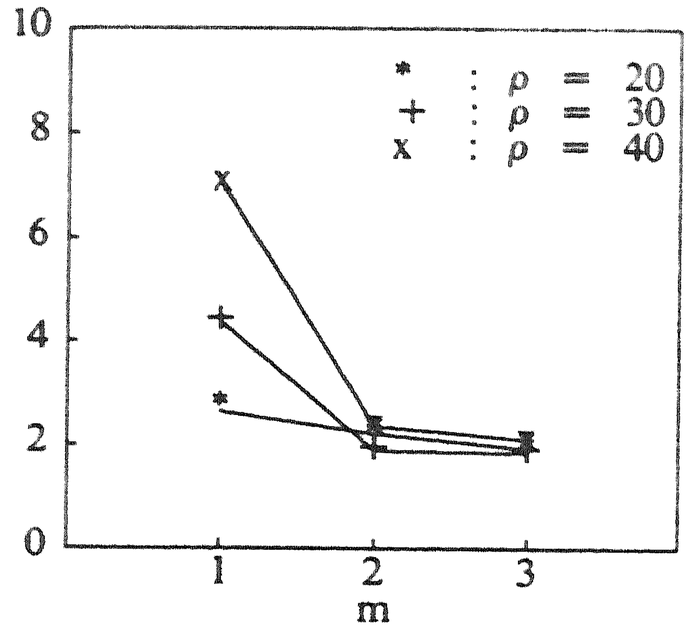

FIGURE 5.1.
Standard dev. in $\hat{\rho}$ (veh/km/lane).

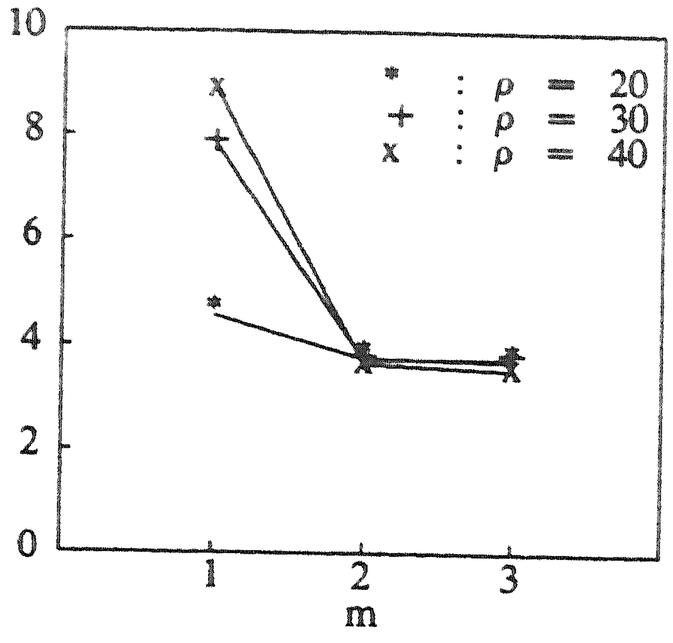

FIGURE 5.2.
Standard dev. in $\hat{\nu}$ (km/h).

	ν^1 (km/h)				
	57.5	67.5	77.5	87.5	97.5
$\tilde{\text{var}}(\rho_1 - \hat{\rho}_1)$	8.3	5.0	3.1	2.9	4.2
$\tilde{\text{var}}(\rho_2 - \hat{\rho}_2)$	10.8	7.0	3.8	3.7	5.6
$\tilde{\text{var}}(\rho_3 - \hat{\rho}_3)$	10.4	7.4	3.9	3.8	5.6
$\tilde{\text{var}}(\rho_4 - \hat{\rho}_4)$	9.4	6.0	3.4	3.1	4.5
$\tilde{\text{var}}(\nu_1 - \hat{\nu}_1)$	28.2	19.0	13.8	16.4	22.2
$\tilde{\text{var}}(\nu_2 - \hat{\nu}_2)$	29.1	19.6	14.1	16.8	23.6
$\tilde{\text{var}}(\nu_3 - \hat{\nu}_3)$	28.5	19.6	14.1	16.7	23.6
$\tilde{\text{var}}(\nu_4 - \hat{\nu}_4)$	30.4	19.2	13.7	16.5	23.5

TABLE 5.3. The effect of speed class location on error variance.

From Table 5.3 the following conclusions may be drawn:

—moving away ν^1 from $\nu^e(\hat{\rho})$ generally increases estimator error variance;
—ν^1 may be far off the center of the speed distribution (>20 km/h) while the error variance remains small;
—the effect is asymmetric: decreasing ν^1 is worse than increasing it. This means that to achieve a uniform accuracy, classes should be narrower at lower speeds and broader at higher speeds.

From the second conclusion it follows that a small number of speed classes will

probably suffice in all circumstances.

5.3. Application of the filter

Now that some basic knowledge about the filter has been obtained, its behaviour on data may be investigated. Performance criteria are proposed in 5.3.1, before reporting on results of filtering simulated data in 5.3.2. It is to be expected that modelling errors play a role in practice: part of the model for freeway traffic is not based on physical laws but on heuristic and limited experimental knowledge about driver behaviour. Therefore experiments with real data are essential to determine the usefulness of the filter in practice. In 5.3.3 sensitivity results on short periods of real data are presented. A validation study on sets of one hour of freeway data is carried out in 5.3.4.

5.3.1. Performance criteria.

We will now discuss several criteria according to which the test results of the next sections are evaluated.

In case the filter is applied to simulated data the real state is available and direct comparison with the estimates is possible. In this case we will use as a measure of the distance between the actual and the estimated sample path:

$$D(\hat{x}_i) = \{\frac{1}{T}\int_{t_0}^{t_0+T} [x_i(t)-\hat{x}_i(t)]^2 dt\}^{\frac{1}{2}} \qquad (5.17)$$

for $i = 1,...,2n$, where

T : length of the period considered;

t_0 : starting time;

x_i : simulated state value (section density or mean speed);

$\hat{x}_i$: estimated state value;

n : number of freeway sections considered.

Unfortunately, in testing with real data no information about the real state variables is available. We can still use the distance D to compare the results of two different filters applied to the same data set, and thereby assess the sensitivity, but it is also necessary to obtain some insight in the actual errors. One may compute the averages of speed and intensity over measuring locations directly from the data (without using the filter) and compare them with weighted filter estimates. Define

$\overline{v}_i^{\delta}(t)$: the harmonic mean of the measured passing speeds during the interval $[t-\frac{1}{2}\delta, t+\frac{1}{2}\delta]$ at location i

$\overline{\lambda}_i^{\delta}(t)$: the number of vehicles that passed location i during the interval $[t-\frac{1}{2}\delta, t+\frac{1}{2}\delta]$, divided by δ (5.18)

$$\hat{\bar{v}}_i(t) := \alpha\hat{v}_i(t) + (1-\alpha)\hat{v}_{i+1}(t)$$

$$\hat{\lambda}_i(t) := (1-\epsilon_i^m+\epsilon_i^f)l_i[\alpha\hat{\rho}_i(t)+(1-\alpha)\hat{\rho}_{i+1}(t)][\alpha\hat{v}_i(t)+(1-\alpha)\hat{v}_{i+1}(t)]$$

for $i=0,...,n$. Here δ is the length of the time period over which the detector data is to be averaged, and α a weighting constant.

A comparison of $\hat{\bar{v}}_i$ and $\bar{v}_i^\delta$ and of $\hat{\lambda}_i$ and $\bar{\lambda}_i^\delta$ gives some indication of the quality of the estimates. Note that we may compute the distance (5.17) for these processes to obtain quantitative results. This procedure is also followed by Cremer [20]. Of course, the results are sensitive to the specific choice for δ and α, so these have to be chosen with care.

As for $\bar{v}_i^\delta$ the following argument may suffice. The mean speed $\hat{v}_i$ is defined over the number of vehicles in section i and $\hat{v}_{i+1}$ over the number in section $i+1$. In all the data sets considered this number will vary between 10 and 40 approximately. This corresponds to an equilibrium intensity of 992 and 2034 veh/h/lane respectively. So, a time period of 18 to 35 seconds is needed for a group of 10 to 40 vehicles to pass over a measuring location. We conclude that if we take $\delta=25$ seconds, the average $\bar{v}_i^\delta$ is taken over about the same group of vehicles as $\hat{\bar{v}}_i$.

For $\bar{\lambda}_i^\delta$ we will take a larger value of δ (1 minute), however, to express our interest in the longer term fluctuations of the intensity.

The value of α follows from Table 5.1. The first experiment with real data, Experiment 5.5, leads to a choice of $\alpha=0.5$, however.

Comparing the estimated speed $\hat{\bar{v}}_i$ and intensity $\hat{\lambda}_i$ with real values in the manner described above is a bit crude. One might argue that what we are doing is to assess the quality of the estimates, $\hat{\bar{v}}_i$ and $\hat{\lambda}_i$, of a nearly optimal filter, by comparing them to the values, $\bar{v}_i^\delta$ and $\bar{\lambda}_i^\delta$, of an inferior estimation procedure. We have to be careful in drawing conclusions from these comparisons. The local estimate criterion may be used to detect systematic errors, however.

A method of assessing the quality of the estimates which is sounder than the previous one consists of computing and maximizing the *likelihood ratio*. This is not attempted here. Instead we study the properties of the *innovations process*. Recall the filter equation (5.3). The process

$$I_t = N_t - N_{t_0} - \int_{t_0}^{t} E[H(X_s)\,|\,\mathcal{F}_s^N]ds \tag{5.19}$$

is called the innovations process and enjoys some important properties.

THEOREM 5.14. *For the innovations process defined by (5.19) the following holds:*

i) $E[I_t] = 0$

ii) $E[(I^j_{t_2} - I^j_{t_1})(I^j_{t_4} - I^j_{t_3})] = 0$,

for $t_1 \leqslant t_2 < t_3 \leqslant t_4$ *and* $1 \leqslant j \leqslant (n+1)m$;

iii) $\mathrm{var}[I_t^j - I_s^j] = \int_s^t E[H^j(X_s)]ds = \mathrm{var}[M_t^j - M_s^j]$,

for $t_0 \leqslant s \leqslant t$ *and* $1 \leqslant j \leqslant (\mathrm{n}+1)\mathrm{m}$.

PROOF.

i) and *ii*) follow from the fact that $\{I_t, t \geqslant t_0\}$ is a martingale with respect to the history $\{\mathcal{F}_t^N, t \geqslant t_0\}$. Note that $N_t = \int_{t_0}^t E[H(X_s)|\mathcal{F}_s^N]ds + I_t$ is the Doob-Meyer decomposition of $\{N_t, t \geqslant t_0\}$ with respect to $\{\mathcal{F}_t^N, t \geqslant t_0\}$.

iii) follows by noting that $\mathrm{E}[(I_t^j - I_s^j)^2] = \mathrm{E}[<I^j, I^j>_t - <I^j, I^j>_s]$ where $<I^j, I^j>$ is the predictable variation of I^j. Now $<I^j, I^j>_t = \int_{t_0}^t E[H^j(X_s)|\mathcal{F}_s^N]ds$.

□

In this theorem *i*) and *ii*) are most important. Violation of *i*) indicates bias and uncorrelated innovation increments indicate non-optimality of the estimates. The sum over j of the variances in *iii*) is a measure of the distance between the predicted and the real output value on the interval $[s,t]$. The three properties in the theorem will be checked during the filter tests.

In applying the innovations later on, we compute I_t^j (componentwise) only at jump times, which means that in fact we are studying the process $\{I^j_{T_n^j}, n \geqslant 0\}$ where $\{T_n^j, n \geqslant 0\}$ are the jump times of $\{N_t^j, t \geqslant t_0\}$. For this process the same properties hold as for the original process, thanks to the *optional sampling theorem* [13], or the Appendix, assuming that we confine ourselves to finite time intervals. This is necessary to satisfy the uniform integrability condition in the theorem. We will not compute the increment of I_t^j over a single jump interval, but over k intervals:

$$I^j_{T^j_{ik}} - I^j_{T^j_{(i-1)k}}, \qquad i = 1,2,\ldots$$

The parameter k will be chosen such that the longer term behaviour of the intensity processes is considered, just as δ was chosen in the local estimate criterion.

The following statistics are computed:

$$\mu_k^j = \frac{1}{k}\left\{\frac{1}{n}\sum_{i=1}^{n}(I^j_{T^j_{ik}} - I^j_{T^j_{(i-1)k}})\right\}$$

$$(\sigma^2)_k^j = \frac{1}{k}\left\{\frac{1}{n}\sum_{i=1}^{n}(I^j_{T^j_{ik}} - I^j_{T^j_{(i-1)k}} - k\mu_k^j)^2\right\}$$

$$\rho_{k,l}^j = \frac{1}{nk(\sigma^2)_k^j}\left\{\sum_{i=1}^{n-l}(I^j_{T^j_{(i+l)k}} - I^j_{T^j_{(i+l-1)k}} - k\mu_k^j)(I^j_{T^j_{ik}} - I^j_{T^j_{(i-1)k}} - k\mu_k^j)\right\}$$

The division by k in the above formulas is based upon the following approximate derivations. Assuming that λ^j is constant and $\hat{\lambda}^j \approx \lambda^j$, we have

$$N^j_{T^j_{i\,k}} - N^j_{T^j_{(i-1)k}} - \int_{T^j_{(i-1)k}}^{T^j_{i\,k}} \hat{\lambda}^j dt \approx k \left[\frac{\lambda^j - \hat{\lambda}^j}{\lambda^j} \right] \tag{5.20}$$

because $E[T^j_{i\,k} - T^j_{(i-1)k}] = k/\lambda^j$. So

$$\frac{1}{k}\left(I^j_{T^j_{i\,k}} - I^j_{T^j_{(i-1)k}}\right) \approx \frac{\lambda^j - \hat{\lambda}^j}{\lambda^j}$$

This means that by μ^j we estimate the *relative error* in $\hat{\lambda}^j$. Furthermore,

$$\mathrm{var}\left(I^j_{T^j_{i\,k}} - I^j_{T^j_{(i-1)k}}\right) \approx k\frac{\hat{\lambda}^j}{\lambda^j} \approx k$$

which means that we expect $(\sigma^2)^j_k$ to be close to one.

Applying the innovations method for each counting process component N^j_t turned out to be problematic in practice: when the real intensity corresponding to N^j_t becomes small and the estimate $\hat{\lambda}^j$ differs from λ^j, the resulting innovations increment is far from zero. To see this consider the expression (5.20). If $\hat{\lambda}^j \neq \lambda^j$ and $\lambda^j \to 0$ then the innovations increment approaches $\pm\infty$. Note that λ^j is the intensity of the counting process corresponding to speed class V^j and may become negligible, even when traffic is dense. In general we are not interested in the relative error when λ^j is small, a small absolute error in $\hat{\lambda}^j$ is allowed. We have just seen that even such a small error in $\hat{\lambda}^j$ may lead to a large error in the innovation increment. It would be preferable to work with the absolute error instead of the relative one.

We have experimented with several modifications of the innovations process, none of which were very satisfying. The best thing to do turns out to be to select periods a priori for which λ^j is larger than some lower bound and compute the innovations over these periods only. This is quite involved, however, and the choice of the lower bound is ad hoc. The conclusion is to use the innovations only when the intensity cannot become small. This means that we only compute them per location, summed over all classes j. Only the quality of the intensity estimate is investigated in this way and we will have to use the *local estimate* method, defined at the beginning of this section, to check for the speed estimate. In the following the statistics corresponding to the cumulative innovations increments will be denoted by μ^i_k, $(\sigma^2)^i_k$ and $\rho^i_{k,l}$ for $i = 0,\ldots,n$.

Finally, we have to note another restriction in the use of the innovations criterion. It has no use trying to optimize the number of speed classes by running different filters and comparing the innovations. The amount of information provided to the filter influences the properties of the innovations process. Assuming that the model is not perfect, more detailed information is more likely to contradict it. Therefore, the more detailed the information is, the more difficult it will be to achieve an uncorrelated process with minimal variance and zero mean. A perfect innovations process in case $m = 1$ may be less desirable than a correlated innovations process in case $m = 2$. The two processes just cannot be

compared. An illustration of this is in Experiment 5.1 in the next section, where the mean, variance and correlation in case m = 1 are comparable to those in case m = 2 (Table 5.5), but the D's are clearly worse (Table 5.4).

5.3.2. Results with simulated data.

The filter generates its estimates of the state of traffic partly on the basis of measurement information and partly on the basis of the model developed in the preceding chapters. Using this model we may generate artificial measurements and feed these into the filter. Details about the simulation procedure and examples were given in Chapter 3. Applying the filter to model generated measurements implies that modelling errors play no role, and one may obtain optimal insight in other errors involved:

- errors due to the 1st or 2nd order approximation to optimal filter (bias);
- errors due to lack of information in the measurements (error variance).

A further advantage of using model generated measurements is the availability of the exact state values for comparison with the estimates. We now present the results of three experiments that illustrate the errors mentioned above.

The second source of error mentioned above was already addressed, in an approximate way, in Section 5.2.3. There it was found that in case of perfectly stationary and homogeneous traffic the asymptotic estimation error could be considerably reduced by using speed information in the estimation procedure. This result is confirmed by the following experiment.

EXPERIMENT 5.1 (ASYMPTOTIC ERROR VARIANCE). The traffic model of Chapter 3 is simulated for one carriageway of a dual two-lane freeway of 4 sections, 0.5 km each, over a period of 15 minutes. There are no on- or off-ramps. The initial values are $\rho(t_0)=30$ veh/km/lane and $v(t_0)=77.5$ km/h and for the entrance boundary a constant intensity of $\lambda_0=2325$ veh/h/lane was prescribed, whereas for the exit stationarity was assumed. Model parameters are chosen according to Table 5.1. There are no counting errors. The first order filter with the improvements that are explained in Section 5.3.3 is applied to the generated measurements in three different cases: m = 1, 2 and 3. Recall that m is the number of speed classes. In case m = 2 or 3 the classes are chosen as in Section 5.2.3: [0.0,77.5) and [77.5,∞) and [0.0,73.4) , [73.4,81.4) and [81.4,∞) respectively. In stationary traffic an equal amount of probability mass falls in each of the classes in this way and in that sense the choice is optimal. The filter is started in the correct state but with a large initial error variance to represent initial uncertainty about the real state. For all three cases the distance D defined by equation (5.17) is tabulated in Table 5.4. Note that the main improvement is in using two speed classes instead of one. Comparing the results here with those of Section 5.2.3, Table 5.2 shows us that the errors in the speeds are comparable and the errors in the densities are somewhat smaller because of the absence of counting errors. Note that Table 5.2 presents variances and Table 5.4 presents standard deviations. The statistics presented in Table 5.5 illustrate the remark in the last

paragraph of the Section 5.3.1.

$D(.)$	m=1	m=2	m=3
$\hat{\rho}_1$	6.4	1.5	1.3
$\hat{\rho}_2$	1.5	0.9	0.9
$\hat{\rho}_3$	2.5	1.5	1.4
$\hat{\rho}_4$	1.9	1.0	1.0
$\hat{\nu}_1$	16.6	4.1	3.3
$\hat{\nu}_2$	8.2	3.7	3.8
$\hat{\nu}_3$	7.7	4.6	4.5
$\hat{\nu}_4$	6.4	2.9	2.8

TABLE 5.4. Estimation errors in Experiment 5.1.

	μ^i_{60}		$(\sigma^2)^i_{60}$		$\rho^i_{60,1}$	
i	m=1	m=2	m=1	m=2	m=1	m=2
0	0.00	-0.05	0.63	1.11	-0.21	-0.33
1	0.04	-0.03	0.64	0.42	0.22	0.18
2	0.05	-0.00	0.61	0.65	0.08	0.08
3	0.00	-0.02	0.59	0.68	0.09	0.15
4	0.01	-0.01	0.58	0.74	-0.11	0.04

TABLE 5.5. Innovation results of Experiment 5.1.

To obtain an impression about the magnitude of bias introduced by the approximations to the optimal filter and about the speed of convergence the following experiment is carried out.

EXPERIMENT 5.2 (BIAS). A series of measurements is generated, not by means of the model, but corresponding to a perfectly stationary and homogeneous traffic stream of 15 minutes. All vehicles drive at constant speed of 77.5 km/h and at a constant distance which is such that the intensity at any location along the freeway equals the equilibrium value (2325 veh/h/lane). Now vehicles pass over section boundaries almost simultaneously and the density is approximately constant over time and equal to 30 veh/km/lane. Starting the first order filter that will result from the experiments in Section 5.3.3, with a large initial uncertainty, we may investigate convergence speed and estimator bias. We compare the cases m=1 and m=2. The final state estimates are shown in Table 5.6 and the final approximate error standard deviations in Table 5.7. We conclude that using two speed classes reduces the bias considerably. The error variances are much smaller also. Convergence of the estimates in Table 5.6 was confirmed by extending the estimation period to 30 minutes. In both case m=1 and m=2 the error variances of the densities have to approach zero. The figures in Table 5.7 indicate that in case m=1 convergence is much slower.

	m=1	m=2
$\hat{\rho}_1$	28.7	30.3
$\hat{\rho}_2$	26.8	29.4
$\hat{\rho}_3$	27.0	30.0
$\hat{\rho}_4$	25.9	29.3
$\hat{v}_1$	82.4	78.2
$\hat{v}_2$	85.0	78.6
$\hat{v}_3$	86.8	78.1
$\hat{v}_4$	88.0	78.6

TABLE 5.6. State estimates of Experiment 5.2.

st.dev.	m=1	m=2
$\hat{\rho}_1$	2.0	0.5
$\hat{\rho}_2$	2.2	0.7
$\hat{\rho}_3$	2.6	0.8
$\hat{\rho}_4$	1.4	0.5
$\hat{v}_1$	9.5	3.4
$\hat{v}_2$	6.9	3.6
$\hat{v}_3$	6.4	3.6
$\hat{v}_4$	5.7	3.1

TABLE 5.7. Estimation errors in Experiment 5.2.

We now show the improved *robustness* of the filter resulting from using speed information.

EXPERIMENT 5.3 (ROBUSTNESS W.R.T. EQUILIBRIUM SPEED). A model based series of measurements is generated like in Experiment 5.1. The first order filter resulting from Section 5.3.3 is applied with an erroneous $v^e(\rho)$-relation: $a=-0.1$ instead of -0.58 and $\rho_{crit}=105$ instead of 27 veh/km/lane. From Figure 5.3 one notes the difference in using two speed classes instead of one. The dotted line represents the simulated speed values. In case m=2 the modelling error is compensated using the speed measurements. We conclude that using a certain amount of speed information is necessary to achieve filter robustness with respect to modelling errors.

EXPERIMENT 5.4 (ROBUSTNESS W.R.T. PASSING SPEED DISTRIBUTION). A model based series of measurements is generated as in the previous experiment, except that now $\lambda^0=1868$ veh/h/lane, which corresponds to $\rho\approx 20$ veh/km/lane and $v\approx 93.4$ km/h. The passage speed distribution is modelled as a Weibull distribution [44]:

$$F(v)=\begin{cases} 1-e^{-\left[\frac{v-a}{b}\right]^c}, & v\geqslant a \\ 0, & v<a \end{cases}$$

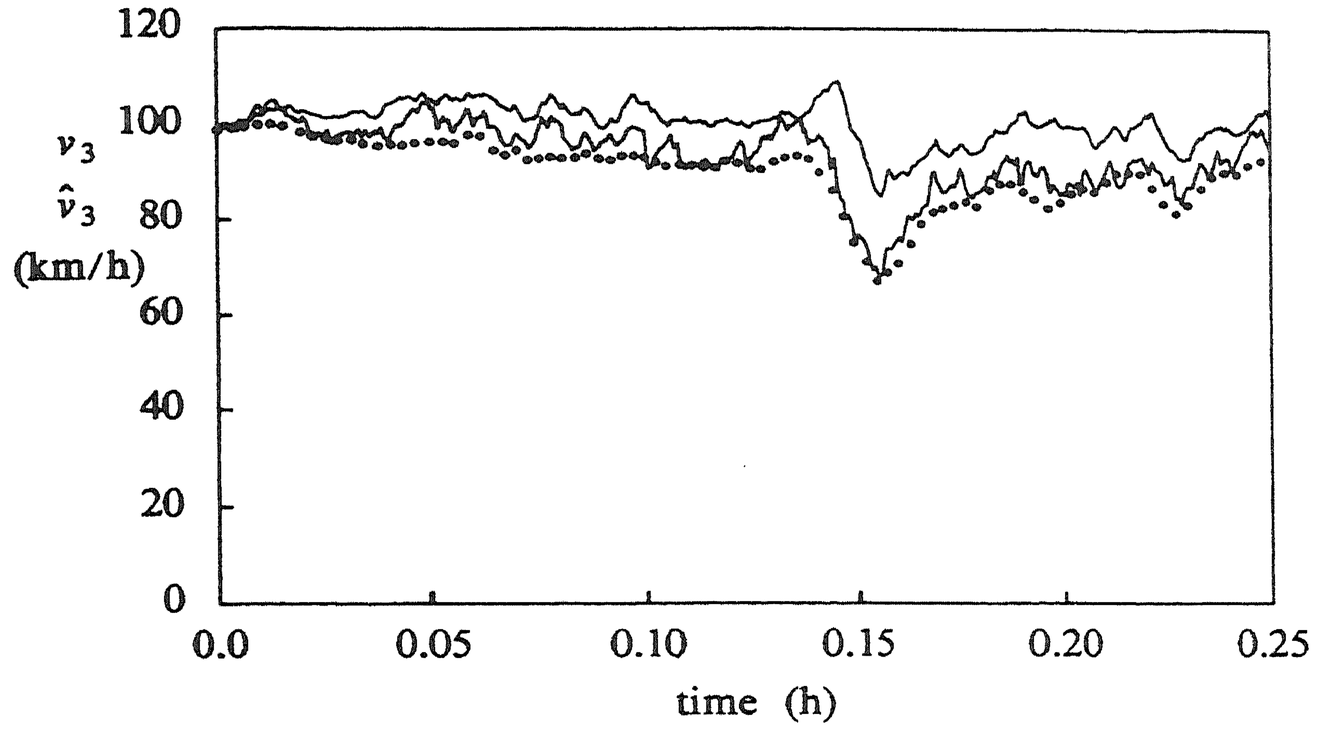

FIGURE 5.3. Simulated and estimated speed trajectories of Experiment 5.3.

with

$$a = \mu - 11.4$$
$$b = 11.8$$
$$c = 1.1$$

to achieve a mean of μ and a standard deviation of 11.8. The mean is modelled as in Chapter 4:

$$\mu = \alpha v_i + (1-\alpha)v_{i+1}$$

at location i. The distribution is such that mean and median do not coincide:

$$\mu - \text{MED} = 2.9.$$

See Figure 5.4 for a plot of the Weibull distribution and a logistics distribution with the same mean and standard deviation. Recall that the filter assumes the passing speed distribution to be logistic. Two speed classes are taken: [0,80) and $[80,\infty)$.

Due to the erroneous passing speed distribution the filter tends to estimate the speed too low. This would lead to an intensity which is too low also, and to compensate this the filter increases traffic density. This is illustrated in Tables 5.8, 5.9 and Figure 5.5. Although speed information is available, because of the fact that the error is in the measurement equation the filter is not able to correct it. In Section 5.3.3. a modification of the filter is proposed which reduces the bias caused by modelling errors in the passing speed distribution and leads to a better performance.

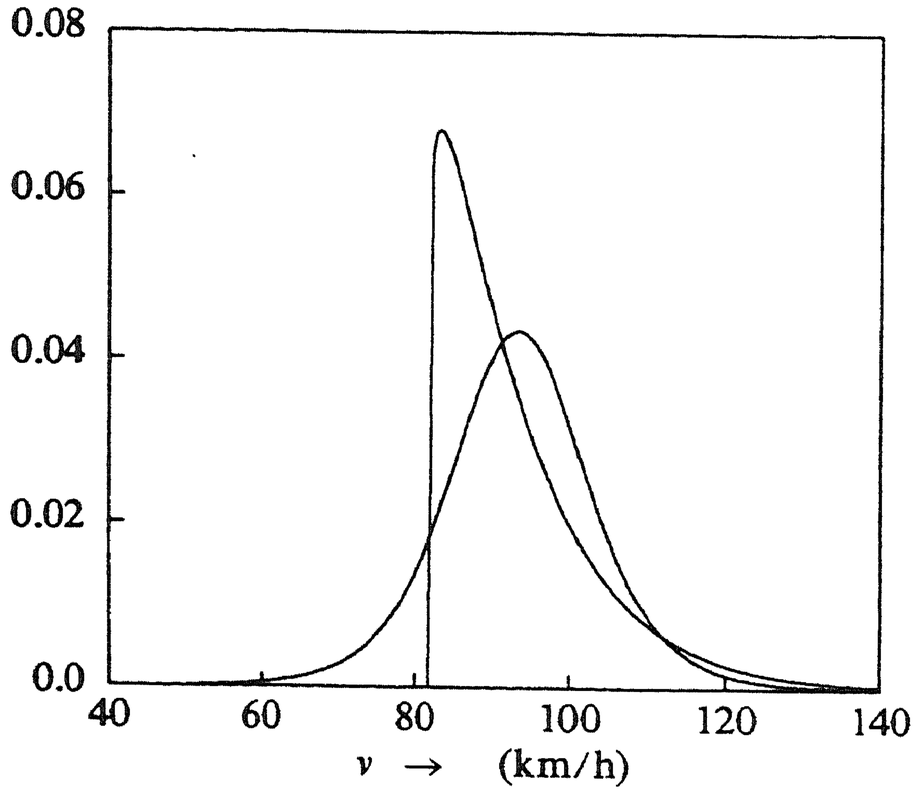

FIGURE 5.4. Normal and Weibull probability density of Experiment 5.4.

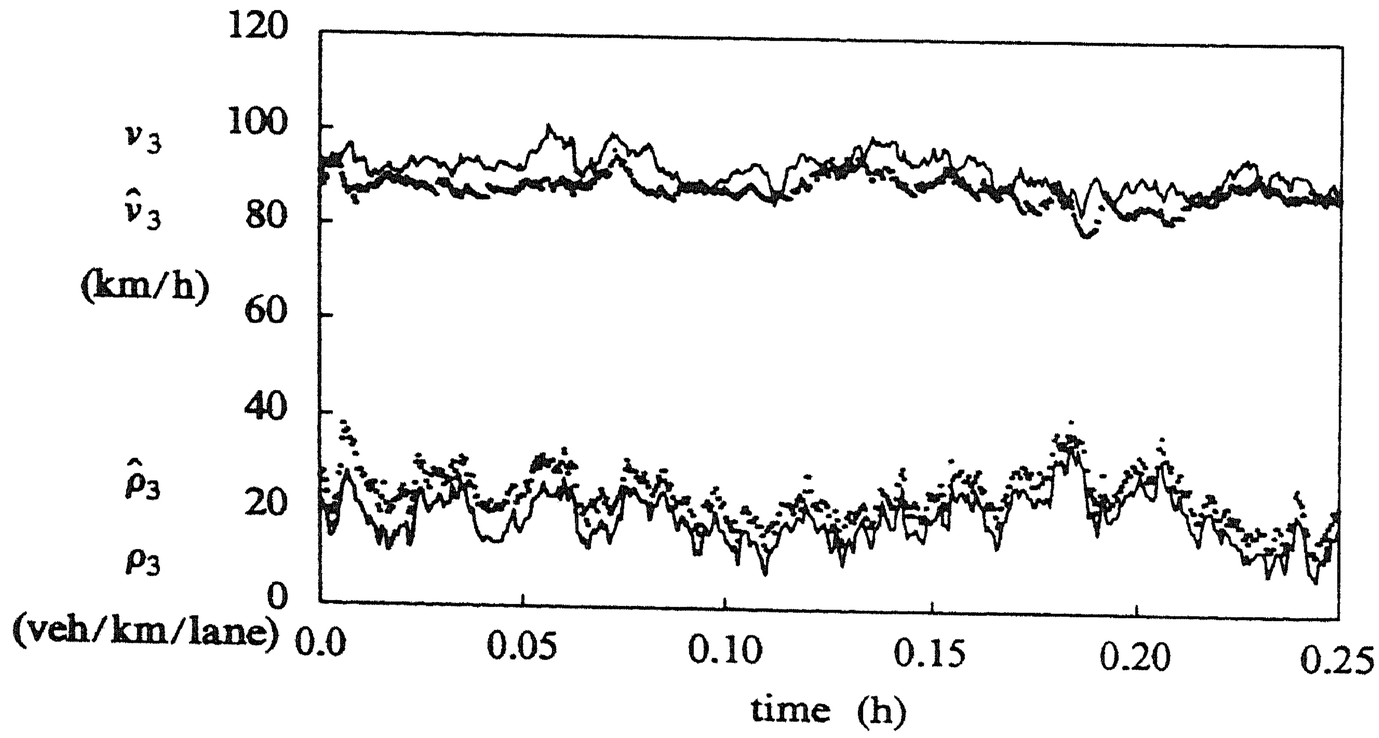

FIGURE 5.5. Simulated and estimated state trajectories of Experiment 5.4.

