Traffic Flow Theory and Simulation

vk4821

Dr. Ir. Serge P. Hoogendoorn
Transportation and Traffic Engineering Section
Faculty of Civil Engineering and Geosciences
Delft University of Technology

Contents

0.1 Microscopic and macroscopic characteristics and tools . . . ix
0.2 Aims and scope of the course . . . x
0.3 Contents of the reader . . . x

I Analytical Framework **1**

1 Demand-supply analysis **5**
1.1 Analytical framework . . . 5
1.2 Initial analysis . . . 6
1.3 Feedback processes . . . 7
1.4 Demand-supply analysis example . . . 8
1.4.1 Base conditions . . . 8
1.4.2 Initial analysis . . . 9
1.4.3 Feedback processes . . . 9
1.4.4 Ramp control . . . 11

II Traffic Flow Characteristics **13**

2 Microscopic and macroscopic traffic flow variables **15**
2.1 Introduction . . . 15
2.1.1 Trajectory of a single vehicle . . . 16
2.1.2 Trajectories of multiple vehicles . . . 18
2.1.3 Applications of trajectories in traffic problem solving . . . 18
2.1.4 Application of trajectories to scheduling problems . . . 18
2.1.5 Mathematical description of trajectories and vehicle kinematics . . . 20
2.2 Time headways . . . 21
2.2.1 Time headways . . . 22
2.2.2 Distance headways . . . 23
2.3 Intensity, density and mean speed . . . 23
2.3.1 Intensity . . . 23
2.3.2 Density . . . 24
2.3.3 Mean speed . . . 25
2.4 Homogeneous and stationary flow conditions . . . 26
2.4.1 Determination of periods with stationary intensity . . . 26
2.5 Relation between local and instantaneous characteristics . . . 27
2.5.1 Relation between instantaneous and local speed distribution . . . 27
2.5.2 Local and instantaneous mean speeds part 1 . . . 28
2.5.3 Local and instantaneous mean speeds part 2 . . . 31
2.5.4 Relation between flow, density and speed revisited . . . 31
2.6 Cumulative vehicle plots and their applications . . . 32
2.6.1 Relation between cumulative flow, intensity and density . . . 33

2.6.2 Trip times and the cumulative flow function . . . 34
2.6.3 Delay times and cumulative flow functions . . . 34
2.6.4 Derivation of conservation of vehicle equation using cumulative flow functions . . . 39
2.7 Definition of q, k and u for a time-space region . . . 39
2.7.1 Calulating the generalised q and k . . . 41
2.8 Measuring methods . . . 43
2.8.1 Passage times, time headways and intensity . . . 43
2.8.2 Distance headways and density . . . 45
2.8.3 Individual speeds, instantaneous and local mean speed . . . 45
2.9 Occupancy rate . . . 47
2.10 Moving observer method . . . 47
2.10.1 Speed distribution as seen by a moving observer . . . 51
2.10.2 Fraction of drivers that is speeding . . . 51
2.11 Summary of main definitions and terminology . . . 52

3 Microscopic flow characteristics 55
3.1 Arrival processes . . . 55
3.1.1 Formulae for parameters, probability terms, and parameter estimation . . . 58
3.1.2 Applications . . . 58
3.1.3 Choice of the appropriate model by using statistical testing . . . 59
3.2 Headway distributions . . . 59
3.2.1 Distribution of headways and the Poisson arrival process . . . 59
3.2.2 Use of the headway distribution for analysis of crossing a street . . . 60
3.2.3 Use of headway distribution to calculate the waiting time or delay . . . 63
3.2.4 Real data and the exponential distribution . . . 64
3.2.5 Alternatives for the exponential distribution . . . 66
3.2.6 Composite headway models: distributions with followers and free drivers . . . 69
3.3 Distance headway distributions . . . 74
3.4 Individual vehicle speeds . . . 75
3.5 Determination of number of overtakings . . . 75
3.6 Dependence of variates (headways, speeds) . . . 77

4 Fundamental diagrams 81
4.1 Introduction . . . 81
4.1.1 Special points of the fundamental diagram . . . 83
4.1.2 Importance of the fundamental diagram . . . 83
4.1.3 Factors influencing the diagram . . . 84
4.2 Models of the fundamental diagram . . . 85
4.2.1 Model specifications . . . 85
4.2.2 Concept of discontinuous diagram . . . 88
4.2.3 Wu's fundamental diagram with capacity drop . . . 89
4.2.4 Diagram for roadway and lane . . . 90
4.2.5 Fundamental diagram based on a car-following model . . . 91
4.3 Studies of the fundamental diagram . . . 94
4.3.1 General points . . . 94
4.3.2 Influence and importance of data collection location . . . 94
4.3.3 Estimation of capacity using the fundamental diagram . . . 96
4.4 Capacity state on a motorway and the effect of rain . . . 97

5 Dynamic properties of traffic flow **99**
5.1 Equilibrium traffic state . . . 99
5.2 Hysteresis and transient states . . . 99
5.3 Metastable and unstable states . . . 102

III Analytical techniques **105**

6 Capacity and level-of-service analysis **107**
6.1 Capacities and level-of-service . . . 107
6.2 Capacity and driver behaviour . . . 109
6.3 Multilane facilities . . . 110
6.3.1 Capacity analysis under ideal conditions . . . 110
6.3.2 Capacity analysis under non-indeal conditions . . . 110
6.4 Ramps . . . 111
6.5 Weaving sections . . . 114
6.5.1 HCM-1965 approach . . . 114
6.5.2 HCM 1985 approach . . . 115
6.6 Dutch approach to motorway capacity analysis . . . 118
6.7 Stochastic nature of motorway capacity . . . 119
6.8 Capacity drop . . . 120
6.9 Capacity estimation approaches . . . 120
6.9.1 Product-limit method to capacity estimation . . . 121
6.9.2 Example application of the PLM method . . . 124
6.9.3 Other applications of the PLM method . . . 125
6.10 Summary . . . 126

7 Queuing analysis **127**
7.1 Deterministic queuing theory . . . 127
7.2 Determination of delay with a queueing model . . . 128
7.2.1 Computations with the queuing model . . . 131
7.3 QUAST-method to take account of variability of congestion . . . 132
7.3.1 QUAST-model . . . 133
7.4 QUAST application results . . . 135
7.4.1 Effect of factors on required capacity . . . 135
7.4.2 Distribution of congestion characteristics . . . 135
7.5 Final remarks . . . 136

8 Shock wave analysis **139**
8.1 Kinematic waves . . . 139
8.2 Shock wave speeds . . . 141
8.3 Temporary blockade of a roadway . . . 142
8.4 Shock wave analysis for non-linear intensity-density curves . . . 143
8.4.1 Temporary blockade revisited . . . 143
8.5 Moving bottlenecks . . . 145
8.5.1 Approach 1 . . . 145
8.5.2 Approach 2 . . . 147
8.6 Shock wave classification . . . 147

9 Macroscopic traffic flow models **149**
9.1 General traffic flow modelling issues . . . 150
9.2 Kinematic wave model and applications . . . 151
9.2.1 Analytical solutions using method of characteristics . . . 151
9.2.2 Application of kinematic wave model for bicycle flows at signalised intersection . . . 158
9.2.3 Implications for fundamental diagram . . . 163
9.3 Numerical solutions to the kinematic wave model . . . 164
9.3.1 Approach to numerical approximation . . . 164
9.3.2 Fluxes at cell-interfaces . . . 165
9.3.3 Simple explicit schemes . . . 165
9.3.4 Flux-splitting schemes . . . 166
9.3.5 Godunov schemes . . . 167
9.4 Higher-order models . . . 170
9.4.1 Derivation of the model of Payne . . . 171
9.4.2 Mathematical properties of the Payne-type models . . . 173
9.4.3 Numerical simulation and high-order models . . . 174
9.5 Alternative modelling approaches and generalisations . . . 174

10 Human factors **175**
10.1 The driving task . . . 175
10.2 Sharing attentional resources . . . 177
10.3 Driver objectives . . . 177
10.4 Modelling driving tasks . . . 178
10.4.1 Car-driver-roadway system diagram . . . 178
10.4.2 Optimal control model . . . 178
10.5 Discrete driver performance . . . 182
10.5.1 Perception-response times . . . 182
10.5.2 Control movement times . . . 183
10.5.3 Response times to traffic control devices . . . 184
10.5.4 Response to other vehicle dynamics . . . 184
10.6 Continuous driver performance . . . 184

11 Longitudinal driving taks models **187**
11.1 Model classification . . . 187
11.2 Safe-distance models . . . 188
11.3 Stimulus-response models . . . 189
11.3.1 Model stability . . . 190
11.4 Psycho-spacing models / action point models . . . 197
11.5 Optimal control models . . . 199
11.6 Fuzzy Logic Models . . . 202
11.7 Cellular automata models . . . 203
11.7.1 Deterministic CA model . . . 203

12 Gap-acceptance theory and models **205**
12.1 Gap acceptance at overtaking . . . 205
12.2 Model with a fixed critical gap for each driver . . . 206
12.3 Example to illustrate the concepts . . . 207
12.4 Overtaking manoeuvre in time-space plane . . . 208
12.5 Studies into overtaking behaviour . . . 209
12.5.1 Calculated required space . . . 210
12.5.2 Studies with vehicle simulators . . . 211
12.6 Estimation of critical gap distributions . . . 211

12.7 Practical results . . . 212
12.7.1 Case 1. Overtaking of long trucks . . . 212
12.7.2 Case 2. Left turn movement at intersection, taking into account many factors . . . 212
12.8 Conclusions and main points . . . 212

IV Traffic Simulation 215

13 Simulation study guidelines 219
13.1 Overview of simulation study steps . . . 220
13.2 Summary of problem definition . . . 221
13.3 Determining objectives of simulation study . . . 221
13.4 Model choice . . . 222
13.5 Setting-up a simulation study . . . 223
13.6 Application of the simulation model . . . 224

14 Microscopic simulation models 227
14.1 General aspects of simulation models . . . 227
14.1.1 General description of simulation . . . 227
14.2 Applications . . . 227
14.2.1 Parts of the model . . . 228
14.2.2 Development of a simulation model . . . 230
14.2.3 Use of a model . . . 231
14.3 Description of simulation model FOSIM . . . 231
14.3.1 General . . . 231
14.3.2 The vehicle process . . . 232
14.3.3 The car-following model . . . 233
14.3.4 Lane change model . . . 234
14.3.5 Calibration and validation . . . 235
14.4 Application of FOSIM . . . 236
14.4.1 Truck overtaking ban . . . 236
14.4.2 Truck ban . . . 238

15 Macroscopic simulation 241
15.1 Traffic processes in METANET . . . 241
15.1.1 Normal motorway links . . . 241
15.1.2 Origin links and destination links . . . 242
15.1.3 Discontinuities . . . 242
15.1.4 Modelling of the Network Nodes . . . 243
15.2 Preparing a METANET simulation study . . . 243
15.2.1 Determine and acquire required input parameters and variables . . . 243

A Green's theorem and its applications 245
A.1 Alternative analytical solution scheme to the kinematic wave model . . . 246

B Mathematical and numerical analysis of higher-order models 251
B.1 Riemann variables . . . 251
B.2 Propagation of disturbances . . . 253
B.3 Instabilities and jam formation . . . 254
B.4 Numerical solutions to the Payne model . . . 256
B.4.1 Payne model in conservative form . . . 256
B.4.2 Flux-splitting numerical solution approach . . . 257

B.4.3 Quasi-linear formulation of Payne model 258

C Gas-kinetic modelling **261**
C.1 Prigogine and Herman model 261
C.2 Paveri-Fontana model 263
C.3 Applications of gas-kinetic models 264

D Higher-order model extensions and generalisations **267**
D.1 Multiclass generalisation of PF-model 267
D.2 Multiclass higher-order model 268

E Pedestrian flow model NOMAD **271**

Introduction

Traffic flow theory pertains to *The Knowledge of Fundamental Traffic Flow Characteristics and Associated Analytical Techniques.* Understanding the basic traffic flow characteristics as well as these techniques is an essential requirement for planning, design, and operation of any transportation system. For example, planners assess traffic and environmental impacts of proposed modifications - either in the infrastructure or in terms of traffic management measures - of the transportation system, which can only be accomplished through demand-supply analysis, which in turn requires an understanding of the flow characteristics as well as their interrelations with the candidate modifications. A designer will for instance need to determine the number of lanes of a particular roadway segment in the network, and will hence need to carefully evaluate the trade-off between traffic flow levels and levels of service. Operators may need to identify locations and causes of problems in the existing infrastructure design and generate operational improvement plans as well as predicting the effect of the latter.

Some important examples of traffic flow characteristics:

- Trajectories of cars, pedestrians, bicycles, etc.
- Time and distance headways, vehicle speeds
- Traffic volumes, densities, time-mean and space-mean speeds

Amongst the analytical techniques are:

- Supply - demand modeling[1]
- Capacity and level-of-service analysis
- Shock wave analysis
- Macroscopic traffic flow modeling
- Queuing analysis
- Microscopic simulation models

0.1 Microscopic and macroscopic characteristics and tools

A key distinction that is made in the study of traffic systems is the distinction between microscopic and macroscopic characteristics and (mathematical and simulation) models. Microscopic characteristics pertain to the individual driver-vehicle unit (DVU). Examples of microscopic characteristics are time headways, individual speeds, and distance headways. Similarly, microscopic models describe the behavior of individual vehicles, for instance in relation to the infrastructure and other vehicles in the flow. That is, they predict the behavior in terms of

[1] As will be argued in the ensuing, nearly all problems relating to the traffic system can be described within the context of a demand-supply framework. Thus, the demand-supply framework is very generic and will serve as the backbone for this reader.

microscopic characteristics. Typical analytical techniques that are based on microscopic traffic data and theory are headway distribution modelling, and microscopic simulation models. On the contrary, macroscopic characteristics pertain to the properties of the traffic flow as a whole (for instance at a cross-section, or at a time instant). Examples of macroscopic characteristics and flow, time mean speed, density, and space-mean speed. Macroscopic models describe traffic flow in terms of macroscopic characteristics of the flow. Macroscopic analytical techniques involve demand – supply analysis, shock wave theory, capacity analysis, and macroscopic traffic flow modelling.

At this point, let us clearly state that a distinction is made between mathematical models and simulation models. A mathematical (or theoretical) model consists of the mathematical equations describing the systems behavior, such as the partial differential equations describing the conservation of vehicles. A simulation model refers to the computer implementation of a mathematical model, as well as all additional simulation model features.

0.2 Aims and scope of the course

Gaining knowledge of fundamental traffic flow characteristics. Developing the ability to analyze traffic operations, focusing of car-based traffic. Gaining knowledge of techniques for traffic analysis and the ability to select and apply the appropriate technique for the problem at hand. Considered techniques include data analysis and statistics, traffic flow models, and macroscopic and microscopic traffic simulation. The course consist of recitations, exercises and two assignments. The first assignment shows how the microscopic simulation model FOSIM is used in a simulation study. The second assignment will either involve working with empirical data.

0.3 Contents of the reader

The reader is roughly divided into four parts. The first part presents the demand-supply framework. This framework is generic for analyzing traffic systems, either using a microscopic or a macroscopic representation of traffic flow. It is shown how basically any problem in traffic analysis can be described within this framework. The second part of the reader describes microscopic and macroscopic characteristics of traffic flow. The third part discusses several analytical techniques that use these empirical findings in order to answer all kind of research questions. Also for the analytical tools, distinction is made between microscopic and macroscopic tools and models: macroscopic models represent traffic flow in an aggregate manner, i.e. as a continuum. Although the representation of traffic may be macroscopic, the rules predicting the dynamics of the traffic flow may be based on driver behavior. Microscopic models represent all vehicles in the traffic flow individually. This does however not necessarily mean that the behavior of the vehicles is based on microscopic rules. Several models describe the behavior of individual vehicles as a function of macroscopic characteristics, such as the traffic density. The last part of the reader discusses simulation models and how to properly conduct a simulation study.

Part I

Analytical Framework

This introduction describes an analytical process that consists of predicting the output or performance of a transportation system as a function of specified inputs. Such an analytical process can take the form of a very simple equation relating the average number of vehicles on a roadway to the mean speed of these vehicles, or can take the form of a complex simulation model. It is clear that for a transportation analyst, knowledge of the system at hand - i.e. the flow characteristics and the analytical techniques - is essential to correctly predict its performance.

Before providing an in-depth description of the analytical framework, let us at this point refer to the chapter IV of this reader, which handles the application of (simulation) models in the analysis of traffic systems in general, and in particular to the assessment of modifications.

Figure 1.1 (page 5) depicts a flowchart from [36], illustrating the analytical process. The flowchart emphasizes the importance of knowing fundamental flow characteristics and basic analytical techniques. Although the process can be represented in many ways, the most common approach is to predict the performance of the traffic system as a function of traffic demand, transport supply, traffic control, and environment information. The predicted performance may include both performance measures relating to the user (travel times, comfort levels), to the system (throughput), or to society (pollution, noise).

The traffic demand is generally specified in terms of demand flow rates, or time headway distributions for the location and time period selected for analysis. The transport supply features are generally converted in capacity values or minimum time headway values. The resulting operations include predictions of speeds, operating speeds, distance headway distributions, and/or density levels. The control elements must be inserted into the supply – demand process and are used to modify predicted performance. Without an understanding of the flow characteristics, these tasks cannot be performed.

The applicability of different analytical techniques is determined by the complexity of the situation and the problem at hand. In some cases, we can tackle the problems using relatively simple models, in other situations, more involved methods are required. In this respect, the analyst should always make a prudent trade-off between research cost and benefits, given the research aims and constraints. Without knowledge of traffic analytical techniques, and the ability to select the most appropriate microscopic or macroscopic method for the problem at hand, an analyst cannot solve the problems in the traffic system adequately.

Chapter 1

Demand-supply analysis

In principle, all analytical techniques of traffic systems are structured in a demand-supply framework. This framework can be at the microscopic level of analysis, in which individual traffic units are studied, or at the macroscopic level of analysis, in which attention is given to groups of traffic units in aggregate form. These demand-supply analytical techniques vary and the analyst must not only learn these various techniques but must develop skills in selecting the most appropriate technique for the problem at hand. Several analytical techniques are described in this reader and include capacity analysis, shock wave analysis, traffic flow models, queuing analysis, and simulation. In part 2 of these notes, only macroscopic characteristics and techniques are considered in detail; part 3 considers microscopic analysis.

This introduction chapter is devoted to presenting an analytical framework for supply-demand analysis. The major elements of the framework are described and their interactions are identified and discussed briefly[1].

1.1 Analytical framework

The analytical framework consist of two processes: an initial process and one or more feedback processes. Fig. 1.1 schematizes the analytical framework. The input of the analytical process is used to predict the performance of the traffic system. Once the initial performance is predicted, the input is modified as a feedback process. These modifications can be user-selected or part of an optimization and/or demand-related process. The remaining sections of this chapter discuss both the initial analysis and the feedback processes.

[1] This chapter is an abbreviated version of chapter 8 of [36].

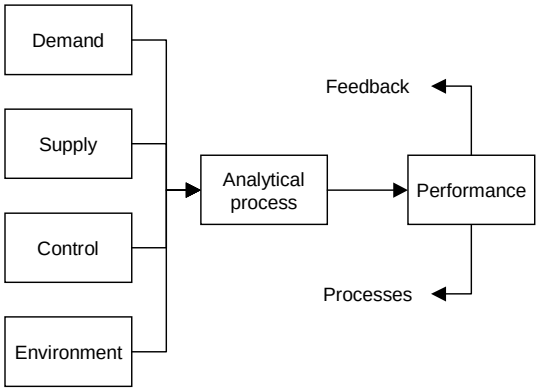

Figure 1.1: An analytical framework for demand-supply analysis of traffic systems

1.2 Initial analysis

The *demand input* represent the number of units that would like to be served. Clearly, the demand is always larger than the number of units that can be served. In microscopic analysis, the demand input is expressed as the arrival times (or equivalently, the headways) of the individual units, while in macroscopic analysis the demand input is expressed as an arrival rate (unit per time interval). It is important to note that in oversaturated conditions, i.e. when the demand exceeds the capacity, more multiple time periods need to be considered. In this way, excess demands in earlier time periods can be transferred and served in later, undersaturated time periods. The time frame for the demand can be for a current point in time, for some future year, or simply a demand modification due to supply or control changes (latent demand).

The *supply input* represents the maximum number of units that can be served. In microscopic analysis, the supply input is expressed as the minimum time headway between units to be served, while in macroscopic analysis the supply input is expressed as the maximum flow rate in unit per time interval. This maximum flow rate is often referred to as the capacity and generally is assumed to remain constant over time periods or traffic states. However, as we have seen, capacity can change between time periods or traffic states due to changes in the demand patterns, vehicle mix, driver characteristics, control states, and / or environment. Often, the analyst begins with the actual physical dimensions of the traffic facility and must translate these features into capacity values, which can be a rather involved process.

The *control input* consists of a set of rules as to how the units interact with each other and with the system. The control can be constant over time, such as car-following rules, rules of service (e.g. FIFO), traffic restrictions, vehicle performance capabilities, and so on. The control can be also time-dependent, such as traffic signals, or during peak periods. The control can be traffic responsive, such as railroad crossings, ramp control, traffic signals, or when drivers modify their behavior due to impatience or frustration. Finally, it must be noted that 'no control' is also a control state that can be encountered.

The *environment input* generally serves to modify the other inputs; namely demand, supply and control. Situations encountered include visibility, weather conditions, pavement conditions, unusual distractions, and others. In practice the performance may be overestimated because the traffic system is assumed to be under 'normal' environmental conditions, that is, under near-perfect conditions.

The analytical technique processes the above-mentioned input data to predict the performance of the traffic system. In the microscopic level of analysis (part 3 of this reader), analytical techniques may include car-following theories, time-space diagrams, stochastic queueing analysis and microscopic simulation. At the macroscopic level of analysis discussed in part 2 of this syllabus, these techniques may include capacity analysis, speed-flow-density relations, shock wave analysis, deterministic queueing analysis, and macroscopic simulation. These analytical techniques can be accomplished manually or using a computer, can utilized mathematical expressions or simulation, and can lead to solutions of a deterministic or a stochastic form.

The objective of demand-supply analysis is to *predict the performance of the traffic system from the viewpoint of the users and from the perspective of the total system.* The users will be interested in their travel times, incurred delays, queueing, comfort, risks, and energy consumption. The system manager will be concerned with system performance, level-of-service, air pollution, noise generation, accident rates, and total transport costs. The formulation often takes the form of a multi-objective function with a set of constraints. For example, at an isolated signalized intersection the objective may be to minimize a combination of delays and stops with the constraint of a maximum individual delay. Generally, as the size and the complexity of the traffic system increases, the formulation becomes more comprehensive and more complex as well.

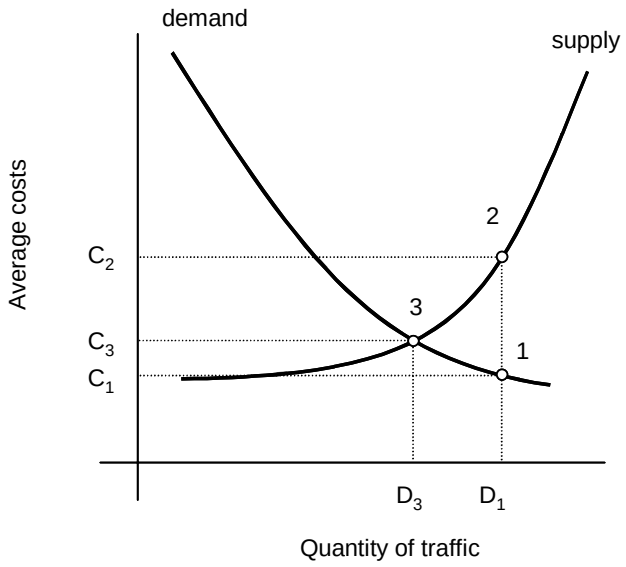

Figure 1.2: Demand-supply equilibrium

1.3 Feedback processes

Once the initial analysis is completed and the performance is predicted, a feedback process (or an iterative process) is generally desired or even required. These feedback processes take various forms and include determining demand-performance equilibrium, answering "what-if" statements, conducting sensitivity analysis, optimizing an objective function, and / or investigating traveler responses due to system changes.

Often the inputs to the analysis, such as demand and supply, are developed independently without knowledge as to how the traffic system will perform. When the initial analysis is completed, the anticipated demand level may not be compatible or appropriate for the resulting predicted performance. In illustration, consider a very simple single-link traffic system with demand-supply cost functions shown in Fig. 1.2. An initial demand level of D_1 was assumed with an anticipated user-cost of C_1. However, once the initial process is completed, a user cost of C_2 (rather than C_1) was predicted, with $C_2 > C_1$. The feedback process requires modifying the demand input until the demand and the specified supply are in equilibrium as denoted in point 3 in Fig. 1.2, where the demand level is reduced to D_3 with an associated average cost of C_3.

A second type of feedback process would be answering "what if" statements. For example, in investigating a rural road, the analyst might want to estimate the change in predicted performance *if* a 500-meter passing lane is added at the midpoint of the upgrade. The analyst would perform the initial process and then in the feedback process, the supply would be modified to include the 500-meter passing lane.

Another type of feedback process would be for *sensitivity analysis*. The analyst would generally supply input modifications either manually or by mathematical formulation. For example, the previous two-lane rural road could be addressed to find the effect of the location of the passing lane from the bottom of the upgrade to the top of the upgrade.

The optimization of an objective function is another form of the feedback process. The analyst normally employs some type of mathematical programming technique. This technique may vary from rigid linear programming approach to a brute-force branch-and-bound approach.

A final feedback process is the investigation of traveller responses to the system changes. Initially, there is equilibrium between demand and supply. However, over time the input to the demand-supply analysis is modified due to demand growth, supply changes, and / or control

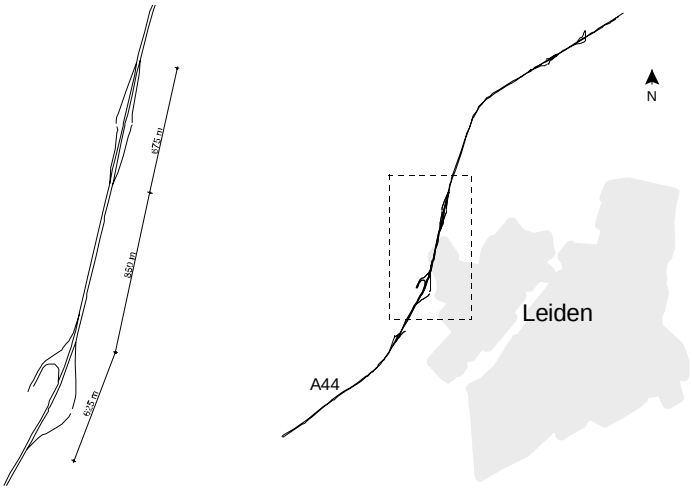

Figure 1.3: A44 near the Dutch city of Leiden

implementation. These input modifications change the predicted performance of the traffic system. The users then may respond by spatial, temporal, modal, or total travel changes. Users may change their paths of travel through the network, resulting in a spatial response. Users may change their time of travel, resulting in temporal responses. Users may change their mode of travel (car pool, bus, etc.), resulting in a model response. Finally, users may eliminate, combine, or create trips, resulting in total travel responses. As users respond in various ways, the input to the demand-supply analysis changes, and correspondingly, the predicted performance changes. This process of user responses and system performance changes becomes an iterative process and requires special procedures in order to reach equilibrium. An added complexity is that changes in travel patterns may modify the previously optimized control strategy. For instance, if a signal timing plan has been optimized, the system users may modify their paths of travel through the arterial network. Users will select cheaper cost routes, and flows change from some links to other links. If a sufficient number of users modify their paths through the network, the previously optimized signal timing plan may not longer be optimal.

1.4 Demand-supply analysis example

In this example, we aim to illustrate the application of the analytical framework for the demand-supply analysis for a motorway corridor. Since we have not yet discussed neither the knowledge, nor the analytical techniques that may be required to solve this problem, the approach is clearly oversimplified.

We will consider three situations: the reference case (or base case), the future situations without modifications (the so-called null-alternative) and an alternative treatment.

1.4.1 Base conditions

The considered roadway section is based on a part of a real motorway: the 2×2 motorway A44 near the Dutch city of Leiden (Fig. 1.3). The figure shows how two connections are situated very near each other. It was decided to consider only the Eastern roadway (see Fig. 1.4). The real roadway geometry has been used; the traffic demands are fictional.

Let us assume that the Table 1.1 describes the average traffic demands during the morning peak hour. We assume that there is a truck percentage of 10%. Furthermore, we assume that under ideal conditions, the capacity of a motorway lane equals 2000 veh/h. As we will see in the ensuing of this reader, the presence of (heavy) trucks will reduce the traffic supply (capacity), or equivalently, increase the traffic flow in person-car equivalents PCE. Assuming a PCE value

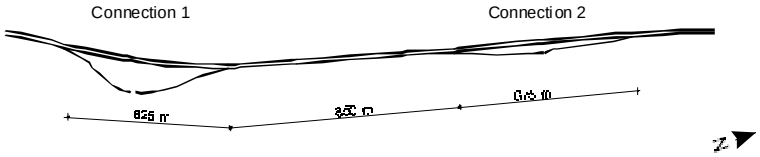

Figure 1.4: The Eastern roadway of the A44 near Leiden

	Destination		
Origin	Main road	Off-ramp 1	Off-ramp 2
Main road	2000	200	500
On-ramp 1	500	-	200
On-ramp 2	900	-	-

Table 1.1: Origin-destination demands (veh/h)

of 2 (which is rather high), the effective capacity per lane (expressed in veh/h) becomes 1900 veh/h.

1.4.2 Initial analysis

The first question that has to be answered is on which part of the motorway problems can be expected. This is the part where the volume-to-capacity ratio is highest (either on the main road, on an on-ramp or an off-ramp). It turns out that the traffic demand on the main road after the second connection equals 3400 veh/h. Since this is lower than the capacity (3800 veh/h), no real problems are expected. If traffic demands would increase equally for all origin-destination pairs, capacity problems are expected at the second on-ramp. The queue that would occur there would spill back, possibly reaching the off-ramp 2. In the chapter discussing shock-wave analysis, we will see that when this occurs, the speed at which the queue propagates increases, and the speed at which the queue spills back increases substantially. Even without shock-wave analysis, this can be understood, as will be seen in the remainder..

1.4.3 Feedback processes

We have now analyzed the reference case in which no problems are expected. In most cases, the traffic analyst would compare the modelling outcomes with empirical data (calibration and validation) to ensure that the representation of the current situation is indeed correct.

In the remainder of this example, we will consider several "what if" scenarios. First, let us assume that the traffic demands grow with 15% for all origin-destination pairs. In this case, the demands at the second connection will be 3910 veh/h (main road: 2875 veh/h; on-ramp 2: 1035), which is higher than the capacity of 3800 veh/h. From experience, we know that congestion mainly occurs on the main road, since the merging flow on the on-ramp appears to be able to merge without much delay. Thus, the 'effective capacity' that is available for the demand on the main road equals 3800 – 1035 = 2765 (capacity bottle-neck – demand on-ramp 2). As a result, a queue starts building upstream of on-ramp 2, propagating upstream. The rate at which the number of vehicles in the queue increases equals 2875 – 2765 = 110 veh/h. Assuming that vehicles in a queue take up approximately 10 m (per lane!), after 1 hour the length of the queue is 10·110/2=550 m. It will then be a matter of minutes before the queue spills back on the off-ramp. More precisely, the queue will spill back to the off-ramp when there will be 2·675/10 = 135 veh in the queue, which would take 135/110 hour. For that point onwards, problems will become more severe since now vehicles that do not have to pass the bottle-neck (on-ramp 2) will have to wait in the queue that is caused by it. How the delays

	Destination		
Origin	Main road	Off-ramp 1	Off-ramp 2
Main road	2000	900	500
On-ramp 1	500	-	1000
On-ramp 2	1400	-	-

Table 1.2: Origin-destination demands (veh/h) for future conditions

of with destination 'main road' relate to the delays of vehicles with destination 'off-ramp 2' compare is not obvious, since it may depend on complicated microscopic processes. However, if we assume that the speeds are equals irrespective of the destination, all vehicles will occur the same delay before arriving at off-ramp 2. How can this problem be solved?

Remark 1 *This situation has been analyzed with the microscopic simulation program FOSIM. It turns out that capacity problems will not always occur at on-ramp 2, but also at the motorway section between on-ramp 1 and off-ramp 2. From the micro-simulation it can be concluded that the inefficiency of the merging process as well as the use of the motorway lanes are the main cause for these problems. Vehicles with destination off-ramp 2 will merge to the right lane before off-ramp 1, causing traffic demands on the lane - on which the vehicles entering via on-ramp 1 need to merge - to exceed the capacity of the lane. In other words, the effective capacity of the merge is lower than the 3800 veh/h due to merging effects.*

Let us now assume that near connection 2, a new residential area is build. As a result, traffic demands on the on-ramp increase during the morning peak hour, and become 500 veh/h. At the same time, near connection 1 a business area is developed. It turns out that the employees working there mostly briefly use the A44, after which they use the underlying network. As a consequence, the traffic demands will increase significantly. The traffic demands from on-ramp 1 to off-ramp 2 increases to 800 veh/h. During the morning peak hour, the business area also attracts a lot of traffic. The traffic demands increases with 700 veh/h.

The highest traffic demand is now attained at the motorway right after the first on-ramp. This means that the upstream section will possibly run into congestion. The total demand equals 4000 veh/h, which is higher than the estimated capacity. Hence, we may expect that congestion will occur on this location. Congestion will mainly occur on the main road (which is usually the case, since merging traffic generally finds their way onto the main road). Then, the occurring queue will spill back upstream, possibly over the first off-ramp.

The number of vehicles that are stored depend on the duration of the oversaturated conditions. Assuming that all merging vehicles are able to flow onto the main road, yields that the remaining capacity for the traffic on the main road equals 3800 – 1500 = 2300 veh/h. Since the demand equals 2500 veh/h, the queue will grow at a rate of 200 veh/h. Assuming that a vehicle in a queue requires 8 m of roadway space – in practise, more space will be needed since the vehicles in the queue are driving – a queue of 200 veh will have a length of 200·8/2=800 m. Hence, the queue will spill back to the off-ramp 1 with an hour. The chapter on queuing analysis will go into detail on how to compute the dynamics of the queue.

One might think that not only the first on-ramp, but also the second on-ramp will pose a problem. According to the traffic demands, the total demand at the second on-ramp equations 3900 veh/h, which clearly also exceeds the capacity. However, since the actual flow at the section after the first on-ramp is less than the traffic demand (namely equal to the capacity of 3800 veh/h), the actual downstream demand will be less than 3800 veh/h, and no congestion will occur.

1.4.4 Ramp control

A possible solution to solve the foreseen queuing problems is the use of ramp-control (ramp-metering). Assume that the objective of the control is to keep traffic on the main road from becoming congested. For the first situation, the flow on on-ramp 2 that is allowed to enter the main road equals 925 veh/h, and hence 1035 – 925 = 110 veh/h need to be held back. The ramp-controller will thus let 925 veh/h merge onto the main road. We can now compute the queue lengths on on-ramp 2 after 1 hour. Note that since the on-ramp is a one-lane road, it has less storage capacity; the physical length of the queue will increase two times as quick.

For the second example (new residential area), a similar ramp-metering strategy can be proposed. In this case however, traffic on the metered on-ramp 1 may decide to use an alternative route, e.g. use the underlying network and drive to the second on-ramp (rat-running). When this occurs, on-ramp 2 will become oversaturated and congestion will occur there. In that case, it will also be prudent to meter the second on-ramp.

Part II

Traffic Flow Characteristics

Chapter 2

Microscopic and macroscopic traffic flow variables

Contents of this chapter. Introduction of vehicle trajectories, time headways (h), distance headways (s), intensity (q), density (k), mean speed (u), local and instantaneous mean speed, harmonic mean speed, relation $q = ku$. Formal definition of stationary and homogeneous states of a traffic flow; definition of q, k and u as a continuous function of position and time; definition of q, k, and u for a surface in the space-time plane; measuring methods, occupancy rate, moving observer method.

List of symbols

x, x_0	m	cross-section, location
t, t_0	s	(initial) time instants
v_i	m/s	speed of vehicle i
a_i	m/s^2	acceleration of vehicle i
γ_i	m/s^3	jerk of vehicle i
h_i	s	headway of vehicle i
s_i	m	distance headway of vehicle i
n	veh	number of vehicles passing cross-section x
T	s	length of time period
q	veh/s	flow rate, intensity, volume
m	veh	number of vehicles on roadway section at instant t
X	m	length of roadway section
k	veh/m	traffic density
u_L	m/s	local speed
u_M	m/s	instantaneous speed
$f_L(v)$, $f_M(v)$	$-$	local / instantaneous speed probability density function
σ_L^2	m^2/s^2	variance local speeds
σ_M^2	m^2/s^2	variance instantaneous speeds
$N(x,t)$	veh	cumulative vehicle count at cross-section x and instant t
β	$-$	occupancy rate
n_{active}	veh	number of active vehicle passings
$n_{passive}$	veh	number of passive vehicle passings

2.1 Introduction

In general, a traffic network consists of intersections and arterials. On arterials of sufficient length the traffic will no longer be influenced by the intersections, and drivers are mainly con-

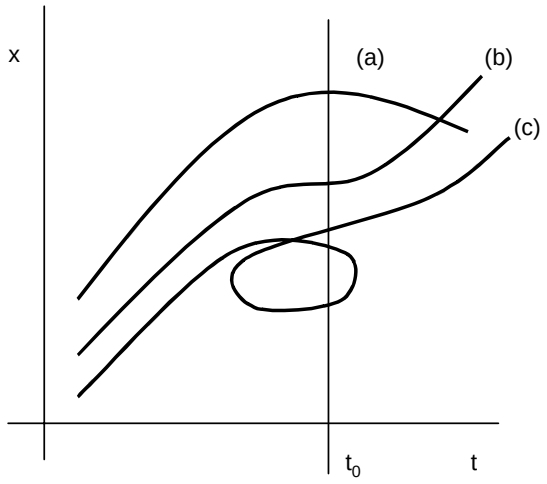

Figure 2.1: Time-space curves: (a) and (b) are vehicle trajectories; (c) is not.

cerned about traffic on the same roadway, either driving in the same or the opposing direction. In these lecture notes it is mainly this situation that is being discussed. The main microscopic variables are *trajectories*, *time headways*, and *distance headways.* The main macroscopic characteristics of a traffic flow are *intensity*, *density*, and *speed*. These variables and some related ones will be discussed in this chapter.

Vehicle trajectories

Very often in the analysis of a particular transportation operation one has to track the position of a vehicle over time along a 1-dimensional guideway as a function of time, and summarize the relevant information in an understandable way. This can be done by means of mathematics if one uses a variable x to denote the distance along the guideway from some arbitrary reference point, and another variable t to denote the time elapsed from an arbitrary instant. Then, the desired information can be provided by a function $x\,(t)$ that returns an x for every t.

2.1.1 Trajectory of a single vehicle

Definition 2 *A graphical representation of $x\,(t)$ in the (t, x) plane is a curve which we call a* trajectory.

As illustrated by two of the curves Fig. 2.1 (adapted from [14]), trajectories provide an intuitive, clear and complete summary of vehicular motion in one dimension. Curve (a), for example, represents a vehicle that is proceeding in the positive direction, slows down, and finally reverses direction. Curve (b) represents a vehicle that resumes travelling in the positive direction after stopping for some period of time. Curve (c) however is not a representation of a trajectory because there is more than one position given for some t's (e.g. t_0). Valid vehicle trajectories must exhibit one and only one x for every t.

Vehicle trajectories or rather, a *set of trajectories*, provide nearly all information concerning the conditions on the facility. As we will see in the ensuing of this chapter, showing multiple trajectories in the (t, x) plane can help solve many problems.

The definition of a trajectory is not complete in the sense that it does not specify which part of the vehicle the position of the vehicle refers to. In fact, none of the traffic flow theory handbooks explicitly specifies whether we consider the front bumper, the rear bumper or the centre of the vehicle. In the remainder of this reader, we will use the *rear bumper of a vehicle* as the reference point, unless explicitly indicated otherwise.

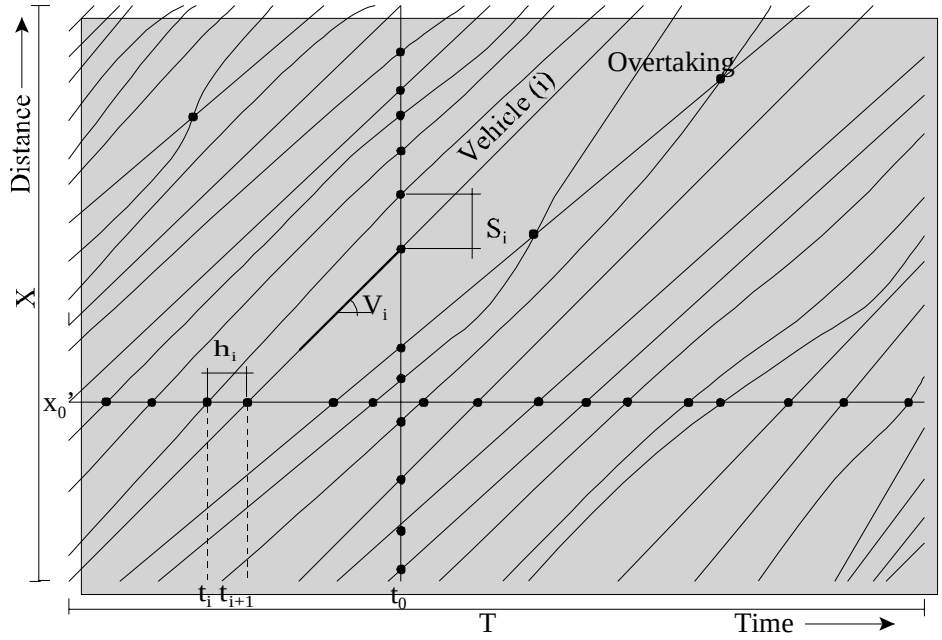

Figure 2.2: Vehicle trajectories.

A second issue is the fact that only a one-dimensional case is considered here. In fact, the position of a vehicle (or a pedestrian, a cyclist) consists of three dimensions x, y, and z.

For vehicular traffic (e.g. cars on a motorway or a bidirectional roadway), the coordinates generally do not refer to real-life coordinates, but are taken *relative to the roadway*, i.e. including the curvature of the latter. That is, x describes the longitudinal position with respect to the roadway, generally in the direction of the traffic. The y dimension, which is only seldom known/shown, describes the lateral position of the vehicle with respect to the roadway. This information thus includes the lane the vehicle is driving on. In fact, for any traffic system where the infrastructure largely determines the main direction of travel, the x and y direction respectively describe the longitudinal and lateral position of the vehicles along the roadway.

For traffic systems where this is not the case – consider for instance pedestrian walking infrastructure – the x and y (and sometimes the z coordinate as well) are given in Cartesian coordinates relative to some reference point $x = y = 0$. In those cases, the definition of x and y directions is more or less arbitrary.

Remark 3 *In traffic flow theory it is customary to show the position on the vertical axis and the time on the horizontal axis. However, in public transport, time is usually displayed vertically (increasing downwards) and position horizontally.*

See Fig. 2.2 for a set of vehicle trajectories for one-way traffic, i.e. the longitudinal position of vehicles along the roadway[1]. All information that the traffic analyst requires can be determined from the trajectories: individual speeds and acceleration, overtakings – where trajectories cross – but also macroscopic flow characteristics, such as densities, intensities, etc.

The speed of a vehicle is the tangent in a point of the trajectory; $v_i = dx_i/dt$; the acceleration of a vehicle is defined by $a_i = d^2x_i/dt^2$. Although these relations are well known, it is important to emphasize that steeply increasing (decreasing) sections of $x_i\,(t)$ denote a rapidly advancing (receding) vehicle; horizontal positions of $x_i\,(t)$ denote a stopped vehicle and shallow segments a slow moving vehicle. Straight line segments depict constant speed motion (with no acceleration)

[1]In the remainder of this reader, we will generally only describe the one-dimensional case (unless explicilty indicated). However, the discussed notions are in most cases easily extended to two or three dimensions.

and curving sections denote accelerated motion; here, the higher the curvature, the higher the absolute value of the acceleration. Concave downwards curves denote deceleration and concave upward (convex) curves denote accelerated motion.

2.1.2 Trajectories of multiple vehicles

On the spot $x = x_0$, i.e. a cross-section, one can observe the time instants that vehicles pass. The differences between successive moments are 'time headways' (h_i), and the speeds at a cross-section are 'local' speeds (v_i) (or spot speeds). Time headways can pertain to the leading vehicle directly in front (i.e. on the same lane), or vehicles which have a different lateral position (i.e. on another lane). For two or three dimensional flows, the definition of time headways is more involved.

On the moment $t = t_0$, one can observe positions of vehicles. The differences between successive positions are 'distance headways' (s_i), and the speeds at a moment are 'instantaneous' speeds v_i. For two or three dimensional flows, the notion of distance headway is less useful, for one since there is no direct relation with the density.

Remark 4 *In English, an important distinction is made between the speed and the velocity of a vehicle. In general, the speed is a scalar describing the absolute speed of a vehicle. The velocity is a one-, two-, or three-dimensional vector that also includes the direction of the vehicle. The latter direction is usually taken relative to the main direction of travel. For one-dimensional flows, the speed is generally equal to the velocity. For two- or three-dimensional flows, the speed is generally not equal the velocity. If* $\mathbf{v}(t)$ *denotes the velocity of a pedestrian* i*, then his/hers speed is defined by* $v(t) = \|\mathbf{v}(t)\| = \sqrt{v_1^2(t) + v_2^2(t)}$.

