Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

S.A. Smulders

Control of freeway traffic flow

Department of Operations Research and System Theory   Report OS-R8817   September

The Centre for Mathematics and Computer Science is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

Control of Freeway Traffic Flow

S.A. Smulders
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

In this report a freeway traffic control problem is considered. Control is exerted by means of the variable speed signs of the Dutch Motorway Control and Signalling System. After determining the precise effect of the advisory speed signals on driver behaviour, models for traffic in one section of a freeway are presented and their stability properties investigated. Based on these models hysteresis type control policies are proposed that optimize the throughput of the freeway section and succeed in postponing congestion. The latter is illustrated by means of simulations.

1980 Mathematics Subject Classification : 90B20, 93E20.
Keywords & phrases: freeway traffic, stochastic control.

Report OS-R8817
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

Contents

1. Introduction 3

2. Homogenising control 4

3. Traffic models and stability 8

3.1. A one dimensional traffic model 9
3.2. A two dimensional traffic model 15

4. Control policy design 20

4.1. Control based on traffic density 20
4.1.1. Infinite horizon criterion 21
4.1.2. Finite horizon criterion 24
4.1.3. Discounted criterion 26
4.1.4. Hysteresis control policy 28
4.2. Control based on traffic density and mean speed 31
4.3. Summary 33

5. Summary and conclusions 35

References 36

1. Introduction

The steady increase of traffic demand on freeways during the past decades in the Netherlands has led to a high rate of congestion. This not only meant long delays for individual drivers but also resulted in a large accident rate. As it turns out the accident rate on a congested freeway is about twice the rate on a freeway with freely flowing traffic [14]. Increasing the capacity of the freeway by increasing the number of lanes is a solution which is not always acceptable or even preferable to alternative approaches.

An alternative approach consists of exerting some kind of control over the flow of vehicles by means of signals, with the objective of avoiding unnecessary congestions and reducing the incident rate. Several control methods have been considered in the literature and have been applied in practice. For an overview of a large range of applications see [25]. Most of the research is concentrated upon *on-ramp control*, regulating the rate with which vehicles enter the freeway, by means of traffic lights [11, 19]. This allows direct control over the number of vehicles on the freeway. By maintaining the traffic volume at a certain level, below capacity, one may reduce the probability of congestion. Detailed algorithms have been proposed in the literature [17], but experiments in practice are not reported often.

An approach that seems to be less well developed is concerned with *rerouting* traffic through a network of freeways and/or secondary roads. One example of a study is [8].

Control by means of variable *speed signals* is not considered very often in literature either. Some experiments with control by means of variable speed limitations are reported in [30] and used in [7] 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 signs is to assess the effect on driver behaviour. In contrast to what apparently was found in [30], in the Netherlands speed limitations are not necessarily followed by drivers. Moreover, the speed signs of the Dutch Motorway Control and Signalling System are not to be interpreted as limitations but as "advisory" speeds [22]. In section 2 of this report we will discuss the effect of these advisory speeds on driver behaviour, by investigating data from experiments in practice [29]. The effects of speed signal control turn out to be far less drastic than proposed in [7]. There an increase of capacity of 21% is assumed to be achievable. According to our data capacity hardly increased, but control nevertheless had a significant positive effect on the stability of the traffic stream.

At the moment the Dutch Signalling System is equipped with an Automatic Incident Detection algorithm (AID) [22], which tries to detect serious disturbances in the traffic stream as soon as possible and automatically generates a suitable set of speed signals for the oncoming traffic. The AID algorithm has resulted in a fair decrease of the number of accidents, but is unable to (and not meant to) avoid the actual occurrence of congestion. In this report we propose a set-up for control of freeway traffic by means of variable speed signs, meant to avoid or at least to postpone congestion. Preliminary control policies are designed as well and evaluated through simulation studies.

In section 2 the effect of control on driver behaviour and traffic stability is investigated. In section 3 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 4 is devoted to control policy design. A hysteresis type of control policy is proposed, based on traffic density alone. Preliminary investigations for control based on traffic density and speed are given. In section 5 conclusions are drawn.

2. Homogenising 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 speed signs that vary too much in place or time, as this would confuse drivers and might be dangerous. These constraints may be difficult to handle mathematically, it may even 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 [23], 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. Clearly, the simpler the strategy that is to be designed, the easier it will be to achieve the desired robustness.

It is for these reasons that we will 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 will be discussed in the following.

The type of control we will consider is called *homogenising control* and is described in [29]. 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 will in almost all cases lead 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 will 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 [27]. 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 in [12] e.g. It is 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, flow differences between the lanes. An illustration of the large differences that exist between lanes may be found in figure 2.2. 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 [28]. Furthermore, not all possible disturbances can be avoided. It is the primary aim of homogenising 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 [22] in 1983. Two freeway stretches, a 2-lane one near Utrecht and a 3-lane one 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 in [29]. 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 homogenising control. The decrease that was measured amounted to about 50% ;
—during control capacity did not decrease. There is some indication for 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 homogenising control is advantageous in that it increases safety and reduces the probability of congestion. The effect is not so much in the homogenisation however, as 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 in explaining 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 earlier 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 2.1 where a plot of the fraction of small time headways in relation to the intensity is copied from [27]. The data is separated per lane. Our measured fractions are added: the "H"-es 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 senseless for intensities below 3000 veh/h.

The difference in character between left and right lane is illustrated once more by this plot. From figure 2.1 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. 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 2.2 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* [2], using the function "density". This involves a histogram-like estimator, a description of which may be found in [26]. Note that, in contrast to what one is likely to expect, the time headways form a nearly uncorrelated process [3] thereby justifying the applied estimation

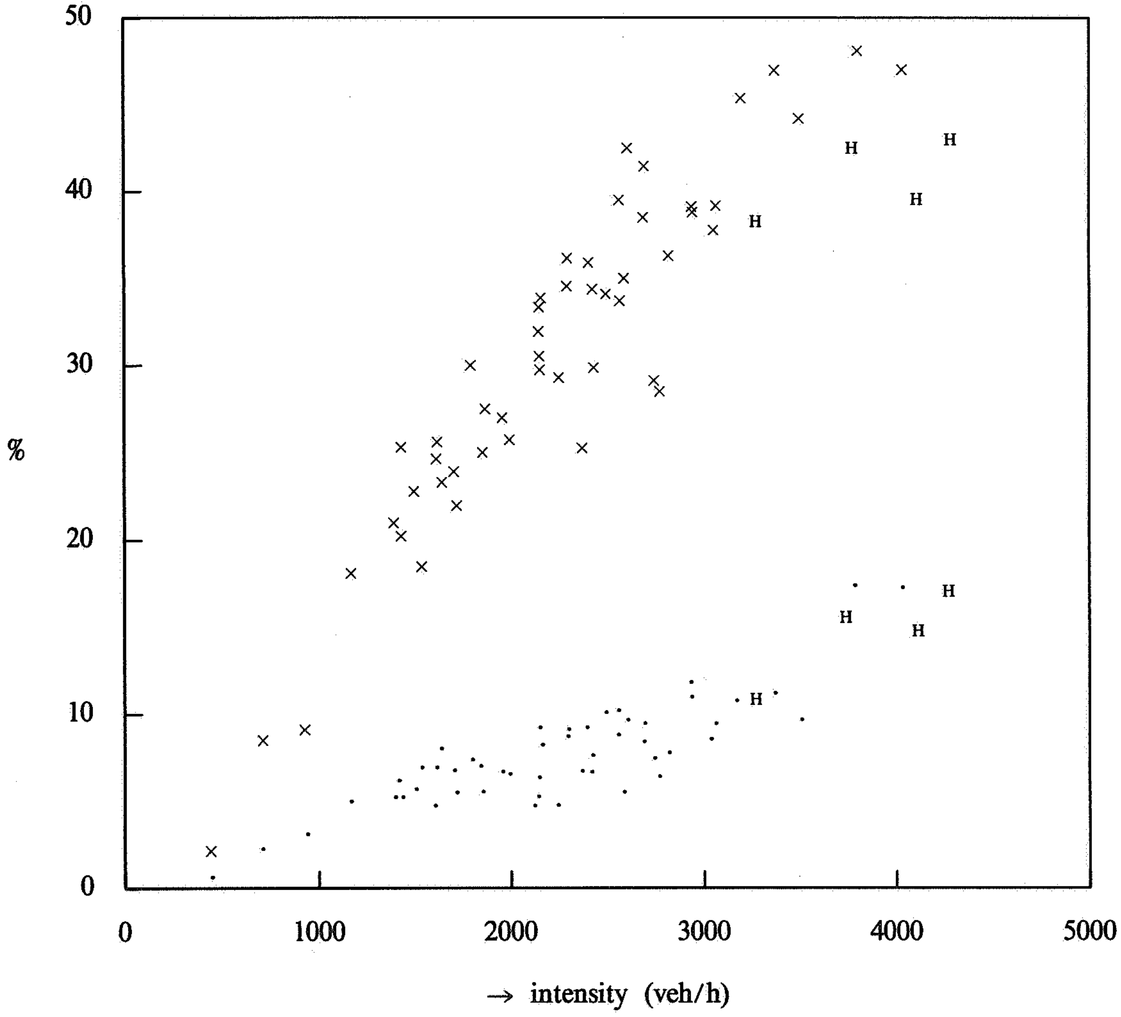

FIGURE 2.1

procedure. The number of data points involved are approximately 2500 for the left lane and 1600 for the right lane.

A consequence of the reduction of small time headways is the decrease of the variance of these variables. For the plotted probability densities this reduction amounts to 16%: from 2.22 to 1.86 s^2. This is to 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. So, both the fraction of larger as well as the fraction of smaller headways is reduced, resulting in a smaller variance.