$D(.)$	
$\hat{\rho}_1$	3.9
$\hat{\rho}_2$	1.4
$\hat{\rho}_3$	5.0
$\hat{\rho}_4$	1.1
$\hat{v}_1$	7.0
$\hat{v}_2$	6.3
$\hat{v}_3$	5.6
$\hat{v}_4$	5.3

TABLE 5.8. Estimation errors in Experiment 5.4.

i	μ^i_{60}	$(\sigma^2)^i_{60}$	$\rho^i_{60,1}$
0	-0.05	0.78	-0.07
1	-0.02	0.42	-0.15
2	-0.06	0.78	-0.41
3	-0.05	0.72	-0.49
4	0.02	1.44	-0.15

TABLE 5.9. Innovation results of Experiment 5.4.

Conclusions:

—estimator bias is considerable in absence of speed information and almost zero in case speed information is used;
—the asymptotic estimation error is reduced considerably when a limited amount of speed information is used;
—using a minimal amount of speed information is necessary to achieve filter robustness with respect to modelling errors.

5.3.3. Sensitivity tests.

In this section results of experiments with the filter on small sets of real data are summarized. The aim of these experiments is to determine the sensitivity of the filter with respect to several parameters, such as T, α, β, γ, ρ_{crit}, ρ_{jam} and allow us to choose approximately optimal values for those parameters to which the filter is sensitive. Recall that these model parameters were chosen partly based upon simulations, partly based upon estimation from freeway data. It is of interest to know whether or not precise knowledge of the parameters is necessary for state estimation purposes. For a detailed report of the sensitivity experiments we refer to Smulders [87].

Sensitivity with respect to several parameters was investigated in the following way: first the distance D was computed for the estimated state trajectory generated with the nominal parameter value and the trajectory generated with the

perturbed parameter. If this distance is large for one of the state variables, the innovations are computed to decide whether the perturbation is beneficial. As the innovations only inform us about the quality of the intensity estimates, the D's of $\hat{\bar{v}}_i$ vs. $\bar{v}_i^{\delta}$ were computed as well.

The sensitivity of the filter heavily depends on the noise variances involved. The variance of the noise in the density equation (3.5) and in the measurement equation (2.20) directly follow from the modelling procedure, but for the variance Σ of the acceleration noise in (3.6) a decision still has to be made. A reasonable value is chosen later on. A small value of Σ leads the filter to rely heavily on the model and put little weight on the measurement information. This is illustrated by Theorem 5.13 of Section 5.2.2 which states that the error variance monotonically decreases with Σ. In contrast, a large Σ will lead the filter to rely on the measurements more than on the model. Consequently, the filter will be less sensitive to the model parameters. To obtain some idea of an upper bound for the sensitivity with respect to model parameters, it is therefore sensible to take a small value for Σ, preferably $\Sigma = 0$.

As was reported by Smulders difficulties occur in filtering with small acceleration noise, especially on nonstationary traffic data [87]. A large bias may be introduced, caused by this small noise and by the approximate nature of the filter. To overcome this difficulty the filter is temporarily simplified: the error covariance matrix P_t^x is set to zero. This leads to the following expression for the filter gain:

$$\Phi = \begin{bmatrix} A \\ O \end{bmatrix}$$

Then the filter equation (5.5) reduces to

$$d\hat{X}_t = \hat{F}(\hat{X}_t)dt + \begin{bmatrix} A(dN_t - \hat{H}(\hat{X}_t)dt) \\ O \end{bmatrix} \tag{5.21}$$

or, in more detail,

$$d\hat{\rho}_i(t) = \frac{1}{l_i L_i}\left\{[dN_{i-1} - dN_i] + (\epsilon_{i-1}^m - \epsilon_{i-1}^f)\hat{\lambda}_{i-1}dt - (\epsilon_i^m - \epsilon_i^f)\hat{\lambda}_i dt\right\} \tag{5.22}$$

$$d\hat{v}_i(t) = [\cdots]dt \tag{5.23}$$

This means that the densities are adjusted according to the measurements only when vehicles enter or leave a section, apart from a small counting error correction, and that the speed approximately follows the model equation. There are no ›nger estimate error corrections based on measurement information. This type approach was also discussed by Van Maarseveen [62]. It should be noted that is simplification is temporary and only carried out for the benefit of the sensi‹vity analysis. Later on the full filter equations are restored and a larger acceleration noise prevents the bias to occur.

It is clear from (5.21) and (5.22) that taking more than one speed class in applying this temporarily modified filter is useless, as observations per class are summed by multiplication with the matrix A in (5.21). So we take $m = 1$. We

start with the parameter values of Table 5.1.

Because of the absence of estimate error correction it is necessary to have a reasonable initial estimate of the traffic state. This estimate is generated by applying the filter with a large acceleration noise to a short period of traffic directly preceding the test period. If during the short period traffic is reasonably stationary reliable estimates are obtained, see Table 5.10.

The data to which the filter was applied consisted of two sets of 15 minutes from the northern carriageway of the A12 dual freeway near Utrecht (km 65.55 to 63.50). This stretch contains no on- or off-ramps and has two lanes. The data was obtained by the Traffic Engineering Division of the Dutch Ministry of Transport by means of the Dutch Motorway Control and Signalling System at May 9, 1983 from 7:10 to 7:25 am and 8:00 to 8:15 am respectively.

The first quarter contains more or less stationary traffic, corresponding to an intensity of 1250 veh/h/lane, whereas during the second quarter instabilities occur. See Figure 5.6 for a plot of 25 second averages of speeds at the successive measurement locations during the second quarter. The dashed lines correspond to a speed of 90 km/h at the specific location and are drawn at distances of 25 km/h. From Figure 5.6 one notes that there are speed drops to 60 km/h, but there is always recovery after some time.

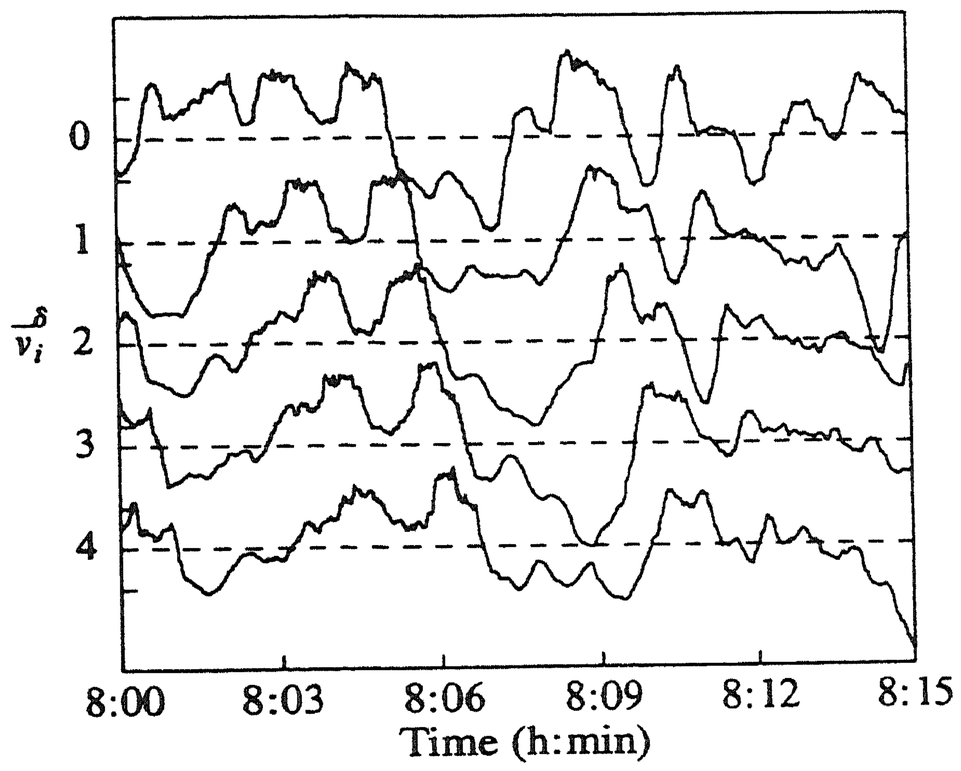

FIGURE 5.6. Averaged speed, measured at consecutive locations.

The intensity during this quarter is 1800 veh/h/lane on the average, about 20 percent below the capacity, but reaches the capacity during short periods. The capacity of this stretch was estimated by Van Toorenburg [94] and found to be approximately 2215 veh/h/lane. The filter is started in the state given in Table 5.10.

i	$\hat{\rho}_i$ (veh/km/lane)	$\hat{v}_i$ (km/h)
1	30 ± 1	86 ± 3
2	13 ± 1	94 ± 4
3	19 ± 1	94 ± 4
4	12 ± 1	93 ± 4

TABLE 5.10. Initial state estimates.

During both quarters weather was dry and clear and no special incidents occurred.

By analysing the counts cumulated over two hours of data we concluded that the relative counting errors among the five measuring sites are zero: all detectors counted about the same vehicle total. One of the sites was calibrated in an earlier study [93] and shown to miss about 1.5 percent of the vehicles that pass. We use this value for all locations.

Detailed results of the experiments carried out with these data sets are given by Smulders [87]. We only present the main results here.

The filter proved to be remarkably insensitive to variation of most of the parameters. This may be explained by the fact that the traffic state is almost completely determined by the density, the mean speed is close to the equilibrium value, and that this density is fairly accurately measured. After the initial uncertainty is reduced, only small counting errors have to be compensated by the filter. The power of the filter apparently is not fully used in the current setup. Omitting measurement information at intermediate locations or using the filter for predictions over short time intervals would put it to a harder test. The first was considered by Cremer [20].

The filter is sensitive to α and γ, however. The following two experiments show this.

EXPERIMENT 5.5 (SENSITIVITY W.R.T. α). Concerning α we note the following. The simulation studies of Chapter 3 led to a value of 0.85 for α. Taking $\alpha = 0.5$ was shown to lead to unsatisfactorily large fluctuations in the section densities ρ_i. In the filter it seems that $\alpha = 0.5$ is to be preferred, as the following argument shows. Consider a bunch of vehicles passing a measuring location i. If α is close to 1.0 the emphasis in $\hat{\lambda}_i$ is on the upstream section. The density $\hat{\rho}_i$ has jumps at moments that vehicles enter section i (that is: pass location $i-1$!) and so $\hat{\lambda}_i$ immediately increases if $\alpha \approx 1.0$. This is somewhat unrealistic because the intensity measured at location i, $\bar{\lambda}_i^{\delta}$, only increases after a certain delay, corresponding to the time it takes a vehicle to pass through section i. In short: $\hat{\lambda}_i$ reacts too early in comparison to $\bar{\lambda}_i^{\delta}$ when α is close to 1.0. A complementary effect occurs when vehicles pass location i. The simplest remedy to this problem is to set α equal to 0.5. This smooths the behaviour of $\hat{\lambda}_i$ and also retards the increase when vehicles pass location $i-1$ and retards the decrease when they pass location i.

The delay in case $\alpha = 0.85$ was confirmed by experiments on real data. See Figure 5.7 for a comparison of $\hat{\lambda}_3$ (dotted) in case $\alpha = 0.85$ with $\hat{\bar{\lambda}}_3^{\delta}$, and Table

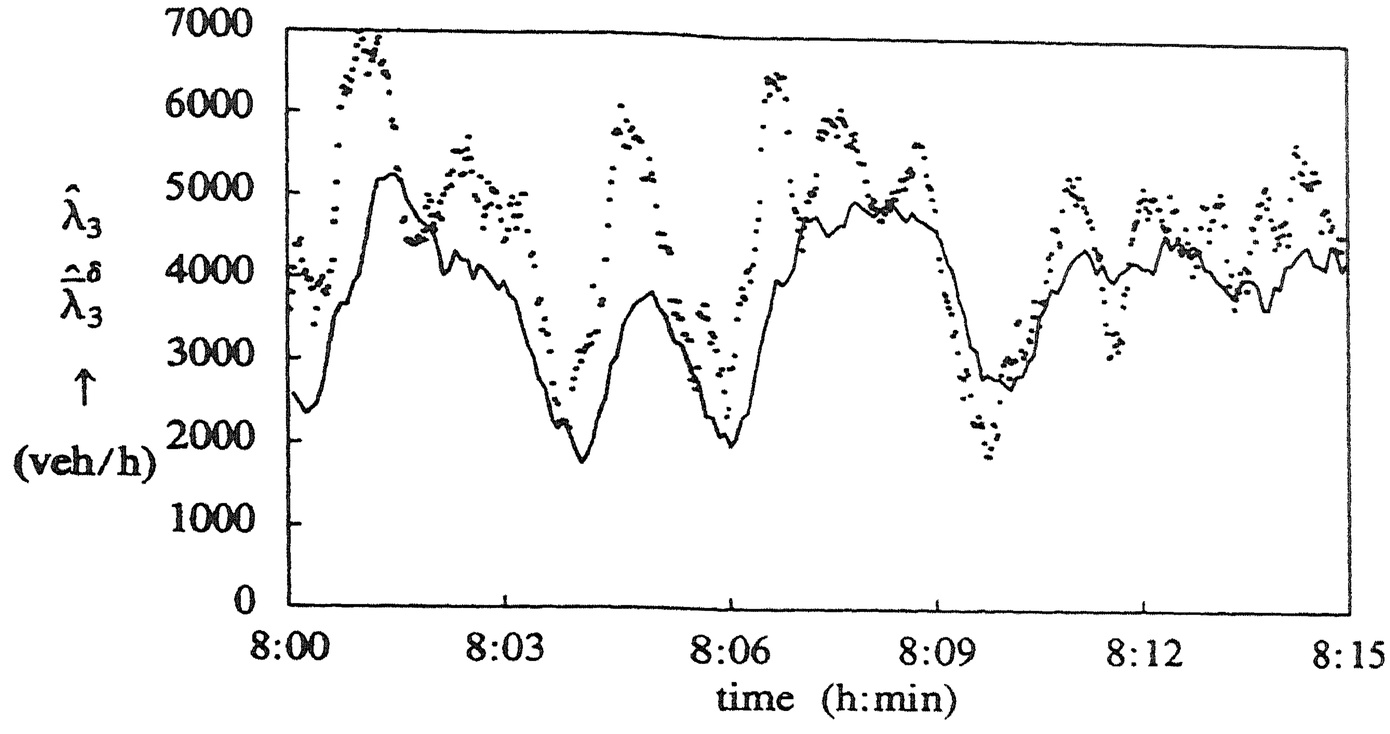

FIGURE 5.7. Estimated and measured intensity of Experiment 5.5.

5.11 for a comparison of the filtered intensity in case $\alpha=0.85$ and 0.5, respectively.

$D(.)$	$\hat{\bar{v}}_0$	$\hat{\bar{v}}_1$	$\hat{\bar{v}}_2$	$\hat{\bar{v}}_3$	$\hat{\bar{v}}_4$	$\hat{\lambda}_0$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$	$\hat{\lambda}_4$
	(km/h)					(veh/h)				
$\alpha = 0.85$ km/h^2	18.6	13.6	8.7	5.8	6.2	1162	1100	941	954	781
$\alpha = 0.5$ km/h^2	18.4	13.2	8.9	5.8	6.2	1163	711	615	556	783

TABLE 5.11. Estimation errors in Experiment 5.5.

The simplified filter used here turned out to be very sensitive to α and $\alpha=0.5$ turned out to be the best choice. Experiments were carried out with $\alpha=0.5$, 0.65 and 0.85. For this reason $\alpha=0.5$ in all experiments to follow. The full filter of 5.3.4 (with the gain Φ no longer set to zero) turns out to be insensitive with respect to α. The reason for this is that in this case compensation of the errors is possible. The filter will adjust $\hat{\rho}$ and $\hat{v}$ to achieve a good innovation process. Errors are therefore not so much to be expected in the intensity estimate as in the density and speed estimates. Experiments have shown that these errors are relatively small. As was noted before the sensitivity computed for the simplified filter gives only an upper bound.

EXPERIMENT 5.6 (SENSITIVITY W.R.T. γ). A slight improvement of the filter performance was achieved by decreasing the anticipation strength γ from 6.5 km/h^2 to 1.0 km/h^2. This results in a faster reaction of the speed to density variations.

The following simple first order analysis confirms this.

Consider the equation for $v_i(t)$ where we assume that the density in the neighbouring sections is constant and equals $\rho < \rho_{crit}$ and the speed in those sections equals $v^e(\rho)$:

$$\frac{dv_i(t)}{dt} = -\frac{1}{T}\left[v_i(t) - v^e(\rho_i(t))\right] - \gamma(L_i l_i)^2\left[\beta\rho_i + (1-\beta)\rho\right]\left[\rho - \rho_i(t)\right] + \frac{l_{i-1}}{L_i l_i} v(v - v_i(t)) \tag{5.24}$$

Linearization about the stable equilibrium point $(\rho, v^e(\rho))$ and assuming that $L_i l_i = 1$ and $l_{i-1} = 2$ leads to

$$\frac{d}{dt}(\delta v_i) = (-\frac{1}{T} - 2v)\delta v_i + (\gamma\rho - \frac{a}{T})\delta\rho_i \tag{5.25}$$

where $\delta v_i = v_i - v^e(\rho)$ and $\delta\rho_i = \rho_i - \rho$ and we have used the expression (4.6) for v^e. Now (5.25) is a linear system with input process $\delta\rho_i(t)$. The response to $\delta\rho_i(t) = \hat{\delta\rho}\sin(\omega t + \phi)$ will be

$$\delta v_i = |H(i\omega)|\ \hat{\delta\rho}\sin(\omega + \phi + \psi)$$

where $\psi = arg[H(i\omega)]$ and H is the transfer function from $\delta\rho_i$ to δv_i. Now

$$H(i\omega) = \frac{\gamma\rho - \frac{a}{T}}{i\omega + \frac{1}{T} + 2v}$$

so

$$\psi = arg[H(i\omega)] = arctan\left[\frac{\omega}{-\frac{1}{T} - 2v}\right] + arg(\gamma\rho - \frac{a}{T})$$

We expect the model (5.24) to be such that an increase of ρ_i leads to a decrease of v_i, possibly after a short delay. This means that ψ should be close to π. In the experiments it turned out that for the given parameters the delay was too large. A necessary condition in our model (5.24) for ψ to be able to approach π is that

$$\gamma\rho < \frac{a}{T}$$

For the parameters in the tests this condition is not satisfied. Decreasing γ or T should lead to better results. According to the condition just mentioned, γ should be less than $58/\rho \approx 2$ km/h^2 if $T = 0.01$ h. And T should be less than $0.09/\rho \approx 0.004$ h if $\gamma = 6.5$ km/h^2.

Decreasing γ was shown to be preferred over decreasing T [87]. The effect of a smaller γ is illustrated in Figure 5.8, where the filtered speed in section 3 of the freeway stretch is plotted (period 8:00-8:15 am), for both $\gamma = 6.5$ (dotted) and $\gamma = 1.0$, together with $\bar{v}_3^{\delta}$. See also Table 5.12.

$D(.)$	$\hat{\bar{\nu}}_0$	$\hat{\bar{\nu}}_1$	$\hat{\bar{\nu}}_2$	$\hat{\bar{\nu}}_3$	$\hat{\bar{\nu}}_4$	$\hat{\lambda}_0$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$	$\hat{\lambda}_4$
	(km/h)					(veh/h)				
$\gamma = 6.5$ km/h^2	19.8	14.7	10.0	7.1	6.6	1174	709	598	537	774
$\gamma = 1.0$ km/h^2	12.1	8.0	5.5	6.1	6.5	807	426	480	444	690

TABLE 5.12. The sensitivity of the filter w.r.t. γ.

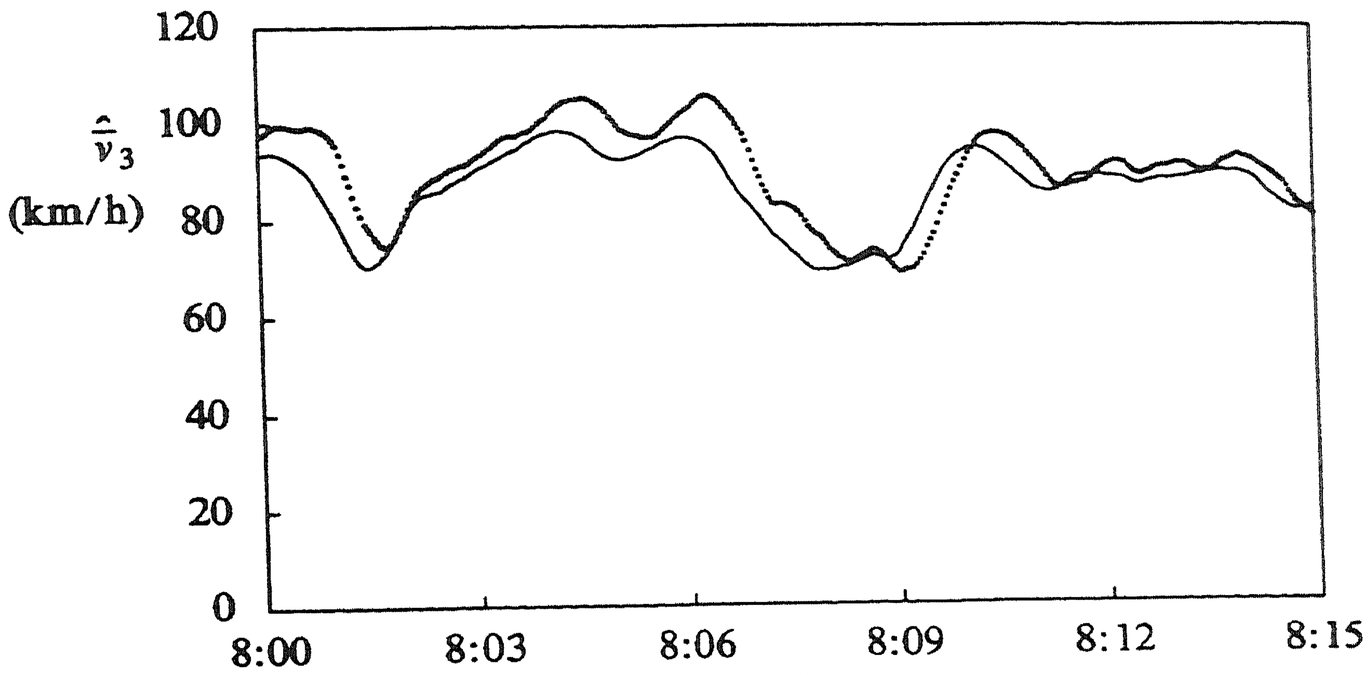

FIGURE 5.8. Measured and estimated speed of Experiment 5.6.

In all tests it turned out that large systematic errors (over 10%) may occur in the estimated intensity over long periods of time. See Figure 5.9 for an illustration of these errors. The continuous line represents $\bar{\lambda}_3^{\delta}$ and the dotted one the filter estimate. From the results of the previous paragraphs it follows that parameter modifications will not remove the bias. Surprisingly enough, even applying the full original filter (5.5) to (5.7), with a large value of Σ to emphasize the measurement information, did not lead to a better performance. This fact indicates that the error is in the measurement equations. Detailed investigation of the data finally led to the conclusion that the bias is caused by a modelling error in the passing speed distribution [87]. This distribution was estimated to depend on the traffic density via its variance, see (4.5) and on the mean speed via its mean. The former relation was estimated over a large number of passages ($\pm$10000), during periods of stationary traffic. In practice large deviations from this relation occur during short periods. It turns out that these deviations are serious enough to prevent the filter from estimating the intensity correctly. Recall that large errors occurred in Experiment 5.4 of

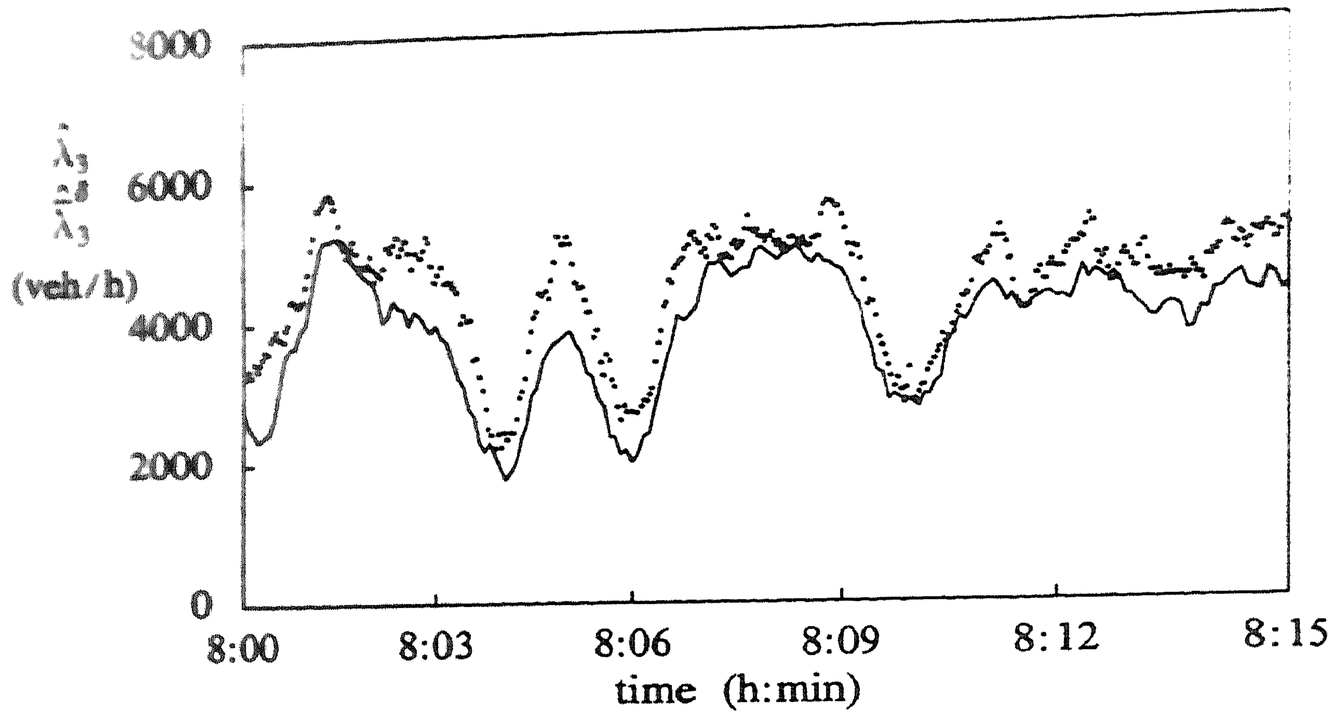

FIGURE 5.9. Measured and estimated intensity of Experiment 5.6.

Section 5.3.2 due to a passing speed distribution modelling error.

The solution to this problem that we propose is to modify the filter in such a way that the density estimates only react to an error in the *total* intensity at a location (summed over all speed classes), instead of reacting to errors in intensities per class. This reduces the sensitivity to passing speed probability modelling errors and amounts to summing the entries in the upper block of Φ_t, corresponding to the density estimates, over all classes. Define

$$\tilde{\Phi}_{i,k\mathrm{m}+j} = \begin{cases} \left(\sum_{j=1}^{\mathrm{m}} \hat{\lambda}_k^j \Phi_{i,k\mathrm{m}+j}\right) / \hat{\lambda}_k & \text{for } 1 \leqslant i \leqslant \mathrm{n} \\ \Phi_{i,k\mathrm{m}+j} & \text{for } \mathrm{n}+1 \leqslant i \leqslant 2\mathrm{n} \end{cases} \tag{5.26}$$

for $k = 0,\ldots,\mathrm{n}$. Then our new filter equations are (5.5) to (5.7), where Φ is replaced by $\tilde{\Phi}$. In practice we compute $\tilde{\Phi}$ as before, but before integrating and updating $\hat{X}_t$ we carry out the summations of (5.26).

The modified filter is applied to the period of unstable traffic and compared with the original filter with Φ no longer set to $\begin{bmatrix} A \\ O \end{bmatrix}$, but restored to its original value. See Table 5.13 for a comparison of the respective results.

	Φ	$\tilde{\Phi}$
i	μ^i_{60}	μ^i_{60}
0	0.03	-0.03
1	0.03	0.01
2	-0.06	0.02
3	-0.07	0.01
4	-0.02	-0.00

TABLE 5.13. Comparison of original and modified filter results.

The modified filter succeeds in reducing the bias considerably. In Section 5.3.4, Experiment 5.7 another comparison of the original filter with the modified one is presented. Figure 5.13 shows that for the modified filter the correlation of the innovation increments is significantly smaller.

Finally, the modified filter is applied to the simulated data of Experiment 5.4. This experiment is concerned with a modelling error in the passage speed distribution. The results of Table 5.14 and 5.15 again show the ability of the modified filter to reduce the bias.

$D(.)$	
$\hat{\rho}_1$	3.8
$\hat{\rho}_2$	1.3
$\hat{\rho}_3$	2.8
$\hat{\rho}_4$	1.0
$\hat{\nu}_1$	6.1
$\hat{\nu}_2$	5.8
$\hat{\nu}_3$	4.8
$\hat{\nu}_4$	4.6

TABLE 5.14. Estimation errors of the modified filter.

i	μ^i_{60}	$(\sigma^2)^i_{60}$	$\rho^i_{60,1}$
0	-0.07	0.78	-0.14
1	-0.03	0.42	-0.23
2	-0.01	0.66	-0.41
3	-0.01	0.66	-0.48
4	0.01	1.44	-0.17

TABLE 5.15. Innovation results of the modified filter.

To conclude this section the modified filter is applied with $\Sigma = 10000$, which leads to a variance of the mean speed measured in practice. A comparison is made between the first and second order approximation to the optimal filter. Table 5.16 shows that the first order filter produces almost the same estimates as the second order one on the set of nonstationary data. The entries of the table

are the distances D between the respective estimated trajectories. Recalling the relative computational effort for the first order filter, we conclude that the first order filter is to be preferred.

$D(.)$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$ (veh/km/lane)	$\hat{v}_1$	$\hat{v}_2$	$\hat{v}_3$	$\hat{v}_4$ (km/h)
1st order vs. 2nd order	0.05	0.01	0.01	0.01	0.4	0.1	0.1	0.1

TABLE 5.16. Estimation errors of first versus second order filter.

Next a comparison is made of results obtained with 4 resp. 2 speed classes. The result in Table 5.17 shows that two classes suffice to produce accurate estimates.

$D(.)$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$ (veh/km/lane)	$\hat{v}_1$	$\hat{v}_2$	$\hat{v}_3$	$\hat{v}_4$ (km/h)
m = 4 vs. m = 2	0.3	0.4	0.2	0.2	2.7	1.9	2.3	2.2

TABLE 5.17. Estimation errors: two versus four speed classes.