Let us close off by illustrating the applications of the use of trajectories, first by recalling an example of [14] showing how the use of the (t, x) plane can help in finding errors in the solution approach.

2.1.3 Applications of trajectories in traffic problem solving

Three friends take a long trip using a tandem bicycle for 2 persons. Because the bike riders travel at 20 km/h, independent of the number of riders, and all three persons walk at 4 km/h, they proceed as follows: to start the journey, friends A and B ride the bicycle and friend C walks; after a while, friend A drops off friend B who starts walking, and A rides the bicycle alone in the reverse direction. When A and C meet, they turn the bicycle around and ride forward until they catch up with B. At that moment, the three friends have complete a basic cycle of their strategy, which they then repeat a number of times until they reach their destination. What is their average travel speed?

The answer to this question is not straightforward, unless one plots the trajectories of the four moving objects on the (t, x) diagram. One finds by inspection that the average speed is 10 km/h. The proof of this is left to the reader as an exercise.

2.1.4 Application of trajectories to scheduling problems

From [14]. This problem illustrates the use of the time-space diagram to analyze the interaction of ships in a narrow canal. The canal is wide enough for only one ship, except for a part in the middle ('the siding'), which is wide enough for two ships so passing is possible. Ships travel at a speed of 6 km/h and should be at least 1.5 km apart when they are moving – expect when traveling in a convoy. When stopped in the siding, the distance between the ships is only 0.25 km. Westbound ships travel full of cargo and are thus given high priority by the canal authority over the eastbound ships, which travel empty. Westbound ships travel in four convoys which are regularly scheduled every 3.5 hours and do not stop at the siding. The problem is now the following

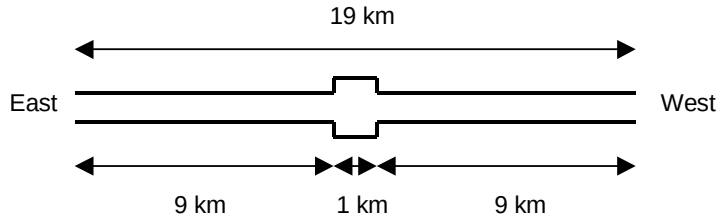

Figure 2.3: Sketch of a canal with an intermediate siding for crossing ships.

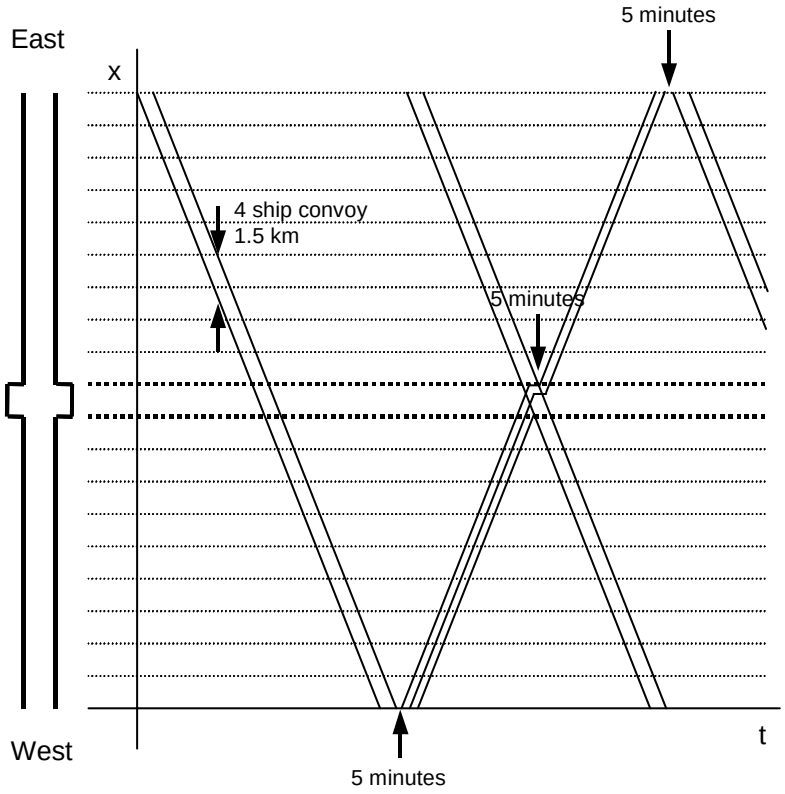

Figure 2.4: Time-space diagram in case of 1 km siding.

1. What is the maximum daily traffic of eastbound ships, and
2. What is the maximum daily traffic of eastbound ships if the siding is expanded to one km in length on both sides to a total of three km.

Note: we assume that eastbound ships wait exactly five minutes to enter either one of the one way sections after a westbound convoy has cleared it. We do not take into account that the ships do not accelerate instantaneously.

To solve the problem, we start by drawing the time-space diagram with the trajectories of the high-priority westbound convoys. See Fig. 2.4. The convoy leaves at the western end at 03:25. Note that since we neglect the size of the ships, the 4 ship convoy takes up 1.5 km distance. The second step is to draw the trajectory of a ship entering the western end of the canal at 3:30, which is the first time a ship can enter given the 5 minute time headway.

Note that the first ship must stop at the eastern part of the siding to yield the right of way to the last ship in the westbound convoy; note also how it makes it within the 5 minute allowance to the eastern end of the canal. The same process is follows successfully with the

second trajectory. The third ship will however not be able to arrive at the western bound of the siding within the 5 min. allowance and it cannot be dispatched. Thus, we find:

$$\text{capacity} = 2(\text{ships per 3.5 hours}) = 13.71 \text{ ships/day} \tag{2.1}$$

It is left as an exercise to determine the capacity in case of the wider siding.

2.1.5 Mathematical description of trajectories and vehicle kinematics

In this section, we recall the mathematical equations describing the kinematics of a vehicle i as a function of time by means of ordinary differential equations. The starting point of our description is the trajectory $x_i(t)$ of vehicle i; the speed v, acceleration a and the jerk γ are respectively defined by the following expressions

$$v_i(t) = \frac{d}{dt} x_i(t) \tag{2.2}$$

$$a_i(t) = \frac{d}{dt} v_i(t) = \frac{d^2}{dt^2} x_i(t) \tag{2.3}$$

$$\gamma_i(t) = \frac{d}{dt} a_i(t) = \frac{d^2}{dt^2} v_i(t) = \frac{d^3}{dt^3} x_i(t) \tag{2.4}$$

Given the initial conditions of vehicle i (in terms of its position, speed, and acceleration), at time $t = t_0$, we can easily determine the following equations of motion:

$$\begin{aligned} x_i(t) &= x_i(t_0) + \int_{t_0}^{t} v_i(s)\, ds && (2.5)\\ &= x_i(t_0) + (t - t_0)\, v_i(t_0) + \int_{t_0}^{t}\int_{t_0}^{s} a_i\left(s'\right) ds' ds && (2.6)\\ &= x_i(t_0) + (t - t_0)\, v_i(t_0) + \frac{1}{2}(t - t_0)^2\, a_i(t_0) && (2.7)\\ &\quad + \int_{t_0}^{t}\int_{t_0}^{s}\int_{t_0}^{s'} \gamma_i\left(s''\right) ds'' ds' ds && (2.8) \end{aligned}$$

The motion of a vehicle can also be described as functions of the position x or the speed v. For instance

$$v(x) = \frac{1}{dt/dx} \Leftrightarrow dt = \frac{dx}{v(x)} \tag{2.9}$$

yielding

$$t_i(x) = t_0 + \int_{x_0}^{x} \frac{1}{v_i(y)} dy \tag{2.10}$$

Alternatively, we can use different definitions to describe the process at hand. For instance, rather that the speed (which by definition describes changes in the position as a function of changes in time), we can define the *slowness* (describing the changes in time per unit distance)

$$w(x) = \frac{dt(x)}{dx} \tag{2.11}$$

yielding the following equations of motion (analogous to equation (2.5))

$$t_i(x) = t_i(x_0) + \int_{x_0}^{x} w_i(y)\, dy \tag{2.12}$$

The kinematics of a vehicle can be modelled by considering the *different forces that act on the vehicle*; once the resultant force F_i acting on the vehicle is known, the acceleration a_i of the vehicle can be easily determined by application of Newton's second law $F_i = m_i a_i$, where m_i denotes the mass of vehicle i. Amongst the most important force terms are the following [14]:

1. Propulsion force F_p: the force that the guideway exerts on the vehicle. It usually varies with time as per the 'driver' input, but is always limited by engine power and the coefficient of friction in the following way

$$\frac{F_p}{m} = a_p \leq g \min\left\{f, \frac{\kappa}{v}\right\} \tag{2.13}$$

where g is the acceleration of gravity, f is a dimensionless coefficient of friction, and κ is the power to weight ratio of the vehicle.

2. Fluidic (air) resistance F_f: the force that air / water exerts on the vehicle. A good approximation is the following

$$\frac{F_f}{m} = -\alpha v_r^2 \tag{2.14}$$

where v_r is the vehicle speed relative to the air of the fluid, and α is the coefficient of drag.

3. Rolling resistance F_r: a force term that is usually modelled as a linear relation with the speed, but is not as important for higher speeds.

4. Braking resistance F_b: this force depends on the force with which the brakes are applied, up to a maximum that depends on the friction coefficient between the wheels and the guideway. Thus, we can write

$$\frac{F_b}{m} \geq -gf \tag{2.15}$$

Note that generally, $F_b = 0$ when $F_p > 0$ (brake and throttle rarely applied simultaneously).

5. Guideway resistance F_g describing the effects of the acceleration due to the earths gravity. When the vehicle is at an upgrade, this force is negative; when the vehicle is on a downgrade, this force is positive.

Let us finally note that for simulation traffic on a digital computer, the continuous time scale generally used to describe the dynamics of traffic flow needs to be discretised and solved numerically. To this end, the time axis is partitioned into equally sized periods k, defined by $[t_k, t_{k+1})$, with $t_k = k\Delta t$. If we then assume that during the interval $[t_k, t_{k+1})$, the acceleration of vehicle i is constant, the time-discretised dynamics of the speed and the location become

$$v_i\left(t_{k+1}\right) = v_i\left(t_k\right) + a_i\left(t_k\right)\Delta t \tag{2.16}$$

$$x_i\left(t_{k+1}\right) = x_i\left(t_i\right) + v_i\left(t_k\right)\Delta t + \frac{1}{2}a_i\left(t_k\right)\Delta t^2 \tag{2.17}$$

Using these approximations, we will make an error of $O(\Delta t)$.

2.2 Time headways

Vehicle trajectories are the single most important microscopic characteristic of a traffic flow. However, only in very special cases, trajectory information is available. In most situations, one has to make due using local observations (i.e. observations at a cross-section x_0).

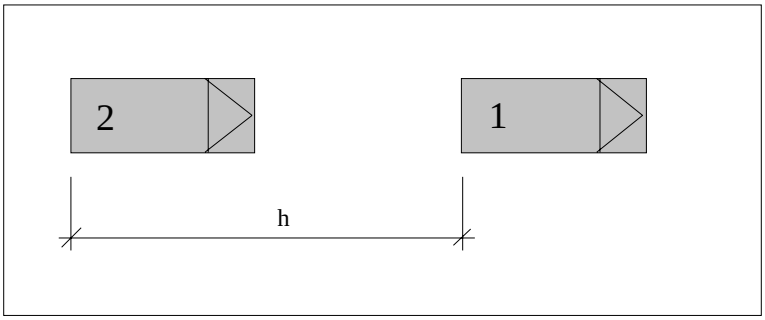

Figure 2.5: Definition of gross time headway h of vehicle 2.

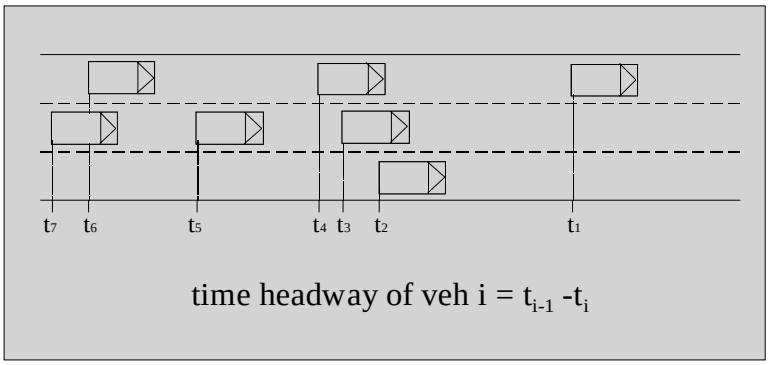

Figure 2.6: Definitions of time headways for a roadway

2.2.1 Time headways

Definition 5 *A time headway of a vehicle is defined as the period between the passing moment of the preceding vehicle and the vehicle considered; see also Fig. 2.2.*

Let h_i denote the time headway of the i th vehicle. The mean time headway equals for a period of length T

$$\overline{h} = \frac{1}{n}\sum_{i=1}^{n} h_i = \frac{T}{n} \tag{2.18}$$

where n denotes the number of vehicles that passed the cross-section during a period of length T. One can distinguish a nett and a gross time headway, as is shown in the definitions below.

Definition 6 *A nett time headway is defined as the period between the passing moments of the rear side of the preceding vehicle and the front of the vehicle considered.*

Definition 7 *A gross time headway (or simply headway) refers to the same reference point of both vehicles, e.g. front or back. Using the rear side of both vehicles has the advantage that the headway of a vehicle is dependent on its own length and not on the length of its predecessor; see Fig. 2.5.*

In traffic flow theory a time headway is usually a gross headway, because then the mean value is known if intensity is known (see Sec. 2.3). Other terms used for time headway are *gap* and *interval*.

Usually headways refer to one lane of traffic. However, they can also be defined for a roadway consisting of two or more lanes. An important consequence of such a definition is that gross and nett headways can be as small as 0.

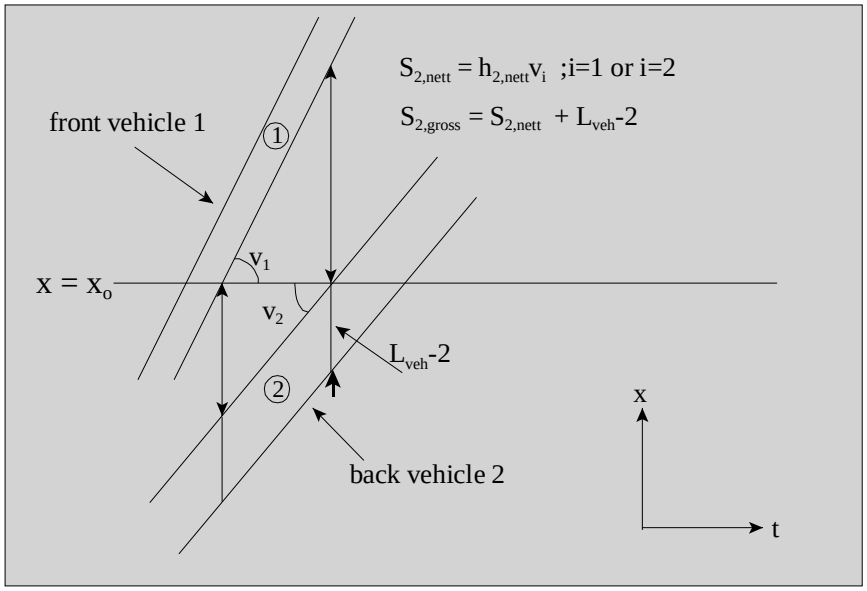

Figure 2.7: Calculation of distance headways from time headways and speeds

2.2.2 Distance headways

Using a similar definition, we can determine the distance headway between two vehicles. On the contrary to the time headway, the distance headway is a *instantaneous variable* defined at a certain time instant.

Definition 8 *A time headway of a vehicle is defined by the distance between the rear bumper of the preceding vehicle and the rear bumper of the considered vehicle at a certain time instant; see also Fig. 2.2.*

If s_j denotes the distance headway of the j th vehicle, then the mean distance headway equals

$$\overline{s} = \frac{1}{m} \sum_{j=1}^{m} s_j = \frac{X}{m} \tag{2.19}$$

where m denotes the number of vehicles that are present on a road of length X at a certain time t.

One can distinguish a nett and a gross distance headway, either including or excluding the length of the vehicle. Fig. 2.7 shows how one can calculate distance headways from local observations, using several possibilities.

2.3 Intensity, density and mean speed

The previous sections described the most important micrscoscopic traffic flow variables. In this section, the main macroscopic – i.e. describing the average behavior of the flow rahter than of each individual vehicle – intensity, density and mean speed, and the relations between them.

2.3.1 Intensity

Definition 9 *The intensity of a traffic flow is the number of vehicles passing a cross-section of a road in a unit of time.*

The intensity can refer to a total cross-section of a road, or a part of it, e.g. a roadway in one direction or just a single lane. Any unit of time may be used in connection with intensity, such as 24 h, one hour, 15 min, 5 min, etc. Hour is mostly used.

Apart from the unit of time, the time interval over which the intensity is determined is also of importance, but the two variables should not be confused. One can express the number of vehicles counted over 24 h in the unit veh/second. The intensity (or flow) is a local characteristic that is defined at a cross-section x for a period T, by

$$q = \frac{n}{T} \quad \text{[number of vehicles / unit of time]} \tag{2.20}$$

From Fig. 2.2 can be deduced that the period T is the sum of the time headways h_i of the n vehicles.

$$q = \frac{n}{T} = \frac{n}{\sum_i h_i} = \frac{1}{\frac{1}{n}\sum_i h_i} = \frac{1}{\bar{h}} \quad \text{where} \quad \bar{h} = \text{mean time headway} \tag{2.21}$$

It goes without saying that in using Eq. 2.21 one should use units that correspond with each other; e.g. one should not use the unit veh/h for q in combination with the unit s for h_i.

The definition of intensity is easily generalized to two or three dimensional flows. It is however important that one realizes that the flow q for a two or three dimensional system is in fact a vector. For instance, for a two dimensional flow, $\mathbf{q} = (q_1, q_2)$ describes the flow q_1 in the longitudinal direction and the flow q_2 in the lateral direction. The respective elements of the flow vector can be determined by considering a lines (for a two-dimensional flow) or a plane (for a three dimensional flow) perpendicular to the direction of the considered element of the flow vector. For instance, when considering the longitudinal component q_1, we need to consider a line perpendicular to this direction (i.e. in the lateral direction) and count the number of vehicles passing this line during time a time period of length T.

2.3.2 Density

Definition 10 *The density of a traffic flow is the number of vehicles present on a unit of road length at a given moment. Just like the intensity the density can refer to a total road, a roadway, or a lane. Customary units for density are veh/km and veh/m.*

Compared to intensity, determining the density is far more difficult. One method is photography or video from either a plane or a high vantage point. From a photo the density is simply obtained by counting m = the number of vehicles present on a given road section of length X.

The density is thus an instantaneous quantity that is valid for a certain time t for a region X. The density is defined by:

$$k = \frac{m}{X} \quad \text{[number of vehicles / unit of length]} \tag{2.22}$$

From figure 2.2 follows that the road length X equals the sum of the 'distance headways' s_i

$$k = \frac{m}{X} = \frac{m}{\sum_i s_i} = \frac{1}{\frac{1}{m}\sum_i s_i} = \frac{1}{\bar{s}} \quad \text{where} \quad \bar{s} = \text{mean distance headway} \tag{2.23}$$

For two or three dimensional flows, the density can be defined by considering either an area or a volume and counting the number of vehicles m that occupy this area or volume at a certain time t. The mean distance headway can in these cases be interpreted as the average area / volume that is occupied by a vehicle in the two or three dimensional flow.

Remark 11 *Strictly speaking, Eqns. (2.21) and (2.23) are only valid if the period T and section length X are precisely equal to an integer number of headways. Practically, this is only relevant when relatively short periods T or section lengths X are used.*

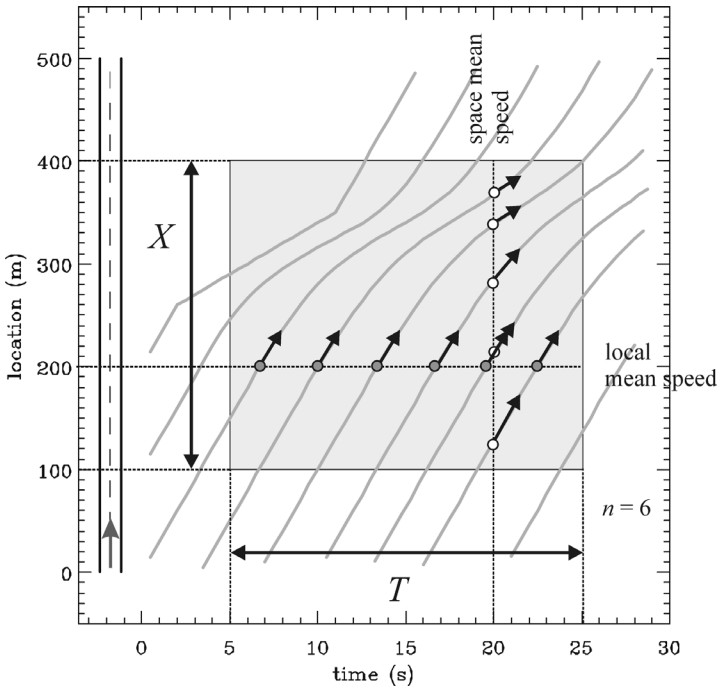

Figure 2.8: Defintion of local mean speed (or time mean speed) and space mean speed

Remark 12 *Intensity and density are traditionally defined as local and instantaneous variables. In the sequel of this chapter, generalized definition of both variables will be discussed. For these generalizations, the relation with the time and distance headways is not retained, at least not for the classical definition of the latter microscopic variables.*

2.3.3 Mean speed

The mean speed can be determined in several ways:

- Suppose we measure the speeds of vehicles *passing a cross-section during a certain period*. The arithmetic mean of those speeds is the so called 'local mean speed' (or mean spot speed; denoted with index L, referring to local).

$$u_L = \frac{1}{n}\sum_{i=1}^{n} v_i \tag{2.24}$$

- Suppose we know the speeds of the vehicles, v_j, that are *present on a road section at a given moment*. The arithmetic mean of those speeds is the-so called 'instantaneous mean speed' (denoted with index M, referring to moment), or 'space mean speed'.

$$u_M = \frac{1}{m}\sum_{j=1}^{m} v_j \tag{2.25}$$

Fig. 2.8 shows the difference between the definitions of the local mean speed (or local mean speed) and the space mean speed. Instantaneous speeds can be determined by reading the positions of vehicles from two photos taken a short time interval apart (e.g. 1 s). This is an expensive method, but it is possible to estimate the instantaneous mean speed from local speeds, as discussed in section 2.5.1.

Note that the notions of local and instantaneous speeds are only meaningful when a number of vehicles is considered; for a single vehicle, both speeds are equal.

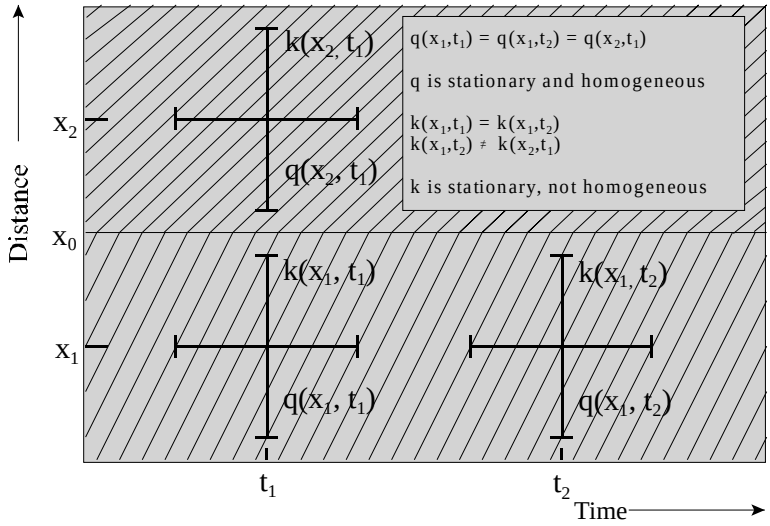

Figure 2.9: Effects of drastic change of road profile at position x_0

2.4 Homogeneous and stationary flow conditions

A traffic flow is composed of vehicles. Movements of different vehicles are a function of position and time (each vehicle has its own trajectory). The characteristics of a traffic flow, such as intensity, density, and mean speed, are an aggregation of characteristics of the individual vehicles and can consequently also be dependent on position and time.

Consider a variable $z(x,t)$. We define this variable z to be:

- Homogeneous, if $z(x,t) = z(t)$; i.e. the variable z does not depend on position.
- Stationary, if $z(x,t) = z(x)$; i.e. the variable z is independent of time.

Example 13 *Figure 2.9 presents a schematic image of vehicle trajectories. At the spot $x = x_0$ the road profile changes drastically, and as a result all vehicles reduce their speed when passing x_0. In this case the distance headways change but the time headways remain the same. This means that intensity q is stationary and homogeneous and density k is stationary but not homogeneous. Figure 2.10 presents vehicle trajectories in another schematized situation. At the moment to the weather changes drastically. All vehicles reduce their speed at that moment. Then the time headways change but the distance headways remain the same.*

2.4.1 Determination of periods with stationary intensity

Intensity is a characteristic that influences many other properties of the traffic flow. When studying such an influence, e.g. on the parameters of a headway distribution, it is advantageous to have periods with a constant or stationary intensity. To determine stationary periods one can apply formal statistical methods, but a practical engineering method is also available.

The number of vehicles that pass a cross-section after a given moment is drawn as a function of time. This can be done using passing moments of every vehicle, but it can also be done with more aggregate data, e.g. 5-minute intensities. A straight part of the cumulative curve corresponds to a stationary period. The next question is 'what is straight enough' but it turns out in practice that this is not problematical. One should choose the scale of the graph with some care; it should not be too large because on a detailed scale no flow looks stationary. *Fig. 1.4 presents an example of application of the method.* Three straight sections and two transition periods between them can be distinguished.

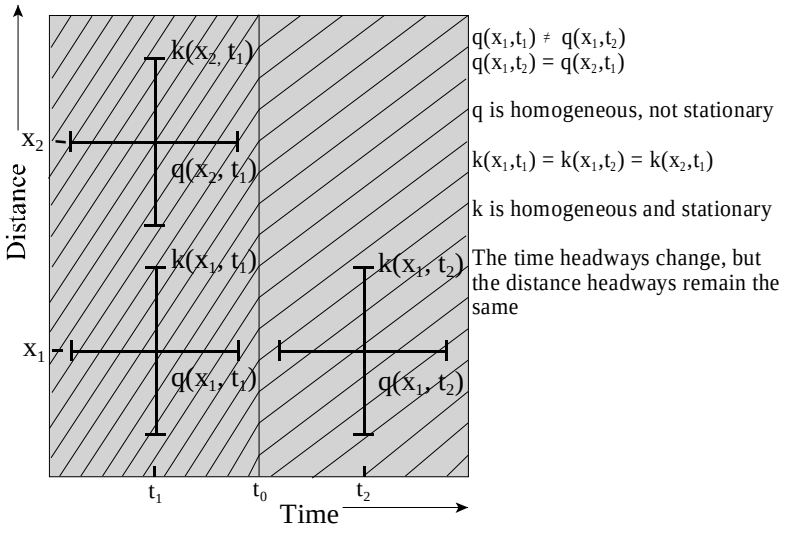

Figure 2.10: Effects of substational weather change at moment t_0

2.5 Relation between local and instantaneous characteristics

Considering a traffic flow in a stationary and homogeneous 'state', the following relation (referred to as the *fundamental relation*) is valid:

$$q = ku \tag{2.26}$$

In words: The number of particles, passing a cross-section per unit of time (q), equals the product of:

- The number of particles present per unit of distance (k); and
- The distance covered by those particles per unit of time (u)

From this general formulation it follows that the relation will be valid for all types of flows, e.g. liquids, gasses, pedestrians, etc. Clearly, when two or three dimensional flows are considered, both the flow and the speed (or rather, velocity) are vectors describing the mean intensity and speed in a particular direction.

2.5.1 Relation between instantaneous and local speed distribution

Let $f_L(v)$ and $f_M(v)$ respectively denote the local and instantaneous speed distribution. Consider a region from x_1 to x_2. Let us assume that the traffic state is homogeneous and stationary, i.e.

$$q(x,t) = q \quad \text{and} \quad k(x,t) = k \tag{2.27}$$

The probability that in the region $[x_1, x_2]$ we observe a vehicle with a speed in the interval $[v, v+dv)$ (where dv is very small) at time instant t, equals by definition

$$(x_2 - x_1)\, k f_M(v) dv \tag{2.28}$$

Consider the period from t_1 to t_2. Then, the probability that during the period $[t_1, t_2]$ a vehicle passes the cross-section x having a speed in the interval $[v, v+dv)$ equals

$$(t_2 - t_1)\, q f_L(v) dv \tag{2.29}$$

Now, consider a vehicle driving with speed v passing cross-section x_1 at time t_1. This vehicle will require $(x_2 - x_1)/v$ time units to travel from x_1 to x_2. Hence:

$$t_2 - t_1 = \frac{x_2 - x_1}{v} \tag{2.30}$$

As a result, the probability that during the interval $[t_1, t_1 + (x_2 - x_1)/v]$ a vehicle driving with speed v passes x_1 is equal to the probability that a vehicle with speed v is present somewhere in $[x_1, x_2]$ at instant t_1. From Eq. 2.29, we can calculate that the probability that a vehicle passes x_1 during the period $[t_1, t_1 + (x_2 - x_1)/v]$:

$$\frac{x_2 - x_1}{v} q f_L(v) dv \tag{2.31}$$

which is in turn equal to eq. 2.28, implying

$$v k f_M(v) = q f_L(v) \tag{2.32}$$

Integrating 2.32 with respect to the speed v yields the following relation between concentration and intensity

$$\int v k f_M(v) dv = k \int v f_M(v) dv = k \langle v \rangle_M = \int q f_L(v) dv = q \tag{2.33}$$

where we have used the following notation to describe the mean-operator with respect to the probability density function of the instantaneous speeds

$$\langle A(v) \rangle_M = \int A(v) f_M(v) dv \tag{2.34}$$

Note that $\langle v \rangle_M$ denotes the mean instantaneous speed u_M.

At the same time, we can rewrite eq 2.32 as $k f_M(v) = q f_L(v)/v$. Again integrating with respect to the speed v, we find the following relation

$$k = q \left\langle \frac{1}{v} \right\rangle_L \tag{2.35}$$

where

$$\langle A(v) \rangle_L = \int A(v) f_L(v) dv \tag{2.36}$$

In combining Eqns. 2.32 and 2.35, we get the following relation between the instantaneous speed distribution and the local speed distribution

$$f_M(v) = \frac{1}{v \left\langle \frac{1}{v} \right\rangle_L} f_L(v) \tag{2.37}$$

Fig. 2.11 shows the probability density function of the local speeds collected at a cross-section of a two-lane motorway in the Netherlands during stop-and-go traffic flow conditions. Note that the local speed and the instantaneous speed probability density functions are quite different. Local speeds are collected at two-lane A9 motorway in the Netherlands during peak hours. Note that the differences between the speed distributions are particularly high for low speeds.

2.5.2 Local and instantaneous mean speeds part 1

Consider the case where data is collected using a presence type detection, e.g. an inductive loop. Assume that besides the passage times of the vehicles, also their speeds are determined. Furthermore, assume that during the data collection periods, traffic conditions are stationary and homogeneous.

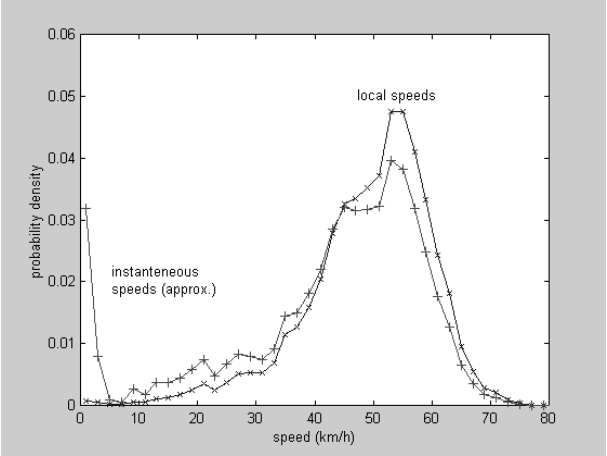

Figure 2.11: Local speed density function and instantaneous speed density function. Local speeds are collected at two-lane A9 motorway in the Netherlands during peak hours.

From the speeds v_i $(i = 1, ..., n)$ collected at the cross-section x, we can determine the so-called *local (or instantaneous) empirical probability density function* $\hat{f}_L$ as follows

$$\hat{f}_L(v) = \frac{1}{n}\sum_{i=1}^{n}\delta(v - v_i) \tag{2.38}$$

where δ is the so-called δ-dirac function implicitly defined by

$$\int a(y)\,\delta(y - x)\,ds = a(x) \tag{2.39}$$

and where n equals the number of vehicles that have passed the cross-section x during the considered period. The local mean speed thus becomes

$$u_L = \langle v \rangle_L = \frac{1}{n}\sum_{i=1}^{n} v_i \tag{2.40}$$

where the mean operator is defined using Eq. (2.36) with $f_L(v) = \hat{f}_L(v)$. According to Eq. (2.37), we now find the following relation between the instantaneous empirical speed distribution $\hat{f}_M(v)$ and the local speed distribution $\hat{f}_L(v)$

$$\hat{f}_M(v) = \frac{1}{v\left\langle \frac{1}{v} \right\rangle_L}\hat{f}_L(v) = \frac{1}{v\frac{1}{n}\sum_{i=1}^{n}\frac{1}{v_i}}\frac{1}{n}\sum_{i=1}^{n}\delta(v - v_i) \tag{2.41}$$

Using this expression, we find the following equation for the instantaneous or space mean speed

$$u_M = \langle v \rangle_M = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{v_i}} \tag{2.42}$$

That is, the space mean speed is equal to the *harmonic average of the speeds collected at a cross-section x during a stationary period.*

The fact that local mean and space mean differ is not only true for speeds *but also for other vehicle characteristics!* We will derive a formula for determining the space mean (of some variable) based on local observations, and also a formula for the reverse procedure. The trick used in these derivations is: define the mean over individual vehicles; divide the traffic flow in uniform sub-flows; use the relation $q = ku$ per sub-flow i; go back to individual values by shrinking the sub-flows to one vehicle.

- From local observations to space mean

Vehicle i passing cross section x has quantitative property z_i. Examples for z are: speed; length; number of passengers; weight; emission rate (= emission per time); etc.

As an introduction, we note that the local mean of z is by definition:

$$z_L = \frac{1}{n} \sum_i z_i \tag{2.43}$$

We can divide the flow in uniform sub-flows q_j , uniform with respect to speed v_j and characteristic z_j. Then the local mean is a mean with intensities q_j as weight:

$$z_L = \frac{\sum_i q_i z_i}{\sum_i q_i} \tag{2.44}$$

Now the derivation: The space-mean value of z is by definition a mean with the densities as weights:

$$z_M = \frac{\sum_i k_i z_i}{\sum_i k_i} \tag{2.45}$$

Replacing k_i by q_i/v_i yields

$$z_M = \frac{\sum_i (q_i/v_i)\, z_i}{\sum_i (q_i/v_i)} \tag{2.46}$$

Take $q_i = 1$, i.e. a sub-flow consists of one vehicle, i.e.

$$z_M = \frac{\sum_i z_i/v_i}{\sum_i 1/v_i} \tag{2.47}$$

Hence the space mean of z can be estimated by taking a weighted sum of local observed characteristics z_i, and the weights are $1/v_i$.

- Reverse, i.e. from space observations to local mean

Assume the following: observed in space are vehicles i with characteristics z_i and v_i. The space mean of z is by definition:

$$z_M = \frac{1}{m} \sum_j z_j \tag{2.48}$$

We divide the total density in uniform subclasses with density k_j, uniform with respect to speed v_j and characteristic z_j.

$$z_L = \frac{\sum_j q_j z_j}{\sum_j q_j} = \frac{\sum_j k_j v_j z_j}{\sum_j k_j v_j} \tag{2.49}$$

Take $k_j = 1$. Then

$$z_L = \frac{\sum_j v_j z_j}{\sum_j v_j} \tag{2.50}$$

Hence the local mean can be estimated from space observations by taking a weighted sum of instantaneous observed characteristics z_j and the weights are v_j.

Example 14 *Many car drivers complain about the high number of trucks on the road during peak hours. Drivers do not observe the traffic flow at a spot but over a road section, whereas road authorities observe and publish truck percentages (TP) based on local observations. Say: local TP = 10%; mean speed of trucks = 80 km/h and mean speed of cars = 120 km/h. Then the TP observed in space is: TPM = (10/80) / {10/80 + 90/120} = 14.3 %. So instead of 1 in 10 vehicles being a truck at a spot, the space proportion is about 1 in 7. The difference is more extreme at a grade where the speed difference between cars and trucks is larger.*

From the example can be concluded that sometimes a characteristic can have a substantially different value at a spot and in space. Which value is more appropriate depends on the application.

2.5.3 Local and instantaneous mean speeds part 2

For general distributions, we have the mean speeds $u_L = \langle v \rangle_L$ and $u_M = \langle v \rangle_M$ are related by:

$$u_L = u_M + \frac{\sigma_M^2}{u_M} \tag{2.51}$$

Proof. This relation can be proven as follows. From eqns. 2.32 and 2.33 we find

$$f_L(v) = \frac{kvf_M(v)}{q} = \frac{kvf_M(v)}{k\langle v \rangle_M} = \frac{vf_M(v)}{u_M} \tag{2.52}$$

By definition, the expected local speed equals

$$u_L = \langle v \rangle_L = \int v f_L(v) dv = \int \frac{v^2 f_M(v)}{u_M} dv \tag{2.53}$$

It can thus be easily shown that

$$u_L = \int \frac{(v-u_M)^2 + 2vu_M - u_M^2}{u_M} f_M(v) dv = \frac{\sigma_M^2}{u_M} + u_M \tag{2.54}$$

where the variance of the instantaneous speeds by definition equals

$$\sigma_M^2 = \int (v-u_M)^2 f_M(v) dv \tag{2.55}$$

■

Relation 2.21 cannot be used to derive u_M from u_L because it contains the parameter σ_M, the standard deviation of the instantaneous speeds, which is not known either. From eq. 2.21 it follows that the local mean speed is larger than the instantaneous mean speed. This fact has got the 'memory aid': 'faster vehicles have a higher probability to pass a cross-section than slower ones'. This memory aid stems from the following situation: Consider a road section on which are present subpopulations of vehicles with the same density k but different speeds u_i. According to the relation $q_i = ku_i$ it is true that vehicles with a higher speed pass more often than those with a lower speed. This situation can be maintained and easily understood if one considers a loop that vehicles drive around and around.

Example 15 *At free flow at a motorway a representative value for the instantaneous mean speed is 110 km/h and for the instantaneous standard deviation (STD) of speeds 15 km/h. Then the local mean speed equals: 110+15²/110 = 112 km/h. At congestion the instantaneous mean speed might be 10 km/h and the instantaneous STD 12 km/h. Then the local mean speed equals: 10+12²/10 = 24 km/h. Hence at free flow the difference between local and instantaneous mean speed can be neglected but at congestion the difference can be substantial.*

2.5.4 Relation between flow, density and speed revisited

In the end, the question remains which of the average speeds should be used in $q = ku$ (Eq. 2.26): the time-mean speed or the space-mean speed? The answer to this question can be found as follows. Considered a group of vehicles that is driving with speed u_i. If we consider a certain section of the roadway, we will denote the density of this group of vehicles by k_i. Since all vehicles are driving at the same speed, equation 2.26 holds. As a consequence, group i will contribute to the total volume q according to $q_i = k_i u_i$. The total traffic volume thus becomes

$$q = \sum_i q_i = \sum_i k_i u_i \tag{2.56}$$

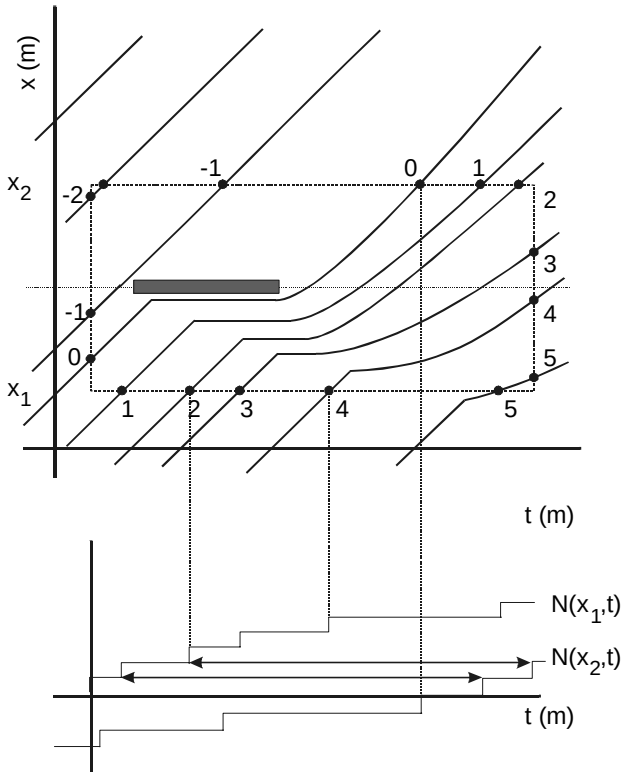

Figure 2.12: Trajectories and cumulative vehicle plot

The total density in the considered roadway section is determined by

$$k = \sum_i k_i \tag{2.57}$$

The mean speed thus equals

$$u = \frac{\sum_i k_i u_i}{\sum_i k_i} = \frac{\sum_i q_i}{\sum_i q_i / u_i} \tag{2.58}$$

Since q_i is the number of vehicles that passes a cross-section per unit time, u is the harmonic average of the speeds collected at the cross-section. Hence, we need to use the space-mean speeds. As the space-mean speed is not easy to obtain an approximation for it has been derived, the harmonic mean of local speeds.

2.6 Cumulative vehicle plots and their applications

A cumulative plot of vehicles is a function $N(x,t)$ that represents the number of vehicles that has passed a cross section x from an arbitrary starting moment. Fig. 2.12 shows a couple of vehicle trajectories which are numbered in increasing order, as well as the cumulative vehicle plots $N(x_1,t)$ and $N(x_2,t)$ determined for two cross-sections x_1 and x_2 as a function of time. Note that each time a vehicle passes either of the cross-section, the respective cumulative vehicle count increases with one. The example shows vehicles being stopped at a controlled intersection. The arrows in the lower figure indicate the travel times of vehicles 1 and 2, including their delay due to the controlled intersection.

Fig. 2.13 shows another plot determined from real-life data.

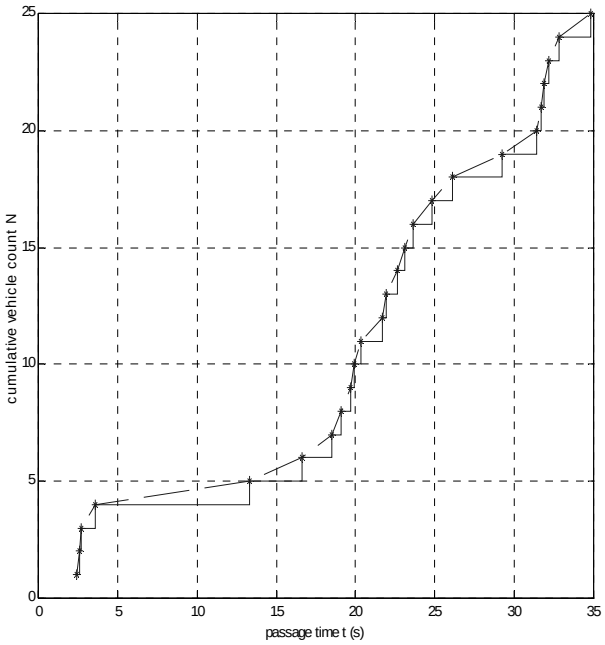

Figure 2.13: Cumulative flow function $N(x,t)$ and smooth approximation $\tilde{N}(x,t)$. Note that for passage times t_j, we have $N\left(x,t_j\right) = \tilde{N}\left(x,t_j\right)$

2.6.1 Relation between cumulative flow, intensity and density

It is obvious that the intensity measured at a certain cross-section x during period t_1 to t_2 equals:

$$q(x, t_1 \text{ to } t_2) = \frac{N(x,t_2) - N(x,t_1)}{t_2 - t_1} \tag{2.59}$$

Since the vehicle are indivisible objects, $N(x,t)$ is a step function. However, in most practical problems it is not needed to have solutions with an accuracy of one vehicle. This allows us to approximate the step function $N(x,t)$ by a smooth function $\tilde{N}(x,t)$ that is *continuous and can be differentiated*. The smooth approximation $\tilde{N}\left(x,t\right)$ is defined such that for the passage times t_j at which the vehicles pass the cross-section x, we have $N\left(x,t_j\right) = \tilde{N}\left(x,t_j\right)$; see Fig. 2.13 for an example.

Taking the limit of eq. (2.59) for $(t_2 - t_1) \to 0$ results in:

$$q(x,t) = \frac{\partial \tilde{N}(x,t)}{\partial t} \tag{2.60}$$

As the position x is a continuous variable, we have now introduced a concept of a local and instantaneous intensity.

