Development of an indicator for traffic flow stability with application to ramp metering

J.A.C.M. Elbers

14th November 2005

Development of an indicator for traffic flow stability with application to ramp metering

Dissertation Committee:

prof. dr. D. Feil	University of Twente, chairman/secretary
prof. dr. ir. E. C. van Berkum	University of Twente, promotor
prof. dr. ir. M. F. A. M. van Maarseveen	University of Twente
prof. dr. ir. B. van Arem	University of Twente
prof. ir. D. Lauwers	University of Gent
prof. dr. G. Wets	University of Hasselt
prof. ir. F.M. Sanders	Delft University of Technology

ISBN 90-5584-074-2

Typeset in LaTeX.

Cover pictures by www.trafficlinq.nl.

Printed by Febodruk BV, Enschede, The Netherlands.

DEVELOPMENT OF AN INDICATOR FOR TRAFFIC FLOW STABILITY WITH APPLICATION TO RAMP METERING

PROEFSCHRIFT

ter verkrijging van
de graad van doctor aan de Universiteit Twente,
op gezag van de Rector Magnificus,
prof. dr. W.H.M. Zijm,
volgens besluit van het College voor Promoties
in het openbaar te verdedigen
op woensdag 30 november 2005 om 13.15 uur

door

JOSEPHUS ARNOLDUS CORNELIS MARIA ELBERS

geboren op 28 juli 1974 te Beuningen.

Dit proefschrift is goedgekeurd door de promotor:

prof. dr. ir. E. C. van Berkum

Acknowledgements

The person responsible for inspiring me for this research is Jaap van Toorenburg. Although he never meant to send me back to University his way of 'putting the microscope on traffic flows' made me see there is much more to traffic than meets the eye on first sight. Once traffic data is gathered in detail, much can be analyzed.

The work reported in this thesis was supervised by prof.dr.ir. Eric van Berkum. He was always there to critically listen to my idea's or to help me through one of my motivation crises. Especially the final year I worked on this dissertation, part-time, was not easy.

I would like to thank Mark Zuidgeest for his help with LaTeX.

Last but not least I would like to thank the students I supervised; Taede Tillema, Natascha Kijk in de Vegte, Sjoerd Linders, Thijs Muizelaar and Martijn Ruijgers. Without their work this thesis would not be what it is now.

Especially Martijn Ruijgers' work was very helpful during the final stages of the development and the implementation of the stability indicator. Discussions with Martijn often ended with the conclusion that more research was to be done or results had to be improved. Therefore I am very grateful to him for his contribution to this thesis.

Contents

Acknowledgements **v**

List of Figures **xii**

List of Tables **xiii**

1 Introduction **1**

1.1 Motivation and objectives . . . 3

1.1.1 Motivation . . . 3

1.1.2 Research objectives . . . 5

1.2 Research setup and scope . . . 6

1.2.1 Contributions . . . 7

1.3 Thesis overview . . . 8

2 Introduction to traffic flows **11**

2.1 Definition of microscopic variables . . . 11

2.2 Definition of macroscopic variables . . . 12

2.3 Fundamental diagram . . . 13

2.4 Relations between microscopic and macroscopic variables . . . 16

2.5 Summary . . . 19

3 Traffic flow modelling **21**

3.1 Deterministic or stochastic . . . 21

3.2 Discrete or continuous . . . 22

3.3 Level of detail . . . 22

3.3.1 Microscopic traffic flow modelling . . . 22

3.3.2 Mesoscopic traffic flow modelling . . . 30

3.3.3 Macroscopic traffic flow modelling . . . 36

3.4 Modelling stability related issues . . . 39

3.4.1 Probabilistic description of traffic breakdown . . . 39

3.4.2 Queue location . . . 39

3.4.3 Propagation of shockwaves . . . 40

3.4.4 Discussion . . . 40

3.5 Summary . . . 41

4 Measuring traffic flows **43**
4.1 Traffic characteristics . . . 43
4.2 Detection methods . . . 44
4.2.1 Loop detection . . . 45
4.2.2 Pneumatic tube traffic detector . . . 46
4.2.3 Traffic cameras . . . 47
4.3 Data collection source, information and application . . . 48
4.3.1 Performance of data collection systems . . . 48
4.4 The Dutch motorway network and traffic flow measurements . 50
4.5 Data conditioning . . . 52
4.5.1 Missing data . . . 52
4.5.2 Data interpolation and extrapolation . . . 52
4.5.3 Adaptive smoothing method . . . 53
4.6 Discussion . . . 55

5 Merging traffic **57**
5.1 Introduction . . . 57
5.2 Merging at on ramps . . . 58
5.2.1 Headway distributions . . . 59
5.2.2 Critical gaps . . . 60
5.2.3 Queueing theory . . . 60
5.3 Summary . . . 65

6 Driver influences **67**
6.1 Driver characteristics . . . 67
6.1.1 General characteristics . . . 69
6.2 Motivation for driving behaviour . . . 72
6.2.1 Vehicle characteristics . . . 72
6.2.2 Road geometry . . . 73
6.2.3 Weather . . . 74
6.2.4 Traffic rules and dynamic traffic management . . . 75
6.3 Measures to homogenize traffic . . . 79
6.4 Summary . . . 82

7 Indicator for traffic stability **83**
7.1 Stability indicator based on traffic reliability . . . 84
7.1.1 Introduction . . . 84
7.1.2 Setup of the indicator . . . 84
7.2 Stability indicator based on traffic synchronization . . . 85
7.2.1 Introduction . . . 85
7.2.2 Setup of the indicator . . . 86
7.3 Stability indicator based on string stability . . . 87
7.3.1 Introduction . . . 87
7.3.2 Setup of the indicator . . . 90
7.3.3 Lookup graphs . . . 95
7.4 Summary . . . 98

8 Test of the stability indicators: case study 101
8.1 Operationalization of indicators . . . 101
8.1.1 Stability indicator based on reliability . . . 101
8.1.2 Stability indicator based on traffic synchronization . . . 102
8.1.3 Stability indicator based on string stability . . . 102
8.2 Data selection . . . 102
8.2.1 Study area . . . 103
8.2.2 Observations . . . 104
8.2.3 Selection of days . . . 105
8.3 Performance of the indicators . . . 107
8.4 Summary . . . 109

9 On ramp control and traffic flow stability 111
9.1 Introduction to ramp metering . . . 111
9.1.1 History of ramp metering . . . 111
9.1.2 Traffic process . . . 113
9.1.3 Metering strategies . . . 116
9.1.4 Some examples of existing ramp metering strategies . . . 121
9.2 Development of an alternative algorithm . . . 123
9.2.1 The stability indicator and merges . . . 123
9.2.2 Elaboration of Alinea . . . 124
9.2.3 The combined algorithm . . . 125
9.3 Simulation . . . 128
9.3.1 Model selection . . . 128
9.3.2 Input and network . . . 130
9.3.3 Performance indicators . . . 132
9.4 Results . . . 132
9.4.1 Results of the simulations . . . 132
9.4.2 Discussion . . . 134
9.5 Summary . . . 134

10 Conclusions and perspectives 137
10.1 Summary and conclusions . . . 137
10.2 Perspectives and future research . . . 140

Bibliography 142

A Headway distribution models 151

B Overview of the simulation network 157

Samenvatting 159

Nomenclature 163

About the author 165

List of Figures

1.1 Headway influence 3
1.2 Video observations 4
1.3 Structure of dissertation. 10

2.1 Microscopic characteristics 12
2.2 Fundamental diagram 14
2.3 Speed-density models 16
2.4 Speed-flow diagram of A22 data 17
2.5 Speed, flow and density as a function of space 17
2.6 Speed headway measurements. 18

3.1 Block diagram of follow the leader car-following model. 23
3.2 Propagation of speed disruption 25
3.3 Perceptual threshold. 27
3.4 Lane changing model structure 29
3.5 Lead and lag gap. 30
3.6 Headway models 35

4.1 Vehicle trajectories and measuring intervals in space and time. 44
4.2 Loop configuration. 45
4.3 Picture: double loop detection. 46
4.4 Picture: camera detection. 47
4.5 Data bank input and output. 48
4.6 MONICA traffic data collection system. 51
4.7 Visualization of 'adaptive smoothing method' 54

5.1 Drivers 'window of opportunity' 58
5.2 Time headway distribution for different hourly volumes. 59
5.3 The influence of merge lane length on merging and gap acceptance 61
5.4 The influence of merge lane length on merging and the resulting time-to-collision. 62
5.5 Queueing system with unlimited waiting capacity. 62
5.6 Erlang-k queueing system 64

6.1 Influences of drivers. 68

6.2 Observation of relative movements. 70
6.3 Influences of DTM on the traffic system. 75
6.4 Signalling and ramp metering 78
6.5 Study area of different types of homogeneous traffic. 79
6.6 Headway distributions 81

7.1 Stability indicator based on traffic reliability 85
7.2 Stability lookup table for synchronized traffic. 86
7.3 Stability indicator based on traffic synchronization 87
7.4 Different levels of local stability 88
7.5 Setup of string stability indicator. 91
7.6 Frequency of gaps 93
7.7 Propagation of speed disruption 94
7.8 Measuring speed effects 95
7.9 String stability indicator for a platoon of vehicles. 96
7.10 Effect of disruptions on the stability indicator 97
7.11 Stability indicator per lane. 99

8.1 Overview of operationalization of the indicators. 103
8.2 Study area between Utrecht and Amsterdam. 104
8.3 Examples of MARE and RCU-data. 105
8.4 Visualization of MARE-data 106
8.5 Speed-flow data from the left lane 107
8.6 Measurement locations 108

9.1 Basic control scheme for motorway control tools 113
9.2 Scheme of fixed-time strategy 117
9.3 Scheme of reactive strategy 117
9.4 Scheme of pro-active strategy 118
9.5 Combined scheme of feed forward and feedback strategy 119
9.6 Alinea control scheme. 123
9.7 Indication of correction of stability indicator on Alinea. 124
9.8 Setup of the combined algorithm. 126
9.9 Buildup of the process time 127
9.10 Stability in time 128
9.11 Correction function. 129
9.12 Research network. 131
9.13 Demand levels 132

A.1 Headway models 156

B.1 Network in Aimsun 158

List of Tables

2.1 Speed density relations 16

4.1 Usability of detection systems 50

6.1 Composition of traffic on Dutch motorways 73
6.2 Results of homogenization tests. 80

8.1 Correlations between stability indicators and congestion. 108

9.1 Results of Alinea. 133
9.2 Results of the combined algorithm. 133

Chapter 1

Introduction

Traffic management has grown from a tool to optimize traffic conditions on links that suffer congestion (e.g. maximizing capacity) to a mean to upgrade robustness and ensure travel times through traffic networks.

In general there are two ways to improve reliability of a traffic network: enhance reliability on the links or create more alternative links. This thesis is not about creating extra reliability (and capacity) to networks by building new links, but on reliability on existing links.

The basis of motorway reliability is predictability: when there is enough certainty that today's capacity of a bottleneck is that of yesterday and the day before the system can be called reliable. Also when today's capacity is only a fraction of yesterday's the system is reliable, as long as it is known in advance. Predicting the traffic process in bottlenecks in a traffic system is not straightforward: events like accidents and bad weather influence bottleneck performance significantly and capacity fluctuates. Prediction can be made easier when processes are stable. Although the term *stability* is used frequently in traffic reports researchers are not unambiguous in the meaning of the word.

The Highway Capacity Manual 2000 does also not contain a definition of (traffic flow) stability. The Oxford dictionary gives the following description of *stability*: 'the quality or state of being steady and not changing or being distributed in any way (= the quality of being stable)'.

This dissertation thesis addresses the need for such a definition, how traffic stability can be measured and used to upgrade Dynamic Traffic Management (DTM) and make it more feasible for it's task: providing a predictable traffic system.

In this chapter the background, the motivation and the objectives of the thesis as well as the scientific and practical contributions of this doctorate studies are described. This research is in the field of traffic dynamics, traffic description, dynamic traffic management, and concentrates on traffic flow stability.

In section 1.1.1 the motivation and research questions of this research are discussed. In section 1.1.2 the research objectives and problems are stated. In section 1.2 the research is outlined and the scientific contribution is discussed. In the final part (section 1.3) of this introduction the subjects which are covered in each chapter of this thesis are outlined.

Background and context

As traffic jams and DTM become more and more common on large stretches of motorways and in traffic networks, the possibility to monitor, understand and influence the breakdown phenomenon on motorways have increased (Lorenz and Elefteriadou, 2001), (Kerner and Rehborn, 1996), (Daganzo, 1981), (Richards, 1956).

Explanatory variables for the variation of capacity in both bottlenecks and on stretches of motorway are found in for instance weather, traffic composition and gradient (Transportation Research Board, 2000). But does this make motorway capacity a deterministic process of external conditions and traffic flow characteristics?

Personal experience [1] led to the conviction that this is not the case. Under apparently the same traffic conditions breakdown sometimes appeared and sometimes not.

The belief of non-deterministic capacity is also displayed in (Lorenz and Elefteriadou, 2001), (AVV, 2000), and (AVV, 1993). In these researches traffic capacity is unravelled into many explanatory variables, but in all cases the remaining distribution of capacity is treated stochastically.

In order to further narrow down the stochastic characteristics of capacity traffic flow stability as an explanatory variable of the breakdown phenomenon in bottlenecks will be focused upon. For the time being (chapter 7 will give more detail) the following definition of traffic stability is used.

Traffic stability is the capacity of a traffic flow to handle disruptions and prevent breakdown. When a flow has more 'capacity to handle disruptions without breaking down' this flow is called more stable.

In this definition a lack of traffic stability results in an increased traffic breakdown probability. For completeness a definition of breakdown is given:

[1] Research on motorway data showed differences in capacity of two motorway stretches, which were identical in alignment, weather conditions, number of lanes, curvature, traffic composition (percentage of heavy vehicle), and distribution over lanes.

Traffic breakdown is the phenomenon that on a motorway the flow rises to the level of capacity and the infrastructure together with the skills of the drivers are no longer able to maintain a high level of service and the speed of traffic drops significantly.

The stated definition of traffic flow stability can be illustrated by comparing two sets of vehicles (figure 1.1).

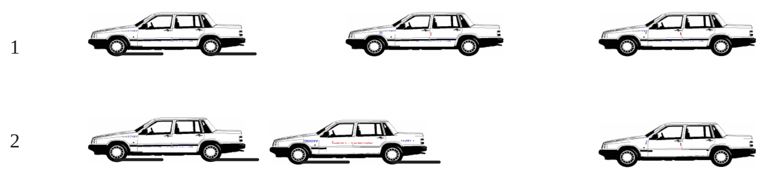

Figure 1.1: The following vehicle in set 1 is less influenced by speed fluctuations than the following vehicle of set 2, because of the distance between the vehicles.

Set 1, which consists of a leader, a follower at a safe distance and a third vehicle, also at a safe distance. Set 2 consists of a leader, a follower that is tailgating and a third vehicle that is driving at more than a safe distance. If the speed of the leader of the first set fluctuates the follower will not be bothered and will hardly have to adjust it's speed. If the speed of the leader of the second set of vehicles fluctuates, the decrease in speed of the leader is followed by a mandatory decrease (in order not to collide with the first vehicle) of the second vehicle, and perhaps even the third vehicle needs to decelerate, because of the limited acceleration capabilities of the first and second vehicle. The first set of vehicles is called more stable, because the disruption in speed of the first vehicle is handled, with smaller probability of propagation through the traffic flow and therefore less probability to result in breakdown of the traffic flow, than the second set of vehicles. So the stability of both sets is not equal, while at an aggregate level the two sets of vehicles are the same.

1.1 Motivation and objectives

1.1.1 Motivation

Travel time on the motorway can vary a lot. When traffic is limited the free travel time for a motorway section can be much less compared to the travel time on the same stretch of road during congestion. Reliability of travel times for the user of the motorway network therefore depends largely on the answer to the question: 'can I expect any traffic jams'. Understanding the breakdown phenomenon, which leads to traffic jams, and being able to influence this phenomenon is therefore a major aspect of motorway reliability.

The initial idea for this research came during a study which was aimed to monitor the behaviour of capacity (Van Toorenburg and Elbers, 2000) on motorways. In static bottlenecks and in shockwaves the capacity in time and place were monitored, together with possible influential characteristics of the traffic flow and external influences. Without going into detail on the methods used in that research; besides the influence of the obvious influences as the weather, the percentage of heavy vehicles there remained differences in capacity which appeared to be due to the distribution of the vehicles (clustering).

A second motivation formed the visual inspections of a motorway bottleneck during an evening peak. A motorway merge in the form of an on ramp was inspected (an on ramp with permanent feed, because a traffic jam formed on both the main road and the on ramp). The traffic in this bottleneck got congested and started moving a lot of times during the evening rush. Although the traffic on the on ramp came from a motorway and not from a crossing with a traffic light the flow was clustered behind trucks before it merged into the main flow (trucks had a very large headway and other vehicles were following impatiently maintaining very short headways behind them). Figure 1.2 shows three area's, which could be easily recognized as starting points of congestion (and from which shockwaves started travelling upstream). The clustering of the vehicles merging and the clusters of vehicles which were created on the left lane prior to the merging area (area 1 in figure 1.2) both had a large influence on the initiation of traffic jams.

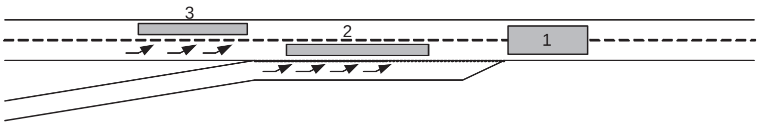

Figure 1.2: The three areas where congestion initiates; 1 when the flow relaxes, 2 when vehicles form the on ramp merge, and 3 when vehicles on the right lane anticipate on the merging traffic by changing into the left lane.

These two experiences led to the conviction that the (microscopic) distribution of vehicles leads to different macroscopic phenomena and characteristics, e.g. capacity.

The capacity of roads is influenced by a lot of factors; e.g. alignment, traffic surrounding as weather and light, and traffic management (Transportation Research Board, 2000). All these influences and their (mostly non-linear) interactions make it difficult to state road capacity as an absolute number. It is easier and more appropriate, to define capacity as a stochastic phenomenon (Lorenz and Elefteriadou, 2001), based on measurements. This does not mean that it is not possible to measure the effects of the influences mentioned above on capacity. Instead of one distribution of the capacity of a bottleneck it can for example be split into two: one at rainy weather conditions and one without any precipitation (AVV, 1999a). In the same way the distribution of capacity

can be split up in a number of situations, as with or without daylight, with or without trucks, morning or evening rush hour, etc.
To explain differences in the breakdown characteristics of traffic flows it was decided to define and measure traffic stability, based on characteristics of traffic reliability, traffic synchronization, and the distribution of traffic (clustering) per lane (string stability). (Definitions of traffic reliability, synchronization and string stability can be found in chapter 7).

It is thought that these three existing theories, related to the traffic breakdown give better results compared to other research approaches into bottleneck capacity. As said, macroscopic approaches and describing capacity as a stochastic process overlooks microscopic traffic phenomena, that underlie the macroscopic behaviour.

Often microscopic traffic information is lot due to the calculation technique used. The product-limit method needs averages in time to calculate capacity and that way microscopic information is lost. The prediction of congestion using cellular automaton is always claimed to be a method, which does not overlook any phenomenon. But if relations are already known they can be put in by pre-processing. That is also the case in the method that is chosen for this research. The breakdown phenomenon can be described in great detail through macroscopic characteristics (flow vs. capacity) but the underlying microscopic characteristics of traffic can explain behaviour which up to now with the macroscopic methods is left unknown.

1.1.2 Research objectives

The main objective of this research is to be able to give better explanations and get better predictions of the onset of traffic jams using data from existing measuring techniques. Macroscopic characteristics, as speed, flow and density do not give enough explanation why in some cases the flow collapses and under apparently the same circumstances survives (Lorenz and Elefteriadou, 2001), (Akcelik and Chung, 1994), (Dijker et al., 1997), (Daganzo, 2002). Distributions of capacity are empirical descriptions of the phenomenon and can be used to explain the behaviour: when two distributions of capacity differ and the only variable is the amount of rain, the effect of this rain on the capacity can be measured. A lot of effects on capacity can be measured through this single-criteria analysis (e.g. weather type or the percentage of trucks). Because capacity differences were measured and the explanatory variables fail to give an answer (first motivation for this research) the distribution of the traffic within the up to now used measurements is checked in more detail. It is not strange that traffic stability (based on microscopic characteristics) is considered.

A second objective is to research how existing theories can be applied to develop measures to better predict congestion, operationalized as indicators.

Related to the second objective is the third objective: application of the measurement tools (indicators) to better predict congestion and test the applicability or usability of the indicator by introducing an alternative ramp metering strategy.

1.2 Research setup and scope

The main motivation is to research for explanatory characteristics of traffic and eventually create a tool which can be used to better predict the onset of congestion and can create better algorithms for dynamic traffic management.

To limit the research effort only traffic on motorways is considered. The definition of a motorway according to the Highway Capacity Manual 2000 is:

A motorway is a multi-lane, divided highway with a minimum of two lanes for the exclusive use of (motorized) traffic in each direction and full control of access without traffic interruptions. There are no signalized or stop-controlled at-grade intersections, and direct access to and from adjacent property is not permitted. Access to and from the motorway is limited to ramp locations.

This limitation of the research does not mean that the theories that are applied and the indicators (operationalizations of the theories) cannot be used for interrupted flows; just that they are not developed and not tested with that eyemark.

Although the term stability is used and defined differently by different researchers ((Lorenz and Elefteriadou, 2001), (Kerner and Rehborn, 1996) or (Ferrari, 1989)) on several occasions in research, three measurable indicators of stability are defined and tested on how good they perform in predicting congestion.

The results of the indicator tests are used to select the best indicator for predicting the onset of traffic jams and this indicator is embedded into a new ramp metering strategy. This algorithm is compared to the Alinea ramp metering strategy. Although more DTM measure protocols can be improved using the indicator, the one which is most obvious related to traffic disruptions (merging vehicles) is chosen.

1.2.1 Contributions

In this paragraph the main contributions of this thesis are listed.

Summary of contributions

Next to the research of how existing theories of the breakdown phenomenon of traffic can be used to predict breakdown and estimate capacity two concrete products are developed within this research:

1. A traffic stability indicator, which attributes to a better prediction of breakdown. Individual vehicle data is used as input and a single number indicates the risk of the traffic flow breaking down.

2. An alternative ramp metering strategy, based on the existing Alinea strategy, together with the developed stability indicator, which is a significant improvement to the Alinea ramp metering algorithm.

Theoretical relevance

The research contains the following, theoretically relevant results.

1. The existing theories on traffic synchronization (Kerner and Rehborn, 1996), and traffic reliability (Ferrari, 1989), both dealing with the breakdown phenomenon are outperformed by the theory on string stability when used to predict the onset of congestion. Both cannot cannot be used to explain or predict the onset of congestion.

2. Both macroscopic and microscopic traffic characteristics are main phenomena related to breakdown at on ramps. This explains the success of the alternative ramp metering strategy. (The macroscopic based Alinea algorithm is enhanced with a, microscopic based, stability indicator).

Practical relevance

It may be clear that the alternative ramp metering strategy can be used for metering improvement. As traffic grows and space gets scarce the need for DTM grows and improvements are necessary; up to today DTM measures have their own individual goal and means. Enhancing DTM measures with a stability indicator can give them a common goal. For instance homogenize measures and ramp metering have contradicting goals; the first wants to make a large gap between vehicles smaller and the seconds might want to use it for a merging vehicle. When conflicts like this are taken care of the effect of dynamic traffic management will be able to improve.

1.3 Thesis overview

This section presents an overview of the chapters in this thesis, which is also illustrated in figure 1.3.

As this thesis is about predicting the onset of traffic jams, the dynamics of motorway traffic need to be introduced. This is done in chapter 2, together with microscopic and macroscopic definitions of traffic characteristics. Also the fundamental diagram and what can be derived from it is described in this chapter.

Different ways to describe or model traffic flow phenomena are presented in chapter 3. What a traffic flow model is, why traffic is modeled, aspects of capacity and the breakdown phenomenon by modelling traffic. Both microscopic and macroscopic modelling approaches are discussed.

Chapter 4 describes how data from traffic flows may be collected and processed. Not all traffic characteristics mentioned in chapter 2 can be measured form the roadside, some have to be calculated. Next to measurements and derived or calculated measurements, interpolation between measurements will prove to be of great importance to get the data that is needed for the development of traffic stability indicators; both for getting richer (more complete) data sets, and to condition data which may be corrupted or false. The collection of traffic flow data on Dutch motorways is discussed.

Chapter 5 focuses on three items. First the window of opportunity of a driver to drive safe, or to prevent a collision with it's predecessor, is studied. Then microscopic characteristics which can be derived from the fundamental diagram are researched; the underlying relations are unravelled and recent mesoscopic findings are discussed. The focus is on the relationships between car spacing, gap distribution and the characteristics speed, flow and density, as they will prove to be relevant for calculating string stability (which is covered in chapter 7). Finally the average queue length and waiting time a vehicle will encounter when entering a congested motorway is deduced.

Chapter 6 is the final introductional chapter. It describes how external influences influence the drivers in a complex system, with relations which are usually not linear and often not independent of each other. The relation between the flow and road geometry, vehicle and driver characteristics, weather and dynamic traffic management are described, as well as their relation to traffic flow stability. Finally microscopic modelling demonstrates that homogenize measures do not always lead to a capacity increase and better merging conditions.

In chapter 7 the concept of traffic stability is dealt with; what is traffic stability and what is the difference between stability and breakdown probability?

The influences and empirical relations described in chapters 5 and 6, together with existing theories on traffic flow reliability (Ferrari, 1988), synchronization of traffic flows (Kerner and Rehborn, 1996) and string stability, are used to develop three stability indicators. Inter-vehicular spacing appears to play an important role in traffic stability and in the prediction of traffic breakdown.

In chapter 8 tests of and comparisons between the indicators are conducted: the three indicators are compared to traffic data and their ability to predict traffic breakdown is tested.

In chapter 9 the stability indicator is incorporated in an enhanced metering strategy. The new algorithm is compared to the existing Alinea algorithm using microsimulation in Aimsun2. The results are positive and the implementation of the enhanced algorithm in real life looks promising.

Finally, in chapter 10 conclusions are drawn and perspectives are discussed.

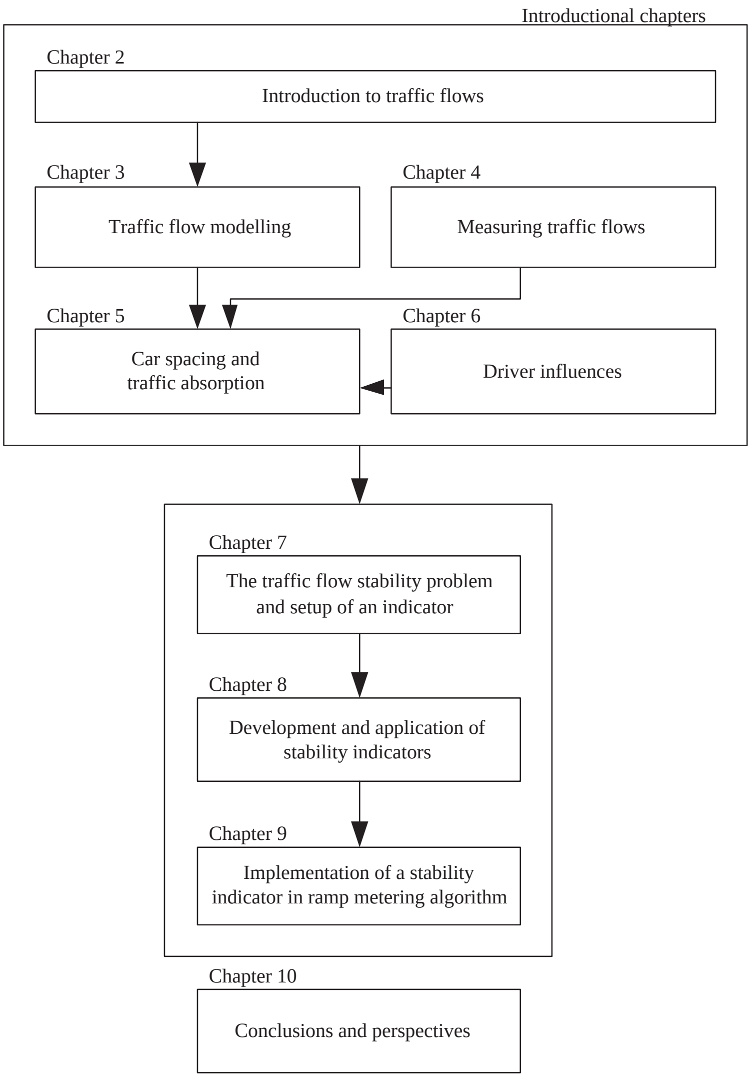

Figure 1.3: Structure of dissertation.

Chapter 2

Introduction to traffic flows

In this chapter the conceptual framework for all the descriptions, hypothesis and models of traffic flows to be described in this thesis is set up. Although traffic consists of individual vehicles it is not only studied as an interaction process at an individual level (microscopically), but also as a flow (macroscopically). Microscopic and macroscopic variables used for traffic descriptions and calculations are presented. Also the fundamental diagram, an illustration of relations between the key macroscopic traffic flow characteristics, is presented.

2.1 Definition of microscopic variables

In microscopic traffic descriptions each vehicle is considered separately. A vehicle's trajectory over a road can be described as a pair $[\hat{x}(t), \hat{y}(t)]$ of points in time in two directions. $[\hat{x}(t)]$ is the direction tangential to the road and $[\hat{y}(t)]$ is the direction at a right angle to the road. The vehicle's speed ($\hat{u}$), acceleration ($\hat{a}$), and jerk ($\hat{j}$) at time t can be calculated by deriving the function of location in the time once, twice or three times (e.g. $\hat{a}(t) = \frac{\mathrm{d}\hat{u}(t)}{\mathrm{d}t} = \frac{\mathrm{d}^2[\hat{x}(t),\hat{y}(t)]}{\mathrm{d}t^2}$).

Length ($\hat{l}$) of a vehicle is the length in meters, measured from bumper to bumper. The 'electronic' length, measured through inductive loops is shorter, because it needs a trigger which is not present at the far front and end of a vehicle, because they don't contain enough metal (e.g. plastic bumpers).

Distance ($\hat{d}$) between two consecutive vehicles is the number of meters between the leading vehicle's rear bumper and the following vehicle's front bumper.

Space ($\hat{s}$) of a vehicle consists of it's own length ($\hat{l}$) and the distance ($\hat{d}$) from it's front bumper to the rear bumper of it's predecessor.

Occupancy ($\hat{o}$) or clearance time is the time in seconds a vehicle needs to pass a cross sectional point.

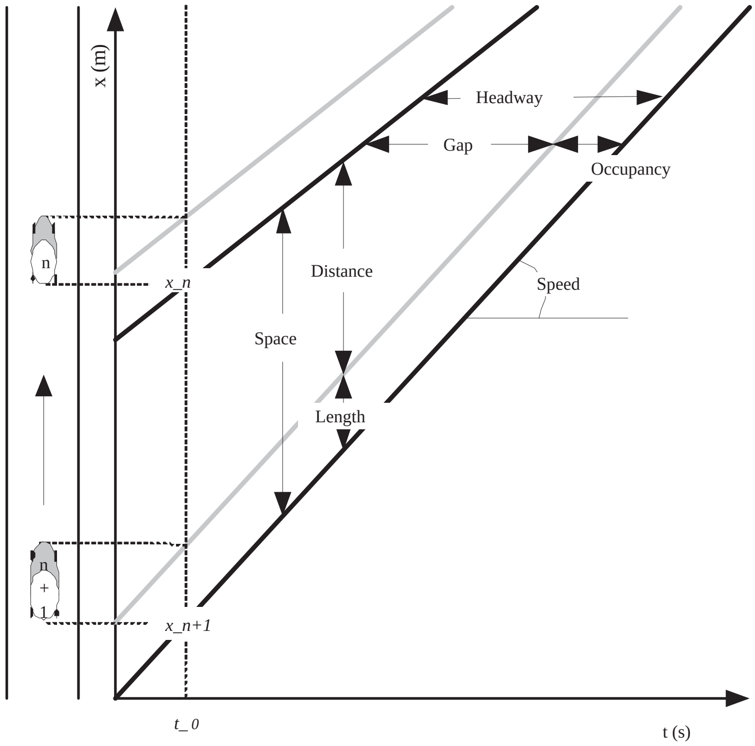

Figure 2.1: A road with two vehicles (n and $n+1$) and microscopic characteristics of vehicle $n+1$.

Gap ($\hat{g}$) between two vehicles is measured at a cross section and is the time in seconds between the passage of the first vehicle's rear bumper and the passage time of the next vehicle's front bumper.

Headway ($\hat{h}$) of a vehicle consist of the gap and to it's predecessor and it's own occupancy (in seconds).

If trajectories ($[\hat{x}(t), \hat{y}(t)] - diagrams$) of vehicles are available all the microscopic variables mentioned above can be calculated, but because of limited measurements this is not always the case (see chapter 4).

2.2 Definition of macroscopic variables

In macroscopic traffic descriptions the traffic flow is treated as if it were continuous and the separate vehicles are not distinguished. The main variables are flow (q), density (k) and speed (u). Not all three variables can be calculated over intervals in time T or intervals in space X, which create time or space related variables:

Flow (q) is the number of vehicles which crosses a certain point over a certain time interval. The flow (also called *volume* or *flux*) is measured over a period T and is denoted as: $q = \frac{n}{T}$, in which n is the number of passed vehicles.

Density (k) is the number of vehicles on a certain road section X at a certain point in time and is denoted as: $k = \frac{n}{X}$ veh/m, in which n is the number of vehicles counted on the road section.

Speed (u) can be calculated over time and space. The *local* speed (u_l) is the average of the speeds of the vehicles which cross a certain point in space over a certain period in time. If n vehicles cross this point in space than the local average is: $u_l = \frac{1}{n}\sum_{i=1}^{n} \hat{u}_i$ (m/s), in which $\hat{u}_i$ is the speed of vehicle i. The *instantaneous* average speed (u_m) is the average speed of all vehicles over a certain road section on a single moment in time. If n vehicles are on a road section X, the instantaneous speed average is: $u_m = \frac{1}{n}\sum_{i=1}^{n} \hat{u}_i$, in which $\hat{u}_i$ is the speed of vehicle i on the road section.

Almost all available traffic measurements are cross-sectional. The reason for this is not that these measurements form a 'richer' data set (from which more characteristics can be calculated), but the effort needed to get an instantaneous average speed is much higher (e.g. a helicopter has to be sent up or camera's have to get mounted to collect video material). An instantaneous average speed can be estimated by averaging the local measured speeds as in: $\tilde{u}_m = \frac{1}{n}\sum_{i=1}^{n} 1/\hat{u}_i$. This may only be done if no large speed fluctuations of vehicles take place over the road that is under investigation. In that case compensation of the overestimation of slow drivers has to be made (van Lint, 2004). This should not be mistaken for the difference between local average speed and instantaneous average speed; this difference is explained above. The fluctuations of a single vehicles speed over a road section however is only measured on cross sections and the fluctuations in between should not be to much.

One of the objectives in this research is the development of tools and theories which are widely applicable; therefore both the stability indicator (chapter 7) and the development of an improved on ramp algorithm (chapter 9) both deal with cross-sectional speed measurements.

The relationships between the macroscopic characteristics flow, density and speed are further discussed in the next section.

2.3 Fundamental diagram

Between the macroscopic variables flow q(veh/s), density k(veh/m), and speed u(m/s) (discussed in section 2.2) on a homogeneous state section (no ramps to enter or leave the motorway, no change in the number of lanes or speed limit, and no change in speed, density or flow over this section), during a fixed time period the continuum relation holds:

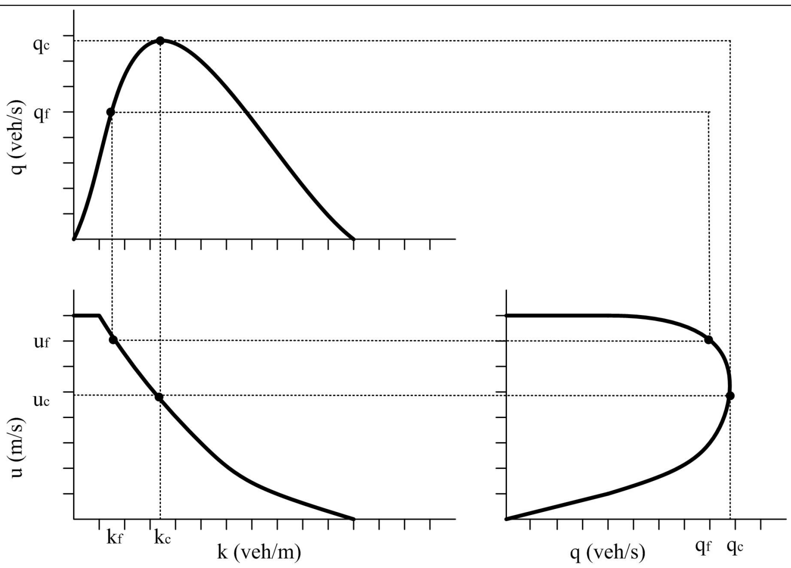

Figure 2.2: Speed-volume-density curves in the fundamental diagram.

$$q = ku \tag{2.1}$$

All variables hold at a certain time at a certain place, but those are left out for convenience purposes. Only when needed they will be used. The (x, t) is left out and reduces the macroscopic traffic flow process to two independent variables (the third can be found through calculation). Between these two there exists an empirical relationship. The relationship was first mathematically described by Greenshields (1934). Because the relation is usual represented in a diagram, it is called the *fundamental diagram.*

The diagram in figure 2.2 shows a possible form of the fundamental diagram. In the figure a point in the free flow regime (k_f, q_f, u_f) and a critical point (k_c, q_c, u_c) are marked. The critical point is the point, that if any more traffic is added to the flow, the flow will collapse and get into the jammed state. The fundamental diagram is not static, it is influenced by road, driver and vehicle characteristics. Some important influences are (in chapter 6 the phenomena behind these influences are described):

- The type of road (Maximum flow (the level of flow at which traffic can travel over a section without breaking down) depends largely on alignment, road width, traffic measures, slope, etc.).

- The composition of the traffic flow (e.g the percentage of passenger cars and heavy vehicles; heavy vehicles have different vehicle characteristics which reduce the maximum flow).

- Driver characteristics (commuters show different driving behaviour (e.g. shorter headway) than long distance traffic).
- Light conditions (darkness or daylight).
- Weather conditions (precipitation).

Two fundamental diagrams may only be compared if they are collected under comparable conditions. The internal and external influences on both microscopic and macroscopic traffic characteristics are quantified in chapter 6.

Speed-density relationship

One of the earliest models of the fundamental diagram is described in Greenshields (1934). The model assumes a linear relationship between speed and density between the point u_f of free flow and k_m of maximum (jammed) density. The equations that formulate this model are:

$$u_m(k) = u_f(1 - k/k_m) \quad \text{(linear)} \tag{2.2}$$

Use $q = ku$ to get the other two relations:

$$q(k) = ku_0 - k^2 u_f/k_m \quad \text{(parabola)} \tag{2.3}$$

and

$$q(u) = u_m k_m - (k_m/u_f)u_m^2 \quad \text{(parabola)} \tag{2.4}$$

The Greenshields model is still used frequently, because of the easy form and calculation possibilities. However, it has some clear omissions, like: the speed goes straight down when density increases from zero, and the density at capacity is far less from half of the density at jammed traffic. These model characteristics do not hold in practice.

In a homogeneous and non-fluctuating steady flow at low concentrations, the average speed u is usually high. Vehicles can reach their desired speed. As the concentration or density k increases, the degree of interaction between vehicles increases and u decreases. This suggests that u is a monotonic decreasing function of k as already seen in figure 2.2 and in equation (2.2). The shape of this relation has been the topic of numerous research occasions over more than seven decades((Greenshields, 1934), (Drew, 1968), (Banks, 1989), and many others). The exact shape of the speed-density relationship, according to several researchers is shown in figure 2.3. The functions are found in table 2.1. The most left part corresponds to free flow traffic conditions. Flow increases while density is increasing up to a maximum, critical concentration k_c. This

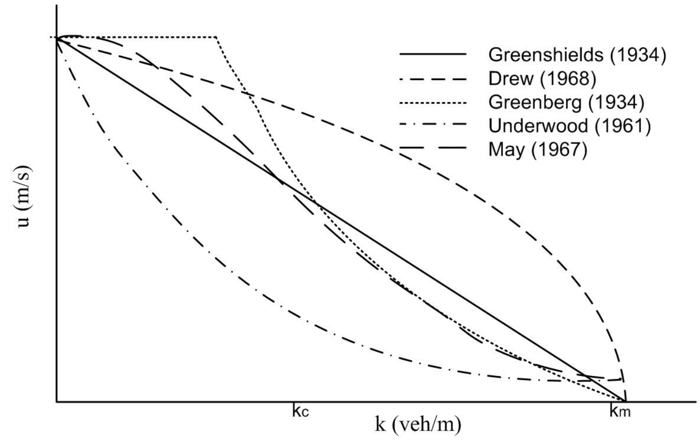

Figure 2.3: Main speed density relations (graphs) according to researchers.

maximum flow is the capacity of the road section. If the concentration increases beyond k_c the flow will start decreasing, vehicles will interact more and the stream will get congested. For one flow there are two different concentrations. Zero flow can mean no traffic or total gridlock (k_m).

Table 2.1: Main speed density relations (functions) according to researchers.