Even using no speed information, m=1, still gives reasonable results, but worse than with m=2: Table 5.18. The possibility of estimating the traffic state from count data alone was already mentioned in 5.2.2, but as was shown in 5.3.2 it is better to include speed information in the estimation procedure for the sake of robustness.

$D(.)$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$ (veh/km/lane)	$\hat{v}_1$	$\hat{v}_2$	$\hat{v}_3$	$\hat{v}_4$ (km/h)
m = 4 vs. m = 1	1.6	0.9	0.9	0.6	9.0	5.7	3.9	4.1

TABLE 5.18. Estimation errors: one versus four speed classes.

Conclusions:

—to reduce the sensitivity to modelling errors in the passing speed distribution a structural modification of the filter was proposed and shown to reduce the bias;
—the anticipation strength factor γ was reduced from 6.5 to 1.0 to obtain a slightly better result;
—the weighting factor α was chosen to be equal to 0.5 to remove a small delay in the reaction of the intensity;
—the filter turned out to be insensitive to most other parameters. This is explained by the fact that density largely explains traffic behaviour. Accurate parameter estimates are therefore not required to achieve good estimation results;
—the first order filter proved to be as accurate as the second order one;
—two speed classes were shown to be sufficient to obtain accurate estimates.

5.3.4. Validation.

In this section the filter that resulted from the tests described in the preceding section is applied to larger data sets of different dates to check the validity of the results. We use the same freeway stretch as before (A12 Utrecht, northern carriageway, km 65.55 to 63.50) and use data from May 9 and May 10, 1983. On both days the weather was dry and clear. The experiment on data from May 9 mainly summarizes results obtained in Section 5.3.3. The experiment on data from May 10 serves as a validation. The filter should work well also on data on which it was not calibrated.

EXPERIMENT 5.7. The period considered is 7:15 − 8:15 am. The first order filter is applied with $\tilde{\Phi}$, using 3 speed classes: [0,80) , [80,100) and [100,∞). The parameter values of Table 5.1 are used except that $\alpha = 0.5$, $\gamma = 1.0$ km/h^2 and $\Sigma = \text{diag}(10000)$. The filter is started with a large initial uncertainty. The results are presented in Table 5.19, 5.20 and 5.21 and Figures 5.10, 5.11, 5.12 and 5.13. Note the reduced correlation in the innovation increments and the small mean errors in the intensity estimates ($\pm 2\%$). The intensity of the entrance turns out to be more difficult to estimate than the others. Figure 5.13 shows the correlation functions of this experiment (points marked "*") and those obtained by running the original filter (points marked "+"), with Φ instead of $\tilde{\Phi}$, on the same data set. There is a clear difference in favour of the modified filter.

Running the filter on this data set required almost 6 minutes of CPU-time on a CDC Cyber 750 mainframe.

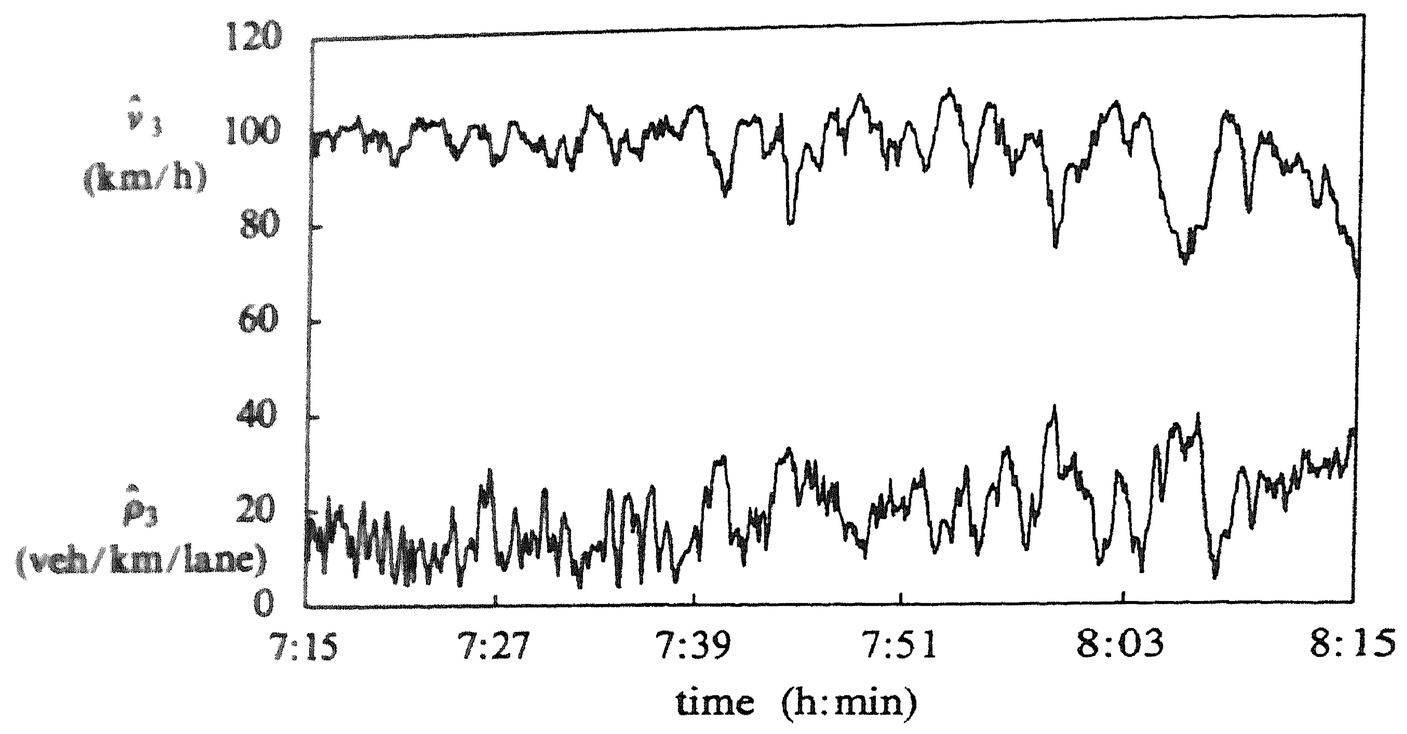

FIGURE 5.10. Estimated state trajectories of Experiment 5.7.

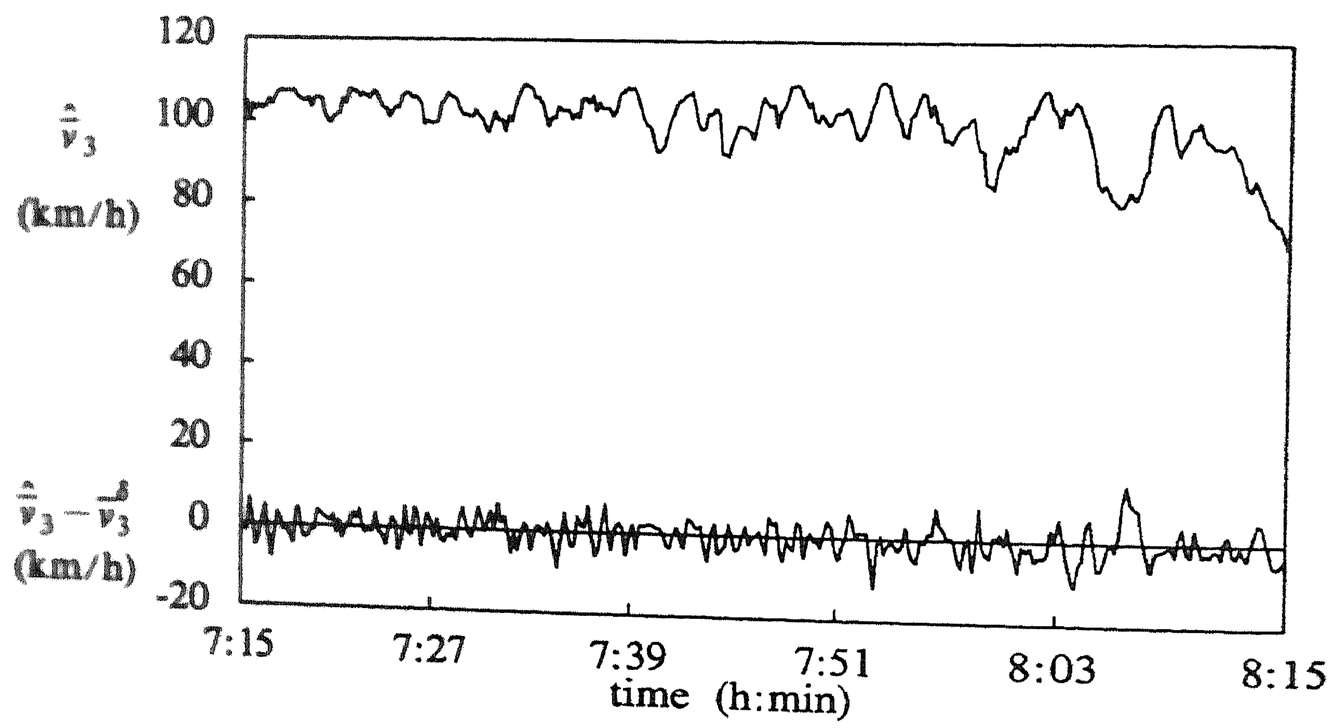

FIGURE 5.11. Estimated speed and error process of Experiment 5.7.

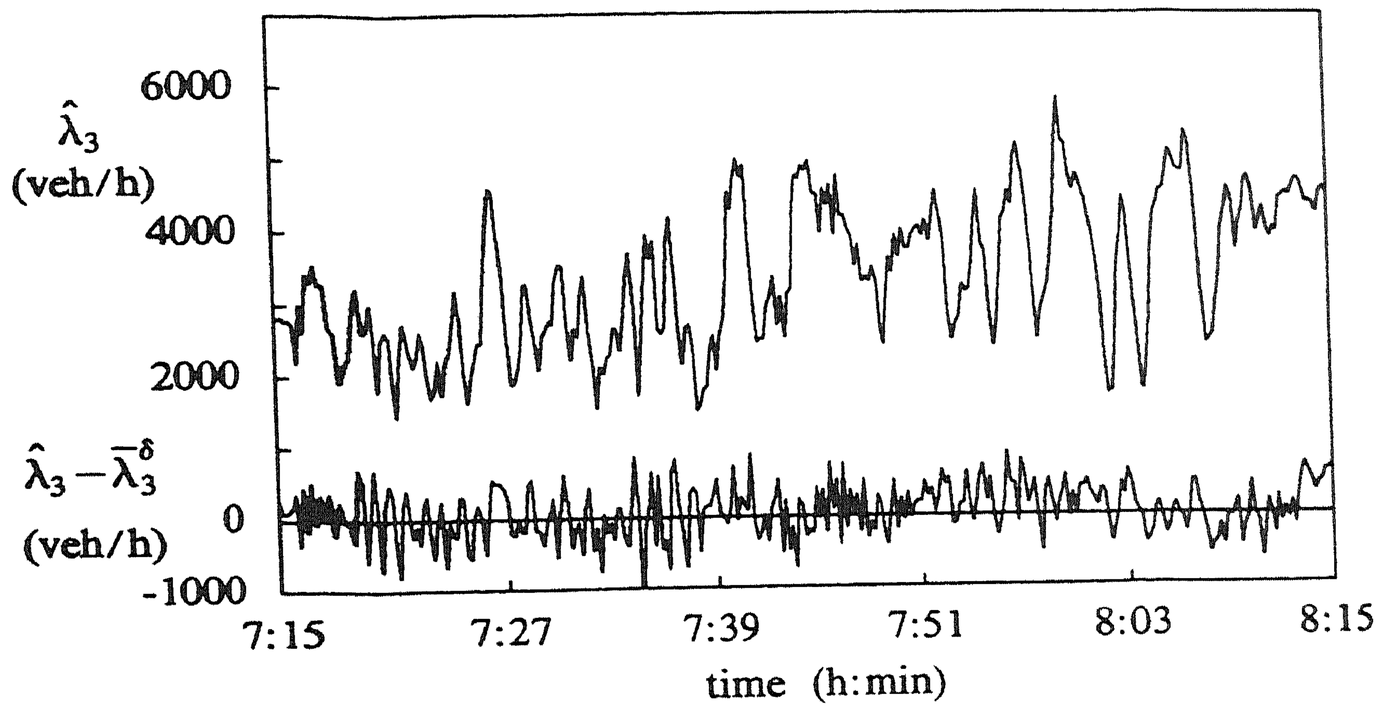

FIGURE 5.12. Estimated intensity and error process of Experiment 5.7.

$D(.)$	$\hat{\bar{v}}_0$	$\hat{\bar{v}}_1$	$\hat{\bar{v}}_2$	$\hat{\bar{v}}_3$	$\hat{\bar{v}}_4$	$\hat{\lambda}_0$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$	$\hat{\lambda}_4$
	(km/h)					(veh/h)				
	5.5	5.1	3.6	3.6	5.1	742	388	399	358	688

TABLE 5.19. Estimation errors in Experiment 5.7.

i	μ^i_{60}	$(\sigma^2)^i_{60}$	$\rho^i_{60,1}$
0	-0.07	1.08	-0.01
1	-0.02	0.32	0.07
2	0.02	0.22	0.04
3	-0.02	0.19	0.13
4	-0.02	0.60	-0.25

TABLE 5.20. Innovation results of Experiment 5.7.

t = 8:15 h	$i=1$	$i=2$	$i=3$	$i=4$
$\tilde{\text{var}}[\hat{\rho}_i - \rho_i]$	1.4	3.6	3.0	1.7
$\tilde{\text{var}}[\hat{v}_i - v_i]$	13.3	15.1	17.6	16.9

TABLE 5.21. Final error variances of Experiment 5.7.

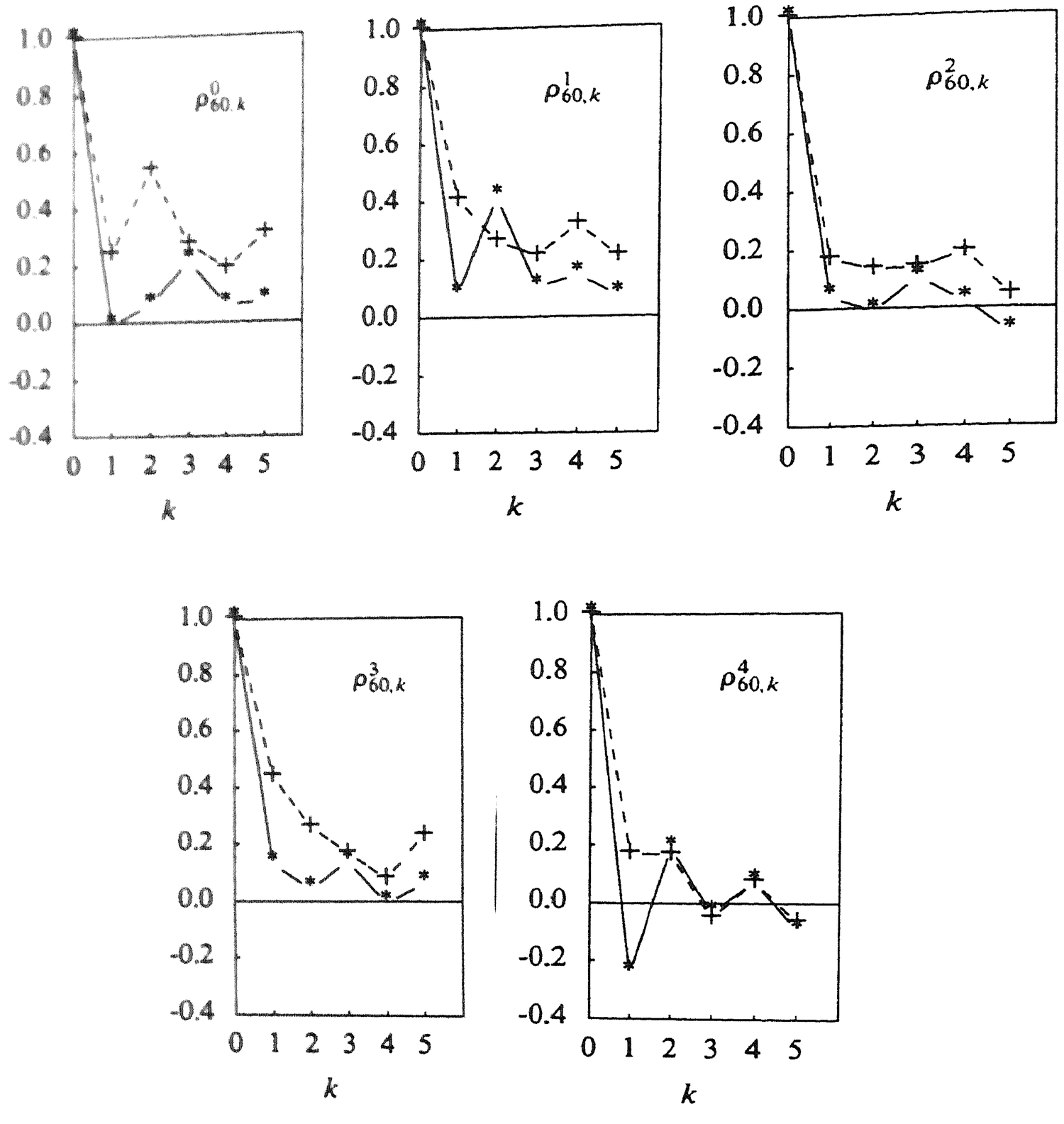

FIGURE 5.13. Innovation correlation of original and modified filters.

EXPERIMENT 5.8. The filter is now applied to data of May 10, 1983 and the period considered is 7:45 – 8:45 am. This time two speed classes are used: [0,80) and [80,∞). The results are in Table 5.22 and 5.23 and Figure 5.14. Despite the fact that traffic is more dense than in the previous experiment (more instabilities) and only two speed classes are used the results are still satisfactory.

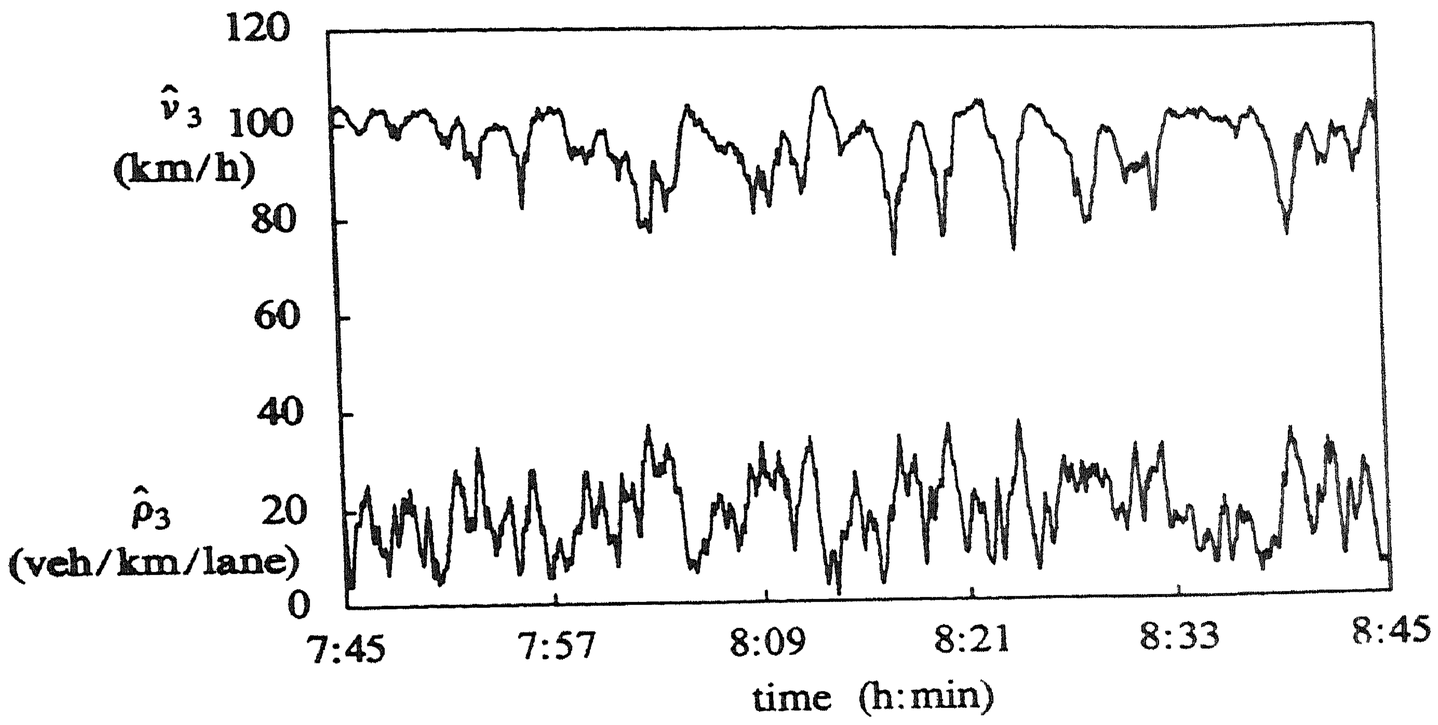

FIGURE 5.14. Estimated state trajectories of Experiment 5.8.

$D(.)$	$\hat{\bar{v}}_0$	$\hat{\bar{v}}_1$	$\hat{\bar{v}}_2$	$\hat{\bar{v}}_3$	$\hat{\bar{v}}_4$	$\hat{\lambda}_0$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$	$\hat{\lambda}_4$
	(km/h)					(veh/h)				
	9.2	5.8	5.1	5.3	4.8	888	427	439	372	784

TABLE 5.22. Estimation errors in Experiment 5.8.

i	μ^i_{60}	$(\sigma^2)^i_{60}$	$\rho^i_{60,1}$
0	-0.06	1.55	-0.32
1	-0.04	0.37	0.01
2	0.01	0.25	0.08
3	0.01	0.20	0.20
4	-0.02	0.71	-0.04

TABLE 5.23. Innovation results of Experiment 5.8.

The previous experiments show the filter to produce satisfactory estimates of traffic speed and density. These estimates may be used in control algorithms, where the control policy is based upon knowledge of the state of traffic. Such a control problem is discussed in the next chapter.

The filtered density and speed values may be used for the benefit of traffic research as well. It is possible to investigate the equilibrium relation between density and speed for instance, and investigate location dependence or the

influence of weather conditions. An example is given in Figure 5.15 where the filter estimates of one hour of traffic data are plotted, of one location along the freeway.

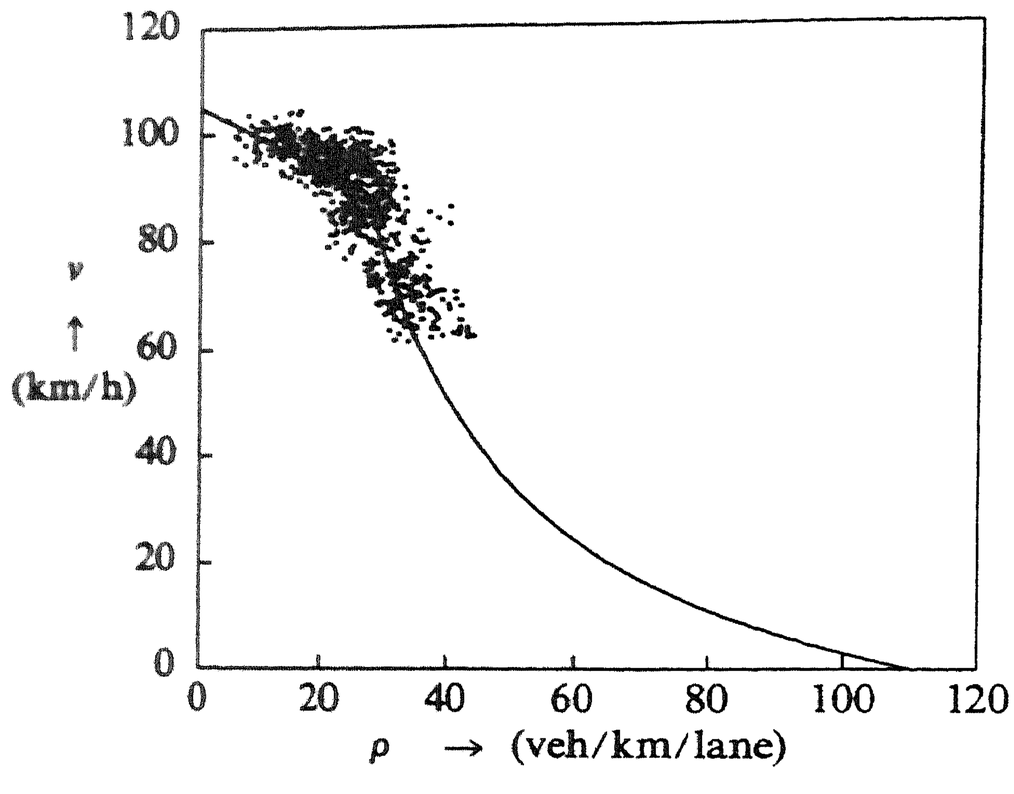

FIGURE 5.15. Phase plot of filter estimates.

Each point corresponds to a $(\hat{\rho},\hat{v})-$pair, there is a point every 2 to 5 seconds. The equilibrium relation as it is modelled is plotted also. The model seems to agree with the real data very well. Note that although the filter uses this modelled relation also, it is quite able to correct errors, as was shown in Experiment 5.3.

Another example is in Figure 5.16 where a large excursion occurs, in which density increases to 60 veh/km/lane and the speed decreases to 40 km/h. The cycle that is apparent in the plot is travelled clockwise. This in accordance with results obtained by e.g. [14].

To conclude this section we mention the result of an experiment in which measurement information from one of the locations are omitted. This is done to test the filter's performance in a more complicated situation. Data is again taken from the four section stretch of freeway near Utrecht. The measurements from the middle location, $i=2$, were left out. The filter turns out to be very sensitive to α. Taking $\alpha=0.5$ as before leads to a large bias in $\hat{\rho}_2$ and $\hat{\rho}_3$. Taking $\alpha=0.85$ gives an accuracy comparable to the one achieved with the filter including the information of location 2, however. See Table 5.24 where the distance between the estimated state trajectories of the filter with and without the measurements of location 2 are presented.

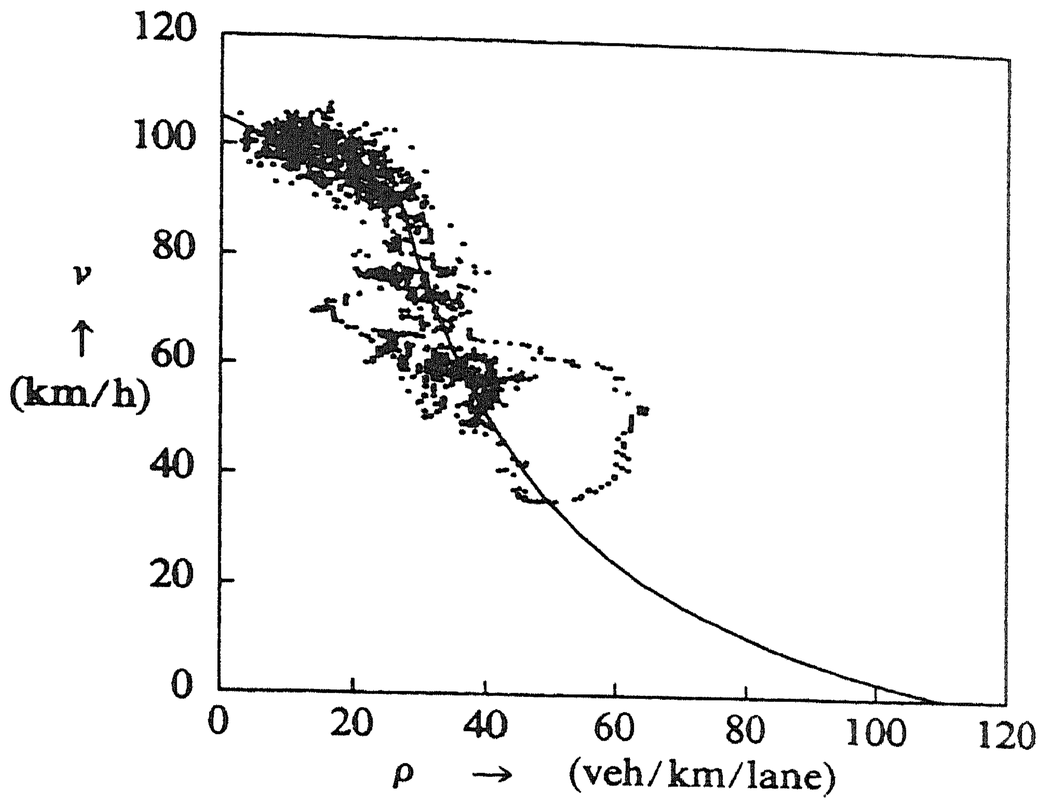

FIGURE 5.16. Phase plot of filter estimates.

$D(.)$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$	$\hat{v}_1$	$\hat{v}_2$	$\hat{v}_3$	$\hat{v}_4$
	(veh/km/lane)				(km/h)			
	1.0	3.3	3.0	1.4	1.4	2.2	2.5	1.1

TABLE 5.24. Estimation errors after omitting measurements of one location.

5.4. CONCLUSIONS

In this chapter an approximate filter has been derived for the recursive estimation of the state of traffic on a freeway. This filter is based upon the model of freeway traffic flow developed in Chapter 2 and uses measurement information from several locations along the freeway. It produces approximations to the optimal conditional expectation estimators.

Detectability and stabilizability of suitable matrix pairs in the filter has been investigated and led to conclusions about the asymptotic behaviour of the error covariance matrix. The effect of using speed information in the filter has been investigated by computing the asymptotic error covariance matrix for several different cases. It is concluded that using a limited amount of speed information is necessary to reduce bias and asymptotic error variance and suffices to produce satisfactory estimates.

The proposed filter was tested against simulated and against real traffic data. Performance was evaluated using three criteria: the integrated squared error criterion, the local estimate criterion and the innovation increments criterion. The

test against simulated data confirmed the results mentioned in the previous paragraph and also showed that using speed information is necessary to obtain robustness with respect to modelling errors. The tests against real data led to a structural modification of the filter to remove bias due to modelling errors in the passing speed probability distribution, and to some parameter changes, resulting in reduced bias and reduced error variance. The filter turned out to be insensitive to most parameters. It is concluded that the evolution of traffic density largely explains traffic behaviour and the speed variable is highly dependent on it. As the density is fairly accurately measured, only initial uncertainty and small counting errors have to be compensated by the filter. Putting the filter to a harder test by omitting information at an intermediate location shows an increase of sensitivity. In case of application of the filter to real data, again, as with simulated data a limited amount of speed information proved to be sufficient to produce satisfactory estimates. A first order approximation to the optimal filter was shown to be sufficiently accurate. A validation of the results was carried out by applying the modified filter to sets of one hour of real data.

We now suggest several subjects for future research.

In order to make the filter more practicable, on- and off-ramps have to be taken into consideration. This may easily be done in case the flows on the ramps are measured.

To reduce the computational effort that is required, the filter implementation should be simplified. A reduction of approximately 80% in computations would result from implementing a constant or parametrized filter gain matrix. This may be deduced from the number of operations needed to compute P_t^x, see Section 5.1. A promising approach seems to be to parametrize with respect to the intensity. The effect of such a modification of the filter on the accuracy of the estimates has to be investigated. Another considerable reduction of computational effort would result from a simplified implementation of the passage speed distribution.

The behaviour of the filter under congestion conditions has to be studied. In case of unstable traffic the observation model for the intensity is no longer correct. This modelling error will result in a bias in the estimates. Some correction mechanism has to be implemented to achieve reasonable estimates. A bias correction procedure was proposed by Willsky [52].