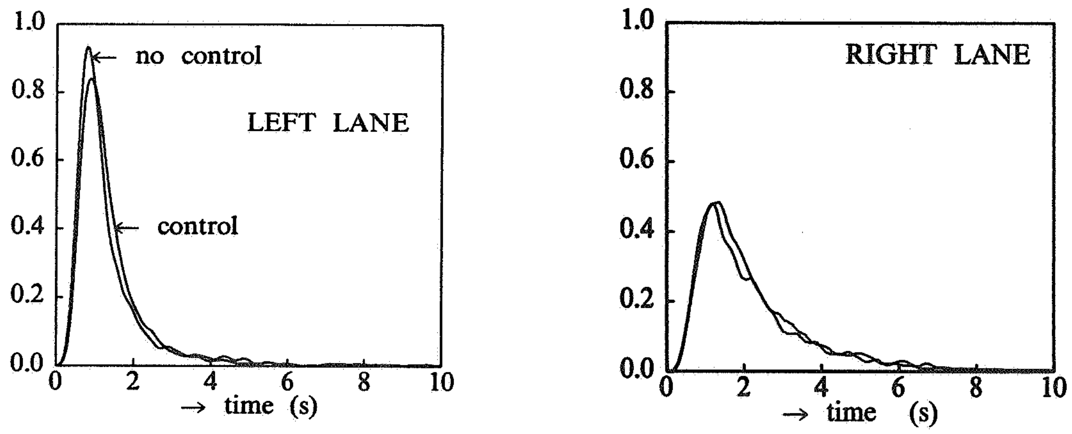

FIGURE 2.2

The mean speed reduction due to 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, 3% on the average. This implies an increase of intensity on the right lane of 3%. Near capacity, the distribution of the intensity over left vs. right lane is about 3:2 so that an increase of total intensity of about 1% may be expected.

Summarising we note that the main effect of homogenising control occurs on the left lane and comes down 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 will be incorporated into traffic models and the increased stability due to control will be theoretically confirmed. The models we will 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 will now investigate the macroscopic implications of the microscopic effects of homogenising 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 [3], 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 [6]:

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

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 $var(N_{t+\Delta t} - N_t)$ 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. So,

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

and

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

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 $var(\Delta N_t)$ leads to a reduction in $var(\Delta\rho_t)$.

Again we used data from the 2-lane freeway near Utrecht to check this theoretical result. The filter developed in [24] 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 $(\text{veh/km/l})^2$ due to control (−25%). Other values of Δt gave analogous results.

Further analysis will be carried out in the next section to asses the effect on the variance of the density itself instead of the increments. For this a simple traffic model will be used. It will be concluded that a reduction of $var(\rho_t)$ 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 $var(\rho_t)$.

3. Traffic models and stability

In this section two traffic models will be presented and the effect of the type of control described in the previous section will be analysed. A model for freeway traffic flow was developed in [23]. There, the state of traffic was 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 was based on the model of H.J. Payne [20], also used by Van Maarseveen [16]. Although this model showed satisfactory behaviour and was able to represent the instabilities that occur in practice, we will not use it here. Instead, we will consider models for one section of freeway only, the simplicity of which allows some analysis such that the control policy design procedure is still feasible. The model for traffic in more than one section may be considered in a later stage, using the insight obtained in the problems that are considered here.

3.1 A one-dimensional traffic model. The simplest model for traffic on a freeway which still incorporates the main features consists of the stochastic differential equation for the density in one section. Denote

ρ_t : the density (number of veh/km/lane) in the section at time t;

v_t : the mean speed (km/h) of the vehicles in the section at time t.

then the model is given by

$$d\rho_t = \frac{1}{Ll}(\lambda_0 - l\rho_t v_t)dt + \sigma dw_t \tag{3.1}$$

$$v_t = v^e(\rho_t) \tag{3.2}$$

Here $v^e(\rho)$ denotes the *equilibrium speed* corresponding to density ρ, as it was defined in [23]:

$$v^e(\rho) = \begin{cases} v_{free} - \alpha\rho & , \quad 0 \leqslant \rho \leqslant \rho_{crit} \\ d(\frac{1}{\rho} - \frac{1}{\rho_{jam}}) & , \quad \rho_{crit} < \rho \leqslant \rho_{jam} \end{cases} \tag{3.3}$$

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, or equilibrium speed as ρ approaches zero (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 were found by direct estimation from freeway data [24]:

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

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

$$\alpha = 0.58 \text{ km}^2/\text{h}$$

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 such as to achieve a realistic value for the variance of ρ_t. From filtered [24] freeway data we found $var(\rho) \approx 50$ (veh/km/lane)^2 which, using a linearized version of (3.1) leads to $\sigma^2 \approx 14000$.

The model given by (3.1) and (3.2) 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 real valued, whereas in practice it is discrete: $0, 1/Ll, \ldots$ 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 earlier:

$$\lambda^{cap} = \max_{0 \leq \rho \leq \rho_{jam}} l\rho v^e(\rho) = 4824 \text{ veh/h}$$

To show that congestion may occur we will first consider the deterministic version of (3.1), (3.2), 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 3.1. 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 will always occur in practice, see figure 3.1.

PROPOSITION 3.1 *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 (3.1), (3.2) has two equilibrium points*

$$\rho_0^s = \frac{v_{free}}{2\alpha} - \sqrt{\left(\frac{v_{free}}{2\alpha}\right)^2 - \frac{\lambda_0}{l\alpha}}$$
$$\rho_0^u = \left(1 - \frac{\lambda_0}{ld}\right) \rho_{jam}$$

of which ρ_0^s *is stable and* ρ_0^u *is unstable;*

ii) if $\lambda_0 = \lambda^{cap}$ *the model (3.1), (3.2) has one unstable equilibrium point*

$$\rho_0^u = \rho_{crit} \ ;$$

iii) if $\lambda_0 > \lambda^{cap}$ *the model (3.1), (3.2) has no equilibrium points.*

PROOF

i) The equilibrium points may easily be found by setting the first term on the right-hand side of (3.1) equal to zero. Then the asymptotic stability of ρ_0^s may be checked by linearisation 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 $x_t \leq \rho_{crit} - \rho_0^s$. Stability follows from $\rho_0^s \leq \rho_{crit}$ 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

$$\frac{dy}{dt} = \frac{d}{L\rho_{jam}} y_t.$$

ii) Analogous to i).
iii) $\lambda_0 - l\rho v^e(\rho) > 0$ for all ρ. □

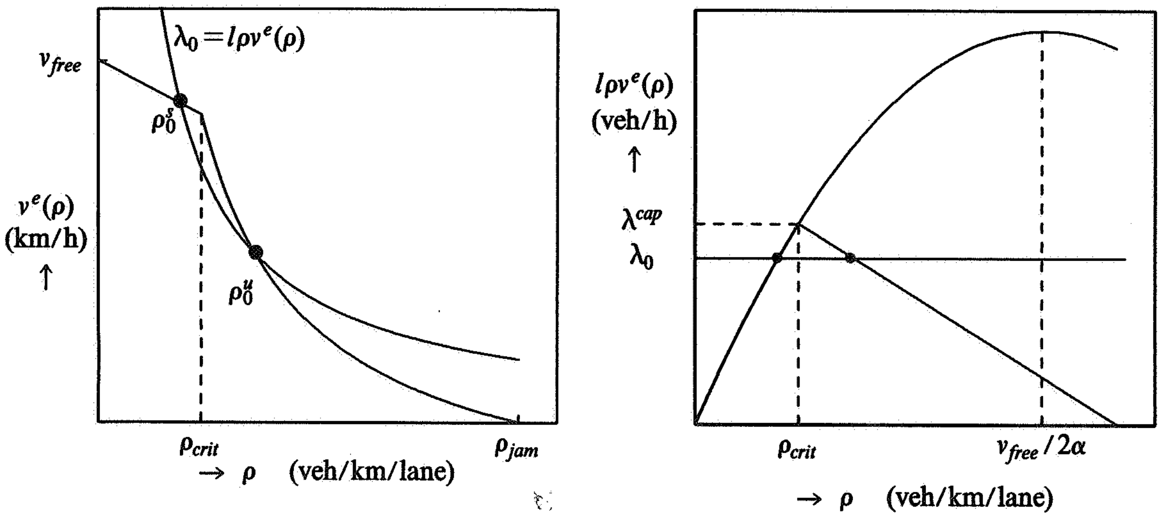

FIGURE 3.1

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 3.1 presents the values of ρ_0^s and ρ_0^u for several values of λ_0.

PROPOSITION 3.2 *Consider the deterministic version of (3.1), (3.2). 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 (3.1) shows that $\frac{d\rho}{dt} > 0$ for all $\rho < \rho_0^s$ and that $\frac{d\rho}{dt} < 0$ for $\rho_0^s < \rho \leq \rho_{crit}$ and so convergence to ρ_0^s is assured for all $\rho \leq \rho_{crit}$. For $\rho > \rho_{crit}$ the model is affine,

$$\frac{d\rho}{dt} = \frac{\lambda_0 - ld}{Ll} + \frac{d}{L\rho_{jam}} \rho_t,$$

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 will 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 (3.1) 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. For a description of this see [18] e.g.

DEFINITION 3.3 *Define the process* $\{\rho_t,\ t\geqslant 0\}$ *on* $[0,\rho_{jam}]$ *by the infinitesimal generator*

$$\frac{1}{2}\sigma^2\frac{d^2}{d\rho^2} + [\lambda_0 - l\rho v^e(\rho)]\frac{d}{d\rho} \tag{3.4}$$

with the boundary at $\rho=0$ *reflecting and the boundary at* $\rho=\rho_{jam}$ *absorbing.*

Now to the effect of the noise term in (3.1). 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

$$\tau = inf\ \{t\geqslant 0\ :\ \rho_t=\rho_{jam}\,|\,\rho(0)\}$$

then it is known [18] that $E[\tau\,|\,\rho(0)]<\infty$, which with $\tau\geqslant 0$, leads to $P(\tau<\infty)=1$. So even at low intensity congestion will eventually occur. This is not as unrealistic as it may look, because if λ_0 is low, $E[\tau\,|\,\rho(0)]$ will be large. Some values of this expectation may be found in table 3.2.