Model	Equation
Greenshields	$u_m = u_f(1 - k/k_m)$
Drew	$u_m = u_f[1 - (k/k_m)^{1/2}]$
Greenberg	$u_m = u_c\ln(k_m/k)$
Underwood	$u_m = u_f\exp(-k/k_c)$
May	$u_m = u_f\exp[-1/2(-k/k_c)^2]$

In which u_m: instantaneous speed, u_c: critical speed, u_f: free speed, k: density, k_c: critical density, k_m: maximum density.

As mentioned, fundamental diagrams are often derived from road side measurements. Speed, flow and occupancy are not as smooth as in theory but nevertheless curves (polynomial, logarithmic or exponential, see table 2.1) are fitted to data as in figure 2.4. When macroscopic relations between variables in for instance a fundamental diagram are presented one should always keep in mind that relations in practice are not as clear the graph suggests.

2.4 Relations between microscopic and macroscopic variables

In this section the relations between microscopic and macroscopic variables is showed. A fundamental diagram based on the following traffic behaviour

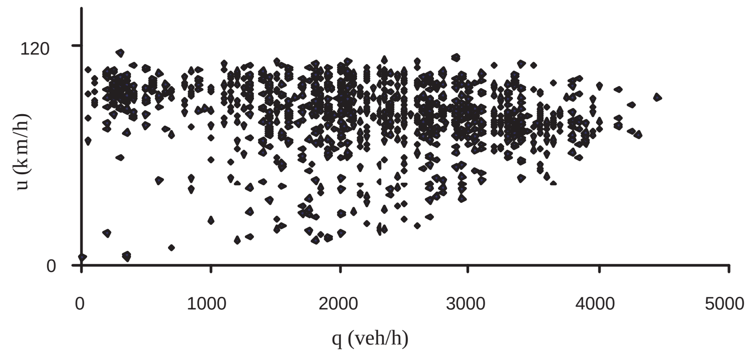

Figure 2.4: Observations in a speed-flow diagram, 1 minute averages, of the A22 motorway, The Netherlands.

is constructed: the shape of the (2-dimensional) flow-speed relation in the fundamental diagram arises from the fact that drivers tend to increase their space as they increase their speed. If spacing is assumed to be proportional to speed to some power n, the following holds:

$$\hat{s}_i = \hat{l}_i + p_1 \hat{u}_i^n \tag{2.5}$$

in which $\hat{s}_i$ equals the space of vehicle i and $\hat{l}_i$ the length of the vehicle and $n > 1$. Speed, flow and density can be expressed in terms of spacing $\bar{s} = \sum_i^m \hat{s}_i/m$ and average vehicle length $\bar{l} = \sum_i^m \hat{l}_i/m$:

$$k = \frac{1}{\bar{s}} \qquad u = \sqrt[n]{((\bar{s} - \bar{l})/p_1)} \qquad q = ku = \sqrt[n]{(\bar{s} - \bar{l})/p_1)}/\bar{s} \tag{2.6}$$

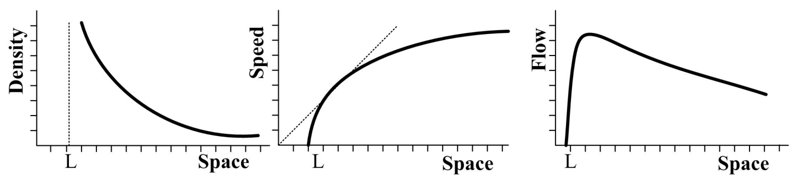

Figure 2.5: Speed, flow and density as a function of space (equation (2.6)).

The most right part of the fundamental diagram relation between flow and speed is for free traffic in which the speeds of the separate vehicles are not influenced by each other.

Space-density relation

The density is the reciprocal to the average space.

Space-speed relation

The assumption $\hat{s}_i = \hat{l}_i + p_1 \hat{u}_i^n$ leads to $u = \sqrt[n]{((\bar{s} - \bar{l})/p_1)}$, which implies a parabolic relationship.
The 'theoretical' assumption does not always match empirical findings. Holland (1998) and Banks (2003) showed empirically that spacing is approximately constant with respect to speed over most of the congested flow regime. Only at large headways the gradient of velocity against spacing is small. See figure 2.6.

This curve was verified with data from the A2 motorway near Amsterdam, see figure 2.6. It is clear that the empirical findings of Holland (1998) and Banks (2003) and the theoretical relation, as seen in figure 2.5, do not hold for the fast lane of the A2 motorway. The motorway data shows a large concentration of points at short headways and high speeds.

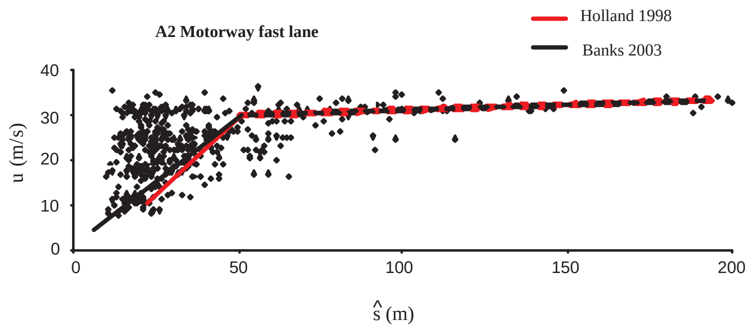

Figure 2.6: Speed headway measurements on the A2 motorway compared to empirical space-speed relationships found by Holland (1998)and Banks (2003). It shows that speed headway relations are location dependent (the A2 data clearly does not fit one of the lines).

Measurements vary clearly. The data shows large variation compared to the two curves of other locations. The fact that most points are neither on the theoretical curve, nor on curves fitted on the data of other locations shows that data aggregation, a very technical and easy process, also plays a role in analyzing traffic data.

Space-flow relation

Equation (2.6) shows that given the assumption about the distance ($\hat{d}_i = p_1 \hat{u}_i^n$) the flow can be formulated as a function of space as: $q = \sqrt[n]{(\bar{s} - \bar{l})/p_1)}/\bar{s}$ (see figure 2.5).

Headway-flow relations are empirically easier to measure and can be recalculated into space-flow relationships by realizing that $\hat{s}_i = \hat{h}_i \hat{u}_i$. Distributions of time headway have been researched extensive and the relation between level of flow and the distribution of headways is discussed in section 5.2.1.

From a single assumption about the spacing of the vehicles macroscopic characteristics as speed, flow and density which comply with traffic behaviour can be deduced.

2.5 Summary

In this chapter microscopic and macroscopic definitions of traffic characteristics, and the relations between them were introduced. They will be used later: in the description of traffic models (chapter 3), when the decision on what characteristics to measure is made (chapter 4) and when the influence of different traffic characteristics on traffic flow stability is discussed (chapters 5 to 7). In chapter 4, when the decision what to measure is made the relations between microscopic and macroscopic traffic characteristics are used.

Microscopic characteristics give information on every single vehicle, whereas macroscopic variables give information on the traffic flow. The relations between the macroscopic variables speed, flow and density can be visualized in a fundamental diagram.

Chapter 3

Traffic flow modelling

A traffic flow model describes traffic operations on a road. It uses mathematical formulas to calculate the operations. The formulas in traffic models can be based on mechanisms that take place between drivers, the road, the vehicle, and other influences (chapter 6). They attempt to describe the processes in the traffic flow.

In this chapter an overview of traffic flow modelling is given. The models can be classified based upon their level of detail and the common distinctions (discrete vs. continuous and deterministic vs. stochastic).

After the overview (in sections 3.1 to 3.3) of the types of traffic flow models different aspects of the breakdown phenomenon will be discussed and what different models comply to the models demands.

Traffic flow modelling started half a century ago when Lighthill and Witham (1955) presented a model, based on the analogy between vehicles in a traffic flow and particles in a fluid. Ever since researchers have improved their modelling techniques and their models.

3.1 Deterministic or stochastic

In a deterministic model there is a deterministic relation between the input, and the output of the model. So, if a traffic flow is simulated using a deterministic model twice, with the same initial data, the output of both simulations is the same. An examples of a deterministic model is METANET (Kotsialos et al., 1998).

A stochastic traffic model contains at least one stochastic variable, which implies that two simulations, using the same input, may give different output, as a result of the stochastic variable. A stochastic variable can be characterized by it's probability density function (p.d.f.), which may be continuous or discrete (histogram).

An examples of a stochastic variable in a model is the patience level in the microscopic simulation program Aimsun2. During the simulation samples are taken from a distribution of the patience level to be used in the simulation. The patience level influences variables as the minimum headway and the gap acceptance, which have a direct effect on the driving behaviour. In different 'runs' of the model, different samples are taken for stochastic variables. Therefor, stochastic models need to be simulated repeatedly and the results need to be averaged in order to be able to draw conclusions.

3.2 Discrete or continuous

As in all models also in traffic modelling, input and output variables (usually in time and space) can be discrete or continuous.

Making variables discrete improves handling and speed of the model: when calculations get complex because of large networks or complex descriptions of driving behaviour the amount of computer capacity is limited when variables are discrete.

3.3 Level of detail

3.3.1 Microscopic traffic flow modelling

In microscopic models the space-time behaviour of individual drivers or vehicles is described explicitly. The basis for this description are rules for interactions between drivers. Because the behaviour of drivers is not the same for all drivers stochastic techniques are used. A wide variety of traffic simulation models using microscopic modelling is available for computer use (e.g. Aimsun2 (Barcelo and Ferrer, 1997), Vissim (PTV, n.d.), Paramics (Quadstone, 2002)). In general, microscopic traffic flow models calculate the status of driver and vehicle at time step $t + \Delta t$ based on their characteristics at time t. For all vehicles the position and speed is calculated over time.

Most microscopic models contain one or more of the modules car-following, lane changing and route choice (discussed in the next section). The car-following and lane-changing modules also form the basis of the development of the stability-indicator in chapter 7. They simulate the two main possible reactions of drivers; slowing down or changing lanes, or a combination of both.

Because this thesis focuses on motorway sections and not on networks the route-choice module is not elaborated extensively and the interested reader is referred to Ortuzar and Willumsen (1990).

Car-following models

In the car-following module of a microscopic traffic flow model the behavioural rule is set out for the reaction of a driver to a speed fluctuation in front (a so called follow the leader model, figure 3.1).

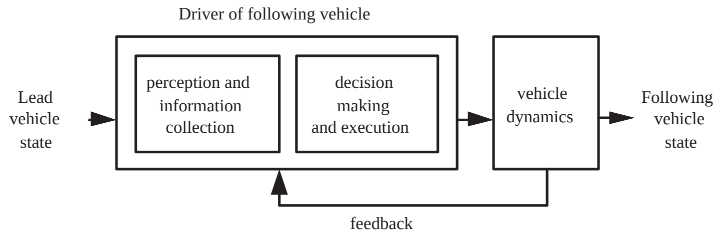

Figure 3.1: Block diagram of follow the leader car-following model.

Three main car-following models are the safe-distance models, the stimulus-response models and the psycho-physical spacing models. These three will be discussed next.

Car following models: safe-distance models

A car following model can be created assuming that drivers will keep a safe distance to their predecessor: Pipes (1953) presented a car following behavioural model which is based on drivers reaction to the vehicle ahead. He assumed that a driver will always keep a distance large enough to drive safely. This distance is at least 1 vehicle for every 16 km/h of the speed at which the vehicle is driving. Represented in a equation:

$$\hat{s}_i = \hat{l}_i(1 + \hat{u}_i/58) \tag{3.1}$$

in which $\hat{s}_i$ equals the space of vehicle i, $\hat{l}_i$ the length of vehicle i and $\hat{u}_i$ is the speed in m/s. In this model the space (and distance) grow linearly with the speed of the vehicles.

The campaign of the Dutch Ministry of Transport, Public Works and Water Management to 'keep at least 2 seconds of headway', can also be represented in a car following rule:

$$\hat{d}_i \geqq 2\hat{u}_i \tag{3.2}$$

in which $\hat{d}_i$ equals the distance of vehicle i, and $\hat{u}_i$ represents it's speed in m/s. In both safe distance models the distance grows linearly with the speed of the vehicles.

Car following models: stimulus-response models

Both safe-distance models mentioned above do not incorporate reactions and reaction times of drivers. In that sense the models are less realistic than stimulus-response models described in May (1990) and Gipps (1981). More complete car-following models are of the form of the speed of a vehicle $(i+1)$ being a function of the speed of the vehicle in front (i), the minimum safe distance ($\hat{d}_i^{\mathbf{s}} = 1/k_c$, with k_c the critical density) and the reaction time (t_r) in the speed function $\hat{u}_{i+1} = f(\hat{u}_i, \hat{d}_i^{\mathbf{s}}, t_{r,i})$, with $t_{r,i}$ the reaction time of the ith vehicle. The reactions of a driver to speed fluctuations of a vehicle in front are according to the principle:

$$response = sensitivity \times stimulus \tag{3.3}$$

the speed difference between the vehicles is the stimulus and the distance between the vehicles, together with the individual characteristics of the drivers form the sensitivity.

This comprehensive approach was (among others) developed by the researchers associated with the General Motors group. They are very well known and important, because of the accompanying comprehensive field experiments. For an elaborate overview of the development of the GM car-following theories see May (1990). The GM team developed five generations of car-following models, all of the form as mentioned above (equation (3.3)), and special cases of the fifth and final model:

$$\hat{a}_{i+1}(t+\Delta t) = \frac{\alpha_{l,m}[\hat{u}_{i+1}(t+\Delta t)]^m[\hat{u}_i(t) - \hat{u}_{i+1}(t)]}{[\hat{x}_i(t) - \hat{x}_{i+1}(t)]^l} \tag{3.4}$$

in which $a_i(t)$ is the acceleration of the ith vehicle at time t, $u_i(t)$ is the speed of the ith vehicle at time t and $x_i(t)$ equals the location of the ith vehicle at time t. The reaction (Δt) time and the sensitivity parameter $(\alpha_{l,m})$ can be determined through field experiments. Typical values are 1.2 to 1.5 seconds for drivers not anticipating any disturbance or down to 0.6 seconds for drivers expecting a disruption of the flow and 10 m/s resp.

Figure 3.2 shows a speed-time and a space-time plot of a platoon of ten vehicles, of which the leading vehicle is confronted with a speed disruption. The leading vehicle decelerates from 30 m/s to 20 m/s with -2 m/s^2 and later accelerates back to 30 m/s with 3 m/s^2. The effect of the leading vehicle's speed disruption is propagated to the vehicles upstream of the leading vehicle, which follow under the GM car-following rules of equation (3.4) $(l = 1, m = 0)$. In this example the inter-vehicular spacing (headway) is 0.8 seconds at the moment the leading vehicle starts with it's speed change. The reaction time of drivers is set at 0.6 seconds. A reaction time of 0.6 seconds can only be realized when drivers are aware of the disruption or expect it in advance. A headway of

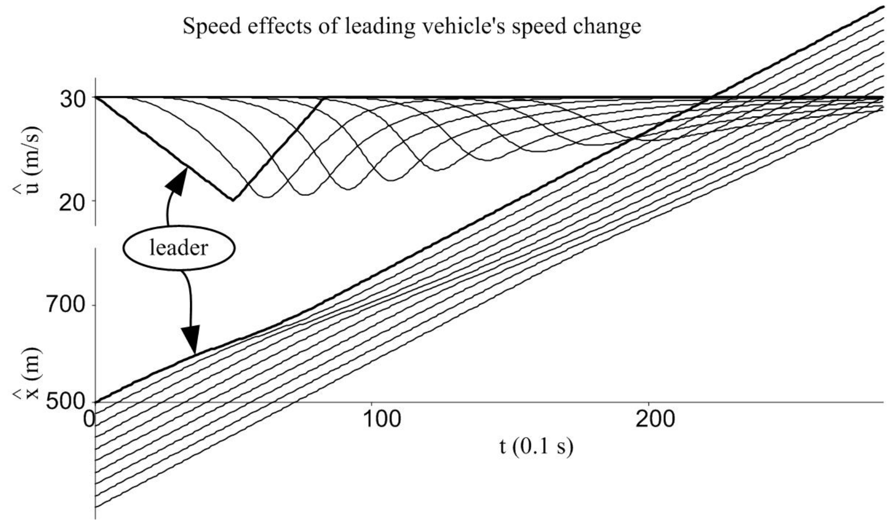

Figure 3.2: Example of propagation of a speed disruption in a platoon calculated using the GM car-following model, visualized in an [x,t] and [u,t] diagram.

0.8 seconds appears to be enough to eventually reduce the speed disruption to zero; the tenth vehicle's speed drop is much less than the first vehicle's speed drop.

In chapter 7 the GM car-following model is used in an operational tool to measure traffic stability, based on headways of a string of vehicles. Different platoon configurations (mean headway and standard deviation) are confronted with a speed fluctuation of the leading vehicle. The probability of an extraordinary manoeuvre (a manoeuvre stronger than the car-following model predicts) being necessary to prevent a possibility of a collision is calculated.

An alternative approach to car-following is found in Gipps (1981); instead of deriving the desired speed and the stimuli for speed changing, he derives a car-following model based on the assumption that every driver sets limits to his desired braking and acceleration rates. The limits are used to calculate a safe distance for each following vehicle. The speed of a vehicle n at time step $t+\tau$, if it is assumed that the driver travels as fast as safety and the limitations of the vehicle permit, is given by:

$$\begin{aligned}\hat{u}_i(t+t_r) \quad = \quad & \min\{\hat{u}_i(t) + 2.5\hat{a}_i t_r(1 - \hat{u}_i(t)/\hat{u}_i^{\mathbf{w}})(0.025 + \hat{u}_i(t)/\hat{u}_i^{\mathbf{w}})^{1/2}, \\ & \hat{b}_{m,i} t_r + \\ & \sqrt{\hat{b}_{m,i}^2 t_r^2 - \hat{b}_{m,i}[2[\hat{x}_{i-1}(t) - \hat{z}_{i-1} - \hat{x}_i(t)] - \hat{u}_i(t) t_r - \hat{u}_{i-1}(t)^2/\hat{b}]}\}\end{aligned} \tag{3.5}$$

in which:

$\hat{a}_{m,i}$ is the maximum acceleration driver i wishes to undertake,

$\hat{b}_{m,i}$ is the most severe braking that driver i can undertake,

$\hat{z}_i$ is the effective size of vehicle i, that is, the physical length plus a margin into which the following vehicle is not willing to intrude, even when at rest,

$\hat{u}_i^{\mathbf{w}}$ is the speed at which the driver of vehicle i wishes to travel,

$\hat{x}_i(t)$ is the location of the front of vehicle i at time t,

$\hat{u}_i(t)$ is the speed of vehicle i at time t, and

t_r is the apparent reaction time, a constant for all vehicles.

$\hat{b}$ is the minimum of -3.0 and $(\hat{b}_{m,i} - 3.0)/2m/s^2$.

when the last part of equation (3.5) is the limiting part congested flow exists with the traffic flowing as fast as the volume of vehicles permits. When the first part is the limiting condition, the traffic flows freely and vehicles will tend towards their desired speeds. This final model developed by Gipps (1981) is used in the microscopic traffic simulation model Aimsun2.

The biggest difference between the GM car-following model (equation (3.4)) and the car-following equation in Gipps (3.5) is that the GM car-following model does not include individual differences in desired speeds. All vehicles follow their leader and tend towards their leaders speed. In congested traffic, or traffic in which all vehicles are driving below their desired speed this will not be a problem as the flow is synchronized (Kerner and Rehborn, 1996), but once vehicles reach speeds, which may be considered as desired for some vehicles, the GM-model becomes less accurate. The GM-model is a typical car-following model; so by definition not fit for free flow, while the Gipps model performs under both conditions.

Nevertheless, because of the very transparent rules and traffic synchronizes prior to breaking down (Kerner and Rehborn, 1996) the GM car-following model is used to setup a table to give insight into the effects of speed disturbances in platoons in chapter 7 and the stability indicator based on string stability is developed.
The calculation time of the GM car-following model is limited and the transparency is high.

Both stimulus response car-following models presented above assume that the driver of the following vehicle react immediately to a change in behaviour of their predecessor however small and at whatever distance this may be (see

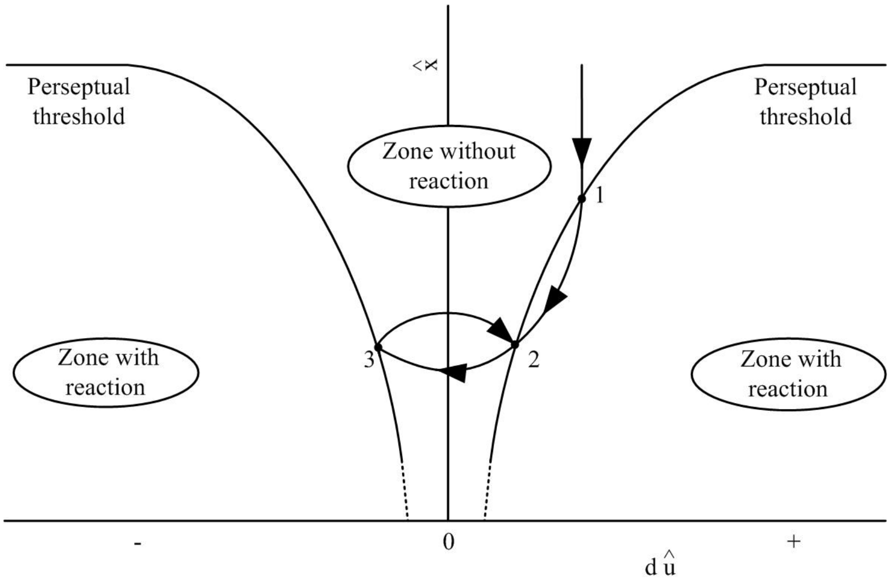

Figure 3.3: Example of trajectory of speed difference and distance between a leader and a follower. 1 = start of reaction (deceleration), 2 = reaction proceeds, but further change is not perceived, 3 = new reaction (acceleration).

equations (3.4) and (3.5)). This effect can be made even more realistic, because both distances and speeds that must be estimated by drivers have their limitations (Cavallo et al., 1997).

Car following models: psycho-physical models

Psycho-physical models include the lack of exact knowledge of drivers as well as the drivers limitations on the stimuli to which they respond, which is present in the stimulus-response models. The basis of these models are two behavioural rules:

1. At large spacing the following vehicle is not influenced by the speed difference. So the 'following' vehicle is not following.
2. When the relative motion is to small, there are combinations of distances and headways at which there is also no response of the following vehicle.

These two behavioural rules imply perceptual thresholds for drivers. Until the combination of speed difference and distance reaches a threshold the following driver will not react to speed differences or distance changes of it's leader; it is only within the thresholds that the follower can judge and will react.

In figure 3.3 the trajectory of a following vehicle approaching a leading vehicle from behind is shown, related to the distance from and speed difference with the leading vehicle. At first the speed difference is constant and the distance is reduced. Then the following vehicle approaches the zone in which differences in speed and the distance are perceived and a reaction starts (point 1). The following vehicle decelerates and reduces the speed difference to a level at which the combination of speed difference and distance is not perceived as one to react upon (point 2). When the driver in the following vehicle notices that the leading vehicle is getting ahead and there is a perceivable speed difference building up, the following vehicle accelerates again (point 3).

Wiedemann (1974) developed the first psycho-physical model. The perception threshold as seen in figure 3.3 marks the distinction between constrained and non-constrained drivers in the free-flow regime.
A full psycho-physical expansion of the GM car-following model is not necessary, because to the results per run do not become more realistic with the incorporation of perceptual thresholds as in figure 3.3. To make larger improvements one should incorporate that the perceptual levels are different for different drivers and even not constant for the same driver over time (Tampère, 2004). One can imagine that these improvements are difficult to make and to calibrate. What can be learned from the psycho-physical analysis of car-following is that above the threshold of 4 seconds is not legitimate to use car-following models. Speed differences at these large distances cannot be judged properly and therefore the GM car-following model can not be used above 4 seconds of headway. In chapter 7 is showed that the main impact on traffic stability is not the differences between people (e.g. reaction time), not the sort of disruption, but is the distribution of headways in the platoon.

Lane changing models

Next to car-following the second important element of microsimulation models is the lane changing model. Lane changing models are not only important for microscopic traffic simulation, but also for developing geometric standards, assessing capacity at weaving sections, and evaluating capacities of and delays at intersections and ramps. Gipps (1986) presented a structure of lane changing decisions for urban roadways. The model focuses on the decision-making process of the driver, considering potential conflicting goals and assuming logical driver behaviour. The decision to change lanes or not in the Gipps model (basis for the development of Aimsun2) can be split up into:

- whether it is physically possible and safe to change lanes without an unacceptable risk of collision,
- the location of objects (vehicle standing still),
- the presence of special purpose lanes such as transit lanes,
- the distance to the driver's intended turning movement (mandatory lane change),

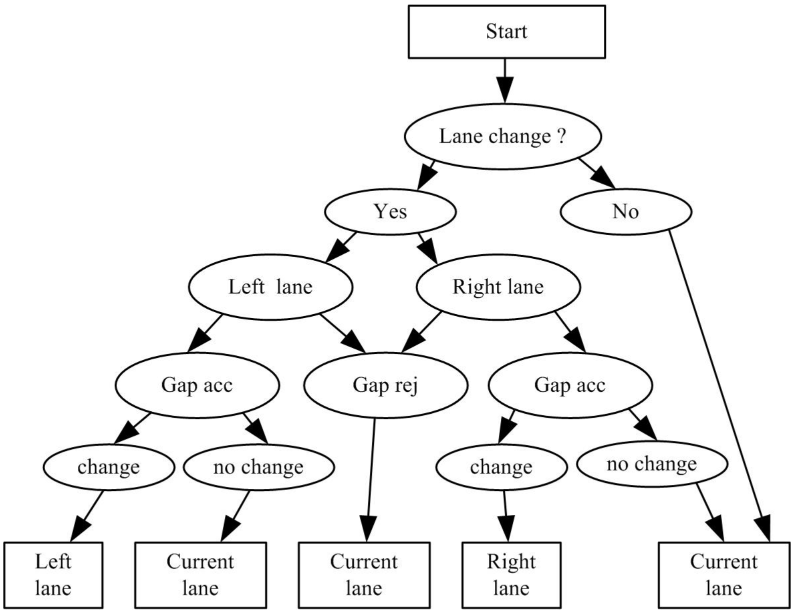

Figure 3.4: Lane changing model structure, after Ahmed et al. (1996).

- legislation (keep right or keep your lane),
- the presence of heavy vehicles, and
- the possibility of gaining a speed advantage.

The model also considers the urgency of the lane changing manoeuvre in terms of distance of the intended turn of the driver. The urgency of the manoeuvre is modelled through the driver's gap acceptance and braking behaviour. However, the model assumes that a lane changing manoeuvre takes place, only when it is 'safe', i.e. when a gap of sufficient size is available in the target lane.

Lane changing models consists of a decision module to perform a lane change, a choice of the target lane and a gap acceptance module. Lane changing modules are always combined, or in need of a gap acceptance model. In gap acceptance models the gap is accepted or rejected by the vehicle that needs to change lanes, based upon speed differences and available space in front and behind.

For modelling lane changing behaviour on motorways a discrete choice framework can be used (Ahmed et al., 1996). Lane changing is modeled as a sequence of three steps: decision to consider a lane change, choice of left or right, and search for an acceptable gap (see figure 3.4).
Whether the gap is acceptable or not depends on the combination of lead gap and lag gap of vehicles driving in the next lane and their speeds.

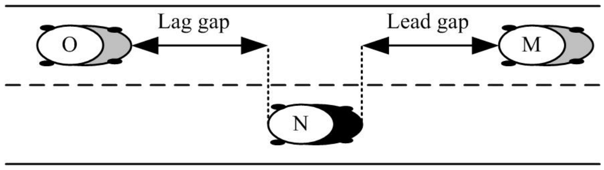

Figure 3.5: Lead and lag gap.

Unfortunately these and other models ((Fritzsche, 1994) and (Yousif and Hunt, 1995)) do not incorporate active cooperation of drivers in adjacent lanes. If a lane change is not 'safe', the vehicle will simply not change lanes, while in real life in those situations (mandatory lane changes on on ramps or due to accidents) vehicles in the adjacent lanes active cooperate to create acceptable gaps (Kita, 1999).

Route choice models

The basic assumption in route choice models is that drivers act rational. Travellers will choose the route from their origin to their destination which is perceived as the less costly. The main factors that influence route costs are: journey time, travel distance, fuel costs, congestion, toll costs and reliability of travel time. Because it is difficult to model all of these, mostly two are modelled: time and money (fuel, toll). The weighted sum of these two becomes the generalized cost used to estimate route choice.

In highway modelling car-following and lane-changing (including gap acceptance) models are far more important than the route choice model, because of lack of route choice. Therefore route-choice models are no further elaborated.

3.3.2 Mesoscopic traffic flow modelling

Mesoscopic traffic flow models do not describe the behaviour or characteristics of individual vehicles, they describe the characteristics of a group of vehicles; usually in a probabilistic manner. The behavioural rules used are deducted from the individual level. The processes in a mesoscopic model are described by various processes (acceleration, deceleration, lane changing), describing the individual level.

Headway distribution models

This section starts with an example of how data observations can be measured and then gives an overview of headway distribution, both for 'free' and for 'constrained' traffic. Later some commonly used headway distributions are shown to calculate merging probabilities.

May (1990) collected an extensive data-set on time headway distributions and did the following general observations.

- Individual time headways are rarely less than 0.5 second (1 to 2 percent).
- Individual time headways are rarely over 10 seconds unless the flow rate is below 15 vehicles per minute.
- The time headway mode is always less than the median, which is always less than the mean. However, they tend to converge as the minute flow rate increases toward capacity.
- The mean time headway tracks the 67 cumulative percentile curve for the entire minute flow rate range.
- The ratio of the standard deviation to the mean time headway approaches 1 under free flow conditions but decreases continuously as the minute flow rate increases.

The set that was considered was clearly not a set of very light traffic. The remarks above are a first screening of measurements; usually the measurements are matched with probabilistic statistics. Besides probability distribution models, there are also some headway distribution models which are based on vehicle and driver behaviour as driver reaction time, driver acceleration and vehicle speed. In the next sections some headway distribution models are compared. One should remember that not all models were developed under comparable conditions and had different research aims.

Arrivals at free flow

Light traffic flows are almost random; the probability of a vehicle passing a certain point during some time interval is independent of the passing times of other vehicles. The Poisson distribution is an appropriate distribution to describe the independent arrivals, which will be proved.

Let the probability of n vehicles passing a point over a time span t is given by $P(n)$. Over the next small time interval δt the probability of n vehicles arriving in a time $t + \delta t$ is the sum (because the arrivals are random) of the probability of n vehicles arriving in a time t and of $n - 1$ vehicles in t and one vehicle in δt, i.e.

$$P_{t+\delta t}(n) = P_t(n-1) * P\{1 \; vehicle \; in \; \delta t\} + P_t(n) * P\{0 \; vehicles \; in \; \delta t\} \quad (3.6)$$

Because the assumption that the flow is random was made, the probability of 1 vehicle arriving in a time interval is the same as the expected number of vehicles to arrive ($q\delta t$). The probability of zero vehicles arriving is one minus the probability that one vehicle arrives ($1 - q\delta t$).

$$P_{t+\delta t}(n) = P_t(n-1)[q\delta t] + P_t(n)[1 - q\delta t] \tag{3.7}$$

From which follows:

$$P_{t+\delta t}(n) = P_t(n) + q\delta t[P_t(n-1) - P_t(n)] \tag{3.8}$$

Differentiating to t shows:

$$\frac{\mathrm{d}P_t(n)}{\mathrm{d}t} = q[P_t(n-1) - P_t(n)] \tag{3.9}$$

The solution of this equation is the Poisson distribution:

$$P_t(n) = \frac{\lambda^n e^{-\lambda}}{n!} \tag{3.10}$$

with λ the (mean) number of vehicles expected to arrive in time t: $\lambda = q\ t$. So for light traffic (arrival rates of vehicles are not affected by other vehicles) the arrival rate can be modelled by the Poisson distribution. Because not the arrival rate, but the distances between drivers form the window of opportunity for merging traffic, in a similar matter the headways of vehicles in a traffic flow are considered.

Headway distribution models for free flow

The distribution of headways can be found in a similar way. Let $P(T > t)$ be the probability that a headway T is larger than some time interval t, then the probability of a T larger than $t + \delta t$is:

$$P(T > t+\delta t) = \underbrace{P(T > t)}_{\textit{probability of the event in T}} * \underbrace{[1 - q\delta t]}_{\textit{probability of no event in } \delta t} \tag{3.11}$$

in which q is the flow. This relation may be rewritten to:

$$\frac{\mathrm{d}P(T > t)}{\mathrm{d}t} = -qP(T > t) \tag{3.12}$$

The solution of this equation is:

$$P(T > t) = e^{-\lambda t} \tag{3.13}$$

with λ the (mean) number of vehicles expected to arrive in time t: $\lambda = q\ t$, so λ can be considered as the flow rate.

The cumulative distribution function $F(t)$ of the Negative exponential distribution is:

$$\begin{array}{ll} F(t) = 0 & (\mathrm{t} < 0) \\ F(t) = 1 - e^{-\lambda t} & (\mathrm{t} \geq 0) \end{array} \tag{3.14}$$

The Poisson distribution for arrival rate and the negative exponential distribution for headways are valid only when traffic flows are light. But they form the basis of headway distribution modelling, as can be found in Cowan (1975), Tolle (1976), and Leutzbach (1988).

The distribution of vehicles and headways is relatively easy to model when traffic is light. When traffic gets heavier, vehicles start interacting and the random process of arrival may no longer be assumed.

Cowan (1975) assumed that the vehicle headway (X) consisted of two components; one for 'tracking' (V) and one for 'free' (U). For free traffic the tracking component equals zero and the distribution is equal to a negative exponential distribution. When a tracking component of τ seconds is added to the negative exponential distribution the result is a displaced (or shifted) negative exponential distribution:

$$\begin{array}{ll} F(t) = 0 & (\mathrm{t} < \tau) \\ F(t) = 1 - e^{-\gamma(t-\tau)} & (\mathrm{t} \geq \tau) \end{array} \tag{3.15}$$

The expected value of the distribution $E(t)$ equals $\tau + 1/\gamma$ and the flow rate $\lambda = \frac{\gamma}{\gamma\tau+1}$. Note that the flow rate λ can be expressed as $1/E(t)$.

Although the shifted negative exponential distribution considers tracking traffic, it still does not fit heavy traffic flows. Two distributions which fit quite well to both free traffic and more dense traffic are Pearson Type III distribution (which include the Erlang distribution) and Log-normal distribution.

One of the generalized mathematical model approaches for the distribution of headways is Pearson type III distribution. The cumulative distribution function of this distribution is:

$$\begin{array}{ll} F(t) = 0 & (\mathrm{t} < 0) \\ F(t) = \int_0^t \frac{\lambda}{\Gamma(K)}[\lambda(t-\alpha)]^{K-1} e^{-\lambda(t-\alpha)} \mathrm{d}t & (\mathrm{t} \geq 0) \end{array} \tag{3.16}$$

where K and α are parameters that affect the shape and the shift of the distribution of the headways (t). If $K = 1$ and $\alpha = 0$ than the p.d.f. equals the negative exponential distribution. If $K = 2, 3, 4, 5$ and $\alpha = 0$ than de p.d.f. equals the p.d.f. of an Erlang distribution (Leutzbach, 1988). All p.d.f.'s with $\alpha = 0$ and positive K are Gamma distributions.

Another distribution which is fit for a wide range of traffic volumes is the Log-normal Distribution. Tolle (1976) tested this with field data from motorways in Ohio, USA. The Log-normal distribution is:

$$\begin{array}{ll} F(t) = 0 & (\text{t}< 0) \\ F(t) = \int_0^t \frac{1}{\sqrt{2\pi}\sigma} e^{-(\ln t-\mu)^2/2\sigma^2} \mathrm{d}t & (\text{t}\geq 0) \end{array} \tag{3.17}$$

in which:

$\mu = \frac{\sum_1^n \ln t_i}{n}$ and $\sigma^2 = \frac{\sum_1^n (\ln t_i - \mu)^2}{n-1}$.

Headway distribution models for constrained flow

Cowan (1975) proposed the Bunched exponential distribution to fit data of constrained flows. This distribution improved the accuracy in the prediction of small arrival headways in the negative exponential distribution and the displaced negative exponential distribution, but retains the simplicity of those models. The model assumes that a portion of vehicles θ, are tracking (following) behind other vehicles at a headway τ. These vehicles are called bunched and the rest of the vehicles is assumed to be travelling freely at a headway greater than τ. The Bunched negative exponential distribution function is:

$$\begin{array}{ll} F(t) = 0 & (\text{t}< \tau) \\ F(t) = 1 - (1-\theta)e^{-\gamma(t-\tau)} & (\text{t}\geq \tau) \end{array} \tag{3.18}$$

Akcelik and Chung (1994) described the model in detail and gave the results of its calibration using real data for single lane traffic streams and simulation data for multi-lane traffic streams. Sullivan and Troutbeck (1994) made a comparison of the Bunched negative exponential distribution and the Double displaced Negative Exponential Distribution developed by Griffiths and Hunt (1991). It is proved that both models are very accurate, but the Bunched exponential model is much simpler. The Double displaced negative exponential distribution is developed by Griffiths and Hunt (1991):

$$\begin{array}{ll} F(t) = 0 & (\text{t}< \tau) \\ F(t) = \int_0^t \phi\lambda_1 e^{\lambda_1(t-\tau)} + (1-\phi)\lambda_2 e^{\lambda_2(t-\tau)} \mathrm{d}t & (\text{t}\geq \tau) \end{array} \tag{3.19}$$

The parameter $\phi(0 < \phi < 0.5)$ is a weighting factor, τ is the displacement factor and λ_1, λ_2 are constants which depend on the traffic flow. Sullivan and

Troutbeck (1994) indicated that the Double displaced negative exponential distribution can model smaller headways more accurately than the Bunched negative exponential model.

Besides single distribution models Composite distribution models exists. They are developed under the assumption that vehicles are either travelling free and are not influenced by the vehicle in front of them or they are following the previous vehicle. A popular composite distribution ((Ovuworie et al., 1980), and (Koshi et al., 1992)) is the Semi-exponential distribution:

$$\begin{array}{ll} F(t) = 0 & (\mathrm{t}< 0) \\ F(t) = (1-\phi)F_1(t) + \phi F_2(t) & (\mathrm{t}\geq 0) \end{array} \tag{3.20}$$

where ϕ is the proportion of the followers, $F_1(t)$ is the distribution for leaders, usually a negative exponential distribution, and $F_2(t)$ is the distribution for followers.

All of the models mentioned above can fit traffic situations well, but the derivation of unknown parameters is complicated. Method of moments, maximum likelihood techniques and hybrid methods can be tried. Griffiths and Hunt (1991) and Sullivan and Troutbeck (1994) gave some descriptions of the methods to get unknown parameters.

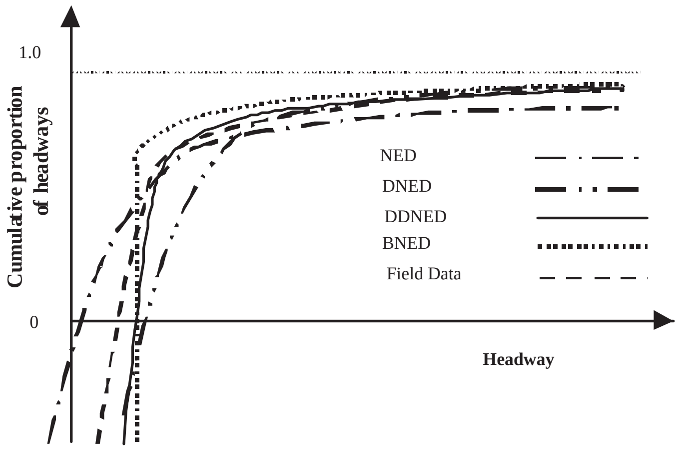

Figure 3.6: Headway models: NED (negative exponential distribution), DNED (displaced negative exponential distribution), DDNED (double -negative exponential distribution)

Headway distribution models are the best known mesoscopic models; cluster models (Botma, 1978) and gas-kinetic models (Hoogendoorn, 1999) are other mesoscopic models.

Conclusions

Most headway distribution models are probability distribution models. The basic model is the negative exponential distribution. This model and Erlang and Pearson Type III can be used in free and light traffic flow conditions. For prediction of small headways vehicles can be divided into two classes, free ones and constrained ones. For constrained vehicles more models are developed, including the Bunched negative exponential distribution model and Double displaced negative exponential distribution model. The first one is simple, while the second is more accurate; the models that well represent the constrained vehicles are more realistic, but the derivation of unknown parameters is complicated.

The main drawback of headway distribution models is that they assume that all drivers are equal (the probability density function are independent of vehicle type, driver type, travel purpose, etc), while these characteristics of drivers actually make a difference (Kijk in de Vegte, 2002). Hoogendoorn (1999) developed headway distribution models for multi-class and multi-lane traffic flow.

Since mesoscopic traffic models do not describe every vehicle individually, mesoscopic simulations are in general less computationally intensive than microscopic simulations. In chapter 7 a mesoscopic headway model and the microscopic GM car following model are used to setup a stability indicator.

3.3.3 Macroscopic traffic flow modelling

Macroscopic traffic flow models do not describe individual vehicles, but deal with flow as an aggregate, as seen in paragraph 2.3. Lighthill and Witham (1955) were the first to recognize the behaviour of a traffic flow in this sense and describe it in a model. Richards (1956) published the same ideas one year later and the macroscopic model they presented is nowadays known as the Lighthill-Witham-Richards model (LWR-model). Since then many macroscopic (continuum) models have been developed. They can be split in first-order and higher-order continuum models.