The filter may be used for several purposes, e.g. in control or detection algorithms or for research purposes. A control problem is considered in the next chapter. The capabilities of the filter in detecting congestion or its performance in combination with an incident detection algorithm are still to be investigated.

Chapter 6

Control of Freeway Traffic Flow

In the introductory chapter the freeway traffic control problem that is the main motivation of our research is described. This control problem will be investigated now. A special type of control, *homogenizing control*, is considered. This measure has been shown to result in an improved flow and greater safety during experiments in practice. Here we develop a control policy for the use of this homogenization measure, based upon knowledge of the state of traffic. An estimate of this state of traffic may be obtained by means of the filter of Chapter 5.

Control by means of variable *speed signs* is not considered very often in the literature. Some experiments with control by means of variable speed limits are reported by Zackor [103] and used by Cremer [20] to show (through simulation studies) the possible advantages of this type of control. One of the main difficulties associated with control by means of speed signals is to assess the effect on driver behaviour. In contrast to what apparently was found by Zackor [103], in the Netherlands speed limits are not necessarily followed by drivers. Moreover, the speed signals of the Dutch Motorway Control and Signalling System are not mandatory but advisory [80]. In the first section of this chapter we discuss the effect of these advisory speeds on driver behaviour, by investigating data from experiments in practice [93]. The effects of speed signal control turn out to be far less drastic than proposed by Cremer [20], where an increase of capacity of 21% is assumed to be achievable. According to our data capacity hardly increases, but control nevertheless has a significant positive effect on the stability of the traffic stream.

In Section 6.1 the effect of control on driver behaviour and traffic stability is investigated. In Section 6.2 two models for traffic in one section of the freeway are presented. The effect of control is incorporated in the models and their stability properties are analysed. Section 6.3 is devoted to control policy design. A hysteresis type of control policy is proposed, based on traffic density alone.

Preliminary results on control based on traffic density and speed are given. In Section 6.4 conclusions are drawn.

6.1 Homogenizing control

There are practical limitations associated with controlling traffic on a freeway, most of them concerned with safety. To mention one example, in case of control by means of speed signals one has to avoid displaying recommendations that vary too much in place or time, as this would confuse drivers and might lead to dangerous situations. These constraints may be difficult to handle mathematically and it may indeed be impossible to quantify them. Care therefore has to be taken in implementing control schemes in practice.

Apart from these constraints there is another problem in the uncertainty involved with freeway traffic. Predicting human behaviour such as the behaviour of drivers on a freeway is a difficult task. Therefore, the mathematical model as developed in Chapters 2 and 3 though showing reasonable behaviour in various traffic situations, is still afflicted with a fair amount of uncertainty. A control strategy based upon such an imperfect model has to possess a large degree of robustness to be feasible.

It is for these reasons that we only consider a specific type of control strategy, which is simple so that it is rather easy to meet safety constraints and achieve robustness. Ease of use and robustness are not the only reasons to restrict attention to the specific type of control, however. There are good reasons to expect a positive effect from the type of control considered here. These reasons are discussed in the following.

The type of control we consider is called *homogenizing control* and is described by Van Toorenburg [93]. With this type of control an identical advisory speed signal is displayed for all lanes at a given set of consecutive stations at the same time. The speed value is chosen out of a finite set and is in correspondence with the actual speed of the traffic stream. In practice this leads in almost all cases to a value of 90 km/h, as this is about the mean speed for intensities near capacity. In some cases 80 or 70 km/h may have to be displayed. The only parameters left to optimize in this type of control are the switch on and switch off times.

The motivation for this type of control stems from the fact that congestion is caused by the severe inhomogeneities of the traffic stream that exist when the intensity approaches the capacity [91]. Shock waves occur, originating in a chain of vehicles closely following each other at high speed. Small disturbances are amplified and may finally lead to a standstill, or worse: an accident, and congestion. This has been investigated theoretically by Köhler [48] e.g. It has been shown that when the headways in a chain of vehicles are below a certain bound, the chain is unstable. The inhomogeneities in the traffic stream readily lead to the small disturbances needed for congestion to set in. Examples of these inhomogeneities are speed differences between consecutive vehicles in one lane, speed differences between the lanes, differences in traffic intensity between the lanes. An illustration of the large differences that exist between lanes may be found in Figure 6.2. The difference between the probability density of time headways of left and right lane is apparent.

The previous discussion suggests two ways to avoid congestion: remove the sources of disturbances (increase homogeneity) or reduce the number of short headways (increase stability). It is also clear that the achievable effect is limited: as intensity approaches capacity drivers are "forced" to drive close, they are competing for the available space and gaps are immediately filled up [92]. Furthermore, not all disturbances can be avoided. It is the primary aim of homogenizing control to increase homogeneity, thereby reducing the number of disturbances.

The type of control presented in the previous paragraphs was applied in practice during experiments with the Dutch Motorway Control and Signalling System in 1983 [80]. A dual two-lane freeway near Utrecht and a dual three-lane freeway near Rotterdam were considered, each about 6 km in length. Data obtained before, after and during the experiments were investigated and the results were reported by Van Toorenburg [93]. The main conclusions that were drawn are as follows:

—the instability of traffic flow, measured as the number of serious speed drops, significantly decreases during homogenizing control. The decrease that was measured amounted to about 50% ;
—during control capacity did not decrease. There is some indication of a small increase (1-2%);
—no significant effect was measured in other traffic characteristics such as mean speed, speed differences, distribution over lanes, etc.

The general conclusion is that homogenizing control is advantageous in that it increases safety and reduces the probability of congestion. The effect is not so much in the homogenization, however, but in the stabilization of the traffic stream. No serious implementation problems were reported during the experiments, illustrating the relative ease of use and robustness of this type of control. The main problem that remains is to explain the established stabilizing effect.

The Traffic Engineering Division of the Dutch Ministry of Transport kindly provided the data from the experiments, of which we used the Utrecht data for further investigations. Eight batches of one hour of traffic each were considered: four with and four without control. These were chosen carefully to assure that the circumstances were similar. The following conclusions were drawn:

—when control is applied a significant reduction occurs in the percentage of small time headways ($\leqslant$1 sec), on the *left* lane. This effect does not occur on the right lane;
—the same holds for the variance of the time headways on the left lane;
—there is a slight but significant reduction in the mean speed, of equal size on both lanes;
—the mean space and time headway on the *right* lane decrease significantly.

We will now comment on the reported results.

The most important result is the decrease of the fraction of small headways on

the left lane. As mentioned before these small headways, corresponding to vehicles driving at close distance and driving at a high speed, are a major reason for the instability of the traffic stream. It is also known that congestion usually sets in on the left lane. The significance of the reduction is illustrated in Figure 6.1, where a plot of the fraction of small time headways in relation to the intensity has been copied from Van Toorenburg [91]. The data is separated per lane.

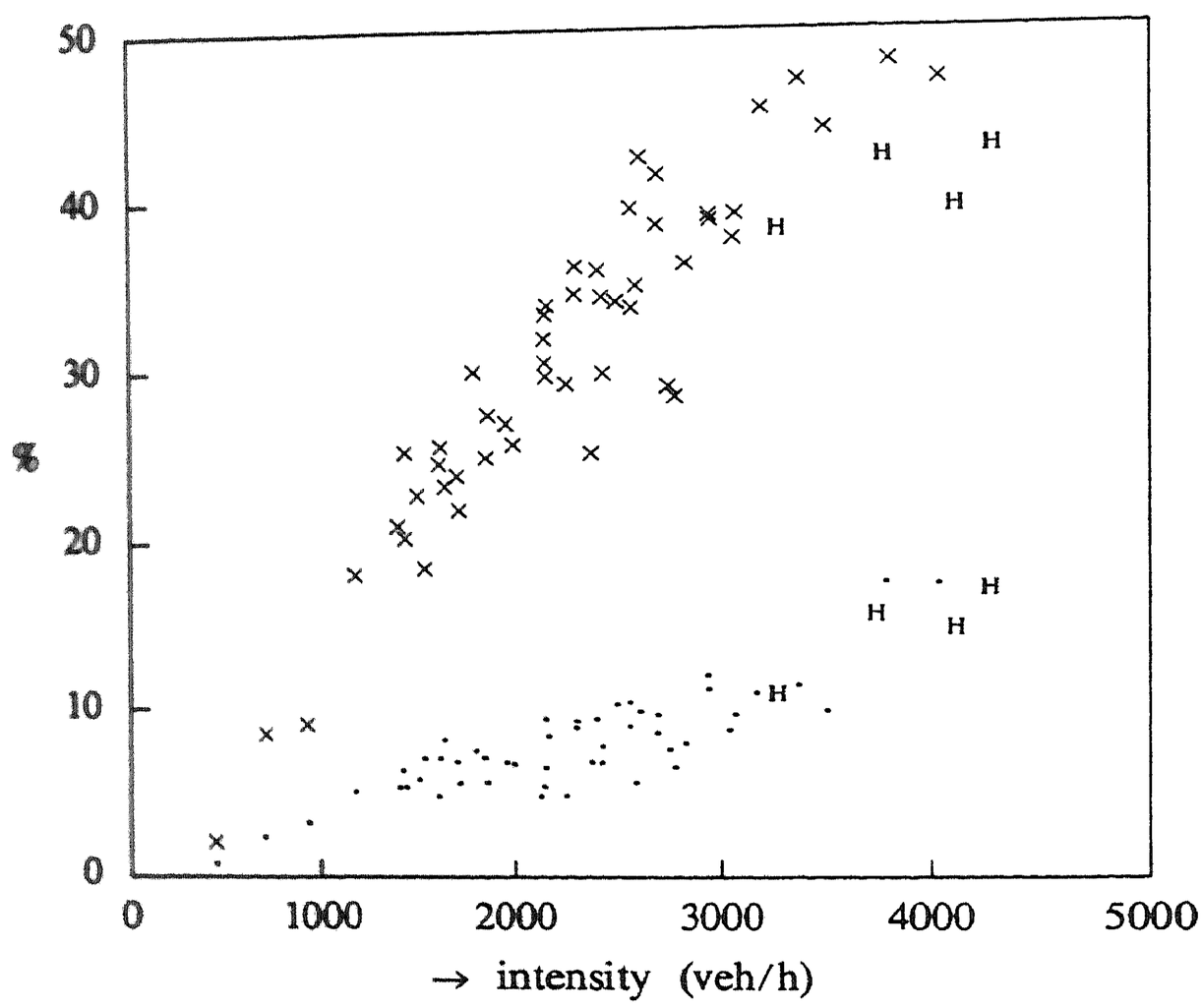

FIGURE 6.1. Fraction of small time headways on a two-lane carriageway.

Our measured fractions are added: the points marked "H" represent the ones with control. The effect on the left lane is clear and amounts to about 10% for intensity values near capacity. The plot also suggests that applying control is useless for intensities below 3000 veh/h, for one carriageway of a dual two-lane freeway.

The difference in character between left and right lane is illustrated once more by this plot. From Figure 6.1 it is seen that inhomogeneity tends to become smaller as intensity increases: the left lane appears to be saturated and the available space on the right lane is taken advantage of. The vehicles present are distributed more evenly over the available space. In practice the ideal state of a perfectly homogeneous stream is only reached after congestion has set in, however,

but then the intensity is far below capacity.

A further illustration of the effect of control is given in Figure 6.2 where the estimated probability densities of the time headways for a data set with and one without control are plotted. The estimation procedure was carried out using the statistical package *S* [5], using the function "density". This involves a histogram-like estimator, a description of which may be found in Chapter 4. Note that in contrast to what one is likely to expect, the time headways form a nearly uncorrelated process [10], thereby justifying the estimation procedure that is applied. The number of data points involved are approximately 2500 for the left lane and 1600 for the right lane.

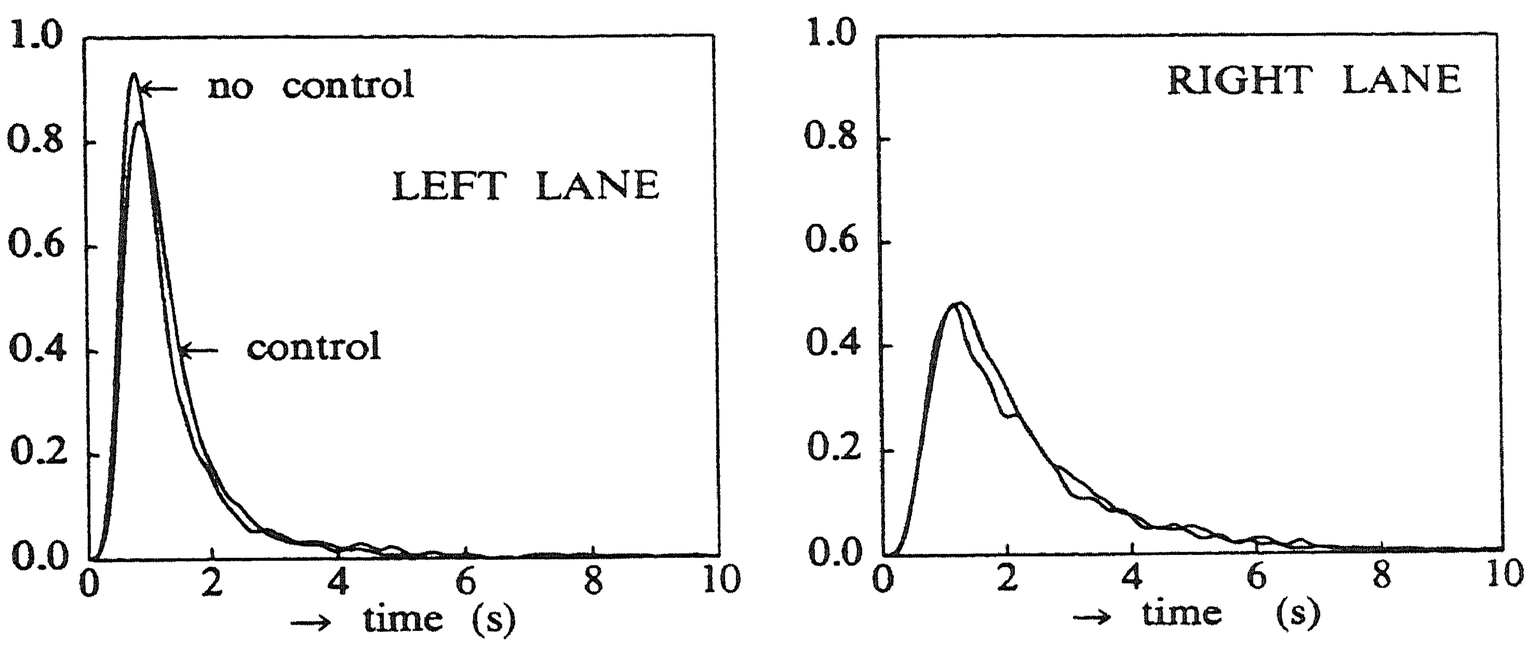

FIGURE 6.2. Estimated time headway probability densities.

A consequence of the reduction of small time headways is the decrease of the variance of these variables. For the plotted probability densities of the left lane this reduction amounts to 16%: from 2.22 to 1.86 s^2. This may be explained by the fact that the vehicles that under control start driving at a larger (safer) distance obtain the extra space they need from the large gaps that still occur, even at high intensities. Gaps of 100 m are not an exception when intensity is near capacity. Hence, both the fraction of larger as well as the fraction of smaller headways is reduced, resulting in a smaller variance.

The mean speed reduction resulting from control was found in all but one data set and varied from 0 to 5% on both lanes.

On the right lane the mean time headway decreased in all cases where control was applied, by 3% on the average. This implies an increase of intensity on the right lane of 3%. Near capacity, the distribution of the intensity between the left and the right lane is about 3:2 so that an increase of total intensity of about 1% may be expected.

Summarizing we note that the main effect of homogenizing control occurs on

the left lane and amounts to a regrouping of the available space and is local: only a fraction of the vehicles is involved. On the right lane the effect is global and leads to an increased intensity.

In the next section the effects of control as we reported them are incorporated into traffic models and the increased stability due to control are theoretically confirmed. The models we use are of a macroscopic type, meaning that traffic flow is described in terms of the aggregate variables density (number of veh/km/lane) and mean speed. Therefore, to conclude this section we now investigate the macroscopic implications of the microscopic effects of homogenizing control.

The reduced time headway variance on the left lane expresses itself in a reduction of the variance of the associated counting processes. The time headways per lane being nearly uncorrelated [10], we may model the counting processes per lane, N_t, as *renewal processes*. Then the process $\Delta N_t = N_{t+\Delta t} - N_t$ is asymptotically normally distributed with mean $\Delta t/\mu$, and the following asymptotic expression holds for the variance [18]:

$$\text{var}(N_{t+\Delta t} - N_t) \approx \frac{\sigma^2}{\mu^3}\Delta t$$

where μ and σ^2 are the mean and the variance of the time headway distribution respectively. From this one expects a reduction of the counting process variance on the left lane to result from the reduction of σ^2, and an increase on the right lane from the reduction in μ. The net effect on the counting process for both lanes is not clear. The expected effects were confirmed by a data analysis: for $\Delta t = 15$ s we have measured a reduction in $\text{var}(N_{t+\Delta t} - N_t)$ from 14.3 to 10.7 on the left lane (-25%) and an increase from 4.2 to 4.7 on the right lane ($+12\%$). The net effect for both lanes was measured to be a reduction from 25.3 to 20.7 (-18%).

Next, the reduction in the variance of the counting processes influences the traffic density in a freeway section. Now

$$d\rho_t = \frac{1}{Ll}(dN_t^{in} - dN_t^{out})$$

is the exact equation for the evolution of the density. Here L is the length of the section in km and l the number of lanes. N_t^{in} counts the number of vehicles entering the section and N_t^{out} the number of leaving vehicles. Hence,

$$\rho_{t+\Delta t} - \rho_t = \Delta\rho_t = \frac{1}{Ll}(\Delta N_t^{in} - \Delta N_t^{out})$$

and

$$\text{var}(\Delta\rho_t) = (\frac{1}{Ll})^2\left\{\text{var}(\Delta N_t^{in}) + \text{var}(\Delta N_t^{out}) - 2\text{cov}(\Delta N_t^{in}, \Delta N_t^{out})\right\}$$

The effect of control on the covariance of the counting processes is not immediately clear. Assuming that the effect is negligible and assuming that the covariance is small or at least positive, the reduction in $\text{var}(\Delta N_t)$ leads to a reduction in

var($\Delta\rho_t$).

Again we used data from the dual two-lane freeway near Utrecht to check this theoretical result. The filter developed in Chapter 5 was applied to a four section part of the freeway and the estimated density of the third section was used. It was found that for $\Delta t = 15$ s the variance of the density increments reduced from 31.1 to 23.3 (veh/km/l)^2 as a result of control (-25%). Other values of Δt gave analogous results.

Further analysis is carried out in the next section to assess the effect on the variance of the density itself instead of the increments. For this a simple traffic model is used. It is concluded that a reduction of var(ρ_t) occurs of the same relative magnitude as for the increments. We again tried to obtain confirmation from real data but failed this time. An explanation may be found in the instationary character of ρ_t, in contrast to the character of $\Delta\rho_t$. The former is very dependent on the accidental traffic demand which fluctuates strongly. Therefore, a large amount of data is needed to obtain an accurate estimate of var(ρ_t).

6.2. Traffic models and stability

In this section two traffic models are presented and the effect of the type of control described in the previous section is analysed. A model for freeway traffic flow has been developed in Chapters 2 and 3. There, the state of traffic is represented by the density (the number of vehicles per km per lane) and mean speed in a series of consecutive freeway sections. The stochastic differential equation for the density followed from a conservation of vehicles argument, the equation for the mean speed is based on the model of Payne [77], also used by Van Maarseveen [62]. Although this model shows satisfactory behaviour and is able to represent the instabilities that occur in practice, we do not use it here. Instead, we consider models for one section of freeway only. The simplicity of these models allows to carry out some analysis and keeps the control policy design procedure feasible. The model for traffic in more than one section may be considered at a later stage, using the insight obtained in the problems that are considered here.

6.2.1. A one-dimensional model: traffic density.

The simplest model for traffic on a freeway that still incorporates the main features consists of the stochastic differential equation for the density in a single section. Denote

ρ_t : the density (veh/km/lane) in the section at time t;

v_t : the mean speed (km/h) in the section at time t,

then the model is given by

$$d\rho_t = \frac{1}{Ll}(\lambda_0 - l\rho_t v_t)dt + \sigma dw_t \tag{6.1}$$

$$v_t = v^e(\rho_t) \tag{6.2}$$

Here $v^e(\rho)$ denotes the *equilibrium speed* corresponding to density ρ, as defined

in Chapter 3:

$$v^e(\rho) = \begin{cases} v_{free} - \alpha\rho & , \quad 0 \leqslant \rho \leqslant \rho_{crit} \\ d(\frac{1}{\rho} - \frac{1}{\rho_{jam}}) & , \quad \rho_{crit} < \rho \leqslant \rho_{jam} \end{cases} \tag{6.3}$$

where

$$d = \frac{v_{free} - \alpha\rho_{crit}}{\frac{1}{\rho_{crit}} - \frac{1}{\rho_{jam}}}$$

to assure continuity of v^e at ρ_{crit}.

The other parameters are defined as follows:

l: the number of lanes of the section;

L: the length of the section (km);

λ_0: the intensity at the entrance (veh/h);

σ: the standard deviation of the noise;

w: a standard Brownian motion (veh/km/lane);

v_{free}: the free speed (km/h);

ρ_{crit}: the critical density (veh/km/lane);

ρ_{jam}: the jam density, at which v^e is zero (veh/km/lane);

Realistic values of v_{free}, ρ_{crit} and α have been found by direct estimation from freeway data in Chapter 4:

$$v_{free} = 105 \text{ km/h}$$

$$\rho_{crit} = 27 \text{ veh/km/lane}$$

$$\alpha = 0.58 \text{ km}^2/\text{h}$$

The other parameters are chosen as follows:

$$l = 2$$

$$L = 0.5 \text{ km}$$

$$\rho_{jam} = 110 \text{ veh/km/lane}$$

The variance σ^2 is chosen to achieve a realistic value for the variance of ρ_t. From filtered freeway data we found $\text{var}(\rho) \approx 50$ (veh/km/lane)^2. Using a linearized version of (6.1) then leads to $\sigma^2 \approx 14000$.

The model given by (6.1) and (6.2) describes the evolution in time of the density of a freeway section when the entrance intensity λ_0 is given. This evolution is subject to considerable noise, modelled by a Brownian motion process. Using

a Brownian motion instead of counting process martingales implies that ρ_t is continuous, whereas in practice it is discrete: 0, $1/Ll$, ... This is an approximation of minor importance. Note that the variance of the noise is not based on a diffusion approximation argument for counting processes. This would lead to a variance that depends on ρ and could be considered in the future.

The model represents the main feature we are interested in, the occurrence of congestion when λ_0 is near the capacity level. The *capacity* λ^{cap} of the freeway section is defined as follows and may be computed for the parameter values mentioned before:

$$\lambda^{cap} = \max_{0 \le \rho \le \rho_{jam}} l\rho v^e(\rho) = 4824 \text{ veh/h}.$$

To show that congestion may occur we first consider the deterministic version of (6.1), (6.2), that is, $\sigma = 0$. In case $\lambda_0 < \lambda^{cap}$ two equilibrium points may be shown to exist, one of which is stable and one of which is unstable. See Figure 6.3. If $\lambda_0 = \lambda^{cap}$ there is one unstable equilibrium and if $\lambda_0 > \lambda^{cap}$ no equilibrium point exists. To prove this we will assume $\rho_{crit} < \frac{v_{free}}{2\alpha}$ holds. This always holds in practice, as Figure 6.3 shows: taking $\rho_{crit} \geqslant \frac{v_{free}}{2\alpha}$ leads to a value of λ^{cap} that is far too large.

PROPOSITION 6.1. *Under the condition* $\rho_{crit} < \frac{v_{free}}{2\alpha}$ *the following holds:*

i) if $\lambda_0 < \lambda^{cap}$ *the deterministic version of the model (6.1), (6.2) has two equilibrium points*

$$\rho_0^s = \frac{v_{free}}{2\alpha} - \sqrt{(\frac{v_{free}}{2\alpha})^2 - \frac{\lambda_0}{l\alpha}}$$

$$\rho_0^u = (1 - \frac{\lambda_0}{ld})\,\rho_{jam}$$

of which ρ_0^s *is stable and* ρ_0^u *is unstable;*

ii) if $\lambda_0 = \lambda^{cap}$ *the model (6.1), (6.2) has one equilibrium point*

$$\rho_0^u = \rho_{crit}.$$

ρ_0^u *is unstable;*

iii) if $\lambda_0 > \lambda^{cap}$ *the model (6.1), (6.2) has no equilibrium points.*

PROOF.

i) The equilibrium points may easily be found by setting the first term on the right-hand side of (6.1) equal to zero. Then the local asymptotic stability of ρ_0^s may be checked by linearization of the model around this point. If $x_t = \rho_t - \rho_0^s$ then

$$\frac{dx_t}{dt} \approx \frac{1}{L}(2\alpha\rho_o^s - v_{free})x_t$$

for $x_t \leqslant \rho_{crit} - \rho_0^s$. Stability follows from $\rho_0^s \leqslant \rho_{crit}$ and the condition $\rho_{crit} < \frac{v_{free}}{2\alpha}$. The instability of ρ_0^u follows immediately from $y_t = \rho_t - \rho_0^u$ and

$$\frac{dy_t}{dt} = \frac{d}{L\rho_{jam}} y_t.$$

ii) Starting from $\rho = \rho_0^u + \epsilon$ with $\epsilon > 0$ it follows from the linearity of the model for $\rho \geqslant \rho_{crit}$ and the analysis of *i*) that ρ_0^u is unstable.

iii) $\lambda_0 - l\rho v^e(\rho) > 0$ for all ρ. □

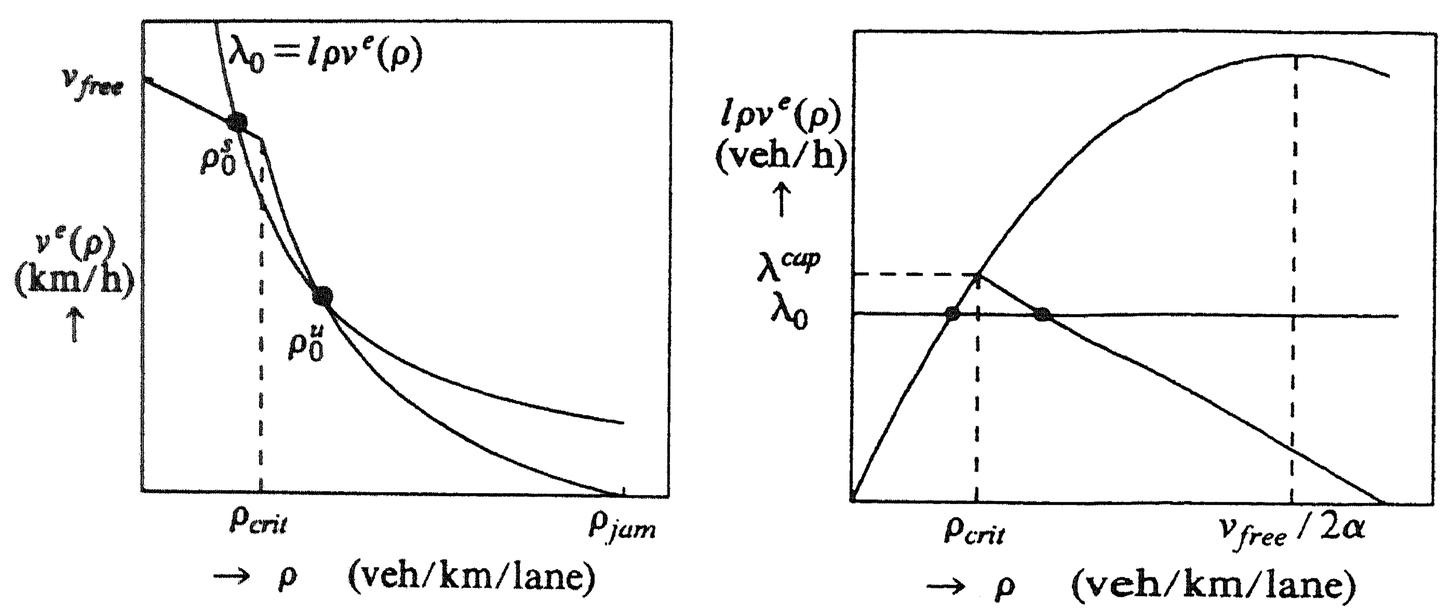

FIGURE 6.3.

The value of ρ_{crit} is to be interpreted as follows: it is the instability point when the entrance intensity is at capacity level. Table 6.1 presents the values of ρ_0^s and ρ_0^u for several values of λ_0.

PROPOSITION 6.2. *Consider the deterministic version of (6.1), (6.2). If* $\lambda_0 < \lambda^{cap}$ *and* $\rho_{crit} < \frac{v_{free}}{2\alpha}$ *the* domain of attraction *of* ρ_0^s *is* $(-\infty, \rho_0^u)$.

PROOF.

Examination of (6.1) (with $\sigma = 0$) shows that $\frac{d\rho}{dt} > 0$ for all $\rho < \rho_0^s$ and that $\frac{d\rho}{dt} < 0$ for $\rho_0^s < \rho \leqslant \rho_{crit}$ and hence convergence to ρ_0^s is assured for all $\rho \leqslant \rho_{crit}$. For $\rho > \rho_{crit}$ the model is affine,

$$\frac{d\rho}{dt} = \frac{\lambda_0 - ld}{Ll} + \frac{d}{L\rho_{jam}}\rho,$$

and an exact expression for the solution may be found to show that ρ converges to ρ_{crit} as long as $\rho<\rho_0^u$. If $\rho>\rho_0^u$ convergence to ∞ follows. Combining the two conclusions completes the proof. □

Before investigating the stochastic model we make a small modification to the original equations. Up to now the domain of ρ is the entire real line, which is unrealistic. In practice ρ is confined to the interval $[0, \rho_{jam}]$. It is possible to modify the equation (6.1) in such a way that this is assured. A more convenient way to do this is to define the process $\{\rho_t, t\geqslant 0\}$ by its *infinitesimal generator* and using suitable boundary conditions. A description of this is given by [65].

DEFINITION 6.3. *Define the process* $\{\rho_t, t\geqslant 0\}$ *on* $[0,\rho_{jam}]$ *by the infinitesimal generator*

$$\frac{1}{2}\sigma^2\frac{d^2}{d\rho^2} + [\lambda_0 - l\rho v^e(\rho)]\frac{d}{d\rho} \qquad (6.4)$$

with the boundary at $\rho=0$ *reflecting and the boundary at* $\rho=\rho_{jam}$ *absorbing.*

Let us now turn to the effect of the noise term in (6.1). Because of the continuous disturbance convergence to ρ_0^s will not occur. Instead, with $\sigma>0$ convergence to ρ_{jam} occurs with probability one. If we define

$$\tau = \inf\{t\geqslant 0 : \rho_t=\rho_{jam} \mid \rho(0)\}$$

then it is known [65] that $E[\tau \mid \rho(0)]<\infty$, which with $\tau\geqslant 0$ leads to $P(\tau<\infty)=1$. Hence, even at low intensity congestion eventually occurs. This is not as unrealistic as it may look, because if λ_0 is low, $E[\tau \mid \rho(0)]$ is large. Some values of this expectation may be found in Table 6.2.

Several stability concepts for stochastic systems have been proposed in the literature, e.g. by Kozin [47], and by Hasminski [43]. In our case all of these lead to trivial conclusions, however: the point ρ_{jam} is stable almost surely and in the p-th mean for all $p\geqslant 1$, and the invariant measure has all its mass at this point. We use the *mean time to congestion* $E[\tau \mid \rho_0^s]$ as a measure of stochastic stability in the sequel therefore.