Several stability concepts for stochastic systems have been proposed in the literature, see e.g. [10, 13]. In our case however, all of these lead to trivial conclusions: 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.

The effect of homogenising control will now be incorporated into the model. In the previous section we have described the effects of control on several traffic characteristics. On the left lane the percentage of short time headways decreased 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 decreased, implying a larger intensity. On both lanes the mean speed decreased slightly.

Directing attention to the effect on the density first, we note that a simple analysis of (3.1), consisting of a linearisation around ρ_0^s, leads to

$$var(\Delta\rho_t) = \sigma^2\Delta t(1 - \frac{1}{2}a\Delta t)$$

where

$$a = \frac{(2\alpha\rho_0^s - v_{free})}{L}.$$

Neglecting possible effects of control on a we see that the reduction of $var(\Delta\rho)$ 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 $var(\Delta\rho)$, which was 20 to 25%. This gives a value of about 11000 under control.

The reduction in speed under control may immediately be expressed in v^e by a reduction of v_{free}. See figure 3.2. 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 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.

The conclusion in [29] that capacity does not decrease and possibly increases slightly together with the speed reduction of the previous paragraph implies that ρ_{crit} will have to

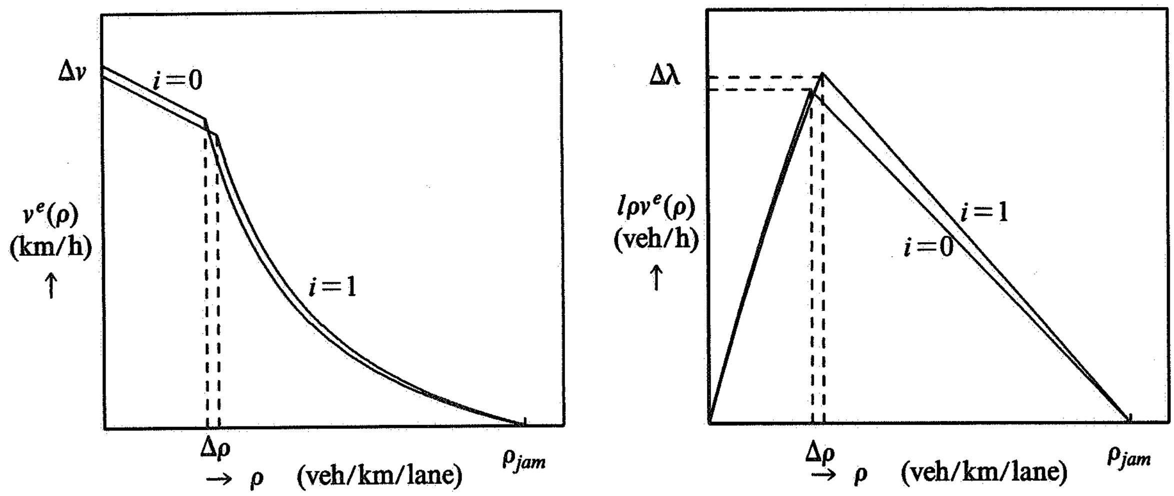

FIGURE 3.2

increase. See figure 3.2. 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:

$$d\rho_t = \frac{1}{Ll}(\lambda_0^i - l\rho_t v_t^i)dt + \sigma^i dw_t \tag{3.5}$$

$$v_t^i = v_i^e(\rho_t) \tag{3.6}$$

Here v_i^e is defined as in (3.3) except that

$$v_{free}^i = \begin{cases} v_{free} , & i=0 \\ v_{free} - \Delta v , & i=1 \end{cases} ,$$

$$\rho_{crit}^i = \begin{cases} \rho_{crit} , & i=0 \\ \rho_{crit} + \Delta\rho , & i=1 \end{cases} .$$

Here i=0 means that no control is applied and i=1 that it is. Furthermore

$$\lambda_0^i = \begin{cases} \lambda_0 , & i=0 \\ \lambda_0 + \Delta\lambda , & i=1 \end{cases} .$$

Concerning the magnitude of the effects:

$$\Delta v = 3 \text{ km/h}$$
$$\Delta \rho = 2 \text{ veh/km/lane}$$
$$\Delta \lambda = 0.01\lambda_0 \text{ veh/h}$$
$$\sigma_0^2 = 14000 \ , \ \sigma_1^2 = 11000$$

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%.

As before we will in fact consider a modification of (3.5) 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 can now be investigated. It is clear that applying control will lead to improved 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 may be deduced from the results of table 3.1 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 3.1

A clear view of the positive effect is obtained from table 3.2 where the value of $E[\tau\,|\,\rho(0)]$ is given for several intensity levels, with and without control. Note that $E[\tau\,|\,\rho(0)]$ depends on the initial value associated with the differential equation (3.5). 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 10% 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 3.2

3.2 A two-dimensional traffic model. The previous model is unrealistic in that the speed reacts instantaneously to a change in density. Instead we now propose

$$d\rho_t = \frac{1}{Ll}(\lambda_0 - l\rho_t v_t)dt + \sigma dw_t \tag{3.7}$$

$$dv_t = -\frac{1}{T}[v_t - v^e(\rho_t)]dt + \mu dz_t \tag{3.8}$$

where

T : the relaxation time (h)

μ : the standard deviation of the noise in the speed equation

z_t : a standard Browian motion (km/h)

It will be assumed that z_t is independent of w_t. We will take

$$T = 0.01 \text{ h}$$

in accordance with the value in [23] 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 [24]. The other parameters have the values as given in subsection 3.1.

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 max $l\rho v^e(\rho)$, 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 3.1: setting the right-hand side of (3.8) to zero leads to $v = v^e(\rho)$. Also, linearisation around these points shows that the first one is stable and the second one unstable. Determining the domain of attraction of the stable equilibrium point requires more effort however. Drawing the phase portrait gives an impression of this domain, see figure 3.3 for an example where $\lambda_0 = 4000$ veh/h. There appears to be a line s, to be called 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 one converges to the unstable equilibrium point. This observation allows us to compute s by parametrising it by ρ, $s(\rho)$, and deriving the differential equation

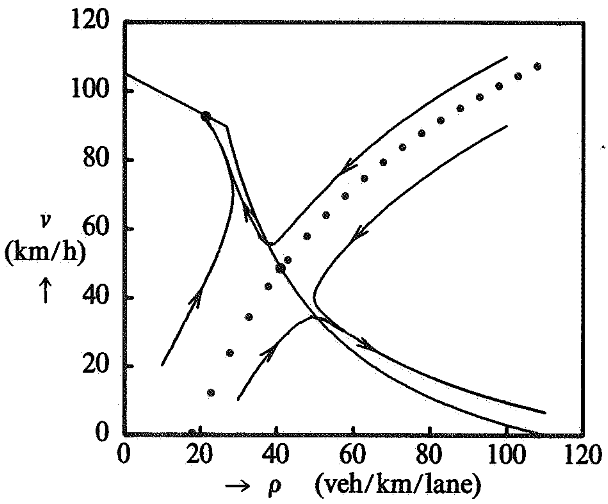

FIGURE 3.3

$$\frac{ds(\rho)}{d\rho} = \frac{Ll}{T}\frac{[v^e(\rho)-s(\rho)]}{[\lambda_0-l\rho s(\rho)]} \tag{3.9}$$

with initial condition

$$s(\rho_0^u) = v^e(\rho_0^u)$$

A small problem occurs in solving (3.9) 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

$$\gamma = \left.\frac{ds(\rho)}{d\rho}\right|_{(\rho_0^u,v^e(\rho_0^u))}$$

Assuming continuity for $\frac{ds(\rho)}{d\rho}$ 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 (3.9) may be numerically integrated to obtain the separator. Plots for several values of λ_0 are presented in figure 3.4. 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 3.5.