First-order continuum models

First-order continuum models use the conservation of vehicles equation:

$$\frac{dk(x,t)}{dt} + \frac{dq(x,t)}{dx} = 0 \tag{3.21}$$

in which k is the density and q the flow at time t and location x.

This equation can be derived by imagining a uniform and continuous road section. On this road are two measuring stations (1 and 2). The number of vehicles entering the road section at point 1 is called N_1 and the number of vehicles leaving the section at point 2 is called N_2. The flow rate at point 1 is $N_1/\Delta t$ and at point 2 is $N_2/\Delta t$ the number of vehicles entering or leaving the section (Δq) in time is $\frac{N_2-N_1}{\Delta t}$ so $\Delta N = \Delta q \Delta t$. The traffic is assumed to be homogeneous distributed over the section. The difference in density (Δk) is $\frac{N_1-N_2}{\Delta x} = \frac{-\Delta N}{\Delta x}$ so $\frac{\Delta q}{\Delta x} = -\frac{\Delta k}{\Delta t}$. Taking the limit ($\Delta x$ and Δt to zero) from this follows the conservation equation (3.21).

If vehicles leave or enter the motorway this equation takes the more general form:

$$\frac{dk(x,t)}{dt} + \frac{dq(x,t)}{dx} = g(x,t) \tag{3.22}$$

where $g(x,t)$ is the generation (or dissipation) rate of vehicles per unit length per unit time.
As said: the solution of the conservation equation, applying it to traffic was first proposed by Lighthill and Witham (1955) and by Richards (1956). Solutions of the equations are the descriptions of the shockwaves.

Solution of the conservation equation and the explanation of shockwaves

To solve equation (3.21) consider the fundamental relationship:

$$q = ku \tag{3.23}$$

Also consider speed a function of the density $u = f(k)$; equation (3.21) can be solved analytically. The conservation equation can be rewritten as:

$$\frac{\partial}{\partial x}(ku) + \frac{\partial k}{\partial t} = \frac{\partial}{\partial x}(kf(k)) + \frac{\partial k}{\partial t} = f(k)\frac{\partial k}{\partial x} + k\frac{\mathrm{d}f\partial k}{\mathrm{d}k\partial x} + \frac{\partial k}{\partial t} = 0 \tag{3.24}$$

which can be rewritten as:

$$[f(k) + k\frac{\mathrm{d}f}{\mathrm{d}k}]\frac{\partial k}{\partial x} + \frac{\partial k}{\partial t} = 0 \tag{3.25}$$

Details of the solution of equation (3.25) are:

- The density k is constant along waves; a wave is the propagation of a change in flow and density along the roadway.
- The waves are straight lines originating from the boundaries of the time-space domain.

- The slope of the waves is:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(k) + k[f(k)] = \frac{\mathrm{d}q}{\mathrm{d}k} \tag{3.26}$$

  which implies that the waves have slope equal to the tangent of the flow-density curve of the point which represents the flow conditions at the boundary from where it originates.

- When two waves intersect, the density should have two values, which is unrealistic. This discrepancy explains the onset of shockwaves.

- The speed of the shockwave is:

$$u_w = \frac{q_d - q_u}{k_d - k_u} \tag{3.27}$$

  where k_d and q_d are downstream characteristics and k_u and q_u are upstream characteristics.

But the descriptions of shockwaves in the LWR-model as the solution of the first order equation has some disadvantages:

- The relations have no anticipation in them; as a consequence of this speed changes, although densities are high near the onset of a shockwave, are immediate.

- The dissolution shockwave and the onset wave both travel with constant speed ($\Delta q/\Delta k$) and therefore the end of the shockwave is rectangular, which is not realistic.

- The model allows flows out of a shockwave equal to the capacity of the road. In practice this outflow is less than capacity. This makes the model unsuitable for modelling highways with dynamic traffic control measures.

- A traffic phenomenon sometimes seen at bottlenecks, stop-and-go-traffic cannot be modelled in the LWR-model.

These disadvantages were reasons to develop better macroscopic traffic flow models; higher order models.

Higher-order continuum models

In order to incorporate some shortcomings mentioned above the continuum model is complemented by $q(u)$ and an equation to describe the acceleration behaviour of vehicles:

$$\frac{\mathrm{d}u}{\mathrm{d}t} = \underbrace{\gamma(u(k) - u)}_{relaxation} \; \underbrace{-\frac{c_0^2}{k}\frac{\partial k}{\partial \mathrm{x}}}_{anticipation} \tag{3.28}$$

The relaxation term allows the delayed adjustment of the stream to a specific speed $u(k)$. The anticipation term allows drivers to change their speed in advance to changes in density lying ahead. The coefficient c_0 corresponds to the speed at which disturbances propagate at very high densities. Besides relaxation and anticipation higher order macroscopic models are able to model start-and-stop waves at traffic jams and hysteresis.

3.4 Modelling stability related issues

This section describes how microscopic or macroscopic models are able to model phenomena which are closely related to traffic stability. Capacity, traffic breakdown, pinpointing the queue location head and the propagation of shockwaves are examples of these phenomena. When these qualities are embedded in the models, the models might prove to be a good basis to model traffic flow stability.

3.4.1 Probabilistic description of traffic breakdown

Traffic breakdown can be described as a probabilistic event (Lorenz and Elefteriadou, 2001). In it's essence this is what stochastic microscopic models do.

First order macroscopic models are deterministic and will not be able to show this behaviour. Higher-order macroscopic models can give information about this behaviour by using different initial conditions (which only differ slightly).

3.4.2 Queue location

After breakdown has occurred a queue will form. How well are microscopic and macroscopic models able to model the queue and it's location?

In the introduction was showed that the head of a queue is not always exactly at the location of the physical disruption, e.g. the on ramp (also in (Van Toorenburg and Elbers, 2000)). The queue head can change location during the existence of the queue, or start at a location downstream of the bottleneck, work it's way up to the bottleneck and stay there. Factors influencing the queue head location are:

- traffic demand (main lane and on ramp)
- the road geometry (length of on ramp, speed limit)
- traffic composition and experience and
- external conditions (weather)

For traffic flow modelling it is important to know the exact location of the queue. Especially in The Netherlands, where on ramps and motorway junctions are close to each other, it is important to know whether one queue will block another flow or not. For this both the queue head and the queue tail need to be known.

Known to the author non of the microscopic models can properly model the displacement of the queue head. The behaviour of drivers in congestion is different, e.g. they accept smaller headways than under other traffic conditions (AVV, 2000). AVV (2000) also suggests that for microscopic models to be able to predict queue tail location differentiations in driver types need to be made to model traffic in congestion better.

Hoogendoorn (1999) shows that a higher order model can be used to show a displacement of a bottleneck over time. It has a variation of about 500m. The relaxation time is needed for the queue to move upstream of where it started. In first order macroscopic models this time is set equal to zero and the exact queue head location can only be approached by the initial queue location. When a homogeneous queue forms, the first order macroscopic model will be able the model the queue length (Treiber and Helbing, 2002). Other forms of queues (non homogeneous) can only be modelled by higher order macroscopic models (in AVV (2000) non-local interaction (Hoogendoorn and Bovy, 2000a) and or building new models (Zhang, 1998) are suggested).

3.4.3 Propagation of shockwaves

Microscopic models are not used for modelling shockwaves, so their ability to do this is still unknown. The ability for macroscopic models to model shockwaves was already discussed in section 3.3.3.

3.4.4 Discussion

Both microscopic and macroscopic models are not able to fully model the complex phenomena as capacity, queue location, breakdown location and time and the propagation of shockwaves on motorways. But combining both is difficult, because of the different characters of the models

The headway distribution models (mesoscopic) overview shows that especially in near congested traffic flows (in which our interest is maximum) it is important to have a good idea on how the headways are distributed.

A combination of mesoscopic and microscopic models might give a better solution to modelling phenomena that surround stability, but before the two are combined the human aspects that should or should not be incorporated are discussed in chapter 6.

In chapter 7 the knowledge on headway distributions in constrained flows will be combined with a microscopic traffic flow simulation. This means that the GM car-following model will be fed with vehicles, which headways are distributed according to the Bunched negative exponential distribution model.

3.5 Summary

In this chapter an overview of the main types of existing traffic flow models was presented, together with their ability to model stability related issues.

For reaching our research goal and to get a better idea on how traffic flows differ and how traffic flow stability as a characteristic of a flow can be used to explain the onset of traffic congestion (breakdown), (parts of) traffic flow models can be used. Both microscopic and macroscopic models cannot incorporate all the phenomena of traffic flows near critical flow and towards breakdown.

Mesoscopic models form a good compromise between realistic modelling and computational possibilities.

Microscopic models are able to show the influences of small disruptions when traffic situations get critical. This is the reason why in the development of the stability indictor a microscopic car-following model (the GM car-following model) and a headway distribution model are used to setup a stability indicator based on string stability.

Both microscopic and macroscopic modelling cannot do without calibration of the results. To do model calibration one needs traffic data, which is never straightforward to collect or to interpret. The most common ways of collecting traffic data in the Netherlands, how these need to be processed, and how to interpret them is covered in the next chapter.
In chapter 7 a microscopic traffic simulation program will be used to estimate the effect of a disruption to the flow in order to make an estimate of the flow's stability. In chapter 9 the microscopic model Aimsun2 is used to compare the Alinea ramp metering algorithm to the enhanced algorithm (with the stability indicator).

Chapter 4

Measuring traffic flows

This chapter shows how the characteristics of traffic, which are important to describe traffic flows, may be collected from the road side. Not all characteristics as defined in chapter 2 can be measured; some must be calculated. Collected traffic data may be corrupted or false and incomplete, therefore data conditioning might be needed.

In the first section (4.1) an overview of traffic characteristics and possible traffic measurements from which they can be deduced is given. Also the effects of the method of measurement are calculated. After that an overview of some of the important and available data collection systems (section 4.2) and the collection systems which are operational on Dutch motorways (section 4.4) are given. Finally some computational ways to condition data sets are presented (section 4.5).

4.1 Traffic characteristics

In chapter 2 microscopic and macroscopic traffic variables have been defined. Before is demonstrated how these variables can be determined, consider the following. A single vehicle's trajectory can be represented by it's space-time $[\hat{x}_i(t), \hat{y}_i(t)]$ trajectory (recall figure 2.1). In this figure the microscopic characteristics of a vehicle are indicated; distance and length, which form space, which are measured in the x-dimension. Gap and occupancy, which form headway are measured in the t-dimension. The speed of a vehicle at a certain time is defined as the derivative in space over time.

At a macroscopic level vehicles are not considered individually. For completeness the definitions of macroscopic variables by considering vehicle trajectories (figure 4.1) are shown. Flow is defined as the number of vehicles which crosses a certain point over a certain time interval. In figure 4.1 the flow is the number of trajectories crossing the 'horizontal rectangle of T by the infinitesimal small dx (which is in our example $= 8/T$). The density is the number of ve-

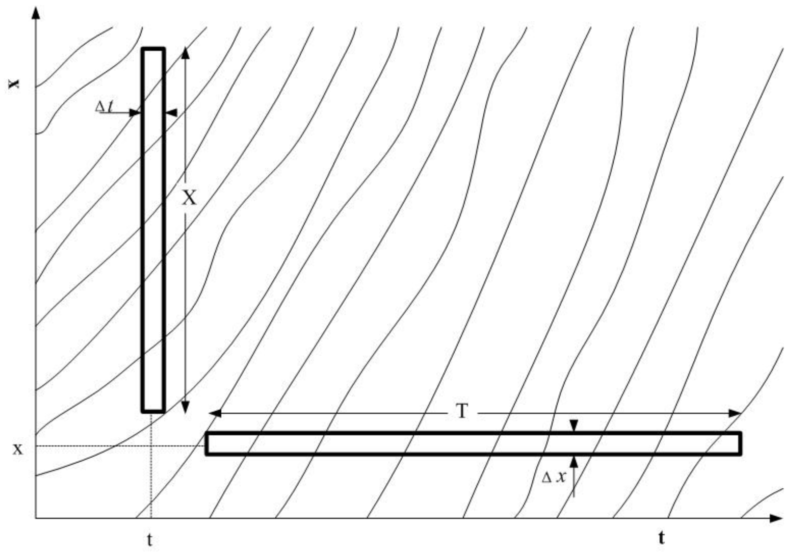

Figure 4.1: Vehicle trajectories and measuring intervals in space and time.

hicles over a stretch of road. The density of the 'vertical' rectangle of X by the infinitesimal small dt equals $5/X$. Local speed is the average of the speeds of the vehicles crossing a certain point (average in the 'horizontal' rectangle). Instantaneous speed is the average of the speeds of the vehicles on a stretch of road at a certain time (average in the 'vertical' rectangle).
Not all characteristics mentioned above can be measured in a straightforward way. Most detection systems are cross sectional (see section 4.2) and can only provide information which is gathered over time. On a microscopic level this means that headway, gap and occupancy can be measured directly, but space, distance and length are calculated by multiplying headway, gap and occupancy with the previous vehicle's or the vehicle's speed. On a macroscopic level the density and the instantaneous average speed require a single or two consecutive aerial photo's. As these are often not available the density is calculated measuring the speed and the flow and using the fundamental relationship $q = ku$. The instantaneous speed is often calculated by averaging the inverse of the speed of the vehicles measured at a cross section.

4.2 Detection methods

Most detection methods are cross-sectional. The reason for this is that these systems are cheap compared to collection of photo or video material from high positions above the ground (e.g. from a helicopter) and are collected on a

continuous basis.
Traffic data uniformity is characterized by consistency, repeatability, accuracy, and precision. In this section three common measurement systems are considered:

- inductive loops,
- pneumatic tubes, and
- traffic camera's.

They are the three most common data collecting systems in The Netherlands. The permanent system collecting data on Dutch motorways uses dual inductive loop detectors and will be dealt with more elaborately, as one of the aims for the stability indicator is to function using existing data collection systems. At the end of this section the conditions under which sensors like radar and video detection can be an alternative for loop detection on motorways are discussed.

4.2.1 Loop detection

Of all detection methods for vehicles in The Netherlands the inductive loop detector is the most common sensor used in traffic management. Loop detectors vary in size, depending on what type of vehicle and which characteristics need to be measured (detection of a bicycle at a traffic light differs from speed of vehicles on the motorway).

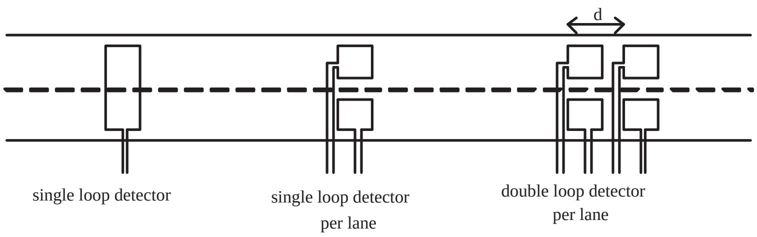

Figure 4.2: Loop configuration. The double loop detector per lane is most common on Dutch motorways.

Single loop detectors can be used for vehicle detection and flow rate (if each lane has it's detector). On Dutch motorways dual loop detectors, which also measure speeds and therefore sometimes called speed traps, of 1.5 meter each and 1 meter apart are most common in the road surface (see figure 4.2). An inductive loop consists of a conducting loop that is put inside the road surface. The loop is connected to an electrical circuit that produces an oscillating current through the loop. If a vehicle enters the loop the metal of the vehicle cause a shift in the current which is measured. From the time difference of the shifts between the first loop and the second the speed of the passing vehicle is calculated. From this speed and the total time in which the shift of the current is running

through the loop, the vehicle's length is calculated. The sum of the times the detector is occupied divided by the total time provides the occupancy.
From this data the average speed and flow are saved each minute (MARE), sometimes divided per lane and vehicle type (MONICA; Monitoring Casco). The detectors are usually located each 500 meter on each lane of the motorway. For research opportunities the individual vehicle data per lane, can be saved. The configuration of the MONICA system is explained in section 4.4.

Dual loop detectors on motorways are static. Traffic flow and individual measurements are done cross-sectional and the behaviour of the flow in between measuring points is not captured. A way to 'capture' macroscopic dynamics as shockwaves is to monitor a string of cross-sections over a road over a certain time (see figure 8.4). Dynamics over the section are not caught and vehicle trajectories are difficult to reconstruct from 2 cross-sectional measurements, 500m apart.

Figure 4.3: Double loop detectors on the A4 motorway. Picture from www.beeldbankvenw.nl

4.2.2 Pneumatic tube traffic detector

Pneumatic tubes were introduced as traffic detection system in the 1960's and are up to today very broadly used, especially for non permanent data collection systems. One (or two if speed is also measured) tube(s) are placed on the road surface (at a distance of about 50 cm from each other). With every axle that crosses the tube, the air in the tube is compressed and from the measurements of these fluctuations in pressure the weight and axle load is estimated and matched with predefined axle load graphs. This way vehicle count, vehicle classification and speed (if two tubes are used) can be calculated. The main advantage of this system is the mobility and the ease at which it can be installed. The main disadvantage is the wear and tear of the system as the tubes are installed on top of the road surface.

4.2.3 Traffic cameras

Traffic camera's as a system to collect traffic data are becoming more and more popular (Klein, 2001); especially for detection of vehicles at traffic lights and on motorways. A video camera is mounted above a roads crossing or the motorway and images are collected and sent to a video processing unit. Image recognition algorithms are used to extract the information needed: e.g. detection of a vehicle to apply for a green light at the traffic lights or speed of a vehicle on the motorway. The main advantages of the traffic camera are the high accuracy and the possibility of visual inspection during or after the collection. The main disadvantages are the high costs of installation and the reduced accuracy during conditions of low visibility (AVV, 2002a).

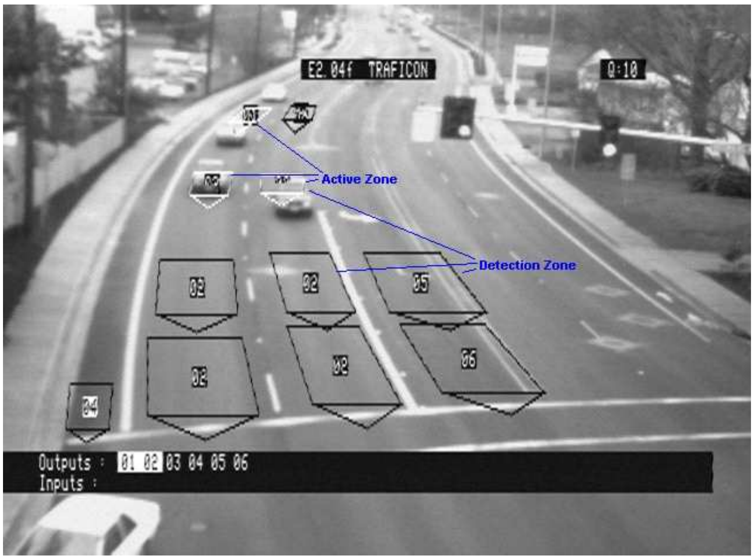

Figure 4.4: Traffic video camera image with virtual loops, courtesy of Traficon.

4.3 Data collection source, information and application

The three main sources of data collection (loops, pneumatic tubes and video camera's) have been discussed. The collection of data is the basis. Data measured can go straight into an (on-line) data bank (mostly in the form of a computer). It is possible to do calculations on data, before it is put into the data bank, e.g. flow and speed measurements can be combined and with the knowledge of fundamental diagrams the density can be calculated. Then the density can be put in the data bank next to the speed and the flow. Applications can get the data from the data bank. The data that is sent in the data bank can also be a copy of what is sent directly to a application (see figure 4.5).

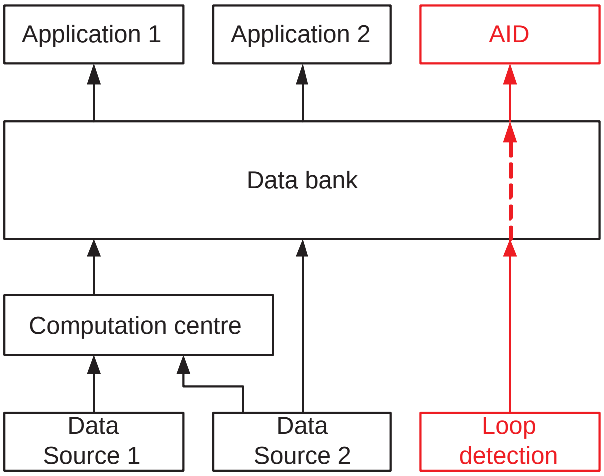

Figure 4.5: Different relationships between data sources and applications.

In the next section the performance of several data collection systems and applications which embed data collection systems are studied.

4.3.1 Performance of data collection systems

In this section a short overview of some traffic data collection systems which can be seen as alternatives for the loop detector, the main data collection method now used on motorways, is presented. The systems are compared on their ability to reach the quality' level that AID acquires of them. Although loop detection is the leading traffic measuring system in The Netherlands, the Dutch Ministry of Transport, Public Works and Water Management is working to explore alternatives for the future (AVV, 2002a). For a comprehensive overview of data collection systems see (Michalopoulos and Hourdakis, 2001) and (Klein, 2001), which cover among others pneumatic tube detectors, infrared detectors, radar detectors, ultrasonic detectors, laser detectors, induction loop detectors, microwave detectors and video detectors.

When alternatives of loop detection are studied functionalities are to be compared. As was mentioned earlier; the primary function of the loop detectors in The Netherlands is to provide speed information of vehicles within the automatic incident detection method. The second purpose is to collect data for the data bank (speed, flow, vehicle length, per lane, per vehicle). This situation in drawn in red in picture 4.5.
In the research into possible alternatives of the inductive loop detection on Dutch motorways (AVV, 2002a),the following questions on data demands were asked:

1. How is the quality of the individual speed data compared to that of the loop detector (percentage missing, or percentage deviation)?
2. What is the quality of the aggregated speed and flow rates (percentage missing, or percentage deviation)?
3. Is it possible to divide the vehicles in vehicle classes?

On the application of AID the following questions were asked.

1. Is it possible for video systems to detect a traffic jam automatically?
2. Can other data (from the video systems) be used for other Dynamic Traffic Management measures?

In a test four systems (three potential alternatives and the loop detector) were tested. The three alternatives were: the *Falcon* radar detector, and *Traficon* and *Autoscope* (two video detection systems). The Falcon radar detector appears to be a good alternative for the loop detector, although the speed measurements at low speeds are not accurate and the vehicles lengths measured are underestimated. Downside is that the Falcon system does not send all the data needed for the monitoring system MCSS (Motorway Control and Signalling System). Both video systems seem able to collect the data needed, but are not fit to cope with changing weather conditions. Therefore these cannot be used in fully automatic systems as the MCSS and the data collection system.

Table 4.1 shows that every technique has the potential to be a good detector, but all techniques have good points but also downsides. When it comes to performance on a certain measurement, the loop detector is still the best. The biggest disadvantage of loop detection is that the loops are static. In case of flexible lanes (a reasonable option for enhancing capacity in the future) this is not adequate and traffic characteristics can better be measured with radar. Up to 2006 the monitoring system of Dutch motorways is elaborated up to 1500 extra km of monitoring, so the availability of data from dual loop detectors is still growing fast, even though a lot of measurement technique developments

Table 4.1: Usability of detection systems. Numbers in brackets are the demands of detection rate. Results from (AVV, 2002a).

	DTM measures							
	Monitoring, Route information, Homogenize measures			MCSS, Automatic Incident Det.		Inc. det., veh. stop	Dyn. lane distr.	
Detection system	Av. Flow (95-99)	Av. Speed (95-99)	Class Veh. (98-99)	Av. Flow (98-99)	Ind. Speed (98-99)	Ind. Speed (98-99)	Check (95)	Explanation
Loop detection	++	++	+	++	++	–	–	Weather independent
Falcon radar	+	+	-	+	+	–	–	Weather independent
Traficon / Autoscope video	+/-	+/-	–	+/-	-	+/-	–	Problems with weather

are in-car.

For all measurement systems goes; although a detection grid of 500 meter is very dense for a motorway section compared to other countries, it might never be dense enough for some applications and therefore interpolation between detection remains beneficial. When good interpolation of traffic characteristics is available one can reduce the number of measuring points without reducing the level of accuracy below a critical level. Also for data conditioning interpolation or extrapolation techniques need to be considered (see section 4.5).

4.4 The Dutch motorway network and traffic flow measurements

The Dutch motorway network is about 2300 km in length and connects the most important economic centres of the country more or less direct (the detour factor is less than 1.4). The network covers the whole country and has a tangential orientation relative to the connected centres. On the 2300 km of motorway are about 260 interchanges, which implies an average entry/exit point every 8 km, and even more dense in urban areas (about every 4 km in Randstad [1]). The Dutch motorway network belongs to the most dense in the world (together with Belgium). Also the length of the motorway network as a percentage of the complete road network is the highest in Europe, 2%. The motorway network is heavily used and very well monitored. 40% of all motor vehicle kilometres driven are driven on the motorway network; so 40% of the kilometres are driven on 2% of the network. Around large cities the motorways are very close to the city, so the roads are also used for short distance trips. The heavy use of the network leads to frequent congestion. This congestion can be measured in vehicle hours lost. This has doubled from 30 million to 60 million from 1990 to 2000.

In order to reduce congestion and increase safety a number of dynamic traffic

[1] Randstad is the name of the densely populated region in the west of the Netherlands, which covers the four largest cities of Amsterdam, Rotterdam, The Hague and Utrecht.

management measures are operational on the network. The most important is the Motorway Control and Signalling System (MCSS), operational on about 1000 km of the motorway network. It consists of speed regulation and lane allocating measures. Extra lanes, only open in rush-hours, route information, and ramp metering are some of the other dynamic traffic management measures. The influences of these measures on the traffic flow characteristics are discussed in section 6.2.4.

MCSS is the system which provides most data on motorway characteristics. It detects vehicles passing every 500 m on each lane, does calculation on it (checking what neighbouring signs show) and presents the traffic a speed limit on dynamic signs on a portal. In case of communication failure the roadside outstation (LS) can work autonomously (figure 4.6).
One of the features is an automatic data logging system. The data is real-time available through MONICA in all regional and the national traffic management centres. Each minute the average speed and flow are saved. The average over all lanes in each direction is also saved automatically.

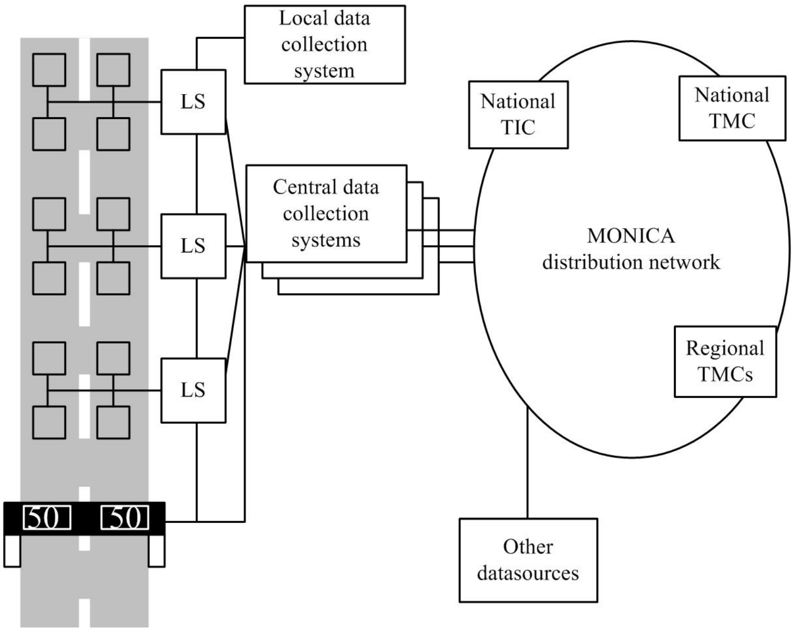

Figure 4.6: The MONItoring CAsco (MONICA) traffic data collection system operational on the Dutch motorway network maintained by the Dutch Ministry of Transport, Public Works and Water Management. Illustration after van Lint (2004). TMC = Traffic Management Centre. TIC = Traffic Information Centre.

The first stretches to be monitored were on parts of the Ring Roads of Amsterdam and Rotterdam and some stretches near important motorway junctions. In 1999 only 15 km of motorway was monitored (next to the MCSS equipped area's) with this system and in 2006 1500 kilometre of motorway (and some im-

portant N-roads) will be monitored next to the motorways which are monitored through the MCSS system. The monitoring system collects data every 0.5 to 5 km, depending on the layout of the road; e.g. no more than one monitoring point is situated between two exits.
One of the research goals is to use existing traffic data collection systems, so no new system has to be build and maintained. Whether cross-sectional data provides enough information or conditioning of some sort is needed is considered in the next paragraph.

4.5 Data conditioning

Data collection failures can be reduced to a minimum by making sure the complete measuring system is well maintained, but there will always be data that has become damaged or disrupted. So before data is used for analysis it should be checked for corrupted or missing measurements and sometimes be 'reconditioned'.

4.5.1 Missing data

The level at which the data is checked is usually the central computer, where the data from road side systems is collected. The first stage is to check data whether it is within the boundaries which are set by physical laws or common sense; e.g. speed on a motorway cannot be negative and length and speed cannot be above a certain level. Usually the collected data which is outside these boundaries is labeled as missing. Other reasons for data to be missing is that is not communicated due to systems failures; e.g. the sensor has been out of order or the communication between sensor and road side station or between road side station and central computer has been out of order for a period of time.

On-line dynamic traffic management systems have default modes or alternative programs in case the input data is missing. In case an on-line system or an analysis needs a full data set interpolation or extrapolation of data can fill the omissions.

4.5.2 Data interpolation and extrapolation

Not only missing data is a reason to interpolate or extrapolate traffic data. Traffic on a motorway is a more or less continuous process which takes place in three dimensions: 2 dimensions in space (along a lane (x-coordinate) and between lanes (y-coordinate)) and the third dimension is time (t-coordinate). Cross sectional traffic measurements as double loop detectors on every lane can only measure a limited part of this continuous process. Only in the dimension time they are continuous; no measurements along a lane or between lanes are done. Measurements are done on all lanes, but vehicle movement from one lane into the next cannot be measured. Aerial photos are cross-sectional in time; they cover a large piece of the x, y plane but also no continuous process is measured; movements from one lane to the next are also not detected using a single aerial photo. An aerial video does give continuous measurements on the data in all three dimensions: x, y and t. Vehicle trajectories (a trajectory of a vehicle describes it's position in time) can be extracted from aerial video material (Hoogendoorn et al., 2003), but this is still a very expensive manner of collecting data. Much more common and less expensive are interpolation and extrapolation techniques to fill in the data that is not measured.

A lot of model techniques have been used to estimate missing or non measured traffic data; piecewise-linear, polynomial functions, splines or wavelets on both microscopic and macroscopic data. There is usually a trade-off between computational effort and accuracy.

Better results can be acquired when the interpolation methods use the characteristics of the traffic flow. Examples of this are autoregressive models, moving averages and neural networks. A technique which performs very well is the adaptive smoothing method, described in Treiber and Helbing (2002) and the next section.

4.5.3 Adaptive smoothing method

On a single cross section the most detailed measurements have information about every single vehicle (time of passage, lane, speed and vehicle length). Most traffic data collection systems give average speed and flow. The most common Dutch motorway traffic data collection system gives 1-minute averages of flow and speed. To interpolate between detection sites (and in time) Treiber and Helbing (2002) developed a new method they called the 'adaptive smoothing method'. A non-linear spatio-temporal low-pass filter was applied to the input detector data. This filter exploits the characteristics of congested data (perturbations travel upstream, while in free traffic information propagates downstream). The results of this filtering technique are velocity and flow as smooth functions of space and time, traffic state visualization, reconstruction of traffic situations from incomplete information, fast identifications of traffic breakdowns and even a short-term traffic forecast.

The method contains the following properties:

- At locations with free flowing traffic, perturbations move in the direction of the flow, with a speed of $\vec{u}_{fr}$. These perturbations should pass the filter.

- In case of congested traffic, perturbations move against the direction of the flow with a speed of $\vec{u}_{co}$. For high densities or low velocities the filter should let perturbations with a velocity near $\vec{u}_{co}$ by.

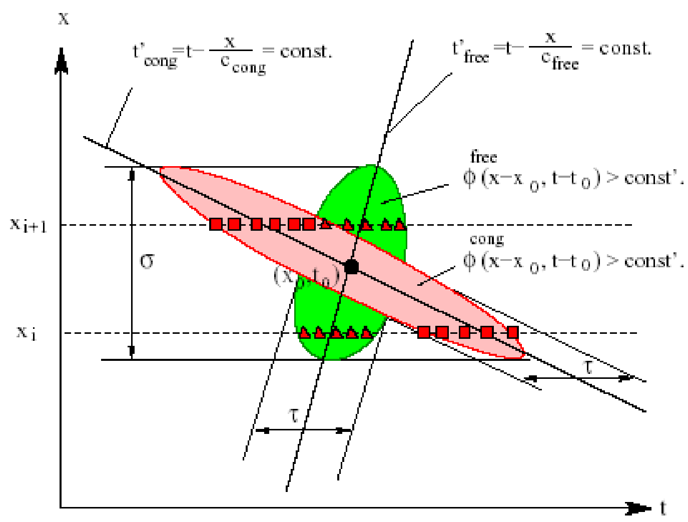

Figure 4.7: Visualization of the effects of linear homogeneous filters (Treiber and Helbing, 2002). The green zone indicates the data used to calculate the characteristics at $[x(t), t]_{t=0}$ under free flow circumstances. The red zone indicates the data used to calculate the characteristics at $[x(t), t]_{t=0}$ under congested circumstances.

- The filter has to smooth out all high frequency fluctuations in the time t on a time scale smaller than τ and fluctuations in space x on a length scale smaller than σ (τ and σ are smoothing factors).

The method provides an estimate for all locations and for all times.

Figure 4.7 demonstrates which data is used to calculate the traffic condition in the centre point $[x(t), t]_{t=0}$. When traffic is flowing free, data from upstream and downstream detectors of the time frames within the green area is used. When traffic is congested, available data within the red area is used for estimating the traffic conditions at point $[x(t), t]_{t=0}$. Thus for the congested and the free flow situation a different kernel is used. The state of the traffic is calculated using free speed and actual speed. Thus the calculated speed for congested and free flow situations determines the state of traffic. More information about the used equations can be found in Treiber and Helbing (2002).

The 'adaptive smoothing method' is used to calculate traffic conditions between measurement points (chapter 9) in order to create an adaptive ramp metering system (based of Alinea and a traffic stability indicator).

4.6 Discussion

This chapter shows that when it comes down to measuring traffic flows there is not always a one on one match between the data needed for research and the data provided by traffic measuring systems. As an example recall that to calibrate macroscopic traffic flow theories and models it would be very helpful if density measurements were available on a large scale. Unfortunately this is not the case. The reason for this is clear: there is not a cost-efficient measuring tool that can measure the data needed. Cross sectional data is broad available. Aerial data, which is expensive and difficult to obtain, is not. therefore traffic density is often estimated using the fundamental relation $q = ku$.

Microscopic traffic flow models would benefit if vehicle trajectories were available on a large scale (Hoogendoorn et al., 2003). As all other traffic data (microscopic, mesoscopic and macroscopic) can be derived from vehicle trajectories it is clear that the effort to obtain this data should be stimulated.

The data needed for the stability indicator based on string stability ranges from microscopic to mesoscopic. It is clear that macroscopic traffic data will not be able to give extra information on stability (the reason for considering traffic stability is that macroscopic variables do not show a one-on-one relation with breakdown frequencies).

It may be clear that in The Netherlands many data collection points, sometimes with very narrow grid length, are available and easy to use. Interpolation techniques as the 'adaptive smoothing method' (Treiber and Helbing, 2002) show better results when measuring points are closer together and on macroscopic data.

One of the research goals is to develop an indicator on traffic stability, using existing data measuring techniques. In this chapter was demonstrated that loop detection is not only the primary source of measurements now, but will still be in the near future. Therefore only data acquired from these loops is used. At an aggregated level the data can be enriched by calculation (adaptive smoothing method), at an individual level this is much more difficult. At the individual vehicle level one is limited to cross sectional information on speed, vehicle length, and time of passage. From this data individual characteristics as headway and gap can be deduced.

In the current configuration the MONICA system does not distinguish between lanes and vehicle classes at all locations. The data from roadside measurements ('raw' data from the roadside station) will be used for the development of the stability indicators in chapter 7. This data is not collected automatically and is also not automatically available at the traffic managements centres. Nevertheless is the availability of this data from all roadside stations guarantied in the near future (RESI-data).

The next chapter will elaborate on what mesoscopic characteristics of the traffic flow are important for traffic flow stability.

Chapter 5

Merging traffic

This chapter focuses on three items. First the window of opportunity of a driver to prevent a collision, when the vehicle in front is reducing it's speed (section 5.1) is discussed.

After that the influences of merging lane length on the position of merging and on the gap acceptance is looked at. Also the time-to-collision of vehicles directly after the merge is discussed.

Finally the average queue length and waiting time a vehicle will encounter when entering a motorway which is under free or constrained flow (section 5.2) is deduced. Those handle traffic stability and focus on traffic disruptions near on ramps.

5.1 Introduction

When a vehicle has to react to a speed fluctuation of its predecessor by braking the gap and the distance the driver has at that moment are the basis for his abilities to prevent a collision.

The gap displays the window of opportunity for a driver to react to the motion of the preceding vehicle (figure 5.1). In order to drive safe, the gap must be larger than the driver's reaction time (assuming that the deceleration characteristics of vehicles are identical). If the reaction time exceeds the gap the reaction to a deceleration of the preceding vehicle will take place downstream of where the leading vehicle started decelerating, which in case of maximum deceleration towards a stand still of the leading vehicle leads to an accident. The space is the physical space the driver has to adapt his vehicle to the speed of the vehicle in front.

In section 2.4 the relations between space and gap (or headway if the occupancy is included) and other traffic relations like speed and flow were showed.

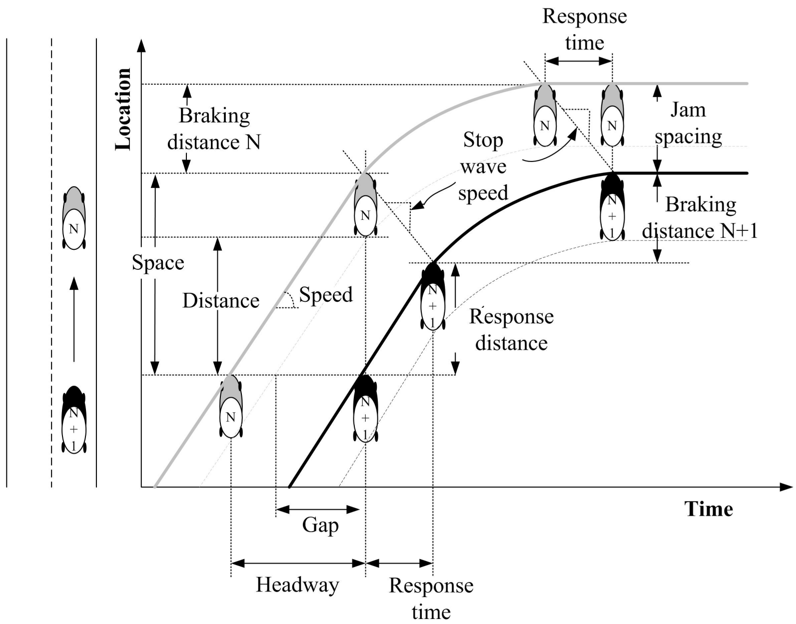

Figure 5.1: Gap and distance determine the drivers 'window of opportunity' in case of a speed reduction of the vehicle in front.

Traffic entering the motorway forms (more or less) a disturbance for traffic travelling on the motorway. How much traffic can enter the motorway is theoretically worked out in the next section, using the availability of critical gaps.

5.2 Merging at on ramps

Because passage times vary inversely with speed, headways and spacing display very different characteristics in flow, congested or not. Higher speeds result in larger spacing as seen above, but larger spacing and higher speeds both have opposite effects on passage times (section 2.4). The gap is the window of opportunity for a driver to react upon the motion of the preceding vehicle. In order to drive safe, the gap must be larger than the driver's reaction time. If the reaction time exceeds the gap the reaction to a deceleration of the preceding vehicle will take place downstream of where the leading vehicle started decelerating, which in case of maximum deceleration of the leading vehicle may lead to an collision.

If no large enough gaps are present, the ability for a driver to react upon decelerations of traffic ahead declines and an unsafe traffic situation appears.

The distributions of headways gives an indication for the possibility for merging and lane changing. The capacity of a section or bottleneck depends on the average minimum headway. At high velocities the difference between the headway and the gap between two vehicles becomes smaller, but for capacity estimates, in which vehicles follow at small distances the difference may not be neglected. However, since the reciprocal of the mean headway is equal to the flow rate, headways remain important from a traffic flow point of view.

5.2.1 Headway distributions

When the distribution of headways is known the flow level can be reconstructed (3600 divided by the average headway). The other way around: a distribution of the headways from traffic volumes cannot be deduced directly. Sadegh-hosseini and Benekohal (2002) collected headways over a wide range of traffic conditions and determined for each data set the minimum, average, median, mode, maximum, standard deviation and average speed (see figure A.1 for their results). They tested both gamma and lognormal models on goodness of fit for wide ranges of flow and found that a shifted lognormal distribution represented for a wide range of volumes the headways best (Hoogendoorn and Bovy, 2000b). Empirical relations between the volume and other characteristics are in figure A.1 in Appendix A.