The effect of homogenizing control is now incorporated into the model. In the previous section we described the effects of control on several traffic characteristics. On the left lane the percentage of short time headways decreases significantly, leading to a smaller variance of these headways. This was then shown to lead to a reduction of the variance of density increments over short time intervals (15 sec.). On the right lane the mean time headway decreases, implying a larger intensity. On both lanes the mean speed decreases slightly.

Directing attention to the effect on the density first, we note that a simple analysis of (6.1), consisting of a linearization around ρ_0^s, leads to

$$\text{var}(\Delta\rho_t) = \sigma^2\Delta t(1 - \tfrac{1}{2}a\Delta t)$$

where

$$a = \frac{(2\alpha\rho_0^s - v_{free})}{L}.$$

Neglecting possible effects of control on a we see that the reduction of $\mathrm{var}(\Delta\rho)$ has to be modelled as a reduction of σ^2: the noise variance. The relative reduction in σ^2 is to be of the same magnitude as the one of $\mathrm{var}(\Delta\rho)$, which was 20 to 25%. This gives a σ^2-value of about 11000 under control.

The reduction in speed under control is immediately expressed in v^e by a reduction of v_{free}. Of course, the speed reduction was only measured for intensity values near capacity. No experimental data for lower intensities are available, for obvious reasons: applying homogenizing control for low intensity values makes no sense. We have assumed that control always leads to a small decrease in speed whenever it is applied.

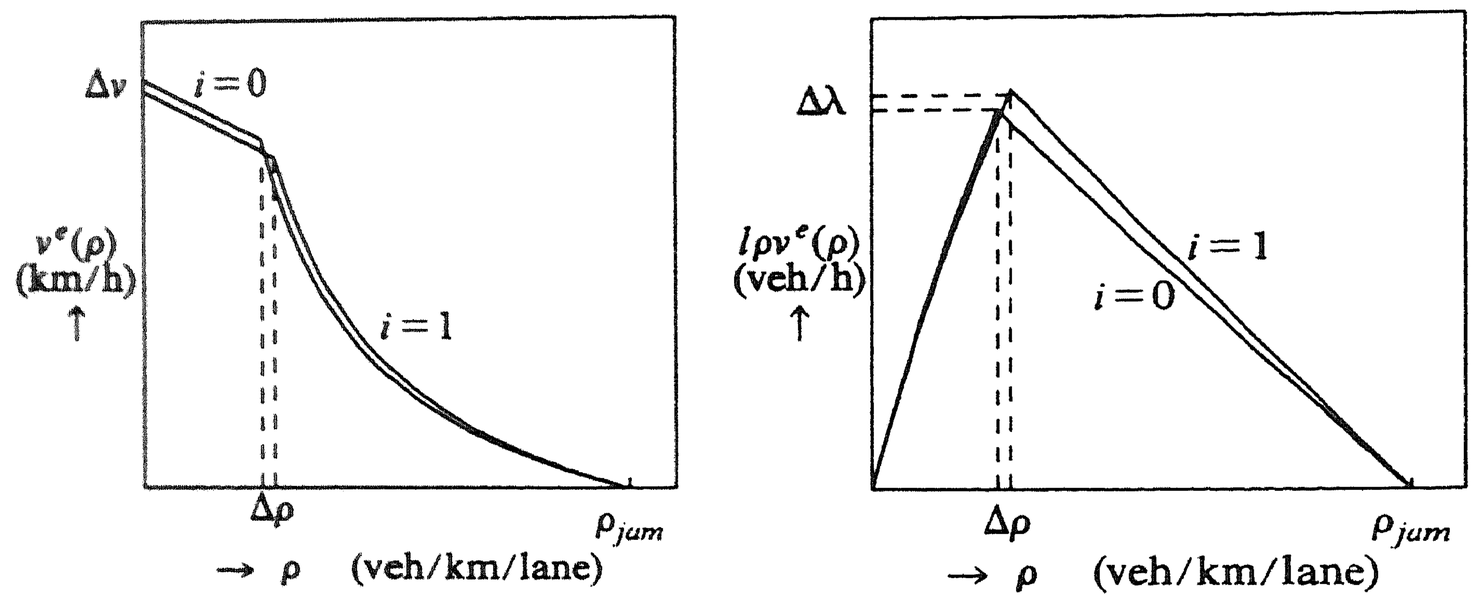

FIGURE 6.4. Equilibrium speed-density and intensity-density relations with ($i = 1$) and without control ($i = 0$).

The conclusion drawn by Van Toorenburg [93] that capacity does not decrease and possibly increases slightly, together with the speed reduction effect of the previous paragraph, implies that ρ_{crit} increases. See Figure 6.4 for a plot of $v^e(\rho)$ and the corresponding plot of $\lambda^e(\rho) = l\rho v^e(\rho)$. Keeping ρ_{crit} constant would lead to a lower capacity.

The slight increase in total flow due to the effects on the right lane may be modelled by a small increase in λ_0.

We now present the model with control:

$$d\rho_t = \frac{1}{Ll}(\lambda_0^i - l\rho_t v_t^i)dt + \sigma^i dw_t \tag{6.5}$$

$$v_t^i = v_i^e(\rho_t) \tag{6.6}$$

Here $i=0$ means that no control is applied and $i=1$ that it is. Furthermore, v_i^e is defined as in (6.3) except that

$$v_{free}^i = \begin{cases} v_{free} & , \quad i=0 \\ v_{free} - \Delta v & , \quad i=1 \ , \end{cases}$$

$$\rho_{crit}^i = \begin{cases} \rho_{crit} & , \quad i=0 \\ \rho_{crit} + \Delta\rho & , \quad i=1 \ , \end{cases}$$

$$\lambda_0^i = \begin{cases} \lambda_0 & , \quad i=0 \\ \lambda_0 + \Delta\lambda & , \quad i=1 \ . \end{cases}$$

From the measurements of the experiments with control, discussed in Section 6.1, the following values for the parameters are obtained:

$$\Delta v = 3 \text{ km/h}$$

$$\Delta\rho = 2 \text{ veh/km/lane}$$

$$\Delta\lambda = 0.01\lambda_0 \text{ veh/h}$$

$$\sigma_0^2 = 14000 \ , \ \sigma_1^2 = 11000.$$

Note that the relative reduction in σ^2 is 21%, and that the increase in capacity due to the given numbers is equal to 2.4%. Interpreting $1/d$, a parameter of $v_i^e(\rho)$, as the reaction time of a driver, as in Section 3 of Chapter 3, we note that this reaction time decreases from 1.13 to 1.07 seconds as a result of control.

As before we consider a modification of (6.5) such that ρ is confined to $[0, \rho_{jam}]$ where 0 is reflecting and ρ_{jam} is absorbing.

The stabilizing effect of control according to this theoretical model may now be investigated. It is clear that applying control leads to a larger degree of stability because of the following facts:

—reduced variance of ρ_t leads to a longer average time for convergence to ρ_{jam};
—the distance between ρ_0^s and ρ_0^u increases because of the effects on speed and capacity.

The latter fact may be deduced from the results of Table 6.1 where the equilibrium points are given, corresponding to control and no control, respectively.

	no control		control	
λ_0 (veh/h)	ρ_0^s (veh/km/lane)	ρ_0^u (veh/km/lane)	ρ_0^s (veh/km/lane)	ρ_0^u (veh/km/lane)
1000	4.9	92.8	5.0	93.6
2000	10.1	75.6	10.4	77.2
3000	15.6	58.4	16.2	60.8
4000	21.6	41.2	22.5	44.4
4800	26.8	27.4	28.0	31.3

TABLE 6.1. Equilibrium points of the one-dimensional traffic model.

A clear view of the positive effect is obtained from Table 6.2 where the value of $E[\tau \mid \rho(0)]$ is given for several intensity levels, with and without control. Note that $E[\tau \mid \rho(0)]$ depends on the initial value associated with the differential equation (6.5). The values presented are for the initial value ρ_0^s. We conclude that for high intensity values congestion may only be postponed somewhat and that for intensities about 20% below capacity the effect has the most significance.

	no control	control
λ_0 (veh/h)	$E[\tau \mid \rho_0^s]$ (min)	$E[\tau \mid \rho_0^s]$ (min)
1000	$9.6 * 10^{10}$	$2.3 * 10^{14}$
2000	$2.2 * 10^{6}$	$2.0 * 10^{8}$
3000	1044	8344
3500	81.15	263.3
4000	15.28	25.82
4400	6.68	8.40
4600	4.94	5.78
4800	3.83	4.25

TABLE 6.2. Mean time to congestion with and without control.

6.2.2. A two-dimensional model: density and speed.
The previous model is unrealistic in that the speed reacts instantaneously to a change in density. Instead we now propose

$$d\rho_t = \frac{1}{Ll}(\lambda_0 - l\rho_t v_t)dt + \sigma dw_t \tag{6.7}$$

$$dv_t = -\frac{1}{T}[v_t - v^e(\rho_t)]dt + \mu dz_t \tag{6.8}$$

where

T : the relaxation time (h)

μ : the standard deviation of the acceleration noise

z_t : a standard Browian motion (km/h)

It is assumed that z_t is independent of w_t. We take

$$T = 0.01 \text{ h}$$

in accordance with the value in Chapter 3 and

$$\mu^2 = 10000$$

in order to achieve var(v)$\approx$40 (km/h)^2 which is measured in practice. This value was also used in Chapter 5. The other parameters have the values as given in Section 6.2.1.

This two-dimensional model is more realistic than the one-dimensional one in that it allows larger fluctuations in the density without congestion occurring. Consider the deterministic model ($\sigma=\mu=0$). Again defining λ^{cap} as the maximum of $l\rho v^e(\rho)$, it is easy to show that when $\lambda_0<\lambda^{cap}$ the equilibrium points are $(\rho_0^s, v^e(\rho_0^s))$ and $(\rho_0^u, v^e(\rho_0^u))$, where ρ_0^s and ρ_0^u are as in Proposition 6.1: setting the right-hand side of (6.8) to zero leads to $v=v^e(\rho)$. Also, linearization around these points shows that the first one is stable and the second one unstable. The unstable equilibrium is a saddle-point, as examination of the eigenvalues of the linearized model shows. Determining the domain of attraction of the stable equilibrium point requires more effort. Drawing the phase portrait gives an impression of this domain, see Figure 6.5 for an example where $\lambda_0 = 4000$ veh/h.

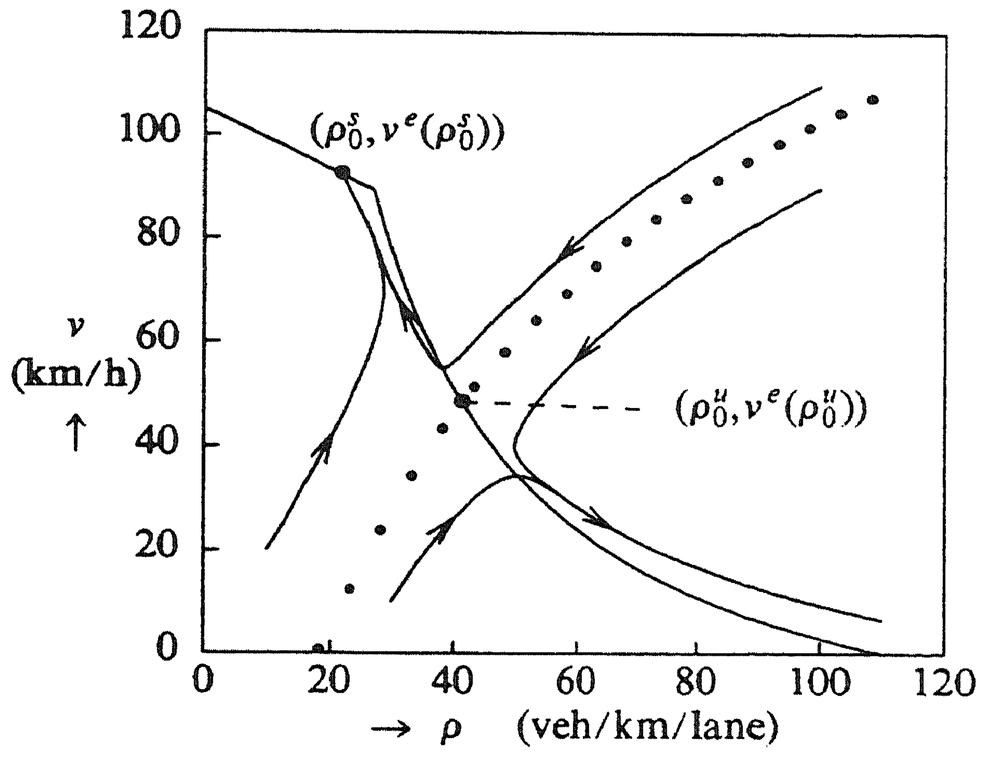

FIGURE 6.5. State trajectories and an approximation of the separator.

There appears to be a line s, which we refer to as the *separator* from now on, which divides the plane in two and is the boundary of the domain of attraction of

$(\rho_0^s, v^e(\rho_0^s))$. The separator consists of two trajectories that meet in $(\rho_0^u, v^e(\rho_0^u))$: starting in a point on the separator convergence to the unstable equilibrium point follows. This observation allows us to compute s by parametrizing it by ρ, $s(\rho)$, and deriving the differential equation

$$\frac{ds(\rho)}{d\rho} = \frac{Ll}{T} \frac{[v^e(\rho) - s(\rho)]}{[\lambda_0 - l\rho s(\rho)]} \tag{6.9}$$

with initial condition

$$s(\rho_0^u) = v^e(\rho_0^u)$$

A small problem occurs in solving (6.9) as the right-hand side is undefined in the initial point. To overcome this we apply the rule of de l'Hôpital

$$\lim_{\rho \to \rho_0^u} \frac{ds(\rho)}{d\rho} = \lim_{\rho \to \rho_0^u} \frac{Ll}{T} \frac{[\dot{v}^e(\rho) - \gamma]}{[-ls(\rho) - l\rho\gamma]}$$

where

$$\gamma = \left. \frac{ds(\rho)}{d\rho} \right|_{(\rho_0^u, v^e(\rho_0^u))}.$$

Continuity of $\frac{ds(\rho)}{d\rho}$ at the unstable equilibrium point then gives us a quadratic equation in γ and the solution

$$\gamma = \frac{\frac{L}{T} - v^e(\rho_0^u) + \sqrt{[\frac{L}{T} - v^e(\rho_0^u)]^2 + \frac{4Ld}{T\rho_0^u}}}{2\rho_0^u}.$$

Now (6.9) may be numerically integrated to obtain the separator. Plots for several values of λ_0 are presented in Figure 6.6. Note that $\lambda_0 = \lambda^{cap}$ is a bifurcation point.

It is of interest to note that if $T \to 0$ (instantaneous reaction of the speed) $\gamma \to \infty$ and so the result is in accordance with the one-dimensional case. If $T \to \infty$ then $\gamma \to 0$: the slower the speed reacts to changes in the density, the larger the fluctuations in the latter may be without leading to congestion. For $\lambda_0 = 4000$ veh/h the separator is plotted for several values of T in Figure 6.7.

The derivation of the equation of the separator has been somewhat heuristic, based on the plot of the phase portrait. An exact proof of the correctness of our procedure is given by Chiang, Hirsch and Wu [16] for systems with a vector field that is continuously differentiable:

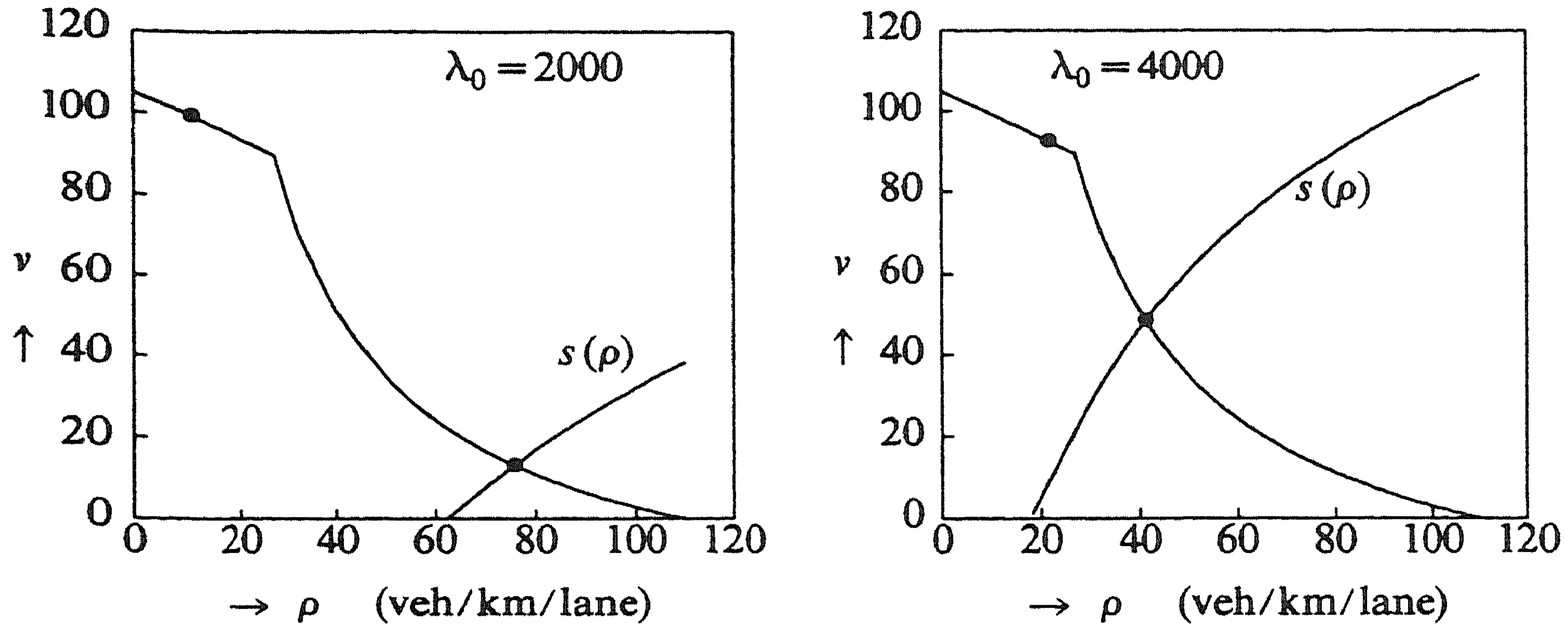

FIGURE 6.6. Computed separators of the two-dimensional traffic model.

THEOREM 6.4 [Chiang et al.]. *The stability boundary of a stable equilibrium point is given by*

$$\bigcup_i A^s(x_i)$$

where $x_1, x_2, \ldots$ are the equilibrium points on the stability boundary and $A^s(x_i)$ is the stable manifold of x_i.

The stable manifold of an equilibrium point is the set of all points that converge

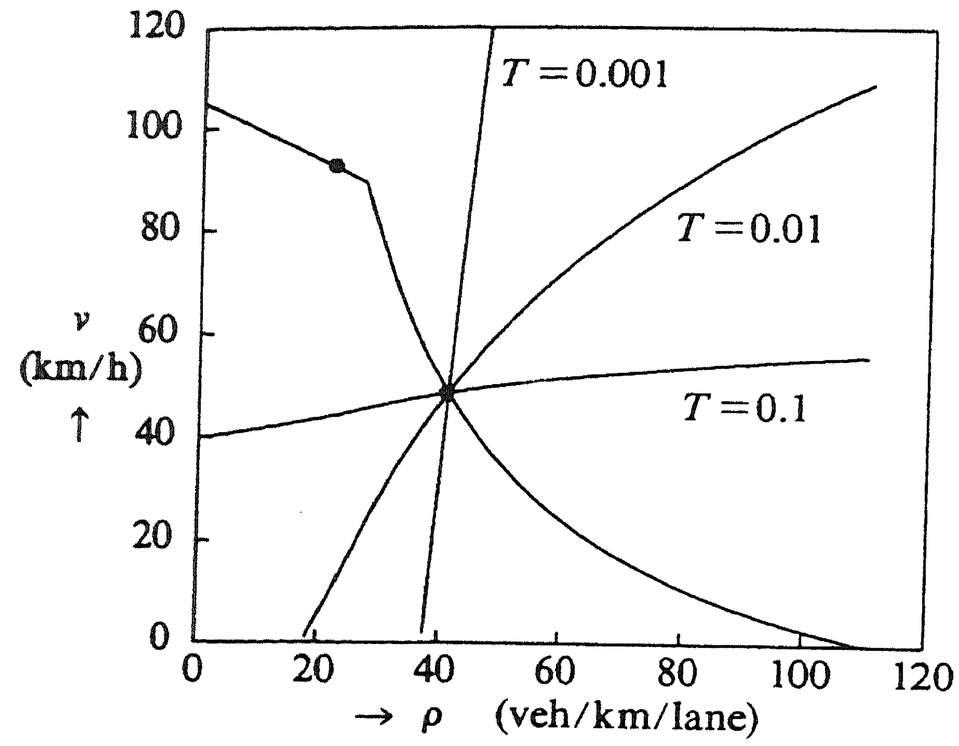

FIGURE 6.7. The separator for different relaxation time values.

to the equilibrium point. In our case the unstable equilibrium point lies on the stability boundary of $(\rho_0^s, v^e(\rho_0^s))$ and its stable manifold consists of the two trajectories ending in $(\rho_0^u, v^e(\rho_0^u))$ itself.

A problem in Theorem 6.4 is that a characterization of equilibrium points that are on the stability boundary is needed. Such characterizations are given by Chiang, Hirsch and Wu [16]. For Theorem 6.4 to be true several nontrivial conditions have to be satisfied. These are not given here. In our model at the point where $\rho = \rho_{crit}$ the vector field is not continuously differentiable, as required. As long as $\lambda_0 < \lambda^{cap}$ no problems are to be expected, however, as the smoothness assumption is mainly used as a sufficient condition for existence of a unique solution of the differential equation. A way to avoid the problem would be to consider a modification of (6.7), (6.8) where the vector field *is* continuously differentiable. This is possible without affecting the essential properties of the model. Some suggestions for a smooth $v^e(\rho)$ are given by Cremer [20]. We do not go into detail here.

Modelling the effect of control exactly parallels the discussion in the previous subsection and is not repeated here. The control decreases the variance in the density equation and modifies $v^e(\rho)$, and is assumed not to have any effect on T and μ.

Again, as in Section 6.2.1 we have to restrict the domain of the diffusion $\{(\rho_t, v_t),\ t \geqslant 0\}$ to be bounded. The domain that is taken is

$$[0,\ \rho_{jam}] \times [0,\ v_{max}]$$

where ρ_{jam} was defined before and

$$v_{max} = 150 \text{ km/h}.$$

The latter choice is rather arbitrary, as is the rectangular shape of the domain.

As $v_t \approx v^e(\rho_t)$ and $v_{free} = 105$ km/h, the choice above is expected not to be crucial, however.

It is natural to take the boundaries to be reflecting, except for the boundary $\rho = \rho_{jam}$. At first we took this boundary to be reflecting also, except for the point $(\rho_{jam}, 0)$. Traffic congestion is then associated with the situation of maximal density and a standstill. This choice turned out to be unfortunate, however, with the diffusion process bouncing back and forth on the boundaries $\rho = \rho_{jam}$ and $v = 0$ for a relatively long time, before ending in the congestion point. It turned out to be more realistic to choose that part of the boundary $\rho = \rho_{jam}$ to be reflecting for which

$$\lambda_0 < l\rho_{jam} v$$

and the other part absorbing. One reason for this is that in the deterministic case the vector field on the boundary points inward in case $\lambda_0 < l\rho_{jam} v$ and outward in case $\lambda_0 > l\rho_{jam} v$. Experimentation with the boundary condition has shown that the problem is not very sensitive to the exact form of the condition, as long as absorption occurs at a subset of nonzero measure which includes $(\rho_{jam}, 0)$.

As for the one-dimensional model we may compute the mean time to congestion for the case with and without control for several intensity levels. This requires the solution of an elliptic partial differential equation in two variables, comparable to the one that will appear in Section 6.3.5:

$$\frac{1}{2}\sigma_i^2 \frac{\partial^2 W}{\partial \rho^2} + \frac{1}{2}\mu^2 \frac{\partial^2 W}{\partial v^2} + \frac{1}{Ll}[\lambda_0^i - l\rho v]\frac{\partial W}{\partial \rho} + \frac{1}{T}[v_i^e(\rho) - v]\frac{\partial W}{\partial v} + 1 = 0 \qquad (6.10)$$

where

$$W(\rho_0, v_0) = E[\tau \mid \rho_0, v_0],$$

with

$$\tau = \inf\{t \geqslant 0 : \rho_t = \rho_{jam} \wedge v_t \leqslant \lambda_0 / l\rho_{jam}\}.$$

The boundary conditions are

$$\frac{\partial W}{\partial \rho}(0, v) = 0, \ \frac{\partial W}{\partial \rho}(\rho_{jam}, v) = 0 \text{ for } v > \lambda_0 / l\rho_{jam}, \ \frac{\partial W}{\partial v}(\rho, 0) = 0,$$

and

$$\frac{\partial W}{\partial v}(\rho, v_{max}) = 0, \ W(\rho_{jam}, 0) = 0 \text{ for } v \leqslant \lambda_0 / l\rho_{jam}.$$

The equation is solved by a special routine [104]. The results are given in Table 6.3.

For λ_0 close to λ^{cap} an accurate solution was obtained. As λ_0 is decreased a more refined grid is needed in the discretization approach to solve (6.10), however, resulting in a larger computational effort. The hardest problem that is solved is the case $\lambda_0 = 4000$ veh/h, with control. Here a grid of about 150×150 points turned out to be necessary, and the amount of memory required led us to use the Cyber 205 supercomputer instead of the Cyber 750 mainframe.

Comparing results to those of Table 6.2 one notes that indeed for fixed λ_0 the time to congestion is larger for the two-dimensional model. The difference is significant and justifies considering this more complicated model instead of the simple one-dimensional version. The relative effect of applying control for a fixed λ_0 is larger for the two-dimensional model also. The most important effect of control occurs at intensities above 4000 veh/h (10% below capacity), whereas in the one-dimensional case it occurred about at 4000 veh/h.

	no control	control
λ_0 (veh/h)	$E[\,\tau \mid \rho_0^s,\ v^e(\rho_0^s)\,]$ (min)	$E[\,\tau \mid \rho_0^s,\ v^e(\rho_0^s)\,]$ (min)
4000	96.9 ± 2.4	246.0 ± ?
4200	41.9 ± 0.35	80.0 ± 1.0
4400	22.4 ± 0.15	34.1 ± 0.7
4600	14.08 ± 0.05	18.5 ± 0.2
4800	9.925 ± 0.005	11.9 ± 0.1

TABLE 6.3. Mean time to congestion for the two-dimensional model.

6.3. CONTROL POLICY DESIGN

In this section control policies are designed of the homogenizing type, based on the traffic models described in Section 6.2. The control corresponding to the one-dimensional model of Section 6.2.1 is extensively discussed in the Sections 6.3.2, 6.3.3 and 6.3.4. Some preliminary investigations concerning control for the two-dimensional model of 6.2.2 are presented in 6.3.5.

6.3.1. Introduction.

It is the aim of control to reduce the instabilities present in traffic flow when intensity approaches capacity. As was shown in Section 6.2 we cannot avoid congestion but can only hope to postpone it. However, we may reduce the probability that congestion occurs within a given time interval. As in practice intensity stays at a high level for a limited time only, control may actually lead to avoidance of congestion.

The advantage of applying control being clear one may wonder why not to apply control always. There are good reasons not to apply control for intensity levels far below capacity:

—when intensity is low, control has no significant effect;
—drivers may become so accustomed to the control measure that they no longer react to it, even in situations where it would be beneficial.

The first point was illustrated in Figure 6.1: when intensity is low there are not enough small time headways left to produce a significant effect. The traffic stream is stable enough already. The second point is very important: the control is based on the supposed effects on driver behaviour and entirely depends on

this. If drivers no longer accept the control measure as reasonable the positive effect is lost. It is very important not to challenge the confidence that drivers have in the control system.

The advantage of control is easily quantified, by the mean time to congestion, for instance (see Table 6.2). But how to quantify the disadvantage of control at low intensities, mentioned in the previous paragraph? One may of course assign a cost to applying control but this is rather arbitrary. After some preliminary investigations with such a cost for control we decided to follow another approach: the control is not applied unless the optimization criterion (to be defined later) significantly decreases. Control is postponed as long as possible, expressing the idea that no control should be applied unless it is strictly necessary. This is discussed in more detail later.

An extension of the work presented here would consist of introducing an auxiliary state variable which measures the *degree of acceptance* of the control measure by the drivers. This variable may range from 1 (full acceptance) to 0 (no acceptance). If control is applied too early, at densities that are too low, the degree of acceptance decreases. The effect of control on the equilibrium speed-density relation and the density noise should decrease with decreasing acceptance. When acceptance is zero, control no longer has any effect. A model for the acceptance variable may be based on experience with acceptance of other traffic measures such as speed limits, alcohol checks etc. This extension of the model would solve the problem mentioned in the previous paragraph, and quantify the disadvantage of applying control always. The control policy design procedure would be considerably more complicated by this extension, however, and is therefore not investigated further.

In the next subsection the criterion that has to be maximized is defined and the maximization problem is investigated. It turns out that attention has to be directed to a finite horizon or to a discounted criterion. The solution to these problems is discussed in Section 6.3.3 for the one-dimensional case, but it is of interest to consider the general multidimensional control design problem here. The existence results may be used to cover the two-dimensional case that is investigated in Section 6.3.5 and to cover future extensions.

The problem that is considered is of the following type: given the n-dimensional diffusion processes $\{x_t^r,\ t \geqslant 0\}$

$$dx_t^r = f^r(x_t^r)dt + g^r(x_t^r)dw_t$$

each of which evolves according to a different *regime* $r \in \{1,...,m\}$, for any series of stopping times $\bar{\theta}=(\theta_0,\theta_1,...)$ with $0=\theta_0<\theta_1<\cdots$, $\lim_{j\to\infty}\theta_j=\infty$ a.s. and a corresponding series of regimes $\bar{r}=(r_0,r_1,...)$ with $r_j \in \{1,...,m\}$ the following process may be defined:

$$dx_t = dx_t^{r_k} \qquad \text{for } \theta_k \leqslant t < \theta_{k+1} \text{ and } k=0,1,... \tag{6.11}$$

The process $\{x_t,\ t \geqslant 0\}$ is defined to be continuous at all θ_k, and x_0 and r_0 are assumed to be given. Corresponding to $\bar{\theta}$ and $\bar{r}$ a control policy $\{u_t,\ t \geqslant 0\}$ may

be defined as follows:

$$u_t = r_k \quad \text{for } \theta_k \leqslant t < \theta_{k+1}, \quad k = 0, 1, \ldots$$

The object is to find a control policy u that optimizes the behaviour of the diffusion $\{x_t, t \geqslant 0\}$ according to some criterion function. The criterion we use is the discounted reward

$$J_u(x_0) = E[\int_0^\tau e^{-ct} h(x_t, u_t) dt \mid x_0] \tag{6.12}$$

where

$$\tau = \inf\{t \geqslant 0 : x_t \notin \Theta\},$$

the *exit time* from an open, bounded set $\Theta \subset \mathbb{R}^n$ with a smooth boundary $\partial\Theta$. The function h represents the reward per time unit and is allowed to depend on the control policy. The discount factor c is assumed to be strictly positive. A heuristic derivation using the well-known dynamic programming principle of R. Bellman leads to the conclusion that the *optimal value function* J^*,

$$J^*(x) = \sup_u J_u(x), \tag{6.13}$$

has to satisfy the *Hamilton-Jacobi-Bellman equation*

$$\max_{1 \leqslant r \leqslant m} (\mathcal{L}^r J^* + h^r) = 0 \qquad \text{for } x \in \Theta \tag{6.14}$$

with boundary condition

$$J_* = 0 \qquad \text{on } \partial\Theta \tag{6.15}$$

where

$$\mathcal{L}^r = \sum_{ij} \frac{1}{2} \sigma_{ij}^r(x)^2 \frac{\partial^2}{\partial x_i x_j} + \sum_i f_i^r(x) \frac{\partial}{\partial x_i} - c,$$

$\sigma = gg^T$, and h^r is the reward per time unit corresponding to regime r.