Now consider two cumulative plots at position x_1 and x_2. Then at time instant t, the average density is:

$$k(x_1 \text{ to } x_2, t) = \frac{\tilde{N}(x_1,t) - \tilde{N}(x_2,t)}{x_2 - x_1} \tag{2.61}$$

Taking the limit of eq. (2.61) for $(x_2 - x_1) \to 0$ leads to:

$$k(x,t) = -\frac{\partial \tilde{N}(x,t)}{\partial x} \tag{2.62}$$

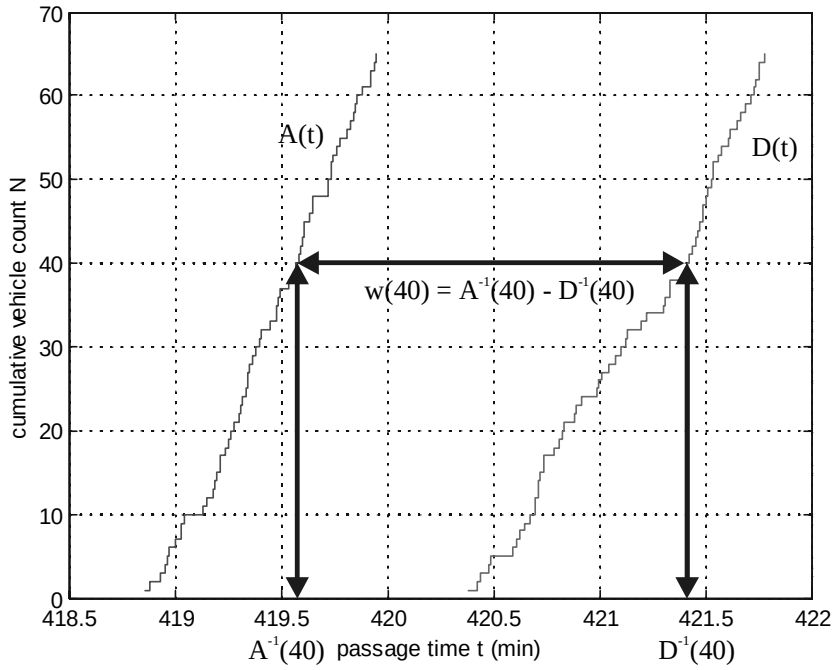

Figure 2.14: Determination of travel time using cumulative vehicle counts

Finally we can define the mean speed at the spot x and instant t as:

$$u(x,t) = q(x,t)/k(x,t) \tag{2.63}$$

Consequently all three main macroscopic characteristics of a traffic flow can be handled as continuous functions of the position x and time t. This property will be used when considering macroscopic traffic flow models.

2.6.2 Trip times and the cumulative flow function

The cumulative flow function has a vast number of applications. It can also be used to determine the *time vehicles need to traverse a certain roadway section* (recall Fig. 2.12). To us further consider this application of the cumulative flow function.

Let $A(t) = N(x_1, t)$ (the *arrival curve*) denote the cumulative vehicle count at the entry of a roadway section x_1; let $D(t) = N(x_2, t)$ (the *departure curve*) denote the cumulative vehicle count at the exit x_2 of a roadway section. The time t at which the N-th vehicle passes the cross-section can be obtained by finding the time t where a horizontal line across the ordinate N meets the crest of a step. Let $N^{-1}(x, N)$ denote the function that returns t for a given N. If vehicles pass the roadway in a *First-In-First-Out* (FIFO) order, then these N-th observations correspond to the same individual, and hence the trip time of the N-th vehicle throught the section $[x_1, x_2]$ equals

$$w(N) = A^{-1}(N) - D^{-1}(N) \tag{2.64}$$

See Fig. 2.14. Note that this relation is not true in case passing is possible.

2.6.3 Delay times and cumulative flow functions

Let us assume that the time τ needed to traverse the empty roadway section is equal for all vehicles. Furthermore, let us assume that vehicle 1 enters an empty roadway. Then, the free trip time τ can be determined by $\tau = D^{-1}(1) - A^{-1}(1)$.

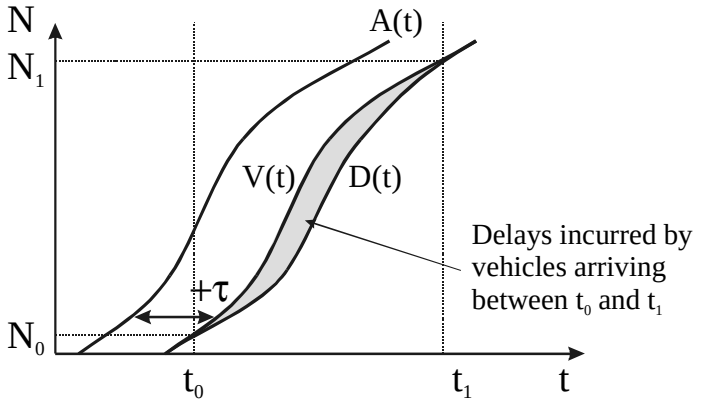

Figure 2.15: Arrival curve $A\,(t)$ (arrivals at entry), arrival curve $D\,(t)$ (arrivals at exit), and the virtual departure curve $V\,(t)$.

Now, let us assume that τ has been determined. We can then define the *virtual arrival curve* (cf. [14]) by translating the arrival curve τ units along the time-axis, i.e.

$$V(t) = A(t - \tau) \tag{2.65}$$

Since horizontal separations in the (t, N) diagram represent time and verticle separations represent accumulation, it is clear that the area enclosed by $V(t)$ and $D(t)$ and any two vertical lines $t = t_0$ and $t = t_1$, is the total delay time incurred by vehicles in the road-section during the period $[t_0, t_1]$. Fig. 2.15 shows the curves $A\,(t)$, $D\,(t)$, and $V\,(t)$, and the total delay W incurred by vehicles arriving at x_2 during the period $[t_0, t_1]$, defined by

$$W = \int_{t_0}^{t_1} [V\,(t) - D\,(t)]\, dt \tag{2.66}$$

The average delay for all vehicles that have passed during the period hence becomes

$$\bar{w} = \frac{W}{V(t_1) - V(t_0)} = \left(\frac{W}{t_1 - t_0}\right)\left(\frac{t_1 - t_0}{V(t_1) - V(t_0)}\right) \tag{2.67}$$

See Fig. 2.15. From eqn. (2.67), we can rewrite

$$\bar{Q} = \bar{\lambda}\bar{w} \tag{2.68}$$

where the average number $\bar{Q}$ of vehicles in the system is equal to

$$\bar{Q} = \frac{W}{t_1 - t_0} \tag{2.69}$$

and where the arrival rate $\bar{\lambda}$ is equal to

$$\bar{\lambda} = \frac{V(t_1) - V(t_0)}{t_1 - t_0} \tag{2.70}$$

The chapter describing queuing theory will provide further discussion on the applications of cumulative curves in traffic theory. At his point, let us provide an example involving delays at an controlled intersection.

Example 16 *Let us show an example of the use of cumulative curves to determine the average waiting time at a controlled intersection. Let R and G respectively denote the red and the green time of the cycle; the total cycle time $C = R + G$. Let λ denote the arrival rate at the controlled*

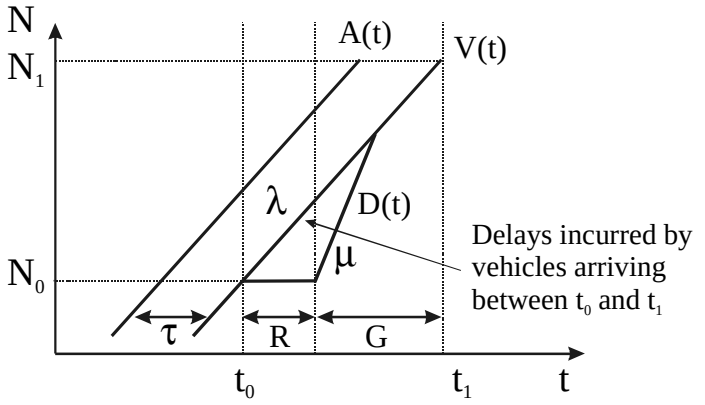

Figure 2.16: Delay at a pre-timed traffic signal

intersection. When the controller is in its red phase, vehicle cannot pass and the queue will grow. During the green phase, μ veh/s can be served. The area between $V(t)$ and $D(t)$ again reflects the total delay W of the vehicles in the system. The total number of vehicles served during one cycle equals $n' = N_1 - N_2$. The number of vehicles that have been delayed is equal to n. From the geometry of the picture, it should be clear that $R = \frac{n}{\lambda} - \frac{n}{\mu}$ and thus that $n = R/\left(\lambda^{-1} - \mu^{-1}\right)$. For a non-saturated intersection, we have $n < n' = \lambda C$, yielding

$$\mu G < \lambda C \tag{2.71}$$

When the signal is over-saturated, this condition is not met and the queue would grow steadily over time. Since $W = nR/2$ and $n = R/\left(\lambda^{-1} - \mu^{-1}\right)$, we obtain $W = \frac{1}{2}\lambda\mu R^2\left(\mu - \lambda\right)$. The long-run average delay per car is thus

$$\bar{w} = \frac{W}{n'} = \frac{1}{2}\frac{\mu R^2}{\left(\mu - \lambda\right)C} \tag{2.72}$$

Example 17 *Fig. 2.17 shows a two-lane rural highway in California, with a controlled intersection at its end (Wildcat Canyon Road)*[2]*. The site has limited overtaking opportunities. More important, there are no entry and exit points. At 8 observation points, passage times of vehicles have been collected and stored. Using these data, the cumulative curves shown in Fig. 2.18 have been determined. In the same figure, we have also indicated travel times. Fig. 2.18 shows how vehicle 1000 experiences a higher travel time than say vehicle 500. Note that at the observer 8 site, a traffic responsive traffic signal is present. Notice how the gradient of the cumulative curve of observer 8 decreases after some time (approximately at 7:15). The reason for this is the growth in the conflicting traffic streams at Wildcat Canyon Road. At approx. 8:40 we see another reduction of the flow. An interesting observation can be made from studying the cumulative curves in more detail. From Fig. 2.19 we can observe clearly the flows during the green phase and the zero flow during the red phase. Notice that before 8:40, the flow during the green phase is approximately constant for all green phases. Reduction in the average flow is caused by a reduction in the length of the green phase as a result of the conflicting streams. From approx. 8:40 onward, the flow during the green phase is reduced. The reason for this is that congestion from downstream spills back over the intersection.Fig. 2.20 shows the shifted cumulative curves. The curves are shifted along the time-axis by the free travel time. The resulting curves can be used to determine the delays per vehicle, as well as the total and average delays.Fig. 2.21 shows*

[2]The data used for this example can be downloaded from the website of Prof. Carlos Daganzo at Berkeley (www.ce.berkeley.edu/~daganzo/spdr.html).

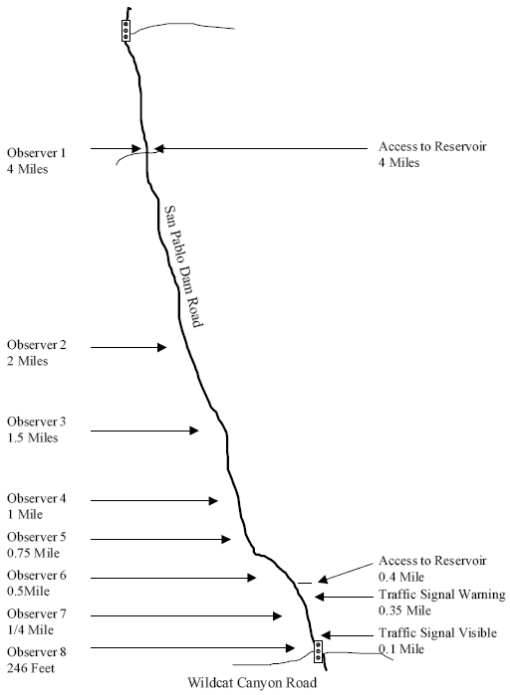

Figure 2.17: Overview of data collection site.

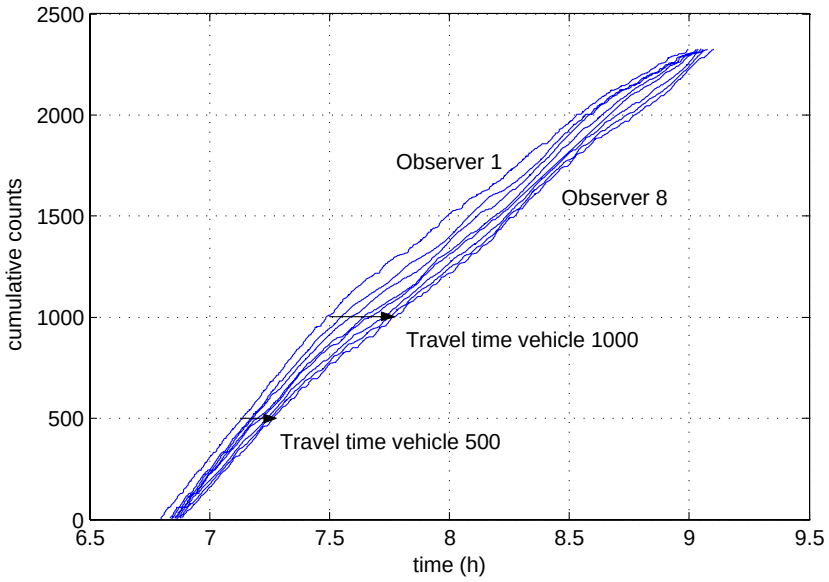

Figure 2.18: Cumulative counts collected at San Pablo site during morning peak period.

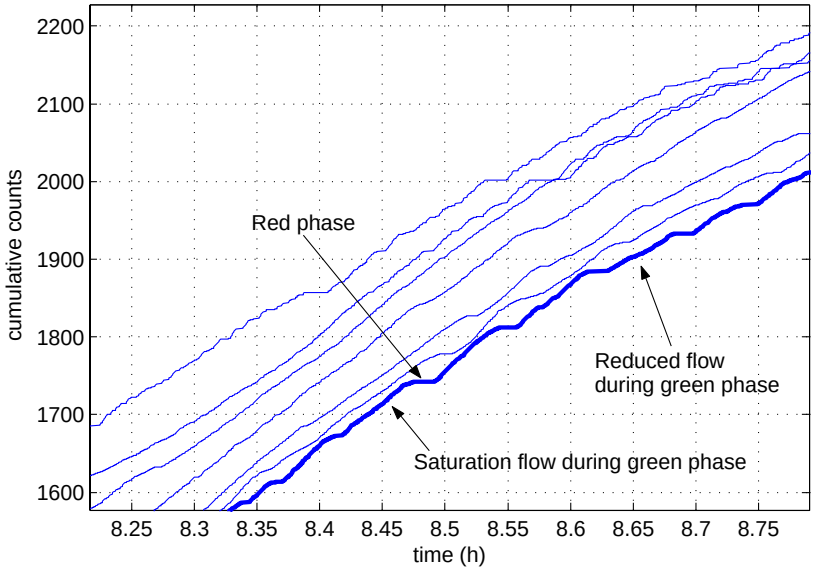

Figure 2.19: Magnification of cumulative curves showing reducting in saturation flow from 8:30 onward.

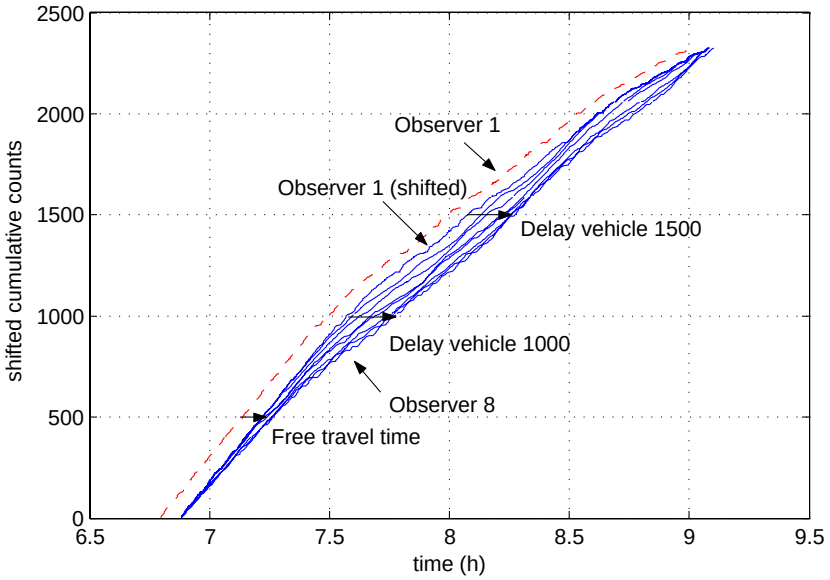

Figure 2.20: Shifted cumulative curves indicating the delays per vehicle.

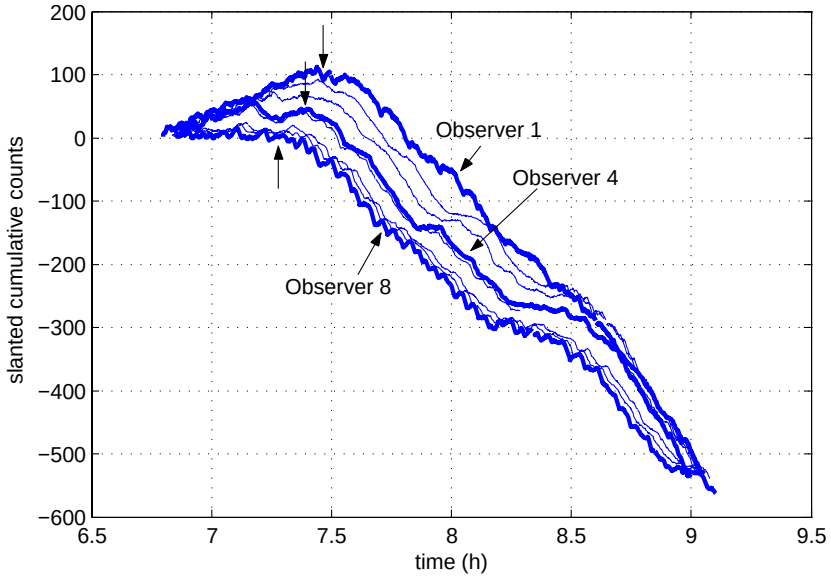

Figure 2.21: Slanted cumulative curves $N'(x_i,t) = N(x_i,t) - q'(t-t_0)$ for $q' = 1300$ veh/h.

the so-called slanted cumulative curves $N'(x,t)$ *defined by* $N'(x,t) = N(x,t) - q'(t-t_0)$ *for some value of* q'*. Using these slanted curves rather than the reguler cumulative curves can be useful to identify changes in the flowrate more easily. The three arrows in Fig. 2.21 show for instance, when the reducting in the flowrate at 7:15 due to increased conflicting flows on the intersection reach observer 4 and observer 1 respectivly. That is, the slanted cumulative curves can be used to identify the speed at which congestion moves upstream. This topic will be discussed in more detail in the remainder of the syllabus.*

2.6.4 Derivation of conservation of vehicle equation using cumulative flow functions

The *conservation of vehicle equations* (also known as the *continuity equation*) is the most important equation in macroscopic traffic flow modelling and analysis. It describes the fact that to no vehicles are lost or generated. The conservation of vehicle equation can be derived using elementary differential calculus. However, the definitions of q and k introduced above, allow a more direct derivation: differentiate eqn. (2.60) to x and (2.62) to t

$$\frac{\partial q(x,t)}{\partial x} = \frac{\partial^2 \tilde{N}(x,t)}{\partial x \partial t} \quad \text{and} \quad \frac{\partial k(x,t)}{\partial t} = -\frac{\partial^2 \tilde{N}(x,t)}{\partial x \partial t} \tag{2.73}$$

and thus yielding the conservation of vehicle equation

$$\frac{\partial q(x,t)}{\partial x} + \frac{\partial k(x,t)}{\partial t} = 0 \tag{2.74}$$

Eq. (2.74) represents one of the most important equations in traffic flow theory. It describes the fact that vehicles cannot be created or lost without the presence of sinks (off-ramps) or sources (on-ramps).

2.7 Definition of q, k and u for a time-space region

The macroscopic traffic flow variables intensity, density and mean speed in the preceding sections referred to either a *cross-section* x *and a period* T (local variable), or, *a road section* X *and*

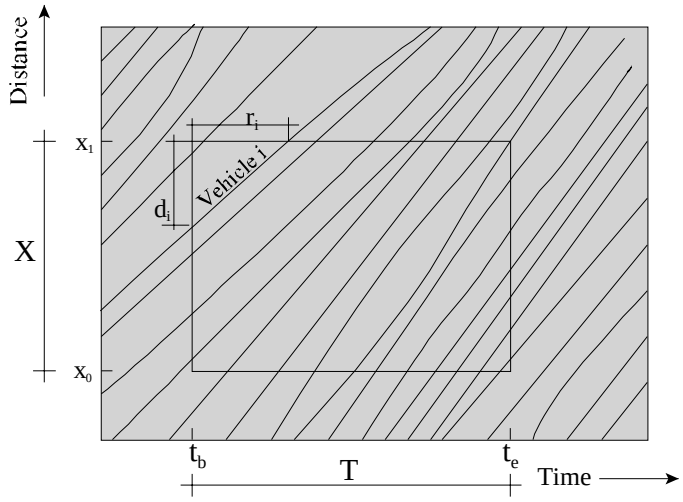

Figure 2.22: Definition of q, k, and u for a time-space region

a moment t (instantaneous variable), or, a *cross-section x and a moment t*. Edie [19] has shown that these variables can be defined for a time-space region and that those definitions are consistent with the earlier ones and make sense. Their main property is that small (microscopic) fluctuations have less influence on the values of the macroscopic variables, which is often an advantage in practice.

Fig. 2.22 shows a time-space rectangle of length X and duration T. In the window marked every vehicle covers a distance d_i and is present during a period r_i. Now intensity can be defined as the sum of all distances d_i divided by the area of the window; in formula:

$$q = \frac{\sum_i d_i}{XT} \tag{2.75}$$

Discussion of this definition:

- The dimension is correct, it is: 1/time.
- Define $x_1 = x_0 + \delta x$ and let δx, that is the height of the window, approach 0. Then the distances d_i become approximately the same for all vehicles; Edie approaches the number entering the window through the lower boundary (n) multiplied by δx and in the denominator $X = \delta x$. Consequently (2.75) approaches n/T, the 'ordinary' (local) intensity.
- Above the rectangle has been shrunken into a cross-section. One can also let it shrink into a moment. Define $t_1 = t_b + \delta t$ and let δt, that is the width of the window, approach 0. Then the distances the vehicles cover in this window are $v_i \delta t$ and the denominator becomes $X \delta t$. Hence eq. (2.75) approaches $\sum_i v_i / X$. This seems a strange definition of intensity but it is a consistent one. It is in fact not more artificial that the local density derived from q/u, which is equal to $\sum_i (1/v_i) /T$.

In analogy with the definition for the intensity, the definition for the density becomes:

$$k = \frac{\sum_i r_i}{XT} \tag{2.76}$$

Deduce yourself that if one lets t_e approach t_b, i.e. letting the window approach a vertical line, the definition of k transforms into the ordinary (instantaneous) definition. Also investigate what happens if one lets the window approach a horizontal line.

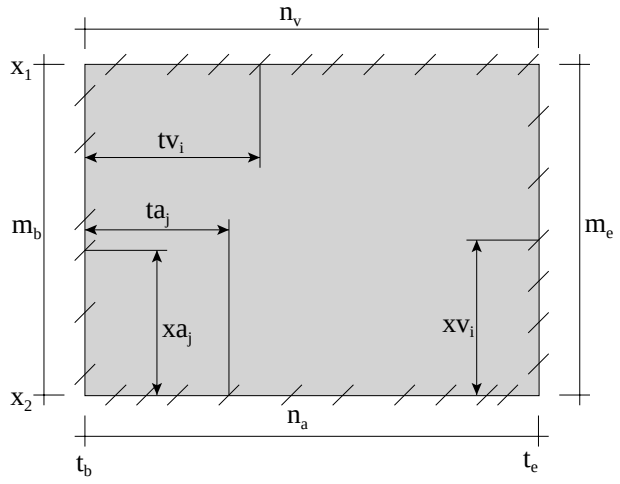

Figure 2.23: Determination of q and k for a time-space region

Finally, the mean speed for a time-space region is defined as:

$$u = \frac{\sum_i d_i}{\sum_i r_i} \tag{2.77}$$

Consequently with these definitions $q = ku$ is valid by definition.

The variable $\sum_i d_i$ is called the 'production' of the vehicles in the time-space region, and $\sum_i r_i$ the total travel time. Customary units are: vehicle-kilometre [veh km] and vehicle-hour [veh h].

Remark 18 *The definitions do not require the window to be a rectangle; they are applicable for an arbitrary closed surface in the time-space plane.*

2.7.1 Calulating the generalised q and k

To determine the generalised intensity and density for a time-space region, only the vehicle coordinates at the border of the area and not the trajectories are needed.

Fig. 2.23 is the time-space area of Fig. 2.22 with only:

- n_a: number of arrivals at cross-section x_0
- n_v: number of departures at cross-section x_1
- m_b: number of positions at moment t_b
- m_e: number of positions at moment t_e

This lead to

$$\text{Production}: K = \sum_i x_i = n_v X - \sum_{j=1}^{m_b} xa_j + \sum_{i=1}^{m_e} xv_i \tag{2.78}$$

$$\text{Travel time}: R = \sum_i t_i = m_e T - \sum_{j=1}^{n_a} ta_j + \sum_{i=1}^{n_v} tv_i \tag{2.79}$$

The required data (the moments ta_j and tv_j , and their numbers n_a and n_v) are simple to determine at the cross-sections x_0 and x_1.

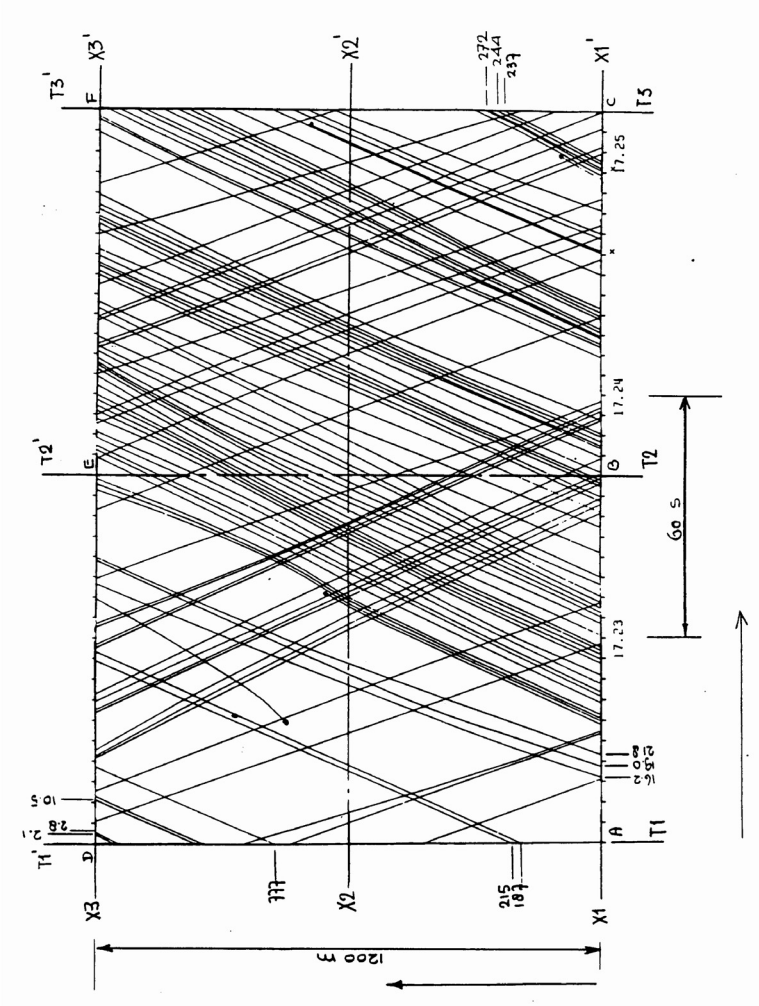

Figure 2.24: Trajectories plot

Determining xa_j and xv_i is clearly more difficult. A possible solution is to identify the vehicles when they pass cross-sections x_0 and x_1, and determine their position at the moments t_b and t_e by means of interpolation. This interpolation becomes more accurate by installing more detector stations between x_0 and x_1 and registering vehicles with an identification.

In principle, vehicles do not appear or disappear unnoticed, and consequently the relation

$$n_a + m_b = n_v + m_e \tag{2.80}$$

holds. This can be used as a check on observations.

Example 19 *Macroscopic characteristics have been derived from a plot of trajectories, depicted in figure 2.24. This is a representation of traffic operation on a two-lane road with vehicles going in both directions. The length of the road section is 1200 m and the considered period lasts 3 minutes. We will consider the traffic in the direction from X1 to X3. Intensities and densities can be determined straightforwardly by counting trajectories crossing lines 'position = a constant' and crossing lines 'time = a constant' respectively. See Table 1 and Table 2 for results.*

Inspecting the numbers shows that especially the density varies substantially between the different moments. The approach of determining intensity and density for a time-space region has as its goal to reduce these strong fluctuations. The moments and positions at the borders of the time-space region have been measured from the plot. The region considered is the rectangle

Cross-section	Period	Period (min)	#veh	q (veh/h)
$X - X1'$	$T1 - T2$	3	65	1300
$X2 - X2'$	$T1 - T3$	3	61	1220
$X3 - X3'$	$T1 - T3$	3	52	1040

Table 2.1: Intensity for 3 cross-sections

Moment	Road section	Length (m)	#veh	k (veh/km)
$T1 - T1'$	$X1 - X3$	1200	7	5.8
$T2 - T2'$	$X1 - X3$	1200	27	22.5
$T3 - T3'$	$X1 - X3$	1200	8	15.0

Table 2.2: Density for 3 cross-sections

ACFD. Result:

$$K = \sum_i x_i = 70.6 \ \textit{veh-km} \quad \textit{and} \quad R = \sum_i t_i \tag{2.81}$$

$$q = \frac{K}{XT} = \frac{70.6}{1.2 \cdot (3/60)} = 1177 \ \textit{veh/h} \tag{2.82}$$

$$k = \frac{R}{XT} = \frac{0.888}{1.2 \cdot (3/60)} = 14.8 \ \textit{veh/km} \tag{2.83}$$

$$u = \frac{q}{k} = \frac{1177}{14.8} = 79.5 \ \textit{km/h} \tag{2.84}$$

Check:

$$n_a + m_b = n_v + m_e \qquad 65 + 7 = 52 + 18 \tag{2.85}$$

2.8 Measuring methods

Real traffic data is one of the most important elements in analysing and improving traffic systems. Different systems have been proposed to collect the different microscopic and macroscopic quantities. This section describes some systems to measure the different traffic flow variables described in this chapter. It is beyond the scope of this course to go into futher detail.

2.8.1 Passage times, time headways and intensity

The *local* traffic variables, passage times, time headways and intensity, can be measured with relative little effort. They are generally determined at a cross-section for each passing vehicle or averaged during a period of time. Detection can be achieved by so-called *infrastructure-based detectors*, which operate at a fixed location. However, moving observer methods are also possible. Let us briefly discuss some well know detection methods.

- *Manual, using a form and a stopwatch.* Advantages of this simple method is that one can use a detailed registration of vehicle types; a disadvantage is the (labour) costs, and the inaccuracies – especially with respect to the microscopic variables.

XT space	n_a	n_v	m_b	m_e	q (veh/h)	k (veh/km)	u (km/h)
ABED	34	14	7	27	938	12.4	75.4
BCEF	31	38	27	18	1415	17.2	82.2

Table 2.3: Time-space q, k and u for smaller areas

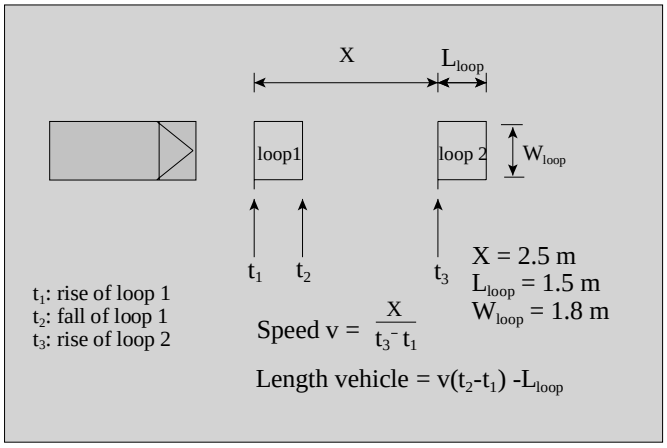

Figure 2.25: Loop configuration as applied on Dutch motorways

- *Pneumatic tube.* The pneumatic tube is one of the oldest traffic detectors available, and consist of a hollow rubber tube that sends a pressure wave when a vehicle runs over it. This device can accurately determine passage times and time headways. A disadvantage is the relatively short time the tubes survive. Note that a tube counts not vehicles but axles. The conversion from axles to vehicles introduces an extra error, especially when many articulate and non-articulate trucks are present in the flow.

- *Induction loops.* The passing of vehicles (iron) changes the magnetic field of the loop that is buried in the road surface. These changes can be detected. The induced voltage shows alternately a sharp rise and fall, which correspond approximately to the passing of the front of the vehicle over the front of the loop and the rear of the vehicle over the rear of the loop; see Fig.2.25. If one installs two loops behind each other on a lane (a 'trap'), then one can determine for each vehicle: passing moment; speed; (electrical) vehicle length. On Dutch motorways it is customary to use two loop detectors; see Fig. 2.25, implying that in principle, individual vehicle variables are available. The figure shows the relation between the individual speed v, the (electrThdsdsdsic) vehicle length, and the different time instant t_1, t_2 and t_3. Fig. 2.26 show how these relations are determined. Also note that there is redundancy in the data, i.e. not all avaiable information is used. The Dutch Ministry only stores the 1-minute or 5-minute *arithmetic* averages (counts and average speeds). This in fact poses another problem, since the arithmetic mean speed may not be an accurate approximation of the instantaneous speed, especially during congested conditions. The Dutch Ministry does however provide the possibility to temporarily use the induction loop for research purposes, and storing individual passing times for each vehicle.
  In the USA often one loop is used because that is cheaper. In the latter case intensity and the so called occupancy rate (see section 2.9) can be measured, but speed information is lacking unless one makes assumptions on the average lengths of the vehicles. In other countries (e.g. France), a combination of single-loop and double-loop detectors is used.
  Non-ideal behaviour of the equipment and of the vehicles (think about the effect of a lane change near the loops) lead to errors in the measured variables. Let us mention that a possible error in speed of 5 % and in vehicle length of 15 %. Especially at low speeds, i.e. under congestion, large errors in the intensities can occur too.

- *Infrared detector.* This family of traffic detectors detects passing vehicles when a beam of light is interrupted. This system thus provides individual passing times, headways and intensities.

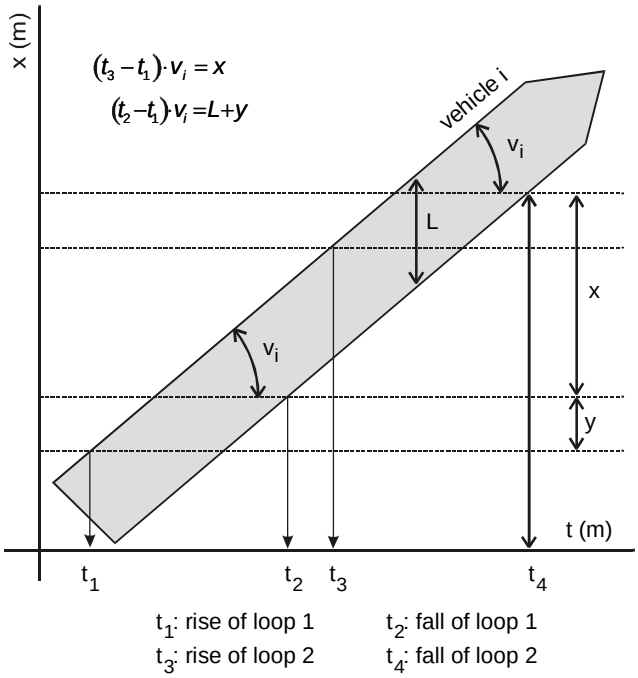

Figure 2.26: Calculating the individual vehicle speed v_i and the vehicle length L from double induction loops.

- *Instrumented vehicles.* Moving observer method, see Sec. 2.10.

2.8.2 Distance headways and density

In principle, distance headways and density can only be measured by means of photos or video recordings from a high vantage point (remote-sensing, see Sec. 2.8.3), that reveal the positions of the vehicles for a particular region X at certain time instants t. That is, the instantaneous variables in general require *non-infrastructure based* detection techniques. However, these method are mostly expensive and therefore seldom used (only for research purposes). Usually distance headways and densities are calculated from time headways, speeds and intensities, or from the the occupancy rate (see section 2.9) is used instead.

2.8.3 Individual speeds, instantaneous and local mean speed

Invidual vehicle speeds can be determined by infrastructure-based detectors and by non-infrastructure based detection methods (probes, remote-sensing). In practice individual vehicle speeds are measured and averaged. Keep in mind that the correct way to average the individual speed collected at a cross-section is using harmonic averages!

- *Radar speedometers at a cross-section.* This method is mostly used for enforcement and incidentally for research.
- *Induction loops or double pneumatic tubes* (discussed earlier under Intensity)
- *Registration of licence number (or other particular characteristics) of a vehicle / vehicle recognition.* Registration / identification at two cross-sections can be used to determine the mean travel speed over the road section in between. Manual registration and matching of the data from both cross-sections is expensive. The use of video and automatic processing of the data is currently deployed for enforcement (e.g. motorway A13 near Kleinpolderplein, Rotterdam).

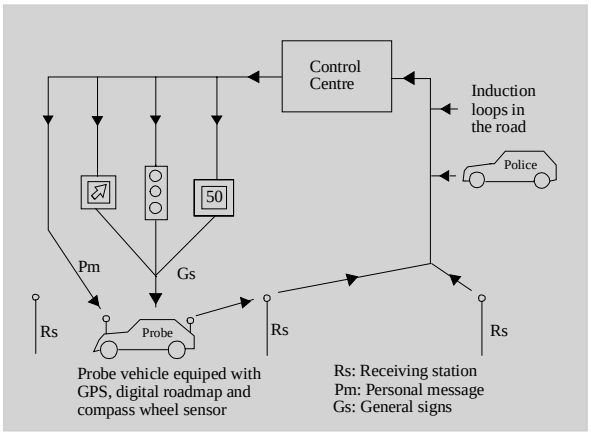

Figure 2.27: Lay-out of probe vehicle system concept

Figure 2.28: Results of vehicle detection and tracking (following objects for consecutive frames) obtained from a helicopter.

- *Probe vehicles.* Probe vehicles are ordinary vehicles with equipment that measures and also emits to receiving stations, connected to a control centre, regularly variables such as position, speed, and travel time over the last (say) 1 km. This results in a very special sample of traffic flow data and its use is under study. Combined with data collected from road stations (with e.g. induction loops), probe data seems very promising for monitoring the state of the traffic operation over a network (where is the congestion?; how severe is it?; how is the situation on the secondary network?; etc); see Fig. 2.27 and [57].

- *Remote-sensing techniques.* Remote sensing roughly pertains to all data collection methods that are done from a distance, e.g. from a plane, a helicopter, or a satelite. In illustration, the Transportation and Traffic Engineering Section of the TU Delft has developed a data collection system where the traffic flow was observed from a helicopter using a digital camera. Using special software, the vehicles where detected and tracked from the footage. By doing so, the trajectories of the vehicles could be determined (and thus also the speeds). It is obvious that this technique is only applicable for special studies, since it is both costly and time consuming. Alternative approaches consist of mounting the system on a high building. Fig. 2.28 shows an example of vehicle detection and tracking.

2.9 Occupancy rate

Earlier it was stated that in the Netherlands induction loops at motorways are installed in pairs but that in the US often one loop stations are used. The use of one loop has brought about the introduction of the characteristic occupancy rate. A vehicle passing over a loop temporarily 'occupies' it, approximately from the moment the front of the car is at the beginning of the loop until its rear is at the end of the loop. Note this individual occupancy period as b_i . For a period T, in which n vehicles pass, the occupancy rate β is defined as:

$$\beta = \frac{1}{T}\sum_{i=1}^{n} b_i \tag{2.86}$$

Per vehicle with length L_i and speed v_i : $b_i = (L_i + L_{loop})/v_i$. Suppose vehicles are approximately of the same length, then b_i is proportional to $1/v_i$, and β can be rewritten to:

$$\beta = \frac{1}{T}\sum_{i=1}^{n} \frac{L_i + L_{loop}}{v_i} = \frac{L_{tot}}{T}\sum_{i=1}^{n} \frac{1}{v_i} = L_{tot}\frac{n}{T}\frac{1}{n}\sum_{i=1}^{n} \frac{1}{v_i} = L_{tot}\frac{q}{u_M} = L_{tot}k \tag{2.87}$$

It appears that β is proportional to the (calculated local) density k if vehicle lengths are equal. However, if a mix of passenger cars and trucks is present, then the meaning of β is less obvious.

Conclusion 20 *If one uses a single induction loop, β is a meaningful characteristic, but if one uses two loops it is better to calculate density k from intensity q and the harmonic mean of the local speeds, u_M.*

2.10 Moving observer method

It is possible to measure macroscopic characteristics of traffic flow by means of so called moving observers, i.e. by measuring certain variables form vehicles driving with the flow. The method is suitable for determining characteristics over a larger area, e.g. over a string of links of a network.

The method produces only meaningful results if the flow does not change drastically. When investigating a road section with an important intersection, where the intensity and/or the vehicle composition change substantially, it is logical to end the measuring section at the intersection.

Moving observer (MO). The MO drives in direction 1 and observes:

- t_1 = travel time of MO over the section
- n_1 = number of oncoming vehicles
- m_1 = number of passive overtakings (MO is being overtaken) minus number of active overtakings (MO overtakes other vehicles)

Remark 21 *If the MO has a lower speed than the stream, then m_1 is positive. m_1 is so to speak the number of vehicles a MO counts, just as a standing observer. The difference is that vehicles passing the observer in the negative direction (vehicles the MO overtakes) get a minus sign.*

The MO drives in direction 2 and observes n_2, m_2, t_2

- q_i = intensity in direction i; $i = 1, 2$
- k_i = density in direction i

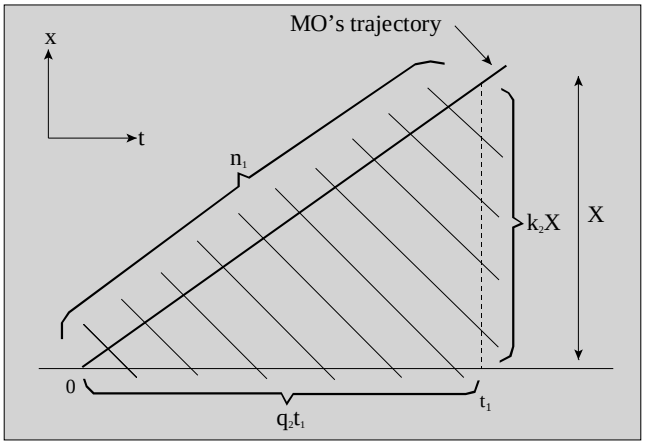

Figure 2.29: Determination number of oncoming vehicles

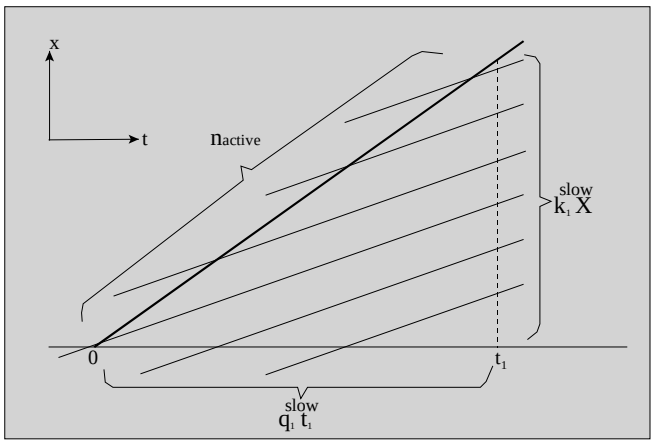

Figure 2.30: Determination of n_{active}

- u_i = mean speed of direction i
- X = section length

Derivation. Number of opposing vehicles

$$n_1 = q_2 t_1 + k_2 X \tag{2.88}$$

And for direction 2

$$n_2 = q_1 t_2 + k_1 X \tag{2.89}$$

To derive the number of overtakings, divide the stream into a part slow than the MO and a part faster than the MO. From Fig. 2.30 follows

$$n_{active} = k_1^{slow} X - q_1^{slow} t_1 \tag{2.90}$$

and Fig. 2.31

$$n_{passive} = q_1^{fast} t_1 - k_1^{fast} X \tag{2.91}$$

$$m_1 = n_{passive} - n_{active} = \left(q_1^{fast} + q_1^{slow}\right) t_1 - \left(k_1^{fast} + k_1^{slow}\right) X \tag{2.92}$$

$$= q_1 t_1 - k_1 X \tag{2.93}$$

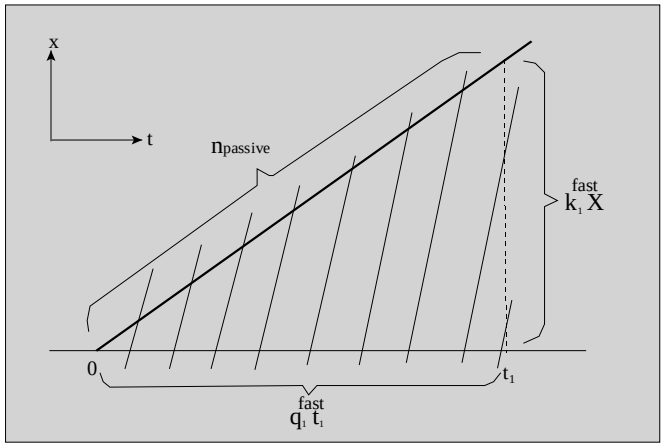

Figure 2.31: Determination of $n_{passive}$

For direction 2 we have the corresponding formula of (2.93)

$$m_2 = q_2 t_2 - k_2 X \tag{2.94}$$

We have 4 equations for the unknowns q_1, q_2, k_1 and k_2. In matrix form

$$\begin{pmatrix} 0 & t_1 & 0 & X \\ t_2 & 0 & X & 0 \\ t_1 & 0 & -X & 0 \\ 0 & t_2 & 0 & -X \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \\ k_1 \\ k_2 \end{pmatrix} = \begin{pmatrix} n_1 \\ n_2 \\ m_1 \\ m_2 \end{pmatrix} \tag{2.95}$$

The equations can be solved easily. Sum row 1 and row 4 of matrix (2.95) $\Rightarrow (t_1 + t_2)\, q_2 = n_1 + m_2$

$$q_2 = \frac{n_1 + m_2}{t_1 + t_2} \tag{2.96}$$

Symmetry $\Rightarrow$

$$q_1 = \frac{n_2 + m_1}{t_1 + t_2} \tag{2.97}$$

Now take row (3) from matrix (2.95) $\Rightarrow$

$$k_1 = \frac{q_1 t_1 - m_1}{X} \tag{2.98}$$

Usually there is more interest in mean speed than in density

$$u_1 = \frac{q_1}{k_1} = \frac{q_1 X}{q_1 t_1 - m_1} = \frac{X}{t_1 - m_1/q_1} \tag{2.99}$$

From the latter equation it can be seen that if the MO drives with the average speed of the stream, i.e. $m_1 = 0$, then u_1 reduces to X/t_1, i.e. the mean speed of the MO. If $m_1 > 0$, i.e. MO is slower than the stream, then $u_1 > X/t_1$.