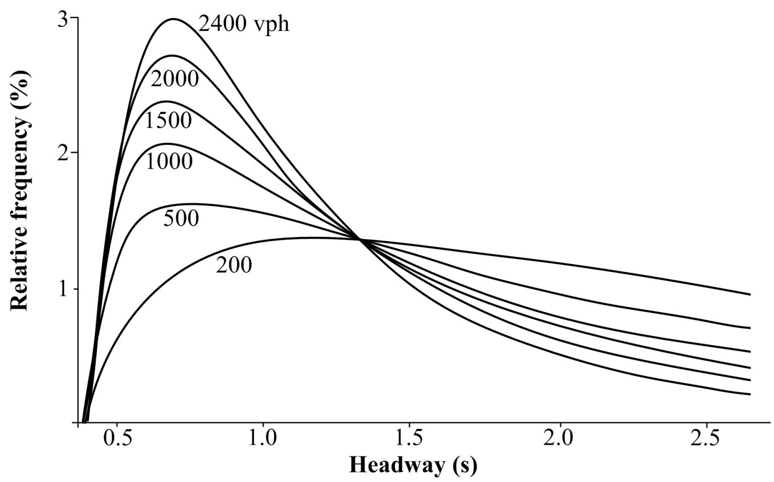

Figure 5.2: Time headway distribution for different hourly volumes. After Sadegh-hosseini and Benekohal (2002).

Spacing, and headway distribution show relations with the three main characteristics. The next paragraph focuses on the distribution of vehicles on the road; both spacing and headway are important characteristics.

5.2.2 Critical gaps

The critical gap is the smallest gap that drivers will accept when merging or crossing a traffic stream. Its actual value will depend on the type of traffic flow, the type of manoeuvre which is needed to merge and the type of merging section. The critical gap varies between motorists and also varies over time for an individual motorist (when long waiting times appear the impatience level rises and smaller gaps get accepted (Minderhout, 1999)).

Traffic behaviour at on ramp merging sections is also influenced by other factors as: congestion level of through traffic, ratio of merging traffic volume to the through traffic volume, lane configurations, sight distance, etc. Because it is variable, it is often measured as the value which is just as often rejected as it is accepted (50th percentile value).

Early efforts for modelling gap acceptance were based on the distribution of the critical gap with no attempt to explain the underlying behaviour ((Herman and Weiss, 1961), (Miller, 1972), and (Van Arem, 1986)). Daganzo (1981) used a probit model to estimate the parameters of a normal distribution of critical gaps to capture the heterogeneity of driver behaviour. Most of the gap acceptance models are aimed at merging traffic at intersections, drivers or pedestrians. The large difference between merging at intersections and motorway on ramps is that at intersections the driver makes choices: what gap to accept, after having observed and judged it. When vehicles start accelerating on the on ramp of a motorway they do not have complete certainty what gaps they are confronted with; which gap are potentially to be accepted or not.

In accepting gaps when merging into the main flow of a motorway the length of the merging lane has a dominant influence. Kita (1993) formulated the problem of gap acceptance at merging points using a binary logit model with the gap length, remaining distance to the end of the acceleration lane, and relative velocity as explanatory variables. Figure 5.3 shows the calibrated distribution of the results of the logit model that predicts the headway after merging as a function of the headway and the distance from the beginning of the merging section. The traffic volume on the right lane of the motorway was 1366 veh/h. It also shows (in red) the cumulative distribution of accepted headways of all merges (form all positions at he merging section).
In the same case study Kita (1993) shows that the time-to-collision after merging also depends of the merge length (figure 5.4).
Most traffic jams on motorways appear at bottlenecks. Merging sections are the most common place for bottlenecks on Dutch motorways. Therefore the most obvious place to measure traffic stability is near bottlenecks (chapter 7).

5.2.3 Queueing theory

Merges of traffic can be modelled microscopic, as in last paragraph. The merge of two traffic flows can also be modelled using queueing theory.

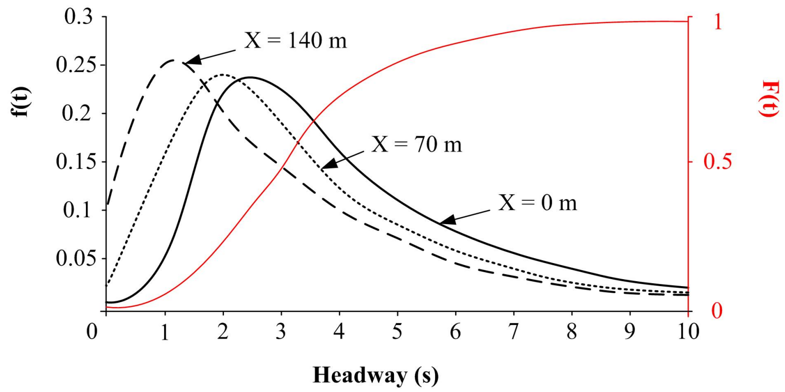

Figure 5.3: The influence of merge lane length on merging and gap acceptance. After Kita (1993).

How long a vehicle has to wait for a stream of free and constrained traffic (Poisson and Erlang distributed) will be calculated.

For these calculations the gap acceptance is held constant and equal for all vehicles. This is a clear simplification, in real life the distribution of accepted gaps shows variation (last section).

Free traffic

Suppose the arrivals at an on ramp are intervals of Poisson a distribution:

$$f(t) = \lambda e^{(-\lambda t)} \tag{5.1}$$

in which λ equals the average arrival rate.

The distribution of gaps at a free flow regime (Exponential distribution, section 3.3.2) is the same accept for a different average:

$$g(t) = \mu e^{(-\mu t)} \tag{5.2}$$

in which μ is the average 'service' rate (the rate at which vehicles merge).

From queueing theory:

Let r be the rate of arrivals over service (λ/μ), then the mean number of customers (L) in a system (drivers in a queue) is:

$$L = m\frac{r}{1-r} \tag{5.3}$$

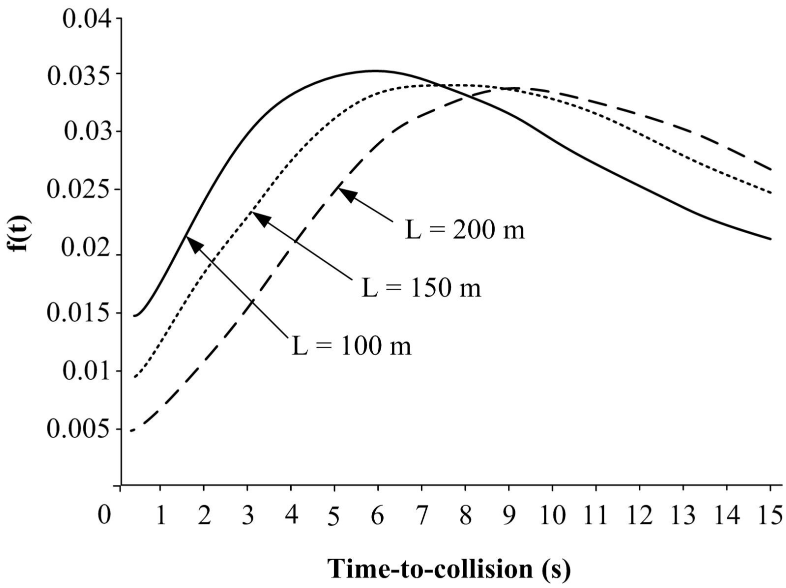

Figure 5.4: The influence of merge lane length on merging and the resulting time-to-collision. After Kita (1993).

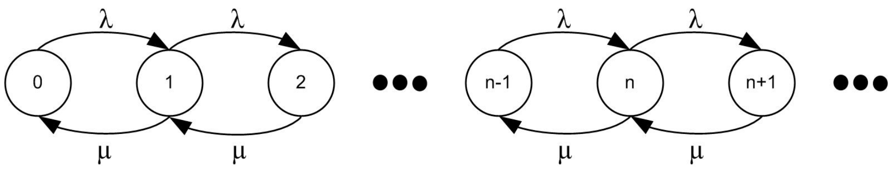

Figure 5.5: Queueing system with unlimited waiting capacity. n equals the number of vehicles in the queue.

and the mean waiting time (W)at the on ramp is:

$$W = m\frac{1}{\mu(1-r)} \tag{5.4}$$

with m constant and usually between 0.5 and 1.

A queueing model can help to solve the M/M/1 (notation from Kendall (1953)) queueing problem described above(see figure 5.5):
The set of time-equations of the M/M/1 system is:

$$\begin{cases} \frac{dp_0}{dt} = -\lambda p_0(t) + \mu p_1(t) \\ \frac{dp_n}{dt} = -(\lambda+\mu)p_n(t) + \lambda p_{n-1}(t) + \mu p_{n+1}(t) \qquad n = 1, 2, 3, ... \end{cases} \tag{5.5}$$

in which $p_n(t)$ equals the probability of n vehicles in the queue.
The solution of this set of equations (Clarke, 1953) is:

$$\begin{aligned} P_n(t) = e^{\lambda+\mu}t \quad & [(\sqrt{\tfrac{\mu}{\lambda}})^{i-n} I_{n-i}(2\sqrt{\lambda\mu}t) \\ & +(\sqrt{\tfrac{\mu}{\lambda}})^{i-n+1} I_{n+i+1}(2\sqrt{\lambda\mu}t) + (1-\tfrac{\lambda}{\mu})(\tfrac{\lambda}{\mu})^n \\ & \textstyle\sum_{k=n+i+2}^{\infty}(\sqrt{\tfrac{\mu}{\lambda}})^k I_k(2\sqrt{\lambda\mu}t)] \end{aligned} \tag{5.6}$$

In which λ is the mean arrival rate, μ is the mean service rate, and $I_v(z)$ is the modified Bessel function, given by:

$$I_v(z) = \sum_{k=0}^{\infty} \frac{(z/2)^{v+2k}}{k!\Gamma(v+k+1)} \tag{5.7}$$

Constrained traffic

A queue at an on ramp when traffic conditions on the motorway are free flow is not very realistic. More realistic are constrained conditions on the motorway. The distribution of gaps at the constrained regime can be modelled using the Erlang distribution (see section 3.3.2):

$$f(t,k) = \frac{(ks)(kst)^{k-1}e^{-kst}}{(k-1)!} \tag{5.8}$$

with t equals time and k depends of the flow rate $(2, 3, 4, ..)$, and s the phase of service.
The M/Ek/1 (notation from Kendall (1953)) queueing model is more difficult to solve than the M/M/1 queueing model:

- Let $p_{n,s}(t)$ be the probability of n vehicles having joined the queue at time t, and the vehicle that is merging is in phase s (s=1,....,k but in constrained traffic s equals 2).
- The service times are assumed to have an Erlang distribution with mean time $1/\mu$ and parameter k, and the arrivals at the on ramp to be random at mean rate λ.
- Let $p_0(t)$ denote the probability that there are no vehicles in line at time t.

From these assumptions the following steady-state balance equations are deduced:

$$\begin{aligned} & 0 = -(\lambda + k\mu)p_{n,s} + k\mu p_{n,s+1} + \lambda p_{n-1,s} && (n \geq 2, 1 \leq s \leq k-1), \\ & 0 = -(\lambda + k\mu)p_{n,k} + k\mu p_{n+1,1} + \lambda p_{n-1,k} && (n \geq 2), \\ & 0 = -(\lambda + k\mu)p_{1,s} + k\mu p_{1,s+1} && (1 \leq s \leq k-1), \\ & 0 = -(\lambda + k\mu)p_{1,k} + k\mu p_{2,1} + \lambda p_0, && \\ & 0 = -\lambda p_0 + k\mu p_{1,1}. && \end{aligned} \tag{5.9}$$

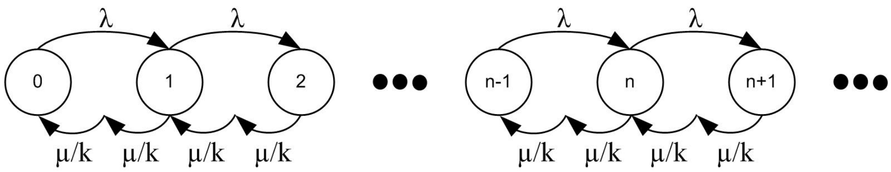

Figure 5.6: Queueing system Erlang-k service rate, k=2 is drawn.

The solution of these equations can be found in (Griffths et al., 2004) but it goes to far to deduct the solution here. The mean queue length and the mean waiting time are deduced as in (Cross and Harris, 1998). The mean waiting time (W_q) of an M/Ek/1 system is:

$$W_q = \frac{k+1}{2k} \frac{\rho}{\mu(1-\rho)} \qquad (\rho = \lambda/\mu). \tag{5.10}$$

The mean queue length (L_q) can be deducted from the waiting time:

$$L_q = \lambda W_q = \frac{k+1}{2k} \frac{\rho^2}{1-\rho}. \tag{5.11}$$

Example 1 Suppose vehicles at an on ramp arrive according to a Poisson distribution, at an average of 15 seconds (4 per minute). Vehicles (from the front of the queue) entering the motorway can enter at a rate of 8 per minute (7.5 seconds) with a standard deviation of 5.3 seconds. The amount of vehicles in the queue is unlimited. Form this is calculated: $\mu = 1/8$ min and $\sigma^2 = 1/(k\mu^2) = (5.3/60)^2 = 1/128$ from which follows that k=2. Thus there is a M/Ek/1 model with $\rho = 4/8$. The waiting time follows from equation (5.10) and is about 6 seconds. The mean queue length follows from equation 5.11 and is about 3/8 vehicle. Because the entrance rate is much higher than the rate of arrival the queue will not get large.

This section has showed how queueing theory can be used to predict waiting times at on ramps. The waiting time is one of the many factors that influence gap acceptance (of which several were discussed in this chapter).

In chapter 9 several combinations of flow and on ramp traffic demand to compare the Alinea and the enhanced Alinea algorithm are used. The queueing models give a theoretical background for choosing proper combinations. This way one does not need to do a lot of testing to be able to find good combinations of the main traffic flow and the on ramp traffic flow.

5.3 Summary

In this chapter the influence of microscopic traffic characteristics (as the distributions of headways or gap acceptance) is underlined. Macroscopic characteristics give a general idea on merging possibility, but merging will only take place after a headway has been approved for merging.

The distribution of vehicles in the traffic flow influences the window of opportunity for a driver to react to a speed reduction of it's predecessor and the distribution of vehicles (in time: headways) determines the rate and waiting time of vehicles entering the motorway.

As on the Dutch motorways the most frequent locations where breakdown occurs are on ramps and motorway merging sections, improvement of the traffic merging can be beneficial for the complete motorway network.

Average waiting times and queue lengths at on ramps can be calculated using queueing theory. Two examples were showed, and the consequences of the difference between free and constrained traffic on the main lanes of the motorway are explained.

The mesoscopic level, headway distribution, has proved to be very appropriate to calculate queue lengths and waiting times. It will be the mesoscopic level in which the developments of the stability indicator in chapter 7 will be tested.

Chapter 6

Driver influences

Drivers play a central role in the traffic system. Next to their personal characteristics (section 6.1) they are limited and influenced by external influences (section 6.2). Also the characteristics of their vehicles (6.2.1) and the road and it's traffic rules (6.2.2) play an important role in the traffic system. External influences as the weather (6.2.3) and DTM-measures (6.2.4) may influence the driver's behaviour directly or through the traffic rules or the road itself (figure 6.1).
Motorway engineers strive towards two goals when designing a motorway: a high level of service (seek to minimize travel time and delays) and provide a high level of safety ((Transportation Research Board, 2000) and (AVV, 1999a)). To provide efficient and safe motorway transportation and do research into traffic stability, a knowledge of the characteristics and the limitations of each of its components is essential. It is also important to be aware of the interrelationships (often non-linear) that exist among the components in order to determine the effect, if any, that they have on each other. A road must be designed to accommodate a wide range of vehicle characteristics and at the same time allow use by drivers with a wide range of physical and psychological characteristics.

This chapter will show how drivers are influenced in their behaviour by their surrounding (internal and external influences). The behaviour of the driver will be of importance for the traffic stability (e.g. stressful behaviour like tailgating or a very large reaction time are characteristics of which is later showed that they reduce traffic stability). Detailed research into traffic stability is only useful when very influential external influences are accounted for.

6.1 Driver characteristics

Although drivers may vary significantly, the design of a road and it's surrounding should be supportive to (almost) all drivers (the use of an average value, such as mean reaction time, is not adequate for a large number of drivers; the 95th percentile might give better design criteria), so they can all fulfil their

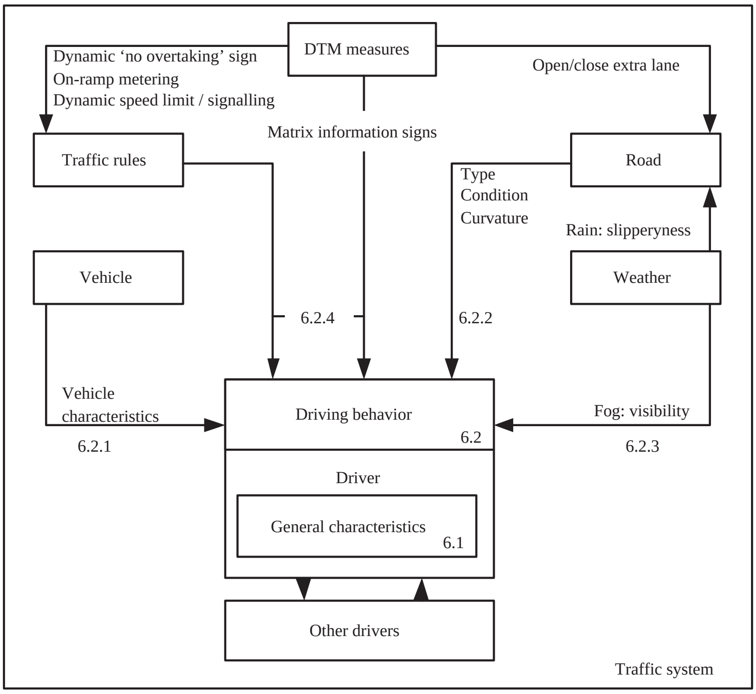

Figure 6.1: Internal and external influences on drivers discussed in this chapter. The numbers indicate the paragraph in which the topic is discussed. Some arrows show an example.

driver tasks. Actions taken by drivers on a road result form their evaluation of (or judgement and reaction to) information they obtain from certain stimuli that they see or hear (Jessurun, 1997). The information observed by drivers on a motorway is constantly changing.

Driver tasks can be split into (TRB, 1992):

- Observe; information which is needed should be within the functional scope; the faster a vehicle drives the more the driver is focussed around a point straight ahead (AVV, 1999a). For elements which are not always visible (lines between lanes in heavy rain) extra information should be present to let the driver be able to make good assumptions. Every deviation from the standard will make it more difficult for the driver to make this connection.
- Judge; he information which is observed must be translated to be able to judge events.
- Decide; decisions should not be to complicated (a string of single decisions is preferred over one multiple decision).

- Starting a manoeuvre.
- Finishing a manoeuvre (which should not be to complex).

The total time from observation to reaction is the key indicator for motorway design. The time which is needed varies with the complexity of the manoeuvre which is needed. When any aspect of the list above takes more time the total length needed for processing information is longer. Individual difference between drivers (e.g. driving style), therefore have a significant impact on driver manoeuvres (Edwards, 1999). Different types of drivers can be distinguished, for example an aggressive or defensive style of driving. Driving behaviour is not constant, but varies over time and space; over a short time, e.g. when the driver is distracted by something or over a long time, because of aging. Also travel motive plays a role in driving behaviour. General driver characteristics can be distinguished from motivation for driving behaviour (figure 6.1). General driver characteristics differ per person but are constant over a trip. The motivation for driving behaviour is influenced by factors as visibility, road type and DTM measures, which vary.

6.1.1 General characteristics

Individual characteristics

Variability among drivers, especially associated with variables such as gender, age, education, health, etc. goes by the name of individual differences. Only few of these are interesting for the characterization of a traffic flow or for modelling a traffic flow. Only variables which directly influence the path and the velocity of the driven vehicle are of interest.

For the purpose of traffic flow analysis, performance differences between men and women may be ignored (Kroemer et al., 1994). Age on itself is also a poor predictor of performance. But changes in visual perception and cognitive performance, which, although there are some exceptions, decline over the years are better indicators.

Visual perception

The feedback a driver gets while driving are visual (‘What traffic situation am I facing?’), acoustic (engine, wind noise) and haptic (vibrations, g-forces). The last two are compensated when the comfort of driving gets below a certain level. The most important, the visual feedback, is limited by human perception. A difference in surrounding can only be noticed when the difference between the original vision and the changed vision is above a certain threshold (Weber, 1905). The size of an image (C) at distance (X) from a driver’s eye of for instance a vehicle in the total surrounding plays an important part. When a relative movement between a driver and his or her predecessor takes place the size of the image on the retina changes. This change in size must happen

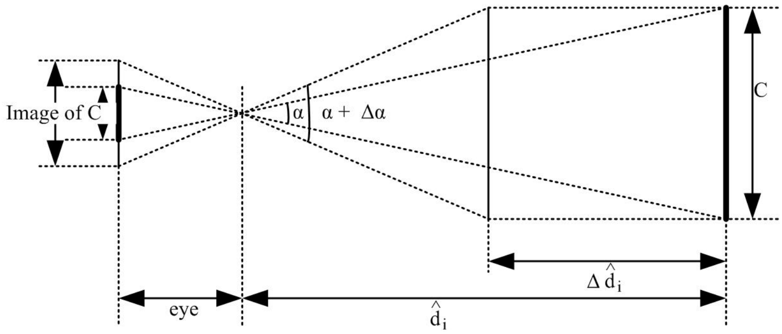

Figure 6.2: Observation of relative movements.

at a certain speed for the driver to be able to witness a change.

The incentives for the change of images can be quantified.
For small angles α (figure 6.2) follows

$$\alpha = \tan\alpha = \frac{C}{\hat{d}_i} \tag{6.1}$$

The angle on the retina (α) equals:

$$\alpha + \Delta\alpha = \frac{C}{\hat{d}_i - \Delta\hat{d}_i} \tag{6.2}$$

The change of the angle on the retina $\Delta\alpha$ can be calculated:

$$\Delta\alpha = \frac{C}{\hat{d}_i - \Delta\hat{d}_i} - \frac{C}{\hat{d}_i} = \frac{C \cdot \Delta\hat{d}_i}{\hat{d}_i^2 - \hat{d}_i \cdot \Delta\hat{d}_i} \tag{6.3}$$

The change of the distance between the leading vehicle and the observer in the following vehicle equals $\Delta\hat{d}_i = \Delta\hat{u}_i\Delta t$ and the angle on the retina equals:

$$\Delta\alpha = \frac{C \cdot \Delta\hat{u}_i \cdot \Delta t}{\hat{d}_i^2 - \hat{d}_i \cdot \Delta\hat{u}_i \cdot \Delta t} \tag{6.4}$$

Divided by Δt the rate at which the angle changes is:

$$\frac{\Delta\alpha}{\Delta t} = \frac{C \cdot \Delta\hat{u}_i}{\hat{d}_i^2 - \hat{d}_i \cdot \Delta\hat{u}_i \cdot \Delta t} \tag{6.5}$$

At comfortable speed changes the value of $\Delta\hat{u}_i$ is small compared to $\hat{d}_i$, so the rate of the change of the angle can be rewritten or estimated by:

$$\frac{\Delta\alpha}{\Delta t} = \frac{C \cdot \Delta\hat{u}_i}{\hat{d}_i^2} \tag{6.6}$$

Recall that $\hat{d}_i = \hat{x}_i - \hat{x}_{i-1}$ and also the car-following rules used in the GM-car-following model (equation 3.4:

$$\hat{a}_{i+1}(t+\Delta t) = \frac{\alpha_{l,m}[\hat{u}_{i+1}(t+\Delta t)]^m[\hat{u}_i(t) - \hat{u}_{i+1}(t)]}{[\hat{x}_i(t) - \hat{x}_{i+1}(t)]^l} \tag{6.7}$$

and notice that the GM-car-following model for $l = 2$ and $m = 0$, which will be used to calculate string stability in chapter 7, is equal to the car-following behaviour one would expect based on the way of human observation.

Changes in visual perception

Older drivers lose visual acuity (Marmor, 1982). Because of scattering in the opting train and light losses visual performance in darkness declines dramatically (Blackwell and Blackwell, 1971). A 55 year old takes 8 times as much time to recover from a glare in a headlight as a 16 year old does, as is literally blind for many seconds after exposure (Fox, 1989). To compensate for the loss of sight in darkness and the extra reaction time as a consequence drivers create a larger distance from their vehicle to the predecessor. In darkness road capacity is on average 5% less than in daylight (Van Toorenburg, 1986).

Changes in cognitive performance

Older drivers experience problems in ignoring irrelevant information and correctly identify hints (McPherson et al., 1988). Forced pace tasks under stressful conditions may disrupt the performance of older drivers. They attempt to compensate this by slowing down (McPherson et al., 1988). A large percentage of older drivers may result in vehicles that lag behind and obstruct the flow. The extra large gaps of these drivers and the smoothing effect they have will probably have a positive effect on traffic stability, but a reduction of capacity in also a consequence.

Hearing perception

The ear receives sound stimuli, which is important to drivers only when warning sounds, usually given out by emergency vehicles, are to be detected. Loss of some hearing ability is not a serious problem and can usually be corrected by a hearing aid.

Observation-reaction time

During all phases of the driver task time elapses. The time that elapses during all phases (from observation to and including execution of the manoeuvre) is called perception-reaction time. The perception-reaction time is an important factor in the determination of the braking distance, which in turn dictates the minimum sight distance needed to come to a complete stop within the length of road which can be overseen. As all visual and cognitive qualities, also the observation-reaction time varies over the drivers and even for the same driver under different circumstances (level of attention).

The influences of changes in visual perception, in the cognitive train and reaction time have a direct influence on the car-following and therefore on string stability.

6.2 Motivation for driving behaviour

6.2.1 Vehicle characteristics

Criteria for geometric design of motorways are partly based on static, kinematic, and dynamic characteristics of vehicles (Garber and Hoel, 2002). Static characteristics include the weight and size of the vehicle; kinematic characteristics involve the motion of the vehicle, without considering the forces that cause the motion. Dynamic characteristics involve the forces that cause the motion of the vehicle. Since nearly all motorways carry both passenger cars and truck traffic, it is essential that the design criteria take into account the characteristics of different types of vehicles. Often a design vehicle is chosen for designing motorways. The vehicle should have characteristics that encompass those of nearly all other vehicles expected to use the motorway. The characteristics of the design vehicle are then used to determine criteria for geometric design. Without going into detail about what demands static and kinematic vehicle characteristics set, a list of some general demands of vehicle characteristics to the road is presented:

- Enough room for accelerating and decelerating for vehicles to change lanes, from the fast to the slow lane and vice versa.
- Take into consideration the parts of the road which cannot be overseen by the driver, because his vehicle prevents vision; especially the dead angles of vehicles are responsible for extra design demands.
- Vehicle driving forces should be compensated to prevent skidding (super elevation in curves).
- The lanes should be wide enough for all the vehicles (curve widening and speed dependent width).

Because the specifications of vehicles change every year the demands these specifications set to the road design change continuous. One should not only require knowledge of the characteristics of the vehicles commonly found in the traffic flow (dimensions, visibility, manoeuvrability, maximum speed, acceleration, deceleration, hill climbing ability, steering cornering and lighting (Lay, 1986b)) but also the composition of traffic is important for the behaviour of the flow. Table 6.1 gives on overview of Dutch traffic over de last 20 years.

Table 6.1: Composition of traffic on Dutch motorways (percentage).

Year	number of counting days	Passenger cars	Trucks	Vans	Busses	Motorcycles
1980	8	84,3	12,1	2,6	0,6	0,4
1985	4	84,4	12,1	2,6	0,6	0,3
1990	2	84,4	12,2	2,5	0,5	0,4
1995	250	81,7	14,2	2,9	0,7	0,5
2000	248	80,3	15,6	3,0	0,7	0,4
2001	250	81,2	14,8	3,0	0,6	0,4

The table shows a clear difference, due to the change of the numbers of counting days. The days counted prior to 1995 apparently were not representative for al of that year. There is no reason found that an inflection took place in the early 90's.

The composition of traffic is one of the factors that influence the onset of congestion. Often platoons form behind trucks, and a traffic flow with a lot of platoons shows different behaviour than one without these platoon (AVV, 1998). The overall percentage of trucks in the Netherlands did not change dramatically over the years, but the characteristics did; especially in the 1980's the trucks became more powerful and their speed was limited.

Not only the composition, also the equipment of the vehicles influences driving behaviour; Minderhout (1999) showed that the impact of driver support systems on motorway capacity are significant.

Behavioural adaptation to more general characteristics of the vehicle are however not found in a recent study of Kijk in de Vegte (2002). Vehicles on the A2 motorway in The Netherlands were divided into two classes; leaders and (very) close followers (tailgating). In the group of close followers a higher percentage of airbags and ABS-braking systems was expected, but not found.

6.2.2 Road geometry

The basic principle of geometric design is that the appearance of the road should clearly indicate to the driver the speed and the path that his vehicle should adopt in order to proceed in comfort, economy and safety. If this is not possible, additional devices such as signing, pavement marking, signals and lighting should be used (Lay, 1986a). Traffic capacity will often dominate urban and arterial designs whereas safety and operating conditions are usually the key factors in rural design.

Motorway design in the Netherlands

Also in The Netherlands the design of motorways has a central idea; of a good functioning road (AVV, 1993). The design should provide enough capacity, comfort and safety. The components road system, driver, and the vehicle should function well together.

The road and its surrounding

Besides that design should supply comfort, safety and economy, the major traffic elements that influence highway design are Average Daily Traffic, Design Hour Volume, Directional distribution, Percentage of Trucks and Design speed. Especially near ramps and at junctions, where most of the relatively 'difficult' decisions have to be made special attention in design is needed to create separate decision points for drivers. Because connections to the motorway are close to each other in the Netherlands (8 km and 4 in the urbanized 'Randstad') most bottlenecks are infrastructure related and the phenomenon of 'the phantom jam' is not a problem in The Netherlands.

6.2.3 Weather

Motorways, with their segregated, single directional flow, are perceived as one of the safest and fastest road environments to drive in. The degree of hazard the driver encounters on these roads is partially the result of influences as weather, and the speed the drivers chose. Changes in weather have an immediate effect on the driving environment. Controlling a vehicle is more difficult in windy conditions, in rain, snow and fog. Wet and slippery road surfaces increase the likeliness of skidding.

Light conditions

In darkness road capacity is on average 5% less than in daylight (Van Toorenburg, 1986). This is an average and the exact percentage is location dependent. (95% confidence interval is between 3 and 7 %). To reduce this sometimes lampposts are placed and the effect is reduced to 3%.

Weather conditions

At less favourable weather conditions capacity will reduce, compared to 'ideal' weather conditions. The reduction follows the increase of headway, and if weather conditions worsen, a reduction in speed as well. The reduction depends on the type of precipitation (light rain, heavy rain, snow or fog), of which the rain is most important in the Netherlands, and the type of road surface (is the asphalt closed or porous to enable to drain the water through the

road surface, instead of over it's surface). The capacity reduction at non-ideal weather conditions is 9% for rainy conditions on an closed asphalt concrete and 6% for rainy conditions on porous asphalt (AVV, 1999a). The decrease of the capacity is partially due to a significant reduction of speed, although the drivers acknowledge the need to reduce their speeds they don't do it enough to compensate for the increased risk of accidents due to the increased stopping distance (Edwards, 1999).
In the development of the stability indicator in chapter 7 can be seen that the reduction of the speed of the drivers plays a role which is less significant than the spacing of the vehicles in case of a speed reduction.

6.2.4 Traffic rules and dynamic traffic management

Next to static traffic rules as the general speed limit, the right of way rules and no right-hand overtaking, Dynamic Traffic Management rules can be issued by the traffic operator.

As definition of Dynamic Traffic Management (DTM) is used:

Dynamic Traffic Management is influencing demand and supply of traffic services in both space and time, and based on the condition of the traffic system. Goal is to get a traffic system which is functioning as good as possible (Van Berkum, 1999).

Traffic managers can make decisions and take DTM measures which influence infrastructure (apply dynamic speed limits, open or close rush-hour lanes, turn ramp metering on or off), driver behaviour (influence route choice through information panels) and the interaction between demand and supply (toll systems with dynamic rates).

DTM measures and their effects

In this section an overview of dynamic traffic management measures in the Netherlands and what the consequences for the traffic flow are is discussed.

Figure 6.3: Model for the influence of traffic measures on the traffic system, after Van Altena and Veling (2000). Notice that this model fits within the model of internal and external influences (figure 6.1).

The effect of (dynamic) traffic management measures can be seen as a result of two processes (figure 6.3); the driver behaviour and the changes of the traffic system. Figure 6.1 shows that the influence of DTM measures on drivers can be 'direct' or through the traffic rules or the road alignment. The driver behaviour is the input for the traffic and transport system and the output is the overall effect on the traffic system.

Next an overview is presented of the most common DTM measures in The Netherlands and their effect on the traffic system according to the Transportation Research Centre of the Dutch Ministry of Transport, Public Works and Water Management.

Signalling

In the Netherlands 1000 kilometre of the 2200 kilometre of motorway is equipped with the Motorway Control and Signalling System (or Motorway Traffic Management system) or in short: signalling system. The reasons that lead to the installation and expansion of the system are (AVV, 1999b):

- Every year the Dutch motorways are more heavily used; traffic flows are growing and traffic jams are increasing. The probability of an accident rises and the possibility to do road works during daytime are decreasing.
- Only a few (new) roads will be built in the coming years and the capacity of existing roads must be used to it's maximum.
- Unforseen situations usually lead to complex accidents. Over and over again it is demonstrated that sudden changes from the usual traffic process lead to a higher accident rate. therefore it is essential to warn the traffic in advance of unusual situations as speed drops due to accidents, traffic jams or road works.

The signalling system measures speed and flow of the passing vehicles by double loop detectors which are embedded in the road surface (chapter 4). Flow disruptions are detected automatically and measures are taken to warn the oncoming traffic by putting speed limits above the road on matrix signs. This is the most important feature, but besides this full automatic function the traffic manager can put measures in the system manually as a fog warning, speed limits, multi signs (e.g. no overtaking for trucks sign), roadblocks or lane blocks in case of road reconstructions, tidal flow lanes and shoulder lane use. The signalling system is the system in which most motorway measures come together.
The effects of (the automatic queue protection warning system part of) signalling are:

- A marginal capacity increase (1-2 %) with a maximum of 5 %,
- Higher traffic flows in jammed traffic (4-5%),
- A reduction of the number of shockwaves (50%),
- A reduction of the number of accidents (15-25 %) and the number of secondary accidents at the tail of the queue (40-50%).

The reduction of the number of shockwaves and the number of secondary accident is a clear indication of the increase of traffic stability.

Ramp metering

Ramp control can be defined as a method of improving overall motorway operations by limiting, regulating and timing the entrance of vehicles from one or more ramps onto the shoulder lane of the main road. At locations where motorway entrance ramps have adequate storage capacity or where the surrounding street network can accommodate additional traffic, ramp control systems can provide substantial operational improvements under certain combinations of traffic demand and motorway capacity (Blumentritt et al., 1981). Ramp metering in The Netherlands is effected with an ordinary three-indication traffic signal which controls the traffic on the on ramp. Chapter 9 will elaborate on ramp-metering. Different algorithms, which have different effects on traffic flow stability are presented. An algorithm, which recognizes the need for traffic flow stability is developed en compared to the Alinea algorithm. The overall results of ramp metering in The Netherlands are:

- Capacity increase (0-5%),
- Travel time reduction (6% overall to 15% during congestion) on main road,
- Travel time increase on the on ramps varies a lot,
- Reduction of on ramp traffic (0-35%),
- Red light running (1-13%).

The effects of ramp metering are very location dependent. The relation between ramp metering and the need for traffic stability is very clear. Especially at on ramps traffic is confronted with disruptions, consisting of vehicles entering the motorway. Especially in those cases it can be beneficial to know whether or not the flow is able to handle these disruptions without becoming less stable or breaking down.

Matrix information signs

Matrix information signs in The Netherlands are mostly used for giving information on the length of traffic jams. On some locations signs are used to predict travel times or give accident information. Matrix information signs, called DRIPs (Dynamic Route Information Panel) are mostly located at positions at which drivers can choose between two alternatives routes; the number of kilometres of jammed traffic on both alternatives is reported. This way the drivers can make their route choice based on up to date traffic information. In The Netherlands over 100 matrix information signs are operational on motorways.

The effects of route information on matrix signs is a change of route choice of a part of the drivers. This effect is stronger outside the rush hour and in time of accidents (drivers are better persuaded to update their route choice if an argument for the traffic jam, e.g. an accident, is given). The effects on the lower roads is minimal, except when traffic jams are very large (on both alternatives). Only when the total capacity in a network is larger than the demand a capacity increase of 5 % can be measured. It must be clear that not a 'real' capacity increase is reached on a single cross-section, but the demand is spread over the network, and less traffic jams are formed. So on a network level it is legitimate to assume that capacity increases (Van Altena and Veling, 2000).

Dynamic 'no overtaking heavy vehicles'

The effects of dynamic 'no overtaking heavy vehicles' have only been researched recently in (AVV, 1998). The effects of the rules were:

- Capacity increase (3%),
- Higher average speeds outside traffic jams (1-2%),
- Higher average speeds inside traffic jams (1-2%),
- Less short headways on left (fast) lane and more short headways on the right (slow) lane.

Based on the effects of tests on some locations the rule was extended and will soon (announcement of Dutch Ministry of Transport, Public Works and Water Management in July 2004) all trucks will be banned from the left lane of the motorway.

A possible downside of the rule could be that more short headways on the right lane (which was one of the effects of the rules) reduces the amount of acceptable gaps for merging traffic. So a possible downside of the measure could be the onset of traffic jams because of non fluent merging of on ramp traffic. This effect is researched in the next paragraph.

Figure 6.4: Signalling and ramp metering. Pictures by www.trafficlinq.nl

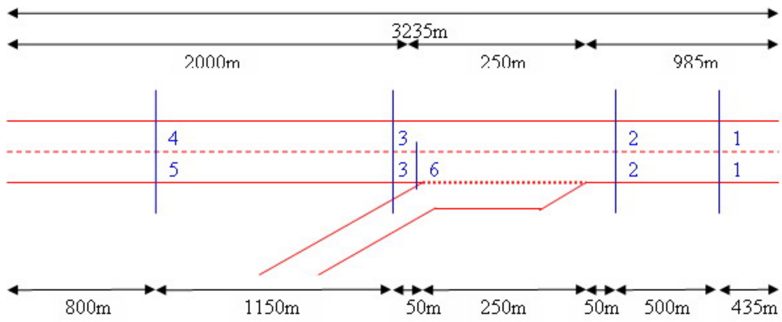

Figure 6.5: Study area of different types of homogeneous traffic.

6.3 Measures to homogenize traffic

One of the leading works in the field of traffic, the Highway Capacity Manual (Transportation Research Board, 2000), does not contain a definition of homogeneous traffic, although the term homogeneous traffic is used often in relation to traffic congestion related problems; the issue of homogenizing traffic has been related to different characteristics of the flow as the distributions of speed, density and flow, lane changes, time-to-collisions on different levels of aggregation.

When homogeneity measures are taken, for instance by introducing a lower speed limit, flow and density become more balanced [1] (Heidemij, 1993). Also speed differences (and time-to-collision) between consecutive vehicles reduce (Grontmij, 1998). However, the main goal of these homogeneity measures, i.e. to increase capacity, is not always achieved ((Heidemij, 1993) and (Grontmij, 1998)). Heidemij (1993) researched the effect of measures to homogenize the traffic flow by comparing 'before' and 'after' fundamental diagrams. The maximum flow had dropped with the measures on some sections and on some sections no significant changes were measured. The total amount of traffic had risen a bit which was in line with the risen amount of traffic jams. The distribution of the vehicles over the lanes of the motorway did get more homogeneous (5% more vehicles on the right lane at the same flow level). No capacity changes near on ramps were measured due to the measures to homogenize traffic.

To test the influence of homogeneity of main road traffic on on ramp traffic and on ramp capacity an on ramp in modelled using Aimsun2. The bottleneck under consideration was a 2-lane stretch of motorway with an on ramp of 300 m, that contains a very high rate of vehicles, at which, when no measures are taken, as a result congestion occurs. The flow rates were 3108 veh/h on the motorway and 856 veh/h on the on ramp. 80% of the vehicles are passenger cars, 10% are light trucks (1.5 passenger car equivalent) and 10% are heavy trucks (2 passenger car equivalent).

[1] A reduction of the differences in flow, speed and occupancy between lanes

Detectors 1 to 3 are used to measure speed and flow in order to calculate capacity using the product-limit estimation method. The detectors 4 and 5 are used to measure the different aspects of homogeneous traffic.

The impact of the different aspects of homogenous traffic were modelled; for five aspects of homogeneity the original capacity of the bottleneck in the case where no measures are taken is compared with the capacity when the flow is homogenized on a specific aspect. The study quantifies the effect on capacity of several aspects of homogeneity. No measures that must be implemented to actually create these levels of homogeneity were considered. The five aspects of homogeneous traffic were implemented in the model separately are:

- Reduction of the speed differences of successive vehicles,
- Reduction of the speed differences between vehicles in different lanes,
- Reduction of the number of lane changes,
- Reduction of the number of clusters of vehicles, and
- Reduction of the number of very small time-to-collisions.