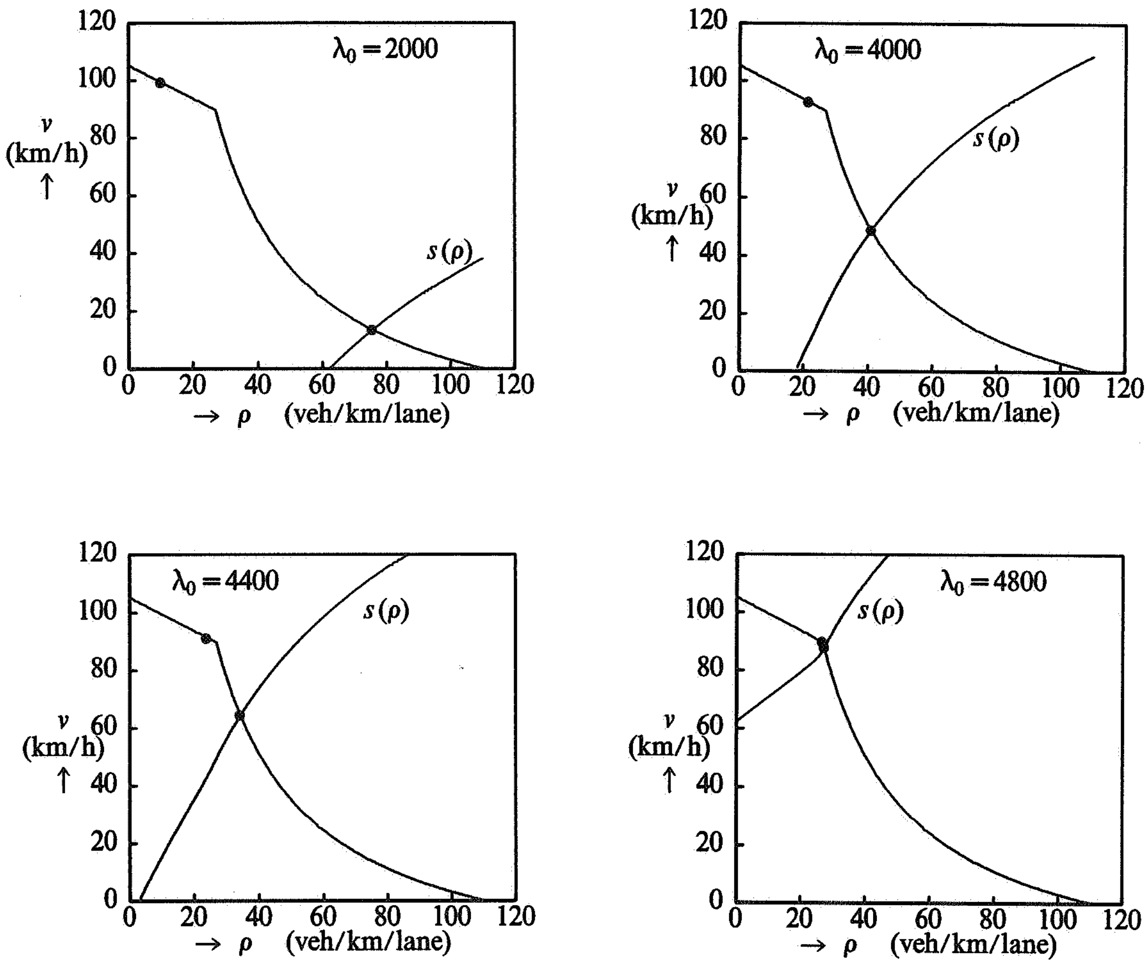

FIGURE 3.4

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 may be found in [5] for systems with a vector field that is continuously differentiable:

THEOREM 3.4 [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 from which convergence to the equilibrium point follows. 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.

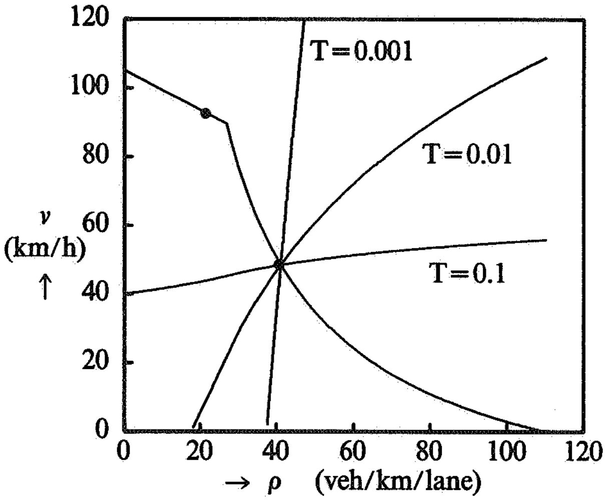

FIGURE 3.5

A problem in theorem 3.4 is that a characterization of equilibrium points that are on the stability boundary is needed. Such characterizations are given in [5]. For theorem 3.4 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 C^1. As long as $\lambda_0 < \lambda^{cap}$ no problems are to be expected however, the smoothness assumption is mainly needed 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 (3.7) (3.8) where the vector field *is* C^1. This is possible without affecting the essential properties of the model. Some suggestions for a smooth $v^e(\rho)$ are given in [7]. We will not go into detail here.

Modelling the effect of control exactly parallels the discussion in the previous subsection and will not be repeated here. The control affects the variance in the density equation and affects $v^e(\rho)$, and is assumed not to affect T and μ.

Again, as in subsection 3.1 we have to restrict the domain of the diffusion $\{(\rho_t, v_t),\ t \geqslant 0\}$ to be bounded. The domain that will be taken is

$$[0, \rho_{jam}]\ x\ [0, v_{\max}]$$

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 however, the choice above is expected not to be crucial.

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, 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 endpoint. 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. A reason for this is that in the deterministic case the vector field on the boundary points inward in the former case and outward in the latter case. Experimenting 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 part of nonzero measure.

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, analogous to the one that will occur in section 4.2:

$$\frac{1}{2}\sigma_i^2\frac{\partial^2 W}{\partial \rho^2}(\rho,v)+\frac{1}{2}\mu^2\frac{\partial^2 W}{\partial v^2}(\rho,v)+\frac{1}{Ll}[\lambda_0^i-l\rho v]\frac{\partial W}{\partial \rho}(\rho,v)+\frac{1}{T}[v_i^e(\rho)-v]\frac{\partial W}{\partial v}(\rho,v)+1=0 \qquad (3.10)$$

where

$$W(\rho_0,v_o) \;=\; E[\tau\,|\,\rho_0,v_0],$$

with boundary conditions

$$\frac{\partial W}{\partial \rho}(0,v)=0,\quad \frac{\partial W}{\partial \rho}(\rho_{jam},v)=0 \text{ for } v>\lambda_0/l\rho_{jam},\quad \frac{\partial W}{\partial v}(\rho,0)=0,\quad \frac{\partial W}{\partial v}(\rho,v_{\max})=0,$$

and $W(\rho_{jam},0)=0$ for $v\leq\lambda_0/l\rho_{jam}$.

The equation is solved by a special routine [31]. The results are given in table 3.3.

For λ_0 close to λ^{cap} an accurate solution was obtained. As λ_0 is decreased however, a more refined grid is needed in the discretisation approach to solve (3.10), 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 120x120 points turned out to be necessary, and the amount of memory required led us to use the Cyber 205 mainframe instead of the usual Cyber 750.

Comparing results to those of table 3.2 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, whereas in the one-dimensional case it occurred about at 4000 veh/h.

	no control	control
λ_0 (veh/h)	$E[\,\tau\,\|\,\rho_0^s,\,v^e(\rho_0^s)\,]$ (min)	$E[\,\tau\,\|\,\rho_0^s,\,v^e(\rho_0^s)\,]$ (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 3.3

4. Control policy design

In this section control policies will be designed of the homogenising type, based on the traffic models described in section 3. The control corresponding to the one-dimensional model of section 3.1 will be extensively discussed in section 4.1. Some preliminary investigations concerning control for the two-dimensional model of 3.2 will be presented in 4.2.

4.1 Control based on traffic density

It is the aim of control to reduce the instabilities present in traffic flow when intensity approaches capacity. As was shown in the previous section 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 will stay 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 will have no significant effect;
—drivers may become so accustomed to the control measure that they will no longer react to it, even in situations where it would be beneficial.

The first point was illustrated in figure 2.1: 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.

Now the advantage of control is easily quantified, by the mean time to congestion e.g. (see table 3.2). 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 will follow another approach: the control will not be applied unless the optimization criterion (to be defined later) significantly decreases. Control will be postponed as long as possible, expressing the idea that no control should be applied unless it is strictly necessary. This will become more clear 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 which 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 in the previous paragraph, and quantify the disadvantage of applying control always. The control policy design procedure would however be considerably more complicated and this extension will therefore not be investigated further.

In the next subsection the criterion that has to be maximised will be defined and the maximisation problem will be investigated. It will turn out that attention has to be directed to a finite horizon or to a discounted criterion. The solution to these problems will be discussed in subsections 4.1.2 and 4.1.3. A refinement of the optimal control policy, meant to reduce the frequency of switching, will be presented in 4.1.4.

4.1.1 Infinite horizon criterion. We will take as the criterion to be maximised

$$V(\rho_0) \;=\; E[\;\int_0^{\tau} l\rho_t v_t^i - \delta I_{\{i=1\}} dt \;\;|\rho_0\;] \tag{4.1}$$

where

$$\tau \;=\; \inf\{t \geqslant 0:\; \rho_t = \rho_{jam}\}\;,$$

the time to congestion. Maximisation takes place over all stationary, piecewise continuous controls:

$$i\;:\;[0,\rho_{jam}] \;\to\; \{0,1\}.$$

Here $i(\rho)=1$ means that control is applied when the density equals ρ. Maximising (4.1) comes down to maximising the total number of vehicles that pass through the section until congestion occurs. This may be achieved by maximising the time to congestion and/or the intensity. Both occur when control is applied and therefore control would always be optimal, if not for the cost associated with control, δ.

Now it may be shown that for each policy $i(.)$ the corresponding criterion function $V(\rho)$ satisfies the following ordinary differential equation:

$$\frac{1}{2}\sigma_i^2\frac{d^2V}{d\rho^2}(\rho) \;+\; \frac{1}{Ll}\;[\lambda_0^i - l\rho v_i^e(\rho)]\frac{dV}{d\rho}(\rho) \;+\; l\rho v_i^e(\rho) \;-\; \delta I_{\{i=1\}} \;=\; 0 \tag{4.2}$$

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 P. Mandl [18] theorem 3, chapter VI. There $V(\rho)$ is shown to be the unique function such that $\frac{dV}{d\rho}(\rho)$ is continuous, and that satisfies (4.2) in every continuity point of $i(\rho)$. The boundary conditions follow from the character of the boundaries of the diffusion process ρ_t: 0 is reflecting and ρ_{jam} is absorbing.

We may use a result from a paper by R. Pliska [21] to show that an optimal control, maximising (4.1) exists and is piecewise constant. The optimal control policy and the associated criterion function were shown by P. Mandl to satisfy the *Hamilton-Jacobi-Bellman* or *dynamic programming equation*

$$\frac{d^2V}{d\rho^2}(\rho) \;+\; \max_{i=0,1}\left\{\frac{2}{\sigma_i^2}\left[\;\frac{1}{Ll}[\lambda_0^i - l\rho v_i^e(\rho)]\frac{dV}{d\rho}(\rho) \;+\; l\rho v_i^e(\rho) \;-\; \delta I_{\{i=1\}}\;\right]\right\} \;=\; 0 \tag{4.3}$$

The optimal policy therefore consists of switching the control on or off at several *switching points*. We will 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 parameter values and will not be 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 will be investigated by means of an example later on, but we will first give some analytical results.