Symmetry $\Rightarrow$

$$u_2 = \frac{X}{t_2 - m_2/q_2} \tag{2.100}$$

Example 22 *Students of the IHE practice the MO-method annually on the Erasmusweg in Den Haag. This road section has two lanes in each direction and is rather narrow with vehicles parked on the roadway. The road section considered has a length of 1.8 km and contains a signalised intersection where intensity changes little. Table 2.4 shows the variation of results of four MO-teams, with and without the time the MO stopped at the intersection included. Fig. 2.32 shows a pretty large variation of the calculated mean speed over 48 MO measurements. In general the MO method requires many repetitions of the trip. The main advantage of the method is that no special equipment is needed. It can be a suitable method for pilot studies.*

MO-team	With stopped time			Without stopped time		
	Q (veh/h)	U (km/h)	K (veh/km)	Q (veh/h)	U (km/h)	K (veh/km)
1	713	44	17	835	51	17
2	630	45	14	715	51	14
3	652	44	15	754	51	15
4	746	49	15	822	53	16

Table 2.4: Results MO-method

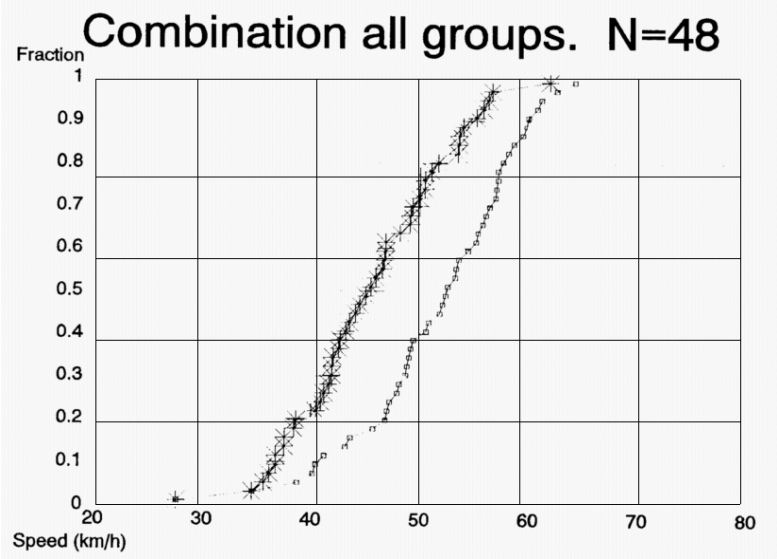

Figure 2.32: Distribution of mean speeds from 48 moving observer trips. Left: 'with stopped time'; Right: 'without stopped time'

2.10.1 Speed distribution as seen by a moving observer

Suppose we have a one directional traffic flow in a homogeneous and stationary state with space speed probability density function (p.d.f.) $f_M(v)$. A moving observer (MO) drives in this traffic flow with a constant speed v_0 and observes the absolute value of the speed of all vehicles that overtake, either active or passive. An example could be a police car measuring these speeds automatically with a radar speedometer. In fact such a speedometer measures the relative speed but we assume that the speed of the police car is added to it.

The question is: how does the p.d.f. of these observed speeds relate to the space p.d.f. of speeds. One point is obvious: the p.d.f. is zero at $v = v_0$ because a vehicle does not encounter vehicles with the same speed, but the rest of the p.d.f. has to be derived.

Proof. Consider a 'class' of vehicles with speed $v < v_0$, with density $k(v)$, and intensity $q(v)$. During period T, covering a distance $X = v_0T$, the MO makes n_1 active overtakings (see Fig. 2.30).

$$n_1(v) = k(v)X - q(v)T = k(v)v_0T - k(v)vT = k(v)T(v_0 - v) \text{ for } v < v_0 \tag{2.101}$$

With respect to vehicles with a speed $v > v_0$, the MO makes n_2 passive overtakings (see Fig. 2.31).

$$n_2(v) = q(v)T - k(v)X = k(v)vT - k(v)v_0T = k(v)T(v - v_0) \text{ for } v > v_0 \tag{2.102}$$

Eqns.(2.101) and (2.102) can be combined into:

$$n(v) = k(v)T|v - v_0| \tag{2.103}$$

The vehicle density as function of speed equals the total vehicle density times the p.d.f. of space speeds $\Rightarrow$

$$n(v) = kf_M(v)T|v - v_0| \tag{2.104}$$

Hence a MO 'sees' vehicles with speed v with an intensity:

$$n(v)/T = kfM(v)|v - v_0| \tag{2.105}$$

The p.d.f. of speeds the MO sees, is the term (2.105) divided by the total intensity, which is the integral of (2.105) over speed:

$$g(v) = \frac{f_M(v)|v - v_0|}{\int f_M(v)|v - v_0|dv} \tag{2.106}$$

■

Example 23 *Fig. 2.33 depicts: 1) $f_M(v)$, a Gaussian p.d.f. with mean of 115 km/h and standard deviation of 15 km/h; (the values are from a measurement at a Dutch freeway with a speed limit of 120 km/h); 2) the corresponding $f_L(v)$; which is not a Gaussian p.d.f. but deviates only little from it at the free flow situation; 3) the p.d.f. as seen from a MO with a speed of 120 km/h. From this figure can be understood that fast drivers often claim that so many drivers drive even faster. The vehicles that overtake them have a mean speed of 136 km/h, whereas the mean speed of all speeds > 120 km/h is 130 km/h.*

2.10.2 Fraction of drivers that is speeding

Just like the p.d.f., the fraction of drivers that is speeding, i.e., has a speed higher than the speed limit, depends on the method of observation

- At a spot:

$$F_L(v) = \int_{v=v_{\text{lim}}}^{\infty} f_L(v)dv \tag{2.107}$$

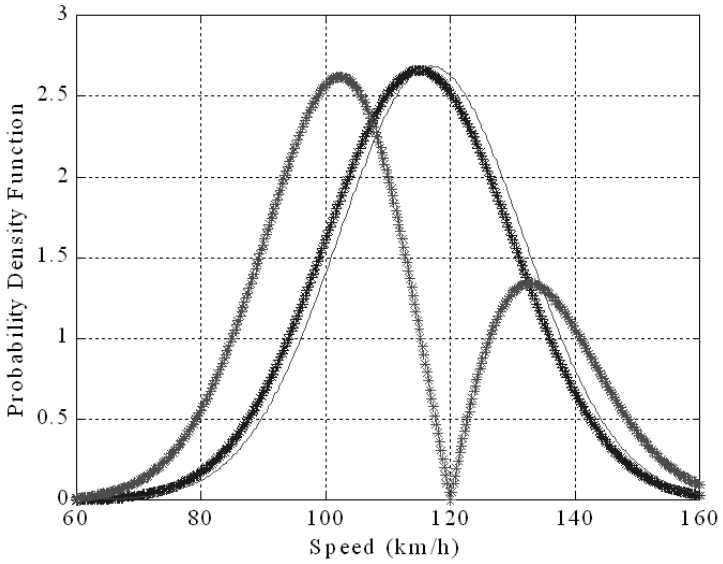

Figure 2.33: Probability density function for speed: local, instantaneous and as seen by moving observer.

- From an aerial photo (instantaneous observation):

$$F_M(v) = \int_{v=v_{\lim}}^{\infty} f_M(v)dv \tag{2.108}$$

- From a MO:

$$F_{MO}(v) = \int_{v=v_{\lim}}^{\infty} g(v)dv \tag{2.109}$$

Example 24 *Assuming a speed limit of 120 km/h, u_M = 115 km/h and STD = 15 km/h (as in the example discussed before) outcomes are: $F_L = 0.40$ and $F_M = 0.34$. Thus we see again that a sample of speeds at a spot has relatively more high speeds than a sample at a moment. The fraction FMO is a function of the speed; at speed 0 it equals the local value and at high speeds it approaches the space value (and reaches it at speed infinity). In between the function reaches a maximum value (0.64 in the example). Hence, naive policeman might think the fraction of speed offenders is very high.*

- The fraction F_q = (relative intensity) / (absolute intensity); it reaches a minimum if v_0 equals the space mean speed of the flow.

- The fraction $F_{MO}F_q$ i.e. the number of offenders a MO sees, divided by the absolute intensity. This function shows clearly that from a MO one sees a high fraction of offenders but the absolute number is relatively small. In fact the highest number of speed tickets (per time) can be obtained with a stationary observation method.

2.11 Summary of main definitions and terminology

Tab. 2.5 gives an overview of the definitions of intensity, density and mean speed. Tab. 2.6 shows an overview of the Dutch, German and English terminology used in literature.

	Spot observations	Aerial photos	Definitions Edie	Continuous variables
Variable	Location (x); period (T)	Section (X); moment (t)	Section (X); period (t)	Location (x); moment (t)
Intensity q (veh/h)	$q = \frac{n}{T}$	$q = u_M$	$q = \frac{\sum_i x_i}{XT}$	$q = \frac{\partial \tilde{N}(x,t)}{\partial t}$
Density k (veh/km)	$k = \frac{q}{u_M}$	$k = \frac{m}{X}$	$k = \frac{\sum_i t_i}{XT}$	$k = -\frac{\partial \tilde{N}(x,t)}{\partial x}$
Mean speed u_M (km/h)	$u_M = \frac{n}{\sum_i v_i}$	$u_M = \frac{\sum_i v_i}{m}$	$u_M = \frac{q}{k}$	$u(x,t) = \frac{q(x,t)}{k(x,t)}$

Table 2.5: Overview definitions of intensity, density and mean speed

Terms in notes	Dutch	German	English
intensity[3]	intensiteit	Stärke für $T = 0$; Intensität für $T > 0$	flow (UK); volume (USA); rate of flow for $T < 1h$ (USA)
density	dichtheid	Dichte für $X = 0$; Konzentration für $X > 0$	density (concentration)
local mean speed	gemiddelde lokale snelheid	Mitterwert lokaler Geschwindigkeiten	time-mean speed or average spot speed
instantaneous mean speed	gemiddelde momentane snelheid	Mittelwert momentaner Geschwindigkeiten	space-mean speed
stationary	stationair	stationär über Zeit	stationary, steady, time-homogeneous
homogeneous	homogeen	stationär über Weg	homogeneous

Table 2.6: Overview of terminology used in literature

Chapter 3

Microscopic flow characteristics

Contents of the chapter. This chapter describes several (distribution) models that describe (the relation between) different microscopic traffic variables (such as time headways, distance headways, etc.). We briefly discuss stochastic arrival processes, headway distributions, and individual speed distributions.

List of symbols

k	veh	arrivals
μ	-	mean
σ	-	standard deviation
h	s	time headway
q	veh/s	traffic intensity
$P(h)$	-	probability distribution function
$p(h)$	-	probability density function of time headway
ϕ	-	fraction of constrained vehicles in composite headway models

3.1 Arrival processes

Intensity varies more, also if the traffic flow is stationary, if the period over which it is observed is smaller. This is due too the fact that the passing of vehicles at a cross-section is to a certain extent a matter of chance. When using shorter observation periods, the smoothing of these random fluctuations reduces. For some practical problems it is useful to know the probability distribution of the number of vehicles that arrive in a short time interval.

Example 25 *The length of an extra lane for left-turning vehicles at an intersection must be chosen so large that in most cases there will be enough space for all vehicles, to prevent blocking of through going vehicles. In such a case it is not a good design practice to take account of the mean value or the 50 percentile. It is better to choose e.g. a 95 percentile, implying that only in 1 out of 20 cases the length is not sufficient.*

Several models describe the distribution of the number of vehicles arriving in a given, relatively short, period. We will discuss three of them.

Poisson-process

If drivers have a lot of freedom they will behave independently of each other. This implies that the passing of a cross-section is a pure random phenomenon. In general this will be the case if there is relatively little traffic present (a small intensity and density) and if there are no upstream 'disturbances', such as signalised intersections, that result in a special ordering of the

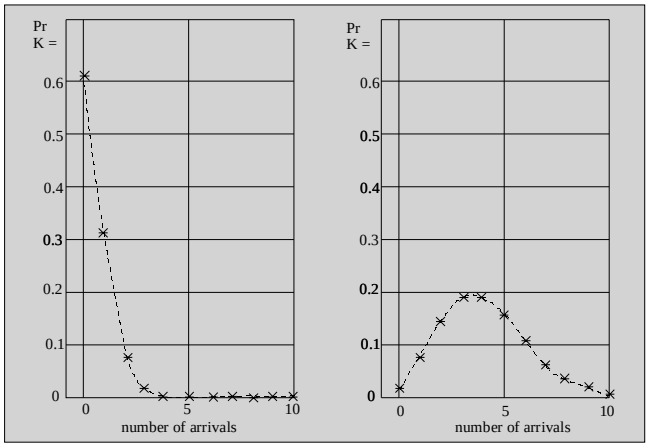

Figure 3.1: Probability function of Poisson with $\Delta t = 20s$; left $q = 90veh/h$ ($\mu = 0.5$); right $q = 720veh/h$ ($\mu = 4$).

vehicles in the stream. The conditions mentioned lead to the so called *Poisson-process*. The probability function of the number of arrivals k is given by:

$$\Pr\{K = k\} = \frac{\mu^k e^{-\mu}}{k!} \quad \text{for} \quad k = 0, 1, 2, ... \tag{3.1}$$

This probability function has only one parameter, the mean μ. Note that the parameter μ need not to be an integer. Fig. 3.1 shows an example of the Poisson probability function at two intensities.

Example 26 *Intensity = 400 veh/h and one wants to know the number of arrivals in a period of 30 s. Then we have: μ = (400/3600)·30 = 3.33 veh.*

A special property of the Poisson-distribution is that the *variance equals the mean*. This property can be used to test in a simple way if the Poisson-process is a suitable model: from a series of observations, one can estimate the mean and the variance. If the variance over the mean does not differ too much from 1, then it is likely that Poisson is an adequate model.

Binomial-process

When the intensity of a traffic stream increases, more and more vehicles form *platoons* (clusters, groups), and the Poisson-process is no longer valid. A model that is suitable for this situation is the model of a so-called *Binomial-process*, with probability function:

$$\Pr\{K = k\} = \left(\begin{array}{c} n \\ k \end{array} \right) p^k (1-p)^{n-k} \quad \text{for} \quad k = 0, 1, ..., n \tag{3.2}$$

Note that in this case, the variance over the mean is smaller than 1. The binomial distribution describes the number of 'successes' in n independent trials, at which the probability of success per trial equals p. Unfortunately this background does not help to understand why it fits the arrival process considered here.

Negative-Binomial-process

In discussing the Poisson-process it has been mentioned that it does not fit the situation *downstream of a signalised intersection*. When this is the case, one can state that high and low intensities follow each other. The variance of the number of arrivals then becomes relatively

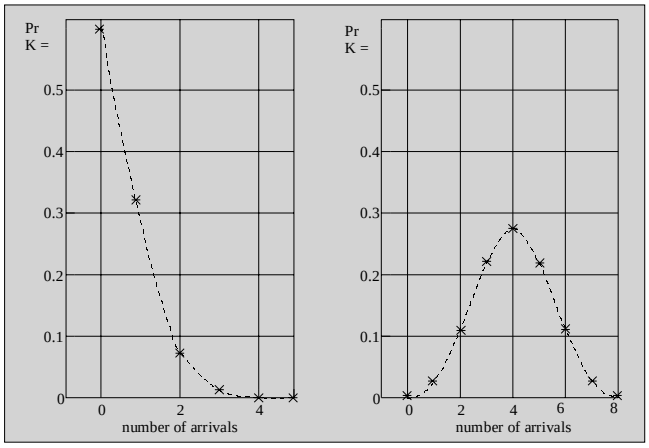

Figure 3.2: Binomial probability function at $\Delta t = 20s$; left $q = 90veh/h$ ($p = 0.1$, $n = 5$); right $q = 720veh/h$ ($p = 0.8,\ n = 8$).

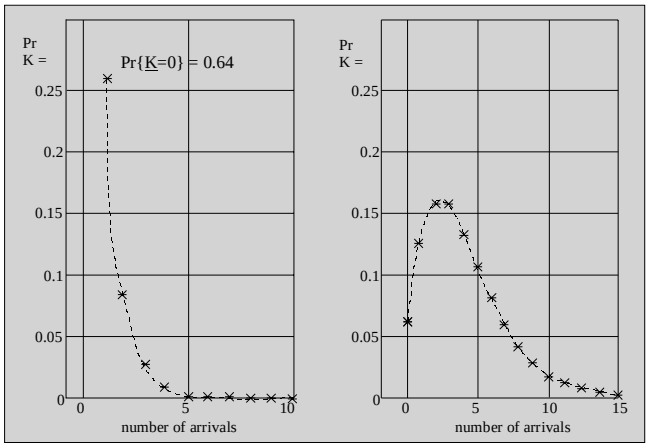

Figure 3.3:

large, leading to variance over mean being larger than 1. In that case the model of a *Negative-Binomial-process* is adequate :

$$\Pr\{K = k\} = \begin{pmatrix} k+n-1 \\ k \end{pmatrix} p^n (1-p)^k \quad \text{for} \quad k = 0, 1, 2, ... \tag{3.3}$$

As with the Binomial-process a traffic flow interpretation of the Negative-Binomial-process is lacking.

Distribution	**Mean**	**Variance**	**St.dev**	**Relative St.dev.**
Poisson	μ	μ	$\sqrt{\mu}$	$1/\sqrt{\mu}$
Binomial	np	$np(1-p)$	$\sqrt{np(1-p)}$	$\sqrt{(1-p)/np}$
Negative binomial	$n(1-p)/p$	$n(1-p)/p^2$	$p\sqrt{n(1-p)}$	$1/\sqrt{n(1-p)}$

Table 3.1: Mean, variance, standard deviation, and, relative standard deviation of the three arrival distributions

3.1.1 Formulae for parameters, probability terms, and parameter estimation

Recursion formulae

These formulae allow to calculate in a simple way the probability of event $\{K = k\}$ from the probability of event $\{K = k - 1\}$.

- Poisson: $\Pr\{K = 0\} = e^{-\mu} \rightarrow \Pr\{K = k\} = \frac{\mu}{k} \Pr\{K = k - 1\}$ for $k = 1, 2, 3, ..$
- Binomial: $\Pr\{K = 0\} = (1 - p)^n \rightarrow \Pr\{K = k\} = \frac{p}{1-p} \frac{n-k+1}{k} \Pr\{K = k - 1\}$
- Neg. Bin.: $\Pr\{K = 0\} = p^n \rightarrow \Pr\{K = k\} = (1 - p) \frac{n+k-1}{k} \Pr\{K = k - 1\}$

Estimation of parameters

From observations one calculates the sample mean m and the sample variance s^2. From these two parameters follow the estimations of the parameters of the three probability functions:

- Poisson: $\hat{\mu} = m$
- Binomial: $\hat{p} = 1 - s^2/m$ and $\hat{n} = m^2/\left(m - s^2\right)$
- Neg. Bin.: $\hat{p} = m/s^2$ and $\hat{n} = m^2/\left(s^2 - m\right)$

3.1.2 Applications

Length of left-turn lane

One has to determine the length of a lane for left turning vehicles. In the peak hour the intensity of the left turning vehicles equals 360 veh/h and the period they are confronted with red light is 50 s. Suppose the goal is to guarantee that in 95 % of the cycles the length of the lane is sufficient. In 50 s will arrive on the average (50/3600) x 360 = 5 veh. It will make a difference which model one uses.

- Poisson: parameter $\mu = 5$ and all probabilities can be easily calculated.
- Binomial with $s^2 = 2.5$: $\hat{p} = (5 - 2.5)/5 = 0.5$ and $\hat{n} = 25/(5 - 2.5) = 10$
- Negative-Binomial with $s^2 = 7.5$: $\hat{p} = 5/7.5 = 0.667$ and $\hat{n} = 25/(7.5 - 5) = 10$

With these parameter values the probability functions and the distributions can be calculated; see Fig. 3.4.

From the distributions can be seen that the Binomial model has the least extreme values; it has a fixed upper limit of 10, which has a probability of only 0.001. Of the two other models the Negative-Binomial has the longest tail. From the graph can be read (rounded values) that the 95-percentile of Binomial equals 7; of Poisson it is 9 and of Negative Binomial it is 10. These differences are not large but on the other hand one extra car requires 7 to 8 extra m of space.

Probe-vehicles

Suppose one has to deduce the state of the traffic stream at a 2 km long road section from probe-vehicles that broadcast their position and mean speed over the last km. It is known that on the average 10 probe vehicles pass per hour over the section considered. The aim is to have fresh information about the traffic flow state every 6 minutes. The question is whether this is possible.

It is likely that probe vehicles behave independently, which implies the validity of the Poisson-distribution. Per 6 minutes the average number of probes equals: $(6/60) \cdot 10 = 1$

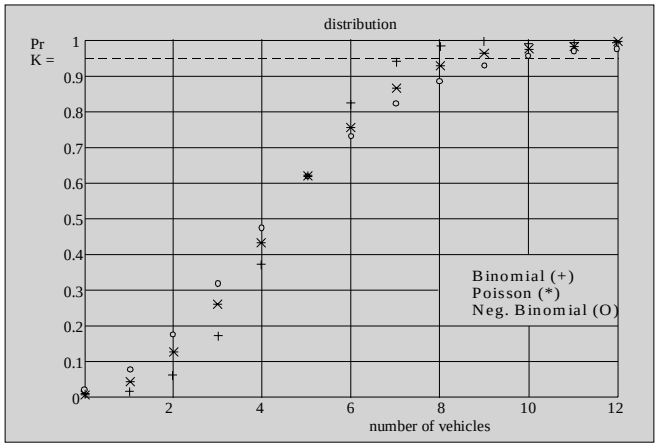

Figure 3.4: Distributions Binomial, Poisson and Negative-Binomial

probe. The probability of 0 probes in 6 minutes then equals: $e^{-1} = 0.37$ and this seems to be much too large for a reliable system.

The requirement is chosen less severe and one decides to update every 20 min. In that case

$$\hat{\mu} = (20/60) \cdot 10 = 3.33$$

and the probability of zero probes = exp[-3.33] = 0.036. This might be an acceptable probability of failure.

Remark 27 *In practice the most interesting point is probably whether the road section is congested. At congestion the probes will stay longer at the section and the frequency of their messages will increase.*

3.1.3 Choice of the appropriate model by using statistical testing

Earlier it has been mentioned that the quotient of sample variance and sample mean is a suitable criterion to decide which of the three distributions discussed, is a suitable model. Instead of using the rule-of-thumb, one may use statistical tests, such as the Chi-square test, to make a better founded choice.

3.2 Headway distributions

3.2.1 Distribution of headways and the Poisson arrival process

From the Poisson arrival process, one can derived a specific distribution of the headways: for a Poisson-process the number of arrivals k in an interval of length T has the probability function:

$$\Pr\{K = k\} = \frac{\mu^k e^{-k}}{k!} = \frac{(qT)^k e^{-qT}}{k!} \tag{3.4}$$

For $k = 0$ follows: $\Pr\{K = 0\} = e^{-qT}$. Remember: $0! = 1$ and $x^0 = 1$ for any x. $\Pr\{K = 0\}$ is the probability that zero vehicles arrive in a period T. This event can also be described as: the headway is larger than T. Consequently $\Pr\{K = 0\}$ equals the probability that a headway is larger than T. In different terms (replace period T by headway h)

$$\Pr\{H > h\} = S(h) = e^{-qh} \tag{3.5}$$

This is the so called *survival probability* or the *survival function S(h)*: the probability that a stochastic variable H is larger than a given value h. On the other hand, the complement is the

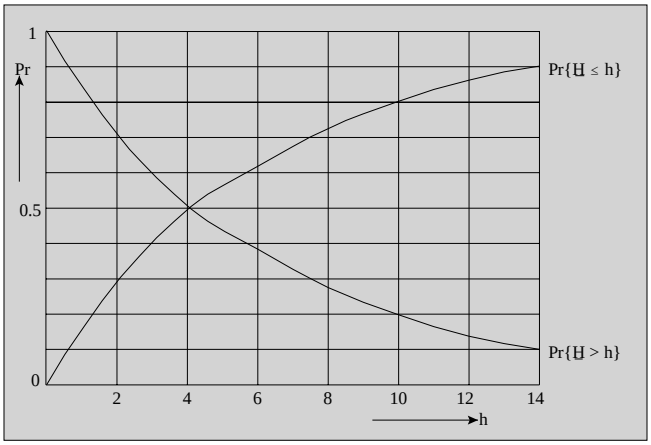

Figure 3.5: Survival function $S(h)$ and distribution function $P(h)$ of an exponential distribution

probability that a stochastic variable is smaller than (inclusive equal to) a given value, the so called distribution function. Consequently, the distribution function of headways corresponding to the Poisson-process is:

$$\Pr\{H \leq h\} = P(h) = 1 - e^{-qh} \tag{3.6}$$

This is the so called *exponential distribution function*. The survival function and the distribution function are depicted in Fig. 3.5 for an intensity $q = 600veh/h = 1/6veh/s$. Consequently the mean headway in this case is 6 s.

From the distribution function $P(h) = \Pr\{H \leq h\}$ follows the probability density function (p.d.f.) by differentiation:

$$p(h) = \frac{d}{dh}P(h) = qe^{-qh} \tag{3.7}$$

The mean value becomes (by definition):

$$\mu = \int_0^\infty hp(h)\,dh = \int_0^\infty hqe^{-qh}dh = \frac{1}{q} \tag{3.8}$$

We see that the mean (gross) headway μ equals the inverse of the intensity q. The variance σ^2 of the headways is:

$$\sigma^2 = \int_0^\infty (h-\mu)^2 p(h)\,dh = \frac{1}{q^2} \tag{3.9}$$

Consequently the variation coefficient, i.e. the standard deviation divided by the mean, equals 1. Note the difference:

- Poisson distribution → Var/mean = 1
- Exponential distribution → St.dev./mean = 1

3.2.2 Use of the headway distribution for analysis of crossing a street

The sequence of headways is governed by a random process. When this process is analysed two points need to be well distinguished. Namely, this process can be considered:

- per event;
- as a process in time.

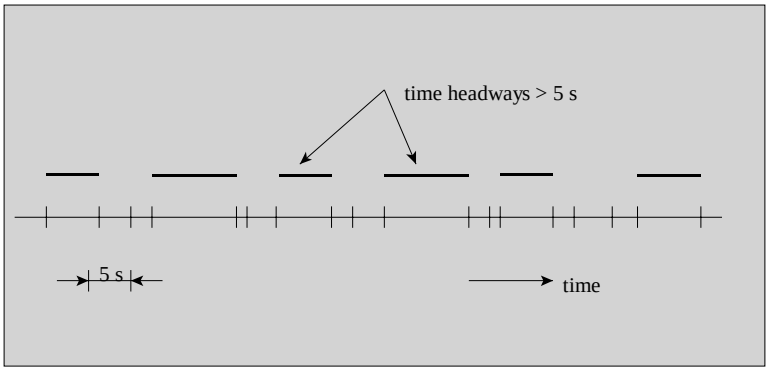

Figure 3.6: Realisation of sequence of gaps with gaps > 5s marked

Example 28 *Suppose only headways of 5 and 10 s are present, each with a probability of 0.5, and also assume that the headways are (stochastically) independent. This means that the events 'headway = 5 s' and 'headway = 10 s' occur on the average with the same frequency and in an arbitrary (random) sequence. Now consider a time axis on which the passage moments of vehicles are indicated, and by that also the gross headways. Then, on average, 5/15 = 1/3 of the time will be 'occupied' by headways of 5 s and 2/3 of the time by headways of 10 s. If one picks an arbitrary moment of the time axis, the probability to hit a 5 s headway is 1/3. This is not equal to the probability that a headway of 5 s occurs, which equals 0.5.*

Suppose a pedestrian wants to cross a road and needs a headway of at least x seconds. *Terminology*: in the context of crossing a street the headway is called a *gap*; the traffic flow offers gaps to the pedestrian, which he/she can either accept or reject (more about gap-acceptance will be discussed in Chapter 12).

The first idea could be that the crossing possibilities are determined by the *frequency of the gaps* larger than x s. However, it is more relevant to consider the *fraction of time* $G_1(x)$ for which the gaps are larger than x seconds? To analyse this, we start from a realisation of the gap offering process on a time axis (see Fig. 3.6). The required fraction G_1 for a given realisation of the process equals the sum of all gaps $> x$ divided by the sum of all gaps for (= period considered).

$$G_1(x) = \frac{\sum h_i | h_i > x}{\sum h_i} = \frac{\frac{1}{n}\sum h_i | h_i > x}{\frac{1}{n}\sum h_i} \tag{3.10}$$

In the nominator of Eq. (3.10) all gaps $> x$ s are summed and in the denominator all complete gaps are summed. If we consider many repetitions of this process, the required fraction $G_1(x)$ is a (mathematical) expectation:

$$G_1(x) = \frac{\int_x^\infty hp(h)\,dh}{\int_0^\infty hp(h)\,dh} = q\int_x^\infty hp(h)\,dh \tag{3.11}$$

The simplification in (3.11) is possible using: mean gap $= 1/q$. For an *exponential distribution* with probability density function $p(h) = e^{-qh}$, one can derive:

$$G_1(x) = e^{-qx}(1+qx) \tag{3.12}$$

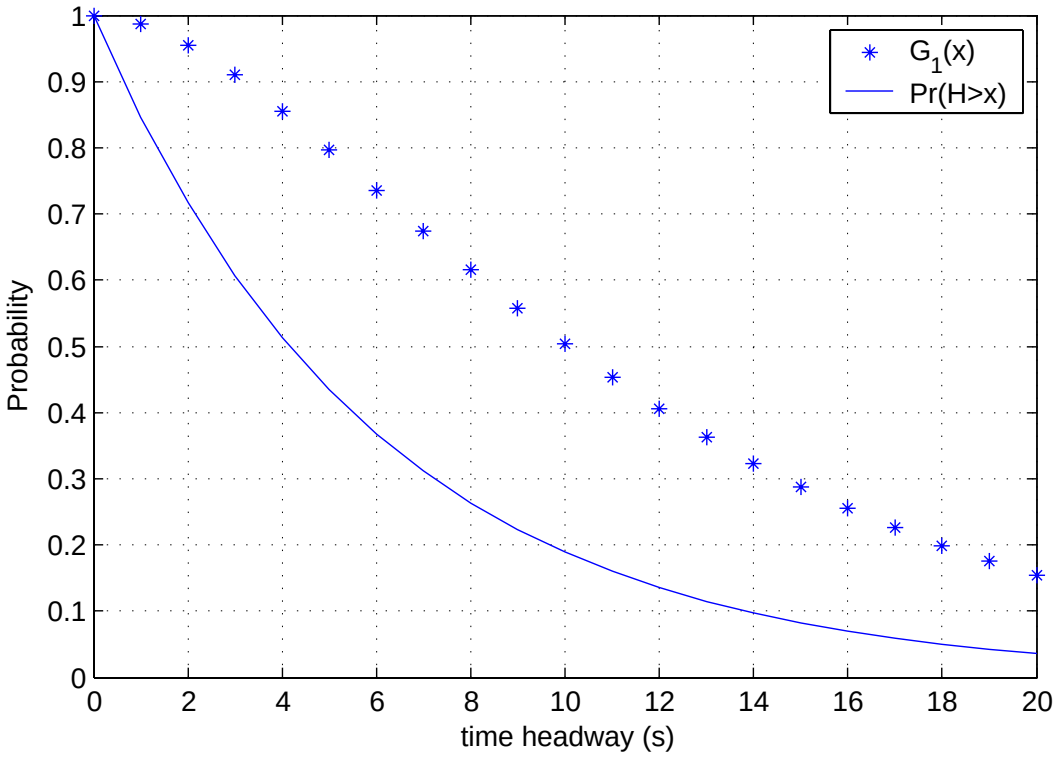

Figure 3.7: Survival function $S(x) = \Pr\{H > x\}$ and fraction of time that $H > x = G_1(x)$ for an exponential distribution of gaps at intensity $q = 600veh/h$.

Proof.

$$q\int_x^\infty hqe^{-qh}dh = q^2\int_x^\infty he^{-qh}dh = q^2\int_x^\infty \frac{h}{-q}d\left(e^{-qh}\right) \tag{3.13}$$

$$= -q\left(he^{-qh}\Big|_x^\infty - \int_x^\infty e^{-qh}dh\right) \tag{3.14}$$

$$= -q\left(he^{-qh}\Big|_x^\infty + \frac{1}{q}e^{-qh}\Big|_x^\infty\right) \tag{3.15}$$

$$= qxe^{-qx} + e^{-qx} = e^{-qx}(1+qx) \tag{3.16}$$

■

From the illustration in Fig. 3.7 follows that the fraction of time a gap is larger than x, is larger than the frequency a gap larger than x occurs.

Fraction of time a 'rest gap' is larger than x

If we analyse the crossing process more precisely, it turns out that for the crossing pedestrian, it is not sufficient to arrive in a gap larger than x s. Rather, he/she should arrive *before the moment the period until the next vehicle arrives, equals x s.* In other words, the so called *restgap* should be larger than x s. To calculate the fraction of time that a restgap is larger than x s, denoted as $G_2(x)$, we consider again a time axis with a realisation of the gap process; see Fig. 3.8.

$$G_2(x) = \frac{\sum(h_i - x)\,|h_i > x}{\sum h_i} = \frac{\frac{1}{n}\sum(h_i - x)\,|h_i > x}{\frac{1}{n}\sum h_i} \tag{3.17}$$

$$= \frac{\int_x^\infty (h-x)\,p(h)\,dh}{\int_0^\infty (h-x)\,p(h)\,dh} = q\int_x^\infty (h-x)\,p(h)\,dh \tag{3.18}$$

For an exponential gap distribution it can be derived that:

$$G_2(x) = e^{-qx} \tag{3.19}$$

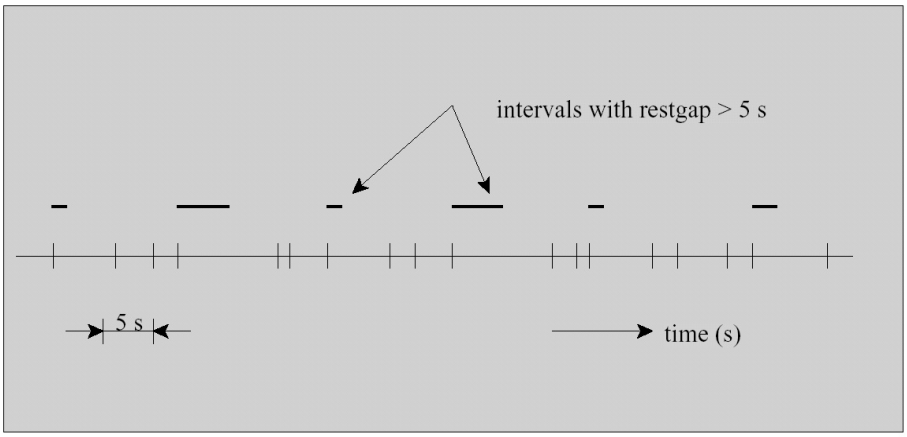

Figure 3.8: Realisation of sequence of gaps with intervals marked restgap > 5 s

Proof.

$$G_2(x) = q\int_x^\infty (h-x)\, qe^{-qh}dh = (\text{set } y = h-x)\; q\int_0^\infty yqe^{-q(y+x)}dy \tag{3.20}$$

$$= qe^{-qx}\int_0^\infty yqe^{-qy}dy = e^{-qx} \tag{3.21}$$

■

Remark 29 *Consequently the probability that a restgap is larger than x equals the probability that a headway, right away, is larger than x. This is a very special property of the exponential distribution, in fact a unique property, sometimes phrased as: 'the exponential process has no memory'.*

3.2.3 Use of headway distribution to calculate the waiting time or delay

If the crossing pedestrian arrives at a moment at which the restgap is too small, how long does he/she have to wait until the next vehicle arrives? Of course no more than x s, but more interesting is: how long will this take on average? This equals the mean of a drawing from a headway probability density $p(h)$ under the condition it is less than x:

$$\delta = \frac{\int_0^\infty hp(h)\,dh}{\int_0^x p(h)\,dh} \tag{3.22}$$

For an exponential distribution it can be derived that:

$$\delta = \frac{1}{q} - \frac{xe^{-qx}}{1-e^{-qx}} \tag{3.23}$$

Proof. Numerator:

$$\int_0^x hqe^{-qh}dh = (\text{partial integration})\; \int_0^x -hd\left(e^{-qh}dh\right) \tag{3.24}$$

$$= -he^{-qh}\Big|_0^x + \int_0^x q\left(e^{-qh}\right)dh = -xe^{-qx} + \left(1-e^{-qx}\right)/q \tag{3.25}$$

Denominator:

$$\int_0^x qe^{-qh}dh = 1-e^{-qx} \tag{3.26}$$

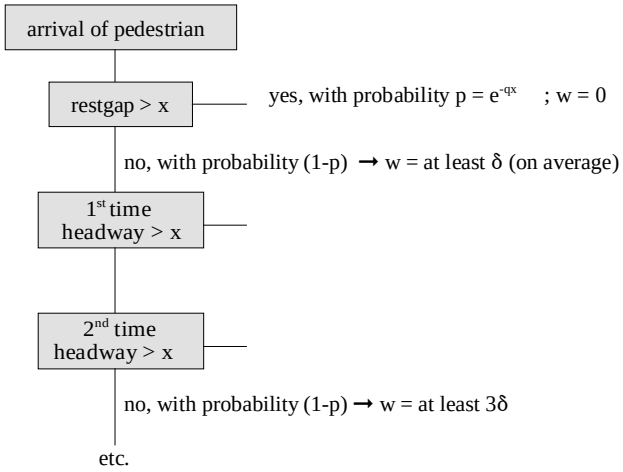

Figure 3.9: Flow diagram of waiting process

■

Next the pedestrian is offered the first full headway. If it is too small, then the waiting time is increased by δ; if it is large enough, then the waiting is over and the crossing can be carried out. The probabilities of both events are known. See Fig. 3.9 for a flow diagram of the process.

Every time a headway is too small, the waiting time is increased by (on average) δ, but the probability that this occurs becomes smaller and smaller; it decreases proportional to $(1-p)k$ where $k = 1, 2, 3, ...$ The expectation of the waiting time W, becomes (first term is p multiplied by zero):

$$E[W] = p \cdot 0 + (1-p)\, p\delta + (1-p)^2\, p2\delta + (1-p)^3\, p3\delta + ... \quad (3.27)$$

$$= p\delta \left[(1-p) + 2(1-p)^2 + 3(1-p)^3 + ... \right] \quad (3.28)$$

$$= \frac{\delta(1-p)}{p} \quad (3.29)$$

Proof. PM $r + r^2 + r^3 + ... = \frac{r}{1-r}$ (Geometric Series)

Differentiate and multiply by $r \rightarrow r\left(1 + 2r + 3r^2 + ...\right) = r\frac{d}{dr}\left(\frac{r}{1-r}\right)$

Thus $r + 2r + 3r + ... = \frac{r}{(1-r)^2}$ ■

For the exponential distribution we subsitute $p = e^{-qx}$, and δ according to Eq. (3.23). This leads to

$$E[W] = \left(e^{-qx} - qx - 1\right)/q \quad (3.30)$$

See Fig. 3.10 for an illustration with $E[W] = f(Q)$ for x = 4, 5, 6, 7 and 8 s.

3.2.4 Real data and the exponential distribution

In general the exponential distribution (ED) of the headways is a good description of reality at low intensities and unlimited or generous overtaking possibilities. If both conditions are not fulfilled, then there are interactions between the vehicles in the stream, leading to driving in platoons (also called clusters and groups).

In that case the ED is fitting reality badly; see Fig. 3.11. The minimum headways in a platoon are clearly larger than zero, whereas according to the ED the probability of extreme small headways is relatively large. The differences between the ED and reality have lead to the use of other headway models at higher intensities. We will first discuss some simple alternatives for the ED model in section 3.2.5. and after that more complex models, dividing the vehicles in two categories, in section 3.2.6.

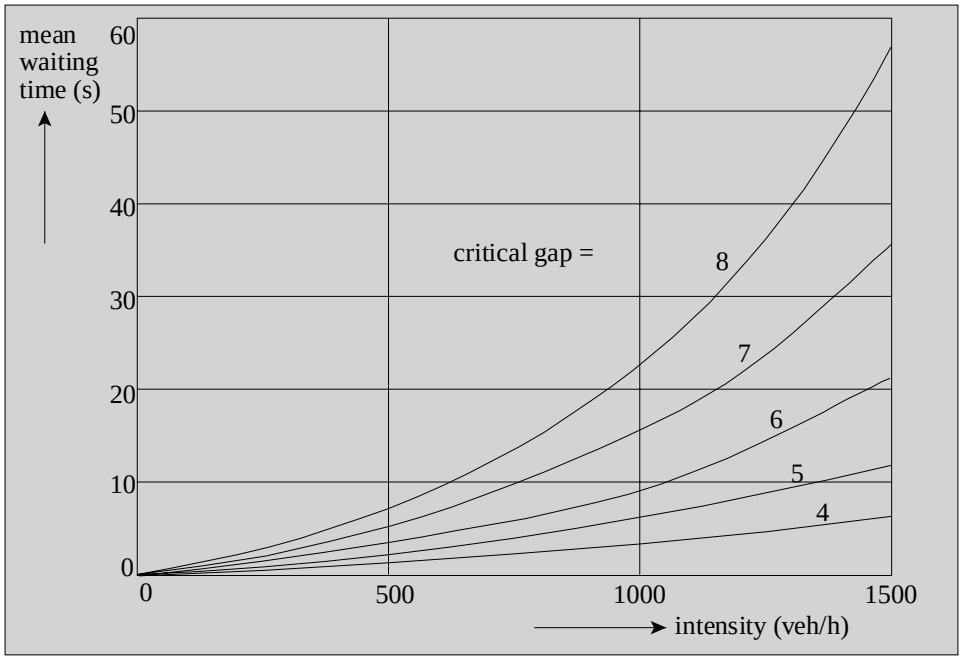

Figure 3.10: Mean delay as a function of the intensity at several critical gaps; gaps have an exponential distribution.

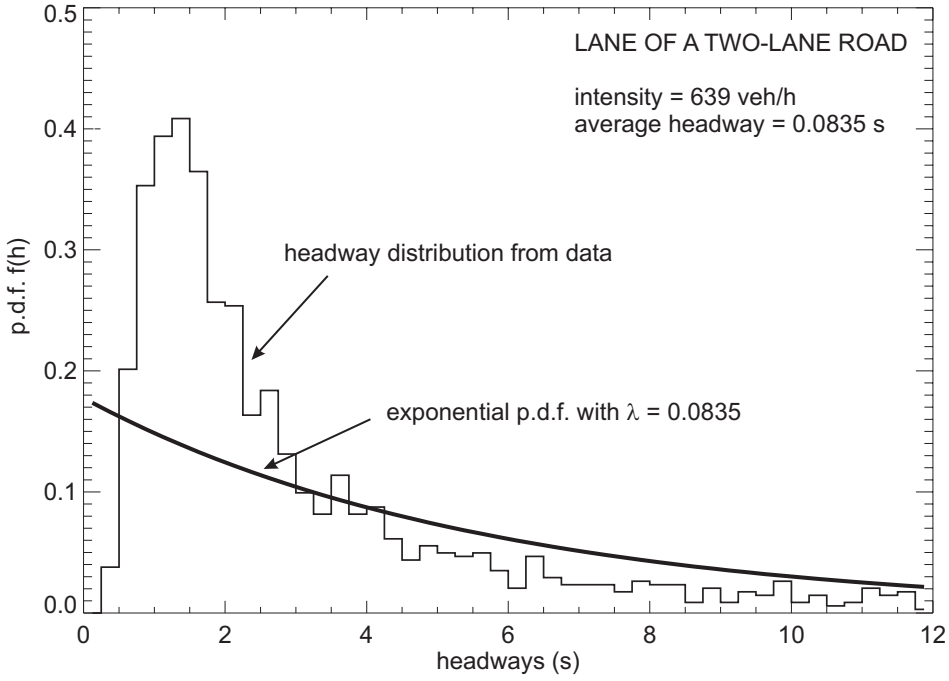

Figure 3.11: Histogram of observed headways compared to the exponential probability density function.