As said, the (Kaplan-Meier) product-limit estimation method was used to measure the bottlenecks capacity.

Table 6.2: Results of homogenization tests. The capacity of the non-homogenized traffic flow is 4106 PCE.

Capacity in PCE (and 95% confidence interval)		
Homogeneity characteristics. Reduction of	*Homogeneous*	*Performance*
speed differences within lanes	3968 (3940-4030)	-3.4%
speed differences between lanes	3881 (3824-3950)	-5.5%
number of lane changes	3988 (3940-4009)	-2.9%
number of clusters of vehicles	3954 (3954-4016)	-2.9%
time-to-collision per vehicle	4016 (3961-4043)	-2.2%

All scenario's show the same results: a reduction of capacity when traffic is more homogeneous (table 6.2). This is contrary to what is common believe: the more homogeneous a traffic flow is, the better it can cope with disruptions. All model outcomes show the same results, i.e. when the traffic flow becomes more homogeneous the capacity of the bottleneck just downstream the on ramp decreases. After inspection of the simulations it appeared that the process of merging from the on ramp onto the motorway becomes a larger problem when the traffic flow is more homogeneous. A possible explanation for this is that homogenizing a traffic flow reduces the number of acceptable gaps for merging traffic. To investigate this the distribution of gaps was measured on the right-hand lane, just prior to the merging section at location 6. The results are found in figure 6.6.

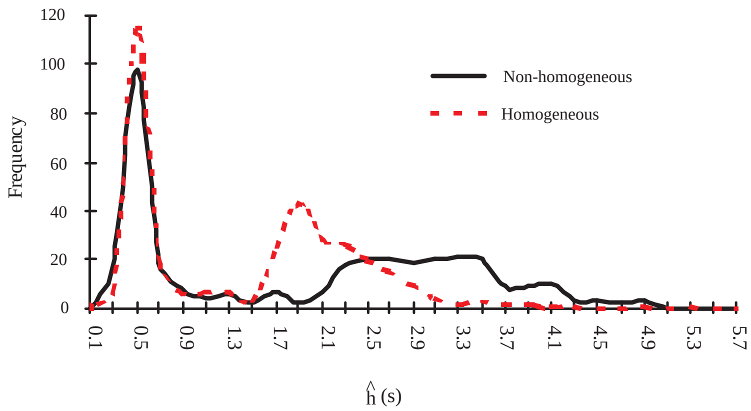

Figure 6.6: Headway distribution on the right lane with and without measures to create homogeneous traffic.

When traffic is congested the homogeneous and non-homogeneous case are almost identical (first peak), but the average headway of the second peak is significantly larger when the traffic is not homogenized. The effect of the difference in probability density function (p.d.f) of headway is obvious; when a vehicle from the on ramp wants to merge it has a larger probability being confronted with an acceptable gap (a gap larger than the minimum acceptable gap).

Although only one on ramp configuration was tested and one model was used, it can be concluded that aspects of homogeneity which are presumed to have a positive effect on capacity do not always have this positive effect at on ramp bottlenecks. Systematic 'redistribution' of vehicles (and gaps) in the main flow influences the number of opportunities for on ramp vehicles to merge into the main flow (in a negative way) and can cause a traffic jam on the on ramp and eventually on the motorway itself. Homogeneity measures and ramp metering are usually taken and installed to keep traffic flowing in bottlenecks. Traditionally these measures do not anticipate on each others effect, but the effect of homogeneity on bottleneck capacity is negative if the traffic on this on ramp is not guided or metered. If homogeneity measures are taken, which are positive for both capacity and safety outside bottlenecks and need a significant period of time and length to sort its effect, ramp metering that incorporates the new homogeneous traffic situation could be a measure to compensate the loss of capacity.

In chapter 9 an enhanced ramp metering algorithm will be presented. The algorithm consists of an existing algorithm enhanced with a module containing a traffic stability indicator. It is tested, and compared to the Alinea ramp metering algorithm.

6.4 Summary

In this chapter an overview of how external influences influence the driver and the traffic system were presented. Drivers interact with other drivers in a complex system, with relations which are usually not linear and often not independent of each other. Next to that they are influenced by:

Vehicle All drivers are limited in their reactions by the abilities of their vehicles. In case of an emergency break the maximum deceleration of a vehicle is very important.

Road The type of the road sets the boundaries for the behaviour of vehicles. Extra lanes, added dynamic or not, give drivers the opportunity to change lanes in case an emergency stop cannot be made. The roughness of the road, influenced by the weather, influences the maximum deceleration, and through that the traffic stability.

Weather Next to through the road, the weather can lead to an immediate increase of the reaction time, because of changed visual conditions.

Traffic rules The traffic rules (speed limit, no overtaking sign, etc.) influence the driver by making drivers behave in a similar way. Especially the dynamic traffic management measures as signalling and ramp metering are used to create a more homogeneous traffic flow. The next chapter shows that homogenized headways have a big influence on string stability.

DTM measures Not only through the traffic rules, but also directly the DTM measures can influence the driver. The system, e.g. matrix information signs, can make people reroute but has no effect on stability on a single link. E.g. dynamic signs, asking drivers to create headways for merging traffic, influence string stability significantly.

Chapter 5 showed that the distribution of headways and gaps in the flow is important for lane changing and deceleration, two important features in a dynamic flow in relation to string stability. In this chapter the influences on drivers were discussed which often have a large influence on the traffic situation.

Microscopic modelling showed that homogenize measures do not always lead to a capacity increase and better merging circumstances. The capacity in a merging section can even decrease because of the homogenization of the traffic flow.

To better understand the phenomenon a stability indicator will be developed (chapter 7), tested (chapter 8) and used to improve ramp metering (chapter 9).

Chapter 7

Indicator for traffic stability

In the previous chapters the components that form a traffic flow and external processes that influence the traffic flow were presented. Models of the traffic flow (chapter 3) gave us ideas how the traffic flow reacts to different situations (e.g. shockwaves). Some traffic characteristics are measured permanently in The Netherlands (chapter 4) and microscopic relationships between other traffic characteristics (chapter 5) than speed, flow, and density may influence traffic stability. Influential aspects (and on what characteristics of the traffic flow they have influence) were discussed in chapter 6.

Several approaches towards a fundamental characteristic in traffic management, capacity, can be taken. A macroscopic approach; capacity appeared not to be a static value, but can be dealt with as if it is a stochastic process. The alternative, a microscopic approach, needs an estimation tool (product-limit estimation method) because of the stochastic behaviour of drivers. The modelling of the different forms of homogeneous traffic in the last chapter shows that microscopic changes in behaviour do not always have the effect on capacity one would expect. To get a better estimate of bottleneck capacity, traffic flow stability is researched and quantified.

Up to this moment the following definition of traffic stability (chapter 1) was used, recall:

Traffic stability is the capacity of a traffic flow to handle disruptions and prevent breakdown. When a flow has more 'capacity to handle disruptions' without breaking down it is called more stable.

This definition and much of the work in the previous chapters deals with traffic stability in a qualitative way. In this chapter traffic stability is quantified by setting up three indicators:

- An indicator based on traffic reliability (section 7.1).
- An indicator based on traffic synchronization (section 7.2).

- An indicator based on string stability (section 7.3).

In the next chapter these stability indicators are tested in a case study.

7.1 Stability indicator based on traffic reliability

7.1.1 Introduction

In 1988 Ferrari presented an analysis of motorway traffic under congested conditions. His aim was to develop a measure of the reliability of the motorway transport system (Ferrari, 1988). He suggests a model of the speed process on a motorway lane under congestion and tests it experimentally. It is shown that the parameters of this model of the speed process together with the traffic density defines in a sufficient matter the behaviour of the complete system under congested conditions.

Ferrari (1988)'s definition of reliability of a traffic stream is: the probability that, within a certain period of time starting from the moment reliability is measured, no 'hazardous' drops in speed occur. The reliability of the traffic flow as a equation is:

$$\phi = \begin{cases} 1 - p_1(Q/1000)^{p_2} \frac{\sigma^2 - p_3}{p_4 - \ln\ k} & \text{if } \sigma^2 > p_3 \text{ and } \ln k < p_4 \\ 0 & \text{if } \ln k > p_4 \\ 1 & \text{if } \sigma^2 < p_3 \text{ and } \ln k < p_4 \end{cases} \tag{7.1}$$

which defines the reliability ϕ of the traffic flow as a function of the flow rate Q, the density k, and the variance σ^2 of the flow rate. The parameters p_1 to p_4 can be estimated using relationships between σ^2 and $\ln k$ and between σ^2 and the speed process. The indicator ϕ denotes the probability that within a certain time span a 'hazardous' drop in speed does not occur.

7.1.2 Setup of the indicator

The theory on reliability was elaborated by Elbers and Van Berkum (2003) into an indicator which provided this probability of a 'hazardous' speed drop (figure 7.1).
The capacity, reliability and safety of a motorway at similar flow rates, does not depend exclusively on its geometric characteristics, but are also largely due to properties of the transport demand and especially to the composition of traffic flow and the distribution of journey lengths over the vehicle population (Ferrari, 1989). So they may vary over the motorway, although the geometric characteristics might be the same. Different driver modes are adapted by drivers with different trip characteristics which results in different lane use and desired speeds. This is confirmed by Kijk in de Vegte (2002). It is possible to increase the reliability of the network by imposing a speed limit at the moment that density almost reaches it's limit value (Ferrari, 1989). In that case drivers

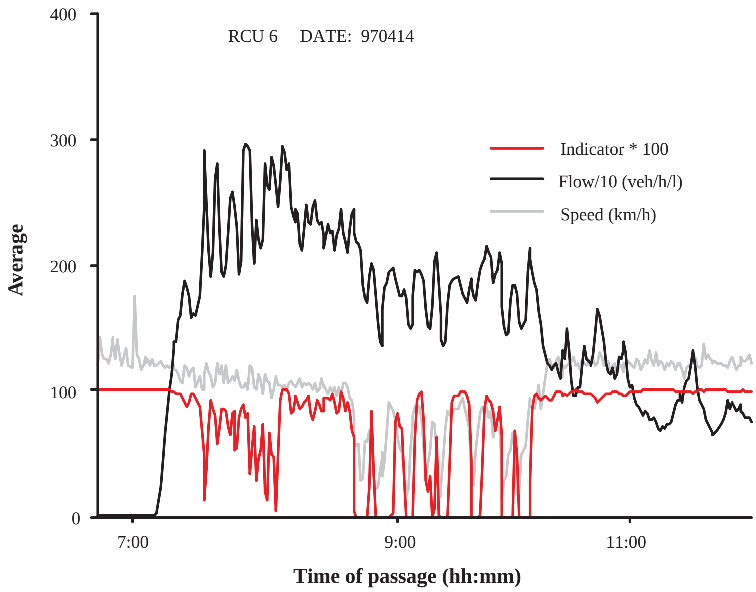

Figure 7.1: Speed, flow and stability indicator based on reliability of flows (0-100%) for the left lane of the three lane motorway A2 in The Netherlands. The performance of the indicator is measured and compared to other stability indicators in chapter 8.

have much less reason to change lanes. This conclusion is experimentally confirmed by Kleinhout (1984) and Papageorgiou (1983). In Ferrari (1994) it is shown that in determining instability of a vehicle platoon the speed fluctuations (amplitude spectrum) of the leading vehicle are of great importance.
In chapter 8 the stability indicator based on the definition of traffic reliability is operationalized and it's ability to predict the onset of traffic jams is tested.

7.2 Stability indicator based on traffic synchronization

7.2.1 Introduction

Experimental investigations of a complexity in traffic shows that flows can be linked to space-time transitions between three qualitative different kinds of traffic: 'free' traffic flow, 'synchronized' traffic flow, and traffic jams (Kerner and Rehborn, 1996). In synchronized flow the density of vehicles is noticeably higher than the corresponding values in free traffic at the same flow level. It may be assumed that in synchronized traffic vehicles are almost not able to pass each other. Kerner and Rehborn (1996) distinguished three different states of synchronized traffic flow:

1. Nearly both stationary and homogeneous states, when both the average

speed and the flow are stationary during a relative long time interval (2-5 minutes).

2. States, where the average speed is nearly stationary during a long time span, but the flow (and density) noticeable changes over this interval. Waves of flow (and density) propagate with positive velocities in this flow. Different waves of density travel in different lanes.

3. Clearly non-stationary and non-homogeneous states; both the average speed and the flow change abruptly. Waves with negative and positive velocities may propagate in such synchronized traffic flow.

7.2.2 Setup of the indicator

This theory was adapted by Elbers and Van Berkum (2003) to create a stability indicator, that indicates how far the state of the traffic flow is from the jam stadium.
A combination of speed fluctuation and flow fluctuation leads to an overall value of the indicator.

In figure 7.2 each combination of speed fluctuation and flow fluctuation can be linked to a value of the indicator. A value of 1 indicates stable traffic.

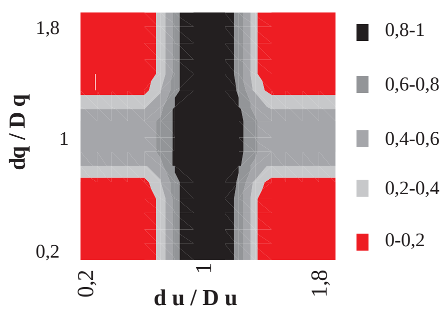

Figure 7.2: Stability lookup table for combinations of speed and flow fluctuations in synchronized traffic. d and D stand for averages over 60 or 300 seconds. Q and u are flow and speed.

An example of the indicator compared to the speed and flow of a motorway is shown in figure 7.3). The three stages of synchronized traffic are used to measure how far the traffic is from actually resulting in jammed traffic.
The three types of traffic, described in Kerner and Rehborn (1996) are often used for qualifying traffic flows, from free to congested. In chapter 8 this indicator is operationalized and tested on it's ability to predict the onset of traffic jams.

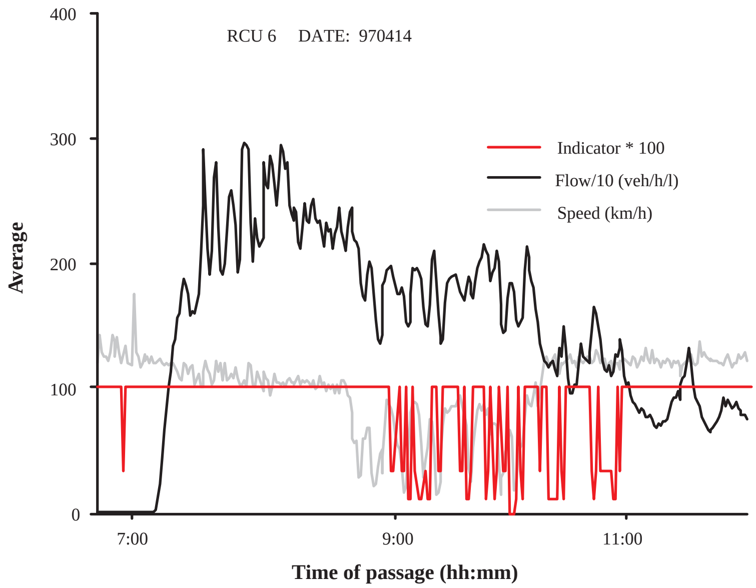

Figure 7.3: Speed, flow and stability indicator based on synchronized flows (0-100%) for the left lane of the three lane motorway A2 in The Netherlands. The performance of the indicator is measured and compared to other stability indicators in chapter 8.

7.3 Stability indicator based on string stability

7.3.1 Introduction

Consider a driver in a heavy traffic flow that is following another vehicle. When the driver is slow to respond (high reaction time) and when the driver responds, over-responds by exerting significantly large acceleration or deceleration rates; this can be characterized as 'unstable' behaviour. So characterized by a high reaction time and a high (or more sensitive) response. The opposite behaviour, a driver who is paying attention, who does not resort to sudden accelerations or decelerations, except in extreme emergencies, which has a low reaction time and a low sensitivity response; this can be called 'stable' behaviour. Thus if the product of reaction time and sensitivity is high, unstable traffic conditions are likely to occur, while if the product is low, stable traffic conditions are more likely to occur. Besides the type of reaction the number of vehicles considered can give information about the stability. A single driver might react in a somewhat wrong way, but a collision might not be the case, while some vehicles upstream of this driver a collision might be inescapable as a result of this behaviour. According to May (1990) there are two issues that play a role in traffic stability: upper limits on car-following parameters, reaction time and sensitivity, and the number of vehicles in car-following mode.

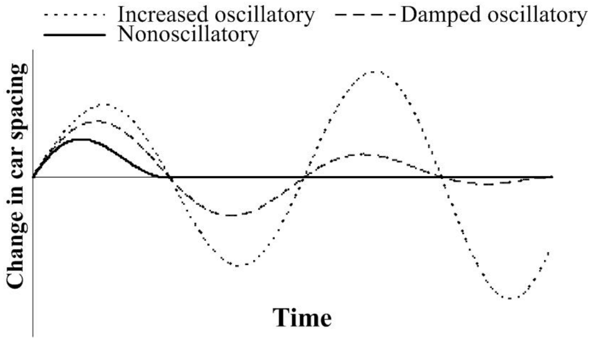

Figure 7.4: Different levels of local stability; increased oscillatory, damped oscillatory or nonoscillatory. After Herman et al. (1959).

This introduction brings us to the definition of three types of traffic stability:

Local stability is (concerned with) the car-following behaviour of two vehicles, one leader and one follower.

String (or asymptotic) stability is concerned with the car-following behaviour of a string of vehicles, one leader and a line of followers.

Traffic flow stability is not concerned with car-following, but with disruptions in macroscopic characteristics of the traffic flow.

Local stability of the General Motors car-following model, equation (3.4), presented in Herman et al. (1959), can be measured by measuring both the sensitivity α and the reaction time t_r and multiplying these two. The product αt_r can be divided into three classes: non-oscillatory, which is more stable than damped oscillatory, which is more stable as increased oscillatory (see figure 7.4).

Herman et al. (1959) also presented norms for asymptotic or string stability. It is divided into damped oscillatory (stable) and increased oscillatory (unstable).

Traffic stability (non-local) can be split up into string stability (or car-following stability) and traffic flow stability ((Swaroop and Rajagopal, 1999), (Yi and Horowitz, 2002), (Yi et al., 2003), (Liang and Peng, 2003), and (Castillo et al., 1993)). String stability is stability with respect to inter-vehicular spacing. Traffic flow stability deals with the evolution of traffic velocity and density (in response to fluctuations of the traffic flow due to vehicles leaving or joining the traffic flow; macroscopic approach). Swaroop and Rajagopal (1999) showed that the spacing policies used in vehicle following control laws (in automatic cruise control) alter the macroscopic behaviour of automated traffic and therefore affect the traffic flow stability. Most macroscopic models of the traffic flow assume the traffic volume, q, to be equal to the product of the traffic density k, and the traffic velocity u; recall that Greenshields (1934) proposed the following relation:

$$u = u_f(1 - \frac{k}{k_m}) \Rightarrow q = u_f k(1 - \frac{k}{k_m}) \tag{7.2}$$

where u_f is the free speed and k_m is the jam, or maximum, density.

Let $\hat{d}_j$, $\dot{\hat{d}}_j$ denote the gap and the rate of change of the gap of the jth vehicle. Let $\hat{d}_j^{\mathbf{d}}$ denote the desired headway of vehicle j and $\epsilon_j = \hat{d}_j - \hat{d}_j^{\mathbf{d}}$ be the difference between the desired headway and the actual headway of vehicle j.

Definition of string stability (Swaroop and Rajagopal (1999)): *A string of vehicles is stable if given $\epsilon > 0$, there exists a $\delta > 0$ such that*

$$\sup_j \max\{|\epsilon_j(0)|, |\dot{\hat{d}}_j(0)|\} < \delta \Rightarrow \sup_j \sup_{t \geq 0} \max\{|\epsilon_j(t)|, |\dot{\hat{d}}_j(t)|\} < \epsilon. \tag{7.3}$$

So when both the difference between the actual and the desired speed and the fluctuations in the size of the gap are not growing for all vehicles the string is called stable.

A definition of traffic flow stability, which can be used is.

Definition of traffic flow stability (Swaroop and Rajagopal (1999)): *Let $u_0(x,t), k_0(x,t)$ denote the initial state of traffic. Let $u_p(x,t), k_p(x,t)$ be the velocity and density perturbations to the traffic, consistent with the boundary conditions and are such that $u_p(x,0) \equiv 0, k_p(x,0) \equiv 0, \forall x \geq x_u$. The traffic flow is stable, if*

1. *given $\epsilon > 0$, there exists a $\delta > 0$ such that*

$$\sup_{x \leq x_u} \{|u_p(x,0)|, |k_p(x,0)|\} < \delta \Rightarrow \sup_{t \geq 0} \sup_{x \leq x_u} \{|u_p(x,t)|, |k_p(x,t)|\} < \epsilon \tag{7.4}$$

   and

2. $\lim_{t \to \infty} \sup_{x \leq x_u} \{|u_p(x,t)|, |k_p(x,t)|\} = 0$

So a perturbation, starting at arbitrary time will eventually damp out.

Both definitions (string stability and traffic flow stability) are used to investigate the relation between intelligent cruise control systems and traffic flow stability in Liang and Peng (2003), Minderhout (1999), and Yi and Horowitz (2002) form the basis on which conclusions about the impact of the systems on the traffic flow are drawn.

As an alternative definition of traffic flow stability a wavefront expansion technique to investigate a nonlinear traffic flow stability criterion can be used (Yi and Horowitz, 2002). Let $k(x,t)$ denote the motorway density, $q(x,t)$ the flow rate and $u(x,t)$ the traffic velocity at position x and time t. By definition, $q(x,t) = k(x,t)u(x,t)$. A model is defined by the conservation law:

$$\frac{\partial k}{\partial t} + \frac{\partial (ku)}{\partial x} = 0 \tag{7.5}$$

An alternative definition of the traffic flow stability, using the conservation equation, is:

Alternative definition of traffic flow (propagation) stability *Let $q_e(x,t) = |k_e(x,t), u(x,t)|^T$ denote the nominal equilibrium state of the traffic system on a highway with length L. Let $q_p(x,t)$ be the perturbed state. The traffic flow $q_e(x,t)$ is **propagation stable** under perturbation traffic state q_p if the spatial gradient of the perturbed state is bounded, i.e. $\parallel \frac{\partial q_p}{\partial x}(x,t) \parallel < \infty$, for $\forall t > 0$, $x\epsilon[0,L]$. If, in addition to the above, $\lim_{t\to\infty} \parallel \frac{\partial q_p}{\partial x}(x,t) \parallel = 0$, then, the traffic state q_e is **asymptotically propagation stable**.*

Yi and Horowitz (2002) further investigated this asymptotic stability and come up with stability conditions for a dynamic system of the form:

$$\frac{\partial u_1}{\partial t} + \alpha u_1 + \beta u_1^2 \tag{7.6}$$

in which all variables are related to the traffic dynamics and the wavefront extension.

In this section different approaches towards stability for a following vehicle (Herman et al., 1959), for a string of vehicles (Herman et al., 1959), (Swaroop and Rajagopal, 1999), and for the traffic (macroscopic description) (Swaroop and Rajagopal, 1999), and (Yi and Horowitz, 2002) were followed.

In the next section an indicator based upon string stability is setup.

7.3.2 Setup of the indicator

In this section a stability indicator based on string stability using the GM car-following model is set up.
The definitions of stability of Swaroop and Rajagopal (1999) and Yi and Horowitz (2002) both focused on propagation of disruption in the flow, speed or density. The problem of these definitions is that assumptions about the initial perturbation need to be made next to the research into the propagation of the

disruption. Therefore it is not known whether the perturbation will lead to an accident or breakdown or not. With the following definition one can formulate disruptions and simulate whether the measured flow would have been able to undergo these disruptions without resulting in an accident.
Recall the definition of traffic stability in the beginning of this chapter.

Traffic stability is the capacity of a traffic flow to handle disruptions and prevent breakdown. When a flow has more 'capacity to handle disruptions' without breaking this flow is called more stable.

Now this definition is worked out using string stability:

Traffic stability A traffic flow is said to be stable if confronted with a (number of) predefined disruption(s) the string stability (reciprocal of accident probability) is less than a certain threshold.

In this definition traffic stability is a combination (multiplication) of disruption probability and propagation probability. If one of the two is zero the stability of the flow will be 1 (100%). Propagation of disturbances finds its cause in driver reactions to traffic situations ahead.

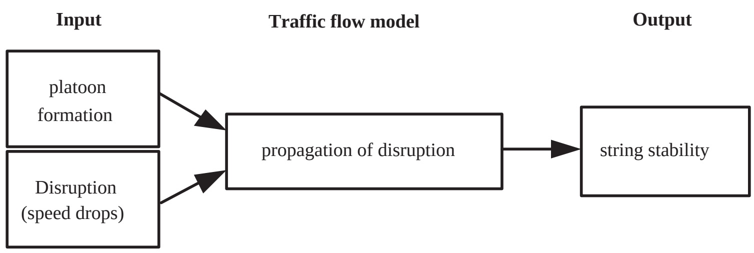

Figure 7.5: Setup of string stability indicator.

Before the traffic flow model which was used to study the propagation of disruptions is presented, the topics which are covered in a microscopic car-following module, of which the research in to reaction time is quite extend, are presented. The reaction time of drivers is often split up into components; the following can be distinguish:

Mental processing time is the time it takes to perceive a signal and to decide upon a response. Mental processing time can be divided into three stages itself.

- Sensation is the time it takes to detect the sensory input. ('There is a contrast change')
- Perception is the time needed to recognize the meaning of the signal. ('The extra contrast is coming from brake lights of the car in front')

- Response selection and programming: is the time to decide upon a response (if needed) and mentally program the movement.

Movement time is the time to perform the muscle movement.

Vehicle response time is the time the vehicle needs to respond to the action of the driver.

The reaction times (mental processing and movement time) are greatly affected by whether the driver is alert to the need to brake in the next moment. Less driver alertness will increase the reaction time significantly. Alertness of the driver can be divided into three classes.

- Expected: the driver is aware of it's surrounding and recognizes the need to apply the brake in the next moment. In this case the total reaction time is 0.7 seconds (0.5 seconds perception time and 0.2 seconds movement time). The other components of the total reaction time are small compared to these two.
- Unexpected: a significant increase in perception time leads to total reaction time of 1.25 seconds (1.05 seconds perception time and 0.2 seconds movement time).
- Surprise: the driver needs extra time to interpret the event and to decide upon the response. The total reaction time becomes 1.5 seconds (1.2 seconds perception time and 0.3 seconds movement time).

Since the implementation of the stability indicator is in traffic which is approaching breakdown drivers are expected to be alert and expect a need to apply the brakes (Green, 2000). Therefore the car following model used to model the string of traffic a reaction time of 0.7 seconds is used.

See figure 7.5. The traffic stability indicator consists of three components: the disruption(s) and the platoon formation as input, the traffic simulation model, which simulates the impact and propagation of the disruptions and the outcome which is a probability on a collision (string stability). The three parts of the indicator are discussed below.

Input, disruptions

Four speed disruptions of the leading vehicle and the effects on it's followers were simulated with the GM-car-following model:

1. *Slow speed drop.* The leading vehicle decelerates from 30 to 20 m/s with -1 m/s^2.
2. *Normal speed drop.* The leading vehicle decelerates from 30 to 20 m/s with -2 m/s^2.

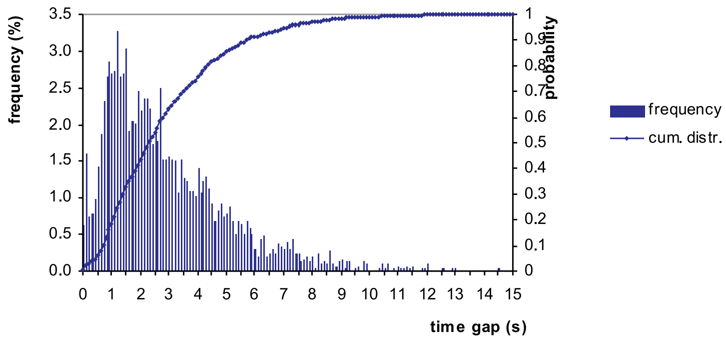

Figure 7.6: Distribution function of gaps from the A2 motorway during peak hour, right lane.

3. *Fast speed drop*: The leading vehicle decelerates from 30 to 20 m/s with -4 m/s^2.

4. *Normal speed drop and recover*. The leading vehicle decelerates from 30 to 20 m/s with -2 m/s^2 and accelerates back to 30 m/s with 3 m/s^2.

Input, platoon formations

A group of ten vehicles driving with 30 m/s were modelled in GM's car-following model and the speed of the first vehicle was disrupted in the sense that is mentioned above. This speed level is chosen, because prior to breakdown the speed of the vehicles is usually synchronized at this level (Kerner and Rehborn, 1996).

Before moving on to the exact form of the simulation model, the results of the simulation of a list of distributions of a platoon of ten vehicles, of which the leader is confronted with disruption number 4, are shown. The platoons headways are created using a measured distribution from the A2 motorway. See figure 7.6.

In the outcome of the indicator the input is described using a mean and a standard deviation (see figure 7.9).

Modelling

The traffic simulation model which is used is the General-Motors car-following model. For an elaborate overview of the development of the GM car-following theories see May (1990). The GM team developed five generations of car-following models, all of the form:

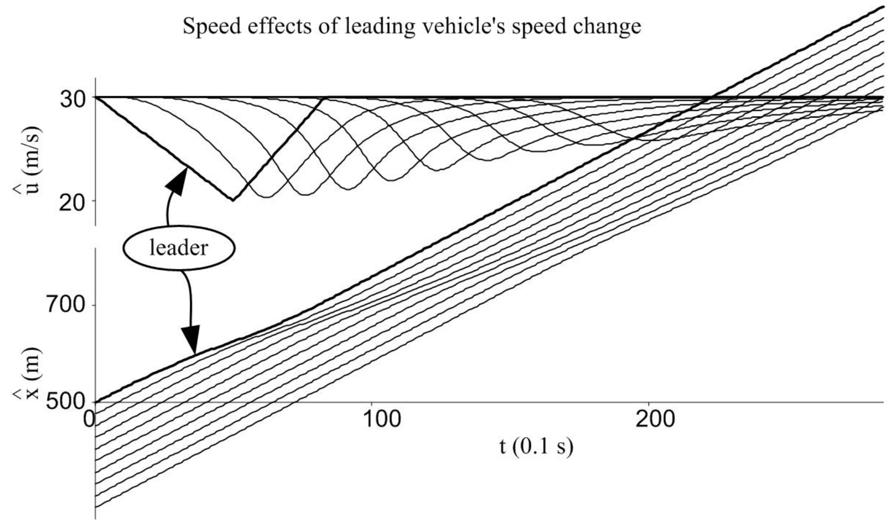

Figure 7.7: Example of propagation of a speed disruption in a platoon calculated using the GM car-following model, visualized in an [x,t] and [u,t] diagram.

$$response = \text{function}(sensitivity, stimuli) \tag{7.7}$$

As said before; the car-following model used for the development uses the function:

$$\hat{a}_{i+1}(t+\Delta t) = \frac{\alpha_{l,m}[\hat{u}_{i+1}(t+\Delta t)]^m[\hat{u}_i(t) - \hat{u}_{i+1}(t)]}{[\hat{x}_i(t) - \hat{x}_{i+1}(t)]^l} \tag{7.8}$$

in which $\hat{a}_i(t)$ is the acceleration of the ith vehicle at time t, $\hat{u}_i(t)$ is the speed of the ith vehicle at time t and $\hat{x}_i(t)$ equals the location of the ith vehicle at time t. A maximum acceleration of 3 m/s^2 and a maximum deceleration of -8 m/s^2 were applied.

As an example figure 7.7 shows a speed-location plot of a platoon of ten vehicles, of which the leading vehicle is confronted with a speed disruption. The leading vehicle decelerates from 30 m/s to 20 m/s with -2 m/s^2 and later accelerates back to 30 m/s with 3 m/s^2. This can be seen as a reaction to a vehicle merging in front of the leading vehicle, because of which it lets go of the accelerator to take some more distance. Later, when the distance is large enough the vehicle accelerates back to it's original speed. The effect of the leading vehicle's speed disruption is propagated to the vehicles upstream of the leading vehicle. In this example the inter-vehicular spacing (headway) is 0.8 seconds at the moment the leading vehicle starts with it's speed change. A headway of 0.8 seconds appears to be enough to eventually damp out the disruption (0.8 is greater

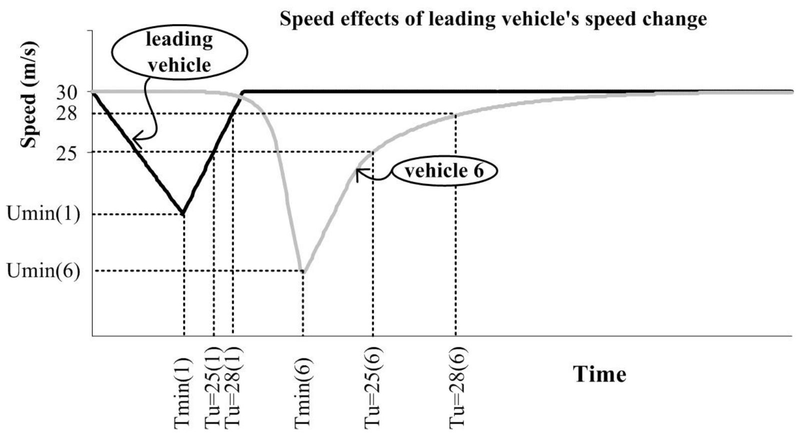

Figure 7.8: Measurements of speed effect of following vehicle 6.

than 0.7, the reaction time). The tenth vehicle's speed drop is much less than the first one.

Comparing speed-time trajectories

Gaps, traffic stability, car following and disruption types have been discussed. In this section the propagation of a disruption, given a certain flow will be quantified.
To measure the effect of the leading vehicle's speed drop a measure for the severity of the speed drop of the following vehicles was developed.

Two speed/time trajectories of vehicles are compared at the moment on which the vehicles drive the same speed, at the minimum speed and at speeds of 25 and 28 m/s with different parameters (figure 7.8). The difference between speed-time curves of vehicles i and i is $\Delta(\hat{u}_i, \hat{v}_j)$.

$$\begin{aligned}\Delta(\hat{u}_i, \hat{v}_j) = \quad & [\min_t(\hat{u}_i) - \min_t(\hat{u}_j)] \cdot [t_{\hat{u}_i=\min_t(\hat{u}_i)} - t_{\hat{u}_j=\min_t(\hat{u}_j)}] \\ & +p_2 \cdot [t_{\hat{u}_i=25} - t_{\hat{u}_j=25}] \\ & +p_3 \cdot [t_{\hat{u}_i=28} - t_{\hat{u}_j=28}]\end{aligned} \tag{7.9}$$

in which $\min_t(\hat{u}_i)$ is the minimum speed of vehicle i, $t(\hat{u}_i = \min_t(\hat{u}_i)$ is the time at which vehicle i reaches it's minimum speed, $t_{\hat{u}_i=25}$ is the time at which vehicle i reaches 25 m/s, and $t_{\hat{u}_i=28}$ is the time at which vehicle i reaches 28 m/s. Notice that parameters p_2 and p_3 are in m/s and $\Delta(\hat{u}_i, \hat{v}_j)$ is in m.

7.3.3 Lookup graphs

With disruption number 4, the minimum speed of the vehicles kept the same when all followers are evenly spaced at a headway of 0.74 seconds (the first

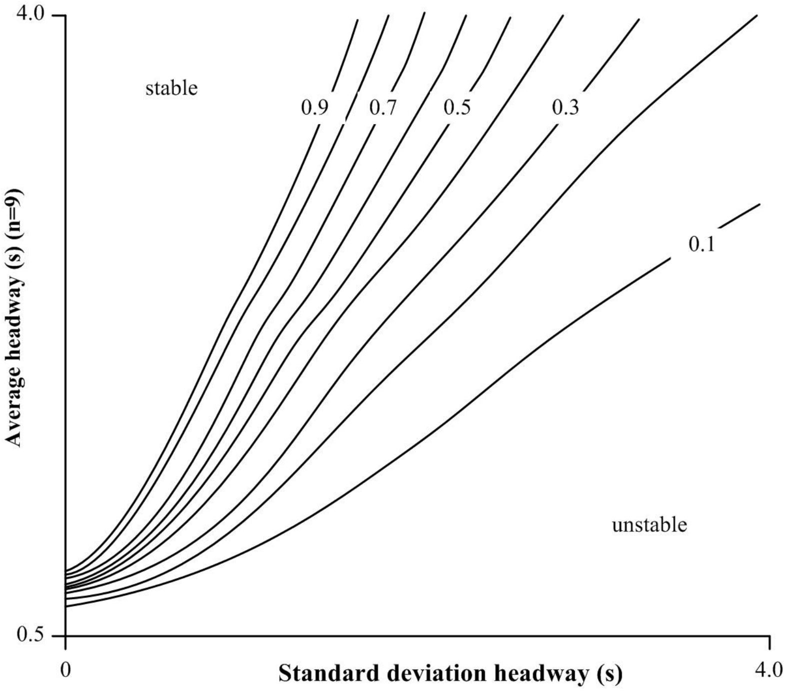

Figure 7.9: The string stability indicator for a platoon of vehicles with an initial speed of 30 m/s and disruption number 4 'normal speed drop and recover'. For the selection of headways drawn the average headway, the standard deviation of the headways and the stability were calculated. The graph shows iso-stability lines at different levels.

part of $\Delta(\tilde{v}_x, \tilde{v}_y)$ is equal to 0). So when a platoon of 10 vehicles is driving at 30 m/s at a headway of 0.74 seconds for all vehicles the flow is neither damped nor increased oscillatory.

With GM's car-following model as presented in the previous paragraph it is easily checked whether or not a platoon of a certain configuration results in a collision when it is confronted with a certain disruption of the leading vehicle. A collision probability is what can be modelled. For each platoon of vehicles measured one can calculate the probability of a collision provided that the sort of disruption is known. But not always all individual headways and speeds are measured and saved. It is much more common to have an average headway and a standard deviation (chapter 4).

A platoon of clusters driving at 30 m/s with different average headway and standard deviation was simulated and the accident rate was calculated. The headways of the followers are randomly taken from a distribution, which has the average and standard deviation as was put in. Because a random selection of nine headways can result in distributions which differ significantly from the original this is repeated (170 times) and averaged, which is the reason why in figure 7.9 the lines are not straight lines.

Figure 7.10 shows the original material of the collision probability for combinations from 0.5 to 4 seconds headway and standard deviations from 0 to 4 seconds for all perturbations (stability equals 1 minus the collision probability). Although technically the data could be extended to average headways larger than 4 seconds this is not done, because of the validity of the results: a car-following model cannot be used when headways reach a certain level, because vehicles will no longer be in car-following mode (Wiedemann, 1974). The speed disruptions as mentioned above in combination with the platoon with a certain average speed and their standard deviation were modelled using the GM-car-following model. The stability can be found in figure 7.10.

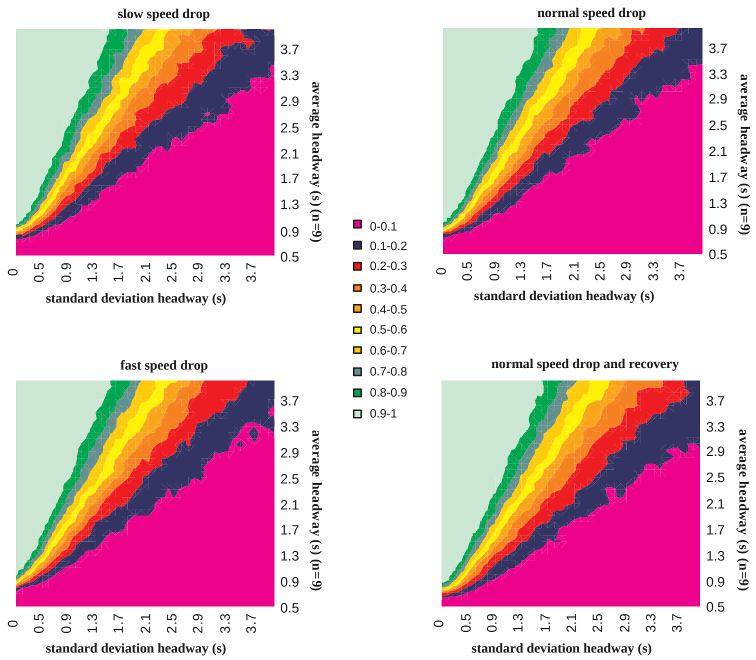

Figure 7.10: Effect of disruptions on the stability indicator, based on string stability calculated with the General-Motors car-following model

Selection of disruption

All four disruptions show a similar result (figure 7.10). The platoon formation appears to be dominant over the type of disruption. For this reason and because of calculation time the stability indicator, which is tested in chapter 8 will use the ‘normal speed drop and recovery’ only.

Stability indicator for multiple lane traffic

The other two indicators are developed to give a single indication of the stability over all lanes. Up to now traffic stability using string stability is calculated for a single lane only. To be able to compare the three indicators in the next chapter, the indicators of the different lanes must be combined. To do this the probability of being able to change lanes is combined with the string stability on all lanes.