Unfortunately, to justify the classical derivation of (6.14), (6.15) it is required that $J^* \in C^2(\Theta)$ and this is not true even for rather standard control problems. Fleming and Rishel [32] introduce some kind of generalized solution of the HJB-equation (6.14) for a restricted class of problems. The definitive answer to most of the open questions was only possible after the introduction of the notion of a *viscosity solution* [19]. These are generalized solutions of the HJB-equation which are not required to be smooth, but only have to be continuous in Θ. The minimum amount of regularity this type of solution still allows is that it has locally $L_n(\Theta)$ – bounded second derivatives (i.e. is a member of $W^{2,n}_{loc}(\Theta)$). A classical, smooth solution of (6.14), if it exists, is always a viscosity solution. Furthermore, given a nontrivial boundary condition like (6.15), there is only one viscosity solution that "solves" (6.14) in the generalized sense. The definition and properties of viscosity solutions are given by Lions [58].

Using the notion of a viscosity solution, P.L. Lions developed a fairly

complete theory for general (non-impulse) control problems in a series of papers [58-60]. The problem we consider is a special case of Lions' problem as in our case the control or action space is a finite set.

In short, using the results of Lions the following statements may be proven under fairly weak conditions on the drift functions f^r, the diffusion functions g^r and the instantaneous throughputs h^r. Firstly, the optimal value function is continuous on the domain: $J^* \in C(\Theta)$. Secondly, J^* is a viscosity solution of (6.14). Finally, if $J^* \in W^{2,p}_{loc}(\Theta)$ for some $p > n$ then (6.14) holds almost everywhere on Θ. Note that the second and the third statement are not equivalent, for a viscosity solution only satisfies (6.14) in a generalized sense. So one notes that only a limited amount of regularity is required and the optimal value function satisfies the HJB-equation after all, except maybe at a set of zero measure.

Under somewhat more restrictive conditions Belbas and Lenhart [54] showed that the HJB-equation (6.14) has a unique solution in the Sobolev space of functions with $L_\infty(\Theta)$-bounded derivatives up to second order, $W^{2,\infty}(\Theta)$, and therefore the following theorem may now be stated:

THEOREM 6.5. *Consider the optimal control problem*

$$\sup_u J_u(x_0)$$

where J is defined in (6.12) and x_0 is the initial state of the diffusion defined by (6.11). Assume that f^r, g^r, h^r have L_2-bounded second derivatives in $\bar{\Theta}$, where Θ is an open, bounded domain in $\mathbb{R}^n$ with a smooth boundary $\partial\Theta$. Assume furthermore that $c > 0$ and a $\kappa > 0$ exists such that for all $\xi \in \mathbb{R}^n$ and all $x \in \Theta$

$$\sum_{i,j} \sigma^r_{ij} \xi_i \xi_j \geq \kappa |\xi|^2.$$

Then $J^* = \sup J_u$ is the unique solution in $W^{2,\infty}(\Theta)$ of the Hamilton-Jacobi-Bellman equation

$$\max_{1 \leq r \leq m} (\mathfrak{L}^r J^* + h^r) = 0 \quad \text{a.e. in } \Theta$$

with boundary condition

$$J^* = 0 \quad \text{on } \partial\Theta.$$

PROOF.
Belbas and Lenhart [54] prove that under the assumptions above the HJB-equation with a Dirichlet boundary condition has a unique solution in $W^{2,\infty}(\Theta)$. According to Lions [58] under the conditions of the theorem if $J^* \in W^{2,p}_{loc}(\Theta)$ for some $p > n$, then it satisfies the HJB-equation. Clearly $W^{2,\infty}(\Theta) \subset W^{2,p}_{loc}(\Theta)$ and the theorem follows.

□

For problems of the type stated in the theorem existence, uniqueness and regularity questions are now settled. The problem that is considered in the next sections differs in some aspects from the problem of Theorem 6.5. The diffusion we

consider is defined on an open, bounded but nonsmooth set Θ, where absorption only takes place at a part of $\partial\Theta$ and reflection occurs at the rest. Moreover, the regularity conditions on the drift and throughput function are not satisfied. As Lions [59] points out, however, and as he shows in his papers, his results may be extended to treat more general cases. We do not investigate this further because the non-standard properties of our problem are not essential. Θ was chosen to be a rectangle for ease of use and presentation, but may just as well be replaced by a smooth set. Also, extending the domain a little may allow us to replace the mixed Neumann-Dirichlet boundary condition by a Dirichlet one. Finally, the $L_2(\Theta)-$unboundedness of the second derivatives of drift and throughput functions in our problem is entirely due to the discontinuous derivative of $v^e(\rho)$ at ρ_{crit}, see (6.3). As we explained in Section 6.2.2 we may replace v^e by a smooth one without affecting the essentials of the problem. The nonsmooth v^e was again chosen for ease of use and presentation.

To conclude let us mention that the existence of a stationary optimal control policy corresponding to the optimal value function J^* is not assured in general. In case $n=1$ an optimal control policy may be shown to exist, however, and to consist of switching from one regime to another whenever the diffusion $\{x_t, t\geq 0\}$ crosses one out of several *switching points*. For an illustration of this see the next subsection.

As was mentioned before, Lions [58-60] treats a more general problem than the one we are interested in. Specific results on the finite control space problem are obtained by Mandl [65] and Pliska [78] for the case $n=1$ and general boundary conditions. Results on the n-dimensional Dirichlet problem with only two control regimes ($m=2$, as in our problem), are given by Evans and Friedman [30], and Brezis and Evans [12]. The Dirichlet problem with general n and m is also treated by Belbas and Lenhart [54], Belbas [6] and Lions [57].

6.3.2 Control based on traffic density.

We now turn to the problem of designing a control policy for the one-dimensional traffic model described in 6.2.1.

The Optimization Problem

We take as the criterion to be maximized

$$V_u(\rho_0) = E[\int_0^\tau l\rho_t^u v_u^e(\rho_t^u) - \delta I_{\{u=1\}}\, dt \mid \rho_0] \tag{6.16}$$

where

$$\tau = \inf\{t\geq 0\colon \rho_t = \rho_{jam}\},$$

the time to congestion. Maximization takes place over all stationary, piecewise continuous controls:

$$u : [0,\rho_{jam}] \to \{0,1\}.$$

Here $u(\rho)=1$ means that control is applied when the density equals ρ. Maximizing (6.16) amounts to maximizing the total number of vehicles that pass through the section until congestion occurs. This may be achieved by maximizing the time to congestion and the intensity. Both occur when control is applied and therefore control would always be optimal, if not for the cost δ associated with control. This cost is rather arbitrary, in the final design procedure of Section 6.3.4 it does no longer play a role, however.

Now it may be shown that for each policy $u(.)$ the corresponding criterion function $V_u(\rho)$ satisfies the following ordinary differential equation:

$$\frac{1}{2}\sigma_u^2\frac{d^2V}{d\rho^2}+\frac{1}{Ll}[\lambda_0^u-l\rho v_u^e(\rho)]\frac{dV}{d\rho}+l\rho v_u^e(\rho)-\delta I_{\{u=1\}}=0 \qquad (6.17)$$

with boundary conditions

$$\frac{dV}{d\rho}(0)=0\,,\ V(\rho_{jam})=0.$$

A proof of this may be found in the book on one-dimensional Markov processes of Mandl [65] Chapter VI, Theorem 3. The drift and diffusion are required to be continuous on $[0,\ \rho_{jam}]$ and the diffusion should be strictly positive, for all controls u. Then $V(\rho)$ is shown to be the unique function such that $dV(\rho)/d\rho$ is continuous, and such that (6.17) holds for every $\rho\in(0,\ \rho_{jam})$ which is a continuity point of $u(\rho)$ (i.e. (6.17) holds almost everywhere), and which satisfies the boundary conditions. The latter follow from the character of the boundaries of the diffusion process ρ_t: 0 is reflecting and ρ_{jam} is absorbing.

The optimal criterion function

$$V^*(\rho)=\sup_u V_u(\rho)$$

was shown by P. Mandl to be the unique solution of the *Hamilton-Jacobi-Bellman* or *dynamic programming equation*

$$\frac{d^2V^*}{d\rho^2}+\max_{u=0,1}\left\{\frac{2}{\sigma_u^2}\left[\frac{1}{Ll}[\lambda_0^u-l\rho v_u^e(\rho)]\frac{dV^*}{d\rho}+l\rho v_u^e(\rho)-\delta I_{\{u=1\}}\right]\right\}=0 \qquad (6.18)$$

with the associated boundary conditions. Note that apparently for this one-dimensional problem $V^*\in C^2$. We may use a result from a paper by Pliska [78] to show that an optimal control, maximizing (6.16), exists and is piecewise constant under the condition that drift, diffusion and reward per time unit are piecewise analytic on $[0,\ \rho_{jam}]$. The optimal policy therefore consists of switching the control on or off at several *switching points*. We restrict attention to right continuous policies in the sequel. Even with this restriction a unique optimal control is not guaranteed to exist. The uniqueness depends on the actual model parameter values; this is not investigated here.

In our problem we expect one switching point to be important: when the density approaches the instability point of the model control is expected to become beneficial. This is investigated by means of an example later on, but we first give some analytical results.

PROPOSITION 6.6. *Assume* $\delta=0$. *Then* $\rho=0$ *is a switching point. Applying control is optimal in a neighbourhood of 0 iff*

$$\frac{\sigma_0^2-\sigma_1^2}{\sigma_0^2} > \frac{\Delta v}{v_{free}}.$$

PROOF.
Define

$$F(\rho)=\frac{2}{\sigma_0^2}\left[\frac{1}{Ll}(\lambda_0-l\rho v_0^e)\frac{dV^*}{d\rho}+l\rho v_0^e\right]-\frac{2}{\sigma_1^2}\left[\frac{1}{Ll}(\lambda_0+\Delta\lambda-l\rho v_1^e)\frac{dV^*}{d\rho}+l\rho v_1^e\right].$$

Then $F(0)=0$ because $\frac{dV^*}{d\rho}(0)=0$, hence $\rho=0$ is a switching point. Furthermore,

$$\lim_{\rho\downarrow 0}\frac{dF}{d\rho}(\rho) = \frac{2}{\sigma_0^2}lv_{free} - \frac{2}{\sigma_1^2}l(v_{free}-\Delta v)$$

because $\lim_{\rho\downarrow 0}\frac{d^2V^*}{d\rho^2}(\rho)=0$, following (6.18) and the boundary condition at $\rho=0$. As control is optimal iff $F(\rho)<0$, the condition stated in the proposition follows.

□

PROPOSITION 6.7. *If* $\delta=0$, $\sigma_1^2=\sigma_0^2$ *and* $\Delta\lambda=0$ *then there are three main switching points:* $\rho=0$, $\rho=110$ *and the intersection of the equilibrium speed curves with and without control:*

$$\bar{\rho} = \frac{v_{free}-\Delta v+\frac{d}{\rho_{jam}} - \sqrt{(v_{free}-\Delta v+\frac{d}{\rho_{jam}})^2-4\alpha d}}{2\alpha}.$$

Apart from these there may be switching points for values of ρ *for which* $dV^*(\rho)/d\rho=Ll$ *holds.*

If there are only three switching points, it is optimal not to apply control until ρ *exceeds* $\bar{\rho}$. *If* $\lambda_0<l\bar{\rho}v^e(\bar{\rho})$ *the switching point* $\bar{\rho}$ *decreases with decreasing* σ_1^2.
PROOF.
A simple calculation shows that $F(\rho)$, defined in the proof of Proposition 6.6, equals

$$F(\rho) = \frac{2}{\sigma_0^2}\left[\frac{1}{Ll}\frac{dV^*}{d\rho}-1\right][l\rho v_1^e(\rho)-l\rho v_0^e(\rho)]$$

and $F(\rho)=0$ iff $\frac{dV^*}{d\rho}(\rho)=Ll$ or $\rho=0$ or $v_0^e(\rho)=v_1^e(\rho)$. The latter condition leads to $\rho=\rho_{jam}$ or the value $\bar{\rho}$ given in the statement above. Assume that $dV^*(\rho)/d\rho<0$. Then $F(\rho)>0$ if $\rho<\bar{\rho}$, hence it is optimal not to apply control

until ρ exceeds this value. Next

$$\left.\frac{dF}{d\sigma_1^2}\right]_{\substack{\sigma_1^2=\sigma_0^2\\ \rho=\bar{\rho}}} = \frac{4}{\sigma_1^4}\left[\frac{1}{Ll}[\lambda_0 - l\bar{\rho}v_1^e(\bar{\rho})]\frac{dV^*}{d\rho} + l\bar{\rho}v_1^e(\bar{\rho})\right]$$

If $\lambda_0 < l\bar{\rho}v_1^e(\bar{\rho})$ and $\frac{dV^*}{d\rho} < 0$ then $\left.\frac{dF}{d\sigma_1^2}\right]_{\substack{\sigma_1^2=\sigma_0^2\\ \rho=\bar{\rho}}} > 0$, and F decreases with decreasing σ_1^2. This implies that control is optimal for $\sigma_1^2 < \sigma_0^2$ in a neighbourhood of σ_0^2.

□

Note that in general $\frac{dV^*}{d\rho}(\rho) < 0$ (starting closer to the unstable equilibrium point increases the probability of congestion), excluding the possibility for $\frac{dV^*}{d\rho}(\rho) = Ll$.

Proposition 6.6 tells us that, under the stated condition, in absence of a control cost it is optimal to apply control, even at low intensity values. For our parameters the condition is satisfied. Computations with $\delta = 0$ have shown that for $\lambda_0 = 2000$ control is optimal for all values of ρ.

Proposition 6.7 shows that there is one important switching point in general, expressing the benefit of control when there is a significant probability of congestion.

Numerical Solution

We now turn to the numerical solution of the equation (6.18). Solving this type of equations is a difficult problem in general. For our relatively simple problem no analytical solution exists, but solving (6.18) numerically poses no problem. If we define $W(\rho)$ as the solution with boundary conditions

$$\frac{dW}{d\rho}(0) = 0\,,\; W(0) = 0$$

then the optimal solution follows from

$$V^*(\rho) = W(\rho) - W(\rho_{jam})$$

Now we may integrate (6.18) from $\rho = 0$ to $\rho = \rho_{jam}$ for W, at every integration step checking whether the term between curly brackets is maximal for $u = 0$ or for $u = 1$. In this way, choosing an integration step that is small enough, we may find the optimal control and criterion value by integrating (6.18) only once! In general some iterative procedure is needed in which optimization and integration steps are carried out alternately, but for the problem considered here this turns out to be unnecessary.

EXAMPLE. For $\lambda_0 = 4600$, $\Delta\lambda = 46$, $\Delta v = 3$, $\Delta\rho = 2$, $\delta = 100$, $\sigma_0^2 = 14000$ and

$\sigma_1^2 = 11000$ we find that the optimal switching points are 27.1 and 48.8. No control is applied for densities below 27.1 veh/km/lane and above 48.8 veh/km/lane.

The major switching point in the example is 27.1. The fact that the control is to be switched off at 48.8 is not interesting. The occurrence of the point at 48.8 may be explained as follows. Once density has exceeded the unstable equilibrium value, congestion occurs, unless a disturbance brings the density back to a value below ρ_0^u. This is more likely to occur if the variance of the disturbances is larger, i.e. when no control is applied. Our main interest is in avoiding congestion, however, and for values far above ρ_{crit} the model may not be realistic.

One-Switch Controls

For other parameter values sometimes more switching points show up. In these cases we always consider the suboptimal policy in which less important switching points are omitted. Consider the previous example. Instead of the optimal policy, in practice the policy of switching the control on for densities above or equal to 27 and off below 27 is applied (we have assumed that $l = 2$ and $L = 0.5$, so ρ will be an integer in practice). Table 6.4, in which the criterion function for the optimal and the suboptimal policy are compared, shows that this policy is nearly optimal.

ρ (veh/km/lane)	$V^*(\rho)$ (veh)	$V^{\bar{\rho}=27}(\rho)$ (veh)
0	397.8	395.8
10	395.8	393.8
20	384.1	382.1
30	337.9	336.0
40	205.6	203.6
50	87.7	85.7
110	0.0	0.0

TABLE 6.4. Comparison of optimal and one-switch controls.

This approach amounts to confining attention to a restricted class of controls: those that correspond to exactly one switching point.

DEFINITION 6.8. *A* one-switch control policy *is defined to be a policy of the form*

$$u(\rho) = I_{\{\rho \geq \bar{\rho}\}},$$

for some $\bar{\rho} \in [0, \rho_{jam}]$. *This means that control is applied only for* $\rho \geqslant \bar{\rho}$.

The value $\bar{\rho}$ is determined by computing the optimal policy first and selecting the main switching point next. These one-switch controls are suboptimal but based

on earlier facts and the example we expect them to be close to optimal.

Up until now an arbitrary cost δ has been assigned when control is applied. It is not clear what value should be chosen for δ as it cannot be interpreted in terms of traffic characteristics. One might use it as a design variable and experiment with different values. The results for $\delta = 100$ and $\delta = 500$ are presented in Table 6.5. Instead of investigating the sensitivity with respect to δ in detail, we propose a design procedure in the next two subsections that avoids introducing such an arbitrary cost.

One notes that when λ_0 is low a problem turns up: the optimal switching point $\bar{\rho}$ decreases with decreasing λ_0. One would expect the opposite to occur, as for lower λ_0 the unstable equilibrium point is higher, thereby allowing large values of the density before congestion occurs. For $\lambda_0 = 2000$ veh/h e.g. the optimal switching point turns out to be 13 veh/km/lane whereas the unstable equilibrium point is about 75.6 veh/km/lane. The stable equilibrium point is 10.4 veh/km/lane which implies that control will often be applied: each time an excursion above 13 takes place. Note that the standard deviation of ρ is about 7 veh/km/lane.

The main reason for this effect is that we have considered an *infinite horizon* criterion:

$$E[\int_0^{\tau} l\rho_t v_t dt \mid \rho_0] = E[\int_0^{\infty} l\rho_t v_t dt \mid \rho_0]$$

For low values of λ_0 the expectation of τ is large, for $\lambda_0 = 2000$ it is about 4 years without and 380 years with control. Clearly, postponing congestion from 4 to 380 years is not very interesting. In practice the probability of congestion is negligible in this case, even without control. Two ways to approach this problem are considered in the following subsections: a finite time horizon may be introduced, or the criterion may be discounted.

$\bar{\rho}$	λ_0 (veh/h)					
(veh/km/lane)	1000	2000	3000	3500	4000	4800
$\delta = 100$	3	5	9	14	22	27
$\delta = 500$	9	13	19	22	26	28

TABLE 6.5. One-switch controls for different values of the control cost.

6.3.3. Finite horizon and discounting.

In the previous section is is concluded that to obtain realistic control policies two different procedures may be followed. The first is to use a finite horizon in the design criterion, the second is to insert a discount factor. Both possibilities are discussed in this section.

Finite Horizon Case

To avoid the problem mentioned in the previous subsection we now consider

$$V^{fin}(\rho_0,t) = \sup_u E[\int_t^{\min(\tau,T)} l\rho_s^u v_u^e(\rho_s^u)ds \mid \rho_0]$$

where

T : time horizon (h)

and τ is the time to congestion, introduced in Section 6.3.2. We take $T=2$ h in the sequel and are interested in $V^{fin}(\rho,0)$. For example, in case $\lambda_0=2000$ veh/h postponing congestion from 4 to 380 years no longer contributes to the criterion value and switching at a low density value is no longer beneficial. The optimal value function $V^{fin}(\rho,t)$ associated with the diffusion process (6.5) and the optimal control policy $u(.)$ satisfy the Hamilton-Jacobi-Bellman equation

$$\frac{\partial V^{fin}}{\partial t} + \max_{u=0,1}\left\{\frac{1}{2}\sigma_u^2\frac{\partial^2 V^{fin}}{\partial\rho^2} + \frac{1}{Ll}[\lambda_0^u - l\rho v_u^e(\rho)]\frac{\partial V^{fin}}{\partial\rho} + l\rho v_u^e(\rho)\right\}=0 \qquad (6.19)$$

The boundary conditions are

$$\frac{\partial V^{fin}}{\partial\rho}(0,t)=0,\quad V^{fin}(\rho_{jam},t)=0,\quad V^{fin}(\rho,T)=0,$$

where the third condition follows from the absence of terminal costs in the criterion.

Solving (6.19) is considerably more difficult than (6.18). It is not easy to carry out the simultaneous maximization and integration of the parabolic partial differential equation. A method that seems to give good results is the approach of Kushner [53]. There the diffusion process is approximated by a finite state Markov chain and then policy or value iteration algorithms are applied. This is not attempted here but is left for future investigations.

We restrict attention to the computation of the criterion value for several elements of the class of one-switch controls introduced in 6.3.2. This avoids the maximization problem in (6.19) and leaves us with the partial differential equation:

$$\frac{\partial V^{fin}}{\partial t} + \frac{1}{2}\sigma_u^2\frac{\partial^2 V^{fin}}{\partial\rho^2} + \frac{1}{Ll}[\lambda_0^u - l\rho v_u^e(\rho)]\frac{\partial V^{fin}}{\partial\rho} + l\rho v_u^e(\rho) = 0$$

where $u(\rho)=I_{\{\rho\geq\bar{\rho}\}}$ and $\bar{\rho}$ is given. This parabolic equation is solved by means of the routine D03PAF from the NAG-library [72]. Starting with the one-switch control corresponding to $\bar{\rho}=0$ we increase $\bar{\rho}$ as long as no deterioration of the criterion function occurs. To be precise: $\bar{\rho}$ is maximized under the condition that for all $\rho\in[0,\rho_{jam}]$ no deterioration of more than 5% should occur in $V^{fin}(\rho,0)$. This may be interpreted as postponing control as long as no significant benefit can be expected from it. There is no longer need for an artificial cost associated with control, and the δ term is omitted. The other parameters are chosen as before.

Results of computations of the optimal $\bar{\rho}$ for several intensity levels are presented in Table 6.6. The values of $\bar{\rho}$ given there are not very precise. An uncertainty of several veh/km/lane should be taken into account, due to the

numerical errors in the solution of the partial differential equation and the fact that V^{fin} is rather insensitive to $\bar{\rho}$ near the optimal value.

λ_0 (veh/h)	1000	2000	3000	3500	4000	4400	4800
$\bar{\rho}$ (veh/km/lane)	96	75	51	29	29	30	31

TABLE 6.6. Optimal one-switch controls: finite horizon case.

Note that now for small λ_0 control will hardly ever be applied, unless something exceptional occurs, like an accident, leading to extreme density values. Note also that for intensities above 3500 veh/h one and the same control policy suffices. In this case the criterion function turns out to be very "flat", insensitive to $\bar{\rho}$, expressing the fact that only a small benefit may be expected from control when intensity is high. This may also explain the increase in $\bar{\rho}$ for $\lambda_0 > 4000$.

Simulation

To illustrate the effect of optimal control we will now present the results of a simulation of the model. The original parameter values are chosen and $\lambda_0 = 4000$ veh/h. The optimal control appears to be $I_{\{\rho \geq 29\}}$. Figure 6.8 shows a realization of ρ_t and Figure 6.9 the corresponding realization of $v^e(\rho_t)$ for the case with and without control. The positive effect is clear: without control congestion occurs at 0.35 h and with control congestion does not occur during the period considered. The respective criterion values are 1320 and 1997 veh. The simulation result shown here is typical in the sense that in most realizations when density is critical, congestion may occur, even when control is applied, though the probability is smaller with than without control. Once congestion is avoided it takes some time before another critical situation turns up, however. This means that control may lead to a relatively large benefit, but that the probability of incurring this benefit is relatively small.

A negative point in the implementation is the frequent switching that takes place: over 600 times in 30 minutes! It is clear that this is not acceptable from a practical point of view, as frequent switching would confuse drivers and might even be dangerous. This problem is addressed in Section 6.3.4.

Discounted Case

Another way to avoid unnecessary control at low intensity values is to give a smaller weight to the benefit as it is obtained further away in the future. To achieve this consider the *discounted* criterion

$$V^{dis}(\rho_0) = \sup_u E[\int_0^\tau e^{-ct} l \rho_t^u v_u^e(\rho_t^u) dt \mid \rho_0] \tag{6.20}$$

where

c : discount factor, ≥ 0.

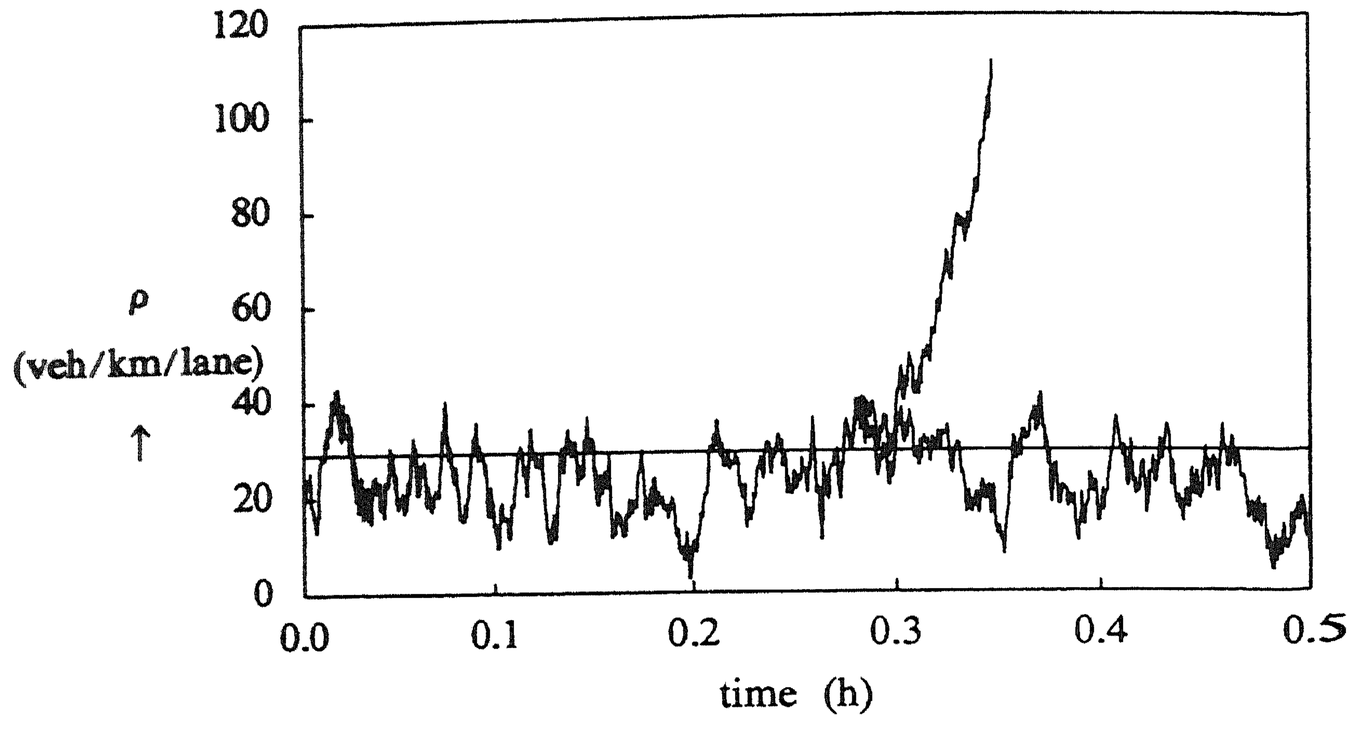

FIGURE 6.8. Model simulation with and without control: traffic density.

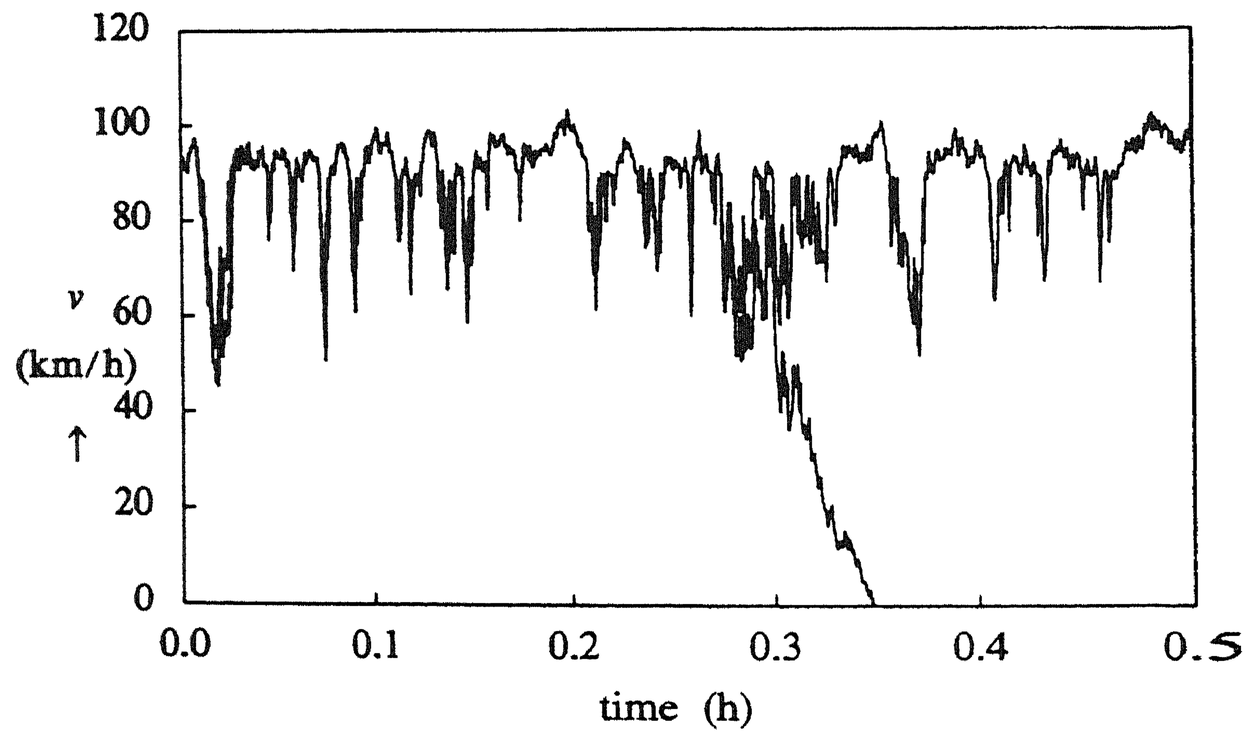

FIGURE 6.9. Model simulation with and without control: mean speed.

We take $c = 0.5$ in order to achieve results that may be compared to the finite horizon ones, because:

$$\int_0^\infty e^{-0.5t} dt = T = 2 \text{ h}$$

in this case. The Hamilton-Jacobi-Bellman equation associated with (6.20) equals [78]

$$\frac{d^2 V^{dis}}{d\rho^2} + \max_{u=0,1} \left\{ \frac{2}{\sigma_u^2} \left[\frac{1}{Ll} [\lambda_0^u - l\rho v_u^e(\rho)] \frac{dV^{dis}}{d\rho} - cV^{dis} + l\rho v_u^e(\rho) \right] \right\} = 0 \qquad (6.21)$$

with boundary conditions

$$\frac{dV^{dis}}{d\rho}(0) = 0, \quad V^{dis}(\rho_{jam}) = 0.$$

Unfortunately, due to the term $cV^{dis}(\rho)$ the trick of turning the boundary value problem into an initial value problem as used in 6.3.2 does not apply here. We therefore follow the approach of the previous subsection and restrict attention to computing $V^{dis}(\rho)$ for several one-switch control policies. The boundary value problem is solved by means of the routine D02HAF of the NAG-library [72]. The results are presented in Table 6.7. Computing the optimal values of $\bar{\rho}$ sometimes turned out to be more involved than for the finite horizon criterion.