Proposition 4.1 *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 $\frac{dV}{d\rho}(0)=0$. 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 (4.3) and the boundary condition at $\rho=0$. Now control is optimal iff $F(\rho)<0$. □

PROPOSITION 4.2 *If* $\delta=0$, $\sigma_1^2=\sigma_0^2$ *and* $\Delta\lambda=0$ *then there are 3 switching points in general:* $\rho=0$, $\rho=110$ *and the intersection of the equilibrium speed curves with and without control:*

$$\bar{\rho} = \frac{v_{free}-\Delta v+\frac{d}{\rho_{jam}}-\sqrt{(v_{free}-\Delta v+\frac{d}{\rho_{jam}})^2-4\alpha d}}{2\alpha}.$$

It is optimal not to apply control until ρ *exceeds this value. If* $\lambda_0<l\bar{\rho}v^e(\bar{\rho})$ *the switching point* $\bar{\rho}$ *decreases with decreasing* σ_1^2.
PROOF
Now

$$F(\rho) = \frac{2}{\sigma_0^2}\left(\frac{1}{Ll}\frac{dV}{d\rho}-1\right)[l\rho v_1^e(\rho)-l\rho v_0^e(\rho)]$$

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. Note that in general $\frac{dV}{d\rho}(\rho)<0$ (starting closer to the unstable equilibrium point will increase the probability of congestion), excluding the possibility for $\frac{dV}{d\rho}(\rho)=Ll$. Note that $F(\rho)>0$ if $\rho<\bar{\rho}$, so 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 $F<0$ if σ_1^2 is decreased, and $i=1$ is optimal. □

Proposition 4.1 tells us that in absence of a control cost it will be optimal to apply control, even at low intensity values. For our parameters the condition is satisfied. Computations have shown that for $\lambda_0=2000$ control is optimal for all values of ρ.

Proposition 4.2 shows that there will be one important switching point in general, expressing the benefit of control when there is a significant probability of congestion.

We will now turn to the numerical solution of the equation (4.3). Solving this type of equations is a difficult problem in general. For our relatively simple problem no analytical solution exists, but solving (4.3) 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 (4.3) from $\rho=0$ to $\rho=\rho_{jam}$ for W, at every integration step checking whether the term between curly brackets is maximal for $i=0$ or for $i=1$. In this way, choosing an integration step that is small enough, we may find the optimal control and criterion value by integrating (4.3) only once! In general some iterative procedure is needed, 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 that interesting. The occurrence of the point at 48.8 may be explained as follows. Once density has exceeded the unstable equilibrium value, congestion will occur, 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. However, our main interest is in avoiding congestion and for values far above ρ_{crit} the model may not be that realistic.

For other parameter values sometimes more switching points show up. In these cases we will 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 will be applied (we have assumed that $l=2$ and $L=0.5$, so ρ will be integer in practice). Table 4.1, in which the criterion function for the optimal and the suboptimal policy are compared, shows that this policy is nearly optimal. This approach comes down to confining ourselves to a restricted class of controls: those that correspond to exactly 1 switching point.

DEFINITION 4.3 *A* one-switch control policy *is defined to be a policy of the form*

$$i(\rho) = I_{\{\rho \geq \bar{\rho}\}}$$

which means that control is applied only for $\rho \geq \bar{\rho}$.

The value $\bar{\rho}$ will be determined by computing the optimal policy first and selecting the main switching point. These one-switch controls are suboptimal but based on earlier facts and the example we expect them to be close to optimal.

ρ (veh/km/lane)	$V^{opt}(\rho)$ (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 4.1

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 however and experiment with different values. The results for $\delta=100$ and $\delta=500$ are presented in table 4.2. Instead of investigating the sensitivity with respect to δ in detail, we will propose a design procedure in the next two subsections which 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:

$$E[\ \int_0^{\tau} l\rho_t v_t dt\ \ |\rho_0\ ]\ =\ E[\ \int_0^{\infty} l\rho_t v_t dt\ \ |\rho_0\ ]$$

For low values of λ_0 the expectation of τ will be large, for $\lambda_0=2000$ it will be 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 will be considered in the following subsections: a finite time horizon may be introduced or the criterion may be discounted.

$\bar{\rho}$ (veh/km/lane)	λ_0 (veh/h)					
	1000	2000	3000	3500	4000	4800
$\delta=100$	3	5	9	14	22	27
$\delta=500$	9	13	19	22	26	28

TABLE 4.2

4.1.2 Finite horizon criterion. To avoid the problem mentioned in the previous subsection we now introduce

$$V^{fin}(\rho_0,t)\ =\ E[\ \int_t^{\min(\tau,T)} l\rho_s v_s ds\,|\,\rho_0\ ]$$

where

T : time horizon (h)

We will take $T=2$ h in the sequel and are interested in $V^{fin}(\rho,0)$. Now for $\lambda_0=2000$ veh/h e.g. postponing congestion from 4 to 380 years does no longer affect 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 (3.5) and the optimal control policy $i(.)$ satisfy the Hamilton-Jacobi-Bellman equation

$$\frac{\partial V^{fin}}{\partial t}(\rho,t) + \max_{i=0,1}\left\{\frac{1}{2}\sigma_i^2\frac{\partial^2 V^{fin}}{\partial \rho^2}(\rho,t) + \frac{1}{Ll}[\lambda_0^i - l\rho v_i^e(\rho)]\frac{\partial V^{fin}}{\partial \rho}(\rho,t) + l\rho v_i^e(\rho)\right\} = 0 \quad (4.4)$$

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 (4.4) is considerably more difficult than (4.3). It is not easy to carry out the simultaneous maximisation and integration of the parabolic partial differential equation. A method which seems to give good results is the approach of H.J. Kushner [15]. There the diffusion process is approximated by a finite state Markov chain and then policy or value iteration algorithms are applied. This will not be attempted here but is left for future investigations.

We will restrict attention to computation of the criterion value for several elements of the class of one-switch controls introduced in 4.1.1. This avoids the maximisation problem in (4.4) and leaves us with the partial differential equation:

$$\frac{\partial V^{fin}}{\partial t}(\rho,t) + \frac{1}{2}\sigma_i^2\frac{\partial^2 V^{fin}}{\partial \rho^2}(\rho,t) + \frac{1}{Ll}[\lambda_0^i - l\rho v_i^e(\rho)]\frac{\partial V^{fin}}{\partial \rho}(\rho,t) + l\rho v_i^e(\rho) = 0$$

where $i(\rho)=I_{\{\rho\geqslant\bar{\rho}\}}$ and $\bar{\rho}$ is given. This parabolic equation is solved by means of the routine D03PAF from the NAG-library. Starting with the one-switch control corresponding to $\bar{\rho}=0$ we will increase $\bar{\rho}$ as long as no deterioration of the criterion function occurs. To be precise: $\bar{\rho}$ will be maximised 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 4.3. 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 4.3

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 too high. This may also explain the increase in $\bar{\rho}$ for $\lambda_0>4000$.

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 4.1 shows a realisation of ρ_t and figure 4.2 the corresponding realisation 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 realisations 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 however, it will take some time before another critical situation turns up. This means that control leads to a relatively small probability for a relatively large benefit.

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 will be addressed in 4.1.4.

4.1.3 Discounted criterion. 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

$$V^{dis}(\rho_0) = E[\int_0^T e^{-ct} l \rho_t v_t dt \mid \rho_0] \tag{4.5}$$

where

c : discount factor, ≥ 0

We will take $c = 0.5$ in order to achieve results that may be compared to the finite horizon ones:

$$\int_0^\infty e^{-0.5t} dt = T = 2 \text{ h}$$

in this case. The Hamilton-Jacobi-Bellman equation associated with (4.5) equals

$$\frac{d^2 V^{dis}}{d\rho^2}(\rho) + \max_{i=0,1} \left\{ \frac{2}{\sigma_i^2} \left[\frac{1}{Ll}[\lambda_0^i - l\rho v_i^e(\rho)] \frac{dV^{dis}}{d\rho}(\rho) - cV^{dis}(\rho) + l\rho v_i^e(\rho) \right] \right\} = 0 \tag{4.6}$$

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 4.1.1 does not apply here. We will therefore follow the approach of the previous subsection and restrict our 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. The results are presented in table 4.4. Computing the optimal values of $\bar{\rho}$ sometimes turned out to be more involved than for the finite horizon criterion. We failed to compute the value for $\lambda_0 = 1000$ veh/h.