On a multi lane motorway a vehicle has another option besides reacting to a speed reduction of it's predecessor by changing it's own speed. The vehicle can change lanes. To incorporate the possibility to avoid a head-tail collision in the existing single lane indicator a single car can change lanes only if this is possible; if there is enough space on the adjacent lane. The space which is needed depends on the speed difference between the cars ((Ahmed et al., 1996), (Kita, 1999), (Minderhout, 1999), (Dijker et al., 1997)). The availability of a gap can be calculated from the distribution of the headways in the target lane and the speed difference between the lanes. If a vehicle A merges from lane x to lane y the speed difference between the two lanes can be reduced by the acceleration or deceleration of the lane changing vehicle and the deceleration of the vehicle in the target lane in front of which the lane changing vehicle will merge. The maximum acceleration is set at 2 m/s^2 and the maximum deceleration is set at -4 m/s^2. Maximum accelerations and decelerations are used, because of the mandatory lane changes (to prevent an accident). For each 6 m/s^2 speed difference between lane x and y the gap that is needed for merging has to be 1 second larger. The gap needed for a mandatory lane change thus depends on six variables: the speed, average headway, and variation of the average headway in both lanes.

The probability of a complete breakdown (over all lanes) is the breakdown probability of the platoon on the lane itself multiplied by the chance that no lane changing is possible. The probabilities of lane change to adjacent lanes are independent, so the probability of breakdown of a middle lane is the breakdown probability of the platoon on the middle lane itself multiplied by the probability that no lane changing is possible to the left and multiplied that no lane change is possible to the right.

An example of the stability indicator on all lanes and the complete motorway can be found in figure 7.11.

7.4 Summary

In this chapter three indicators on traffic stability were developed. The first is based on Ferrari (1988)'s definition of traffic reliability. The second is based on small fluctuations in speed and flow within synchronized traffic. The third

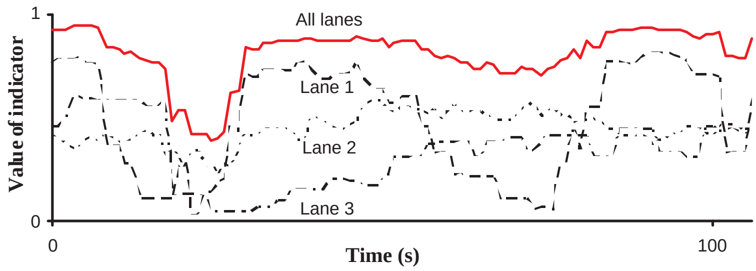

Figure 7.11: Stability indicator per lane of a three lane motorway and stability indicator all for all lanes together in time. An indicator close to 1 means little space and time to manoeuvre in case of a speed disturbance (unstable).

is based on string stability and uses the GM car-following model to calculate it(section 7.3).

In the next chapter the stability indicators are operationalized and compared to the actual onset of traffic jams on a three lane motorway.

Chapter 8

Test of the stability indicators: case study

In this chapter the three stability indicators, which were developed in chapter 7, are tested on their ability to predict the onset of traffic jams. Individual vehicle data from a motorway was collected and used to measure these abilities. The indicators are operationalized to give a traffic jam probability, which is compared to the actual onset of traffic jams.

8.1 Operationalization of indicators

In this section the development from theory to an operational indicator is described for the three indicators. How the indicators were applied to the traffic measurements is also described.

8.1.1 Stability indicator based on reliability

Recall the equation which describes Ferrari (1989)'s 'reliability' of a traffic flow.

$$\phi = \begin{cases} 1 - p_1(q/1000)^{p_2} \frac{\sigma^2 - p_3}{p_4 - \ln\ k} & \text{if } \sigma^2 > p_3 \text{ and } \ln k < p_4 \\ 0 & \text{if } \ln k > p_4 \\ 1 & \text{if } \sigma^2 < p_3 \text{ and } \ln k < p_4 \end{cases} \tag{8.1}$$

which provides the reliability ϕ of the traffic flow as a function of the flow rate q, the density k, and the variance σ^2 of the flow rate. The indicator ϕ denotes the probability that within a certain time span a 'hazardous' drop in speed does not occur.

To apply this equation to the individual vehicle data density, flow rate and the variance, σ^2 of the flow rate were calculated. Flow rate and density were based on the last 100 vehicles that passed a cross-section. The data set consists of the passing times of vehicles; every vehicle's passing is treated as an 'event'. The average density, flow and variance of the flow are recalculated for each event (moving average over 100 vehicles). The calculated density, flow and variance are therefore constant in between vehicle passages and do not decrease over time. The reliability indicator is calculated using equation (8.1) and the moving averages for q, σ^2, and k.

8.1.2 Stability indicator based on traffic synchronization

In Kerner and Rehborn (1996) the definitions of phases of synchronized traffic are all based on averages over time. To create an indicator based on this theory the traffic data is averaged over time (a small time-span to locate short fluctuations(60s =ds) and a long one (300s=DS) to locate trends). The difference between the average over a relative long period (5 minutes) and a short period (1 minute) are the basis for the synchronization indicator. A combination of speed fluctuation and flow fluctuation leads to an overall value of the indicator (as was explained in section 7.2).

8.1.3 Stability indicator based on string stability

The indicator developed in chapter 7 resulted in graph 7.5 in which, for a combination of average and standard deviation of the headway, the indicator can be found. For each 'event' (vehicle passage) the platoon data is updated (average and variation) and the indicator is looked up in the table. To get an indicator for the complete lane, the ability to change lanes is incorporated, using the lane changing probabilities. To make the string stability indicator comparable to the other indicators, the indicator is developed for multiple lanes, as described in section 7.3.3.

The three indicators (originating from different theories; the reliability indicator, indicating the probability of no hazardous speed drop breaking down over a period of a number of minutes, the synchronization indicator, that measures how far the traffic is from resulting in a traffic jam, and the string stability indicator, measuring the probability that a certain disruption results in a traffic jam) are tested for the accuracy in which they predict the onset of congestion.

8.2 Data selection

To test the three indicators for their ability to predict the onset of traffic jams data from a Dutch motorway was selected. The selection complied to the following demands:

1. The presence of traffic jams; preferable with a static bottleneck to locate the exact location of the traffic breakdown and get frequent breakdowns

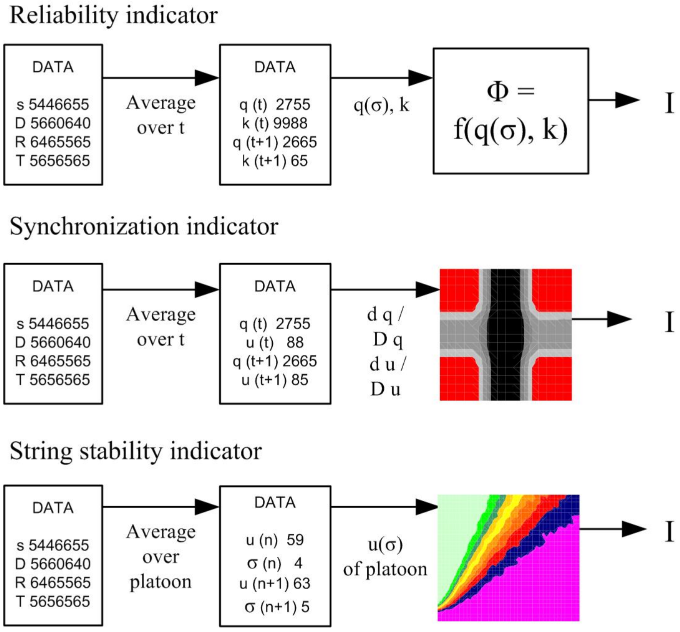

Figure 8.1: Overview of operationalization of the indicators.

at the same location. No demands were set to the traffic situation prior to breakdown; all onsets of traffic jams were selected.

2. Fine meshed data collection system; to get data which is disaggregated enough for all three indicators. Aggregation can always be done, but for disaggregation assumptions need to be made. Because the indicator based on string stability needs individual vehicle data the location must be equipped with measurement tools which are able to collect individual vehicle data.

8.2.1 Study area

A site was found appropriate between Utrecht and Amsterdam, on the A2 motorway. This motorway is 3 lanes wide in each direction and has a heavy on ramp (Vinkeveen) in the direction of Amsterdam (see figure 8.2).

This on ramp creates a static bottleneck from which shockwaves move upstream (see figure 8.4). Data of this section was available in the form of individual vehicle data (RCU data) and one minute aggregates (MARE data) (see figure 8.3).

The speed limit on the road section is 120 km/h, which is the general speed limit on Dutch motorways. The terrain is level and the traffic is not influenced by the small grades due to viaducts. The curvature of the motorway also does

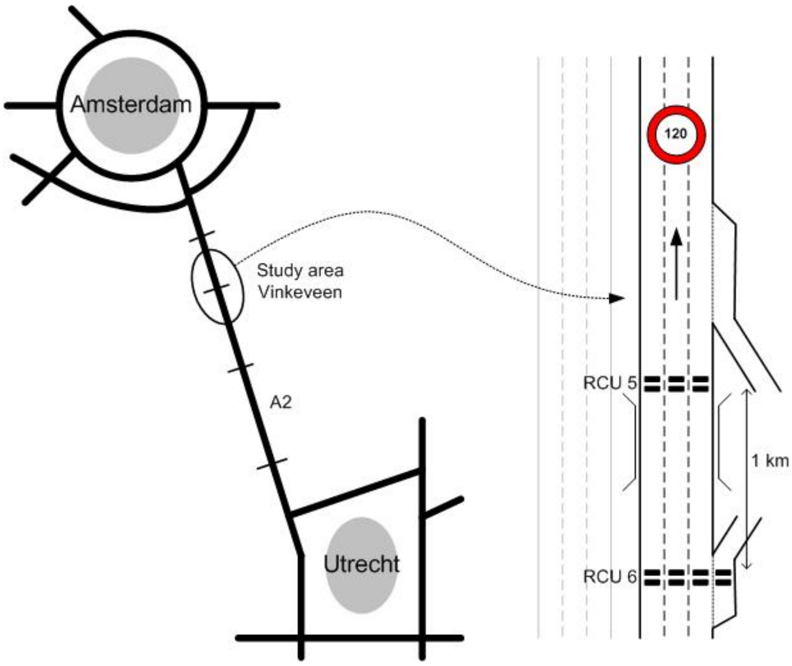

Figure 8.2: Study area between Utrecht and Amsterdam.

not limit traffic behaviour.

In the peak hours the main flow, just upstream of the Vinkeveen on ramp (which is responsible for the traffic jam) can run up to 7850 veh/h over three lanes based on one-minute averages and 6190 veh/h based on 15 minute averages. Into this flow the on ramp traffic (1200 veh/h) has to merge. Although data is available from Utrecht to Amsterdam, the focus is on detector locations 5 and 6. 5 is just prior to the on ramp and 6 is 1 km upstream of location 5.

8.2.2 Observations

The A2 motorway has, just as two thirds of the Dutch motorway system dual loop detectors approximately each 500 meters on all lanes and occasionally on an on or of ramp. On all three lanes of the A2 the vehicle's speed and length are detected. The measurements are aggregated to one minute averages of speed and flow and are saved (MARE-data). This MARE-data was used for global inspection and selection (paragraph 8.2.3) of the more detailed RCU-data. The Research Communication Unit-data is the original measurement of the MARE-data, before it is aggregated and contains individual vehicle measurements. See figure 8.3 for a segment of a MARE and a RCU data file.

MARE data

Road number A2E Flow A & B
11-jun-98 Agr= 1 min Time step = 1 min

Hm pos->	456	462	467	473	479	485	493	497
TIME								
6:31	3600	4200	3800	4450	4300	4000	4550	3550
6:32	4450	3650	3550	3050	3550	3350	4000	3300
6:33	3650	3700	3400	3100	2700	3400	3350	3400
6:34	3900	3400	3600	3050	3000	2600	3900	3250
6:35	3550	3650	3550	3050	3500	3050	2650	3050
6:36	3700	3650	3950	3800	3650	3350	1900	3000
6:37	3500	3950	3700	3700	3700	3500	2200	2300
6:38	3700	3850	3950	3550	3500	3300	2350	1300
6:39	3850	4000	4200	4250	3800	3850	3400	1550
6:40	3700	3500	3300	3800	3350	3700	3250	2400
6:41	3400	4200	4300	3850	3950	3250	3600	2450
6:42	4100	3600	3100	3900	3050	3900	3950	2900

50 vehicles passed between 6:33:01 and 6:34:00 at position 479

RCU data

DATE	TIME	RCU	LANE	SPEED	LENGTH
970409	64900613	11	1	115	513
970409	64944933	11	1	160	318
970409	64950360	11	1	130	360
970409	64956557	11	1	120	256
970409	65158708	11	1	116	337
970409	65203143	11	1	105	279
970409	65204517	11	1	105	370
970409	65250099	11	1	123	295
970409	65506689	11	1	132	327
970409	65507970	11	1	130	386
970409	65513057	11	1	147	341
970409	65555854	11	1	115	391

characteristics of an individual vehicle

Figure 8.3: Examples of MARE and RCU-data.

8.2.3 Selection of days

The data from the A2 motorway was originally collected as part of the evaluation study of a speed checking system. The system measured the average speed of all vehicles between kilometre 54.5 and 51.3 for a period of 2 times 6 months.

Global inspections were made making location-time-speed and location-time-flow diagrams of the MARE data. In these diagrams one can get an overview of global characteristics of the road on time as speed and flow. Based on these graphs days were selected which showed congestion starting at the Vinkeveen on ramp (for an example see figure 8.4).

The week from the 14th to the 18th of April 1997 was chosen to test the predictability of the indicators. This selection is not based on the type of traffic jams, or on the frequency; all days show more or less the same traffic jams. The selection is based on the availability of complete data sets (no missing data).

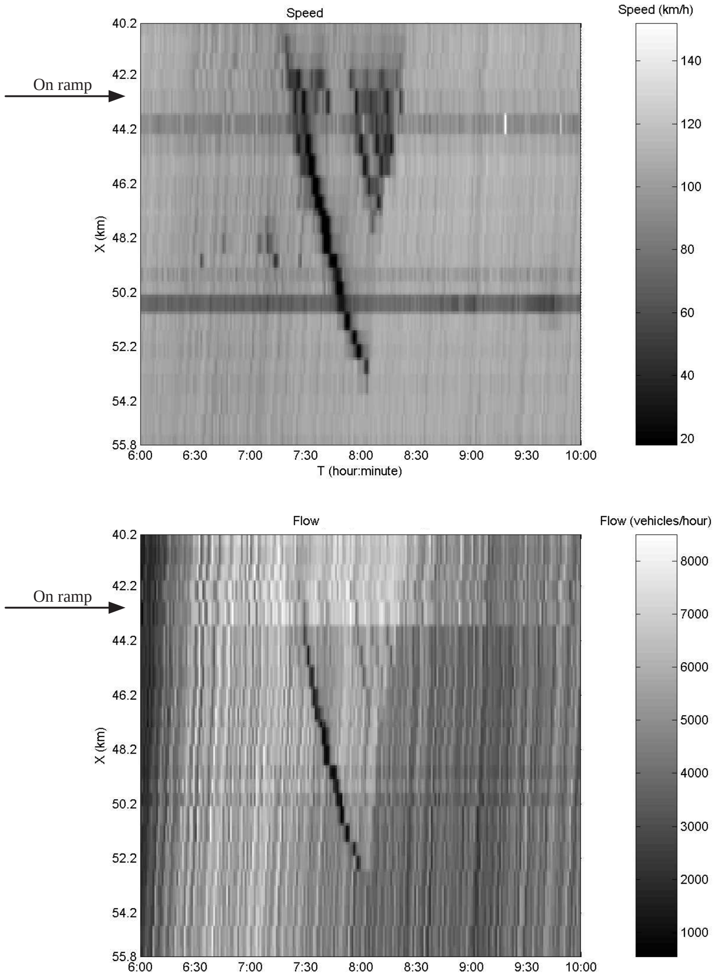

Figure 8.4: Examples of visualization of MARE-data. Traffic jams start near the on ramp and travel upstream to bridge a large part of the section Amsterdam-Utrecht.

8.3 Performance of the indicators

The performance of the indicators is measured by calculating the correlation between the value of the indicators and the moment of breakdown. To quantify the breakdown the fundamental diagram of the motorway is used to check at what speed the flow breaks down (after (Lorenz and Elefteriadou, 2001)).

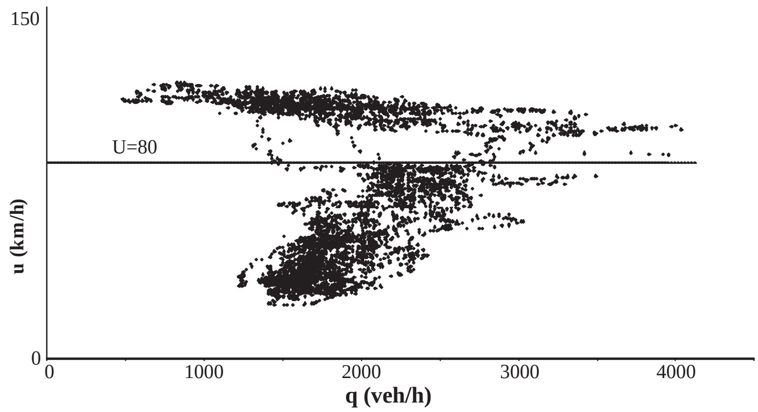

Figure 8.5: A2 motorway speed-flow data from the left lane, which indicates that breakdown has occurred when speeds reach under 80 km/h.

From figure 8.5 it is clear that if the speed of a vehicle is more than 80 km/h breakdown has not jet occurred. Data of lane 1, the fast lane, is used, because the traffic jam starts in that lane (the first speed drop).
A workable definition of breakdown which can be derived from the reasoning above and which will be used to quantify breakdown is:

Breakdown on A2 motorway A location at a certain time is called 'broken down' when a vehicle is measured, driving 80 km/h or less. The label 'broken down' remains for 5 minutes to prevent high frequency switching between breakdown and no breakdown.

The values of the three indicators and the breakdown indicator are correlated in time and on two spatial cross-sections (5 and 6): the first is between the breakdown indicator and the three stability indicators on the same location at the same time. The second is between the breakdown indicator and the stability indicators at one kilometre upstream of the jam location at different time intervals.
This is done at time intervals of 0, 40 and 60 seconds (which corresponds with infinite speed, 25 m/2 and 17 m/s). Speed dependent time intervals will probably result in better performances, but from this all three indicators suffer and the results will be used to compare the indicators to each other and the absolute value will give a general idea of the performance.

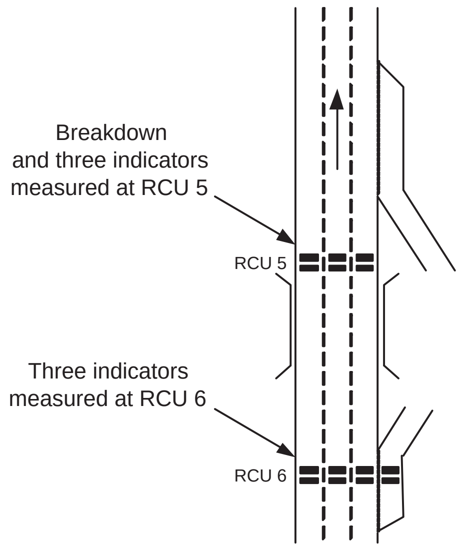

Figure 8.6: At both measurement locations and with different time-lags, all three indicators are calculated.

To treat indicators that over- or under estimate the onset of traffic jams the same, the data set was chosen to have 50% of the time a jammed status. This meant that the research time interval was expanded in front and behind the time-interval. This was never more than 15 minutes, because in the jammed period the percentage of jammed traffic (according to the indicator) is almost 50%.

The results of the correlations between the indicators and the jam indicator are in table 8.3.

Table 8.1: Correlations (n=40.000) between stability indicators (at locations 5 and 6) and congestion (at location 5). BI = Breakdownindicator, RI = Reliability Indicator, SI = Synchronisation Indicator, SSI = String Stability Indicator. The breakdown indicator is 1 in 39% of the time.

Indicator	BI	BI ($\Delta t = 40$)	BI ($\Delta t = 60$)
RI (loc 5)	-0.77		
SI (loc 5)	-0.43		
SSI (loc 5)	-0.78		
RI (loc 6)	-0.52	-0.52	-0.51
SI (loc 6)	-0.51	-0.55	-0.56
SSI (loc 6)	-0.70	-0.70	-0.70

The correlations between the congestion indicator and the stability indicators at location 5 (just upstream of the on ramp) vary between -0.43 for the Synchronization indicator and -0.77 and 0.78 for the Reliability and the String stability indicator. The Synchronization indicator clearly under performs at a single cross-section.

When the stability indicators from a cross section 1 kilometre upstream of the bottleneck (RCU 6) are correlated to the breakdown indicator at RCU 5 it is clear that both the Reliability and the Synchronization indicator under perform.

Conclusions

The outcome of the correlations between the indicator and the actual onset of traffic jams should not be interpret to rigorous. They do not give information of where the indicators failed: were they to early, to late, or a combination of both. Nevertheless the correlations show clearly that the string stability indicator outperforms both others.

All three indicators can probably be fine tuned, to give a better prediction of the onset of traffic jams. The risk of doing this however is that the tuning will make the indicators less general and more (location and jam type) specific.

At this stage it was chosen to implement the string stability indicator in an ramp metering algorithm and test it's performance and leave the fine tuning of the stability indicator for future research.

8.4 Summary

In this chapter three stability indicators (based on reliability of traffic flow (Ferrari, 1988), synchronization of traffic flows (Kerner and Rehborn, 1996), and string stability a.o. (Swaroop and Rajagopal, 1999) using the GM car-following model), were operationalized and tested on their ability to predict the onset of congestion.

Data was collected from a motorway with a static bottleneck and correlations between the indicators and the breakdown indicator were calculated over different time intervals and over a single location and two adjacent locations.

The string stability indicator clearly outperforms the other two indicators. It gives an indication of the onset of traffic jams.

The string stability indicator can give extra information on the breakdown phenomenon (correlation of 0.7) and in the next chapter the Alinea ramp metering algorithm will be combined with the stability indicator in order to create a better ramp metering algorithm.

Chapter 9

On ramp control and traffic flow stability

This chapter shows how the stability indicator, developed in section 7.5, is used to enhance the Alinea ramp metering algorithm. It starts with a short introduction ramp metering and the traffic processes near an on ramp. It is showed that the level of detail at which the traffic process can be studied (microscopic or macroscopic) directs the enhancements that can be made to the ramp metering algorithm. After this an enhanced Alinea algorithm is developed based on the existing Alinea ramp metering algorithm and the stability indicator developed earlier. Then both algorithms are tested in a simulation environment and the altered algorithm shows a significant improvement. This chapter contains material from Ruijgers (2005).

9.1 Introduction to ramp metering

Ramp control can be described as a method of improving overall motorway operations by limiting, regulating and timing the entrance of vehicles from one or more ramps onto the shoulder lane of the main road. At locations where motorway entrance ramps have adequate storage capacity or where the surrounding street network can accommodate additional traffic, ramp control systems can provide substantial operational improvements under certain combinations of traffic demand and motorway capacity (Papageorgiou, 1991).

9.1.1 History of ramp metering

During the past fifty years car-ownership, and thereby the demand for infrastructure, increased tremendously. Governments started expanding the traffic network to facilitate this demand, until resistance against it grew, leading to an increased awareness of the environmental influences. As a result of this resistance, governments gradually changed their policies from expanding infrastructure to optimizing utilization of existing infrastructure. Nowadays traffic policies to increase accessibility in many western countries are based on three types of measures: utilization, price-policies and expansion. Ramp metering is a clear example of a utilization measure.

Ramp metering strategies were first implemented in 1961 in Chicago (USA) to create smoother and safer merging circumstances for vehicles merging from an on ramp on to the motorway, and to minimize disruption of the mainline flow by merging vehicles. These objectives have not changed over the years, however, the way in which it is realized has quite changed. Where the first ramp metering was performed by a policeman who manually controlled the traffic flow on the on ramp by pre-specified metering rates, today ramp metering strategies may be integrated with other measures, and metering rates are often calculated in real-time, using advanced algorithms or artificial neural networks to anticipate at the actual traffic conditions as good as possible. Although several studies proved that ramp metering is able to reduce travel time, it is certainly not the ultimate solution for traffic congestion; it is only a way to improve local (and as a result possibly global) traffic conditions. This is possible to a certain (time and place depending) amount; beyond this level benefits of ramp metering are limited.

Basic control scheme for motorway traffic

The basic control scheme for motorway traffic (figure 9.1), developed in Papageorgiou (1991) is used to show how the ramp metering measure can be applied to control the traffic flow. Later this scheme is used to visualize the Alinea and the enhanced algorithm.

Process is the traffic process at the on ramp, which consists of vehicles merging, giving way, changing lanes, etc.

Nonpredictable disturbances , for instance accidents or the weather, influence the traffic process.

Predictable disturbances are demand or origin-destination rates.

Process measurements are the measurements which are needed to steer the on ramp algorithm (as speed, flow, and occupancy).

Control strategy is the core of the ramp metering algorithm. The predictable data and the process measurements are combined with the goal (e.g. no traffic disruptions on the motorway) and a metering strategy (rate) is calculated.

Input is the actual traffic light situated at the on ramp and it's metering rate. It has a direct influence on the traffic process.

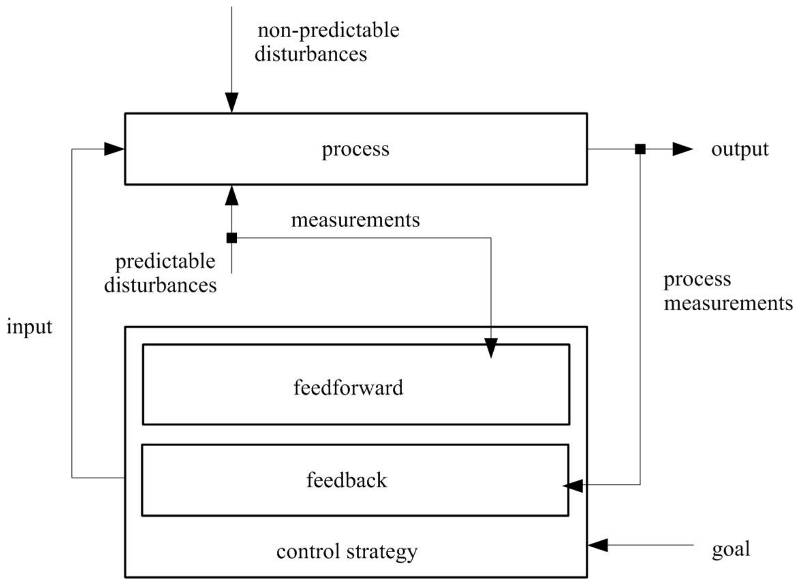

Figure 9.1: Basic control scheme for motorway control tools; after Papageorgiou.

9.1.2 Traffic process

Traffic processes can be modelled at various levels of data aggregation. The most aggregated level is the macroscopic level, where vehicle stream characteristics like flow, density and average speeds are used. The opposite is the microscopic level, where individual vehicle characteristics like headways and car-following behaviour are considered. The process of interest in this thesis is the interaction between on ramp and freeway vehicles due to the mandatory merging actions of on ramp vehicles. As individual vehicles are concerned here, it would be logical to look at the process at a microscopic level and to base the ramp-metering strategy on microscopic relationships. However, almost all currently used ramp metering strategies are based on macroscopic assumptions. Reason for this is that microscopic data is much more sensitive to small disturbances, because individual vehicle characteristics are used. On such a detailed level it is much harder to react adequately to changes of the traffic conditions than it is on a macroscopic level. In this paragraph the macroscopic assumptions and drawbacks for use in ramp metering strategies will be discussed first. After that, possibilities and drawbacks of microscopic modelling will be considered.

Macroscopic considerations

Greenshields (1934) described traffic as a continuous fluid. This fluid can be described by three parameters, which are all continuous functions of space and time: speed, density and flow. The fundamental diagrams (section 2.3) suggest that capacity is one specific value.

But capacity can vary between roads and during time. In the previous chapters was showed that variations between roads are due to differences in geometry like (for example) horizontal and vertical alignment, number of lanes, lane width and the presence of on ramps and off ramps. During time, capacity will also vary at the same location, as a result of differences in e.g. vehicle types, driver types, weather and incidents.

Even if macroscopic circumstances were identical, there appears to be a range of possible flow values where traffic enters the congested regime. Elefteriadou and Lerworawanich (2002) studied the differences between the theoretical definitions and the situation in reality. They concluded the following:

- The maximum sustained flow at a certain freeway section varies and does not necessarily occur in conjunction with breakdown, more general, for equal flows breakdown may or may not occur.
- Flows at the moment of breakdown vary and can be lower than maximum observed flows or capacity flow.
- The probability of breakdown increases (as expected) with increasing flow rate but breakdown flow (for freeways) can vary between 1500 and 2300 vehicles per hour per lane.

Obviously, the transition at the specific flow level does not exist and thus is this value less accurate than it seems, as a whole range of values of maximum flow and critical densities are possible. To obtain a proper capacity definition it should be based on breakdown. Elefteriadou and Lerworawanich (2002) recommend that the definition of breakdown should be based on a speed threshold and an associated minimum duration (which are both site-specific values). In general:

Definition of breakdown (Elefteriadou and Lerworawanich, 2002) The moment at which speed drops during a pre-defined (site-specific) period of time below a pre-defined (site-specific) threshold.

Also recall the definition of reliability of a traffic stream as suggested by Ferrari (1988).

Definition of reliability (Ferrari, 1988) The probability that, within a certain period of time starting from the moment when the reliability is measured, no 'hazardous' drops in speed occur.

Elefteriadou and Lerworawanich (2002) state that a capacity definition should be based on the breakdown flow, which is the average flow over five (or fifteen) minutes prior to the moment of breakdown. Advantages of such a definition are:

- it is a conservative estimate (breakdown flow is lower than maximum pre-breakdown flow and maximum discharge flow);
- it is consistent with the current implication of being the transition from non-congested to congested regime;
- it can be associated to the probability of breakdown.

Most currently used ramp metering strategies are based on macroscopic data, and try to keep demand below capacity, or try to reach the critical density (i.e. the density where capacity occurs in the flow-density diagram). So they use a deterministic value of capacity. Considering the conclusions of the section above in ramp metering perspective, strategies based on traditional definitions of capacity will not be as robust as they seem. First, capacity is a time varying value and is therefore hard to determine. Second, even under identical (macroscopic) circumstances, a single capacity value does not cover the transition between free-flow and congested states, as it should be considered as a probability. By using a deterministic value, in some occasions no breakdown will occur at capacity, while in other occasions breakdown has occurred before the expected capacity is reached. Obviously, the macroscopic level lacks detail to predict the exact moment of breakdown and therefore it might be useful to consider the merging process on a microscopic level.

Microscopic considerations

As merging takes place one vehicle after the next it can be considered as a microscopic phenomenon, and a microscopic ramp metering strategy would therefore be expected. But almost all currently used strategies are based on macroscopic relationships. One example of a microscopic strategy is the release-to-gap strategy, which tries to synchronize the release of vehicles with detected freeway gaps. This method, first developed in the 1960's, is hardly used anymore, as it appeared to be too difficult to predict gap-propagation and other variables with sufficient accuracy. The large varieties within data of successive vehicles and the instability of it are probably the reasons that hardly any ramp metering strategy uses microscopic relationships. Nevertheless the widely used macroscopic data is insufficient to predict the moment of breakdown accurately. Based on microscopic data the probability of breakdown can be calculated and this might provide a useful addition to existing macroscopic strategies.

First it has to be discussed how traffic behaviour is modelled on a microscopic scale. Microscopic density characteristics are based on car-following models, which development started in the early 1950s. Car-following models are based on the assumption that every driver has a desired speed, which he tries to establish continuously. Microscopic traffic flow modelling and the General-Motors car-following model were discussed in section 3.3.1.

Conclusion

The assumption of many ramp metering strategies to release (by metering rate) just as many vehicles as possible to get or keep demand just below capacity is basically good, as it leads to optimum usage of the traffic facilities. However, as capacity is not a deterministic value it can only lead to sub-optimal results. Sometimes breakdown will occur although demand has not reached (assumed) capacity, while in other occasions the distribution of headways leads to very stable circumstances and metering rate can be increased although (assumed) capacity has been reached.

Considering these findings a microscopic addition to existing macroscopic strategies is potentially beneficial. Additionally, as merging is a microscopic process, it would be logical to base ramp metering strategies on microscopic assumptions.

9.1.3 Metering strategies

The assumptions and theories on which on ramp metering strategies are based form the main difference between the various strategies and lead automatically to specific demands on data gathering and process input.

For a good overview of existing ramp metering strategies, it is helpful to divide the strategies into classes. The following classification is used (after Kotsialos and Papageorgiou (2004)):

- fixed time ramp metering strategies;
- reactive ramp metering strategies;
- pro-active (non-linear optimal) ramp metering strategies.

The classes are based on the way data is used to calculate the output of the ramp metering strategies. More specific: whether historical or actual data is used to determine variable and parameter values.

Fixed time strategies

Fixed time strategies (figure 9.2) are static strategies, as they do not depend on actual traffic state. Based on historical data derived from field tests, a schedule is determined, which specifies metering rates over time. Metering rates can vary across the day and between days, but they are not adjusted to actual traffic conditions, neither do they try to reach actual optima. Peculiar circumstances like bad weather (low capacity) or big events (high demand) might lead to metering rates that are far from optimal. Capabilities of fixed-time strategies to smooth the traffic process are limited. Nevertheless they can adequately be used to influence route choice.

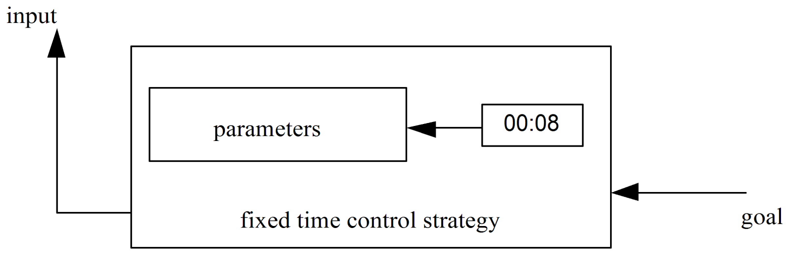

Figure 9.2: Scheme of fixed-time strategy. Control scheme based on basic control scheme, Papageorgiou.

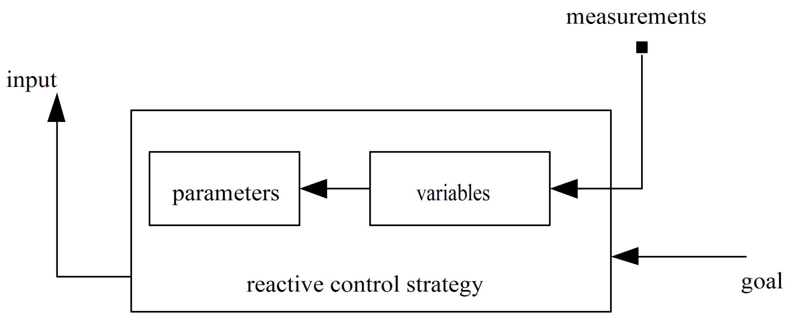

Figure 9.3: Scheme of reactive strategy. Control scheme based on basic control scheme, Papageorgiou.

Reactive strategies

Reactive ramp metering strategies (figure 9.3) are dynamic strategies, because they compare actual traffic conditions to pre-specified parameters like capacity or desired occupancy. As these strategies are able to react on actual traffic conditions they perform better in preventing congestion and reducing travel time than fixed time strategies.
Most important drawback of reactive strategies is that pre-set control objectives (e.g. capacity or critical occupancy) are used, which are assumed to be optimal. However, as different variables (like for instance weather conditions) can influence these quantities, these control objectives might not be optimal at all in specific circumstances.

Pro-active strategies

To overcome the drawback of the reactive strategies (pre-set control objectives) pro-active strategies try to adjust their thresholds to actual conditions by estimating them in real-time. The used information about traffic conditions is much more comprehensive, as incidents, non-linear flow dynamics and queue constraints are considered. Although pro-active ramp metering strategies theoretically have the potential to outperform reactive strategies, they are more

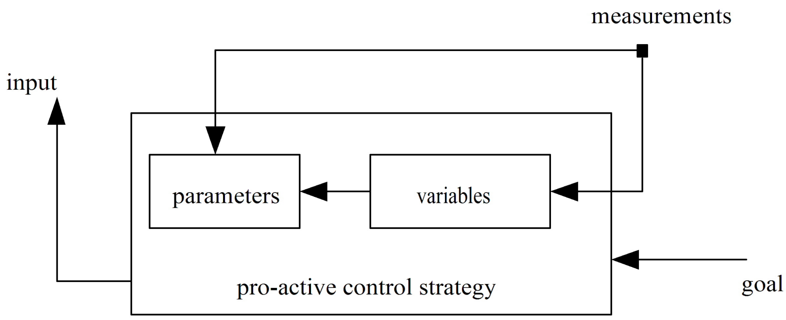

Figure 9.4: Scheme of pro-active strategy. Control scheme based on basic control scheme, Papageorgiou.

difficult to calibrate since they need more data to make reliable predictions, and therefore they have not been implemented yet.

Feed forward and feedback strategies

In the previous sections a classification was made based on the differences in application of parameters and variables. One static and two dynamic strategies were mentioned; fixed-time, reactive and pro-active strategies. Strategies can also be classified using, the variables that are used; the input or the output of the traffic process. This leads to two ways of considering the traffic process, feed forward or feedback (see figure 9.5), both types can be used single or combined. As shown in figure 9.1 the traffic process in the vicinity of an on ramp is determined by disturbances (predictable and non-predictable) and the input, which is in this case the influence of the ramp meter.

Feed forward strategies try to measure the predictable disturbances to determine the influence of it on the desired process state. If a difference between the predicted state and the desired state is measured, the control strategy tries to influence the predicted state, increasing or decreasing the metering rate.
Feedback strategies compare the actual state (i.e. measurements of the process output) with a desired state.

Both strategies have different advantages and drawbacks as shown by the following example. Imagine a vehicle that is released from the on ramp is driving at the acceleration-lane and merges into a relative small gap. The vehicle behind it overreacts and brakes very severe. The following driver drives with a (too) small headway and a collision occurs, leading to total breakdown. Typically, feed forward strategies consider the difference between demand and capacity, and they assume that if flow does not exceed capacity, the traffic flow process in the vicinity is stable and metering rate can be further increased, to reach an optimum throughput. As input (freeway flow) is measured, this type of strategy is very accurate in determining how much traffic can be added. However, due to the collision, the traffic process is far from desirable and very

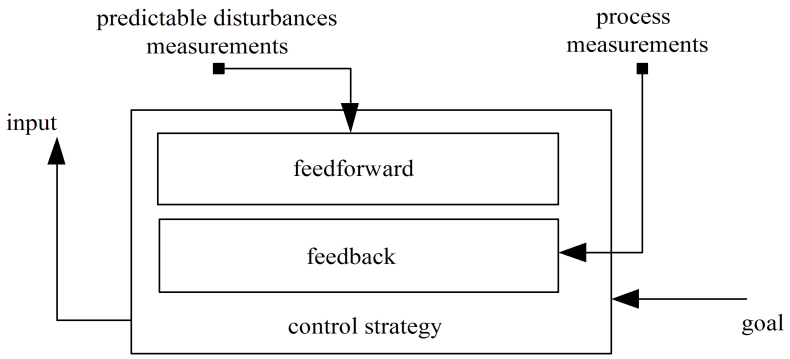

Figure 9.5: Combined scheme of feed forward and feedback strategy. Control scheme based on basic control scheme, Papageorgiou.

different from the assumed. One of the biggest drawbacks of feed forward strategies is that they are based on predictions, which inevitably include uncertainty. As feedback strategies consider the result of the process, nothing has to be predicted. As soon as the collision has occurred, the control strategy will determine a strong reduction of flow on the downstream detectors and concludes that breakdown has occurred. Then it can adjust its metering rates. An important drawback of feedback strategies is that they react on the process output and therefore are always too late to change the process input (e.g. metering rate).

Both feed forward and feedback strategies are used (single or in combination) in current ramp metering strategies. Various research results do not lead to a conclusion that either feed forward or feedback strategies lead to better results. Based on the theoretical consideration mentioned above, a combination of both strategies might lead to best results, as the main drawbacks of both types (difficult prediction and late reaction) can be covered by the other strategies.

Basic principles of ramp metering strategies

Except for the release-to-gap strategy, the output of all metering strategies (i.e. input for the traffic process) is metering rate. This implies that green, amber and red period occur in proportions that are only gradually changed by the control strategy if necessary. The actual traffic state is determined by detectors which measurements are aggregated to intervals of about 30 seconds to one minute. In this way strong fluctuations in the variables due to differences between individual vehicles are minimized. For each time period a new desired metering rate is calculated. This desired metering rate might be chosen as new metering rate, but often it is averaged with the current rate, to avoid a nervous metering rate tendency.

In classical strategies, metering rate is only determined by the red period, which is variable. Green period and amber (if used) period were static, and took about 3 seconds altogether. This short static period is chosen because it enables only one car to leave the on ramp queue. In modern strategies green and amber period are not static anymore, but are triggered by detectors just downstream the stop line. The duration of green and amber does not influence metering rate, since the requested metering rate will be achieved by adjusting the red period, taking into account this duration. Metering rates can vary between 180 vehicles per hour and 750 vehicles per hour (AVV, 2002b) for a single lane on ramp and up to 1450 vehicles per hour for an on ramp of two lanes. Metering rates below 180 vehicles per hour will lead to increased violation of the red signal, as drivers have to wait too long. If metering rate is increased beyond 750 vehicles per hour, drivers will get confused because too short red and green periods have to be used.