λ_0 (veh/h)	2000	3000	3500	4000	4400	4800
$\bar{\rho}$ (veh/km/lane)	70	42	28	28	29	31

TABLE 6.7. Optimal one-switch controls: discounted case.

We see the same behaviour qualitatively as for the finite horizon criterion.

6.3.4. Hysteresis control.
The simulation results in the previous subsection show that for the optimal control policy a large number of switchings may occur in a short time period. This is an annoying feature and therefore a modified policy is proposed here.

The Optimization Problem

One way to reduce the number of switchings would consist of introducing a *switching cost* in the criterion function. The corresponding control problem turns out to be quite different from the one without switching costs, and is studied extensively in the literature. Consider again the general control problem presented in section 6.1 of controlling a diffusion $\{x_t, t \geqslant 0\}$ by switching among m different regimes. The reward function considered there is now replaced with

$$J_u(x_0, r_0) = E\left[\int_0^\tau e^{-ct} h(x_t, u_t) dt - \sum_{k=1}^{\infty} e^{-c\theta_k} p(r_{k-1}, r_k) \mid x_0, r_0\right]$$

where $p(r_{k-1},r_k) \geqslant 0$ is the switching cost corresponding to a switch from regime r_{k-1} to regime r_k. It is assumed that $p(r,s)=0$ if and only if $r=s$. The problem may be formulated as a problem of *impulse control* by introducing an auxiliary state variable $\{y_t,\ t \geqslant 0\}$ and rewriting the diffusion model as follows:

$$dx_t = \bar{f}(x_t,y_t)dt + \bar{g}(x_t,y_t)dw_t$$

$$dy_t = \sum_{k=1}^{\infty} \{r_k - r_{k-1}\}\delta(t-\theta_k)dt$$

where

$$\bar{f}(x,r)=f^r(x), \quad \bar{g}(x,r)=g^r(x)$$

and δ_r is an impulse of size r. The state variable y represents the current regime. Besides an initial value for the diffusion state x the initial regime r has to be specified also: $y_0=r_0$.

General impulse control problems were first studied by Bensoussan and Lions [7]. An extensive treatment is given in [8]. Our problem is a specific case in that the size of the impulses is restricted to come from a finite set, one of the states is not driven by noise and that only one state is directly affected by the impulses. This case is treated as an example in Section 1.4 of Chapter 1 in [8].

Application of the dynamic programming principle to impulse control problems does not lead to a Hamilton-Jacobi-Bellman equation like (6.14), but to the Quasi Variational Inequalities (QVI's):

$$\mathfrak{L}^r J_r^* \leqslant h^r$$

$$J_r^* \geqslant \max_{1 \leqslant s \leqslant m} \{J_s^* - p(r,s)\}$$

$$(\mathfrak{L}^r J_r^* + h^r)(J_r^* + \max_{1 \leqslant s \leqslant m} \{J_s^* - p(r,s)\}) = 0$$

for $r,s \in \{1,...,m\}$, on Θ with the boundary condition

$$J_r^* = 0$$

on $\partial\Theta$ for the Dirichlet problem. Note that J_r^*, the optimal reward function, now not only depends on the initial diffusion state, but also on the initial regime: starting at a non-optimal regime immediately leads to a switch and incurring of the associated cost, and therefore to a different reward function. If the switching costs are all zero the dependence on the initial regime disappears, as we saw in Section 6.1, and in that case the QVI's reduce to the HJB-equation (6.14).

The QVI's may be interpreted as a set of coupled Variational Inequalities, which occur in optimal stopping problems, and the impulse control problem may be interpreted as a series of stopping problems.

As for the non-impulse control problem, literature is available on the specific case that we are interested in. Results on uniqueness, existence and regularity of solutions of QVI's are given by Belbas and Lenhart [54], Evans and Friedman [30], and Lions [60]. References to literature treating the problem directly, without introducing the impulse control, are [15, 26]. The latter two references

also treat the existence question for stationary optimal controls, and compute examples for the one-dimensional case. As it turns out, introducing switching costs leads to a *hysteresis* type of control: a switching point as it occurs in absence of switching costs splits up into two points when these costs are introduced. The optimal control then consists of switching to one decision when the larger of the two points is crossed in upward direction, and returning to the original decision only after crossing the smaller point in the downward direction. See Figure 6.10.

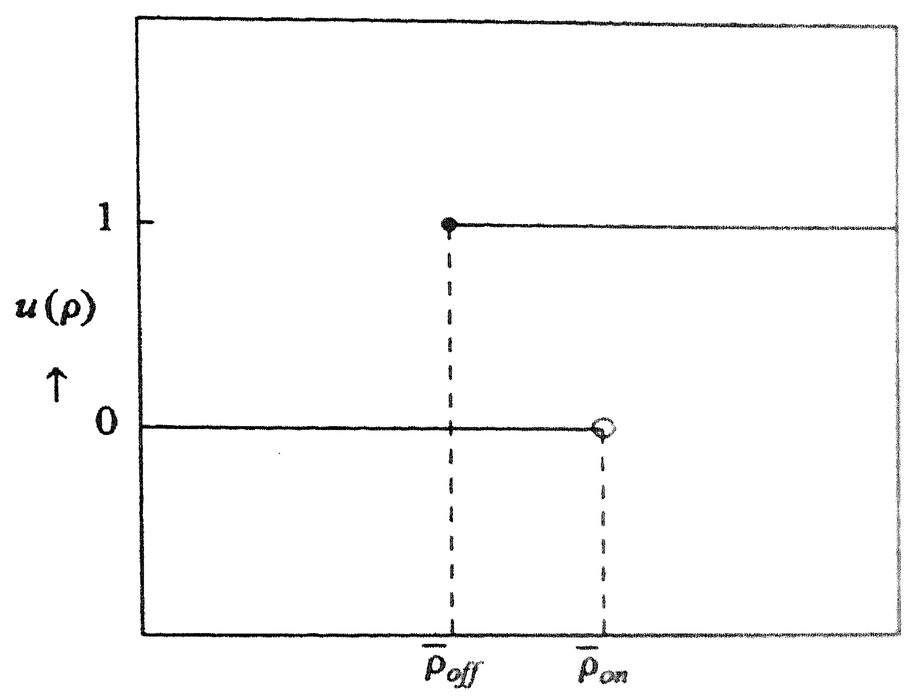

FIGURE 6.10. The hysteresis control policy.

To describe this there is need for an auxiliary state variable that memorizes the state of control: is control being applied at the moment or not? Denote this variable by s. Then the following sets may be introduced:

$$C_0 = \{(\rho,s): \rho\in[0,\bar{\rho}_{on}),\ s=0\} \tag{6.22}$$

$$C_1 = \{(\rho,s): \rho\in[\bar{\rho}_{off},\rho_{jam}],\ s=1\} \tag{6.23}$$

$$S_{01} = \{(\rho,s): \rho\in[\bar{\rho}_{on},\rho_{jam}],\ s=0\} \tag{6.24}$$

$$S_{10} = \{(\rho,s): \rho\in[0,\bar{\rho}_{off}],\ s=1\} \tag{6.25}$$

C_0 and C_1 are called the *continuation sets* and S_{01} and S_{10} the *switching sets*. The hysteresis policy is now completely determined by the continuation and switching sets. As long as $(\rho,s) \in C_0$ control is not applied, so $u(\rho)=0$. As long as $(\rho,s) \in C_1$ control is applied, so $u(\rho)=1$. As soon as (ρ,s) enters S_{01}, $u(\rho)$ switches from 0 to 1 and as soon as (ρ,s) enters S_{10}, $u(\rho)$ switches from 1 to 0. The introduction of the hysteresis clearly enables a reduction of the frequency of switching.

The assignment of costs to switching is ad hoc in our problem, and it is not clear how to choose them or interpret them. Moreover, the policy iteration procedure proposed by Doshi [26] is quite complicated except for linear models. We would like to satisfy a condition like not switching off the system before m minutes after switching it on have gone by. To achieve this one might experiment with several values for the costs. We follow a more direct approach instead.

We consider the hysteresis type of control of Figure 6.10 and design it according to the following procedure:

i) we compute $\bar{\rho}_{on}$ according to the original criterion and procedure of Section 6.3.3. If $\bar{\rho}_{on} = \rho_{jam}$ take $u(\rho) \equiv 0$, otherwise proceed to *ii*);

ii) $\bar{\rho}_{off}$ is taken to be the solution of the problem

$$\max_{0 \leq \bar{\rho}_{off} \leq \bar{\rho}_{on}} \bar{\rho}_{off}$$

under the condition

$$P(\theta \leq m) \leq 0.05$$

where m is a given time duration and θ *the time to the first occurrence of an* on-off-on *cycle*:

$$\theta = \inf\{t \geq 0:\ \rho_t = \bar{\rho}_{on} \wedge \rho_s = \bar{\rho}_{off},\ \textit{for some } s,\ 0 \leq s \leq t \mid \rho_0 = \bar{\rho}_{on}\}$$

If no $\bar{\rho}_{off}$ exists that satisfies the condition $\bar{\rho}_{off}$ is taken to be zero.

The motivation for this procedure is given next.

The design procedure is split in two parts to assure that control is applied as soon as it is necessary: $\bar{\rho}_{on}$ is determined by maximizing the original criterion, neglecting the undesired effect of frequent switching. The frequency of switching is next reduced by postponing switching off the control. The criterion for the determination of $\bar{\rho}_{off}$ is to be interpreted as limiting the occurrences of on-off-on cycles that take less than, or precisely m time units. With a probability of 95% there is at most 1 such a cycle per m time units. This clearly reduces the number of switchings. It does not exclude the possibility of a short on-off or off-on sequence, however. It is allowed to switch the control off (on) after it has recently been switched on (off) as long as one may expect the next switch to be necessary only much later.

Working with the criterion of reducing on-off sequences instead of on-off-on sequences was considered in preliminary investigations, but turned out to give an unsatisfactory control policy in case λ_0 is small. If the intensity is small the stable equilibrium point ρ_0^s is far from $\bar{\rho}_{on}$ (see Table 6.6) and control is only applied in exceptional circumstances: a large excursion of ρ_t. Once ρ_t is that large either convergence to ρ_{jam} occurs or the traffic stream recovers: ρ will

fluctuate around the stable equilibrium point again with a negligible probability of returning to $\bar{\rho}_{on}$. The recovery, if it takes place, occurs within a short time, far shorter than m will usually be chosen. Imposing the restriction of the procedure proposed above on the cycle on-off here would lead to the policy of maintaining control once a large excursion has occurred that did not lead to a traffic jam. This is clearly unrealistic. Taking into account the small probability of returning to $\bar{\rho}_{on}$ as in our proposed approach does not give this unrealistic result.

Considering off-on cycles instead of on-off-on cycles leads to equivalent results when λ_0 is low, but for $\lambda_0 > 3500$ veh/h one obtains $\bar{\rho}_{off} = 0$. Once an excursion has occurred control is always applied. This is not unrealistic, but the results of Table 6.8 might be preferable in that a possible change in λ_0 is anticipated better.

The problem to be solved in *ii*) above is well-defined, as is shown in the following theorem:

THEOREM 6.9. *Let the process* $\{\rho_t, t \geq 0\}$ *on* $[0, \rho_{jam}]$ *be defined by (6.5) with 0 reflecting and* ρ_{jam} *absorbing and let u(.) be the hysteresis control policy corresponding to the continuation and switching sets (6.22) to (6.25), with* $\bar{\rho}_{on} \in [0, \rho_{jam})$ *and let* $m > 0$. *Assume also that* $\sigma_0, \sigma_1 > 0$. *Then either* $P(\theta \leq m) > 0.05$ *for all* $\bar{\rho}_{off} \in [0, \bar{\rho}_{on}]$, *or a unique* $\bar{\rho}_{off}$, *satisfying the condition* $P(\theta \leq m) \leq 0.05$, *exists.*

PROOF.

From the pathwise continuity of $\{\rho_t, t \geq 0\}$ it follows that for $0 < \bar{\rho}_{off} \leq \bar{\rho}_{on}$, $0 < \Delta \leq \bar{\rho}_{off}$:

$$A_\Delta(m) \subset A_0(m)$$

where

$$A_\Delta(m) = \{\rho_t = \bar{\rho}_{on}, \rho_s = \bar{\rho}_{off} - \Delta, \textit{for some } s \textit{ and } t \textit{ with } 0 \leq s \leq t \leq m \mid \rho_0 = \bar{\rho}_{on}\}$$

$$A_0(m) = \{\rho_t = \bar{\rho}_{on}, \rho_s = \bar{\rho}_{off}, \textit{for some } s \textit{ and } t \textit{ with } 0 \leq s \leq t \leq m \mid \rho_0 = \bar{\rho}_{on}\}$$

i.e., A_0 is the set of realizations, starting at $\bar{\rho}_{on}$, that touch on $\bar{\rho}_{off}$ at least once and return to $\bar{\rho}_{on}$ within the time interval $[0,m]$, and A_Δ the analogue for $\bar{\rho}_{off} - \Delta$. Here the inclusion is proper because the sets are nonempty and $\Delta > 0$. So, the probability of a cycle $(\bar{\rho}_{on}) - (\bar{\rho}_{off} - \Delta) - (\bar{\rho}_{on})$ occurring within the interval $[0,m]$ is strictly smaller than the probability of a cycle $(\bar{\rho}_{on}) - (\bar{\rho}_{off}) - (\bar{\rho}_{on})$ occurring in the same interval. Note that, because of the pathwise continuity and the Markov property of $\{\rho_t, t \geq 0\}$, the time associated with the cycle $(\bar{\rho}_{on}) - (\bar{\rho}_{off} - \Delta) - (\bar{\rho}_{on})$ is larger than the time associated with the cycle $(\bar{\rho}_{on}) - (\bar{\rho}_{off}) - (\bar{\rho}_{on})$. The conclusion therefore is that $P(\theta \leq m)$ decreases strictly and monotonically with decreasing $\bar{\rho}_{off}$. As this probability is continuous in $\bar{\rho}_{off}$ and for $\bar{\rho}_{off} = \bar{\rho}_{on}$ it equals 1, it is clear that either for a unique $\bar{\rho}_{off} > 0$ $P(\theta \leq m) \leq 0.05$ holds, or for $\bar{\rho}_{off} = 0$ $P(\theta \leq m) > 0.05$. □

Analytically solving $\bar{\rho}_{off}$ is not possible. Computing first passage times requires applying (inverse) Laplace transforms on solutions of an ordinary differential equation [3]. In our case this equation does not permit an analytic solution. Instead of approaching the problem numerically we decided to estimate the switch-off density by means of simulation. A value for $\bar{\rho}_{off}$ is guessed and (6.5) is simulated, starting in $\bar{\rho}_{on}$, until either an on-off-on cycle or convergence to ρ_{jam} has occurred. Out of a large set of realizations (250-500) generated in this way the fraction in which a cycle occurred within m time units is computed. This is an estimate of $P(\theta \leqslant m)$. If this fraction is smaller than 0.05 $\bar{\rho}_{off}$ is increased, otherwise it is decreased. If the fraction is close to 0.05 the procedure is finished. In practice ρ is discrete so a finite series of attempts will lead to the right $\bar{\rho}_{off}$.

The results of the simulations are presented in Table 6.8, the values of $\bar{\rho}_{on}$ are the ones obtained in 6.3.3 with the finite horizon criterion.

λ_0 (veh/h)	1000	2000	3000	3500	4000	4400	4800
$\bar{\rho}_{off}$ (veh/km/lane)	77	56	36	2	5	8	11
$\bar{\rho}_{on}$ (veh/km/lane)	96	75	51	29	29	30	31

TABLE 6.8. Optimal hysteresis-control policies.

In the simulation experiment presented before ρ_t exceeded 29 ($=\bar{\rho}_{on}$) at t=0.01 h and only became less than 5 ($=\bar{\rho}_{off}$) for a period of 36 seconds near t=0.02 h. This means that the number of switchings is reduced from over 600 to 3. The criterion value is hardly different when we apply the hysteresis control instead of the one-switch control: 2005 instead of 1997 veh.

An alternative to the above procedure would consist of using a *timer*, which e.g. prevents switching off control for a period of m minutes from the moment is switched on. This procedure has the slight disadvantage that in case the density does decrease fast and becomes low, within m minutes, the system is not switched off. This is typical of an open-loop type of control like using a timer, and does not occur in the closed-loop approach proposed by us.

6.3.5. Control based on both traffic density and mean speed.

In the previous subsections some optimal and suboptimal control policies were computed using various performance criteria, based on a model for traffic density only. As discussed in Section 6.2, a model that accounts for the delayed reaction of speed to density fluctuations is more realistic: (6.7), (6.8). Such a model allows larger excursions of the density without convergence to ρ_{jam} occurring. This results in a more realistic control policy, now depending on both ρ and v. The phase plane is divided in two regions and control is switched on or off when the speed-density pair crosses the boundary between the two regions. The approximate form of this boundary may be deduced from the separators of Figure 6.6. In Figure 6.11 the separator for $\lambda_0 = 4000$ veh/h is plotted as well as a guess for the switching curve (dashed). Control is to be applied if (ρ, v) is below or on the curve. The vertical dotted line represents the switching curve corresponding to the policy based on density alone.

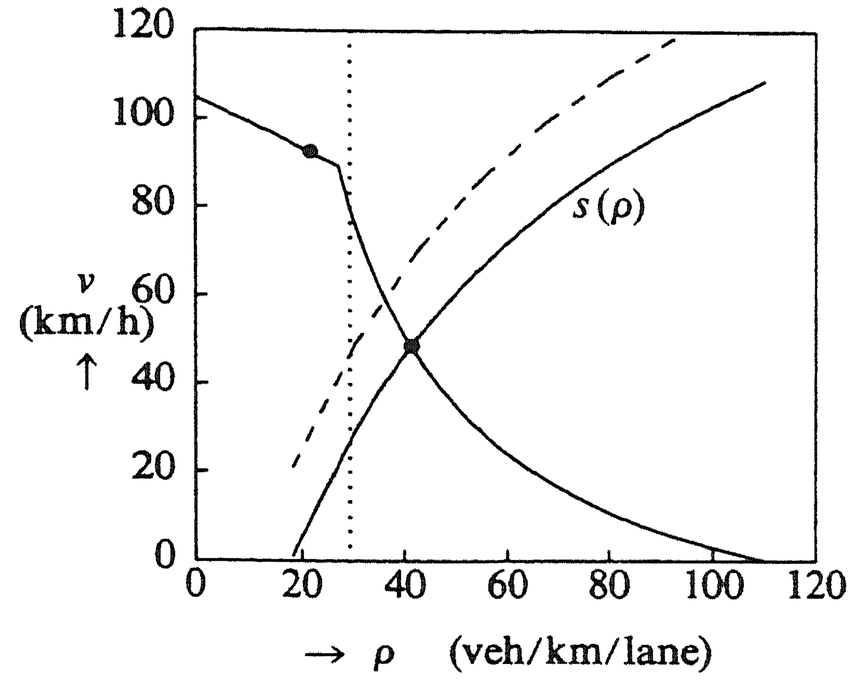

FIGURE 6.11. Comparison of one- and two-dimensional control.

For the one-dimensional model and the infinite horizon criterion (6.16) we have been able to compute the exact optimal switching policy $u(\rho)$ numerically. For the more realistic finite horizon or discounted criterion this turned out to be much more involved and we only computed suboptimal policies. Even the infinite horizon problem for the two-dimensional model leads to an elliptic partial differential equation and poses the same problems as in 6.3.3. A promising approach to solving this type of problem is the approximation procedure justified by Kushner [53], but this is not investigated here. We refer to Schuit [85] for some interesting results obtained with this approximation approach. We only present a limited illustration of the idea expressed in Figure 6.11 and restrict ourselves to computing a suboptimal solution for the discounted criterion (6.20). The domain of the two-dimensional diffusion process is restricted to $[0,\rho_{jam}] \times [0,v_{max}]$ like in Section 6.2.2. We choose $v_{max} \gg v_{free}$ in order to achieve a good approximation of the unrestricted process. Note that v does generally not deviate much from $v^e(\rho)$ and that therefore the probability that $v > v_{free}$ is small. The boundaries $v = v_{max}$, $v = 0$, $\rho = 0$ are chosen to be reflecting, and of $\rho = \rho_{jam}$ the part where $v \leqslant \lambda_0 / l\rho_{jam}$ is absorbing and the rest reflecting just as in 6.2.2.

Optimization of control policies corresponding to a single curve in the (ρ, v) plane, like the one in Figure 6.11 still is too involved. We therefore only compare the criterion corresponding to some simple policies.

The criterion value function $V^{dis}(\rho, v)$ corresponding to such policies satisfies the following equation:

$$\frac{1}{2}\sigma_u^2 \frac{\partial^2 V^{dis}}{\partial \rho^2} + \frac{1}{2}\mu^2 \frac{\partial^2 V^{dis}}{\partial v^2} + \frac{1}{Ll}[\lambda_0^u - l\rho v]\frac{\partial V^{dis}}{\partial \rho} + \frac{1}{T}[v_u^e(\rho) - v]\frac{\partial V^{dis}}{\partial v}$$

$$-cV^{dis} + l\rho v = 0 \tag{6.26}$$

with boundary conditions

$$\frac{\partial V^{dis}}{\partial \rho}(0,v)=0,\ \frac{\partial V^{dis}}{\partial \rho}(\rho_{jam},v)=0 \text{ for } v>\lambda_0/l\rho_{jam},\ \frac{\partial V^{dis}}{\partial v}(\rho,0)=0,$$

$$\frac{\partial V^{dis}}{\partial v}(\rho,v_{max})=0,\ V^{dis}(\rho_{jam},v)=0 \text{ for } v\leqslant\lambda_0/l\rho_{jam}.$$

The equation (6.26) was solved using the solver of De Zeeuw [104] for $\lambda_0=4200$ veh/h, $c=0.5$ and the parameter values of Section 6.2.2.

To start with we take policies that only use the value of the density for control: $u(\rho)=I_{\{\rho\geqslant\bar{\rho}\}}$, and vary $\bar{\rho}$. As it turns out, the optimal criterion function is achieved by taking $\bar{\rho}=28$. See Table 6.9 for the criterion value in $(\rho,v)=(20,90)\approx(\rho_0^s,v^e(\rho_0^e))$.

$\bar{\rho}$ (veh/km/lane)	0	26	27	28	29	30	110
$V^{dis}(20,90)$	3444	4017	4063	4096	4058	3912	2200

TABLE 6.9. Comparison of the criterion value for different controls.

The optimal value corresponds to the one obtained using the one-dimensional model.

It is of interest to investigate whether postponing control is allowed without serious degradation of the criterion function, when v is large enough. This is motivated by the form of separator of Figure 6.11. For this we tried the following policy which consists of postponing control when $v>v_{crit}$:

$$u(\rho)=\begin{cases}0 \text{ if } \rho\leqslant 28 \wedge v\leqslant v_{crit}\\ \quad \text{or } v>v_{crit} \wedge v>1.66\rho+v_{crit}-46.5\\ 1 \text{ elsewhere}\end{cases}$$

The boundary between the domain of no control and the domain of control consists for this type of policy of a vertical line at $\rho=28$ veh/km/lane for $v=0$ km/h to v_{crit} km/h, and at the latter point it continues as a straight line with coefficient of direction 1.66. This is the coefficient of direction of the separator at the unstable equilibrium point.

The result for several values of v_{crit} is given in Table 6.10.

v_{crit} (km/h)	85	90	95	100
$V^{dis}(20,90)$	3947	4037	4080	4093

TABLE 6.10. Comparison of the criterion value for different controls.

We may conclude that indeed postponing control is allowed for large v-values and in that sense the latter control policy is an improvement over the ones which only use the density information. Further experiments need to be carried out to find even better policies. This is not investigated here.

To reduce the number of switchings the one-switch control policy would again have to be replaced by a hysteresis-type of policy as in 6.3.4.

6.4. Conclusions

In this chapter a freeway traffic control problem has been addressed, motivated by the Dutch Motorway Control and Signalling System. This system allows the display of advisory speed signals to drivers and is aimed to improve traffic flow and safety. Speed measurements and vehicle counts are available at several locations along the freeway. Applying the filter algorithm developed in Chapter 5 makes estimates of traffic density and mean speed of a series of freeway sections available. Based upon this estimate of the state of traffic and using a model for traffic flow, an optimal control policy for the speed signals may be designed.

In Section 6.1 the effect on driver behaviour of displaying an identical speed signal at a given set of consecutive gantries is investigated. Based on experiments with this "homogenizing control" in 1983, it was concluded in an earlier report that the stability of the traffic stream improved significantly [93]. We conclude that (for one carriageway of a dual two-lane freeway) this improved stability has to be attributed mainly to a significant reduction in the fraction of small time headways on the left lane during control. Further effects of control are a small decrease of mean speed, a small increase of capacity, and a small increase of flow on the right lane.

In Section 6.2 two traffic models are discussed, describing the evolution of traffic density and mean speed in one section of a freeway. The simplest model assumes the mean speed to react instantaneously to density changes, in the second model the speed reacts with a certain delay. The stability properties of both models have been investigated and the effect of homogenizing control has been incorporated into the models. The stabilizing effect has been demonstrated by computing the mean time to congestion for the cases with and without control.

In Section 6.3 the design of an optimal control policy was discussed. After discussing the general control problem, the one-dimensional traffic model was used as a basis for the design procedure. Starting off with an infinite horizon criterion representing the throughput of the freeway section, it was shown that one-switch control policies are nearly optimal and that a finite horizon or a discounted criterion should be used to obtain realistic results. Both for a finite horizon criterion and a discounted criterion one-switch control policies were computed for a range of traffic intensity levels. The positive effect of control was illustrated by means of the criterion value, the mean time to congestion and by simulation. To reduce the frequency of switching and to obtain a feasible control law the computed one-switch control policies were extended to hysteresis control policies by computing a separate switch-off value for traffic density. For the two-dimensional model only very limited computations were carried out due to the complexity of the problem.

Applying the control policy as it was designed in this section requires knowledge of the parameters of the model, of λ_0 and of the variables ρ_t and v_t at any time moment. The model parameters were identified in previous chapters.

Traffic density and mean speed may be estimated from the measurement information available through the Dutch Motorway Control and Signalling System, using the filter algorithm developed in Chapter 5. This algorithm also produces an estimate of λ_0 by multiplication of $\hat{\rho}$ and $\hat{v}$ according to the stationarity and homogeneity of flow approximation. This intensity estimate $\hat{\lambda}_0$ shows high frequency fluctuations, see for example Figure 5.12. What we are actually interested in are the long-term fluctuations of the intensity, the increase during the rush hour and return to night-time flow, for example. To obtain this we could apply an exponential smoothing procedure to $\hat{\lambda}_0$:

$$d\bar{\hat{\lambda}}_0(t) = \frac{1}{\kappa}[l\hat{\rho}_l\hat{v}_l - \bar{\hat{\lambda}}_0(t)]dt$$

We propose to take $\kappa = 3$ min to obtain a interval length of about 10 minutes in the averaging procedure. Note that the results of Table 6.8 suggest that the smoothing of λ_0 is mainly important to detect an increase of λ_0 above 3500 veh/h. For values of λ_0 larger than 3500 veh/h the control policy hardly changes.

The combination of the hysteresis policy of Section 6.3.4 and the smoothing procedure presented here nicely decomposes the freeway control problem in a fast and a slow mode part. The slow mode, gradually changing traffic demand, is represented by $\bar{\hat{\lambda}}_0$ and the fast mode, the temporary dynamics, by $\hat{\rho}$ or $(\hat{\rho}, \hat{v})$.

Further extension of the research reported here might consist of the design of a control policy based on a traffic model for several freeway sections. In this way shock wave like phenomena could be accounted for and the effect of control on traffic density might be modelled more accurately. Moreover, part of the effect of control may be modelled as a decrease of the anticipation strength in a model for several sections (This parameter strongly influences the density fluctuations). This may be more realistic than modelling the effect as a reduction of density variance only. It is therefore to be expected that the design of a control policy based on a traffic model for several sections will lead to a more realistic result. A model of four sections would have to be considered, as this would fit in nicely in the set-up of the Motorway Control System. In this system every gantry (outstation) may collect information from five measuring locations in the immediate neighbourhood, see Figure 6.12. Because of the limited capacity of the communication lines between the outstations and the control centre, most of the data processing will have to be carried out locally. This includes filtering of traffic counts and speed measurements by the algorithm developed in Chapter 5. The estimates of traffic density and mean speed for all four sections generated by this algorithm will have to be classified according to some optimal rule. This rule could be designed in a way similar as in this chapter, on the basis of a traffic model for four sections of freeway. In this way at every outstation a decision is made whether it is advisable to switch on the signals at the specific station, represented by a 0 or a 1. This 0 or 1 is then sent to the control centre where coordination of all the decisions from the different gantries has to take place, resulting in a final decision whether to apply homogenizing control or not and if

so, which gantries to include in the control measure.

As mentioned in Section 6.3.1, the models presented here might be extended by the introduction of a variable representing the degree of confidence that drivers have in the control measure. Applying control too often, in situations where it is not necessary would lead to a loss of confidence.

Another approach to the signalling control problem posed here is suggested by the results of Section 6.1. It would consist of considering a microscopic freeway traffic model instead of the macroscopic ones of Section 6.2. Such a microscopic model could be a car-following type of model, for traffic on the fast lane of the freeway only. Control would have to be based upon this model and upon the time headways which are directly available from the measurements. This would circumvent the filtering procedure necessary in our approach and allow a more direct modelling of the effect of control. Work along these lines which uses passage speed rather than time headways is reported by Ferrari [31].

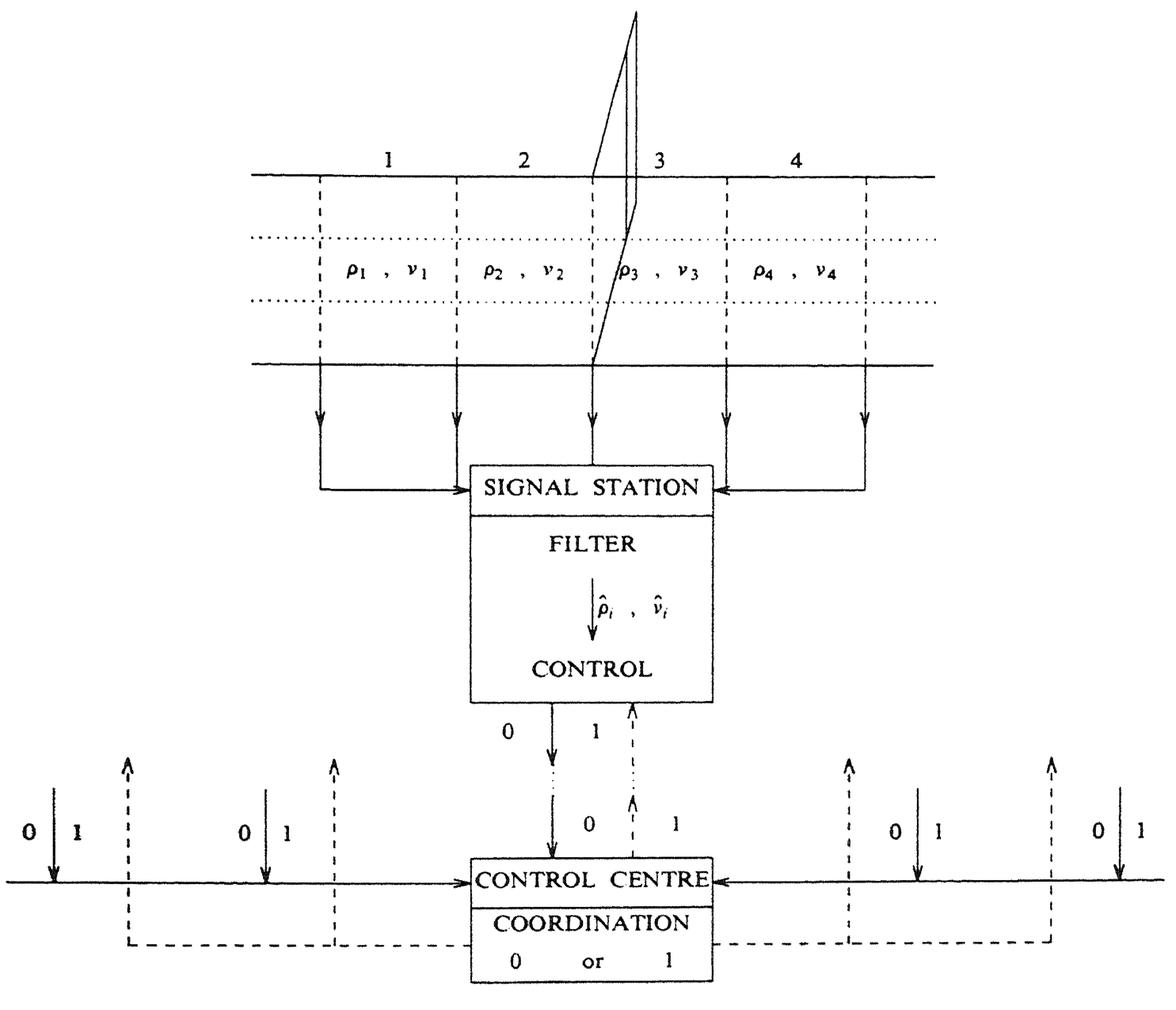

FIGURE 6.12. The proposed setup of the freeway control system.