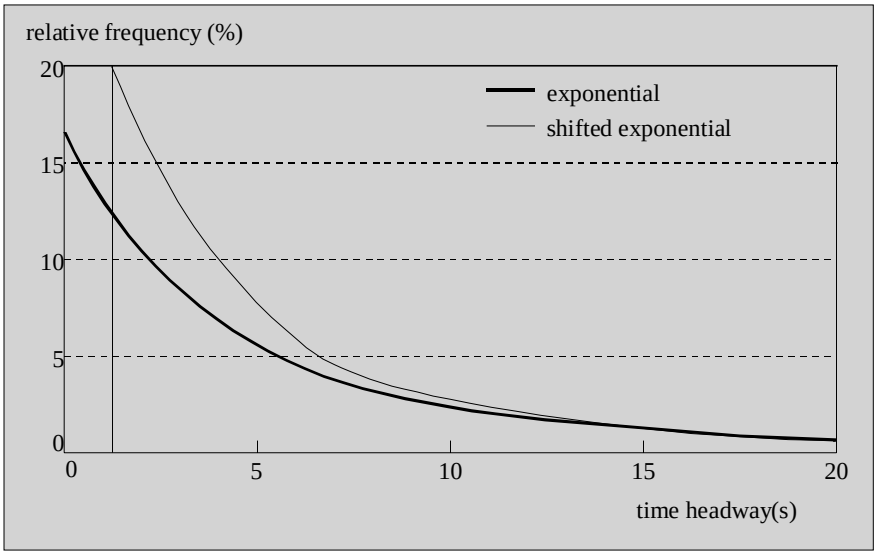

Figure 3.12: Non-shifted and shifted exponential probability density functions.

3.2.5 Alternatives for the exponential distribution

- Shifted exponential distribution

The shifted exponential distribution is characterised by a minimum headway h_m, leading to the distribution function:

$$\Pr\{H \leq h\} = 1 - e^{-\lambda(h-h_m)} \quad \text{with} \quad \lambda = \frac{q}{1-qh_m} \quad \text{for} \quad h \geq h_m \tag{3.31}$$

and the probability density function:

$$p(h) = 0 \quad \text{for} \quad h < h_m \tag{3.32}$$

$$p(h) = \lambda e^{-\lambda(h-h_m)} \quad \text{for} \quad h \geq h_m \tag{3.33}$$

The mean value is

$$\mu = h_m + \frac{1}{\lambda} \tag{3.34}$$

the variance equals:

$$\sigma^2 = \left(\frac{1}{\lambda}\right)^2 \tag{3.35}$$

and the variation coefficient

$$\frac{\sigma}{\mu} = \frac{1}{1+\lambda h_m} \tag{3.36}$$

which is always smaller than 1 (recall that for the ED, $\frac{\sigma}{\mu} = 1$) as long as $h_m > 0$. Fig. 3.12 depicts a non-shifted and a shifted probability density with the same mean value ($\mu = 6$ s) and a minimum headway h_m= 1 s.

In practice it is difficult to find a representative value for h_m ; moreover the abrupt transition at $h = h_m$ does not fit reality very well; see Fig. 3.11.

- Erlang distribution

A second alternative for the ED model is the Erlang distribution with a less abrupt function for small headways.

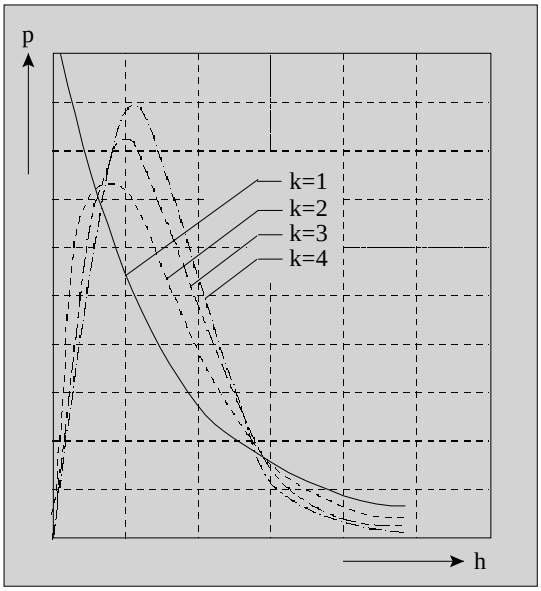

Figure 3.13: Erlang probability densities

The Erlang distribution function is defined by

$$\Pr\{H \leq h\} = 1 - e^{-kh/\mu} \sum_{i=0}^{k} \left(\frac{kh}{\mu}\right)^i \left(\frac{1}{i!}\right) \tag{3.37}$$

and the corresponding probability density:

$$p(h) = \frac{h^{k-1}}{(k-1)!} \left(\frac{k}{\mu}\right)^k e^{-kh/\mu} \tag{3.38}$$

Note that for $k = 1$ we have $p(h) = (1/\mu)e^{-h/\mu}$. Consequently the *exponential distribution is a special Erlang distribution* with parameter $k = 1$. For values of k larger than 1 the Erlang p.d.f. has a form that better suits histograms based on observed headways. This is illustrated by Fig. 3.13. which shows densities for $\mu = 6$ s and $k = 1$, 2, 3 and 4.The mean of an Erlang distribution $= \mu$, the variance equals μ^2/k and consequently the coefficient of variation equals $\frac{1}{\sqrt{k}}$ which is smaller or equal to 1.

- Lognormal distribution

A second alternative headway distribution is the lognormal distribution.

Definition 30 *A stochastic variable X has a lognormal distribution if the logarithm $ln(X)$ of the stochastic variable has a normal distribution.*

This can be applied as follows: if one has a set of observed headway h_i, then one can investigate whether $x_i = \log h_i$ has a normal distribution. If this is the case, then the headways themselves have a lognormal distribution. The p.d.f. of a lognormal distribution is:

$$p(h) = \frac{1}{h\sigma\sqrt{2\pi}} e^{-\frac{\ln^2\left(\frac{h}{\mu}\right)}{2\sigma^2}} \tag{3.39}$$

with mean μ^*

$$E(H) = \mu^* = \mu e^{\frac{1}{2}\sigma^2} \tag{3.40}$$

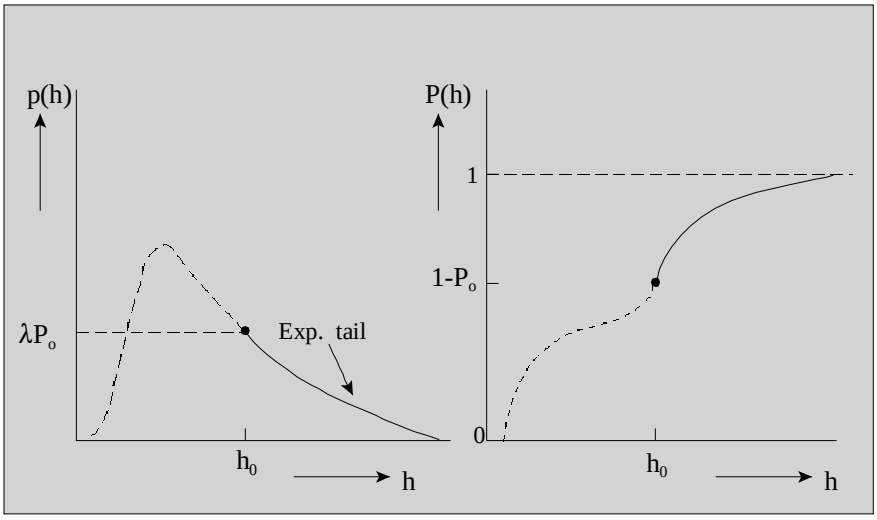

Figure 3.14: Probability density function and distribution of headways according to Exponential Tail Model

and variance $(\sigma^*)^2$

$$var\,(H) = (\sigma^*)^2 = \mu^2 e^{\sigma^2}\left(e^{\sigma^2} - 1\right) \tag{3.41}$$

The coefficient of variation thus equals

$$C_v = \frac{\sigma^*}{\mu^*} = \frac{\mu e^{\frac{1}{2}\sigma^2}\sqrt{\left(e^{\sigma^2} - 1\right)}}{\mu e^{\frac{1}{2}\sigma^2}} = \sqrt{e^{\sigma^2} - 1} \tag{3.42}$$

In contrast to the previous discussed distributions, the coefficient of variation of the lognormal distribution, can be smaller as well as larger than 1.

Remark 31 *The parameters of the p.d.f., μ and σ are not the mean and st. dev. of the lognormal variate* ***but of the corresponding normal variate****. If μ^* and $(\sigma^*)^2$ are given, μ and σ^2 follow from:*

$$\mu = \frac{\mu^*}{\sqrt{1 + C_v^2}} \quad and \quad \sigma^2 = 2\ln\left(\sqrt{1 + C_v^2}\right) \tag{3.43}$$

In fact all alternatives for the exponential headway distribution discussed so far, are based on selecting a p.d.f. in which small headways have a small frequency and there is a long tail to the right.

Exponential tail model For estimating capacity we need a model for the p.d.f. of the empty zone. In contrast to that only large headways are relevant for the overtaking opportunities on two-lane roads. All headway less than, say 10 s, are too small for an overtaking and the precise distribution of those headways is not relevant. However, it is relevant how the frequency of large headways depends on intensity. In preparing backgrounds for new design guidelines the Transportation Laboratory of TU Delft has developed a headway model that neglects the small headways, the so called Exponential Tail Model (ETM). It has the following p.d.f.:

$$p\,(h) = \begin{cases} \lambda P_0 e^{-\lambda(h-h_0)} & h \geq h_0 \\ \text{not defined} & h < h_0 \end{cases} \tag{3.44}$$

with parameter P_0 is the probability that $h > h_0$. The survival function is:

$$\Pr\{H > h\} = P_0 e^{-\lambda(h-h_0)} \quad \text{for} \quad h > h_0 \tag{3.45}$$

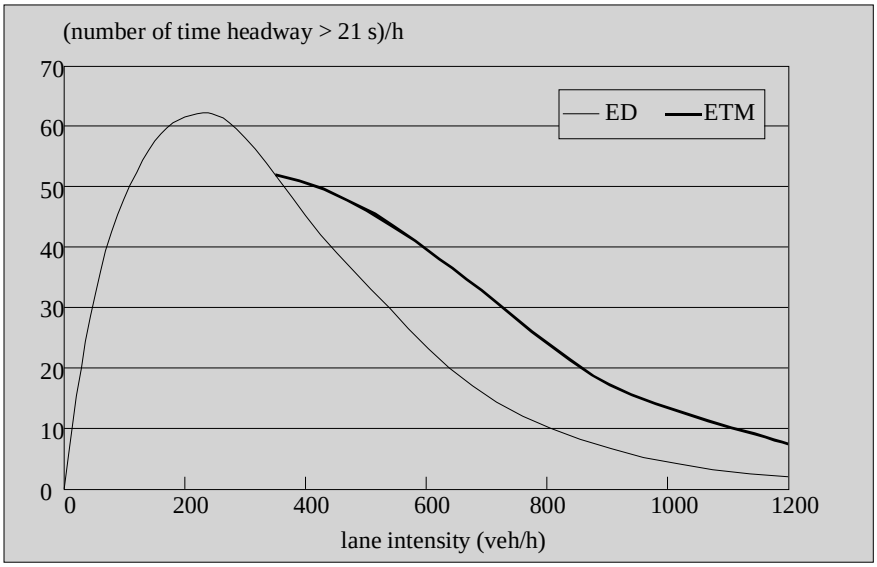

Figure 3.15: 'Intensity' of headways larger than 21 s as a function of lane intensity. according to exponential distributed headways (ED model) and according to exponential tail model (ETM).

The parameter h_0 has been set at 10 s, to be sure that the headways are from free drivers and consequently are very likely exponentially distributed. Based on observed headways at given intensities the two model parameters P_0 and λ can be estimated. In a second step of the analysis the relation between those parameters and intensity has been determined. The results were:

$$P_0 = -0.286 - 8.24q \tag{3.46}$$

$$\lambda = 0.0314 + 0.475q \tag{3.47}$$

Substitution of those two relations in eq. (3.45) leads to (for simplicity the logarithm of eq. (3.45) has been used):

$$\ln\left(\Pr\{H > h\}\right) = 0.028 - 3.49q - 0.0314h - 0.475qh \text{ (q in veh/s and h in s)} \tag{3.48}$$

Using this equation it is easy to calculate the probability of a certain (large) headway. In fact this is nearly as easy as when using the exponential distribution of headways, but it corresponds much better to reality.

See Fig. 3.15 for an illustration with a target headways of 21 s, which is a representative value for a gap that is acceptable for an overtaking at a two-lane road. It can be concluded that the new model offers much more overtaking opportunities than the exponential model.It should be added that the ETM is not a good description for intensities smaller than about 300 veh/h. However, for these intensities the simple exponential model is a good description. For both models it is a requirement that no substantial upstream disturbances (e.g. signalised intersections) have influenced the traffic stream.

3.2.6 Composite headway models: distributions with followers and free drivers

Comparisons of observed histograms of headways and the simple models discussed earlier have often lead to models badly fitting data. This has been an inspiration to develop models that have a stronger traffic behaviouristic background than the simple ones discussed above.

In so-called *composite headway models*, it is assumed that drivers that are obliged to follow the vehicle in front (because they cannot make an overtaking or a lane change), maintain a

certain minimum headway (the so-called *empty zone* or *following headway*). They are in a constrained or following state. If they have a headway which is larger than their minimum, they are called free drivers. Driver-vehicle combinations are thus in either of two states: free or constrained. As a result, the p.d.f. $p(h)$ of the headways has two components: a fraction ϕ of constrained drivers with p.d.f. $p_{fol}(h)$ and a fraction $(1-\phi)$ with $p_{free}(h)$

$$p(h) = \phi p_{fol}(h) + (1-\phi)\, p_{free}(h) \tag{3.49}$$

The remaining problem is how to specify the p.d.f. of both free drivers and constrained drivers. Different approaches have been presented in the past. Let us discuss the most important ones.

Composite headway model of the Branston type

Several theoretical models have been used to determine expressions for the free and the constrained headway distributions. The approach discussed here is relatively straightforward and was first proposed by [27]. It is based on the idea that the total headway H can be written as the sum of two other random variates, the empty zone X and the free headway U. The distribution of the empty zone X is described by the p.d.f. $p_{fol}(x)$. The p.d.f. describing the free headway is given by the following expression

$$p_{free}(u) = \phi\delta(u) + (1-\phi)\,\lambda e^{-\lambda h} \tag{3.50}$$

where $\delta(u)$ is the δ-dirac function defined by $\int f(y)\,\delta(y-x)\,dy = f(x)$. The term $\lambda e^{-\lambda h}$ again denotes the exponential distribution, which is valid for independent arrivals (free flow). The rational behind this expression is that drivers which are constrained (with probability ϕ) have a free headway which is equal to zero. Since the total headway H is the sum of X and U, it's p.d.f. $p(h)$ can be determined by convolution

$$\begin{aligned} p(h) &= \int p_{fol}(s)\, p_{free}(h-s)\, ds && (3.51)\\ &= \int p_{fol}(s)\left(\phi\delta(h-s) + (1-\phi)\,\lambda e^{-\lambda(h-s)}\right) ds && (3.52)\\ &= \phi p_{fol}(h) + (1-\phi)\,\lambda e^{-\lambda h}\int_0^h p_{fol}(s)\, e^{-\lambda s} ds && (3.53) \end{aligned}$$

which thus implies

$$p_{fol}(h) = p_{fol}(h) \text{ and } p_{free}(h) = \lambda \int p_{fol}(h-s)\, e^{-\lambda h} ds \tag{3.54}$$

Model Cowan type III [13]

Cowan's model [13] assumes that all constrained drivers have the same headway h_0 and all free drivers have a headway that is distributed according to a shifted exponential distribution (shifted with h_0).

$$p(h) = \phi\delta(h-h_0) + (1-\phi)\, H(h-h_0)\,\lambda e^{-\lambda(h-h_0)} \tag{3.55}$$

$$P(h) = H(h-h_0)\left(\phi - (1-\phi)\left(1-e^{-\lambda(h-h_0)}\right)\right) \tag{3.56}$$

with delta-function δ and Heaviside-step-function (or unit step function) H; see Fig. 3.16

Although this model is certainly not completely realistic (constrained vehicles do not all have the same minimum headway), it turns out to be a good description for a class of practical problems. The value of the minimum headway h_0 is in the order of 2 to 3 s.

If intensity increases, the fraction of constrained vehicles ϕ has to increase too and in the limiting case, i.e. when intensity reaches capacity, ϕ should have the value 1. The capacity than equals $1/h_0$.

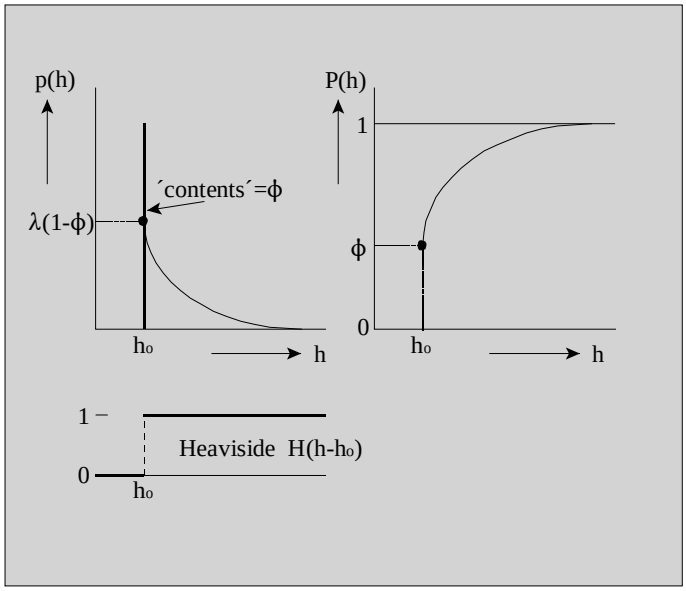

Figure 3.16: P.d.f. $p(h)$ and distribution function $P(h)$ of Cowan's model type III

If we assume further that with the increasing of ϕ, the parameter h_0 does not change, then it is possible to use this model for capacity estimation – as well as the other composite headway models presented in the remainder of this section. This could be carried out by estimating the parameter h_0 at traffic flow states at which capacity has not yet been reached.

Cowans model type III is not used for this purpose because it is too schematised for that procedure. We have only discussed this application to explain the principle of the capacity estimation method. The method is applied using the more detailed headway distribution model of Branston, discussed in the next session.

Model of Branston [7]

In Cowans model it was assumed that all constrained vehicles had the same headway. Branston assumes that the headways of constrained vehicles have their own p.d.f. and derives a model of which a main characteristic is that the p.d.f. of the free drivers is overlapping with the p.d.f. of the constrained drivers.

The model has a traffic flow background which was used in [7] to derive the model: consider a one-directional traffic stream on a roadway, which is so wide that practically speaking overtaking possibilities are unlimited. The headways, referring to the total roadway, then should have an exponential distribution.

Now consider this wide road narrowing to one lane; consequently overtaking possibilities become zero. Vehicles arriving at the transition sometimes have to wait until they can enter the one lane section. It is assumed that the *intensity is smaller than the capacity* of the one lane section, otherwise no equilibrium state is possible. For his situation one can derive (see [7]) a stochastic queueing model, that implies the following headway distribution:

$$p(h) = \phi p_{fol}(h) + (1-\phi)\,\lambda e^{-\lambda h} \int_0^h p_{fol}(\eta)\, e^{\lambda \eta} d\eta \tag{3.57}$$

with $p_{fol}(h)$ an arbitrary probability density function.

In this queueing model the parameter ϕ is not free but depends on $p_{fol}(h)$ and λ. Branstons trick has been to define ϕ as a free parameter, arguing that in this way a realistic headway distribution is described, because in reality overtaking possibilities are between zero and unlimited.

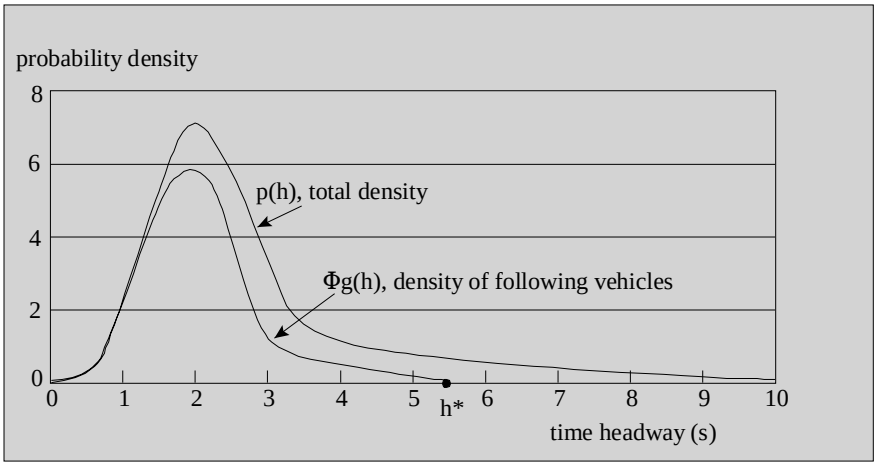

Figure 3.17: Branston's model for the probability density of headways

The p.d.f. of Branstons model has been depicted in Fig. 3.17. A special property of the model is that rather small headways can be free ones; this might occur in reality due to overtakings.

It is likely that $p_{fol}(h)$ has an upper limit, i.e. there are no constrained vehicles with a headway larger than, say, h^*; $p_{fol}(h) = 0$ for $h > h^*$. This implies that in that region the total p.d.f. has an exponential character, the so called exponential tail of the headway p.d.f. In fact parameter h^* is the most important parameter of the model because it is difficult to estimate.

Example 32 Estimating capacity with composite headway distribution models. *According to the principle explained in the preceding section the capacity can be estimated. To carry out this procedure the components of the headway model have to be estimated, based on a sample of observed headways. The procedure consists of two steps:*
Step 1. *Determining boundary value h^*. Parameter h^* is the headway where the exponential tail starts. To determine h^* the logarithm of the survival probabilities is plotted against the headway. As long as the tail is exponential, the result will be a straight line; because:*

$$P(h) = 1 - e^{-qh} \rightarrow S(h) = 1 - P(h) = e^{-qh} \rightarrow \log(S(h)) = -qh \tag{3.58}$$

Practical procedure: sort the headways in ascending order; result denoted as: $h_{(i)}$, $i = 1, 2, 3, ..., n$. Then $i/(n+l)$ is the estimated probability a headway is smaller than h_i and $1 - i/(n+1)$ is the estimated probability a headway is larger than h_i; the survival probability. Consider: $y_{(i)} = \ln[l - i/(n+l)]$ as a function of the headway h_i; see Fig. 3.18. The headway at which the function starts to deviate from a straight line is the boundary value h^*. Studies of the Transportation Laboratory of TU Delft, with data from two lane rural roads, have lead to values for h^* between 4 and 6 s. On motorways this value is smaller; drivers feel less constrained at the same small headway, because they have more lateral freedom to manoeuver.
Step 2: Estimation of the other parameters of Branston's model. *The other parameters ϕ, λ and $p_{fol}(h)$ can be estimated by new methods developed by [23] and [51]. Discussion of these methods falls outside the scope of this subject.*
Fig. 3.19 shows an example of the result of the method with data from a two-lane rural road. In general application of this method results in rather high capacity values. The conjecture is that drivers at an intensity of say, 1/2 capacity, have a different (a shorter) constrained headway than they have at capacity.

Buckley's model [10] One of the drawbacks of the Branston model is the fact that the total headway H is composed of the free headway U and the following headway X ($H = X+U$), where the former follows the exponential distribution. However, for independent arrivals - e.g. when

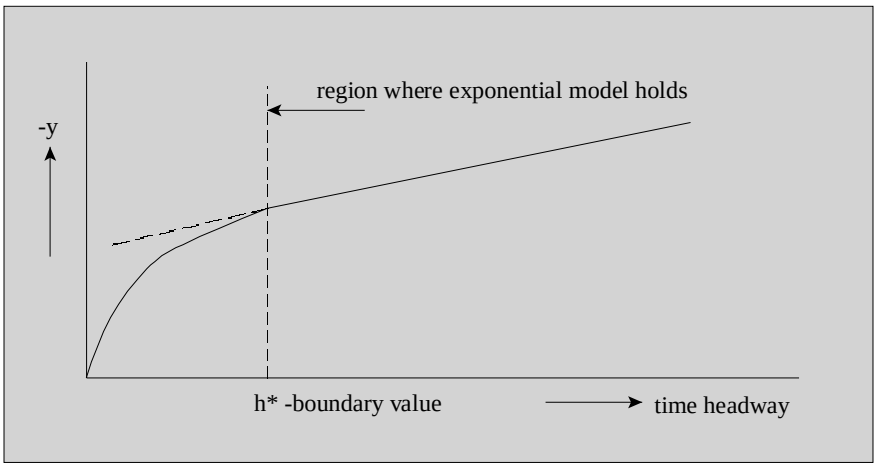

Figure 3.18: Method to determine headway value h^*

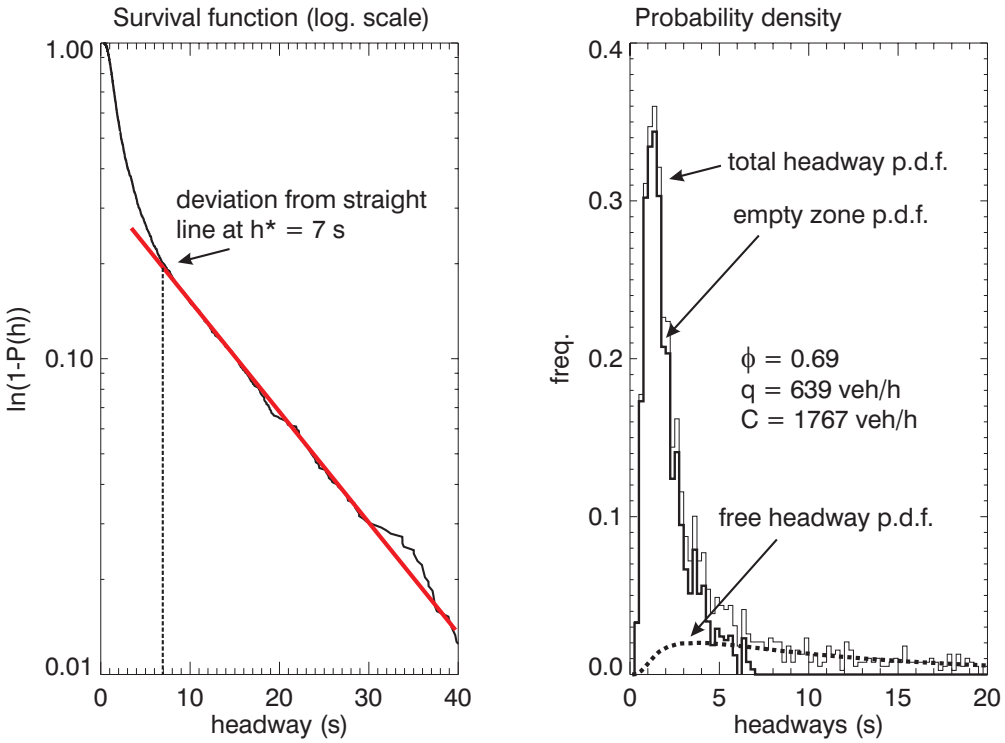

Figure 3.19: Results of application of capacity estimation.

traffic flow conditions are free - the *total headway* H would follow the exponential distribution, and drivers would not consider their empty zone. In the model of Buckley [10], this problem does not occur: drivers either follow at their respective empty zone X or at the free headway U, i.e. $H = \min\{U, X\}$. Let $p_{fol}(w)$ denote the probability density function of W. Let C denote the random variate describing the state of the vehicle. By convention, let $C = 1$ and $C = 2$ respectively denote unconstrained and constrained vehicles. The mathematical formulation of model is given by the following equation:

$$p(h) = \phi p_{fol}(h) + (1 - \phi) p_{free}(h) \tag{3.59}$$

where $p_{fol}(h)$ and $p_{free}(h)$ respectively denote the headway probability density functions of the followers (constrained vehicles) and the leaders (free driving vehicles); ϕ denotes the fraction of followers. Assuming no major downstream distributions, the headway distribution of the free drivers can be proven to be exponential in form. As a direct result, for sufficiently large headways we can write for $p(h)$

$$p\left(h\right) = (1 - \phi) p_{free}\left(h\right) = A\lambda \exp\left(-\lambda h\right) \text{ for } h > h^* \tag{3.60}$$

Here h^* is a threshold headway value corresponding to a separation beyond which there is no significant probability of interactions between vehicles, i.e. none of the vehicles is following; λ (arrival rate for free vehicles) and A (the so-called normalization constant, given by $A = \int_0^\infty \lambda \exp(-\lambda s) p_{fol}(s) ds$; see [11]) are parameters to be determined from headway data. For headway values $h < h^*$ the exponential form is no longer valid and must thus be corrected to take account of the following vehicles. This correction is effected *by removing from the exponential distribution the fraction of vehicles that have preferred following times larger than* t, since the assumption is that no vehicle will be found at less than its preferred following headway. This fraction $\pi(h)$ is given by

$$\pi(h) = \int_h^\infty p_{fol}\left(s\right) ds \tag{3.61}$$

Hence, we write

$$(1 - \phi) p_{free}(h) = A\lambda \exp\left(-\lambda h\right) \left[1 - \pi(h)\right] = A\lambda \exp\left(-\lambda h\right) \left[1 - \int_h^\infty p_{fol}\left(s\right) ds\right] \tag{3.62}$$

Using $p_{fol}\left(h\right) = \frac{p(h) - (1-\phi) p_{free}(h)}{\phi}$ we can easilty obtain for $p_{free}\left(h\right)$ the following *integral equation*

$$(1 - \phi) p_{free}\left(h\right) = \frac{A\lambda}{\phi} \exp\left(-\lambda h\right) \int_0^h \left[p\left(s\right) - (1 - \phi) p_{free}\left(s\right)\right] ds \tag{3.63}$$

The parameters A and λ can be evaluated from the observed headways in the range $h > h^*$, where Eq. (3.60) applies. Then the integral equation can be solved numerically subject to the constraint $\phi = \int_0^\infty g_1\left(s\right) ds$ to yield the quantity ϕ and the function $(1-\phi) p_{free}\left(h\right)$ in the range $h < h^*$.

3.3 Distance headway distributions

Most results discussed for time headways in the earlier sections are also valid, with some modifications, for distance headways. It is more easy to observe time headways than distance headways, for the same reason as it is more easy to observe intensity than density.

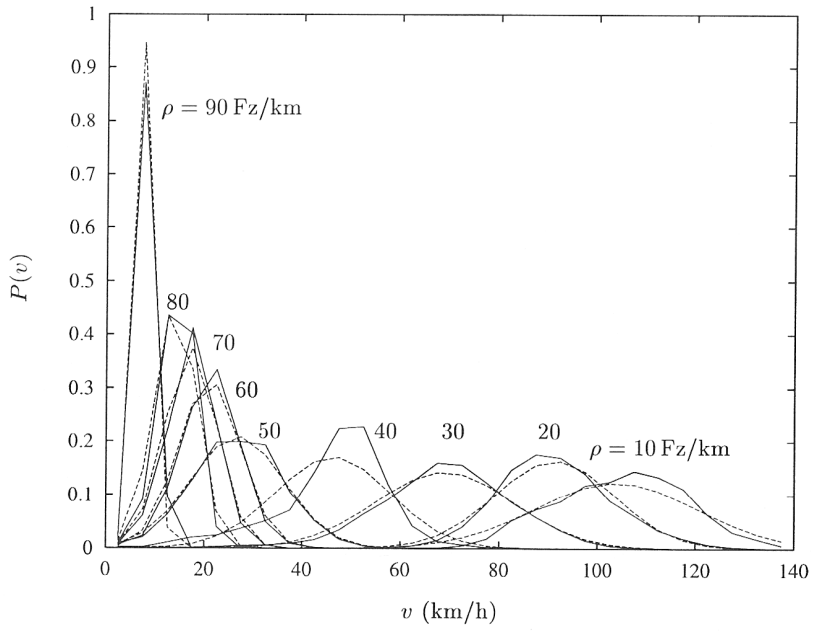

Figure 3.20: Probability density functions of local speeds collected at the A9 two-lane motorway in the Netherlands, for different density values. The distributions are compared to Normal probability density functions (dotted lines).

3.4 Individual vehicle speeds

Just as time and distance headways, speeds have a continuous distribution function. From observations and analysis it follows that speeds usually have a Normal (or Gaussian) distribution. That means the p.d.f., with parameters mean μ and standard deviation σ is:

$$p(v) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(v-\mu)^2}{2\sigma^2}} \tag{3.64}$$

If the histogram of observed speeds is clearly not symmetric, then the Lognormal distribution is usually a good alternative as a model for the speed distribution. As with the mean speed, that could be defined locally and instantaneously, one can consider the distribution of speeds locally and instantaneously.

3.5 Determination of number of overtakings

Vehicles on a road section usually have different speeds, which leads to faster ones catching up the slower ones, and a desire to carry out overtakings. One can calculate the number of desired overtakings from the quantity of traffic and the speed distribution. It depends on the overtaking possibilities to what extend desired overtakings can be carried out.

Consider a road section of length X during a period of length T. Assume the state of the traffic is homogeneous and stationary. Assume further that the instantaneous speed distribution is Normal with mean μ and standard deviation σ. Then the number of desired OT's is:

$$n = XTk^2\sigma/\sqrt{\pi} \tag{3.65}$$

The greater part of this equation is obvious. The number of OT's is:

- linear dependent on the time-road region one considers (term X multiplied by T);

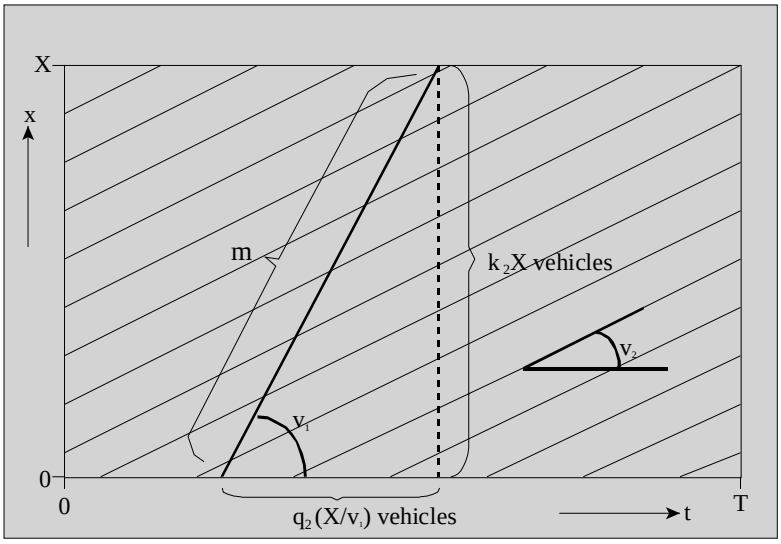

Figure 3.21: Schematised vehicle trajectories to determine number of overtakings

- increases with the quantity of traffic squared (term k^2);
- is larger if the speeds are more different (term σ).

Only the constant term $1/\sqrt{\pi}$ is not obvious.

Proof. Derivation of eq. (3.65). In a road-space region (X multiplied by T) a vehicle drives with speed v_1 through a vehicle stream with intensity q_2 and uniform speed v_2 ($v_1 > v_2$);see Fig. 3.21. At the road section the vehicle overtakes m vehicles. (we apply here the conservation of vehicle trajectories over a triangle: what comes in over the left and lower side must go out over the right side).

$$m = k_2 X - q_2 \left(X/v_1 \right) \tag{3.66}$$

Suppose we have not one vehicle with speed v_1 but an intensity q_1 (with density k_1). During a period of T that makes $q_1 T$ vehicles, that each overtake m slower vehicles. The total number of OT's then becomes:

$$n = q_1 T(k_2 X - q_2(X/v_1)) \tag{3.67}$$

Putting the terms q_2 and X in eq. (3.67) outside the brackets, leads to:

$$n_1 = q_1 q_2 XT(k_2/q_2 - 1/v_1) = q_1 q_2 XT(1/v_2 - 1/v_1) \tag{3.68}$$

An alternative can be derived by replacing the intensities, substituting $q_i = k_i v_i$ ($i = 1, 2$):

$$n_2 = k_1 v_1 T(k_2 X - k_2 v_2 (X/v_1)) = XTk_1 k_2 (v_1 - v_2) \tag{3.69}$$

Generalisation to continuous variables: In stead of considering two densities k_1 and k_2 with corresponding speeds v_1 and v_2, one can generalise towards many densities k_i with speeds v_i. Then one can sum over all classes i or take continuous variables and integrate. We then get a multiple integral:

$$n_2 = XT \int_{v_1=a}^{b} \int_{v_2=v_1}^{b} k^2 p_M \left(v_i \right) p_M \left(v_j \right) \left(v_i - v_j \right) dv_i dv_j \tag{3.70}$$

Substituting a Normal p.d.f. for $p_M(v)$, one can calculate equation (3.65). ■

The quality of operation on a separate cycle path is assumed to depend on the number of overtakings

On most cycle paths overtaking possibilities are rather abundant and it is assumed that the quality of operation is negatively influenced by the number of OT's a cyclist has to carry out (active OT) or has to undergo (passive OT).

Example 33 *Consider a one-way path of 1 km length during 1 hour. The bicycle (bic) intensity = 600 bic/h and the moped (mop) intensity = 150 mop/h. From observations it is known that speeds of bics and mops have, with a good approximation, a Normal distribution:* u_{bic} *= 19 km/h;* σ_{bic} *= 3 km/h;* u_{mop} *= 38 km/h;* σ_{mop} *= 5 km/h.*
Then can be calculated: OT[bic-bic] = (600/19)2 3/$\sqrt{\pi}$ *= 1688 and OT[mop-mop] = (150/38)2 5/*$\sqrt{\pi}$ *= 44. The number of OT[mop-bic] can be calculated with the earlier derived eq. (3.69). It is valid when both speed distribution do not have overlap and in practice that is the case. OT[mop-bic] = (600/19) (150/38) (38 – 19) = 2341.*
The outcomes of the calculation show clearly that the mopeds are responsible for an enormous share in the OT's. The operational quality for the cyclists will increase much if mopeds do not use cyclepaths.

Remark 34 *The measure to forbid mopeds to use cycle paths inside built-up areas has been implemented in the Netherlands in 2001.*

3.6 Dependence of variates (headways, speeds)

When analysing stochastic variables (briefly variates) such as headways and speeds, usually their distribution is the first point of interest. A secondary point is the possible dependence of sequential values. This is relevant for the determination of confidence intervals for estimated parameters and also it is of importance when generating input streams for a simulation model.

For a sample of n independent drawings the following rule holds: the standard deviation of the mean is the standard deviation of the population divided by $\sqrt{n}$. However, if the elements of the sample are not independent but positively correlated, then this rule does not hold anymore and the st. dev. of the mean becomes larger. From practical studies it is known that sequential time headways of vehicles are usually independent or at most very weakly dependent. This is certainly not true for speeds, which is obvious if the headways are small. A vehicle in a platoon has a speed that is very dependent on the speed of its leader. This is illustrated in Fig. 3.22: inside the platoons the speeds vary less than between platoons.This effect has been used in a procedure to determine the headway that separates constrained and free driving. One considers the mean of the absolute relative speeds (relative speed is the speed of a vehicle considered minus the speed of its predecessor) and depicts it as a function of the headway. Usually then a boundary valuer is visible in the graph; see Fig. 3.23.

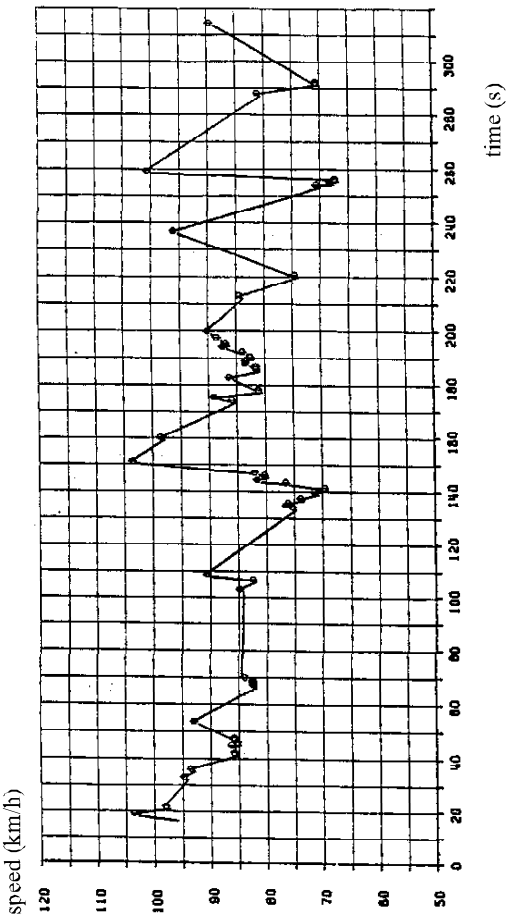

Figure 3.22: Time series of speeds observed at a cross-section to illustrate platooning; data from Dutch two lane road

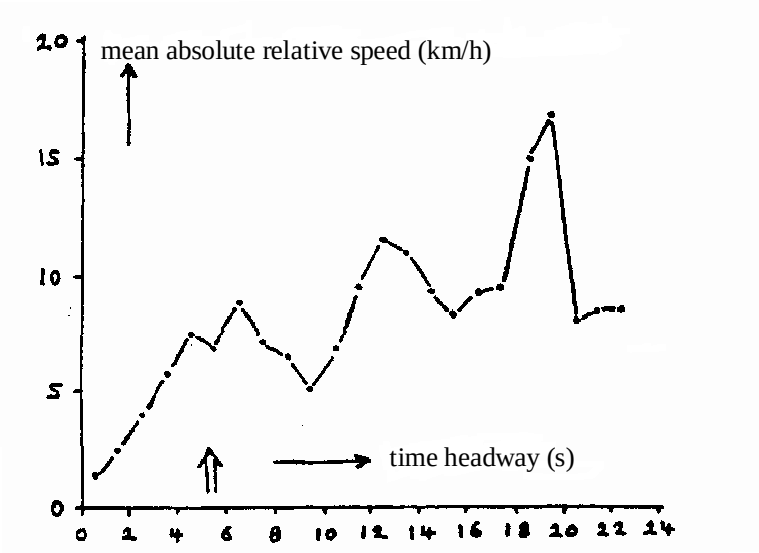

Figure 3.23: Results of processing headways and speeds of a cross-section to determine the boundary between constrained and free driving; data from a Dutch two-lane rural road; lane intensity = 680 veh/h; sample of 1200 vehicles. Outcome: boundary value $\approx$ 5 s

Chapter 4

Fundamental diagrams

Contents of this chapter. This chapter introduces the concept of the fundamental diagram. Some models used in practice will be discussed, also one with a so called capacity drop. This capacity drop has implications for using the diagram for capacity estimation. Also the link with a simple model for an indivudual driver is discussed. Finally some aspects of studies about the diagram will be discussed and a practical result of the effect of rain on the diagram.

List of symbols

q	veh/s	flow rate, intensity, volume
k	veh/m	traffic density
u	m/s	instantaneous speed
q_c	veh/s	critical flow / capacity
k_c	veh/m	critical density
k_j	veh/m	jam density
u_c	m/s	speed at critical density

4.1 Introduction

It is reasonable to assume that drivers will on average do the same under the same average conditions: if drivers are driving in a traffic flow that has a certain speed u, they will on average remain the same distance headway s with respect to the preceding vehicle. This implies that *if we would consider a stationary traffic flow, it is reasonable to assume that there exists some relation between the traffic density* $k = 1/s$ *and the instantaneous mean speed* u (and since $q = ku$, also between the density and the flow, or the flow and the speed). This relation – which is sometimes refered to as the equilibrium situation – will depend on the different properties of the road (width of the lanes, grade), the composition of the flow (percentage of trucks, fraction of commuters, experienced drivers, etc.), external conditions (weather and ambient conditions), traffic regulations, etc.

In traffic flow theory the relations between the macroscopic characteristics of a flow are called 'fundamental diagram(s)'. Three are in use, namely:

- intensity - density $q = q(k)$
- speed - density $u = u(k)$
- speed - intensity $u = u(q)$

It is important to understand that these *three relations represent the same information*: from one relation one can deduce the other two; See Fig. 4.1.