In some occasions the capacity that is achieved by the ramp meter is insufficient to accommodate on ramp traffic demand. This will lead to increasing queue lengths with high delay times and possible interference with surface street traffic as a result. In these occasions it might be useful to implement a strategy based on two vehicles per green period. An increase of about 200 vehicles can be achieved by releasing two vehicles per green period (Ross, 1991). In the Netherlands an experiment with two cars per green period was carried out at an on ramp to the A28 near Utrecht. Varying traffic demand biases results of the field-research of this experiment and therefore no conclusions can be drawn about the influence on traffic conditions. The simulation part of this study showed that freeway conditions are hardly affected by the two-vehicles-per-green strategy, while queue length decreased significantly. Although further (field) research is necessary, preliminary results suggest that two-vehicles-per-green can be advantageous in case of high on ramp demands and little buffer space (short on ramp). If two vehicles per green-time is chosen, it is important to use a consequent strategy for a specific on ramp; either one or two vehicles-per-green-period should be chosen, and this choice should be made clear to drivers by traffic signs to avoid confusion. Some strategies only operate during pre-specified periods of the day (typically the peak periods), but most strategies measure traffic data all day. This does not mean that actual ramp metering will be in use all day, since the need for ramp metering is compared to certain thresholds. If the ramp meter is not operating, metering lights are switched off or the green light is switched on continuously. The first option is preferred, to make it clear for drivers that the system is not operating, and no sudden red period has to be expected.

Discussion

In the sections above, two differences between coordinated and local strategies were discussed. Although in theory coordinated strategies should be able to perform better than non-coordinated strategies, it is questionable whether this will be true in reality, since several drawbacks can be mentioned.

Much research has been done to determine the effects of ramp metering, and some focussed specifically on the comparison of different strategies. Zhang et al. (2001) tested five different ramp-metering strategies. Two of them, ALINEA and Zone are local strategies, the others, Bottleneck and SWARM (two versions), are coordinated strategies of which the original local strategy was replaced by ALINEA. Results of this research state that no significant performance differences exist between these strategies (except for SWARM with five-step ahead prediction which performs worse), considering total travel times. A study by Hasan and Ben-Akiva (2002) concluded that FLOW (a coordinated strategy) performs better than ALINEA (local strategy) if the on ramp is hindered by the spill back of downstream congestion. If there is no downstream bottleneck ALINEA performed equal or better. Obviously it can not be concluded that coordinated strategies are better than local strategies, at least not for total travel time. This is confirmed by field operational tests Papageorgiou et al. (1997) of the ALINEA strategy and its coordinated form METALINE on Boulevard Peripherique in Paris, where results on total travel time, total travel distance and mean speed of (local) ALINEA were even slightly better than (coordinated) METALINE. Although some research suggests that optimal use of ramp metering will only be possible with system wide integrated strategies (Kotsialos and Papageorgiou, 2004), this is not confirmed by the research results mentioned above.

With the increasing computing strength, and increasing possibilities of gathering accurate data about route choice and traffic conditions, it is not unlikely that in the future coordinated strategies will outperform local strategies. Till date however, this is not the case, and therefore this thesis concentrates on optimizing local traffic performance, which will in most situations lead to better system performance as well.

9.1.4 Some examples of existing ramp metering strategies

In this section three frequently used ramp metering algorithms are described: a fixed time algorithm, a demand capacity algorithm (feed-forward)(RWS-C) and the Alinea on ramp algorithm (feedback).

Fixed time control

The fixed time control uses historical data to decide how much traffic can be added to the main flow. The metering rate is set with a cycle time of $1/flow$. This algorithm cannot react to real time traffic situations.

RWS-C

The RWS-C strategy is a demand-capacity strategy, which is extensively used in the USA and in the Netherlands; it is based on measuring the demand (upstream) of the merge area and comparing this to the capacity of the bottleneck, determined by historical data. Since the measurements of volume alone is not enough to determine whether the motorway is congested or not, the occupancy (or speed) downstream of the on ramp is used. If the occupancy is above (or the speed below) a certain threshold the algorithm will go to a minimum metering rate. When occupancy is below (or speed above) a threshold level the metering rate ($r(t)$) is determined by comparing the demand (flow upstream of on ramp) with the capacity:

$$r(t) = \begin{cases} Cap - q_{in} & \text{if } \; O_{out} < O_{cr} \;\; \text{or} \;\; u_{out} < u_{out} \\ r_{min} & \text{otherwise} \end{cases} \tag{9.1}$$

with
$r(t)$ the metering rate, Cap the bottleneck capacity, $q_{in}(t)$ the flow upstream of the bottleneck, which is heading towards it, O_{out} the occupancy downstream of the bottleneck, O_{cr} the critical occupancy, u_{out} the speed downstream of the bottleneck, u_{cr} the critical speed, and r_{min} the minimal metering rate.
When downstream of the bottleneck the traffic is not congested the metering rate is determined by the access capacity (Bottleneck capacity minus (upstream) demand). When the bottleneck is congested (occupancy above threshold or speed below threshold) the metering rate is reduced to the minimal metering rate. Because the predictable disturbances are measured and used in the algorithm the strategy is called a feed forward strategy.

Alinea

The third example is the Alinea ramp metering algorithm. It is a feed back strategy. The main aim of on ramp strategies is to maintain capacity flow downstream of the bottleneck or merge area, and for this reason this control is based on downstream measurements:

$$r(t) = r(t-1) + K_R[O_{des} - O_{out}(t)] \tag{9.2}$$

where O_{des} is a preset desired value (typically O_{cr}) and K_R is a regulation parameter. The algorithm lets more vehicles on the motorway when occupancy is lower than that at capacity (desired occupancy) and reduces the metering rate when the level has been reached.
Figure 9.6 shows the Alinea algorithm operates: the output from the traffic process is measured (occupancy downstream of the on ramp and queue length at the on ramp). The metering rate is reduced when the occupancy is to high (over the critical level) or increased when the occupancy has not reached that level. When the on ramp queue capacity is small one might decide on an override algorithm when the queue gets to large and release a lot of vehicles on the main road.

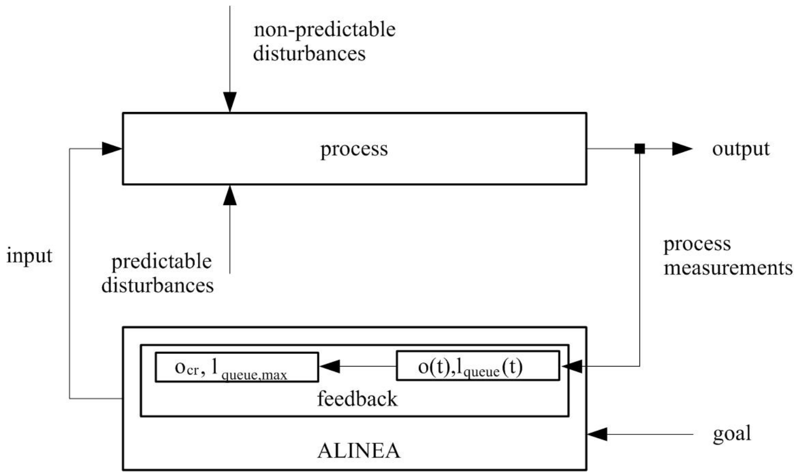

Figure 9.6: Alinea control scheme.

9.2 Development of an alternative algorithm

9.2.1 The stability indicator and merges

Although ramp metering has been studied and implemented for four decades, section 9.1 showed that further improvements should be possible. Progress can be made by optimizing various parts of the control scheme, e.g. when ramp metering strategies have the capability to predict breakdown and capacity. Most currently implemented ramp metering strategies are based on the assumption that it is possible to estimate or predict the value of macroscopic quantities like capacity for a certain cross-section (section 9.1. In reality however, capacity appears to be stochastic and it should therefore be considered as a probability. If the assumptions of a ramp metering strategy are limited to macroscopic quantities, essential information about the configuration of the traffic stream is left out of consideration. As breakdown is highly related to microscopic phenomena (e.g. merging and car-following) it might be better to use microscopic theories to determine the probability of breakdown occurrence. Strategies that are purely based on microscopic theories however, have turned out to be inadequate, as they are influenced by small disturbances too much. Hence, by combining microscopic and macroscopic theories in one strategy, the drawbacks of both theories might be overcome by the advantages of the other. This way the performance of ramp metering can be increased. Parts of the development of the combined ramp metering algorithm are developed in cooperation with M. Ruijgers.

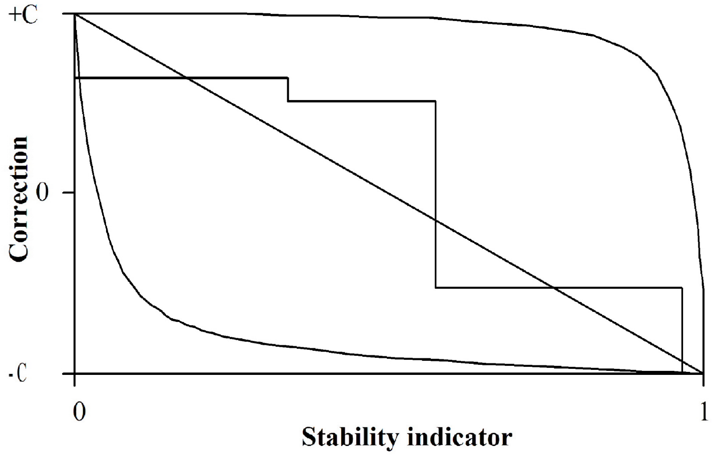

Figure 9.7: 4 out of many possible correction functions of stability indicator on Alinea. When traffic is unstable, the cycle time of the ramp meter will become larger. The exact correction function will be filled in later and depends on the fluctuations of the stability indicator.

9.2.2 Elaboration of Alinea

Microscopic based ramp metering algorithms can show to much fluctuations. A simple way to avoid this in a new strategy is to take Alinea as the basis for metering and correct it for the stability indicator value. An indication (not the exact shape, which will be showed later) can be found in figure 9.7; when the stability indicator is near 0 (stable traffic) the metering rate may be increased with a percentage (C). When the stability indicator is near 1 the metering Alinea calculated should be reduced with a percentage. The values at which the correction takes place and what effect the value of the indicator has are to be set (the shape of the graph has to be set). The figure shows an example in which two thresholds for stability (L_{min} and L_{max}) are set and in between no correction is made.

Constraints of Alinea

Ramp metering is primarily used to improve traffic conditions for freeway traffic, and therefore attention has to be paid to the possible disadvantages for on ramp traffic. To limit the negative consequences of ramp metering to on ramp traffic various problems were handled.

- Thresholds for turning on and off.
- Queue override strategies.
- Cycle time limitations.

- The impact of the measures of effectiveness used.

Calibration of Alinea

For optimal usage of the capabilities of ALINEA the parameters will be obtained from comparison of the simulation results with road-side measurements.

For the parameter setting, the regulator parameter, the interval length, the detector locations and the desired occupancy were calculated (Ruijgers, 2005) using either experiments or literature . The desired occupancy used is 22%, the detector is located at 260 m downstream of the merging section, the time interval is 60 seconds and the parameter used is 70 veh/h.

9.2.3 The combined algorithm

In the previous sections relevant aspects of the new strategy were discussed. In section 9.1 it was substantiated what the main problems of current ramp metering strategies are and how these can be overcome by combining macroscopic and microscopic theories. A layout of the new strategy, which consists of two elements, was presented. One part is a macroscopic component (ALINEA) which is used to calculate cycle times based on downstream occupancy. The second part is a microscopic component (including a string stability indicator) which is used for fine tuning of the cycle times by a correction based on upstream stability. In this section all components will be integrated into a new strategy. The control scheme of the new strategy, including all elaborated elements is shown in figure 9.8.
The 'predictable disruptions' are the vehicles which are measured upstream (the average and deviation of their headways $\bar{h}$ and $h(\sigma)$). The stability indicator is based these headway measurements. The data for the Alinea algorithm is measured downstream of the bottleneck. In the stability indicator the influence of the location of the detectors and the level of headways at which a group of vehicles is not longer considered a platoon is also integrated.

Prediction time

Main idea of the microscopic component of the combined algorithm is that if local upstream traffic conditions are stable it is beneficial to release more vehicles than suggested by ALINEA. When the traffic flow is unstable less vehicles should be permitted to merge. This is achieved by correcting the proposed cycle time. It depends on the stability course in time and space how important (or even essential) it is for a single on ramp vehicle to merge exactly in front of this string. As was mentioned in section 9.1 release-to-gap strategies have been implemented in the past but appeared to be unsuccessful. Reasons for this failure were large varieties of variables in the strategy. From this can be concluded that a new strategy can only be successful if stability does not fluctuate with every vehicle that passes the detection loops on the motorway. It is not necessary for an on ramp vehicle to merge exactly in front of a string,

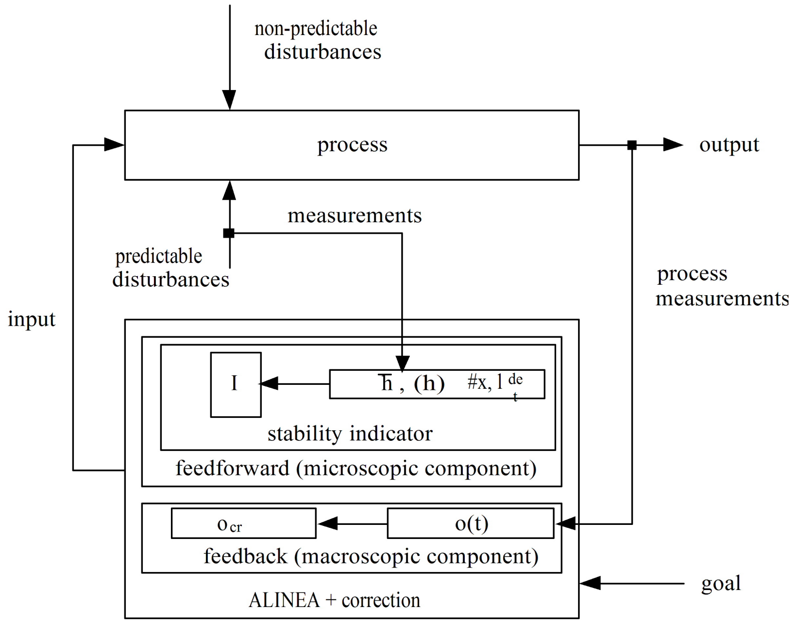

Figure 9.8: Setup of the combined algorithm.

as merging one or two vehicles more upstream or downstream implies merging in front of a string with approximately the same stability (only one or two out of ten headways differ).

In this section the optimal detector location is be estimated (after Ruijgers (2005)). To minimize the effect of lane changing and speed changes on the measured stability, the stability detector should be chosen as much downstream as possible. However, the time period (process time) between the measurement of the first vehicle of the string and the arrival of the string and on ramp vehicle at the merging zone should be taken into account. The detector location is determined by the product of motorway speed and process time. Process time consists of: measure time, calculation time, and acceleration time (figure 9.9).

Measure time (needed to calculate the stability) is the time between the detector crossing of the rear of the first vehicle of a string up to the crossing of the front of the tenth vehicle. It depends on the local freeway flow on the right lane. Higher flow levels will reduce the measure time and as a result, the necessary distance between the detector and on ramp can be smaller.

Calculation time is a negligibly part of the process time as the only quantities that have to be calculated are the average and standard deviation of the measured time gaps within a string and the resulting stability, which will be converted into a correction time to ALINEA.

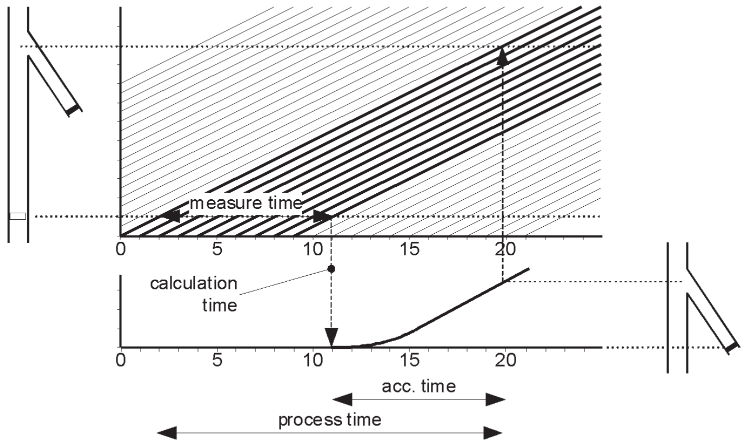

Figure 9.9: Buildup of the process time; after Ruijgers 2005.

Acceleration time is the time it takes for an on ramp vehicle to reach the merging zone, from the moment the ramp-meter turns green. Slow accelerating vehicles or inattentive drivers need more time to reach the merging zone, thus if for example a truck is waiting for the ramp meter, the optimal detector location is further upstream.

The same process time requires different detector locations for different freeway speeds. If motorway speed is high, the vehicle trajectories in figure 9.9 will be steep, resulting in a longer required distance between the stability detector and merging zone. Clearly, it is essential to determine one detector location, because (apart from the conceptual consequences of using multiple detectors) it will be too expensive to use multiple detectors. Therefore the detector location is chosen such that the probability of a successful merge in front of the measured vehicle string is maximal and is based on visual inspection of the simulated network and traffic flows. The optimal distance between the stability detector and the ramp nose appears to be about 600 meters upstream of the ramp nose. For a full description of this see Ruijgers (2005).

Correction function

A comparison of stability and speed did not reveal an obvious relationship between stability and breakdown. Nevertheless a pattern was found for stability in time (Ruijgers, 2005). Consider figure 9.10 which shows the stability in time for the simulated period.

Stretches of about half a minute of stable circumstances are separated by shorter periods of unstable behaviour. This is encouraging, because it means that the correction function can be applied. Correcting ALINEA's cycle time can only be successful if stability does not change within a few seconds from

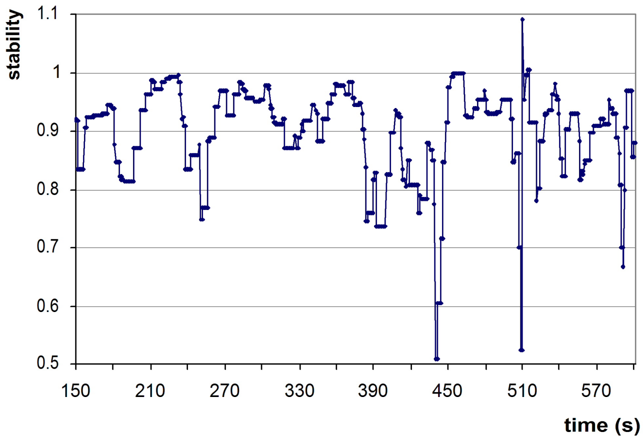

Figure 9.10: Stability in time, after Ruijgers 2005.

very stable to unstable. When stability is between 0.9 and 1.0 a large drop into unstable regions seldom follows. Between 0.8 and 0.9 a transition zone can be seen in which stabilities can both increase and decrease very fast. Finally, stabilities below 0.8 only exist for a short time. This is no evidence that a pattern indeed exists, especially since also time periods can be found which show more disordered observations. Based on these findings a correction function is designed, which is shown in figure 9.11.

For unstable situations, the cycle time will be increased with either one or two seconds. For stabilities within the transition zone between 0.8 and 0.9 no correction will be applied. Finally, for very stable observations cycle time will be decreased with one second.

9.3 Simulation

This section shows the setup of a simulation environment in which the Alinea algorithm and the combined algorithm can be compared. The selection of the computer model the simulation input (OD-levels and network) are after Ruijgers (2005).

9.3.1 Model selection

Boxill and Yu (2000) compared 81 traffic simulation programs (microscopic, mesoscopic and macroscopic). A first comparison was made based on five criteria, discerning whether the program:

- uses credible theories

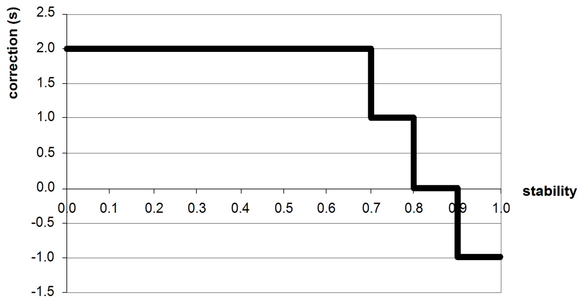

Figure 9.11: Correction function.

- has been tested in real world
- is able to output measures of performance
- incorporates at least one ITS feature
- is obtainable by public

All five criteria are relevant in this project to select the best program for evaluating the performance of the new strategy, but other criteria are important as well. The program should:

- enable the implementation of new ramp metering strategies
- be microscopic
- able to model motorway traffic

Out of the 81 simulation programs, only three satisfy all eight conditions. These are:

- AIMSUN
- Paramics
- VISSIM

It is recognized that this approach does not guarantee that all appropriate programs are considered, as the study of Boxill and Yu dates from 2000 and improvements are made to most programs frequently.

Model choice

Of the three mentioned programs AIMSUN is used in a lot of traffic studies at Twente University. As experiences with AIMSUN are positive, and no obvious arguments in favour of one of the other programs were found, it was decided to use this program for the comparison of the combined strategy to Alinea. In Ruijgers (2005) the model selection can be read in detail. The topics which were considered are:

- possibility to implement new ramp metering algorithms
- realism of the lane change model
- consequences of the used car-following theory

The considerations are based on theories applied in AIMSUN as described in (Barcelo and Ferrer, 1997) and results of the calibration and validation process.

9.3.2 Input and network

Scenario's

Within this study three scenarios will be compared:

- no-control scenario
- (traditional) ALINEA scenario
- combined strategy scenario

The no-control scenario will be used for the calibration process of ALINEA and for determining network specific quantities like capacity. For the actual comparison of ALINEA and the combined strategy, the second and third scenario will be used. Furthermore, the no-control scenario will be used to put possible travel time savings in perspective.

Network

The simulation network consists of one long homogeneous three-lane freeway section, with one single on ramp. The freeway has a total length of 6 kilometres and the on ramp is located at 4 kilometres from the beginning. In this way the formation and propagation of shockwaves both upstream and downstream from the on ramp as well as the build up of queues are enabled. In appendix B an overview of the network is given, including all detector locations.

The merging area is designed based on recommendations by the Ministry of Transport (AVV, 2002b) for ramp metered on ramps and guidelines that are used in The Netherlands for motorway design. The minimum upstream on ramp length is 200 meters to create enough buffer capacity for ramp meter queues. In the simulation the length of the on ramp is much longer, to ensure

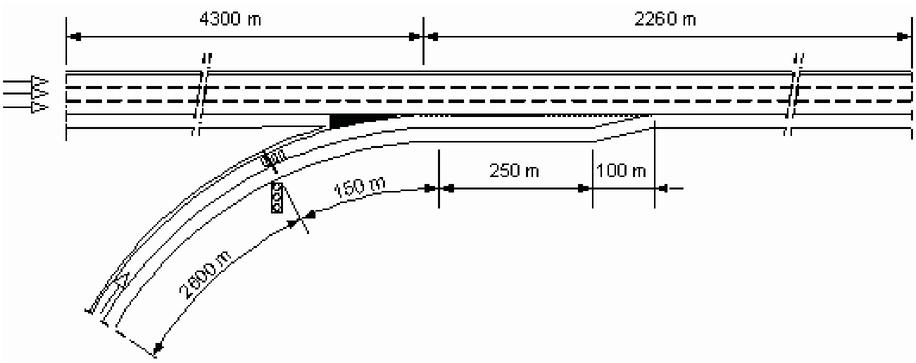

Figure 9.12: Research network.

that the tail of the on ramp queue will not reach the network boundaries. The acceleration length of 250 meters between ramp meter and on ramp nose is necessary to provide enough time for trucks to gain speed.

The network (figure 9.12) is equipped with detectors to measure data as input for the ramp metering strategies and as output for calibration and evaluation. In reality (if possible) existing detectors will be used to avoid unnecessary construction duties to the freeway traffic and thus the choice of detector location is limited. As these drawbacks do not apply to the simulation network, detector locations are optimal.

OD-patterns

Two demand levels are used to compare the travel times of the three scenarios. They are based on the demand pattern of figure 9.13. Simulations are performed for 90 minutes. After an initial period of moderate demand, demand increases to high demand, which will be sustained for 30 minutes. Then demand is reduced to the original moderate demand level, and finally forty minutes of low demand follow to let the queue dissolve.

Two demand patterns (Demand 1 and Demand 2) are used, of which the 100% flow on the freeway is respectively 7250 veh/h and 7900 veh/h. On ramp demand is for every period 10% of freeway demand. Truck rate is 10% of the total demand consisting of 7.5% normal trucks and 2.5% long trucks.

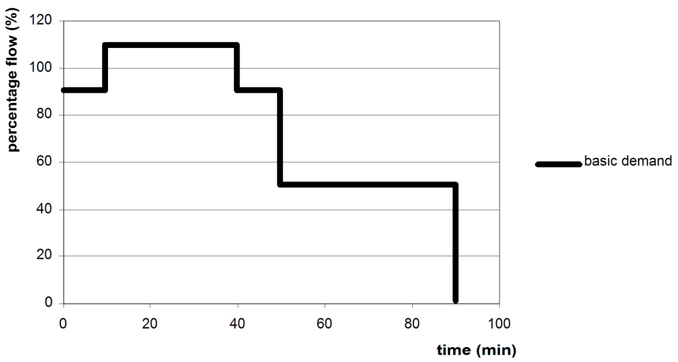

Figure 9.13: Demand levels as a percentage of the demand. Demand 1 = 7250 veh/h, demand 2 = 7900 veh/h.

9.3.3 Performance indicators

The main objective of the combined strategy is to further reduce travel times. To verify this objective the following Measures Of Effectiveness (MOE's) are defined:

- average freeway travel time
- average on ramp travel time
- average travel time of all traffic

9.4 Results

Replication of simulation runs

The three scenario's (no control, Alinea algorithm, combined algorithm) were modelled for the two demand levels (demand 1 and 2). To get statistical significant results 50 runs were done for each simulation (for the significance test see Ruijgers (2005)). All scenario's are based on the same random seed numbers.

9.4.1 Results of the simulations

Before the results of the combined strategy are presented, the travel time savings of ALINEA compared to the no-control scenario are shown in table 9.1.

Clearly, ramp metering is especially effective during the peak demand pattern (demand 2). The large increase in travel time for on ramp vehicles (more than six minutes per vehicle) is due to the experimental design in which practical

Table 9.1: Results of the Alinea metering algorithm compared to no metering (in seconds.

	motorway traffic	on ramp traffic	average
Demand 1	-10	+76	-2
Demand 2	-101	+385	-56

constraints like maximum cycle times and queue override strategies were excluded. These results do not necessarily hold in practise: they can be a starting point from which the Alinea algorithm can be compared to the combined algorithm.

Table 9.2: Results of the combined algorithm compared to the Alinea metering algorithm for different detector distances. Detector numbers can be found in figure B.1 in appendix B. Detector 84 is 470m from the ramp nose, detector 76 is 650m from the ramp-nose; the detectors are equally distributed.

	detector	motorway traffic	on ramp traffic	average
Demand 1	76	0	+4	0
	78	0	0	0
	80	0	+9	1
	82	0	0	0
	84	0	+8	+1
Demand 2	76	-1	-53	-6
	78	-1	-47	-5
	80	-1	-81	-8
	82	-1	-66	-7
	84	-1	-50	-5

For Demand 1 the effect of the new strategy is rather small and adds travel time to the traffic on the on ramp. This might be explained by the small effect that was gained by implementation of traditional ALINEA compared to the no-control scenario. Apparently, traffic conditions are such that ramp metering does hardly yield a profit.

Clearly, the combined strategy has positive effect for demand pattern 2, which has a larger peak demand. All used detector locations lead to an improvement. A reason for this might be the fact that due to speed and flow differences optimal detector location varies and, as a result, on the detector locations more upstream or downstream, the same string characteristics are measured too, though less frequent.

For Demand 2 the original average travel time saving per vehicle was 56 seconds. If the combined strategy is used, the total travel time savings can run up to 65 seconds, which means that ramp metering effectiveness (for this demand pattern) is increased with 15%.

Another improvement in the combined algorithm is that it reduces both motorway and on ramp travel times.

Obviously the combined strategy can increase ramp metering performance, but not for all demand levels.

9.4.2 Discussion

Further research can probably increase the performance of the combined strategy more. The stability indicator can be elaborated and in the previous chapter suggestions on how to improve the indicator theoretically were made.

A string stability indicator in which stabilities are defined for all combinations of average and standard deviation might give more insight in the relationship between stability and breakdown. If a relationship is found, the correction function of the ramp metering algorithm can be improved. The correction function that was used in this thesis was based on an indistinct pattern of stability in time. It is expected that improvement of the stability indicator will lead to a further increase of ramp metering effectiveness.

Another method to improve the algorithm is to test and adjust it in a more realistic research environment or in practise. The travel time savings of table 9.2 are based on a research environment in which individual drivers can not complain about queue lengths and waiting times, has no queue override nor red light neglect. Therefore the constraints to the ramp metering strategies that are essential in practice could be excluded. Consequently, delay times of the on ramp vehicles are much larger than they are in reality, and therefore effectiveness of ramp metering in general, and the combined strategy in particular will probably be smaller than suggested by table 9.2. Nevertheless, the combined strategy, which contains macroscopic and microscopic elements outperforms ALINEA for on ramp traffic without adding extra travel time to the highway traffic.

9.5 Summary

In this chapter the traffic flow stability indicator was used to improve ramp metering.

Ramp metering usually uses only macroscopic traffic characteristics (density or flow) while the stability indicator uses more microscopic characteristics (headway distribution of a small group of vehicles).

Algorithms that are used for ramp metering can also be split into feedforward and feed back algorithms. The Alinea ramp metering algorithm (feed back) was combined with the stability indicator (feed forward). The combined algorithm used Alinea as the basis of the metering rate and added (or reduced) the number of vehicles allowed on the motorway when the traffic flow was stable (or unstable) according to the stability indicator. The combined (Alinea +

stability indicator) algorithm performs 15% better than the Alinea algorithm; it reduces the average travel time on both the motorway and the on ramp compared to the travel times with the Alinea ramp metering algorithm. Alinea (compared to no metering) reduces travel times on the motorway and adds travel time to traffic on the on ramp. Especially under heavy traffic conditions the combined algorithm proves to be a significant improvement.

Chapter 10

Conclusions and perspectives

10.1 Summary and conclusions

This research is conducted because a phenomenon was left unexplained in Van Toorenburg and Elbers (2000). In that research two road sections with the same number of lanes, up grade level, percentage of truck drivers, and the same weather conditions showed a clear difference in capacity.

Visual inspection of merging traffic shows that the influence of clusters on the throughput of a bottleneck is large.

Research objectives

Three research objectives are phrased:

- The main objective of this research is to give better explanations and get better predictions of the onset of traffic jams using data from existing measuring techniques.
- The second objective is to research how existing theories on traffic breakdown can be applied to develop measures to better predict congestion, operationalized as indicators.
- The third objective is to apply the measurement tools (indicator) to improve ramp metering.

Traffic stability

Although the term 'stability' in relation to traffic is used often, the Highway Capacity Manual (Transportation Research Board, 2000) and the Dutch Capacity on Motorways (AVV, 1999a), do not give a definition. Nevertheless 'stability' is often associated with homogeneous traffic. This dissertation describes how different types of stability can be defined, measured and used to help predict the onset of congestion.

The traffic system exists of quite some, mostly non-linear and complex, relations between drivers and their surrounding. The surrounding consists, among others, of other drivers, the road and it's condition, the vehicle, the (static) traffic rules, the weather and dynamic traffic management (DTM). DTM is often aimed at homogenizing the different aspects (speed or the distribution of vehicles) of the traffic flow. Examples of this are speed checking and ramp metering.

A microscopic simulation proves that homogenizing traffic will not always increase capacity. Simulation of an on ramp shows that if traffic on the main road is homogenized (in different aspects), the inhomogeneous flow has more gaps fit for merging and therefore a bottleneck can have a larger capacity.

From these simulations is concluded that traffic flows should not be homogenized near on ramps. The fluctuations of the density form the basis for good merging conditions.

The stability of a traffic flow is defined as the capacity to handle disruptions.

Three types of stability can be distinguished in literature:

- local stability is concerned with the car-following behaviour of two vehicles;
- string stability is concerned with the car-following behaviour of a string of vehicles, one leader and a line of followers
- traffic *flow* stability is not concerned with car-following, but with disruptions in macroscopic characteristics of the traffic flow.

To measure stability as defined above three stability indicators were operationalized and tested:

1. based on traffic reliability, as in (Ferrari, 1989). Traffic is called reliable when no hazardous speed drop will occur.
2. based on traffic synchronization of traffic flows prior to breakdown, as described in (Kerner and Rehborn, 1996). The probability of the onset of a traffic jam is based upon speed and flow fluctuations.
3. based on string stability; a platoon of vehicles is simulated microscopically and checked for it's robustness to handle a certain speed disruption without the flow breaking down.

The three stability indicators were tested on data from the A2 motorway, and the indicator based on string stability appears to be better than the other two in predicting the onset of traffic jams (correlation of 0.7, while the others show 0.5).

The stability indicator based on string stability was implemented into a ramp metering algorithm and the performance of this algorithm was compared and found better than the performance of the Alinea ramp metering algorithm in a simulation environment.

Data collection

One of the goals of the research is to find better predictions and explanations for the onset of traffic jams, *using existing traffic data collecting systems.* All the data used in this research is available from common Dutch motorway monitoring systems.

When vehicle trajectories had been widely available, microscopic variables as position, speed, headway and distance could have been deducted from them. That way it would be easy to deduct characteristics as string stability. This is not the case; up to now trajectories can only be gathered through expensive methods (making images from helicopters or other flying platforms) and will not be collected on a large scale for years. This research uses data from loop detection on the motorway (double loops per lane, each 500m). Individual vehicle data which is collected with loop detection is only available in some research area's but soon this individual data collection system will be available on most parts of Dutch motorways.

Results

Both the development of the stability indicators and the implementation into the ramp metering algorithm show that microscopic distribution of the vehicles (clustering) plays an important role in the breakdown phenomenon. This was also suggested by the microscopic modelling of an on ramp at a highway, which was filled with homogeneous or less homogeneous traffic.

Disruptions in a traffic flow depend on the distribution of gaps and reaction times of the drivers. If gaps are to short or reaction times to long a following vehicle will apply it's brakes downstream of where it's leader did and the deceleration has to be stronger to prevent a collision. The result is a shockwave of vehicles which are breaking more and more intense.

The stability indicator is used to create a combined on ramp metering algorithm. The Alinea algorithm is corrected by the stability indicator: low (and high) stability were used to reduce (or increase) the metering rate which is calculated by Alinea in a simulation environment (Aimsun). The combined algorithm was compared to the original Alinea algorithm and it proved to be 15% more efficient; reducing travel times on the motorway and the on ramp.

Overview of conclusions

- Although there are clear definitions the term stability is still not always used in a proper way. Definitions in the Highway Capacity Manual can improve this situation.

- A stability indicator based on string stability can contribute to a better prediction of breakdown.

- The stability indicator can be used for a better and more effective use of ramp metering.

- Homogenizing traffic near an on ramp will not increase the on ramps' capacity. The results in this research even imply that capacity will reduce.

The research shows that the traditional choice between microscopic or macroscopic approaches to traffic flow questions is not suitable for all problems. Apparently there are traffic phenomena which call for more than just measures based on macroscopic behaviour of the traffic flow. The initial starting point of this research, the belief that traffic flows cannot be distinguished by macroscopic characteristics only, especially when the traffic situation is critical, is found to be true. Both the stability indicator, as the combined ramp metering algorithm show that adding or replacing microscopic knowledge to or for macroscopic characteristics makes differences between traffic flows, which initially seemed identical, measurable.

10.2 Perspectives and future research

In the effort of getting more vehicles through a road network in the same time and increase travel time reliability, the predictability of the onset of traffic jams in bottlenecks plays a significant role. Next to postponing the onset of a traffic jam, also upgrading the level of throughput and the level of traffic stability seem important to keep traffic flowing. A stability indicator can give information on what the consequences are when more and more traffic is fed to a bottleneck like an on ramp.

The string stability indicator, developed in this research, appears to be the most promising stability indicator to give a better prediction of breakdown in bottlenecks. It offers a fast and easy way of calculation. It is created using a single disturbance and all vehicles heave the same initial speed and reaction. Even with these simplifications (a lot of detail is neglected) the indicator outperforms other indicators (using density, flow and average speed) which should be able to predict the onset of congestion.

Indicator improvements

The indicator can get more realistic when the actual passage times and the actual speeds can be put in prior to simulating the disturbance and calculating the string stability.

Also the disturbance of the first vehicle can be varied according to what the most expected disturbance is, e.g: the 'normal speed drop and recovery' in case of an on ramp can be replaced by a 'slow speed drop' in case of a reduced speed limit. Multi-lane calculation now uses the probability to change between lanes, but does not incorporate the effects of the merges to the stability indicator of the individual lanes. Incorporation of this can improve the multi-lane indicator. Also, at this moment the indicator is developed for each lane and, based on speed differences and headway distribution, lane changing possibilities are calculated. The difference between an early lane change (e.g. the second vehicle in line changes lane because of the disruption) or a late one (e.g. the ninth vehicle in line changes lane because of the disruption) is neglected at this stage. This problem could be dealt with, when a lane changing model is incorporated in the existing car-following model and all lanes are modelled simultaneously in a microscopic model. This will cost a serious calculation effort: for each vehicle over all lanes crossing the measuring point the model must be re-run. Also real-time implementation of the current stability indicator sets serious demands on communication and calculation speed.

It is difficult to predict how much the performance of the indicator will improve, but as the distribution of the vehicles is dominant over the type of disruption one may expect a significant improvement. Real-time availability of the indicator needs a lot of computer and communication capacity, especially when instead of a look-up graph all situations (speed and headway per vehicle) are used as input. A large increase of the number of vehicles in the platoon definition has a downside that further upstream of the potential bottleneck measurements need to be taken, which makes it more difficult to predict the configuration of the vehicles at the location of potential disruption.

Ramp metering as a case study was chosen to incorporate the stability indicator. The reason for this is that in merging sections disruptions surely occur, because of mandatory lane changes. The incorporation of the stability indicator into the ramp metering algorithm was tested using micro simulation; the algorithm needs to be tested in practice as well.

Implementation of the indicator in other dynamic management tools could also be beneficial. Two examples are:

Dynamic Route Information Signs These can give the drivers information as 'stable' or 'non-stable' which attributes to their travel time predictability.

Signalling At levels of critical stability the signalling can help to influence traffic; homogenize over lanes and reduce speeds, which results in more stable traffic. Notice that this is only beneficial on sections which do not have an on ramp bottleneck.

The stability indicator developed in this research can be upgraded on different aspects. The suggestions for this done earlier in this section or making the indicator available real-time require high calculation and communication performance.

Implementation outside of the domain of ramp metering looks promising, but on-line implementation of the current, easy to calculate, stability indicator in an ramp metering algorithm has higher expectations, because of the model results.

Bibliography

Ahmed, K. I., Ben-Akiva, M. E., Koutsopoulos, H. N. and Mishalani, R. G. (1996). Models of freeway lane changing and gap acceptance behavior, *Proceedings of the 13th International Symposium on Transport and Traffic Theory*, Lyon, France.

Akcelik, R. and Chung, E. (1994). Calibration of the bunched exponential distribution od arrival headways, *Road & Transportation Research* **3**(1): 42–59.

AVV (1993). Guidelines for the design of motorways (in Dutch: Richtlijnen voor het Ontwerpen van Autosnelwegen, *Research report*, Ministry of Transport, Public Works and Water Management, Transportation Research Centre, The Hague, The Netherlands.

AVV (1998). Monitoring no overtaking for trucks (in dutch), *Research report*, Ministry of Transport, Public Works and Water Management, Transportation Research Centre, Rotterdam, The Netherlands.

AVV (1999a). Hand book for capacities of motorway sections (in dutch: Capaciteitswaarden infrastructuur autosnelwegen, handboek), *Research report*, Ministry of Transport, Public Works and Water Management, Transportation Research Centre, Rotterdam, The Netherlands.

AVV (1999b). Mtm-2 general introduction (in dutch: Mtm-2 algemene inleiding), *Research report*, Ministry of Transport, Public Works and Water Management, Transportation Research Centre, Rotterdam, The Netherlands.

AVV (2000). Traffic flow operations during congestion, *Research report*, Ministry of Transport, Public Works and Water Management, Transportation Research Centre, Rotterdam, The Netherlands.

AVV (2002a). Alternatives for loop detection (in dutch: Alternativen voor lusdetectie. resultaten en conclusies), *Research report*, Ministry of Transport, Public Works and Water Management, Transportation Research Centre, Rotterdam, The Netherlands.

AVV (2002b). Manual for the preparation and realization of on-ramp metering (in Dutch: Handleiding voorbereiding en uitvoering toeritdoseerinstallaties), *Research report*, Ministry of Transport, Public Works and Water Management, Transportation Research Centre, The Hague, The Netherlands.

Banks, J. H. (1989). Freeway speed-flow-concentration relationships: more evidence and interpretations, *Transportation Research Record* **1225**: 53–60.

Banks, J. H. (2003). Average time gaps in congested freeway flow, *Transportation Research* **37**(A): 539–554.

Barcelo, J. and Ferrer, J. (1997). Aimsun2: Advanced interactive microscopic simulator for urban networks, *User's manual*, Universitat Politecnica de Catalunya.

Blackwell, O. and Blackwell, H. (1971). Visual performance data for 156 normal observers of various ages, *Journal of Illumination Engineering* **1**(1): 3–13.