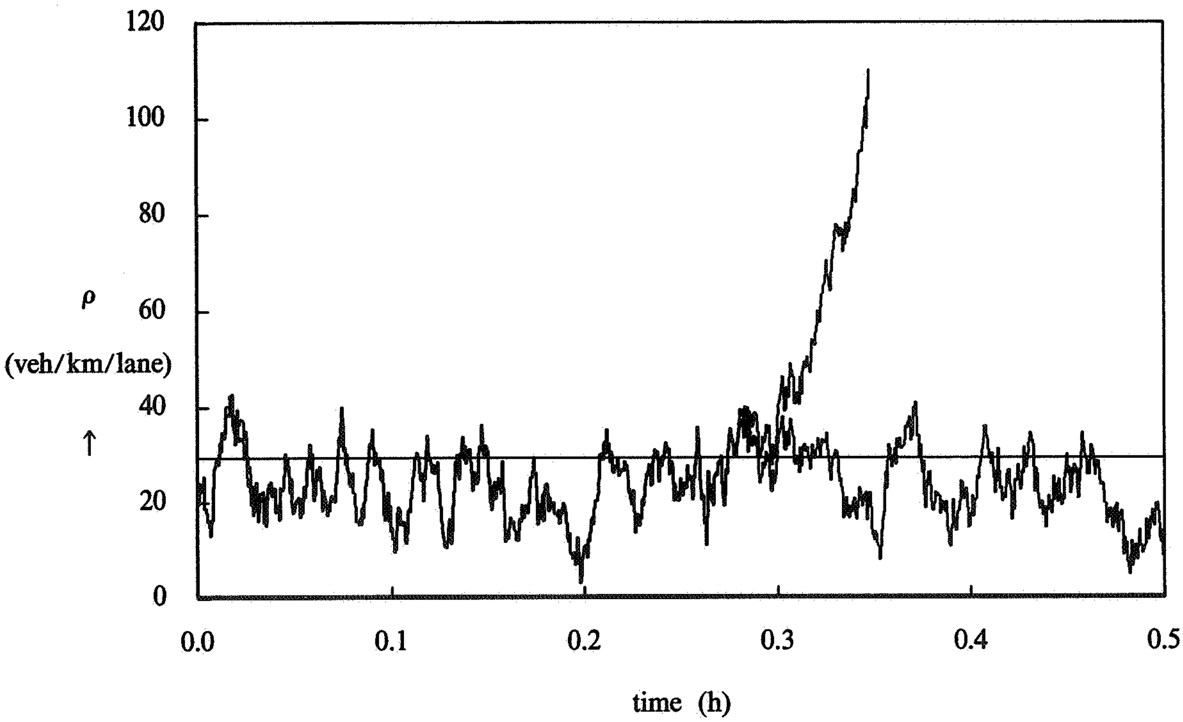

FIGURE 4.1

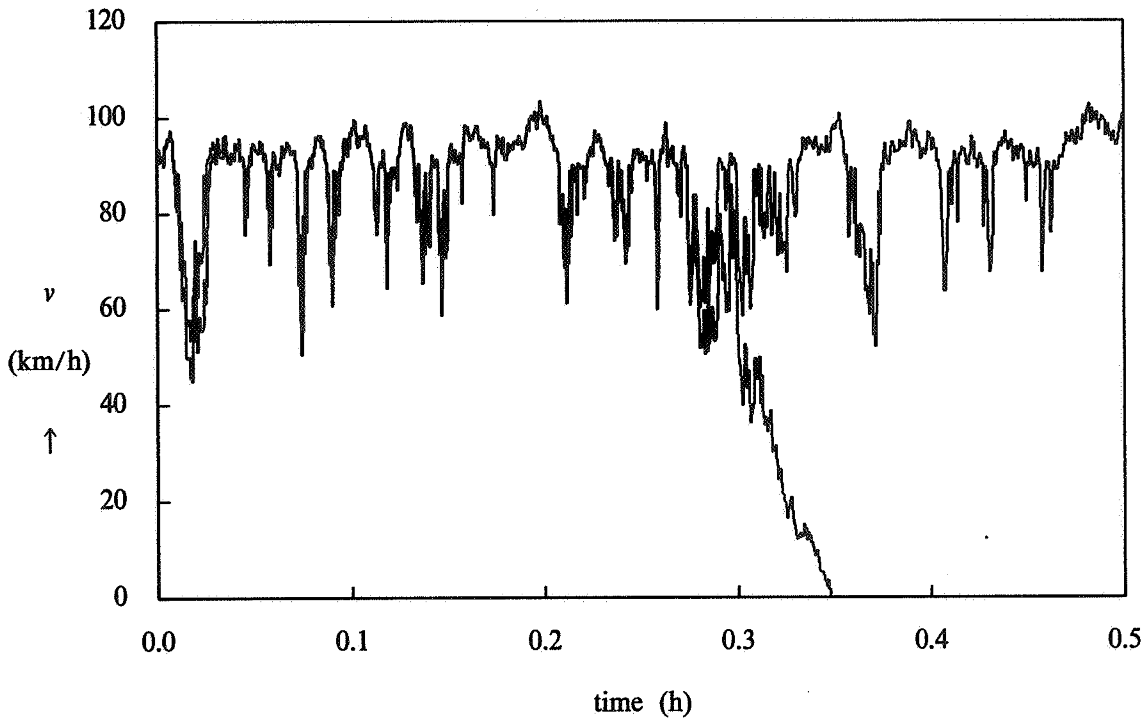

FIGURE 4.2

λ_0 (veh/h)	1000	2000	3000	3500	4000	4400	4800
$\bar{\rho}$ (veh/km/lane)	*	70	42	28	28	29	31

TABLE 4.4

We see the same qualitative behaviour as for the finite horizon criterion.

4.1.4 Hysteresis control policy. In the literature optimisation problems where one decision out of a finite set has to be chosen are addressed, introducing costs for switching from one decision to another. H. Chernoff and A. Petkau [4] have considered this problem for the case of a Brownian motion process. 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 4.3.

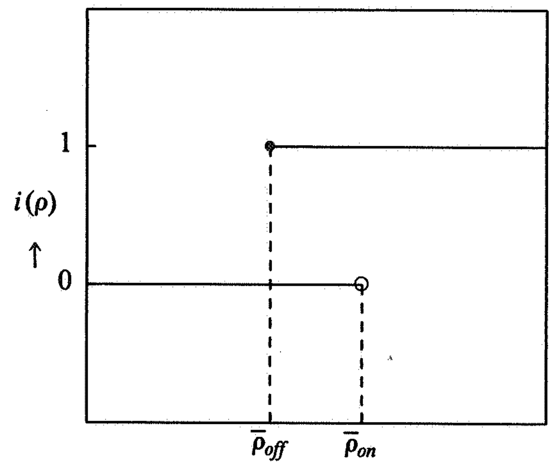

FIGURE 4.3

It is clear that there is need for an auxiliary state variable which 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:

$$C_0 = \{(\rho,s):\ \rho\in[0,\bar{\rho}_{on}),\ s=0\} \tag{4.7}$$

$$C_1 = \{(\rho,s):\ \rho\in[\bar{\rho}_{off},\rho_{jam}],\ s=1\} \tag{4.8}$$

$$S_{01} = \{(\rho,s):\ \rho\in[\bar{\rho}_{on},\rho_{jam}],\ s=0\} \tag{4.9}$$

$$S_{10} = \{(\rho,s):\ \rho\in[0,\bar{\rho}_{off}],\ s=1\} \tag{4.10}$$

C_0 and C_1 are called the *continuation* 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 $i(\rho)=0$. As long as $(\rho,s) \in C_1$ control is applied, so $i(\rho)=1$. As soon as (ρ,s) enters S_{01}, $i(\rho)$ switches from 0 to 1 and as soon as (ρ,s) enters S_{10}, $i(\rho)$ switches from 1 to 0.

The hysteresis clearly enables a reduction of the frequency of switching. Now the assignment of costs to switching is ad hoc in our problem, it is not clear how to choose them or interpret them. We would like to satisfy a condition like not switching off the system before m minutes after switching it on have gone by, or something of the kind. To achieve this one might experiment with several values for the costs. We will follow a more direct approach instead.

We will consider the hysteresis type of control of figure 4.3 and design it according to the following procedure:

i) we compute $\bar{\rho}_{on}$ according to the original criterion and procedure of subsection 4.1.2. and 4.1.3. If $\bar{\rho}_{on}=\rho_{jam}$ take $i(\rho)\equiv 0$, otherwise proceed to *ii*);

ii) $\bar{\rho}_{off}$ is taken to be the solution of the problem

$$\max_{0\leqslant\bar{\rho}_{off}\leqslant\bar{\rho}_{on}} \bar{\rho}_{off}$$

under the condition

$$P(\theta\leqslant m)\leqslant 0.05$$

where m is a given time duration and θ the time to the first occurrence of an on-off-on *cycle*:

$$\theta = \inf\{t\geqslant 0:\ \rho_t=\bar{\rho}_{on} \wedge \rho_s=\bar{\rho}_{off},\ \text{for some } s,\ 0\leqslant s\leqslant t \mid \rho_0=\bar{\rho}_{on}\}$$

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 maximising 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 will be at most 1 such a cycle per m time units. This clearly reduces the number of switchings. It does not however exclude the possibility of a short on-off or off-on sequence. 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 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 4.3) and control will only be 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 mentioned restriction 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 cycli leads to about the same results as for on-off-on cycli when λ_0 is low, but for $\lambda_0>3500$ veh/h $\bar{\rho}_{off}=0$. Once an excursion has occurred control will always be applied. This is not unrealistic, but the results of table 4.5 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 4.4 *Let the process* $\{\rho_t,\ t\geqslant 0\}$ *on* $[0,\rho_{jam}]$ *be defined by (3.5) with 0 reflecting and* ρ_{jam} *absorbing and let u(.) be the hysteresis control policy corresponding to the continuation and switching sets (4.7) to (4.10), with* $\bar{\rho}_{on}\in[0,\rho_{jam})$ *and let* $m>0$. *Assume also that* $\sigma^0,\sigma^1>0$. *Then either* $P(\theta\leqslant m)>0.05$ *for all* $\bar{\rho}_{off}\in[0,\bar{\rho}_{on}]$, *or a unique* $\bar{\rho}_{off}$, *satisfying the condition* $P(\theta\leqslant m)\leqslant 0.05$, *exists.*