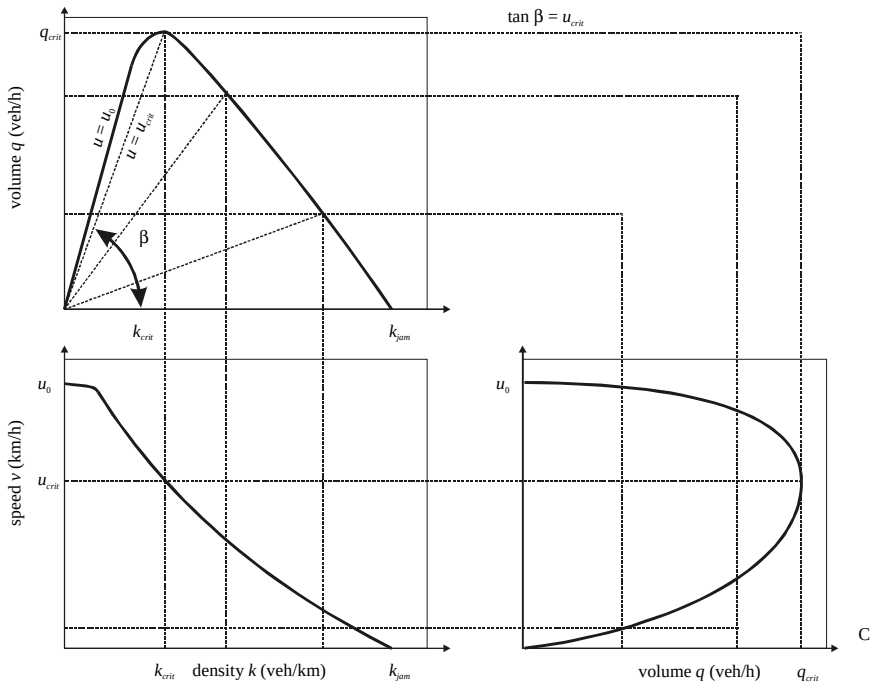

Figure 4.1: Three interrelated forms of the fundamental diagram

Remark 35 *Sometimes density k is replaced by occupancy β but this does not influence the character of the relations substantially.*

It is tempting to assume that the fundamental diagram implicates causality amongst density, flow and speed. One might argue that the average speed u of the drivers is determined by the density k, i.e. $u = u(k)$ (or equivalently, the speed u is determined by the average distance headways $1/k$). However, care should be taken when making such inferences, as is illustrated by the following example.

Consider a situation when traffic conditions are congested, and overtaking opportunities are limited, drivers will be following the vehicle in front and will thus adopt to the speed of the latter to do so. In other words, the follower can only control the distance s with respect to the vehicle in front. It can thus be argued that under these circumstances the distance s is a function of the speed u. In fact, the following relation has been proposed in the literature to express the relation between the distance s and the speed u

$$s = s_0 + T_r u \tag{4.1}$$

where s_0 denotes the minimal gap between two vehicles, and T_r denotes the reaction time. On average, this relation implies the following relation between the speed and the density

$$k = k(u) = \frac{1}{s_0 + T_r u} \tag{4.2}$$

or

$$q = ku = \frac{u}{s_0 + T_r u} \tag{4.3}$$

Since by definition the average speed will always be smaller than the free speed, i.e. $u \leq u_0$, from Eq. (4.3) we can conclude that the flow is bounded from below, i.e. $q \geq \frac{u_0}{s_0 + T_r u_0}$ and can thus not be valid for dillute traffic conditions, where the flow is very small.

On the contrary, assuming free flow conditions, the speed u is generally not determined by the vehicle in front, but rather by the free speed u_0 and an occasional interaction between a

slower vehicle (assuming sufficient passing conditions). Then, the density is a function of the flow q and the free speed u_0, i.e.

$$k = k(q) = \frac{q}{u_0} \tag{4.4}$$

Note that derivation of the fundamental relation from a microscopic model is discussed later in this chapter.

4.1.1 Special points of the fundamental diagram

Special points of the diagram are:

- (mean) Free speed u_0; this is the mean speed if $q = 0$ and $k = 0$; it equals the slope of the function $q(k)$ in the origin;
- Capacity q_c; this is the maximal intensity, sometimes called critical intensity;
- Capacity density or critical density k_c ; i.e. the density if $q = q_c$;
- Capacity speed u_c ; i.e. the mean speed if $q = q_c$;
- Jam density k_j ; i.e. the density if $u = 0$ and $q = 0$.

Note that $u(q)$ is a two-valued function, i.e. at one value of q there are two possible values for u, and k. Note also that capacity is not a special point of the function $u(k)$.

Notation 36 *The part of $q(k)$ with a constant speed is called the 'stable region' of the diagram*[1]*. As soon as speed decreases with increasing density, one enters the 'unstable region'.*

Notation 37 *The region in which densities are greater than the capacity density, is called 'congestion region' or 'congestion branch'. The entire region with $k < k_c$ is sometimes called 'free operation' and the congestion region 'forced operation'.*

4.1.2 Importance of the fundamental diagram

- Definition of Level of Service (LOS) - the LOS - a yard stick for the quality of the traffic operation - is related to the fundamental diagram. The LOS is discussed further in chapter 6.
- The diagram contains also information about the relation between intensity and travel time; this is used in traffic assignment. Fig. 4.2 shows the relations between density and intensity on the one hand and travel time on the other.
- Determination of the capacity q_c, which is an important parameter in traffic planning and control, is sometimes based on a specific form of the diagram, as will be shown in section 4.2.3.
- Often the fundamental diagram is a basic element of more comprehensive models that describe the traffic operation on a network, e.g. METANET, or describe assignment and traffic operation in combination, e.g. 3DAS [15].
- On the other hand microscopic models, e.g. the model FOSIM, can be used to derive fundamental diagrams.

[1] In chapter 4 on shockwave analysis, this namegiving will become clear.

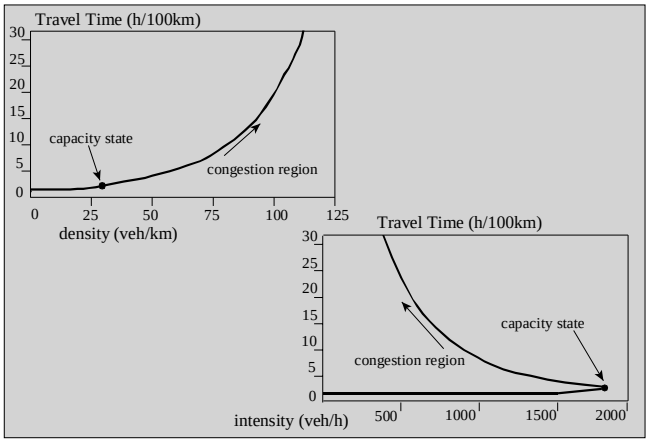

Figure 4.2: Travel time per distance as a function of density and intensity

4.1.3 Factors influencing the diagram

The fundamental diagram is not a physical law but is dependent on characteristics of:

1. the road;
2. the drivers and their vehicles;
3. conditions such as lighting and weather.

In practical terms the most important factors are:

- Type of road and, for a given type, characteristics such as lane width, curvature, grades, and quality of surface.
- Vehicle composition; this is normally made operational by a division of the flow between cars and trucks (in the USA recreational vehicles are often taken into account as a separate vehicle category).
- Type of traffic with respect to travel purpose and familiarity of the drivers with the road. E.g., home-work travel over relatively short distances shows a different behaviour than long distance holiday traffic.
- Measures, such as speed limits, either permanent or dynamic.
- Lighting conditions, e.g. darkness (with or without road lighting) and daylight.
- Weather conditions: dry, rain, fog.

To give some indication of the influence of these different aspects, Tab. 4.1 shows an overview of the capacity reductions for different weather and ambient conditions.

Over the years vehicles have changed (better road-holding, better brakes, more comfort) and so have drivers, especially their experience in handling dense traffic at relatively high speeds. These factors have led to an increase in capacity of motorways during the last few decades; estimates range from 0.4 to 1.0 % per year[4].

[4] American Highway Capacity Manual: under ideal conditions (road, traffic, external factors) the capacity of one lane of the motorway has increased from 2000 veh/h in 1950 to 2400 veh/h in 2000, corresponding to a growths of 0.37 % per year.

Circumstances	Capacity
Ideal conditions	100%
Darkness (no illumination)	95%
Darkness (with illumination)	97%
Dense Asphalt Concrete (DAC) [2]with rain	91%
Open Asphalt Concrete (OAC) with rain[3]	95%
DAC / rain / darkness	88%
DAC / rain / darkness / illumination	90%
OAC / rain / darkness	91%
OAC / rain / darkness / illumination	92%

Table 4.1: Capacity reduction effects

4.2 Models of the fundamental diagram

Looking through the literature one finds a lot of models for the diagram. They are based on:

- *Driving behaviour theories*; e.g. car-following theory. By aggregating the individual microscopic behaviour, one can obtain the macroscopic fundamental diagram. The aggregation can be done analytically or by means of a microscopic simulation model. It's obvious that keeping distance is especially relevant for the congestion region. At free operation a mixture of keeping distance and free driving determines the diagram. Analogies with a queueing system exist; the queueing is the driving behind a slower leader that cannot be overtaken immediately. Wu's diagram in the next section uses this analogy.
- *Empirical observations*. Using a 'curve fitting' technique a function is determined that fits the data well and preferably does not contradict elementary conditions. The importance of this type of model has increased with the abundant availability of data. The models of Smulders and De Romph in the next section are Dutch examples of this development.
- *Analogy with flow phenomena from different fields of science*, e.g. the mechanics of a stream of gas of fluid. The importance of these type of models has decreased because the knowledge about the behaviour of drivers has increased and it is really different from gas molecules and fluid particles. This type of model will not be included in these notes; interested readers are referred to [18].

Often an analysis starts with the function $u(k)$ because that function is monotonic. It is always advisable to determine the other relations $q(k)$ and $u(q)$, too, and inspect them for consistency and plausibility. Because the derivative of $q(k)$ is the speed of the kinematic wave, this function is also useful for checking the model.

4.2.1 Model specifications

- Fundamental diagram of Greenshields [22].

This is certainly one of the oldest models of traffic flow theory. It is based on the simple assumption that mean speed decreases linearly with density. The model was validated using 7 (seven!) data points obtained by aerial photography.

$$u(k) = u_0 \left(1 - \frac{k}{k_j}\right) \tag{4.5}$$

with u_0 = free speed and k_j = jam density. The function $q(k)$ follows simply:

$$q = ku \Rightarrow q(k) = ku_0 \left(1 - \frac{k}{k_j}\right) = ku_0 - k^2 u_0/k_j \text{ (parabola)} \tag{4.6}$$

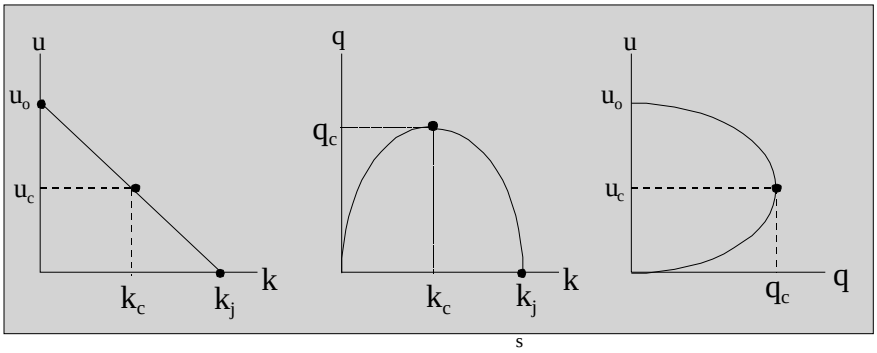

Figure 4.3: Greenshields' fundamental diagram

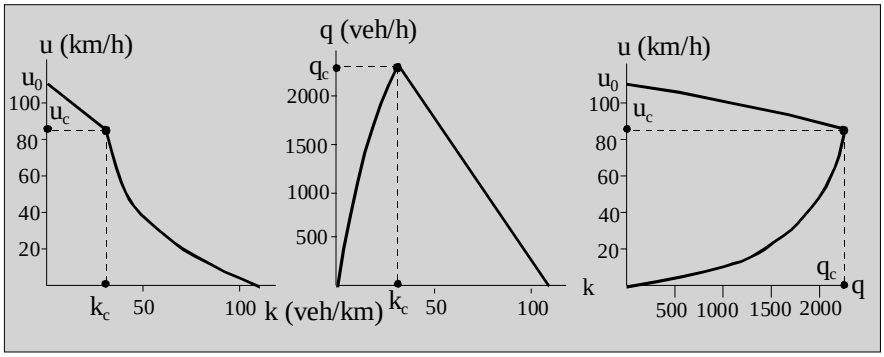

Figure 4.4: Smulders' fundamental diagram

The critical values for q, k, and u follow from the condition that at the capacity point:

$$\frac{dq}{dk} = 0 \Rightarrow \frac{dq}{dk} = u_0 - 2ku_0/k_j = 0 \Rightarrow k_c = k_j/2 \tag{4.7}$$

and thus, by substition into eqn. (4.5)

$$u_c = u_0\left(1 - k_c/k_j\right) = u_0/2 \tag{4.8}$$

For the capacity, we find

$$q_c = k_c u_c = \frac{1}{4} u_0 k_j \tag{4.9}$$

The function $q(u)$ is also simple to derive: substitute in (4.5) $k = q/u$

$$u = u_0\left(1 - (q/u)/k_j\right) \Rightarrow 1 - u/u_0 = (q/u)/k_j \tag{4.10}$$

or

$$q(u) = uk_j - (k_j/u_0)\, u^2 \tag{4.11}$$

Greenshields' model is still used very much because of its simplicity; see Fig. 4.3. However, for motorways it deviates from present reality since u does not decrease that strongly for small densities, and the fact that capacity density is much smaller than half of the jam density (0.2 would be a better number).

- Fundamental diagram of [49]

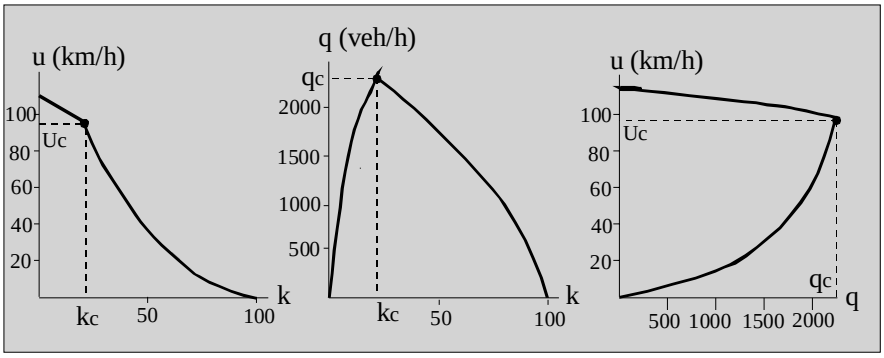

Figure 4.5: De Romphs' fundamental diagram for a 3-lane roadway

In developing a dynamic macroscopic model for traffic flow on motorways, Smulders introduced the diagram: (see Fig 4.4):

$$u(k) = \begin{cases} u_0\,(1 - k/k_j) & \text{for } k < k_c \\ \gamma\,(1/k - 1/k_j) & \text{for } k > k_c \end{cases} \tag{4.12}$$

Parameters are: u_0, k_c, k_j, γ. The parameter γ is not free but follows from the requirement that $u(k)$ is continuous at point $k = k_c$. It is left to the reader to derive that $\gamma = u_0 k_c$.

Smulders chose for a Dutch motorway: $u_0 = 110km/h$, $k_c = 27veh/km$, and $k_j = 110veh/km$ (all variables per lane). Capacity then becomes: $q_c = 2241veh/h$.

Remark 38 *For this diagram the requirement $dq/dk = 0$ in the capacity point does not necessarily hold.*

- Fundamental diagram of De Romph

De Romph (1994) has generalised Smulders' diagram to:

$$u(k) = \begin{cases} u_0\,(1 - \alpha k) & \text{for } k < k_c \\ \gamma\,(1/k - 1/k_j)^{\beta} & \text{for } k > k_c \end{cases} \tag{4.13}$$

At first glance there are 6 parameters: u_0, k_c, k_j, α, β, γ. The required continuity at $k = k_c$ reduces the number of independent parameters to 5.

The reader is invited to derive that: $\gamma = u_0\,(1 - \alpha k_c)\,/\,(1/k_c - 1/k_j)\,\beta$. Using extended measurements on the ring road of Amsterdam, the parameters of this model have been estimated for different elements of the motorway; e.g. 2-lane roadways, 3-lane roadways; single lane on-ramps; roadways in a tunnel; etc.

Example 39 *For a 3-lane roadway the results (average values for one lane) are:*

- **Example 40**
  - $u_0 = 110km/$, $k_c = 23veh/km$, $\alpha = 5.710^{-3}$, *and* $\beta = 0.84$.
  - *Assumed is:* $k_j = 100veh/km$. *Then it can be calculated:*
  - $\gamma = 1672$, *and capacity* $q_c = 2215veh/h$; *see Fig. 4.5.*

Remark 41 *The speed of the kinematic wave at jam density for this model is minus infinity, which is far from reality. Derivation:*

$$q(k) = \gamma k\,(1/k - 1/k_j)^{\beta} \Rightarrow \frac{dq}{dk} = \gamma\,(1/k - 1/k_j)^{\beta} - \gamma k^{-1}\beta\,(1/k - 1/k_j)^{\beta-1} \tag{4.14}$$

If $k \to k_j$ then the first term of $dq/dk \to 0$, but the second term $\to \infty$ because $\beta < 1$. β has to be < 1 because otherwise the congested branch of $q(k)$ is convex, which is not feasible either. Moreover, De Romph has found $\beta < 1$ based on data.

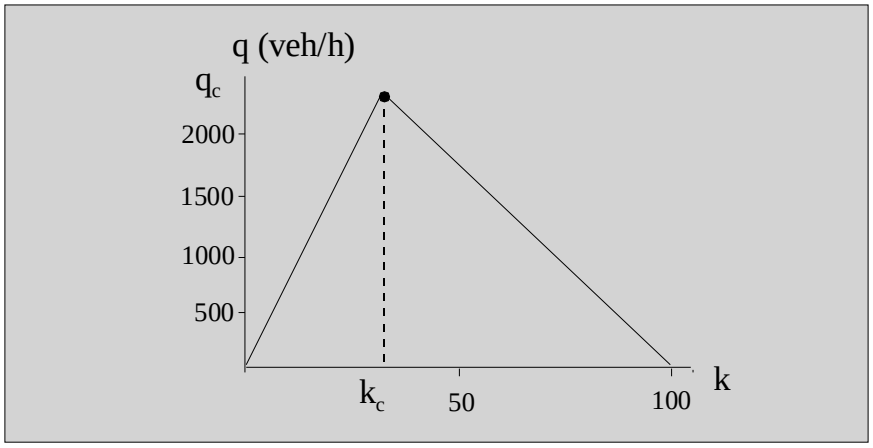

Figure 4.6: Daganzo's fundamental diagram

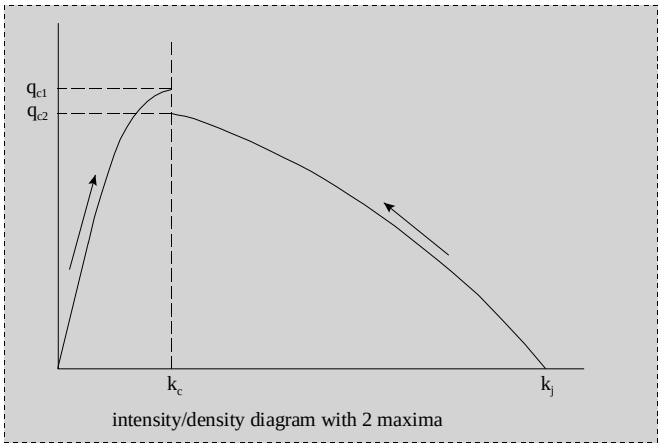

Figure 4.7: Example of a fundamental diagram with a discontinuity

- Schematised fundamental diagram of Daganzo [14]

When developing new models for traffic operation one needs a fundamental diagram that is simple but represents the essential properties of the traffic flow correctly. The simple parabola model of Greenshields is an example fulfilling these requirements. Daganzo has introduced an alternative in which the function $Q(k)$ is represented by two straight lines; see Fig. 4.6. It is left to the reader to determine the relations $u(k)$ and $u(q)$.This model has three parameters: u_0, q_c (or k_c) and k_j. It is remarkable that the diagrams of De Romph and Daganzo are rather similar if one chooses the parameter values correctly.

4.2.2 Concept of discontinuous diagram

Edie [19] was the first researcher to indicate the possibility of a discontinuity in the diagram around the capacity point. This idea is based on the observation that a traffic stream with increasing density (starting from stable or free flow) reaches a higher capacity value ('free flow capacity', q_{c1}) than a traffic stream starting from a congested state (in the extreme case from a standing queue) that ends in the so called 'queue discharge capacity', q_{c2}. This idea is illustrated in Fig. 4.7 in which a so called 'capacity drop' is present.

4.2.3 Wu's fundamental diagram with capacity drop

Recently Wu [58] has developed a model for the diagram with a capacity drop, based on assumptions about microscopic behaviour. In this model two regimes are distinguished: free flow and congested flow. Free flow has densities from $k = 0$ up to $k = k_1$. Congested flow has a density range from $k = k_2$ up to the jam density k_j. Both regimes are overlapping in terms of density range, i.e. $k_1 > k_2$.

Main assumption for free flow state

In this state it is assumed the traffic flow is a mixture of free driving vehicles with mean speed u_0 and platoons with speed u_p. If the fraction of free driving vehicles (in terms of density) is p_{free}, then the fraction of vehicles in platoon is $(1 - p_{free})$, and the overall space mean speed equals:

$$u = p_{free}u_0 + (1 - p_{free})u_p \tag{4.15}$$

Parameters: $p_{free} = 1$ for $k = 0$ and $p_{free} = 0$ for $k = k_1$.

It can be argued, based on queueing theory, that for a two lane roadway (for one-directional traffic) p_{free} decreases linearly with density, i.e.

$$p_{free} = 1 - k/k_1 \tag{4.16}$$

For a three-lane roadway $p_{free} = 1-(k/k_1)^2$ and in general the exponent equals the number of lanes minus 1, i.e. for n lanes:

$$u(k) = (1 - (k/k_1)^{n-1})u_0 + (k/k_1)^{n-1}u_p \tag{4.17}$$

It is left to the reader to derive that (4.17) implies $u(k)$ starts flatter if the roadway has more lanes.

At density k_1 every vehicle drives in a platoon with speed u_p. This implies a relation between the nett time headway, h^f_{nett}, in this platoon, the effective vehicle length $(1/k_j)$ and speed. It is left to the reader to derive:

$$k_1 = \left(u_p h^f_{nett} + \frac{1}{k_j}\right)^{-1} \tag{4.18}$$

Given parameters u_p and k_j, either h^f_{nett} or k_1 is a free parameter of the model. Wu has chosen h^f_{nett} because it can be observed in practice more easily than k_1, especially if it is assumed that h^f_{nett} is a constant parameter for all free flow states.

Assumption for congested flow state

In congested flow every vehicle is (more or less) in a car-following state and maintains a constant nett time headway, h^c_{nett}, over the density range $k_2 < k < k_j$. This assumption implies a straight line for the congested part of the function $q(k)$. Hence in this aspect the model is the same as Daganzo's and Smulders'.

Derivation:

$$s_{nett} = uh^c_{nett} \quad \text{with } u = \text{mean speed} \tag{4.19}$$

$$s_{gross} = 1/k_j + uh^c_{nett} \tag{4.20}$$

$$k = \frac{1}{s_{gross}} = \frac{1}{1/k_j + uh^c_{nett}} \tag{4.21}$$

Solving for u leads to

$$u = \frac{1}{h^c_{nett}} \left(\frac{1}{k} - \frac{1}{k_j} \right) \quad \text{and apply } q = ku \tag{4.22}$$

yielding

$$q = \frac{1}{h^c_{nett}} \left(1 - \frac{k}{k_j} \right) \tag{4.23}$$

Parameter k_2: This parameter can be determined by assuming that the maximum speed of the congested state is at most equal to the speed that corresponds to 100 percent platooning at free flow state. Consequently using eq. (4.21):

$$k_2 = \frac{1}{1/k_j + u_p h^c_{nett}} = \left(u_p h^c_{nett} + \frac{1}{k_j} \right)^{-1} \tag{4.24}$$

Note the similarity of eq. (4.24) and eq. (4.20).

Example 42 *The model has 5 parameters. Typical values (standardized to lane values) for a two-lane roadway with 100 % cars are: u_0 = free flow speed = 110 km/h; u_p = speed of free flow platoon = 80 km/h; k_j = jam density = 150 veh/km; h^f_{nett} = nett time headway in free flow platoon = 1.2 s; and h^c_{nett} = nett time headway at congested flow = 1.6 s. Because the nett time headway at free flow is smaller than at congestion the two branches of the fundamental diagram in the u-k plane do not coincide. In the example the free flow capacity is 2400 veh/h and the discharge capacity 1895 veh/h (21 % less); see Fig. 2.8. Capacity drops found in practice are usually smaller.*

Remark 43 *The capacity drop offers a substantial possible benefit of ramp metering. If one can control the input flows such that the intensity on the main carriageway does stay below, say, 2200 veh/h per lane, than most of the time the smaller discharge capacity is not relevant. However, in practice the benefits of ramp metering are less than according to this reasoning.*

4.2.4 Diagram for roadway and lane

The interpretation of diagrams for separate lanes of a roadway and for the total roadway is not straightforward. The diagram for the roadway is not just the sum of the diagrams of the lanes because the distribution of the traffic over the lanes plays a role.

The mean speed as a function of intensity on a roadway of a motorway, the upper branch of $u(q)$, is often remarkably flat until rather high intensities. This can be partly explained by the changing distribution of the roadway intensity over the lanes. We explain this mechanism by means of a schematised situation.

Consider a two-lane roadway of a motorway with two types of vehicles; with free speed v_1 and v_2. The fast ones will use the left lane more and more with increasing roadway intensity. Denote

- q_t = roadway intensity
- q_r= intensity on right lane and q_l = intensity on left lane
- v_1 = speed of fast vehicles and v_2 = speed of slow vehicles
- c = slope of $u(q)$ for right lane
- α = fraction of q_t on right lane

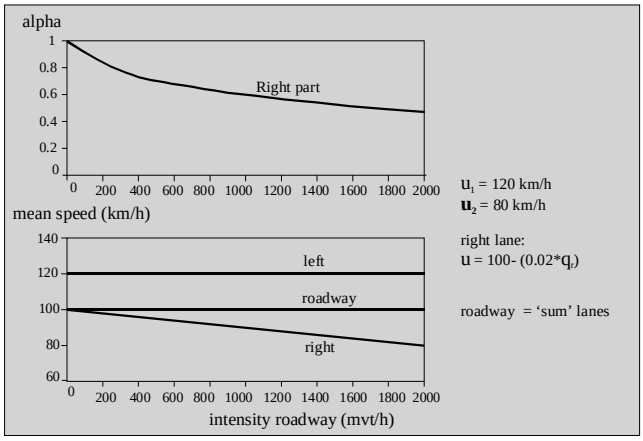

Figure 4.8: Distribution of roadway densities over the lanes and part of the fundamental diagram for roadway and lanes

The distribution of vehicles over v_1 and v_2 is 1:1. When q_t is near 0, all vehicles use the right lane and with increasing q_t more and more fast ones will use the left lane. We can divide the roadway intensity as follows:

$$q_t : \begin{cases} \text{Right} & \alpha q_t : \begin{cases} 0.5q_t \text{ (slow)} \\ \alpha q_t - 0.5q_t \text{ (fast)} \end{cases} \\ \text{Left} & (1-\alpha)\, q_t \text{ (fast)} \end{cases} \tag{4.25}$$

yielding

$$u_r = \{0.5v_2 + (\alpha - 0.5)v_1\}/\alpha \tag{4.26}$$

We require that for the right lane the mean speed decreases linearly with its intensity:

$$u_r = 0.5\,(v_1 + v_2 - cq_r) \tag{4.27}$$

The question is which function $\alpha\,(q_t)$ brings about correspondence between (4.26) and (4.27). Set (4.26) = (4.27) and substitute $q_r = \alpha q_t$

$$\{(\alpha - 0.5)\, v_1 + 0.5v_2\}\,/\alpha = 0.5\,(v_1 + v_2) - c\alpha q_t \tag{4.28}$$

This is a second degree equation for α with the solution:

$$\alpha = \frac{-\delta v + \sqrt{\delta v^2 + 8cq_t\delta v}}{4cq_t} \quad \text{with} \quad \delta v = v_1 - v_2 \tag{4.29}$$

It turns out that $\alpha\,(q_t)$ corresponds qualitatively with reality.

Example 44 $v_1 = 120km/h$, $v_2 = 80km/h$, $c = 0.02$. *At* $q_t = 2000veh/h$ *all fast vehicles drive on the left lane and all slow ones drive on the right lane; Note:* u_r *is linear as function of* q_r *but not as function of* q_t*.*

4.2.5 Fundamental diagram based on a car-following model

Consider a situation in which vehicle i drives behind vehicle $i-1$. Vehicle i considers a gross distance headway of s_i. Both vehicles have the same speed v. Driver i includes the following in determination of the gross distance headway

1. Vehicle $i-1$ may suddenly brake and come to a complete stop
2. Driver i has a reaction time of T_r seconds
3. Braking is possible with an deceleration of a (with $a > 0$)
4. When coming to a full stop behind the preceding vehicle, the net distance headway between the vehicles $i-1$ and i is at least d_0
5. The decelaration of the vehicle $i-1$ is α times the deceleration of vehicle i

It should be clear that the parameter α is somehow a measure for the aggressiveness of the driver: larger values of α imply that the follower assumes more abrupt deceleration of the follower, which will result in larger distance headways. It is left to the reader to determine that the minimal distance s_i to the leader of vehicle i

$$s_i + v_i^2 - \frac{1}{2\alpha a_i} = L + d_0 + v_i T_r + \frac{v_i^2}{2a_i} \tag{4.30}$$

or in words: the gross distance headway + braking distance of the leader = length of the follower + margin + reaction distance + braking distance follower.

If we assume that $s_0 = L + d_0$ (gross stopping distance headway = vehicle length + safety distance margin) then we have

$$s_i = s_0 + v_i T_r + \frac{v_i^2}{2a_i}\left(1 - \frac{1}{\alpha}\right) \tag{4.31}$$

To transform this microscopic relation into a macroscopic description of traffic flow, we need to make the following substitutions: replace the gross distance headway s_i by the inverse of the density $1/k$, replace the individual speed v_i by the flow speed u and replace the gross stopping distance headway s_0 by the inverse of the jam density $1/k_j$. We get

$$\frac{1}{k} = \frac{1}{k_j} + uT_r + \frac{u^2}{2a}\left(1 - \frac{1}{\alpha}\right) \tag{4.32}$$

Since $q = ku$, we get the following expression for the relation between flow and speed

$$q(u) = \frac{u}{\frac{1}{k}} = \frac{u}{\frac{1}{k_j} + uT_r + \frac{u^2}{2a}\left(1 - \frac{1}{\alpha}\right)} \tag{4.33}$$

Fig. 4.9 shows the resulting relation between the speed and the flow. Note that the speed is not restricted in this model.

Let us now show how we can determine the capacity from the car-following model. Fig. 4.9 shows how the flow rate depends on the speed for a large range of speeds for a certain set of parameters. We can now distinguish two situations

1. The flow speed u is – within certain boundaries – free and the drivers can choose it arbitrarily. This is the case for motorway traffic and it appears from practise that the maximum flow can be attained in this situations
2. The maximum flow speed u_m is prescribed and is lower that the optimal speed u_c yielding the capacity. In that case, the capacity is given by $q(u_m)$ (Eq. (4.33)).

If we first consider situation 1, then we can determine the speed u_c for which capacity results by taking the derivative of q with respect to u and setting it to zero. After checking that this indeed yields the maximum (and not the minimum), we can determine that the optimal (critical) speed, capacity and critical density are given by

$$u_c = \sqrt{\frac{2}{k_j}\frac{a}{(1-1/\alpha)}} \qquad q_c = \frac{u_c k_j}{2 + T_r u_c k_j} \qquad k_c = \frac{k_j}{2 + T_r u_c k_j} \tag{4.34}$$

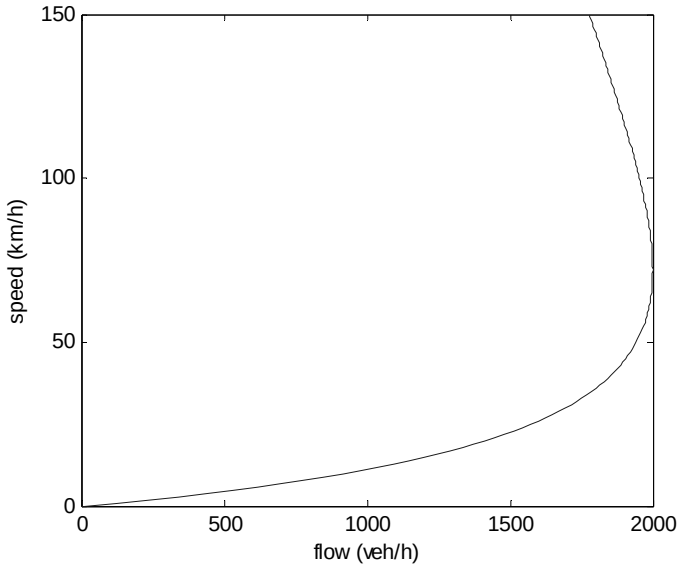

Figure 4.9: Flow - speed curve $u\,(q)$ derived from simple car-following strategy

Sensitivity of capacity

We can also express q_c as a function of α, T_r, k_j and a to determine the effect of changes in the parameters on the capacity

$$q_c = \frac{1}{T_r + \sqrt{\frac{2}{k_j}\frac{1-1/\alpha}{a}}} < \frac{1}{T_r} \tag{4.35}$$

From Eq. (4.35), we observe that the capacity increases when

- the reaction time T_r decreases;
- the jam-density k_j increases (shorter cars);
- the braking deceleration a is larger (improved braking system, e.g. ABS);
- the parameter α is closer to 1 (drivers are less cautious).

Furhtermore, it is remarkable that the critical speed u_c is not a function of the reaction time T_r. How can that be explained?

From Eq. (4.34) we can see that when $\alpha \to 1$ that $u_c \to \infty$; this implies that we need to restrict our choice for α by assuming that $\alpha > 1$. Clearly, this does not hold when there is a speed limit.

The model provides us an explanation for the following observations: it turns out that as time goes by, capacity increases steadily. For example, in the USA the capacity has increases from 2000 veh/h (in 1950) to 2400 veh/h (in 2000) *under ideal circumstances*. Using the derived model, we can explain this 20% increase. It is left to the reader to verify that the parameter α has the biggest influence on the capacity. We can thus explain the capacity increase by a decrease in α (drivers are more daring and are more experienced), a shorter reaction time T_r and higher deceleration a.

Brick wall following strategy

A special case of the model described here results when the following driver assumes that the leading vehicle is able to abruptly come to a full stop. This implies that the leader may have

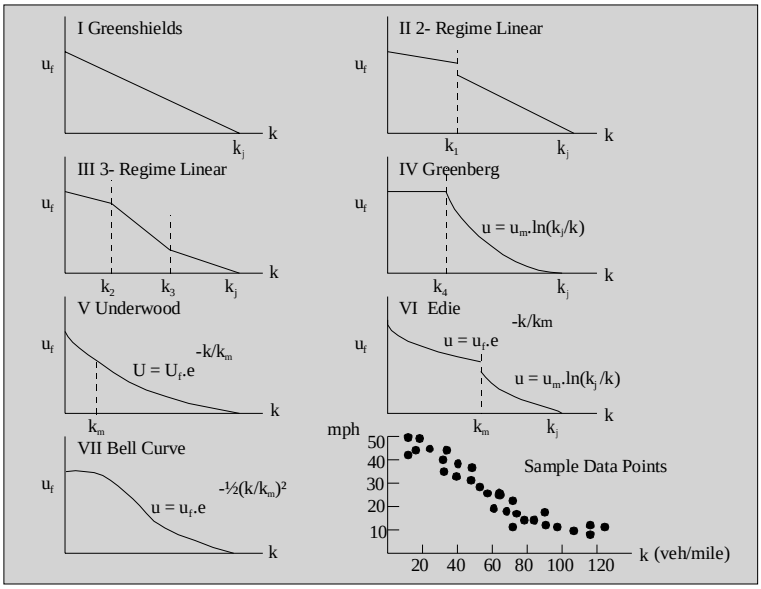

Figure 4.10: Models for $u(k)$ and a set of data points

an infinite deceleration – in practise, a very large deceleration. Consider for instance a car that crashes into a stopped truck. This pessimistic scenario is especially important for systems in which not the individual drivers have responsibility, but the operator (e.g. Combi-road or people movers).

4.3 Studies of the fundamental diagram

4.3.1 General points

If one wants to determine the fundamental diagram for a road section the following points are relevant:

- Does one need the complete diagram or only a part of it; e.g. only the free operation part ($k < k_c$) or only the congestion branch. A more fundamental point is whether it is possible to determine the complete diagram at one cross-section; see next section.
- Is the road section homogeneous? If this is the case, one can do with observations at a single cross-section. Otherwise road characteristics are variable over the section and a method such as the moving observer might be suitable.
- Period of analysis: If this is chosen too short, random fluctuations will have too much influence; if it is too long then it is questionable whether the state of the flow is stationary over the period. In practice the balance between randomness and stationarity has led to periods of 5 to 15 minutes.
- Finally, one has to estimate the parameters of the model chosen. This is mostly done by using a regression technique.

Fig. 4.10 (from [17]) shows that a given set of data points can be used to fit quite a few different models. In general models without too many parameters are preferred.

4.3.2 Influence and importance of data collection location

In order to get quantitative data about a fundamental diagram one has to carry out measurements at selected sites and periods. This will be illustrated with the data one can get when

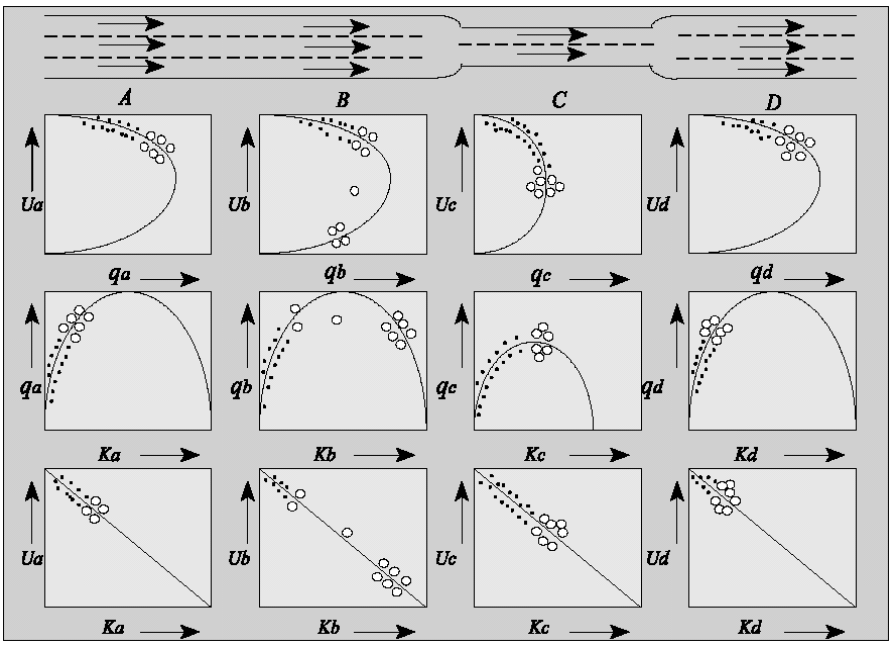

Figure 4.11: Traffic flow conditions at different cross-sections for a under and oversaturated bottle-neck

taking measurement around an overloaded bottle-neck. Fig. 4.11 depicts a roadway of a motorway of 3 lanes with a bottle-neck (b-n) section of 2 lanes wide. Measurements will be carried out at 4 cross-sections:

A This cross section is so far upstream the bottle-neck that congestion due to an overloading of the b-n will not reach it.

B This cross-section is closer to the b-n than A and congestion will reach it.

C A cross-section inside the b-n.

D A cross-section downstream of the b-n.

The fundamental diagrams of Greenshields have been assumed to hold for all cross-sections. They are the same for cross-sections A, B and D. For cross-section C the form is similar but the capacity and jam density are 2/3 of the values at A. We assume that the intensity increases gradually from a low value to a value that is just a little smaller than the capacity of the b-n; this capacity is $2C_0$ with C_0 = capacity of one lane. The data points resulting from such a demand pattern are depicted as * in the diagrams. At cross-section A, B and D intensity is not higher than 2/3 of the capacity. This means free operations with high speeds. At cross-section C the capacity is nearly reached. In a realistic case this would mean somewhat lower speeds but here that is not the case because Greenshield's diagram is used.

Now we assume demand increases further until a value that equals 2.5 times C_0 and discuss which data points we will get on the 4 cross-sections.

A It has been assumed that congestion will not reach this cross-section, so the state of the flow remains free. The data points, depicted by open circles, are in a range of 4000 to 5000 vehicles per hour and speeds remain high.

B When the higher demand reaches the beginning of the b-n, congestion will start, move upstream and reach cross-section B after some time. Before that moment data points

are still on the free flow part of the diagram and afterwards on the congestion part. The intensity then equals (on average) the capacity of the b-n and the mean speed equals the speed corresponding to this intensity according to the congested part of the diagram (in this case around 30 km/h).

C In the b-n intensity is limited to the capacity value of the b-n and the state does not change much relative to the state at the end of the demand increase discussed before.

D Here the intensity is not larger than $2C_0$ because the b-n does not let through more vehicles and traffic operation remains free.

When demand is reduced to low values the process will develop in reverse order. Looking to the total results it appears that at cross-section A and D only free traffic operation can be observed. In the b-n one can observe the total free part of the diagram. Most information about the diagram is seen at cross-section B. But it should be realised that also here the information about the congested part of the diagram is rather limited. To collect data about the complete congested part of the diagram, one requires a b-n with a capacity varying from zero to, in this case, $3C_0$.

Concluding:

- The site of data collection determines which traffic flow states one can observe.
- Only in a bottle-neck one can observe a long lasting capacity state. If one wants to estimate capacity and has observations carried out at sites not being a bottle-neck, some form of extrapolation is always required.

An important consequence of the discussion above is that the capacity state can only occur in a bottle-neck. If one wants to determine the capacity from observations, not carried out in a bottle-neck, some form of extrapolation is needed.

4.3.3 Estimation of capacity using the fundamental diagram

In the light of the capacity drop two capacities do exist; free flow capacity and queue discharge capacity. Both values are relevant and require a different estimation method.

Estimating free flow capacity

Capacity can be estimated by fitting a model of the function $q(k)$ to data. The calculated maximum of $q(k)$ is an estimate of capacity. It turns out that this model is not suitable for nowadays motorway traffic, where it does not hold that the model for the free flow part of $q(k)$ has a maximum with the derivative $dq/dk = 0$ at capacity.

An alternative method is to assume a value of the capacity density k_c, fit a curve to a model for the free flow part of $q(k)$ and estimate the capacity by $q(k_c)$. It is obvious that the outcome of the estimation depends on the value of k_c and this makes the methods more suitable for comparisons than for absolute values. The procedure has been applied in determining the effect on capacity of road lighting; see [6].

"To estimate a capacity shift due to road lighting a before-and-after study set-up has been chosen. In its most simple form this implies estimating the average capacity of a roadway before and after installing of lighting. However, the disadvantages of such a 'naive' before-and -after study are great because there are so many factors that could have an influence on capacity and be different in both periods. Therefore the capacities have also been estimated under daylight conditions in before and after periods.

This set-up also enables determining the effect on capacity of darkness, without and with lighting, with daylight capacity as the reference situation. Such a comparison can give additional information on the relative contribution of lighting to a capacity improvement.

	Daylight		Darkness	
	Before	After	Before	After
Site I NB Treatment	100	99.1	90.3	92.9
Site I SB Treatment	100	96.8	93.5	95.4
Site II Treatment	100	100.4	95.5	97.1
Site III Comparison	100	98.8	95.1	94.4

Table 4.2: Set-up of study. Cells with double lining indicate cases with road lighting

In total, data from 3 motorway sites has been used. At Site I (motorway A50) data was available from both directions (North Bound=NB and SB), which was analysed separately. At the second motorway, A12, only data from one roadway was available. However, along the same roadway (approximately 14 km downstream) data was available from a road section with lighting in before and after period, which has been used as an extra comparison group. The total set-up is shown in Table 4.2, in which double lined cells represent cases with road lighting." Results in the table are presented as an index. The estimated capacities of the cells 'Daylight-Before' have been set at 100. Overall the positive effect of road lighting on capacity could be convincingly shown but it is rather small (2.5 % at the 2-lane roadway and 1.7 % at the three lane roadway).

Estimating queue discharge capacity

This capacity can only be estimated in an overloaded bottle-neck. Ideally three measuring stations are required:

- Upstream of the bottle-neck; data from this cross-section can be used to determine if overloading occurs (mostly mean speed is used as a criterion).
- Downstream of the bottle-neck; the state of the traffic flow should be free at this cross-section. If this is not the case, then a more severe bottle neck is present downstream.
- At the bottle-neck; intensities at this spot are capacity measurements if the traffic state upstream is congested and the state downstream is free.

Remark 45 *At an overloaded bottle-neck intensity at all three measuring stations is the same. It is an advantage to use the intensity at the downstream section to estimate capacity because traffic flow is more smooth than at the other stations and this reduces measuring errors.*

Remark 46 *The fact that large parts of motorway networks all over the world exhibit daily overloading is beneficial for the knowledge of capacity values.*

4.4 Capacity state on a motorway and the effect of rain

Finally an illustration of the fundamental diagram by means of data from a motorway (roadway with three lanes) near Rotterdam under two conditions: dry and fair weather, and rainy conditions; see Fig. 4.12 and 4.13.