Chapter 7

Conclusions

In this thesis a freeway traffic control problem has been addressed and solved. The control problem consists of designing an optimal policy for the variable speed signs of the Dutch Motorway Control and Signalling System. The aim is to reduce the instabilities of traffic flow that occur if demand is high, and thereby to reduce the probability of congestion and increase safety.

As a first step towards solving the control problem we presented a mathematical model for freeway traffic in Chapter 2. It consists of a set of nonlinear stochastic differential equations, driven by counting process martingales and Brownian motions. The freeway is divided into sections and for each section two state variables are defined, traffic density (veh/km/lane) and mean speed (km/h). This model is based upon the model developed by Van Maarseveen, but we extend it to account for both vehicle count data and passing speed measurements.

We investigated model behaviour in Chapter 3, by means of stochastic simulation. Three basic traffic situations were considered, ranging from low to high density. The simulations resulted in a modified model, which showed reasonable behaviour in all three circumstances: stationary behaviour when traffic density is low, instabilities when density takes on critical values, and stop-start waves when traffic is congested.

To obtain an estimate of the passing speed probability density function incorporated in the traffic model, an analysis of freeway traffic data was carried out in Chapter 4. The data were obtained through the signalling system. Several hours of data were divided into intervals of approximately stationary traffic by means of a statistical sliding trend test. For each level of traffic demand the probability density was estimated using a modified histogram estimator. It was concluded that the probability density could well be approximated by a normal density, where the mean and variance decrease with increasing traffic intensity. Besides

the probability density some parameters of the equilibrium relation between speed and traffic density were estimated from the data as well.

Based upon the model developed so far we presented a filter in Chapter 5. This algorithm estimates the state of traffic, consisting of traffic densities and mean speeds of all freeway sections, from the available count and speed measurement data. The theoretically optimal filter was approximated and extensively tested against simulated and real traffic data. A first order approximation to the theoretically optimal filter turned out to be sufficiently accurate. It was also concluded that using a minimal amount of speed information is necessary to reduce bias and error variance and to achieve robustness with respect to modelling errors, and that this minimal amount suffices to produce satisfactory estimates. The filter was structurally modified to reduce bias and error variance caused by variability of the speed probability density function.

In Chapter 6 we addressed the actual control problem. The homogenization measure that is considered proved to be successful in reducing traffic instabilities during experiments in 1983. Investigating the data obtained during these experiments we concluded that the increased stability has to be attributed mainly to a significant reduction of the fraction of short time headways on the fast lane during control. A one- and a two-dimensional model for freeway traffic in one section, based upon the model of Chapter 2 were presented next. In the two-dimensional model the speed behaviour is relaxed, whereas in the one-dimensional model the speed is identical to the equilibrium value. Both models were shown to represent the unstable character of traffic flow, by investigation of their stability properties. The effect of control was incorporated in these models and the stabilizing effect of control was demonstrated by comparing the mean time to congestion for the case with and without control.

For the design of an optimal control policy we proposed a criterion which measures the expected total throughput of a freeway section till the moment of congestion. Based upon the policies obtained with this criterion, a hysteresis type of control policy was proposed. This policy consists of switching on control in case traffic density exceeds a critical value, and switching off control again as soon as traffic density becomes lower than another, smaller critical value. The hysteresis prevents switching the control on and off too often.

The hysteresis policy was shown to be capable of postponing congestion considerably and increasing the flow slightly. The homogenization measure turns out to be the most effective for traffic intensity values 10 to 15% below capacity. The policy is based upon exact knowledge of the traffic state at each moment. This information is not available, but the filter developed in Chapter 5 may be used to obtain an estimate of the state.

Several suggestions for further study are given next.

First, the modelling of traffic flow requires further attention. Several modifications of the form of the anticipation term of the model may be considered. Their effect on model behaviour has to be investigated. Furthermore, the effect of merging and lane changing at on- and off-ramps may be investigated and modelled. As all of the freeways in the Netherlands have on- and off-ramps every few kilometers, this extension is a necessary prerequisite for the applicability of

the results of our research. Also, an extensive identification procedure using both vehicle count data and speed measurement information may be carried out to obtain better estimates of model parameters.

Second, several simplifications of the filter, reducing the computational effort, may be studied. One suggestion is to parametrize the error covariance matrix of the filter by the traffic intensity variable. It is expected that the behaviour of the filter will not deteriorate much when using such an approximate covariance matrix instead of the exact one. The parametrization would reduce the amount of computation by 80%. Furthermore, the behaviour of the filter in case of congested traffic requires further study, as in that case modelling errors are present in the observation equation: the approximate relation between intensity, speed and density no longer holds.

Third, control policy design may be based upon more accurate models of freeway traffic, describing the behaviour on a freeway of several sections, thereby including anticipation effects and shock waves. The effect of control may be modelled more accurately this way and the resulting control policy will be more realistic. The solution of the optimization problem requires investigation of efficient numerical approximation schemes. Also, the models may be extended by introducing a variable representing the degree of confidence that drivers have in the control measure.

Finally, the problems of on-ramp control and incident detection may be considered. These require a different design procedure but are expected to result in policies that are as easy to implement as the homogenization policy. The main computational effort is in the estimation of the state of traffic by the filter of Chapter 5. Once these estimates are available, control is relatively easy to implement.

Appendix

Counting Processes and Martingales

In this appendix only a few definitions and theorems are given, for a thorough review of the theory see for example [13, 17, 24].

We assume to be given a *probability space* $(\Omega, \mathcal{F}, P)$ where Ω is the set of elementary events, $\mathcal{F}$ the σ-algebra of events and P a probability measure on $\mathcal{F}$. Next we introduce the *family* of σ-algebras $\{\mathcal{F}_t,\ t \geqslant t_0\ \}$ which we assume to satisfy the following conditions:

(1) $\mathcal{F}_t \subset \mathcal{F}$ for all $t_0 \leqslant t$

(2) $\mathcal{F}_s \subset \mathcal{F}_t$ for all $t_0 \leqslant s \leqslant t$

(3) $\bigcap_{s>t} \mathcal{F}_s = \mathcal{F}_t$ for all $t_0 \leqslant t$

(4) $\mathcal{F}_{t_0}$ contains all sets with P-measure 0.

The σ-algebra $\mathcal{F}_t$ may be thought of as being generated by one or more stochastic processes up to and including t. The second condition is the essential one: it describes the increase of information with time.

We now come to a special kind of stochastic process.

DEFINITION A.1. *The process* $\{m_t,\ t \geqslant t_0\}$ *is called a* martingale *with respect to the family of* σ*-algebras* $\{\mathcal{F}_t,\ t \geqslant t_0\}$ *if it satisfies the following conditions:*

(1) m_t *is* $\mathcal{F}_t$*-measurable for all* $t_0 \leqslant t$

(2) $m_{t_0} = 0$ *except possibly for a set of P-measure zero*

(3) $E[|m_t|] < \infty$ *for all* $t_0 \leqslant t$

(4) $E[m_t \mid \mathcal{F}_s] = m_s$ *except possibly for a set of P-measure zero, for all* s,t with $t_0 \leqslant s \leqslant t$.

The first condition states that at all time instants all the information about m_t has to be contained in $\mathcal{F}_t$. In practice this means that $\{m_t,\ t \geqslant t_0\}$ is (one of) the process(es) that generates $\{\mathcal{F}_t,\ t \geqslant t_0\}$. In this case m_t is said to be *adapted* to the family $\{\mathcal{F}_t,\ t \geqslant t_0\}$. The second condition is not a necessary one, it is just a convention. The fourth condition states the martingale condition. Conditioned upon the information we have at time s we expect the process to be constant in the mean over all realisations. In particular, from (4) it follows that $E[m_t] = E[m_{t_0}]$ for all $t \geqslant t_0$. We can also define *sub-* and *supermartingales* by replacing the equality sign in (4) with $\geqslant$ and $\leqslant$ respectively.

Let us now introduce a counting process.

DEFINITION A.2. *The process* $\{n_t,\ t \geqslant t_0\}$ *is called a* counting process *if it satisfies the following conditions*

(1) $n_{t_0} = 0$ *except possibly on a set of P-measure zero*

(2) *the sample paths are piecewise constant and right continuous except possibly on a set of P-measure zero*

(3) *the jumps are unit jumps.*

A well-known counting process is the *Poisson process* where the intervals between jump times are independent and exponentially distributed with constant parameter λ.

Clearly, a counting process is an increasing process and under the assumption of a finite number of jumps in any finite time interval it is a submartingale with respect to any family to which it is adapted.

THEOREM A.3. *Let* $\{n_t,\ t \geqslant t_0\}$ *be a counting process and* $\{\mathcal{F}_t,\ t \geqslant t_0\}$ *a family of* σ*-algebras. If*

(1) $n_t < \infty$ *except possibly on a set of P-measure zero, for all* $t \geqslant t_0$

(2) n_t *is* $\mathcal{F}_t$*-measurable for all* $t \geqslant t_0$

(3) $E[n_t] < \infty$ *for all* $t \geqslant t_0$

then $\{n_t,\ t \geqslant t_0\}$ *is a submartingale with respect to the family* $\{\mathcal{F}_t,\ t \geqslant t_0\}$.

Under some technical conditions, not to be repeated here, submartingales can be uniquely decomposed into two processes with special characteristics. This is the so called *Doob-Meyer decomposition*. Counting processes satisfy these conditions and we have the following theorem:

THEOREM A.4. (Doob-Meyer decomposition) *Let* $\{n_t,\ t\geqslant t_0\}$ *be a counting process that is a submartingale with respect to the family* $\{\mathcal{F}_t,\ t\geqslant t_0\}$. *Then there exists a unique (right continuous)* $\mathcal{F}_t$*-martingale* $\{m_t,\ t\geqslant t_0\}$ *and a* $\mathcal{F}_t$*-predictable increasing process* $\{a_t,\ t\geqslant t_0\}$ *such that*

$$n_t = a_t + m_t \qquad \text{for all } t\geqslant t_0.$$

The predictability of the process $\{a_t,\ t\geqslant t_0\}$ is a mathematical concept. A useful sufficient condition for a process to be predictable is that the process is adapted and left continuous, or the limit of left continuous adapted processes.

If the process $\{a_t,\ t\geqslant t_0\}$ turns out to be absolutely continuous, it may be written as

$$a_t = \int_{t_0}^{t} \lambda_s ds \qquad \text{for all } t\geqslant t_0,$$

where $\{\lambda_t,\ t\geqslant t_0\}$ is a nonnegative process which is $\mathcal{F}_t$-measurable for all $t\geqslant t_0$. Hence in this case we may write the decomposition as

$$n_t = \int_{t_0}^{t} \lambda_s ds + m_t$$

or

$$dn_t = \lambda_t dt + dm_t.$$

The process $\{\lambda_t,\ t\geqslant t_0\}$ is called the *rate* or *intensity* process of $\{n_t,\ t\geqslant t_0\}$.

The following definition is useful for the derivation of the filter in Chapter 5.

DEFINITION A.5 (quadratic variation). *Let* $\{x_t,\ t\geqslant t_0\}$ *be a square integrable martingale with respect to* $\{\mathcal{F}_t,\ t\geqslant t_0\}$. *Then the following Doob-Meyer decomposition holds for* $\{x_t^2,\ t\geqslant t_0\}$:

$$x_t^2 = \langle x,x\rangle_t + \bar{m}_t$$

where the process $\{\langle x,x\rangle_t,\ t\geqslant t_0\}$ *is called the* predictable quadratic variation *of* $\{x_t,\ t\geqslant t_0\}$ *with respect to* $\{\mathcal{F}_t,\ t\geqslant t_0\}$, *and* $\{\bar{m}_t,\ t\geqslant t_0\}$ *is an* $\mathcal{F}_t$*-martingale.*

For two square integrable martingales $\{x_t,\ t\geqslant t_0\}$, $\{y_t,\ t\geqslant t_0\}$ the

predictable covariation can be defined analogously:

$$x_t y_t = \langle x,y \rangle_t + \overline{\overline{m}}_t.$$

The next theorem is also referred to in Chapter 5.

THEOREM A.6 (optional sampling theorem). *Let* $\{x_t,\ t \geqslant t_0\}$ *be a right-continuous* $\mathcal{F}$*-submartingale and let* τ, σ *be finite stopping times,* $\sigma \leqslant \tau$. *If either*

i) σ *and* τ *are bounded by a finite constant*

or

ii) $\{x_t,\ t \geqslant t_0\}$ *is uniformly integrable*

then

$$E[x_\tau \mid \mathcal{F}_\sigma] \geqslant x_\sigma.$$

REFERENCES

1. E. ALBERTI and G. BELLI (1978). Contributions to the Boltzmann-like approach for traffic flow - A model for concentration dependent driving programs, *Transportation Research*, 12, 33-42.
2. F.C. ANDREWS (1973). A statistical theory of traffic flow on highways: IV Semi-empirical steady state theory, *Transportation Research*, 7, 233-241.
3. A.T. BARUCHA-REID (1960). *Elements of the theory of Markov processes and their applications*, McGraw-Hill, New York.
4. S. BEAN and C.P. TSOKOS (1982). Bandwidth selection procedures for kernel density estimates, *Commun. Statist. - Theor. Meth.*, 11, 1045-1069.
5. R.A. BECKER and J.M. CHAMBERS (1986). *S: an interactive environment for data analysis and graphics*, Wadsworth, Belmont.
6. S.A. BELBAS (1981). Weak solutions of the Hamilton-Jacobi-Bellman equation, *System and Control Letters*, 1, 168-172.
7. A. BENSOUSSAN and J.L. LIONS (1973). Nouvelle formulation de problèmes de contrôle impulsionnel et applications, *C.R. Acad. Sc. Paris*, 276, 1189-1192.
8. A. BENSOUSSAN and J.L. LIONS (1982). *Contrôle impulsionnel et inéquations quasi-variationelles*, Dunod, Paris.
9. L. BREIMAN and R.L. LAWRENCE (1973). Time scales, fluctuations and constant flow periods in uni-directional traffic, *Transportation Research*, 7, 77-105.
10. L. BREIMAN, R.L. LAWRENCE, D. GOODWIN, and B. BAILEY (1977). The statistical properties of freeway traffic, *Transportation Research*, 11, 221-228.
11. L. BREIMAN, W. MEISEL, and E. PURCELL (1977). Variable kernel estimates of multivariate densities, *Technometrics*, 19, 135-144.
12. H. BREZIS and L.C. EVANS (1979). A variational inequality approach to the Bellman-Dirichlet equation for two elliptic operators, *Archive Rational Mechanics Analysis*, 71, 1-13.
13. P. BRÉMAUD (1981). *Point processes and queues. Martingale dynamics*, Springer Verlag, New York.
14. A. CEDER (1979). Stable phase-plane and car-following behavior as applied to a macroscopic phenomenon, *Transportation Science*, 13, 64-79.
15. H. CHERNOFF and A.J. PETKAU (1978). Optimal control of a Brownian motion, *SIAM Journal on Applied Mathematics*, 34, 717-731.
16. H.D. CHIANG, M.W. HIRSCH, and F.F. WU (1988). Stability regions of nonlinear autonomous dynamical systems, *IEEE Transactions on Automatic Control*, AC-33, 16-27.
17. K.L. CHUNG and R.L. WILLIAMS (1983). *Introduction to stochastic integration*, Birkhauser, Boston.
18. D.R. COX (1962). *Renewal Theory*, Methuen & Co. Ltd., London.

19. M.G. CRANDALL and P.L. LIONS (1983). Viscosity solutions of the Hamilton-Jacobi equation, *Transactions of the American Mathematical Society*, 277, 1-42.
20. M. CREMER (1979). *Der Verkehrsfluss auf Schnellstrassen*, Springer Verlag, Berlin.
21. M. CREMER and A.D. MAY (1985). *An extended traffic model for freeway control*, Research report, Institute for Transportation Studies, University of California, Berkeley.
22. M. CREMER and M. PAPAGEORGIOU (1981). Parameter identification for a traffic flow model, *Automatica*, 17, 837-843.
23. M. DELECROIX (1980). Sur l'estimation des densités d'un processus stationnaire à temps continu, *Public. de l'Institut de Statist. de l'Univ. de Paris*, 25, 17-40.
24. C. DELLACHERIE and P.A. MEYER (1975). *Probabilités et Potentiel*, Hermann, Paris.
25. L. DEVROYE (1987). *A course in density estimation*, Birkhäuser, Boston.
26. B.T. DOSHI (1981). Optimal switching among a finite number of Markov processes, *Journal of Optimization Theory and Applications*, 35, 581-610.
27. L.C. EDIE (1961). Car-following and the steady-state theory for non-congested traffic, *Operations Research*, 9, 66-76.
28. L.C. EDIE and E. BAVEREZ (1967). Generation and propagation of stop-start traffic waves, in *Vehicular Traffic Science, Proceedings 3rd International Symposium on the Theory of Traffic Flow*, ed. L.C. Edie, R. Herman, R. Rothery, Elsevier, New York.
29. L.C. EDIE, R. HERMAN, and T.N. LAM (1980). Observed multilane speed distribution and the kinetic theory of vehicular traffic, *Transportation Science*, 14, 55-76.
30. L.C. EVANS and A. FRIEDMAN (1979). Optimal stochastic switching and the Dirichlet problem for the Bellman equation, *Transactions of the American Mathematical Society*, 253, 365-389.
31. P. FERRARI (1988). The reliability of the motorway transport system, *Transportation Research*, 22B, 291-310.
32. W.H. FLEMING and R. RISHEL (1975). *Deterministic and stochastic optimal control*, Springer, Berlin.
33. N.H. GARTNER (1985). Demand-responsive traffic signal control research, *Transportation Research*, 19A, 369-373.
34. N.H. GARTNER (1987). Simulation study of OPAC: a demand-responsive strategy for traffic signal control, in *Proceedings of the 10th International Symposium on Transportation and Traffic Theory*, 233-249, ed. N.H. Gartner, N.H.M. Wilson, Elsevier, New York.
35. D.C. GAZIS and L.C. EDIE (1968). Traffic Flow Theory, *Proceedings of the IEEE*, 56, 458-471.
36. D.C. GAZIS, R.HERMAN, and R.W. ROTHERY (1961). Nonlinear follow-the-leader models of traffic flow, *Operations Research*, 9, 545-567.

37. M.R. GEVERS, R.R. BITMEAD, I.R. PETERSEN, and R.J. KAYE (1986). When is the solution of the Riccati equation stabilizing at every instant?, in *Frequency Domain and State Space Methods for Linear Systems*, 531-540, ed. C.I. Byrnes, A. Lindquist, Elsevier, New York.
38. J.D. GIBBONS (1976). *Nonparametric methods for quantitative analysis*, Holt, Rinehart and Winston, New York.
39. B.D. GREENSHIELDS (1934). A study of traffic capacity, *Proceedings of the Highway Research Board*, 14, 448-477.
40. M.S. GREWAL and H.J. PAYNE (1976). Identification of parameters in a freeway traffic model, *IEEE Transactions on Systems, Man and Cybernetics*, 6, 176-185.
41. OECD ROAD RESEARCH GROUP (1975). *Research on traffic corridor control*, Organisation for Economic Co-operation and Development, Paris.
42. F. HALL, B.L. ALLEN, and M.A. GUNTER (1986). Empirical analysis of freeway flow-density relationships, *Transportation Research*, 20A, 197-210.
43. R.Z. HASMINSKI (1980). *Stochastic stability of differential equations*, Sijthoff & Noordhoff, Alphen aan de Rijn.
44. N.L. JOHNSON and S. KOTZ (1970). *Distributions in statistics. Continuous univariate distributions-2*, Houghton Mifflin Company, Boston.
45. H. KANO (1987). Existence condition of positive-definite solutions for Algebraic Matrix Riccati Equation, *Automatica*, 23, 393-397.
46. M. KOSHI, M. IWASAKI, and I. OHKURA (1983). Some findings and an overview on vehicular traffic flow characteristics, in *Eighth International Symposium on Transportation and Traffic Theory*, ed. V.F. Hurdle, E. Hauer, University of Toronto Press, Toronto.
47. F. KOZIN (1969). A survey of stability of stochastic systems, *Automatica*, 5, 95-112.
48. U. KOEHLER (1974). Stability of vehicle platoons, in *Proceedings of the Sixth International Symposium on Transportation and Traffic Theory*, 39-56, ed. D.J. Buckley, Elsevier, New York.
49. J.L. DE KROES, P. DONK, and S.J. DE KLEIN (1983). *Evaluatie van de externe effecten van het VerkeersSignaleringsSysteem voor Autosnelwegen 1983*, Report.
50. V. KUCERA (1972). A contribution to matrix quadratic equations, *IEEE Transactions on Automatic Control*, AC-17, 344-347.
51. V. KUCERA (1973). A review of the matrix Riccati equation, *Kybernetika*, 9, 42-61.
52. A. KURKJIAN, S.B. GERSHWIN, P.K. HOUPT, A.S. WILLSKY, E.Y. CHOW, and C.S. GREENE (1980). Estimation of roadway traffic density on freeways using presence detector data, *Transportion Science*, 14, 232-261.
53. H.J. KUSHNER (1977). *Probability methods for approximations in stochastic control and for elliptic equations*, Academic Press, New York.

54. S.M. LENHART and S.A. BELBAS (1983). A system of nonlinear partial differential equations arising in the optimal control of stochastic systems with switching costs, *SIAM Journal on Applied Mathematics*, 43, 465-475.
55. P.A.W. LEWIS and G.S. SHEDLER (1979). Simulation of nonhomogeneous Poisson processes by thinning, *Naval Research Logistics Quarterly*, 26, 403-413.
56. M.J. LIGHTHILL and G.B. WHITHAM (1955). On kinematic waves. II: A theory of traffic flow on long crowded roads, *Proceedings of the Royal Society, Series A*, 229, 317-345.
57. P.L. LIONS (1981). Résolution analytique des problèmes de Bellman-Dirichlet, *Acta Mathematica*, 146, 151-166.
58. P.L. LIONS (1983). Optimal control of diffusion processes and Hamilton-Jacobi-Bellman equations. Part 2: Viscosity solutions and uniqueness, *Communications in Partial Differential Equations*, 8, 1229-1276.
59. P.L. LIONS (1983). Optimal control of diffusion processes and Hamilton-Jacobi-Bellman equations. Part 1: The dynamic programming principle and applications, *Communications in Partial Differential Equations*, 8, 1101-1174.
60. P.L. LIONS (1983). Optimal control of diffusion processes and Hamilton-Jacobi-Bellman equations. Part 3: Regularity of the optimal cost function, in *Nonlinear partial differential equations and their applications*, 95-205, ed. H. Brezis, J.L. Lions, Pitman Advanced Publishing Program, Boston.
61. G. LOGIE (1980). *Glossary of Transport*, Elsevier, Amsterdam.
62. M.F.A.M. VAN MAARSEVEEN (1982). *Application of martingales in stochastic systems theory - Surveillance and control of freeway traffic flow*, Thesis, Twente University of Technology, Enschede.
63. M.F.A.M. VAN MAARSEVEEN (1983). A martingale approach to estimation and control of traffic flow on motorways, in *Control in Transportation Systems, Proceedings of the 4th IFAC/IFIP/IFORS Conference*, 203-210, ed. D. Klamt, R. Lauber, Baden-Baden.
64. H. MAHMASSANI and P. MICHALOPOULOS (1985). Traffic Science: Perspectives on future research, *Transportation Research*, 19A, 377-380.
65. P. MANDL (1968). *Analytical treatment of one-dimensional Markov processes*, Springer-Verlag, Berlin.
66. P.S. MAYBECK (1979). *Stochastic models, estimation and control: Volume 1,2 and 3*, Academic Press, New York.
67. P.G. MICHALOPOULOS, D.E. BESKOS, and Y. YAMAUCHI (1984). Multilane traffic flow dynamics: some macroscopic considerations, *Transportation Research*, 18B, 377-395.
68. A.J. MILLER (1961). A queueing model for traffic flow, *Journal of the Royal Statistical Society*, B-23, 64-75.
69. C. MOLER, M. VANBEGIN, and P. VAN DOOREN (1985). *MATLAB-SC Users' Guide*, Technical Note N 166, Philips Research Laboratory, Brussels.

70. M. MORF, B. LÉVY, and T. KAILATH (1978). Square-root algorithms for the continuous time linear least-square estimation problem, *IEEE Trans. Automatic Control*, AC-23, 907-911.
71. P. MUNJAL and J. PAHL (1969). An analysis of the Boltzmann-type statistical models for multi-lane traffic flow, *Transportation Research*, 3, 151-163.
72. NAG. *Fortran Library Manual*, The Numerical Algorithms Group Ltd., Oxford (UK).
73. Y. OGATA (1981). On Lewis' simulation method for point processes, *IEEE Transactions on Information Theory*, IT-27, 23-31.
74. Y. OSHMAN and I.Y. BAR-ITZHACK (1985). Eigenfactor solution of the matrix Riccati equation - a continuous square root algorithm, *IEEE Trans. Automatic Control*, AC-30, 971-979.
75. M. PAPAGEORGIOU (1983). *Applications of automatic control concepts to traffic flow modeling and control*, Springer Verlag, Berlin.
76. S.L. PAVERI-FONTANA (1975). On Boltzmann-like treatments for traffic flow. A critical review of the basic model and an alternative proposal for dilute traffic analysis, *Transportation Research*, 9, 225-235.
77. H.J. PAYNE (1971). Models of freeway traffic and control, in *Mathematical Models of Public Systems. Simulation Council Proceedings*, 51-61.
78. S.R. PLISKA (1973). Single-person controlled diffusions with discounted costs, *Journal of Optimization Theory and Applications*, 12, 248-255.
79. M. POUBELLE, I.R. PETERSEN, M.R. GEVERS, and R.R. BITMEAD (1986). A miscellany of results on an equation of Count J.F. Riccati, *IEEE Transactions on Automatic Control*, AC-31, 651-654.
80. H. REMEIJN (1982). *The Dutch Motorway Control and Signalling System*, Traffic Engineering Division, Dutch Ministry of Transport, The Hague.
81. P.I. RICHARDS (1956). Shock waves on the highway, *Operations Research*, 4, 42-51.
82. T.J. RICHARDSON and R.H. KWONG (1986). On positive definite solutions to the algebraic Riccati equation, *System & Control Letters*, 7, 99-104.
83. M. ROSENBLATT (1956). Remarks on some nonparametric estimates of a density function, *Annals of Mathematical Statistics*, 27, 832-837.
84. W. RUEMELIN (1982). Numerical treatment of stochastic differential equations, *SIAM Journal on Numerical Analysis*, 19, 604-613.
85. J.N.T. SCHUIT (1988). *Numerical solution of the Hamilton-Jacobi-Bellman equation for a freeway traffic flow control problem*, Note, Centre for Mathematics and Computer Science, Amsterdam.
86. S.A. SMULDERS (1987). Modelling and filtering of freeway traffic flow, in *Proceedings of the 10th International Symposium on*

Transportation and Traffic Theory, 139-158, ed. N.H. Gartner, N.H.M. Wilson, Elsevier, New York.

87. S.A. SMULDERS (1988). *Filtering of freeway traffic flow,* Report OS-R8806, Centre for Mathematics and Computer Science, Amsterdam.

88. C.E. DE SOUZA, M. GEVERS, and R.R. BITMEAD (1986). Riccati equations in optimal filtering of nonstabilizable systems having singular state transition matrices, *IEEE Transactions on Automatic Control*, AC-31, 831-838.

89. D. TABAK (1973). Application of modern control and optimization techniques to transportation systems, in *Control and Dynamic Systems*, 345-434, ed. C.T. Leondes, Academic Press, New York.

90. R.A. TAPIA and J.R. THOMPSON (1978). *Nonparametric probability density estimation*, The John Hopkins University Press, Baltimore and London.

91. J.A.C. VAN TOORENBURG (1980). *De stationaire verkeersstroom op de 2x2-strooks autosnelweg,* Report, Traffic Engineering Division, Dutch Ministry of Transport, The Hague.

92. J.A.C. VAN TOORENBURG (1982). *De mens in het autosnelwegverkeer,* Report, Traffic Engineering Division, Dutch Ministry of Transport, The Hague.

93. J.A.C. VAN TOORENBURG (1983). *Homogeniseren. Effekt van aangepaste adviessnelheid op de verkeersafwikkeling,* Report, Traffic Engineering Division, Dutch Ministry of Transport, The Hague.

94. J.A.C. VAN TOORENBURG (1986). *Praktijkwaarden voor de capaciteit,* Report, Traffic Engineering Division, Dutch Ministry of Transport, The Hague.

95. J. TREITERER and J.A. MYERS (1974). The hysteresis phenomenon in traffic flow, in *Proceedings of the 6th Internatonal Symposium on Transportation and Traffic Theory*, 13-38, ed. D.J. Buckley, Elsevier, New York.

96. W. UITTENBOGAARD (1986). Verkeer en wegen door de eeuwen heen, in *Verkeerstechniek tussen onderzoek en praktijk*, 11-35, VUGA Uitgeverij BV, Den Haag.

97. M. VERHAEGEN and P. VAN DOOREN (1986). Numerical aspects of different Kalman filter implementations, *IEEE Transactions on Automatic Control*, AC-31, 907-917.

98. E.J. WEGMAN (1970). Maximum likelihood estimation of a unimodal density, *Annals of Mathematical Statistics*, 41, 2169-2174.

99. G.B. WHITHAM (1974). *Linear and nonlinear waves*, John Wiley & Sons, New York.

100. R. WIEDEMANN (1974). *Simulation des Strassenverkehrsflusses*, Universität Karlsruhe, Germany.

101. H.K. WIMMER (1985). Monotonicity of maximal solutions of algebraic Riccati equations, *System and Control Letters*, 5, 317-319.

102. W.M. WONHAM (1968). On a matrix Riccati equation of stochastic control, *SIAM Journal on Control and Optimization*, 6, 681-697.

103. H. ZACKOR (1972). Beurteilung verkehrsabhängiger Geschwindigkeitsbeschränkungen auf Autobahnen, *Forschung Strassenbau und Strassenverkehrstechnik*, 128.

104. P.M. DE ZEEUW (1988). *Matrixdependent prolongations and restrictions in a blackbox multigrid solver,* Report NM-R8801, Centre for Mathematics and Computer Science, Amsterdam.

INDEX