PROOF

From the pathwise continuity of $\{\rho_t,\ t\geqslant 0\}$ it follows that for $0<\bar{\rho}_{off}\leqslant\bar{\rho}_{on}$, $0<\Delta\leqslant\bar{\rho}_{off}$:

$$A_\Delta(m)\ \subset\ A_0(m)$$

where

$$A_\Delta(m)\ =\ \{\omega\,|\ \rho_t=\bar{\rho}_{on},\ \rho_s=\bar{\rho}_{off}-\Delta,\ \textit{for some s and t with}\ 0\leqslant s\leqslant t\leqslant m\,|\,\rho_0=\bar{\rho}_{on}\}$$

$$A_0(m)\ =\ \{\omega\,|\ \rho_t=\bar{\rho}_{on},\ \rho_s=\bar{\rho}_{off},\ \textit{for some s and t with}\ 0\leqslant s\leqslant t\leqslant m\,|\,\rho_0=\bar{\rho}_{on}\}$$

i.e., A_0 is the set of realisations, 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 analogon 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 $\{\rho_t,\ t\geqslant 0\}$, 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 $P(\theta\leqslant m)$ 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 $\bar{\rho}_{off}>0$ $P(\theta\leqslant m)\leqslant 0.05$ holds, or for $\bar{\rho}_{off}=0$ $P(\theta\leqslant m)>0.05$. □

Solving for $\bar{\rho}_{off}$ is not possible analytically. Computing first passage times requires applying (inverse) Laplace transforms on solutions of an ordinary differential equation, see [1]. 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 (3.5) 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 realisations (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 4.5, the values of $\bar{\rho}_{on}$ are the ones obtained in 4.1.2 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 4.5

In the simulation experiment presented earlier ρ_t exceeded 29 $(=\bar{\rho}_{on})$ at t=0.01 h and only become 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 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.

4.2 Control based on 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 was discussed in section 3, a model which accounts for the delayed reaction of speed to density fluctuations is more realistic: (3.7), (3.8). Such a model allows larger excursions of the density without convergence to ρ_{jam} occurring. This will result in a more realistic control policy, now depending on both ρ and v. The phase plane is divided in two regions and control will be switched on or off when the speed-density pair crosses the boundary between the two regions. The approximative form of this boundary may be deduced from the separators of figure 3.3. In figure 4.4 the separator for λ_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 dashed line represents the switching curve corresponding to the policy based on density alone.

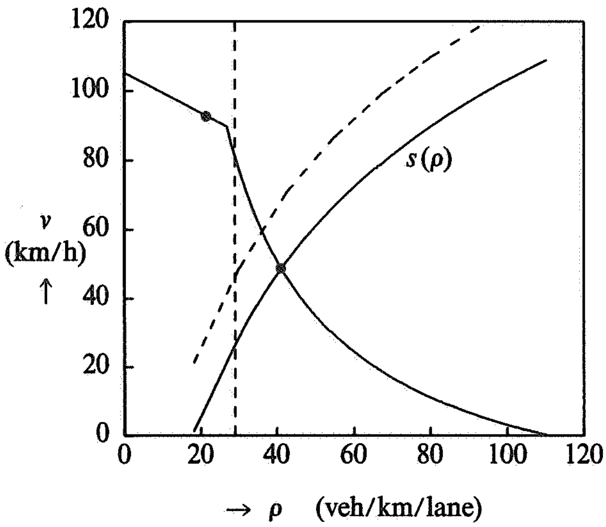

FIGURE 4.4

For the one-dimensional model and the infinite horizon criterion (4.1) we have been able to compute the exact optimal switching policy $i(\rho)$ numerically. For the more realistic finite horizon or discounted criterion this turned out to be much more involved and we only computed suboptimal policies. Now the infinite horizon problem for the two-dimensional model already leads to an elliptic partial differential equation and poses the same problems as in 4.1.2 and 4.1.3. A promising approach to solving this type of problem is the approximation procedure justified by H.J. Kushner [15], but this will not be investigated here. We will only present a limited illustration of the idea expressed in figure 4.4 and restrict ourselves to computing a suboptimal solution for the discounted criterion (4.5). The domain of the two-dimensional diffusion process is restricted to $[0,\rho_{jam}]$ x $[0,v_{\max}]$ like in section 3.2. We choose $v_{\max} >> v_{free}$ in order to achieve a good approximation of the unrestricted process. Note that v will 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 $v \leq \lambda_0 / l\rho_{jam}$ is absorbing and the rest reflecting just as in 3.2.

Optimizing control policies corresponding to a single curve in the (ρ,v) plane, like the one in figure 4.4 still is too involved. We will only compare the criterion corresponding to some simple policies therefore.

The criterion value function $V^{dis}(\rho,v)$ corresponding to such policies satisfies the following equation:

$$\frac{1}{2}\sigma_u^2\frac{\partial^2 V^{dis}}{\partial \rho^2}+\frac{1}{2}\mu^2\frac{\partial^2 V^{dis}}{\partial v^2}+\frac{1}{Ll}[\lambda_0^u - l\rho v]\frac{\partial V^{dis}}{\partial \rho}+\frac{1}{T}[v_u^e(\rho)-v]\frac{\partial V^{dis}}{\partial v}$$

$$-cV^{dis}+l\rho v \;=\; 0 \tag{4.11}$$

with boundary conditions

$$\frac{\partial V^{dis}}{\partial \rho}(0,v)=0,\quad \frac{\partial V^{dis}}{\partial \rho}(\rho_{jam},v)=0 \text{ for } v>\lambda_0/l\rho_{jam},\quad \frac{\partial V^{dis}}{\partial v}(\rho,0)=0,$$

$$\frac{\partial V^{dis}}{\partial v}(\rho,v_{\max})=0,\quad V^{dis}(\rho_{jam},v)=0 \text{ for } v\leq\lambda_0/l\rho_{jam}.$$

The equation (4.11) was solved using the solver of [31] for $\lambda_0=4200$ veh/h, $c=0.5$ and the parameter values of section 3.2.

To start with we take policies that only use the value of the density for control: $u(\rho)=I_{\{\rho\geq\bar{\rho}\}}$, and vary $\bar{\rho}$. As it turns out, the optimal criterion function is achieved by taking $\bar{\rho}=28$, see table 4.6 for the criterion value in $(\rho,v)=(20,90)\approx(\rho_0^s,v^e(\rho_0^s))$.

$\bar{\rho}$ (veh/km/lane)	0	26	27	28	29	30	110
$V^{dis}(20,90)$	3444	4017	4063	4096	4058	3912	2200

TABLE 4.6

The optimal value corresponds with the one obtained using the one-dimensional model.

Now 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 4.4. For this we tried the following policy which consists of postponing control when $v > v_{crit}$:

$$u(\rho) = \begin{cases} 0 & \text{if } \rho\leq 28 \wedge v\leq v_{crit} \\ & \text{or } v>v_{crit} \wedge v>1.66\rho+v_{crit}-46.5 \\ 1 & \text{elsewhere} \end{cases}$$

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 4.7.

v_{crit} (km/h)	85	90	95	100
$V^{dis}(20,90)$	3947	4037	4080	4093

TABLE 4.7

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 will have to be carried out to find even better policies. This will not be 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 analogous to section 4.1.4.

4.3 Summary. In this section the design of an optimal control policy was discussed. Starting off with a simple one-dimensional model and an infinite horizon criterion it was shown that one-switch control policies are nearly optimal and that a finite horizon or discounted criterion should be used to obtain realistic results. For both types of criteria suboptimal one-switch control policies were computed for a range of intensity levels. The positive effect of control was illustrated by the criterion value, the mean time to congestion and by means of simulation. To reduce the frequency of switching and obtain a feasible control law the computed one-switch policies were extended to so-called hysteresis control policies, by computing a separate switch-off value for the density. For the two-dimensional model only a very limited investigation was 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 already identified in [23]. 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 [24]. 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 [24]. 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 e.g. To obtain this we could apply an exponential smoothing procedure to $\hat{\lambda}_0$:

$$d\bar{\lambda}_0(t) \;=\; \frac{1}{\kappa}[l\hat{\rho}_t\hat{v}_t - \bar{\lambda}_0(t)]dt$$

We propose to take κ=3 min to obtain a time interval of about 10 minutes in the averaging procedure. Note that the results of table 4.5 suggest that the smoothing of $\hat{\lambda}_0$ 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 4.3 and the smoothing procedure presented here nicely decomposes the freeway control problem in a fast and a slow mode part. The slow mode, the gradual changing traffic demand, is represented by $\bar{\lambda}_0$ and the fast mode, the temporary dynamics, by $\hat{\rho}$ or $(\hat{\rho},\hat{v})$.

An interesting extension of the results in this report might be to consider a model of traffic in more than one section. The model presented in [23] may be used for this. In a model for several sections shock waves are represented, which are absent in a one section model. Moreover, part of the effect of control may be modelled as a decrease in 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 4 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 (signal station) may collect information from 5 measuring locations in the immediate neighbourhood, see figure 4.5.