The road section considered is more or less a bottle-neck with a length of 4 km. At a speed of 70 to 80 km/h drivers stay 3 to 3.5 minute in this traffic flow state. Apparently they are capable to deliver a high performance over this relatively short period, leading to high capacities. It is questionable if drivers are able to maintain this over a period of, say, one hour.

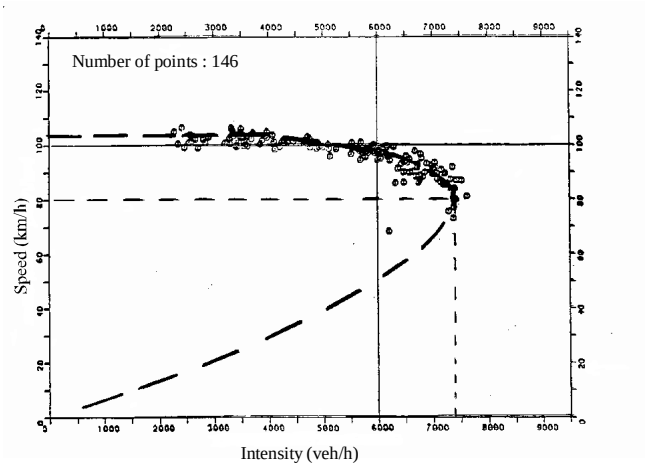

Figure 4.12: Relation speed-intensity in dry weather

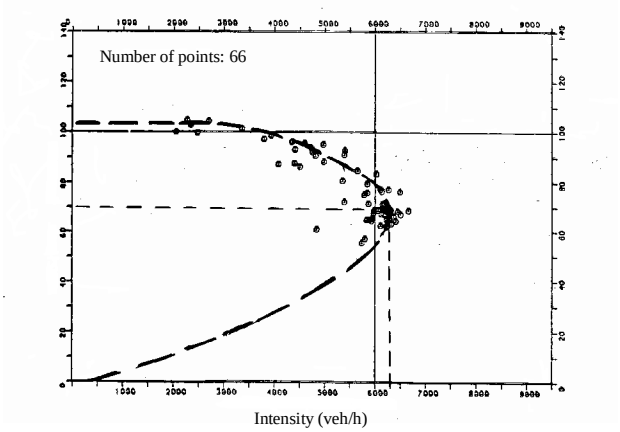

Figure 4.13: Relation speed-intensity in case of rain

MO-team	Roadway (3 lanes)			Average per lane			
	q_c	u_c	k_c	q_c	k_c	$\bar{s}$	$\bar{h}$
	veh/h	km/h	veh/km	veh/h	veh/km	m	s
Dry	7400	80	92.5	2460	30.8	32.4	1.46
Rain	6300	70	92.7	2100	30.9	30.9	1.71

Table 4.3: Characteristics capacity state at dry and wet weather

Chapter 5

Dynamic properties of traffic flow

Summary of the chapter. The preceding chapters have introduced the fundamental properties of traffic flow. Using the correct analytical techniques, such as queuing analysis, shock wave theory, and macroscopic traffic flow models, these fundamental flow characteristics can be provide valuable inisights into the stationary and even dynamic behaviour of traffic flow. However, insight into the dynamic (or rather, transient) properties of traffic flows is also needed for the traffic analyst.

In this chapter, we do provide some insight into these dynamic properties. The main point of interest here are the different types of congestion that been identified and there dynamic properties.

List of symbols

k	veh/m	density
u	m/s	speed
$u^e(k)$	m/s	speed as function of density
$x_i(t)$	m	position of vehicle i as function of time t
T_r	s	reaction time

5.1 Equilibrium traffic state

The different microscopic processes that constitute the characteristics of a traffic flow take time: a driver need time to accelerate when the vehicle in front drives away when the traffic signal turns green. When traffic conditions on a certain location change, for instance when the head of a queue moves upstream, it will generally take time for the flow to adapt to these changing conditions.

Generally, however, we may assume that given that the conditions remain unchanged for a sufficient period of time – say, 5 minutes – traffic conditions will converge to an *average state*. This state is ofter referred to as the *equilibrium traffic state*.

When considering a traffic flow, this equilbrium state is generally expressed in terms of the fundamental diagram. That is, when considering traffic flow under stationary conditions, the flow operations can – on average – be described by some relation between the speed, the density and the flow. This is why the speed - density relation is often referred to as the equilibrium speed $u^e(k)$.

5.2 Hysteresis and transient states

From real-life observations of traffic flow, it can be observed that many of the data points collected are not on the fundamental diagram. While some of these points can be explained by

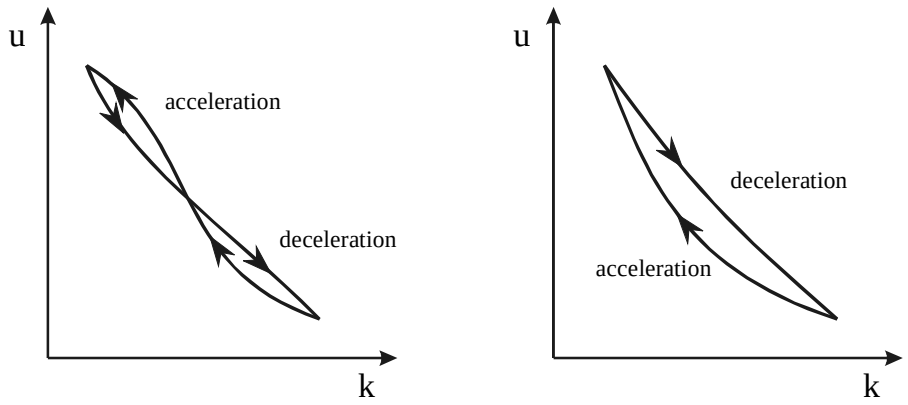

Figure 5.1: Examples of non-equilibrium curves for different data collection sites determined from [54] (left) and [35] (right)

stochastic fluctuations (e.g. vehicles have different sizes, drivers have different desired speeds and following distances), a number of researchers have suggested that these differences are structural, and stem from the dynamic properties of traffic flow. That is, they reflect so-called *transient states* (i.e. changes from congestion to free flow (*acceleration phase*) or from free flow to congestion (*deceleration phase*)) of traffic flow. In turns out that generally these changes in the traffic state are not on the fundamental diagram, that is, we have $u \neq u^e(k)$. In other words: if we consider the average behaviour of drivers (assuming stationary traffic conditions) expressed in the speed-density relation $u^e(k)$, observed mean speeds will generally not be equal to the to the 'equilibrium' speed. The term 'equilibrium' reflects the fact that the observed speeds in time will converge to the equilibrium speed, assuming that the average conditions remain the same. That is, the average speed does not adapt instantaneously to the average or equilibrium speed.

Fig. 5.1 shows examples of non-equilibrium traffic states, where we have indicated *acceleration* and *deceleration* phases. From these figures it can be concluded that no general structure can be found: acceleration phases can either lie above the fundamental diagram or below; this holds equally for deceleration curves. In general, a single acceleration or deceleration curve could not be established for generic situations.

In sum, the following could be concluded:

1. There is more than one non-equilibrium curve
2. Acceleration and deceleration are assymmetric
3. Phase trajectories form hysteresis loops
4. Both acceleration and deceleration curves are nearly smooth
5. Mixing acceleration and deceleration flow generates discontuity

In explaining the position non-equilibrium (transient) states (i.e. either above or below the speed-density curve), [59] distinguishes different transient states, namely:

1. anticipation dominant phase (phase A)
2. balances anticipation and relaxation phase (phase B)
3. relaxation dominant phase (phase C)

He asserts that drivers are usually aware of traffic conditions further downstream. This implies that their will react on downstream high density disturbances before they reach the disturbances (or the disturbance reaches them). In this case, the anticipation is stronger than the relaxation (phase A). On the contrary, under different circumstances a driver may not see that a disturbance is coming until it reaches him / her. In these cases, the response is retarded, and relaxation is dominant. Also mixes of these cases may occur.

These conjectures have been put into a microscopic model. Let $x_i(t)$ denote the position of vehicle i. For the *anticipation dominant phase A*, we assume that the the speed vehicle i is determined by the density k at position $x_i(t) + \Delta x$, i.e.

$$\frac{d}{dt}x_i(t) = v_i(t) = u^e\left(k\left(x_i(t) + \Delta x, t\right)\right) \tag{5.1}$$

Using a first-order approximation, we can easily show that

$$k\left(x_i(t) + \Delta x, t\right) = k\left(x_i(t), t\right) + \Delta x\frac{\partial k}{\partial x}\left(x_i(t), t\right) + O\left(\Delta x^2\right) \tag{5.2}$$

and thus that

$$\begin{aligned} v_i(t) &= u^e\left(k\left(x_i(t), t\right) + \Delta x\frac{\partial k}{\partial x}\left(x_i(t), t\right) + O\left(\Delta x^2\right)\right) &(5.3)\\ &= u^e\left(k\left(x_i(t), t\right)\right) + \Delta x\frac{\partial k}{\partial x}\left(x_i(t), t\right)\frac{du^e}{dk} + O\left(\Delta x^2\right) &(5.4)\end{aligned}$$

In neglecting the second-order term, we get

$$v_i(t) = u^e\left(k\left(x_i(t), t\right)\right) + \Delta x\frac{\partial k}{\partial x}\left(x_i(t), t\right)\frac{du^e}{dk} \tag{5.5}$$

Note that $\frac{du^e}{dk} \leq 0$. Now consider the following situation: a vehicle driving into a region with increasing traffic density, i.e. $\frac{\partial k}{\partial x} > 0$. In that case, the speeds $v_i(t)$ of the vehicles satisfy $v_i(t) \leq u^e\left(k\left(x_i(t), t\right)\right)$ i.e. the speeds are less than the equilibrium speeds: when drivers anticipating on deteriorating traffic conditions downstream, their speed will be less than the equilibrium speed. On the contrary, in situations where drivers are in a region of decreasing traffic density, we have $\frac{\partial k}{\partial x} < 0$ and thus $v_i(t) \geq u^e\left(k\left(x_i(t), t\right)\right)$. That is, drivers anticipate on the improved traffic conditions downstream causing their speeds to be higher than the equilibrium speed.

For the *relaxation dominant phase C*, the speed of vehicle i is determined by the density $k\left(x_i, t\right)$; since the speed, we have

$$\frac{d}{dt}x_i(t + T_r) = v_i(t + T_r) = u^e\left(k\left(x_i(t), t\right)\right) \tag{5.6}$$

To again show the effect of retarded reaction to the traffic conditions, we can use the Taylor approximation

$$v_i(t + T_r) = v_i(t) + T_r\frac{dv_i}{dt} + O(T_r) \tag{5.7}$$

which in turn leads to the following approximation

$$\frac{dv_i}{dt} = \frac{u^e\left(k\left(x_i(t), t\right)\right) - v_i(t)}{T_r} \tag{5.8}$$

The retarded acceleration thus causes an exponential relaxation to the equilibrium speed, rather than an instantaneous one. When a driver moves into a region with increasing densities, the speed will be higher than the equilibrium speed; on the contrary, when moving into a region of decreasing densities, the speed will generally be less than the equilibrium speed.

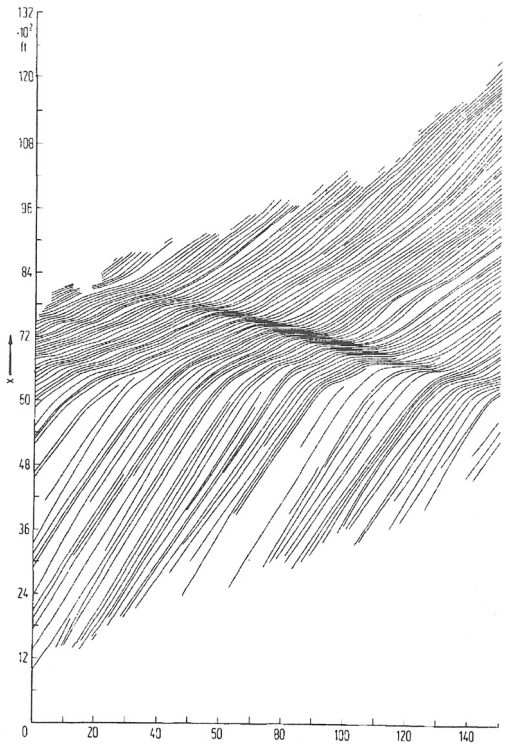

Figure 5.2: Formation of congestion in unstable flow conditions

In combining both phases A and C, we get the combined / mixed phase

$$\frac{d}{dt}x_i\left(t+T_r\right)=u^e\left(k\left(x_i\left(t\right)+\Delta t,t\right)\right) \tag{5.9}$$

It is beyond the scope of this reader to go in further detail; we refer to [59]. It will be shown in chapter 9 that these expressions form the basis for the Payne model, and thus that the emergent flow characteristics can be described using these resulting system of partial differential equations.

5.3 Metastable and unstable states

Several authors have shown that traffic flows are in fact not stable. This implies that for specific traffic flow conditions, small disturbances in the traffic flow may grow into congestion. Fig. 5.2 from [53] shows the formation of congestion in case of unstable traffic flow conditions. From the figure, we can clearly see how a small disturbance grows into a complete, upstream moving jam.

Interesting observations can be made from the San Pablo dataset described ealier. In [add reference to Daganzo here] it is shown how the flow oscillations generated by a traffic signal (the Wildcat Canyon Road) apper to damp out within half a mile of the bottleneck. However, other oscillations arise wihtin the queue farther upstream (at varied locations) which grow in amplitude as they propagate in the upstream direction. In other words, the queue appears to be stable near the bottleneck and unstable far away of the bottleneck.

In [32], transitions between the different phases of traffic is discussed and explained. Based on empirical studies of German motorways, three states of traffic flow are distinguished

1. *Free flow*, which defition is equal to the traditional definition of free flow or unconstrained traffic flow.

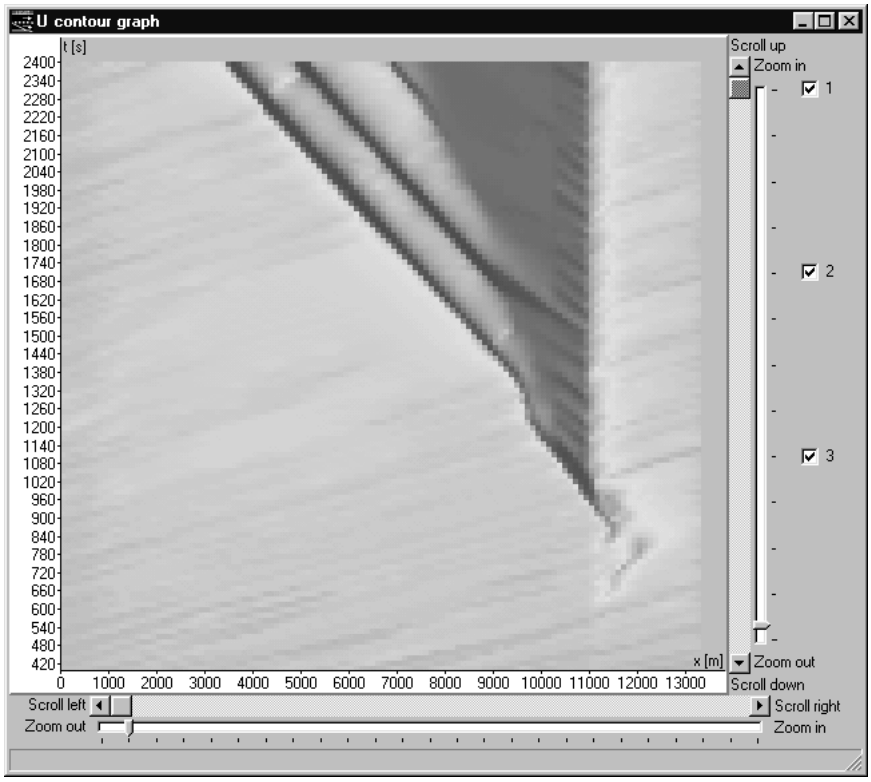

Figure 5.3: Transition from synchronised flow to wide moving jams. Results are obtained from simulations using the microscopic simulation model FOSIM.

2. *Synchronised flow*, which reflects congested (or constrained) flow operations in case traffic inside the congestion is still moving. The synchronised state is characterised by little differences in the average speeds between the lanes and the fact that it in metastable or unstable.

3. *Moving traffic jam* ('wide jam'; [32]) describes a jam in which the traffic is standing still. The traffic jam moves upstream with a fixed speed (approxmately 16 km/h). This speed is determined by the way vehicle accelerate away from the traffic jam.

Synchronised flow occurs among other things at bottle-necks (on-ramp, off-ramp, merge) when traffic demand exceeds infrastructure supply. Over time, the region of synchronised flow may spill back upstream. When the density inside the synchronised flow region is sufficiently high, small disturbances will propagate upstream (this will be explained in detail in chapter 9). The instability of the synchronised state will cause the disturbance to grow in amplitude over time when travelling upstream. In time, the disturbance will turn into a moving traffic jam, given that the region of synchronised flow extends far enough[1]. Fig. 5.3 shows the results of application of the FOSIM model for an oversaturated bottle-neck (in this case, an on-ramp). At the entry of the bottle-neck (at $x = 11150$), a region of synchronised flow forms. Here, the densities are high and the speed is low (albeit non-zero). As the synchronised flow region moves upstream (the dynamics of which can be predicted using shock wave analysis, see chapter 8), the small (stochastic) fluctuations in the flow grow into wide moving jams. FOSIM predicts that these wide moving jams travel upstream with a near constant speed.

Another issue that is address by [32] is the fact that in synchronised flow, drivers will not typically choose one particular distance headway for a certain speed. It is argued in [32] that as a result of this, measurements collected during synchronised flow will generally not lie on a

[1] This is why Kerner refers to this region as the 'pinch region'.

line in the flow-density plane, but will rather form a region. We will not discuss this issue in detail in this reader.

Part III

Analytical techniques

Chapter 6

Capacity and level-of-service analysis

Summary of the chapter. When predicting the performance of a traffic facility, an important question is how much traffic the facility can carry. The fact that the first capacity studies of highway facilities date back as far as 1920 indicates that the issue has been of interest to builders and operators alike for many years. Recently, the field of capacity analysis has been extended to include level-of-service. That is, current analysis represents the trade-off between the quantity of traffic a facility can carry and the resulting level-of-service offered to the user of the facility.

This chapter focusses on capacity of uninterrupted and interrupted freeway sections, and is a summary of the relevant chapter of [36], which in turn describes largely the approach adopted in the HCM [4]. In the Netherlands, a different approach has been used. The basis for the Dutch approach is the application of the microscopic simulation model FOSIM to a large number of bottle-neck situations (on-ramps, weaving sections, etc.) for a large variety of flow conditions. On the contrary to the HCM approach, the Dutch approach does not explicitly consider level-of-service (LOS). Capacity analysis of signalised intersections is an important issue, but is beyond the scope of this course. In the ensuing chapters, we will see the importance of the notion of capacity to estimate travel times and delays using queuing models.

List of symbols

c	veh/s	capacity
c_j	veh/s	lane capacity under ideal conditions; design speed level j
N	-	number of lanes
f	-	capacity reduction factors
h_i	s	time headways
SF_j	veh/s	service flow rate at LOS i
γ	-	weaving influence factor
u	m/s	speeds
v	veh/s	traffic volume (intensity)

6.1 Capacities and level-of-service

Capacity is usually defined as follows [36]

Definition 47 *The maximum hourly rate at which persons or vehicles can reasonably be expected to traverse a point or uniform section of a lane or roadway during a given time period (usually 15 minutes) under prevailing roadway, traffic, and control conditions.*

Although we adopt this definition, it is stressed that several aspects make a practical single definition of capacity complicated. These complications are among other things due to the capacity drop phenomenon, the differences between the capacity of a motorway link (or multilane

facility, basic motorway segment), a motorway bottle-neck (on-ramps, off-ramps, weaving sections), and the stochastic nature of the capacity. The capacity drop has already been discussed in section 4.2.2.

In the US, typical values of the capacity of a freeway with a design speed of 60 or 70 miles/h is 2000 veh/h/lane under ideal conditions; in Europe and especially in the Netherlands, capacities under ideal circumstances are much higher, around 2400 veh/h. Ideal conditions in this case imply 12-foot lanes and adequate lateral clearances; no trucks, buses, or recreational vehicles in the traffic stream; and weekday or commuter traffic. When ideal conditions do not exist, the capacity is reduced. The *Highway Capacity Manual* [4], [5] proposes using the following example relation to express the influence of non-ideal conditions

$$c = c_j N f_w f_{HV} f_p \tag{6.1}$$

where

c	=	capacity (veh/h)
c_j	=	lane capacity under ideal conditions with design speed of j
N	=	number of lanes
f_w	=	lane width and lateral clearance factor
f_{HV}	=	heavy vehicle factor
f_p	=	driver population factor

For example, if ideal conditions existed along a three lane directional motorway having a design speed of 70 mph, the capacity would be

$$c = 2000 \cdot 3 \cdot 1 \cdot 1 \cdot 1 = 6000$$

However, normally ideal conditions do not exist, and in the US typical lane capacities are around 1800 veh/h. In Europe, and in particular in the Netherlands, much higher capacities are encountered. Furthermore, several studies have shown that other factors (such a weather and ambient conditions) also influence the capacity significantly. For instance, section 4.4 showed that the effect of rain on the capacity yields a factor of $f_{weather} = 0.85$; the effects of ambient conditions f_{light} are shown in Tab. 4.2 on page 97; the effect of rain for different road surfaces is shown in Tab. 4.1; other factors are discussed in Sec. 4.1.3.

Capacity is a measure of maximum route productivity that does not address the traffic flow quality or the level-of-service to the users. The *level-of-service* (LOS) reflects the flow quality as perceived by the road users. These flow quality aspects for drivers on the motorway is closely related to the experienced travel times (or travel speeds), the predictability of future traffic conditions (e.g. travel speed, waiting times), and experienced comfrot of the trip (number of stops, required accelerations and decelerations, ability to drive at the desired speed).

To include the user-related traffic flow quality aspects, the concept of *service volume* has been introduced. The service volume SF has a definition which is exactly like capacity except that a phrase is added at the end: *"while maintaining a designated level-of-service"*. In the HCM [4], [5], six service levels ranging from service level A to F are distinguished. Table 6.1 (and Fig. 6.1) shows the definitions of these levels of services. In illustration, if one wishes to operate this particular section of freeway at LOS C, the volume-capacity ratio should be limited to 0.77, and speeds over 54 mph and lane densities of less than 30 veh/mile per lane should results. Speed characteristics, density characteristics, and the relation between these characteristics have been and will be discussed elsewhere in this syllabus. Note that the values in the Netherlands are very different from the values shown in table 6.1. Moreover, the concepts are not only applicable to freeway traffic flow operations, but for instance also in the analysis of pedestrian walking facilities, such as railway stations, sport stadiums, etc.

In the remainder of this chapter, we will focus on different types of motorway facilities, starting with uninterupted multilane roads. Note that for other facilities, the analytical tools are similar.

LOS	Flow conditions	v/c limit	Service volume	Speed	Density
			(veh/h/lane)	(miles/h)	(veh/mile)
A	Free	0.35	700	≥ 60	≤ 12
B	Stable	0.54	1100	≥ 57	≤ 20
C	Stable	0.77	1550	≥ 54	≤ 30
D	High density	0.93	1850	≥ 46	≤ 40
E	Near capacity	1	2000	≥ 30	≤ 67
F	Breakdown	Unstable		< 30	> 67

Table 6.1: Level of service for basic freeway sections for 70 km/h design speed.

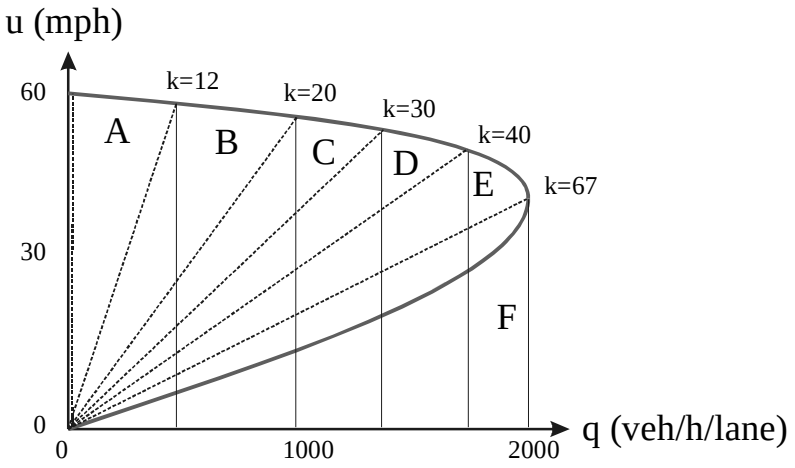

Figure 6.1: Speed-flow relation for a multilane facility for 70mph design speed (from [36])

6.2 Capacity and driver behaviour

Before discussing how the notion of capacity can be applied to basic motorway segments and bottle-necks, let us first describe how the capacity relates to the characteristics of the traffic flow or rather of the driver vehicle combinations in the flow. Recall that for a single lane of the roadway, the flow q can be determined from the headways h_i as follows

$$q = \frac{1}{\frac{1}{n}\sum_i h_i} \tag{6.2}$$

When a roadway lane operates at capacity, this thus implies that most drivers follow each other at the *minimum time headway* (empty zone), say h_i^*. Thus, we have for the capacity of a lane

$$c = \frac{1}{\frac{1}{n}\sum_i h_i^*} \tag{6.3}$$

Note that this relation indicates clearly that the capacity is related to driver behaviour, which explains how aspects like the vehicle fleet composition, lane width and lateral clearance factor, weather conditions, etc., will affect capacity, namely by (changing) the behaviour of drivers. For instance, trucks drivers generally need a larger headway with respect to their leader, due to the length of the truck, as well as larger safety margins for safe and comportable driving.

For multilane facilities, besides the car-following behaviour, the distribution of traffic over the roadway lanes will determine the capacity. Ideally, during capacity operations, all lanes of the roadway are utilised fully, that is, all driver-vehicle combinations are following their leader at the respective minimal headway h_i^*. In practise however, this is not necessarily the case, since the lane distribution will depend on the lane demands and overtaking opportunities upstream of the bottle-neck.

6.3 Multilane facilities

By definition, multilane facilities have two or more lanes available for use (for each direction of travel). The key is that multilane facilities provide uninterrupted flow conditions away from the influence of ramps or intersections. They are often referred to as basic motorway segments. In the approach proposed by the HCM, first capacity analysis under ideal conditions is performed, followed by capacity analysis under non-ideal circumstances. Ideal conditions satisfy the following criteria [4]:

- Essentially level and straight roadway
- Divided motorway with opposing flows not influencing each other
- Full access control
- Design speed of 50 mph or higher
- Twelve-foot minimum lane widths
- Six-foot minimum lateral clearance between the edge of the travel lanes and the nearest obstacle or object
- Only passenger cars in the traffic stream
- Drivers are regular users of such facilities

6.3.1 Capacity analysis under ideal conditions

The speed-flow relationships for multilane facilities have been discussed in chapter 4. These diagrams relate our three scales (flow, density, speed) that are important in LOS analysis. The average speed is an indication of the LOS provided to the users. Traffic flow is an indication of the quantity of traffic that can use the facility. The density is an indication for the freedom of movement of the users. It is noted that the upper density boundary of LOS E (of 67 veh/mile/lane) occurs at the capacity value. Only one congested state in considered in the 1985 HCM LOS classification.

For multilane facilities, the basic equation needed for capacity (and LOS) analysis under ideal conditions is

$$SF_i = \left(\frac{v}{c_j}\right)_i (c_j N) \tag{6.4}$$

where

SF_i = maximum service flow rate for level of service i
j = design speed
c_j = lane capacity under ideal conditions with design speed of j
N = number of directional lanes
$(v/c_j)_i$ = maximum volume-to-capacity ratio for LOS i

The eqn. (6.4) can be used in three ways: 1) by solving for SF_i, the maximum service flow can be determined for a given designed multilane facility under specified LOS requirement; 2) by solving for $(v/c_j)_i$, the LOS can be detemrined for a given designed multilane facility carrying a specific service flow rate. Finally, 3) by solving for $(c_j N)$, the design of a multilane facility can be determined when the LOS and the service flow are specified.

6.3.2 Capacity analysis under non-indeal conditions

The starting point for capacity and LOS analysis for multilane facilities under less than ideal conditions is to go back to eqn. (6.4). Clearly, the factor $(c_j N)$ should be reduced by some factor or a series of factors (compare eqn. (6.1)). Each factor would represent one non-ideal

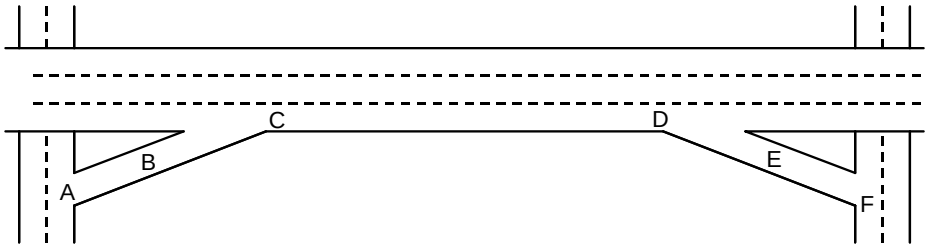

Figure 6.2: Typical motorway configuration (from [36])

condition listed in section 6.3, and multiplied to get a composite reduction factor. It should be noted that in multiplying these factors, we implicitly assume that these factors and independent and that their combined independent effects are multiplicable. In any case, eqn. (6.4) becomes

$$SF_i = \left(\frac{v}{c_j}\right)_i (c_j N)\,(f_1 \times f_2 \times ... \times f_n) \tag{6.5}$$

where $f_1, ..., f_n$ are reduction factors for non-ideal conditions. In the 1985 HCM, four reduction factors are proposed for multilane facilities, namely

1. the width reduction factor f_W, describing the reduction in capacity due to less than ideal lane widths and side clearances,
2. the heavy-vehicle reduction factor f_{HV}, describing the reduction in the capacity (in veh/h/lane!) due to the presence of heavy vehicles *under different vertical alignment conditons*,
3. the driver population factor f_P reflects the reduction in capacity due to the presence of non-regular users, and
4. the environment factor f_E to consider the reduction in capacity due to the lack of a median and/or the lack of access control.

6.4 Ramps

Ramps are sections of roadway that provide connections from one motorway facility to another motorway facility or to another non-motorway facility. Entering and exiting traffic causes disturbances to the traffic on the multilane facilities and can thus affect the capacity and the LOS of the basis motorway segments. Fig. 6.2 shows a typical (schematised) motorway configuration where an on-ramp is followed by an off-ramp. On each ramp, three locations must be carefully studied.

Location A is the entrance to the on-ramp and is affected by the ramp itself and/or by the at-grade intersection. Since the dimensions and the geometrics at location A are (normally) better or at least as good as that of location B, the effect of the physical on-ramp will be studied further at location B. Normally, the at-grade intersection controls the entrance to the on-ramp, and the potential restrictions this causes will not be studied in this course.

Location B is on the on-ramp itself and its capacity is affected by the number and the width of the lanes, as well as the length and the grade of the on-ramp. As long as the on-ramp demand is smaller than the on-ramp capacity, LOS is not really a concern. The reason for this is the relative short length of the ramps.

Locations E and F are "mirror images" of locations A and B in an analytical sense. Location E is the off-ramp itself; similar to the on-ramp, the LOS is not really a concern for the off-ramp. Location F is at the exit of the off-ramp where it connect to a crossing arterial at an at-grade

LOS	Merge flow rate	Diverge flow rate
A	≤ 600	≤ 650
B	≤ 1000	≤ 1050
C	≤ 1450	≤ 1500
D	≤ 1750	≤ 1800
E	≤ 2000	≤ 2000
F	–	–

Table 6.2: Allowable service flow rates for merging and diverging areas (passenger cars per hour) from 1985 HCM.

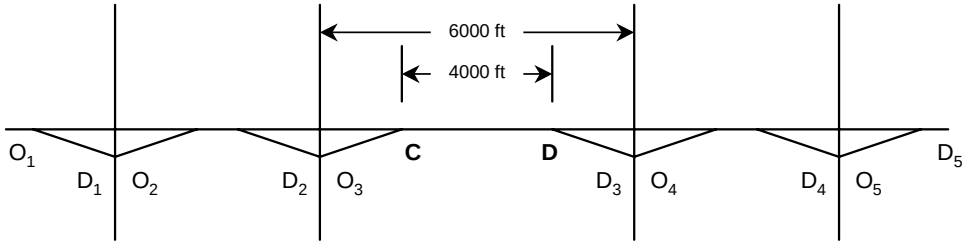

Figure 6.3: Extended typical motorway ramp configuration from [36].

intersection. An important difference between locations A and F is the location of the queues if they exit. At location A, any queues will extend into the at-grade intersection, whereas at location F, any queues will extend up the off-ramp and - if serious enough - into the multilane facility.

Locations C and D are the merge and the diverge areas and require special analytical procedures. The concept and basic principles presented by [36] are straightforward: the substance of the analytical procedure is to compare the actual demands in the merge and the diverge areas with the allowable service flow rates. This comparison is then used to determine the resulting LOS.

Table 6.2 shows an example of allowable service flow rates for merging and diverging areas for ideal conditions for various levels of service. Note that the upper limit of LOS E corresponds to the capacity of the rightmost lane under ideal conditions, which in this case equals 2000 passenger-cars per hour. As noted in table 6.2, the LOS of merge and diverge areas diminish as traffic demands in the rightmost lane increase. These allowable service flow rates should be reduced when non-ideal conditions are considered, using the reduction factors employed for basic multilane facilities. If the capacities and levels-of-service of the basic multilane motorway segment between the merge and the diverge area have been computed, the multilane service flow rates divided by the number of lanes in the basic segment can be used as the allowable lane service flow rates in the merge and diverge analysis.

The major difficulty is in estimating the traffic demands in the rightmost lane. The key to he solution is to consider that traffic in the rightmost lane is made up of subgroups of traffic each having an unique origin and destination along the multilane facility. Fig. 6.3 can be used to illustrate this approach. Table 6.3 shows all possible OD movements. Note that not all OD movements will pass through the merge and diverge areas and can thus be ignored in our analysis. The remaining nine OD movements can be combined into four groups: through, entering, exiting and local. Each will now be addressed in order to determine its share of the traffic demand in the rightmost lane in the vicinity of the merge and diverge areas is question. For demonstration purposes, the distance between the on-ramp nose and the off-ramp gore is assumed to be 4000 feet and its share of traffic in the rightmost lane will be calculated at 1000-foot intervals.

OD	D_1	D_2	D_3	D_4	D_5
O_1	-	-	Exiting	Through	Through
O_2	-	-	Exiting	Through	Through
O_3	-	-	Local	Entering	Entering
O_4	-	-	-	-	-
O_5	-	-	-	-	-

Table 6.3: Possible motorway OD movements.

Through traffic demand	Motorway lanes		
veh/h	8	6	4
≥ 6500	10	-	-
6000-6499	10	-	-
5500-5999	10	-	-
5000-5499	9	-	-
4500-4999	9	18	-
4000-4499	8	14	-
3500-3999	8	10	-
3000-3499	8	6	40
2500-2999	8	6	35
2000-2499	8	6	30
1500-1999	8	6	25
≤ 1499	8	6	20

Table 6.4: Possible motorway OD movements.

Through traffic is traffic that enters the motorway at least 4000 feet upstream of the merge area and exits the freeway at least 4000 feet downstream of the diverge area. Tab. 6.3 shows which OD movements are combined and classified as through traffic, assuming interchange spacing on the order of 1 mile. The question is now to determine how much of this traffic will be in the rightmost lane. May [36] proposes using tables that describe the perceptage of traffic in the rightmost lane for n-lane motorway facilities. Tab. 6.4 shows an example from [36]. The percentages shown in this table are assumed to be constant between the on-ramp and the off-ramp.

Entering traffic is that traffic that enters the motorway in the merge area (location C) and has a destination that is beyond the diverge area (location D); see Tab. 6.3. All entering traffic is in the rightmost lane in the merge area and as the traffic moves farther and farther downstream, a smaller and small proportion remains in the rightmost lane. Fig. 6.4a shows some figures describing the percentage of entering traffic on the rightmost lane. Fig. 6.4b shows the percentage of exiting traffic on the rightmost lane. Finally, local traffic is traffic that enters in the merge area (location C) and exits in the diverge area (location D). Generally, it is assumed that local traffic remains in the rightmost lane.

In sum, the total traffic in the rightmost lane at various locations can be determined by the sum of through, entering, exiting and through traffic. The demand on the rightmost lane is subsequently compared with the allowable service flow rates (such as those given in Tab. 6.2). The highest demand in the vicinity of the merge area is compared with the allowable merge service flow rates, and the highest demand in the vicinity of the diverge area is compared with the allowable diverge service flow rates. The resulting level of service can then be determined.

Although the principles set forth earlier for capacity and LOS analysis of merging and diverging areas are straightforward, their applications can be complicated and tedious. The complications can be caused by unusual ramp geometrics and are particulary difficult at near capacity or oversaturated situations. The 1985 HCM [4] contains many monographs that can

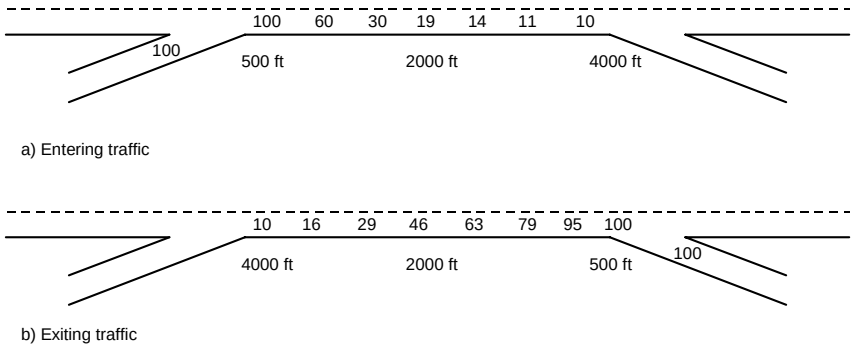

Figure 6.4: Percentage of entering and exiting traffic in rightmost lane, from [36].

be used to estimate the LOS provided in the merge and diverge areas under a wide variety of geometric configurations.

6.5 Weaving sections

Traffic entering and leaving multilane facilities can also interrupt the normal flow of basic motorwau segments by creating weaving sections.

Definition 48 *Weaving is defined as the crossing of two (or more) traffic streams traveling in the same direction along a significant length of motorway without the aid of traffic control devices.*

Weaving vehicles that are required to change lanes cause "turbulance" in the traffic flow and by so doing reduce the capacity and the LOS of weaving sections. Thus, analytical techniques are needed to evaluate this reduction.

6.5.1 HCM-1965 approach

A variety of weaving analysis techniques are available and are being used. Still, it has been recognised that further research on the capacity and LOS of weaving sections is very important. In this section, we show one specific approach to analyse a weaving area in order to show the important factors and arising complications. We will only consider one specific type of weaving section, namely the one shown in Fig. 6.5. Here v_{o1} denotes the heavy flow from A to C (outer flow 1), and v_{o2} denotes the light flow from B to D (outer flow 2). Neither of these flows is a weaving movement; their sum $v_{nw} = v_{o1} + v_{o2}$ is referred to as the total non-weaving flow. The flow from B to C and A to D cross each other's path over a certain distance and are referred to as weaving flows. The higher weaving flow is indicated by v_{w1}; the lower weaving flow is referred to as v_{w2}; their sum $v_w = v_{w1} + v_{w2}$ is referred to as the total weaving flow. The length of the weaving section is denoted by L.

In the approach of the HCM-1965, one first needs to determine if the weaving causes more than the normal amount of lane changing. For instance, when the weaving length L is large and the total weaving flow is small, then only the normal amount of lane changing is expected and the roadway section is "out of the realm of weaving". In the 1965 HCM, the following equation expresses the service flow rate for a specific weaving section

$$SF = \frac{v_{w1} + \gamma v_{w2} + v_{o1} + v_{o2}}{N} \tag{6.6}$$

where

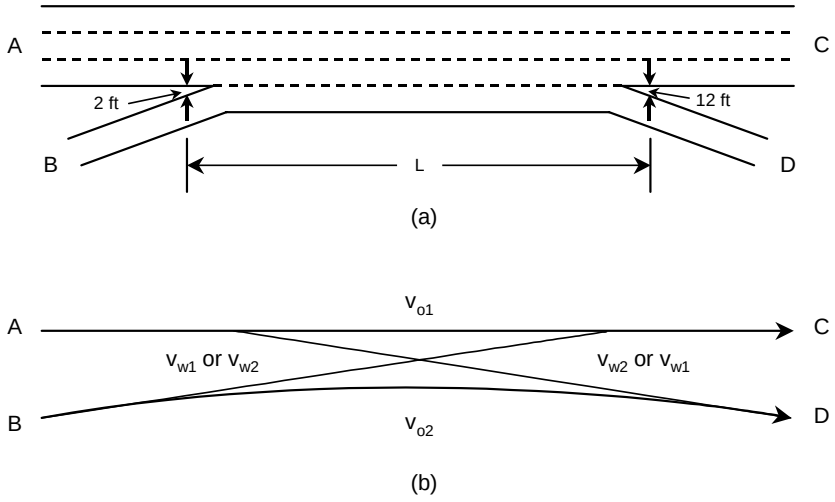

Figure 6.5: Typical simple weaving section (from [36]).

SF = service flow rate
N = number of lanes in the weaving section
γ = weaving influence factor

The weaving infleunce factor γ is a function of the total weaving traffic demand v_w and the length of the weaving section L (see example Fig. 6.6).

6.5.2 HCM 1985 approach

In the 1985 HCM [4], a more comprehensive approach to analyse the LOS is presented. In the approach, three types of weaving sections are distinguished (A, B, and C), as well as the distinction between unconstrained and constrained operations. based on field study results, 12 multiple regression equations where proposed predicting the speed of weaving and non-weaving vehicles. Using these speed predictions, the LOS can be determined.

Weaving sections types

The weaving sections are distinguished based on the required lane changing maneouvres of the weaving vehicles. Type A weaving sections (see Fig. 6.7) require that each weaving vehicle is required to make one lane changing movement, although more than one lane change may be required is weaving vehicles on are not in the correct lane at the start of the weaving section. The minimum number of lane changes equals $v_{w1} + v_{w2}$; the minimum rate of lane changes is equal to $(v_{w1} + v_{w2})/L$.

Type B weaving sections (see Fig. 6.8) require that one waving movement may be accomplished without making any lane changes, while the other movement requires one lane change. This design can be very effective if the minor weaving flow v_{w2} is relatively small. The minimum number of lane changes equals v_{w2}; the minimum lane changing rate equals v_{w2}/L.

Type C weaving sections (see Fig. 6.9) require that one waving movement may be accomplished without making a lane change, and the other waving movement requires at least two or more lane changes. This can be an effective design if the second weaving flow is small, but it can havy very adverse effects if the second weaving flow is too large, the number of lange changes is large, and the weaving length is too short. The minimum number of lane changes equals $2v_{w2}$ (or more if more than two lane changes are required); the minimum lane changing rate is equal to $2v_{w2}/L$.

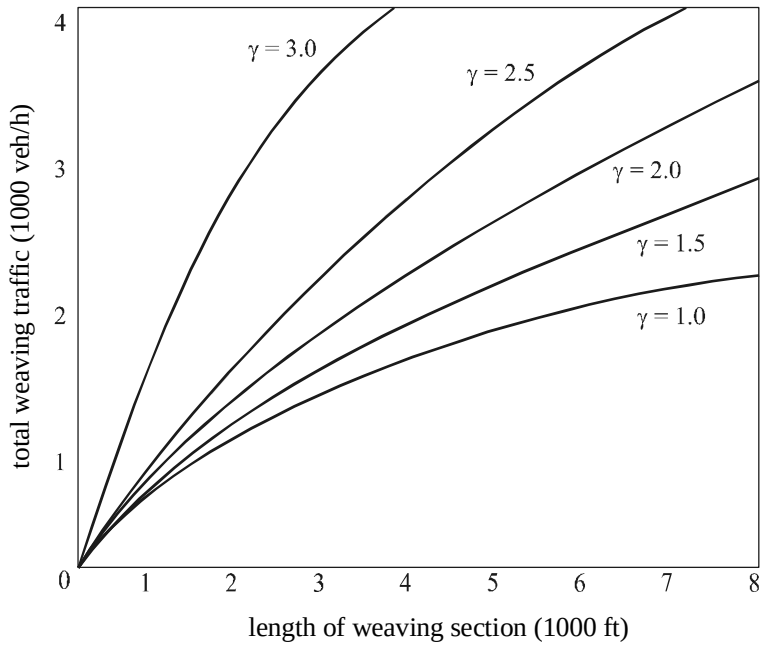

Figure 6.6: Weaving influence factor γ as a function of the length of the weaving area and the amount of weaving traffic

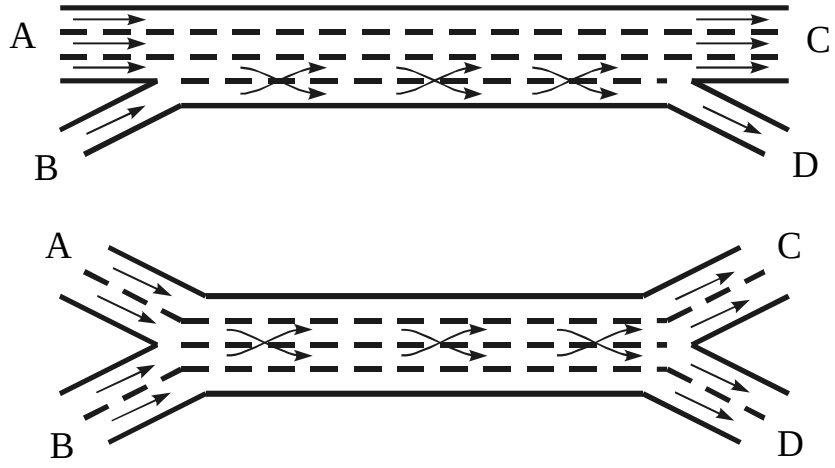

Figure 6.7: Examples of weaving area configuration A.