Blumentritt, C. W., Pinnel, C., McCasland, W. R., Ross, D. W. and Glazer, J. (1981). Guidelines for selection of ramp control systems, *Natl. Coop. Highway Research Report 232*, Highway Research Board.

Botma, H. (1978). State-of-art-report trffic flow models (in dutch), *Research report r-78-40*, SWOV.

Boxill, S. A. and Yu, L. (2000). An evaluation of traffic simulation maodels for supporting its development, *Technical report*, Centre for Transportation Training and Research, Texas Southern University.

Castillo, J. M., Pintado, P. and Benitez, G. (1993). The reaction time of drivers and the stability of traffic flow, *Transportation Research* **28 B**: 35–60.

Cavallo, V., Mestre, D. and Berthelon, C. (1997). Time-to-collission judgements: Visual and spatio-temporal factors, *Traffic & Transport Psychology, Theory and Application.*

Clarke, A. B. (1953). The time-dependent waiting line problem, *University of Michigan Report M720- 1R39*, University of Michigan.

Cowan, R. J. (1975). Useful headway models, *Transportation Research* **9**: 371–375.

Cross, D. and Harris, C. M. (1998). *Fundamentals of Queueing Theory*, John Wiley & sons, Inc, New York, USA.

Daganzo, C. F. (1981). A model for the structure of lane changing decisions, *Transportation Research* **20 B**(5): 403–414.

Daganzo, C. F. (2002). A behavioral theory of multi-lane traffic flow. part i: Long homogeneous freeway sections, *Transportation Research* **36 B**: 131–158.

Dijker, T., Bovy, P. H. L. and Vermijs, R. G. M. M. (1997). Car-following under non-congested and congested conditions, *Research report vk2206.301*, Delft University of Technology, Traffic Engineering Section, Delft, The Netherlands.

Drew, D. R. (1968). *Traffic Flow Theory and Control*, McGraw-Hill, New York, USA.

Edwards, J. B. (1999). Speed adjustment of motorway commuter traffic to inclement weather, *Transportation Research* **2 F**(2): 1–14.

Elbers, J. A. C. M. and Van Berkum, E. C. (2003). Results of dtm should be measured (in dutch: Effectmeting dvm geen overbodige luxe, *Verkeerskunde* **4**: 40–44.

Elefteriadou, L. and Lerworawanich, P. (2002). Defining, measuring and estimating freeway capacity, *Transportation Research Board (TRB) 2002 Annual Meeting CD-ROM*, Washington, USA.

Ferrari, P. (1988). The reliability of the motorway transport system, *Transportation Research* **22B**(4): 291–310.

Ferrari, P. (1989). The effect of driver behaviour on motorway reliability, *Transportation Research* **23B**(2): 139–150.

Ferrari, P. (1994). The instability of motorway traffic, *Transportation Research* **28B**(2): 175–186.

Fox, M. (1989). *Elderly Driver's Perceptions of Their Driving Abilities Compared to their Functional Visional Perception Skills and Their Actual Driving Performance.*, Assessing the driver ability of the Elderly, Hayword Press, New York, USA.

Fritzsche, H. (1994). A model for traffic simulation, *Traffic Eng. Control* **35(5)**: 317–321.

Garber, N. J. and Hoel, L. A. (2002). *Traffic and Highway Engineering*, Brooks/Cole, Thomson Learning, Bill Stenquist, California, USA.

Gipps, P. G. (1981). A behavioural car-following model for computer simulation, *Transportation Research* **15B**: 105–111.

Gipps, P. G. (1986). A model for the structure of lane changing decisions, *Transportation Research* **20B**(5): 403–414.

Green, M. (2000). How long does it take to stop? methodological analysis of driver perception-brake times, *Transportation Human Factors* **2**: 195–216.

Greenshields, B. D. (1934). A study in highway capacity, *Highway Research Record* **14**: 448–477.

Griffiths, J. D. and Hunt, J. G. (1991). Vehicle headway in urban area's, *Traffic Engineering and Control* pp. 448–462.

Griffths, J., Williams, J. and Leonenko, A. (2004). Transient solution for m/ek/1 system, *Internet page*, Cardiff University.

Grontmij (1998). Evaluation of average speed checking on the a2 motorway, *Report V&I-98005780*, De Bilt, The Netherlands.

Hasan, M. and Ben-Akiva, M. (2002). Evaluation of ramp control algorithms using microscopic traffic simulation, *Transportation Research c* **10**: 229–256.

Heidemij (1993). Evaluation of the test homogenization on the a2 motorway, *Report 642/BA93/A360/05990*, The Netherlands.

Herman, J. J. and Weiss, G. H. (1961). Comments on the highway crossing problem, *Operations Research* **9**: 828–840.

Herman, R., Montroll, E. W., Potts, R. and Rothery, R. W. (1959). Analysis of stability in car-following, *Operations Research* **1**: 86–106.

Holland, E. N. (1998). A generalised stability criterion for motorway traffic, *Transportation Research B* **32**: 141–154.

Hoogendoorn, S. and Bovy, P. (2000a). Modelling multiple user-class traffic flow, *Transportation Research* **B**(34): 123–146.

Hoogendoorn, S. P. (1999). *Multiclass Continuum Modelling of Multiclass Traffic Flows*, PhD thesis, Trail Thesis series, Delft University of Technology, Delft, The Netherlands.

Hoogendoorn, S. P. and Bovy, P. H. L. (2000b). State-of-the-art of vehicular traffic flow modelling, *Report*, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands.

Hoogendoorn, S. P., van Zuylen, H. J., Schreuder, M., Gorte, B. and Vosselman, G. (2003). Microscopic traffic data collection by remote sensing, *82nd TRB annual meeting*.

Jessurun, M. (1997). *Driving through a road environment*, PhD thesis, Groningen University, Groningen, The Netherlands.

Kendall, D. G. (1953). Stochastic processes occurring in the theory of queues and their analysis by the method of imbedded markov chains, *Ann. Math. Statist. Soc. Ser.* **24**: 338–354.

Kerner, B. S. and Rehborn, H. (1996). Experimental properties of complexity in traffic flow, *Physical Revieuw* **53B**(5): R4275–R4278.

Kijk in de Vegte, N. (2002). Characteristics of cluster drivers (in dutch: Kenmerken van clusterrijders, *Research report TT02-08*, University of Twente, Centre for transport studies / Traffic Test, Enschede/Veenendaal, The Netherlands.

Kita, H. (1993). Effect of merging lane length on the merging behavior at expressway on-ramps, *Transportation and Traffic Theory*, pp. 37–51.

Kita, H. (1999). A merging-giveway interaction model of cars in a merging section: a game theoretic analysis, *Transportation Research* **33A**: 305–312.

Klein, L. A. (2001). *Sensor technologies and data requirements for ITS*, Artech-House, Boston, USA.

Kleinhout, J. J. (1984). Motorway control and signalling: The test of time, *Traffic Engineering Control* (25): 193–197.

Koshi, M., Honda, H. and Mori, H. (1992). Optimization of isolated traffic signals, *International Conference on Road Traffic Signalizing*, pp. 5–8.

Kotsialos, A., Diakaki, C., Pavlis, Y., Papageorgiou, M. and Vardaka, G. (1998). Short term traffic forecasting using the macroscopic simulation tool metanet, *Proceedings of the DACCORD Short-term Forecasting Workshop*, pp. 13–29.

Kotsialos, A. and Papageorgiou, M. (2004). Motorway network traffic control systems, *European Journal of Operational Research* **152**: 321–333.

Kroemer, K. H. E., Kroemer, H. B. and Kroemer-Ebert, K. E. (1994). *Ergonomics: How to design for Ease and Efficiency*, Prentice-Hall, Englewood Cliffs, New York, USA.

Lay, M. G. (1986a). *Handbook of road technology, volume 1, Planning and pavement*, Transportation studies volume 8, Gordon and Breach science publishers, London, England.

Lay, M. G. (1986b). *Handbook of road technology, volume 2, Traffic and Transport*, Transportation studies volume 8, Gordon and Breach science publishers, London, England.

Leutzbach, W. (1988). *Introduction to the Theory of Traffic Flow*, Springer-Verlag, Berlin Heidelberg.

Liang, C. Y. and Peng, H. (2003). String stability of adaptive cruise controlled vehicles, *TRB 2003 Annual Meeting CD-ROM*, Washington, USA.

Lighthill, M. J. and Witham, G. B. (1955). On kinematic waves. ii. a theory of traffic on long crowded roads, *Proceedings of the Royal Society A*, London, England, pp. 371–345.

Lorenz, M. R. and Elefteriadou, L. (2001). Defining freeway capacity as a function of the breakdown probability, *Transportation Research Board (TRB) 2001 Annual Meeting CD-ROM*, Washington, USA.

Marmor, M. (1982). Aging and the retina, *in* R. Sekuler, D. Kline and K. Dismukes (eds), *Aging and Human Visual Functions*, Alan R. Liss, New York, USA.

May, A. D. (1990). *Traffic flow fundamentals*, Englewood Cliffs, Prentice-Hall, England.

McPherson, K., Michael, J., Ostrow, A. and Shaffron, P. (1988). Physical fitness and the aging driver - phase i, *Technical report*, AAA Foundation of Traffic Safety, Washington D.C., USA.

Michalopoulos, P. and Hourdakis, J. (2001). Review of non-intrusive advanced sensor devices for advanced traffic management systems and recent advances in video detection, *Proceedings of the Institution of Mechanical Engineerings Part I: Journal of Systems and Control Engineering*, number 215(14), pp. 345–355.

Miller, A. J. (1972). Nine estimators of gap acceptance parameters, *Proceedings, 5th International Symposium on the Theory of Traffic Flow and Transportation*, New York, USA, pp. 215–235.

Minderhout, M. M. (1999). *Supported Driving: Impacts on Motorway Traffic flow*, PhD thesis, Trail Research School, Delft, The Netherlands.

Ortuzar, J. D. and Willumsen, L. G. (1990). *Modelling Transport*, Wiley, Sussex, England.

Ovuworie, G. C., Darzentas, J. and Mcdowell, M. R. C. (1980). Free movers, followers and others: a reconsideration of headway distribution, *Traffic Engineering and Control* pp. 425–428.

Papageorgiou, M. (1983). *Applications of automatic control concepts to traffic flow modelling and control*, Springer, Berlin, Germany.

Papageorgiou, M. (1991). *Concise encyclopedia of traffic & transportation systems*, Pergamon Press, Frankfurt, Germany.

Papageorgiou, M., Haj-Salem, H. and Middelham, F. (1997). Alinea local ramp metering: Summary of field results, *Transportation Research Board (TRB) 1997 Annual Meeting CD-ROM*, Washington, USA.

Pipes, L. A. (1953). An operational analysis of traffic dynamics, *Journal of Applied Physics* **24**: 274–281.

PTV (n.d.). Vissim - traffic flow simulation, *Technical report*, PTV.

Quadstone (2002). Paramics v4.0 system overview, *Technical report*, Quatstone, limited, http:// www.paramics-online.com, Edinburgh, Scotland.

Richards, P. I. (1956). Shock waves on the highway, *Operations Research* pp. 42–51.

Ross, P. (1991). *On-ramp control: Realization principles in: M. Papageorgiou: Concise encyclopedia of traffic & transportation systems*, Pergamon Press, Frankfurt, Germany.

Ruijgers, M. (2005). A ramp metering strategy based on string stability, *Technical report*, Enschede, The Netherlands.

Sadeghhosseini, S. and Benekohal, R. F. (2002). Headway models for low to high volume highway traffic, *TRB 2002 Annual Meeting CD-ROM*, Washington, USA.

Sullivan, R. J. and Troutbeck, R. J. (1994). The use of cowan's m3 headway distribution for modelling urban traffic flow, *Traffic Engineering and Control* pp. 445–450.

Swaroop, R. and Rajagopal, K. R. (1999). Intelligent cruise control systems and traffic flow stability, *Transportation Research* **7C**: 329–352.

Tampère, M. J. (2004). *Human-Kinetic Multiclass Traffic Flow Theory and Modelling. With Application to Advanced Driver Assistance Systems in Congestion*, PhD thesis, Trail Research School, Delft, The Netherlands.

Tolle, J. E. (1976). Vehicular headway distributions: Testing and results, *Transportation Research Record* **567**: 56–64.

Transportation Research Board (2000). *Highway Capacity Manual*, National Research Council, Washington, USA.

TRB (1992). Traffic flow theory: A state-of-the-art-report, *Special report 165*, Transportation Research Board - National Research Council, Washington D.C., USA.

Treiber, M. and Helbing, D. (2002). Reconstructing traffic dynamics, *Cooper@tive Tr@nsport@tion Dyn@mics (www.Traffic Forum.org)* **1**: 3.1 – 3.24.

Van Altena, J. and Veling, I. (2000). Effect of measures on roadsections, *Technical report*, Ministry of Transport, Public Works and Water Management, Transport Research Centre, Rotterdam, The Netherlands.

Van Arem, B. (1986). *A stochastic model of traffic delay of merging traffic (in Dutch)*, PhD thesis, University of Twente, Enschede, The Netherlands.

Van Berkum, E. C. (1999). *Everything under control (in Dutch: alles onder controle)*.

van Lint, H. (2004). *Reliable Travel Time Prediction for Freeways*, PhD thesis, Trail Thesis series, Delft University of Technology, Delft, The Netherlands.

Van Toorenburg, J. A. C. (1986). Praktijkwaarden voor de capaciteit, *Technical report*, Ministry of Transport, Public Works and Water Management, Transport Research Centre, Rotterdam, The Netherlands.

Van Toorenburg, J. A. C. and Elbers, J. A. C. M. (2000). Capaciteitsgedrag in bottlenecks, *Technical report*, Ministry of Transport, Public Works and Water Management, Transport Research Centre, Rotterdam, The Netherlands.

Weber, E. H. (1905). Tastsinn un gemeingefuhl, *in* R. Wagner (ed.), *Handworterbuch der Physiologie*, Leipzig, Germany.

Wiedemann, R. (1974). Simualtion des straßenverkehrsflußes., *Technical report*, Institute for Traffic Engineering, University of Karlsuhe, Karlsruhe, Germany.

Yi, J. and Horowitz, R. (2002). Macroscopic traffic flow stability for adaptive cruise controlled (acc) vehicles, *Proceedings of the 41st IEEE Conference on Decision and Control*, Las Vegas, Nevada, USA.

Yi, J., Lin, H., Alvarez, L. and Horowitz, R. (2003). Stability of macroscopic traffic flow modeling through wavefront expansion, *Transportation Research* **37B**: 661–679.

Yousif, S. and Hunt, J. (1995). Modelling lane utilisation on britisch dual-carriageway roads: effects on lane-changing, *Traffic Eng. Control* **36(12)**: 680–687.

Zhang, M. (1998). Analyses of non-equilibrium traffic flow, *Transportation Research* **32B**: 485 – 498.

Zhang, M., Kim, T., Nie, X., Jin, W., Chu, L. and Recker, W. (2001). Evaluation of on-ramp control algorithms, *Technical report*, Institute of Transport Studies, University of California, Davis, California.

Appendix A

Headway distribution models

Headway distribution models

Recall the definition of headway as the sum of a vehicle's occupancy and the time gap to it's predecessor. Headway distribution models are mesoscopic, because they describe the distribution of the headways of individual vehicles and not of every vehicle separately.

This section starts with an example of how data observations can be measured and then gives an overview of headway distribution, both for 'free' and for 'constrained' traffic. Later some commonly used headway distributions are shown to merging probabilities.

May (1990) collected an extensive data-set on time headway and from that the following general observations were done.

- Individual time headways are rarely less than 0.5 second (1 to 2 percent).
- Individual time headways are rarely over 10 seconds unless the flow rate is below 15 vehicles per minute.
- The time headway mode is always less than the median, which is always less than the mean. However, they tend to converge as the minute flow rate increases toward capacity.
- The mean time headway tracks the 67 cumulative percentile curve for the entire minute flow rate range.
- The ratio of the standard deviation to the mean time headway approaches 1 under free flow conditions but decreases continuously as the minute flow rate increases.

The set they considered was clearly not a set of very light traffic. The remarks above are a first screening of measurements; usually the measurements are matched with probabilistic statistics. Besides probability distribution models, there are also some headway distribution models which are based on vehicle and driver behaviour as driver reaction time, driver acceleration and vehicle speed. In the next sections some headway distribution models are compared. One should remember that not all models were developed under comparable conditions and had different research aims.

Arrivals at free flow

Light traffic flows are almost random; the probability of a vehicle passing a given point during some interval is independent of the passing times of other vehicles. The Poisson distribution is an appropriate distribution to describe the independent arrivals. Let the probability of n vehicles passing a point over a time span t is given by $P(n)$. Over the next small time interval δt the probability of n vehicles arriving in a time $t+\delta t$ is the sum (because the arrivals are random) of the probability of n vehicles arriving in a time t and of $n-1$ vehicles in t and one vehicle in δt, i.e.

$$P_{t+\delta t}(n) = P_t(n-1) * P\{1 \; vehicle \; in \; \delta t\} + P_t(n) * P\{0 \; vehicles \; in \; \delta t\} \quad \text{(A.1)}$$

Because the flow is assumed to be random, the probability of 1 vehicle arriving in a time interval is the same as the expected number of vehicles to arrive $(q\delta t)$. The probability of 0 vehicles arriving is 1 minus the probability that 1 vehicle arrives $(1 - q\delta t)$.

$$P_{t+\delta t}(n) = P_t(n-1)[q\delta t] + P_t(n)[1 - q\delta t] \quad \text{(A.2)}$$

From which follows:

$$P_{t+\delta t}(n) = P_t(n) + q\delta t[P_t(n-1) - P_t(n)] \quad \text{(A.3)}$$

Differentiating to t shows:

$$\frac{\mathrm{d}P_t(n)}{\mathrm{d}t} = q[P_t(n-1) - P_t(n)] \quad \text{(A.4)}$$

The solution of this equation is the Poisson distribution:

$$P_t(n) = \frac{\lambda^n e^{-\lambda}}{n!} \quad \text{(A.5)}$$

with λ the (mean) number of vehicles expected to arrive in time t: $\lambda = q\ t$. So for light traffic, for which may be assumed that arrival rates of vehicles are not affected by other vehicles, the arrival rate can be modelled by the Poisson distribution. Because not the arrival rate, but the distances between drivers form the window of opportunity for merging traffic, in a similar matter the headways of vehicles in a traffic flow are considered.

Headway distribution models for free flow

The distribution of headways can be found in a similar way. Let $P(T > t)$ be the probability that a headway T is larger than some time interval t, then the probability T larger than $t + \delta t$is:

$$P(T > t+\delta t) = \underbrace{P(T > t)}_{probability\ of\ the\ event\ in\ T} * \underbrace{[1 - q\delta t]}_{probability\ of\ no\ event\ in\ \delta t} \tag{A.6}$$

in which q is the flow. This relation may be rewritten to:

$$\frac{\mathrm{d}P(T > t)}{\mathrm{d}t} = -qP(T > t) \tag{A.7}$$

The solution of this equation is:

$$P(T > t) = e^{-\lambda t} \tag{A.8}$$

with λ the (mean) number of vehicles expected to arrive in time t: $\lambda = q\ t$, so λ can be considered as the flow rate.

The cumulative distribution function $F(t)$ of the Negative exponential distribution is:

$$\begin{array}{ll} F(t) = 0 & (\text{t} < 0) \\ F(t) = 1 - e^{-\lambda t} & (\text{t} \geq 0) \end{array} \tag{A.9}$$

The Poisson and the negative exponential distribution are valid only when traffic flows are light. But they form the basis of headway distribution modelling, as can be found in Cowan (1975), Tolle (1976), and Leutzbach (1988).
The distribution of vehicles and headways is relatively easy to model when traffic is light. When traffic gets heavier, vehicles start interacting and the random process of arrival can no longer be assumed.

Cowan (1975) assumed that the vehicle headway (X) consisted of two components; one for 'tracking' (V) and one for 'free' (U). For free traffic the tracking component equals zero and the distribution is equal to a negative exponential distribution. When a tracking component of τ seconds is added to the negative exponential distribution the result is a displaced (or shifted) negative exponential distribution:

$$
\begin{aligned}
&F(t) = 0 && (\mathrm{t}< \tau) \\
&F(t) = 1 - e^{-\gamma(t-\tau)} && (\mathrm{t}\geq \tau)
\end{aligned}
\tag{A.10}
$$

The expected value of the distribution $E(t)$ equals $\tau + 1/\gamma$ and the flow rate $\lambda = \frac{\gamma}{\gamma\tau+1}$. Note that the flow rate λ can be expressed as $1/E(t)$.

Although the shifted negative exponential distribution considers tracking traffic, it still does not fit heavy traffic flows. Two distributions which fit quite well to both free traffic and more dense traffic are Pearson Type III distribution (which include the Erlang distribution) and Log-normal distribution.

One of the generalized mathematical model approaches for the distribution of headways is Pearson type III distribution. The cumulative distribution function of this distribution is:

$$
\begin{aligned}
&F(t) = 0 && (\mathrm{t}< 0) \\
&F(t) = \int_0^t \frac{\lambda}{\Gamma(K)}[\lambda(t-\alpha)]^{K-1}e^{-\lambda(t-\alpha)}\mathrm{d}t && (\mathrm{t}\geq 0)
\end{aligned}
\tag{A.11}
$$

where K and α are parameters that affect the shape and the shift of the distribution of the headways (t). If $K = 1$ and $\alpha = 0$ than the p.d.f. equals the negative exponential distribution. If $K = 2, 3, 4, 5$ and $\alpha = 0$ than de p.d.f. equals the p.d.f. of an Erlang distribution (Leutzbach, 1988). All p.d.f.'s with $\alpha = 0$ and positive K are Gamma distributions.

Another distribution which is fit for a wide range of traffic volumes is the Log-normal Distribution. Tolle (1976) tested this with field data form motorways in Ohio, USA. The Log-normal distribution is:

$$
\begin{aligned}
&F(t) = 0 && (\mathrm{t}< 0) \\
&F(t) = \int_0^t \frac{1}{\sqrt{2\pi}\sigma}e^{-(\mathrm{ln}t-\mu)^2/2\sigma^2}\mathrm{d}t && (\mathrm{t}\geq 0)
\end{aligned}
\tag{A.12}
$$

in which:

$\mu = \frac{\sum_1^n \mathrm{ln}t_i}{n}$ and $\sigma^2 = \frac{\sum_1^n(\mathrm{ln}t_i-\mu)^2}{n-1}$.

Headway distribution models for constrained flow

Cowan (1975) proposed the Bunched exponential distribution to fit data of constrained flows. This distribution improved the accuracy in the prediction of small arrival headways in the negative exponential distribution and the displaced negative exponential distribution, but retains the simplicity of those models. The model assumes that a portion of vehicles θ, are tracking (following) behind other vehicles at a headway τ. These vehicles are called bunched and the rest of the vehicles is assumed to be travelling freely at a headway greater than τ. The Bunched negative exponential distribution function is:

$$\begin{array}{ll} F(t) = 0 & (\mathrm{t}< \tau) \\ F(t) = 1 - (1-\theta)e^{-\gamma(t-\tau)} & (\mathrm{t}\geq \tau) \end{array} \tag{A.13}$$

Akcelik and Chung (1994) described the model in detail and gave the results of its calibration using real life data for single lane traffic streams and simulation data for multi-lane traffic streams. Sullivan and Troutbeck (1994) made a comparison of the Bunched negative exponential distribution and the Double displaced Negative Exponential Distribution developed by Griffiths and Hunt (1991). It is proved that both models are very accurate, but the Bunched exponential model is much simpler. The Double displaced negative exponential distribution is developed by Griffiths and Hunt (1991):

$$\begin{array}{ll} F(t) = 0 & (\mathrm{t}< \tau) \\ F(t) = \int_0^t \phi\lambda_1 e^{\lambda_1(t-\tau)} + (1-\phi)\lambda_2 e^{\lambda_2(t-\tau)}\mathrm{d}t & (\mathrm{t}\geq \tau) \end{array} \tag{A.14}$$

The parameter $\phi(0 < \phi < 0.5)$ is a weighting factor, τ is the displacement factor and λ_1, λ_2 are constants which depend on the traffic flow. Sullivan and Troutbeck (1994) indicated that the Double displaced negative exponential distribution can model smaller headways more accurately than the Bunched negative exponential model.

Besides single distribution models Composite distribution models exists. They are developed under the assumption that vehicles are either travelling free and are not influenced by the vehicle in front of them or they are following the previous vehicle. A popular composite distribution ((Ovuworie et al., 1980), and (Koshi et al., 1992)) is the Semi-exponential distribution:

$$\begin{array}{ll} F(t) = 0 & (\mathrm{t}< 0) \\ F(t) = (1-\phi)F_1(t) + \phi F_2(t) & (\mathrm{t}\geq 0) \end{array} \tag{A.15}$$

where ϕ is the proportion of the followers, $F_1(t)$ is the distribution for leaders, usually a negative exponential distribution, and $F_2(t)$ is the distribution for followers. All of the models mentioned above can fit traffic situations well, but the derivation of unknown parameters is complicated. Method of moments, maximum likelihood techniques and hybrid methods can be tried. Griffiths and Hunt (1991) and Sullivan and Troutbeck (1994) gave some descriptions of the methods to get unknown parameters.

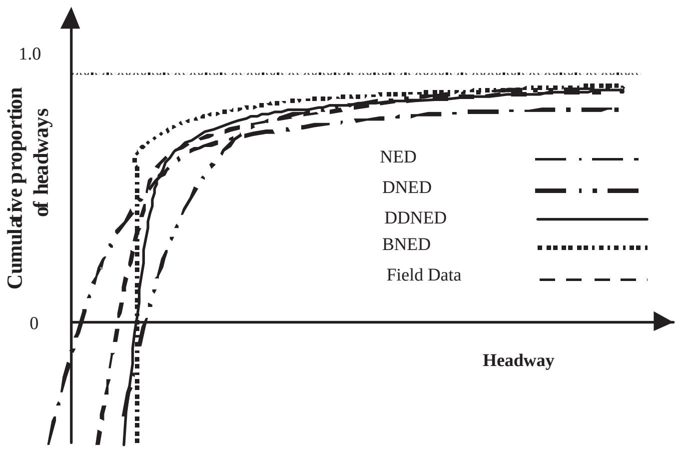

Figure A.1: Headway models: NED (negative exponential distribution), DNED (displaced negative exponential distribution), DDNED (double -negative exponential distribution)

Appendix B

Overview of the simulation network

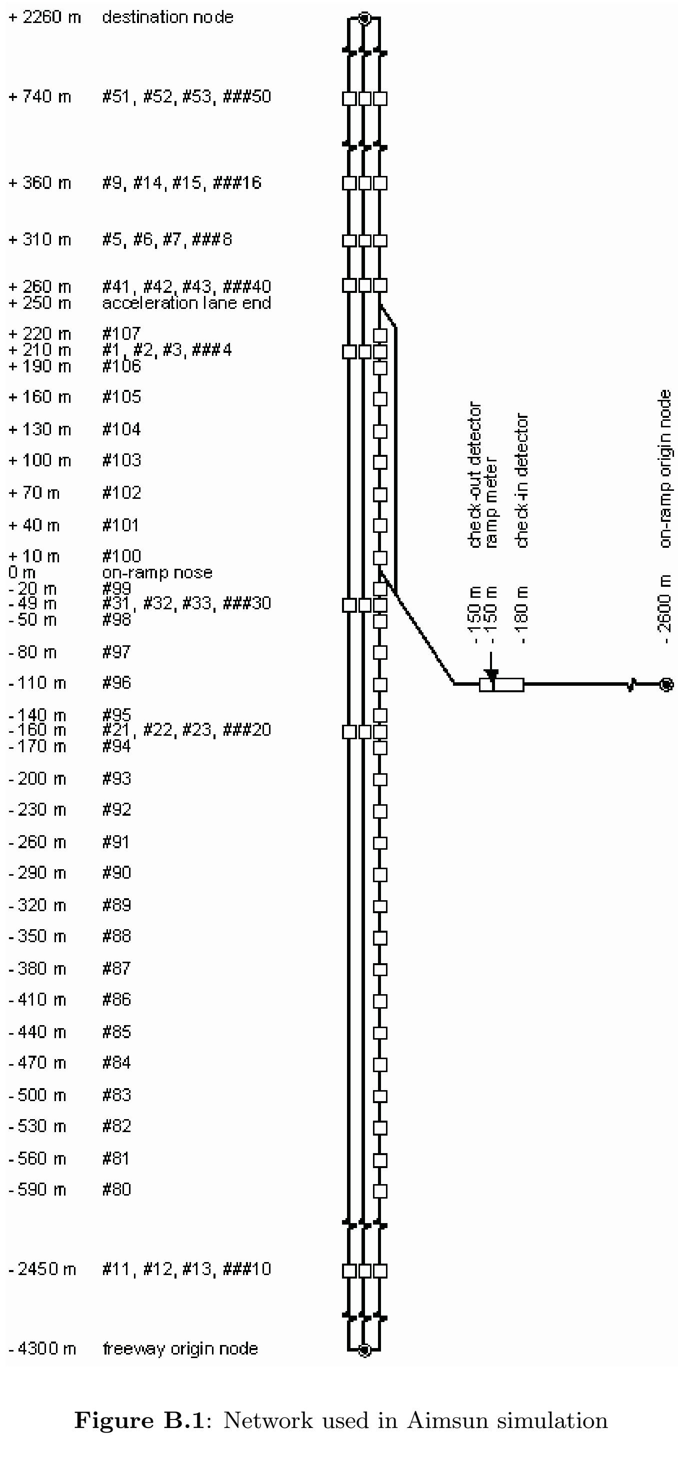

Figure B.1: Network used in Aimsun simulation

Samenvatting: Verkeersstabiliteit en Toerit Doseer Installaties

Dit proefschrift beschrijft hoe stabiliteit van verkeer kan worden gedefiniëerd en gemeten teneinde het ontstaan van files beter te voorspellen. Er is een stabiliteitsindicator ontwikkeld welke is toegepast om een bestaande dynamisch verkeersmanagement (DVM) maatregel, de Toerit Doseer Installatie (TDI), te verbeteren.

Aanleiding

Aanleiding voor dit onderzoek is een studie naar capaciteitsverdelingen, uitgevoerd door Van Toorenburg en Elbers (2000). Daarin wordt op een tweetal wegvakken met een gelijk aantal stroken, vergelijkbaar hellingspercentage en percentage vrachtverkeer en (onder) gelijke weersomstandigheden, desondanks een significant verschil in capaciteitsverdeling gevonden tussen beide wegvakken.

Uit het analyseren van videomateriaal is de invloed van clusters verkeer op de verkeersafwikkeling in een bottleneck groot gebleken.

Doelen

Drie doelen zijn geformuleerd:

- Het hoofddoel is een betere voorspelling te kunnen geven van het ontstaan van file, waarbij gebruik gemaakt wordt van bestaande meetsystemen.

- Het tweede doel is bestaande theorieën over het ontstaan van file om te zetten in operationele indicatoren.

- Het derde doel is het toepassen van de beste indicator om een bestaand TDI-algoritme te verbeteren.

Stabiliteit

Hoewel de term stabiliteit in relatie tot verkeer regelmatig wordt gebruikt, wordt in de toonaangevende werken die de verkeerskundig ingenieur tot zijn beschikking heeft, de Highway Capacity Manual (Transportation Research Board, 2000) en de Capaciteitswaarden Infrastructuur Autosnelwegen (AVV, 1999a), geen definitie gegeven van deze term. De algemene voorstelling van stabiel verkeer is: verkeer dat beweegt met kleine snelheidsverschillen tussen de individuele voertuigen en homogeen verdeeld is tussen en op de rijstroken. Hierin schuilt de veronderstelling dat homogeen verkeer stabiel is.

Het verkeer bestaat uit een groot aantal, veelal niet-lineaire en complexe relaties tussen bestuurders en hun omgeving. Deze omgeving bestaat o.a. uit andere weggebruikers, de conditie van de weg, het voertuig, de (statische) verkeersregels, het weer en DVM. Veel DVM maatregelen hebben als doel het homogeniseren van verschillende aspecten van de verkeersstroom (TDI voor dichtheid, Trajectcontrole voor snelheid). Dat dit niet altijd tot het gewenste resultaat hoeft te leiden blijkt uit een in dit onderzoek uitgevoerde modelstudie waarbij invoegend verkeer op een autosnelweg wordt geconfronteerd met verschillende vormen van gehomogeniseerd verkeer op deze weg. Verkeer is hierbij gehomogeniseerd wat betreft plaats, strook en snelheid. De *niet* homogene stroom blijkt meer geschikte hiaten te bezitten voor invoegend verkeer, wat leidt tot een hogere capaciteit van de bottleneck.
Stabiliteit van een verkeersstroom wordt gedefinieerd als de mogelijkheid om een verstoring op te vangen.

In de literatuur worden in het algemeen drie soorten stabiliteit onderscheiden:

- lokale stabiliteit, welke afhankelijk is van het voertuigvolggedrag tussen twee opeenvolgende voertuigen;
- ketenstabiliteit, op basis van het voertuigvolggedrag van een reeks voertuigen; en
- verkeersstroomstabiliteit, waarbij niet het individuele voertuigvolggedrag wordt bekeken, maar verstoringen op macroscopische wijze worden beschouwd.

Voor het meten van de bovengenoemde definitie van stabiliteit ('opvangen van een verstoring') is een drietal indicatoren ontwikkeld:

1. op basis van verkeersbetrouwbaarheid (Ferrari, 1989). Verkeer wordt betrouwbaar genoemd als geen snelheidsafname onder een bepaald niveau voorkomt (verkeersstroomstabiliteit).
2. op basis van een theorie over het synchroniseren van verkeersstromen voordat file ontstaat (Kerner and Rehborn, 1996). Hierbij wordt de kans op het ontstaan van een file berekend op basis van veranderingen van snelheid en intensiteit in de tijd (verkeersstroomstabiliteit).

3. op basis van ketenstabiliteit; van een peloton voertuigen wordt middels microsimulatie berekend of het robuust genoeg is om een bepaalde verstoring op te vangen.

Uit onderzoek op data van de snelweg A2, nabij Vinkeveen, blijkt dat de stabiliteitsindicator op basis van ketenstabiliteit (indicator 3) duidelijk beter in staat is om files te voorspellen dan de alternatieven (correlatie van 0.7 versus 0.5 met het tijdstip van het ontstaan van de file).

Data verzameling

Een van de doelen van dit onderzoek is het verbeteren van de voorspelling van het ontstaan van file op basis van meetgegevens die reeds verzameld worden. De intentie was niet op zoek te gaan naar nieuwe meetsystemen. Het ontwikkelen van een drietal stabiliteitsindicatoren en het testen van hun voorspellend vermogen gaat dan ook uit van bestaande meetsystemen. De beste van de drie stabiliteitsindicatoren is gebruikt bij het verbeteren van het bekende toeritdoseeralgoritme Alinea.

Van voertuigtrajectoriën (plaats-tijd diagram van voertuigen) kunnen microscopische voertuigkarakteristieken als plaats, snelheid, volgafstand en volgtijd worden afgeleid. Het gebruik van voertuigtrajectoriën zou het berekenen van de stabiliteit kunnen vereenvoudigen vanwege beschikbaarheid van data op ieder tijdstip en locatie. Trajectoriën kunnen echter alleen met behulp van (een nog) zeer kostbare meetmethode over grote afstanden worden gemeten (bijvoorbeeld vanuit helikopters of andere zwevende platforms). Dit onderzoek gaat uit van bestaande meetmethoden en data afkomstig van lusdetectoren op de autosnelweg (dubbel per strook, elke 500 m). Deze zijn op het Nederlandse snelwegennet in groten getale aanwezig. Individuele voertuiggegevens, verzameld met lusdetectoren zijn slechts beschikbaar uit een beperkt aantal studiegebieden, maar zullen in de nabije toekomst vanaf grote delen van het Nederlandse snelwegennet beschikbaar komen.

Resultaten

Zowel uit de ontwikkeling van de stabiliteitsindicatoren als de inpassing van de ketenstabiliteitindicator in het Alinea TDI-algoritme blijkt dat microscopische verkeersprocessen een grote rol spelen bij het invoegproces.

Ook het capaciteitsonderzoek naar meer of minder homogene verkeersstromen bij toeritten onderschrijft dit. Toename van de homogeniteit van de verkeersstroom op de hoofdrijbaan resulteert in een geringere kans dat een voertuig zonder problemen kan invoegen.

Verstoringen in de verkeersstroom zijn voor een groot deel afhankelijk van variaties in volgtijden en reactiesnelheden van bestuurders. Korte volgtijden en lange reactietijden, waardoor een volgend voertuig stroomafwaarts van waar zijn voorganger ging remmen gaat remmen, resulteren in schokgolven van steeds heftiger remmende voertuigen.

De stabiliteitsindicator gebaseerd op ketenstabiliteit geeft een duidelijk betere indicatie voor het ontstaan van file dan de andere onderzochte indicatoren.

Zoals gezegd is de beste stabiliteitsindicator gebruikt om een TDI-algoritme te ontwikkelen. Het bestaande Alinea algoritme is hiertoe aangepast, middels een stabiliteitscorrectie: als de stabiliteit duidelijk hoog of juist laag is wordt de doseerfrequentie die Alinea geeft naar boven of naar beneden bijgesteld. Dit levert een extra reductie van de verliestijd op van 15 %, waarbij zowel op de hoofdrijbaan als op de toerit de rijtijden afnemen. De performance van het Alinea TDI-algoritme kan dus met zo'n 15% worden verbeterd wanneer de stabiliteitsindicator wordt geïntegreerd.

Conclusies

- Hoewel er definities van stabiliteit bestaan is er geen eenduidig gebruik van de term. Het opnemen van de term in toonaangevende literatuur kan hiervoor een oplossing bieden.
- Met de ontwikkelde verkeersstabiliteitsindicator op basis van ketenstabiliteit kan het tijdstip van het ontstaan van file beter worden voorspeld.
- De verkeersstabiliteitsindicator kan een bijdrage leveren aan een effectiever gebruik van TDI-maatregelen.
- Het homogeniseren van verkeer nabij toeritten resulteert niet in een hogere capaciteit. De resultaten van dit onderzoek geven zelfs aanleiding te veronderstellen dat deze wordt gereduceerd.

Nomenclature

	Vehicle characteristics
$\hat{l}$	vehicle length in meters
$\hat{d}$	vehicle's distance to it's predecessor in meters
$\hat{s}$	space from the vehicles front to it's predecessors front in meters
$\hat{o}$	vehicle's occupancy on a cross section
$\hat{g}$	gap between a vehicle and it's predecessor in seconds
$\hat{h}$	headway from a vehicle's front to it's predecessors front in seconds
$\hat{u}$	vehicle's speed in meters per second or kilometres per hour
$\hat{a}$	vehicle's acceleration, derivative of speed in time
$\hat{j}$	vehicle's jerk, derivative of acceleration in time
$\hat{x}_i$	one of the characteristics above off the ith vehicle
$\hat{x}^{\mathbf{s}}$	the characteristics above under safe circumstances
$\hat{x}^{\mathbf{j}}$	the characteristics above under jammed circumstances (no probability of collision)
$[\hat{x}_i(t), \hat{y}_i(t)]$	the coordinates of the position of the ith vehicle at time t

Traffic characteristics

$q(x,t)$	flow on position x at time t
q	flow
q_m	maximum flow
q_c	critical flow
q_f	free flow
$k(x,t)$	density on position x at time t
k	density
k_m	maximum density
k_c	critical density
k_f	density at free traffic
k_0	density at zero flow (0 or maximum density)
$o(x,t)$	occupancy on position x at time t
o_cr	critical occupancy
$u(x,t)$	speed on position x at time t
u	speed
u_l	local speed
$u_{l,m}$	maximum local speed
$u_{l,c}$	critical local speed
$u_{l,f}$	local speed at free traffic
u_m	instantaneous speed
$u_{m,m}$	maximum instantaneous speed
$u_{m,c}$	critical instantaneous speed
$u_{m,f}$	instantaneous speed at free traffic

Other notations

t	time
T	period of time
x	position
X	distance (interval of position)
$\tilde{x}$	estimate of x
x_w	x which is wished for
$C_{x,t}$	capacity at location x and time t
$TC_{r,d,i}$	travel costs at road r, over distance d and vehicle i
$D\hat{u}_{i,j}$	speed difference between vehicle i and j
$\vec{u}_{fr}$	speed of information at free traffic
$\vec{u}_c$	speed of information at critical traffic
$\vec{u}_{co}$	speed of information at congested traffic
s	second
m	meter
h	hour

About the author

Bart Elbers was born in 1974 in Beuningen (Gld.), The Netherlands. He finished his secondary school at the Pax Christi College in Druten in 1992. In 1998 he received a master's degree in mathematics from the University of Utrecht. His master thesis was on collections of conflict points in globe- and hyperbolic geometry.

In the years he spent in Utrecht he became interested in politics and traffic. The second interest became his job when he started to work as a traffic consultant at Grontmij, Traffic & Infrastructure in De Bilt, The Netherlands. After two years he started a part-time job at Transpute traffic consulting and started his Ph.D. at the University of Twente.

From 2000 to 2004 Bart was affiliated as a research assistant with the Centre for Transport Studies at the University of Twente under the leadership of Professor Eric van Berkum. In this period he did research on traffic flow stability which resulted in this Ph.D. dissertation.

After the summer of 2004 he went back to his consulting profession at Transpute and finished his research in his spare time.

At this moment Bart's main research at Transpute consists of predicting traffic delay in networks due to road maintenance. In his spare time he is putting a lot of time in his second hobby: local politics.