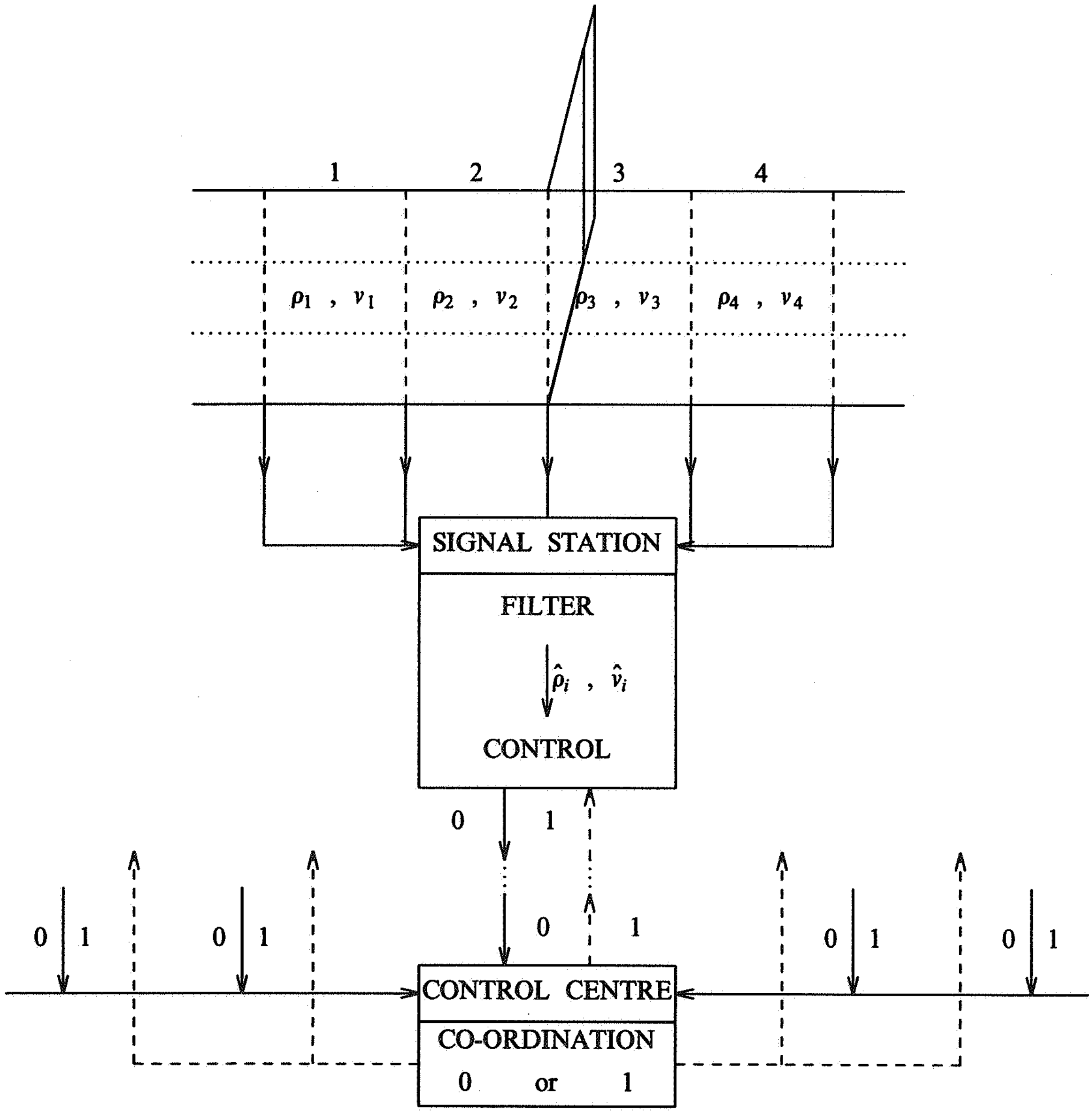

FIGURE 4.5

Because of the limited capacity of the communication lines between the signal stations 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

[24]. The estimates of traffic density and mean speed for all 4 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 report, on the basis of a traffic model for 4 sections of freeway. In this way at every signal station a decision is made as to 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 homogenising control or not and if so, where to apply it.

5. Summary and conclusions

In this report a freeway traffic control problem is 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 a filter algorithm as of [24] makes available estimates of traffic density and mean speed of a series of freeway sections. Based upon this approximative knowledge of the state of traffic and using a model for traffic flow, an optimal control policy for the signal settings may be designed.

In section 2 of this report 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 "homogenising control" in 1983 it was concluded in an earlier report [29] that the stability of the traffic stream improved significantly. In section 2 we conclude that (for a 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 3 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 were investigated and the effect of homogenising control was incorporated into the models. The stabilizing effect was demonstrated by computing the mean time to congestion for the cases with and without control.

In section 4 the design of an optimal control policy was discussed. In the first part 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 suboptimal one-switch control policies were computed for a range of traffic intensity levels. The positive effect of control was illustrated by means of simulation. To reduce the frequency of switching and to obtain a feasible control law the computed one-switch control policies were extended to so-called 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. Computing the criterion value function corresponding to a single policy requires the solution of an elliptic partial differential equation, which is quite involved. Carrying out the maximization to obtain the optimal policy simultaneously with the criterion value function is even more complicated. An interesting approach which seems to give good results is the one described in [15]. This will be investigated in the near future.

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. Also, this would nicely fit in the control set-up as described in 4.3. This would consist of considering four freeway sections in the neighbourhood of a signal station, carrying out filtering of the measurement data of the five associated measurements sites locally, as well as implementing the control logic. At the control centre the global coordination of the decisions of the separate signal stations would be taken care of.

As was mentioned in section 4.1, 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 2. It would consist of considering a microscopic freeway traffic model instead of the macroscopic ones of section 3. 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 in [9].

ACKNOWLEDGEMENTS

I would like to thank Jan van Schuppen, the members of the "gebruikerscommissie" and the Traffic Engineering Division of the Dutch Ministry of Transport for their valuable comments and support during the last year, and Paul de Zeeuw for his numerical advice and assistance in solving the partial differential equations I encountered.

These investigations are supported by the Technology Foundation.

REFERENCES

1. A.T. BARUCHA-REID (1960). *Elements of the theory of Markov processes and their applications*, McGraw-Hill, New York.
2. R.A. BECKER and J.M. CHAMBERS (1986). *S : an interactive environment for data analysis and graphics*, Wadsworth.
3. L. BREIMAN, R. LAWRENCE, D. GOODWIN, and B. BAILEY (1977). The statistical properties of freeway traffic flow, *Transportation Research*, 11, 221-228.
4. H. CHERNOFF and A.J. PETKAU (1978). Optimal control of a Brownian motion, *SIAM Journal of Applied Mathematics*, 34, 717-731.
5. 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.
6. D.R. COX (1962). *Renewal Theory*, Methuen & Co. Ltd., London.
7. M. CREMER (1979). *Der Verkehrsfluss auf Schnellstrassen*, Springer Verlag, Berlin.
8. M. CREMER and S. FLEISCHMANN (1987). Traffic responsive control of freeway networks by a state feedback approach, in *Proceedings of the 10th International Symposium on Transportation and Traffic Theory*, 357-376, ed. N.H. Gartner, N.H.M. Wilson, Elsevier, New York.
9. P. FERRARI (1988). The reliability of the motorway transport system, *Transportation Research*, 22B, 291-310.
10. R.Z. HASMINSKI (1980). *Stochastic stability of differential equations*, Sijthoff & Noordhoff, Alphen aan de Rijn.
11. L. ISAKSEN and H.J. PAYNE (1973). Freeway traffic surveillance and control, *Proceedings IEEE*, 61, 526-536.
12. 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.
13. F. Kozin (1969). A survey of stability of stochastic systems, *Automatica*, 5, 95-112.
14. J.L. de Kroes, P. Donk, and S.J. de Klein (1983). *Evaluatie van de externe effecten van het VerkeersSignaleringsSysteem voor Autosnelwegen 1983,* Report.
15. H.J. Kushner (1977). *Probability methods for approximations in stochastic control and for elliptic equations*, Academic Press, New York.
16. M.F.A.M. van Maarseveen (1982). *Application of martingales in stochastic systems theory - Surveillance and control of freeway traffic flow,* Thesis, Technical University of Twente, Enschede.
17. M.S. Mahmoud and S.Z. Eid (1988). Optimization problems of freeway traffic control problems, *Optimal Control Applications and Methods*, 9, 37-49.
18. P. Mandl (1968). *Analytical treatment of one-dimensional Markov processes*, Springer-Verlag, Berlin.
19. M. Papageorgiou (1983). *Applications of automatic control concepts to traffic flow modeling and control*, Springer Verlag, Berlin.
20. H.J. Payne (1971). Models of freeway traffic and control, in *Mathematical Models of Public Systems. Simulation Council Proceedings*, 51-61.
21. S.R. Pliska (1973). Single-person controlled diffusions with discounted costs, *Journal of Optimization Theory and Applications*, 12, 248-255.
22. H. Remeijn (1982). *The Dutch Motorway Control and Signalling System*, Traffic Engineering Division, Dutch Ministry of Transport, The Hague.
23. S.A. Smulders (1986). *Modelling and simulation of freeway traffic flow,* Report OS-R8615, Centre for Mathematics and Computer Science, Amsterdam.
24. S.A. Smulders (1988). *Filtering of freeway traffic flow,* Report OS-R8806, Centre for Mathematics and Computer Science, Amsterdam.
25. D. Tabak (1973). Control and optimization in transportation systems, in *Control and Dynamic Systems*, 345-434, ed. C.T. Leondes.
26. R.A. Tapia and J.R. Thompson (1978). *Nonparametric probability density estimation*, The John Hopkins University Press, Baltimore and London.
27. J.A.C. van Toorenburg (1980). *De stationaire verkeersstroom op de 2x2-strooks autosnelweg,* Report, Traffic Engineering Division, Dutch Ministry of Transport, The Hague.
28. J.A.C. van Toorenburg (1982). *De mens in het autosnelwegverkeer,* Report, Traffic Engineering Division, Dutch Ministry of Transport, The Hague.
29. J.A.C. van Toorenburg (1983). *Homogeniseren. Effekt van aangepaste adviessnelheid op de verkeersafwikkeling,* Report, Traffic Engineering Division, Dutch Ministry of Transport, The Hague.
30. H. Zackor (1972). Beurteilung verkehrsabhaengiger Geschwindigkeitsbeschraenkungen auf Autobahnen, *Forschung Strassenbau und Strassenverkehrstechnik*, 128.
31. P.M. de Zeeuw (1988). *Matrix dependent prolongations and restrictions in a blackbox multigrid solver,* Report NM-R8801, Centre for Mathematics and Computer Science, Amsterdam.