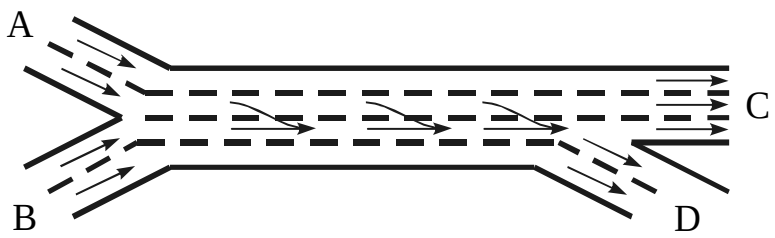

Figure 6.8: Example of weaving area configuration B.

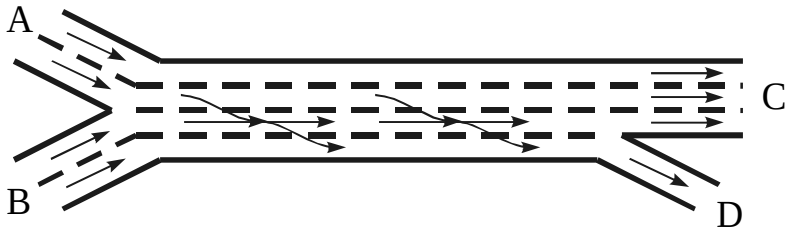

Figure 6.9: Example of weaving area configuration C.

Conf.	N_w	N_w (max)
Type A	$2.19N \cdot \left(\frac{v_w}{v_w+v_{nw}}\right)^{0.571} \left(\frac{(100L)^{0.234}}{S_W^{0.438}}\right)$	1.4
Type B	$N\left(0.085 + 0.703\frac{v_w}{v_w+v_{nw}} + \left(\frac{234.8}{L}\right) - 0.018\left(S_{nw} - S_w\right)\right)$	3.5
Type C	$N\left(0.761 - 1.1L_H - 0.005\left(S_{nw} - S_w\right) + 0.047\frac{v_w}{v_w+v_{nw}}\right)$	3.0

Table 6.5: Criteria for unconstrained and constrained operations of weaving sections. S_{nw} and S_w respectively denote the speed of the non-weaving and weaving vehicles.

Constrained and unconstrained operations

The 1985 HCM approach also distinguishes constrained and unconstrained operations. If the weaving configuration in combination with the traffic demand patterns permits the weaving and non-weaving vehicles to spread out evenly across the lanes in the weaving section, the flows will be somewhat balanced between lanes and the operation is more effective and is classified as unconstrained. On the contrary, if the configuration and demand limit the ability of weaving vehicles to occupy their proportion of available lanes to maintain balances operations, the operations is less effective and is classified as constrained. Consider for instance the weaving section shown in Fig. 6.5: if the flow from A to C is relatively light and the other flows are relatively heavy, the lanes on the left side of the weaving section will be underutilised and the lanes on the right side will be overutilised. Such imbalanced or constrained operations will result in weaving vehicles travelling at lower speed (hence lower LOS) and non-weaving vehicles travelling at a higher speed.

Determination of the type of operation is done by comparing two variables, namely N_w (number of lanes that must be used by weaving vehicles in order to achieve balanced or unconstrained operations) and N_w (max) (maximum number of lanes that may be used by weaving vehicles for a given configuration). If $N_w < N_w$ (max), the operation is defined as unconstrained, while if $N_w > N_w$ (max) the operation is defined as constrained. Based on empirical observations, procedures to compute these variables are shown in Tab. 6.5.

The next step in the analysis is to select appropriate multiregression type equations for prediction weaving and non-weaving speeds based on types of weaving configurations and types of operations. Again, empirically derived equations have been determined and can be found in the 1985 HCM. These are listed in Tab. 6.6. These parameters can subsequently be substituted in the following equation

$$S_w \text{ (or } S_{nw}) = 15 + \frac{50}{1 + a\left(1 + \frac{v_w}{v_w+v_{nw}}\right)^b \left(\frac{v}{N}\right)^c / L^d} \tag{6.7}$$

Note that the speeds of weaving and non-weaving vehicles are also required to decide between constrained and non-constrained operations, yet these speeds have not yet been determined. The suggested approach is to first assume unconstrained operations, calculate weaving and non-weaving speeds and then use the equations in Tab. 6.5 to see if the assumption of unconstrained

Conf.	Parameter values for S_w				Parameter values for S_{nw}			
and operation	a	b	c	d	a	b	c	d
A - unconstrained	0.226	2.2	1.00	0.9	0.020	4.0	1.30	1.00
A - constrained	0.280	2.2	1.00	0.90	0.020	4.0	0.88	0.60
B - unconstrained	0.100	1.2	0.77	0.50	0.020	2.0	1.42	0.95
B - constrained	0.160	1.2	0.77	0.50	0.015	2.0	1.30	0.90
C - unconstrained	0.100	1.8	0.80	0.50	0.015	1.8	1.10	0.50
C - constrained	0.100	2.0	0.85	0.50	0.013	1.6	1.00	0.50

Table 6.6: 1985 HCM parameter values for determination of speeds of weaving and non-weaving traffic.

LOS	Minimum S_w	Minimum S_{nw}
A	55	60
B	50	54
C	45	48
D	40	42
E	35	35
F	30	30

Table 6.7: Level of service criteria for weaving sections.

operations is correct. If not, the process is repeated assumig constrained operations. The final step in determining the LOS of the weaving section is to enter Tab. 6.7 with the predicted weaving and non-weaving speeds.

6.6 Dutch approach to motorway capacity analysis

The Dutch design procedure focuses on the achievable capacity of a weaving section. The level of service is not estimated: as a rule of thumb, the weaving segment design should be based on a demand-capacity ratio equal or below 0.8. This value is accepted by roadway designers as a good design value for freeway facilities in the Netherlands.

The Level of Service of a freeway facility or network can be determined by additional procedures that take into account the probability of congestion over a year. The background of the method is a cost-benefit analysis, given the distribution of capacity and expected traffic demand pattern over a year. Currently, new guidelines for the Level of Service of freeways are being developed using the average travel speed over a 10-20 km traject as variable. However, these Dutch approaches for level of service calculation cannot be compared to the LOS definition in the HCM, as they relate to a facility or network, rather than the segment based HCM approach. Capacity is defined similarly as in the HCM, except that a different analysis time interval is used. Dutch guidelines consider a 5-minute time interval, because this value is used in the simulations using the Dutch freeway trafic flow simulator FOSIM. Capacity refers to a *pre-queue capacity value*.

As an example, let us consider determination of the capacity of a weaving section. The notation used in describing weaving sections is straightforward. Two types of weaving sections are distinguished: symmetrical and asymmetrical types:

- A symmetrical weaving section '2+1' indicates that leg A has 2 incoming lanes, and leg B has 1 lane. The section is said to be symmetrical, because the number of lanes in legs C and D are equal to that of A and B respectively.
- A weaving section '2+2 $\rightarrow$ 3+1' indicates that legs A and B have 2 incoming lanes each, and the weaving section geometry is asymmetrical, with leg C having 3 lanes and leg D

having 1 lane.

By means of microscopic simulation, several tables have been established giving capacity values as function of:

- Weaving configuration types: symmetrical 1+1, 2+1, 3+1, 4+1, 2+2, 3+2, 4+2 (which are similar to the HCM weaving section type A), and several asymmetrical designs.
- Free flow speed (120 km/h or 75 mph)
- Truck proportion (with a range of 5-15%);
- Length of weaving segment (the range depends on the type of weaving segment – lenghts considered are: 100 – 1000 m)
- Weaving ratio of the leg with the smallest incoming flow (LR) (range depends on the type of weaving segment)

The capacity values in the tables are expressed in vehicles/hour. This value can be converted into pcph by using a pcu-truck value of 1.5 according to the Dutch guidelines. The overview table is published in the Dutch guidelines for freeway capacity, the so-called *CIA-manual*. In 2001 the asymmetrical types (which can be compared with HCM type C) were also studied using the microscopic simulation model FOSIM. For these types also capacity tables have been established, and are available. The calibrated and validated microscopic simulation model FOSIM was used to calculate capacity values for a wide range of weaving segment designs. The following procedure was repeated for every scenario (weaving designs varying in weaving section geometry, truck proportion, weaving segment length) in order to determine the capacity values:

1. Run simulation for a specific scenario until congestion occurs upstream or on weaving segment; Stop simulation and determine maximum 5-minute flow rate.
2. Repeat step 1 50-100 times applying different random number seed values. This results in a distribution of capacity values.
3. Determine median capacity value of the performed simulation runs. This value is denoted as 'capacity'. An important aspect to be considered is the approach followed in the simulation of a scenario. In all the simulations the weaving flow rate is equal for both legs. Thus the flow rate on the origin-destinations AD and BC was nearly equal.

6.7 Stochastic nature of motorway capacity

Maximum flows (maximum free flows of queue discharge rates) are not constant values, and vary under the influence of several factors, as was disussed in this and previous chapters. Factors influencing that capacity are among other things the composition of the vehicle fleet, the composition of traffic with respect to trip purpose, weather-, road-, and ambient conditions, etc. These factors affect the behaviour of driver vehicle combinations and thus the maximum number of vehicles that can pass a cross-section during a given time period. Some of these factors can be observed and their effect can be quantified. Some factors can however not be observed directly. Furthermore, differences exist between drivers implying that some drivers will need a larger minimum time headway than other drivers, even if drivers belong to the same class of users. As a result, the minimum headways h_i^* will not be constant values but follow a distribution function (in fact, the empty zone distribution $p_{fol}(h)$), as was explained in chapter 2. As a result, capacity will also be a random variable following a specific distribution. The shape of this distribution depends on among other things the capacity definition and measurement method / period. In most cases, a Normal distribution will can be used to describe the capacity.

6.8 Capacity drop

In section 4.2.2, we have discussed the existence of two different maximum flow rates, namely *pre-queue capacity* and *queue discharge rate* respectively. Each of these have their own maximum flow distribution.

Definition 49 *We define the pre-queue maximum flow as the maximum flow rate observed at the downstream location just before the on-set of congestion (a queue) upstream. These maximum flows are characterised by the absence of queues or congestion upstream the bottle-neck, high speeds, instability leading to congestion on-set within a short period, maximum flows showing a large variance [37].*

Definition 50 *The queue discharge flow is the maximum flow rate observed at the downstream location as long as congestion exists. These maximum flow rates are characterised by the presence of a queue upstream the bottle-neck, lower speeds and densities, a constant outflow with a small variance which can sustain for a long period, however with lower flow rates than in the pre-queue flow state [37].*

Both capacities can only be measured *downstream* of the bottle-neck location. Differences between the two capacities (so-called *capacity drop*) are in the range of -1% to -15%. Different explanations for the capacity drop can be and have been given. Dijker [16] argues that the main reason is the preference for larger headways if drivers experience congested conditions. Differences between acceleration and deceleration behaviour may also contribute to this phenomenon.

6.9 Capacity estimation approaches

To determine the capacity of a bottle-neck or a basic motorway segment, appropriate capacity estimation techniques are required. These techniques can be classified in techniques that do not require capacity observations and those who do. The former methods, which are based on free flow traffic *and* constrained traffic measurements are generally less reliable than methods using capacity measurements. If the capacity state has not been reached and a capacity estimation must be performed, the following approaches are applicable:

1. *Headway distribution method.* The observed headway distribution is used to determine the distribution of the minimum headway h_i^*, which in turn is used to estimate a single capacity value (no distinction between pre-queue capacity and queue-discharge rate). See chapter 2.

2. *Fundamental diagram method.* This approach uses the relationship between speed and density or flor rate to estimate the capacity value. A functional form needs to be selected and assumptions about the critical density must be made. See section 4.3.3.

Methods using explicitly capacity flows sometimes use additional flow measurements in order to get an improved capacity estimate. Some methods do not distinguish between queue and pre-queue maximum flows.

1. *Selected maxima method.* Measured flow rate maxima are used to estimate a capacity value or distribution. The capacity state must be reached during each maxima selection period. Approach should be applied over a long period.

2. *Bimodal distribution method.* This method may be applied if the observed frequency distributions of the flow rates exhibit a clear bimodal form. The higher flow distribution is assumed to represent capacity flows.

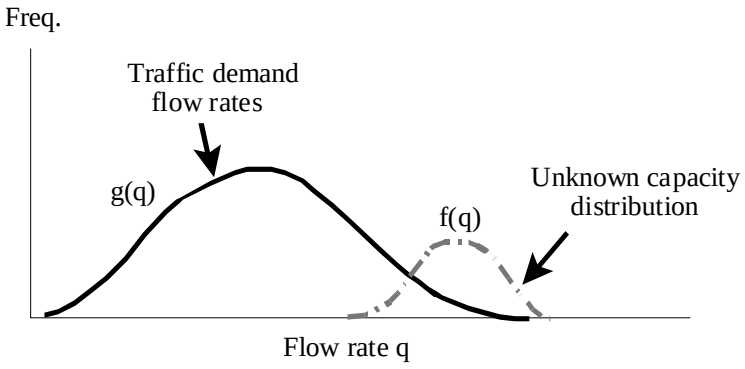

Figure 6.10: Example of probability density functions of flow rates divided into demand and capacity flows.

3. *Queue discharge distribution method.* This is a very straightforward method using queue discharge flow observations to construct a capacity distribution or a capacity value. The method requires additional observations (for instance, speeds upstream of the bottle-neck) to determine the congested state.

4. *Product-limit method.* This method uses below-capacity flows together with capacity flows to determine a capacity value distribution. Speed and / or density data is needed to distinguish the type of flow measurement at the road section upstream of the bottle-neck.

6.9.1 Product-limit method to capacity estimation

Formally, an observation is said to be 'right censored' at flow rate q if the unknown capacity value c of the observation is only known to be greater than or equal to q. This type of censoring is very common in lifetime data, and obviously apparent in capacity estimation problems. Mostly, in lifetime data analysis continuous models and data are considered.

Let $i = 1, 2, ...n$ indicate the observation period and n the total number of observed periods. A traffic flow rate in a bottle-neck section at observation period i is denoted with q_i. Simultaneous observations at the upstream section gives information over the traffic conditions. We denote δ_i as an *indicator for the type of measurement at the bottle-neck*, based on the observations at the upstream section. Two conditions are distinguished:

- $\delta_i = 0$ (capacity flow at bottle-neck in period i, i.e. congested conditions upstream);
- $\delta_i = 1$ (traffic demand flow at bottle-neck in period i, uncongested conditions upstream)

Each observation period i is assumed to have its own specific capacity value. However, we can only measure a capacity value if the bottle-necks capacity has been reached. If the bottle-necks capacity has not been reached, we are observing traffic demand. Nonetheless, this traffic demand value can be used in the Product-Limit capacity estimation procedure since it gives valuable information about the location of the capacity: the capacity value will be higher than the observed volume.

Based on this principle, we can present the maximum likelihood for a sample of observations. Firstly, let us denote functions (and their properties) which will be used in the estimation. See also Fig. 6.10.

Capacity observations are assumed to be *identically and independently distributed* with probability density function $f(q)$ and survival function $S(q)$. The survival function $S(q)$ is equal to $1 - F(q)$, where $F(q)$ is the cumulative distribution function of $f(q)$, i.e. $F(q) = \int_{-\infty}^{q} f(x)\, dx$.

The independence requirement will be met by selecting observation periods between 5 and 15 minutes.

Traffic demand observations do also have a cumulative distribution function, survival function and p.d.f., denoted respectively with $G(q)$, $K(q)$ and $g(q)$. Their shape strongly depends on the total observation time and hours of the day selected for analysis, therefore the choice of a functional form and estimation of its parameters is not relevant. The problem now is to estimate the actual – unknown – capacity distribution given the capacity survival function $S(q)$ and p.d.f. $f(q)$. Then, for a sample the likelihood is:

$$L = \prod_{i=1}^{n} f\left(q_i\right)^{1-\delta_i} S\left(q_i\right)^{\delta_i} \tag{6.8}$$

This expression can be easily understood by noticing that when an observation q_i is a capacity observation ($\delta_i = 0$), this will occur with probability $f\left(q_i\right)$; if not ($\delta_i = 1$), then the flow is less than the capacity, which occurs with probability $\Pr\left(C \geq q\right) = S\left(q\right)$. In other words, each observed capacity contributes a term $f(q)$ to the likelihood, and each below-capacity volume contributes a term $S(q)$.

Parametric Product Limit Method

Let θ denote the parameters of the capacity distribution function. Then, the likelihood function for sample q_i can be written as a function of θ. By maximizing the likelihood we can estimate the parameters θ of the capacity distribution, i.e., we can determine the parameters θ of the probability distribution function $f\left(q\right) = f\left(q;\theta\right)$ – and thus also of the survival function $S\left(q\right) = S\left(q;\theta\right)$ – as follows

$$\theta = \arg\max\left\{L\left(\theta\right)\right\} \tag{6.9}$$

In most cases, we would use the natural logaritm of the likelihood function $L\left(\theta\right)$, i.e.

$$\tilde{L}\left(\theta\right) = \ln L\left(\theta\right) = \sum_{i=1}^{n}\left\{\left(1-\delta_i\right)\ln f\left(q_i;\theta\right) + \delta_i \ln S\left(q_i;\theta\right)\right\} \tag{6.10}$$

To use this expression, one first needs to determine a good functional form of the probability density function of the capacity. In most cases, a Normal distribution with mean μ and standard deviation σ (i.e. $\theta = (\mu, \sigma)$) will be a good first approximation.

Non-parametric Product Limit Method

For roadway capacity estimation a *non-parametric form* of the capacity survival function S may be used, rather than the parameterised one presented in the previous sections. This is prefered over the parameterised one, since there is no real evidence for the choice of a particular functional form of the capacity survivor function. If there have been only capacity flow observations in a sample of size n, the empirical survival function $\hat{S}_n\left(q\right)$ is defined as:

$$\hat{S}_n\left(q\right) = \frac{1}{n}\left\{\text{Number of observations } i \text{ with } q_i \geq q\right\} \tag{6.11}$$

This cumulative frequency function decreases by $1/n$ just after each observed 'lifetime'. One may represent observed capacity values as observed lifetimes and use this equation to estimate the roadway capacity. We may choose the 50% percentile (median) as the representative capacity value, or any other percentile point.

A straightforward approach would involve only using capacity observations ($\delta_i = 0$). The resulting approach would be a special case of the generic approach described here, the Product Limit Method (PLM). When dealing with both censored and uncensored data, some modification is necessary, since the number of capacity values greater than or equal to q will not

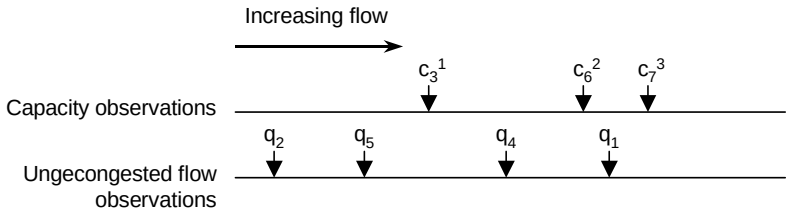

Figure 6.11: Capacity and uncongested flow observations

generally be exactly known. This modified form is called the Product-Limit estimate of the survival function, and is given by

$$\hat{S}_{PLM}(q_i) = \prod_{j=1}^{k} \frac{m_j - d_j}{m_j} \quad j = 1, ..., k \text{ and } 1 \leq k \leq n \tag{6.12}$$

In Eq. (6.12), index j indicates the ordering of the observed capacity flow rates c_i according to increasing size. Since index i indicates the observation period, we denote the size-ordered capacity values c_i with an extended index j, thus c_i^j. Obviously, since there must be at least one capacity observation to apply the method and there will not be more capacity observations than the total number of observation periods, $1 \leq j \leq n$. The total number of capacity observations is denoted with k, and is smaller than or equal to n. Furthermore in equation (6.12), m_j is the sum of

1. the number of capacity observations that *are at least as high as* c_i^j (thus including c_i^j) *and*
2. the number of uncongested observations q_i that have a higher value than c_i^j

At last, d_j represents the number of (exactly equal) capacity values observed at ordering index j, *which is mostly equal to one.* The multiplication is performed from the lowest capacity value observation $j = 1$ to the highest capacity value observation k, thus j increases from 1 to k. *This multiplication makes clear that only at observed capacity flow rates a survival probability is calculated.* The decrease at these points is not equal to $1/n$, as with the empirical distribution, but depends on the number of observations with flow rates q higher than capacity value c_i^j. This number m_j includes higher-value censored observations q_i, but higher capacity values as well.

Let us illustrate formula (6.12) with a sample of 7 observations (in Fig. 6.11 below). Four uncongested flow rates q were observed, and three capacity flows c. Thus $n = 7$ and $k = 3$. It appears that $d_j = 1$ since no equal capacity observations occurred. When applying equation (6.12) to the example shown in Fig. 6.11, the first calculation to be performed is for capacity observation c_3^1. This is the *lowest observed capacity value*, denoted with index $j = 1$. The observation was made in period 3. At this capacity value, there are *two uncongested flow rates* (q_2 and q_5) observed, and two capacity values (c_6^2 and c_7^3) having higher flow rates. Thus, including the observed capacity value, this means $m_1 = 5$. Now the probability that the capacity is at least as high as value c_3^1 can be calculated using Eq. (6.12), and equals $\frac{4}{5}$. For the next capacity value c_6^2 the same procedure applies. Now $m_2 = 3$, which would yield a survival probability of $\frac{4}{5} \cdot \frac{2}{3} = \frac{8}{15}$. The last calculation to be performed in this example is for capacity observation c_7^3. Since $m_3 = 1$ the multiplication result in a zero value, which is according to the expectations.

The motivation for the (discrete) Product-Limit estimate is essentially the same as that for the continuous approach. That is, the estimate $\hat{S}_{PLM}(c)$ is built up as a product, and each term in the product can be thought of as an estimate of the conditional probability of 'surviving' capacity flow rate c_i^j, given survival till just prior to c_i^j. It will be noted that when there is no censoring, the equation reduces to the ordinary empirical survivor estimate $\hat{S}_n$ given above.

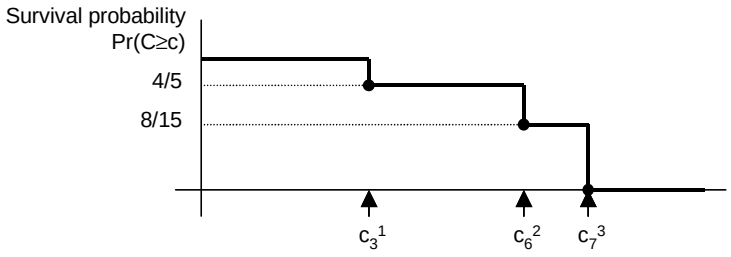

Figure 6.12: Estimated survival probabilities using PLM method.

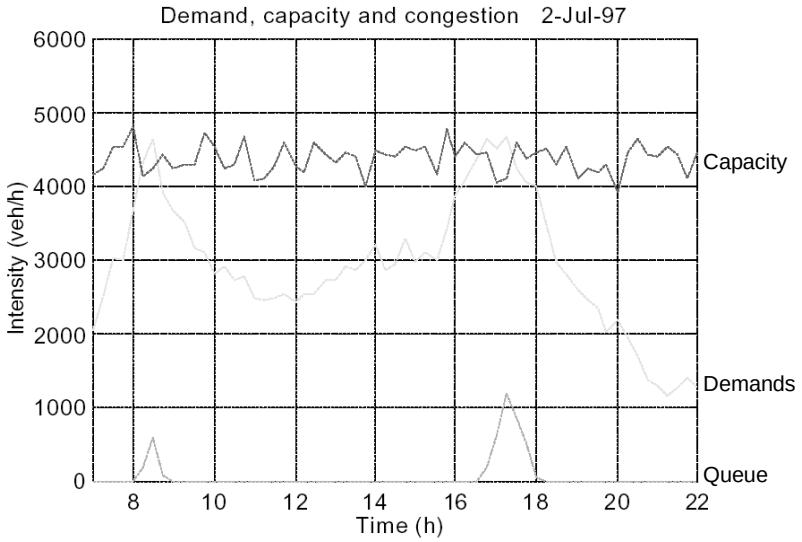

Figure 6.13: Simulation example to application of the PLM method

To effectively assess results when using PL estimates it is desirable to have an estimate of the variance. It can be shown that the variance estimate for this estimate equals

$$var\left[\hat{S}_{PLM}\right] = \hat{S}^2_{PLM} \sum_{j=1}^{k} \frac{d_j}{m_j\left(m_j - d_j\right)} \tag{6.13}$$

6.9.2 Example application of the PLM method

In order to demonstrate the validity of the PLM we are forced to use a simulated case. Fig. 6.13 shows three curves, respectively, capacity, traffic demand and congestion (queue) curve describing the flow characteristics of a sample day for the constructed case. The applied mean capacity value was set at 4400 vehicles/hour, which is a representative value for two-lane freeways in the Netherlands. A normal distribution with a standard deviation of 5% was chosen to generate stochastic road capacity values for each 15-minute period. The 15-minute interval is considered long enough to assume identically and independently distributed capacity values.

Apart from the capacity, the applied traffic demand curve has stochastic characteristics as well. However, account is taken of the correlation between the deviation from the mean demand in succeeding time intervals, smoothing the path. Queueuing directly results from the excess demand with respect to the prevailing capacity. The impact of the stochastic nature of both capacity and traffic demand on the congestion severity yields that capacity values are in the range of 3800 to 5000 veh/h. Therefore, traffic demand sometimes exceed capacity in which case congestion does occur.

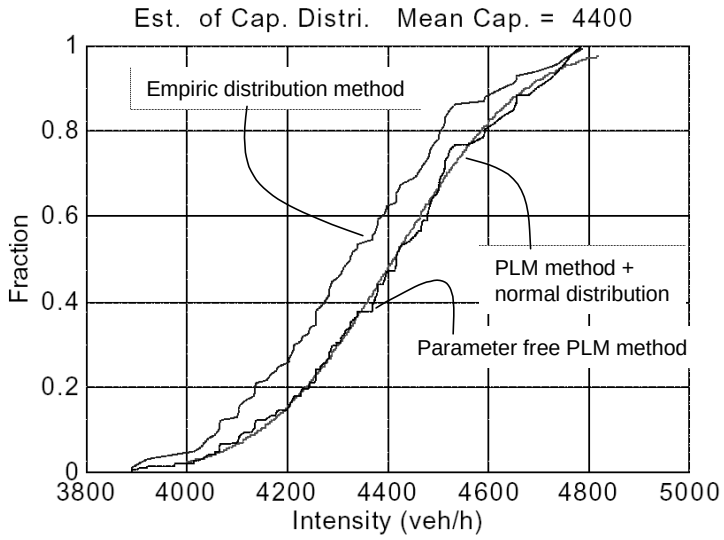

Figure 6.14: Estimates for the capacity distribution functions based different estimation methods.

The capacity value distribution was estimated based on three methods. The 'measurement' configuration is according to that of one cross-section in the bottle-neck section. Since no backward wave disturbance was simulated, it was suitable for our analysis purpose. The measurement interval equals that of the generated traffic data, that is 15-minutes.

In the simulated example, about 8% of the observed flow rates were capacity flows. The range of capacity flow values strongly overlaps with the higher range of uncongested flow values.

Three capacity distributions were calculated:

- Empirical Distribution, which is the distribution of all observed capacity flow rates;
- Estimated distribution of capacity using PLM and normally distributed capacity values;
- Estimated distribution of capacity using PLM and a parameter free distribution of capacity values;

These three estimated cumulative distributions are shown in Fig. 6.14 where one can read from the figure the differences in the capacity estimate at any desired fraction. In order to be regarded as a sound outcome, the estimated capacity values at fraction 0.5 should be close to our input value of 4400 veh/h. By looking at the three values; the empirical, PLM-normal and PLM-parameter free curves, median capacity values of approximately 4300, 4400, 4400 veh/h respectively can be found. It can be concluded that – the easily and generally applied – empirical distribution of capacity values *underestimates the 50%-percentile true value significantly.* Both forms of the Product-Limit Method, however, result in good estimations of mean capacity.

6.9.3 Other applications of the PLM method

The PLM method is not only applied for capacity estimations: it is generally applicable to problems where censored observations are present. Consider for instance the notion of *free speeds of a population of drivers*: for some drivers, we know that they are driving at their free speed, while for other drivers (constrained by the vehicle in front), we may assume that their free speed is higher than their current speed. The main problem is then to distinguish between censored (constrained drivers) and uncensored (free drivers) data, since the estimation results will be sensitive to a correct disinction between the two. A similar problem occurs in *gap acceptance analysis*: only the gaps that are accepted are monitored.

6.10 Summary

This chapter serves as an introduction to the analytical techniques for capacity and level-of-service determination of critical elements of the motorway system. Heavy emphasis has been placed on the use of the 1985 Highway Capacity Manual. However, the field of capacity analysis is not limited to motorway facilities but also includes other land transport modes as well as air and water transportation.

Chapter 7

Queuing analysis

Summary of the chapter. In this chapter the deterministic queuing model will be introduced briefly and applied to determine congestion probabilities at a single bottle-neck over a year. This is of importance because the Dutch method to assess quality of traffic operation on motorways is based on this model.

List of symbols

D	veh/s	arrival rate
C	veh/s	capacity / queue discharge rate
P_c	-	fraction of vehicle experiencing congestion
R_t	s	collective delays
R_{mean}	s/veh	mean delay per vehicle
F	-	demand multiplication factors
ε	-	errors
f	-	capacity multiplication factors
n	veh	number of vehicles in queue

7.1 Deterministic queuing theory

The queueing model to determine delay is not a realistic description of the real traffic process, the main deviation being that vehicles are stored vertically in a queue. Nevertheless with this model the delay can be calculated correctly. In chapter 2, we have introduced the notion of the cumulative vehicle plot and its applications. Let us briefly recall some of the key elements of this chapter, and introduce some new ones along the way. Recall that $\tilde{N}(x_1, t)$ denotes the arrival curve at some cross-section x_1. $\tilde{N}(x_2, t)$ denotes the departure curve at some other cross-section x_2. The arrival time of vehicle i at cross-section x_1 is denoted by $A^{-1}(i)$; its departure time by $D^{-1}(i)$. Clearly, the travel time of vehicle i equals – assuming that overtaking is not possible –

$$R_i = N^{-1}(i; x_2) - N^{-1}(i; x_1) \tag{7.1}$$

The *collective travel times* for all vehicles $i = 1, ..., N$ then equals

$$R = \sum_{i=1}^{N} R_i = \sum_{i=1}^{N} \left[N^{-1}(i; x_2) - N^{-1}(i; x_1) \right] \tag{7.2}$$

which approximately equals the area between the arrival and the departure curve (i.e. neglecting the small errors made at the boundaries), i.e.

$$R = \int_{t=0}^{T} \left[N(x_1, s) - N(x_2, s) \right] ds \tag{7.3}$$

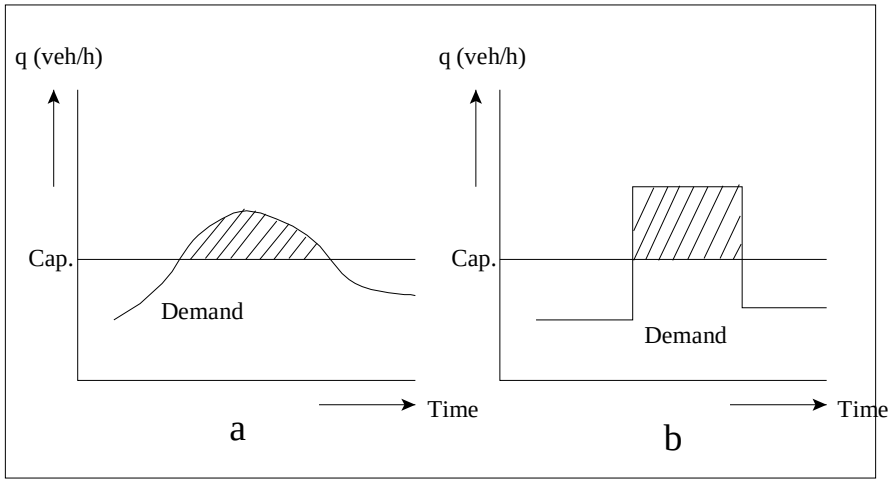

Figure 7.1: Temporarily oversaturation: (a) realistic; (b) schematised

Note that overtaking does not compromise the determination of the collective travel times R, but overtaking does imply that the individual travel times cannot be adequately determined. In chapter 2 we have also shown how we can determine the delays experienced by drivers / collective delays by introducting the virtual departure curve. This curve is equal to the arrival curve, translated along the t axis.

7.2 Determination of delay with a queueing model

Let us consider a bottle-neck situation sketched in Fig. 7.2. As long as the demand flow is less than the capacity flow, we can determine the traffic flow conditions using the fundamental diagram. But what happens when the demand exceeds the capacity? Let us consider both situations in detail.

Situation A: no oversaturation. As long as the demand D is smaller than the bottle-neck capacity C there is no congestion. On the entire road, free flow conditions exists $q(x,t) = D$. In the bottle-neck, a small delay may be incurred due to the higher densities.

Situation B: oversaturation. When the demand D exceeds the capacity C, vehicles can not pass the bottle-neck without experiencing delays. The next chapter on shock wave analysis with provide an in-depth discussion of the dynamics of the traffic flow conditions in this situation. However, we can also approximate the queue dynamics by using queuing theory, as is described in the remainder of this chapter.

To this end, it is assumed that all vehicles drive unhindered (without delay) to the position where the bottle-neck begins. They wait at this spot, as said before in a vertical queue, for their turn to drive through the bottle-neck. The growth of the queue depends on the difference between demand D and capacity of the bottle-neck C. The queue starts growing when for the first time D is larger than C. This process is usually visualised in a graph with time on the horizontal axis and the cumulative number of vehicles on the vertical axis (we have already addresses cumulative arrival curves in chapter 2). The so called 'arrival' curve is a line with demand D as slope. The 'departure' curve has the capacity C as slope; see Fig. 7.3 - at least when there is a queue.

The distance between the arrival curve and the departure curve at a given moment (the vertical distance) equals the length of the queue (expressed in number of vehicles). The distance between the arrival curve and the departure curve for a given arriving vehicle (the horizontal distance) equals the time the vehicle spends in the queue, assuming the queue discipline is 'First in, first out' (FIFO) implying no overtaking.

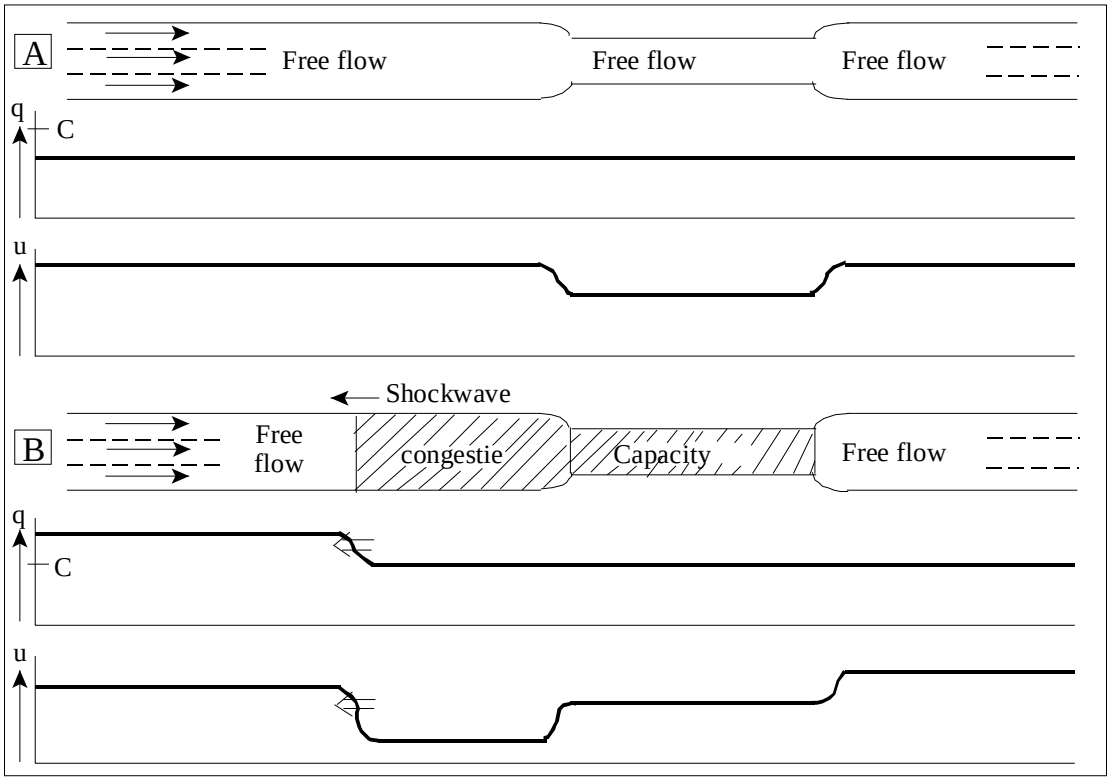

Figure 7.2: Characteristics of a traffic flow on a road with a bottleneck

Remark 51 *Both the length of the queue and the time spent in the queue are not realistic but refer to the vertical queue.*

When D becomes smaller than C, the slope of the arrival curve becomes smaller than the slope of the departure curve. When both curves intersect, the congestion is finished.

The delay equals the area between the arrival and the departure curve, because the area equals the sum of all horizontal distances between the arrival and the departure curve.

It is evident that congestion does not end at the moment D becomes smaller than C; the queue that has been built still has to be broken down. The rate of decay, like the growth rate, depends on the difference $D - C$. The (vertical) queue is at its maximum at the moment D becomes smaller than C.

This model determines the (collective) delay correctly, but not how the queue is present over the road upstream of the bottle-neck. In reality vehicles experience their delay by slowing down and possibly stopping now and then for short periods. It is important to understand that the delay only depends on the capacity of the bottle-neck and the undisturbed (by definition) demand as a function of time. This holds as long as the congestion itself does not induce extra disturbances which lead to a lower capacity.

Fig. 7.4 illustrates that the delay is independent on the form of the vehicle trajectories in congestion between cross-section x_1, upstream of the congestion, and cross-section x_2, the begin of the bottle-neck. The delay is the difference between the moment the vehicle drives through the beginning of the b-n and its virtual moment of arrival at the same cross-section. From the figure can also be concluded that overtakings in congestion do not change the collective delay. The smaller delay of the overtaker is compensated by the larger delay of the vehicle being overtaken.

Remark 52 *A part of the delay is not accounted for by the queueing model. It is the delay due to the fact that the vehicles usually have to decelerate when they enter the b-n. However, if the queueing has any relevant size, then this extra delay can usually be neglected.*

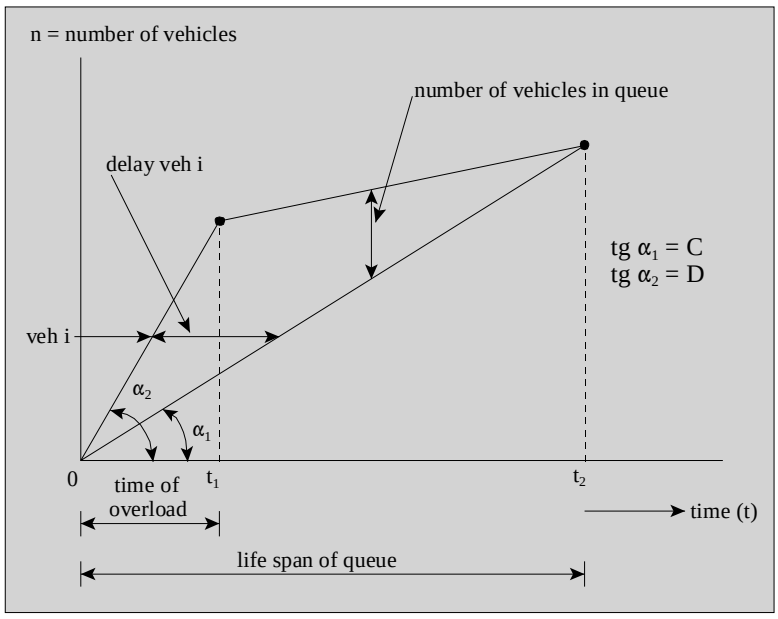

Figure 7.3: Modelling with vertical queues. C denotes the capacity of the bottleneck and D denotes the traffic demand during 'overload'.

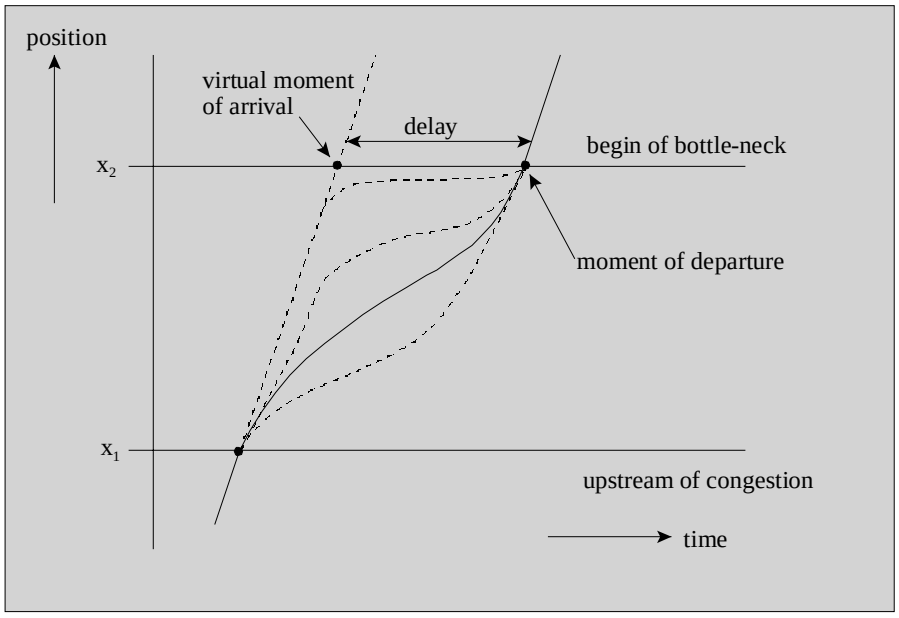

Figure 7.4: Possible vehicle trajectories in congestion.

The relatively simple queueing model has been used in a Dutch approach to take account of congestion and its fluctuation over a year in deciding how much capacity is needed. It will be discussed in the remaining part of this chapter. The main drawback of the approach is that the queue is modelled as a vertical queue, i.e. the amount of space taken up by the queue is not considered explicitly.

7.2.1 Computations with the queuing model

With the deterministic queuing model, we can derive several relations that provide insight into the characteristics of queuing and congestions.

Situation: from time $t = 0$ the capacity is exceeded. This oversaturation lasts until $t = t_1$, after which the demand is reduced to some value below C. This means: between 0 and t_1, the queue increases; after t_1 the queue reduces again. Let $N_D(t)$ denote the cumulative arrival curve; q_1 denotes the flow rate between 0 and t_1; q_2 denotes the flow rate after t_1. C denotes the capacity. We aim to determine: the time t_2 at which the queue has dissappeared; the duration T of congestion; the number of vehicles M that experienced congestion; the collective delay R, the mean delay R_{mean} and the maximum individual delay $R_{\max}$.

Computational steps:

1. Auxilary point (cumulative vehicle count at time t_1): $N(t_1) = t_1 D_1 = t_1 \tan\alpha_1 = t_1 q_1$
2. Maximum queue length $Max = N(t_1) - Ct_1$
3. Maximum delay (at time t_1) via

$$\tan\alpha_1 = \frac{Max}{R_{\max}} \qquad R_{\max} = \frac{Max}{C} = \frac{t_1(q_1 - C)}{C} \tag{7.4}$$

4. Duration of congestion T by intersection of the cumulative demand curve

$$N_D(t) = \begin{cases} q_1 t & 0 < t < t_1 \\ q_1 t_1 + q_2(t - t_1) & t > t_1 \end{cases} \tag{7.5}$$

and supply curve

$$N_C(t) = Ct \tag{7.6}$$

yields

$$q_1 t_1 + q_2(T - t_1) = CT \rightarrow T = \frac{(q_1 - q_2)\, t_1}{C - q_2} \tag{7.7}$$

5. Total number of vehicles M that has been in the queue

$$M = CT \tag{7.8}$$

6. Total collective loss R

$$R = \frac{Max}{2} T \tag{7.9}$$

or

$$R = \frac{M}{2} R_{\max} \tag{7.10}$$

7. Mean time loss

$$R_{mean} = \frac{R}{M} \tag{7.11}$$

or

$$R_{mean} = \frac{1}{2} R_{\max} \tag{7.12}$$

Note that R_{mean} does not depend on q_2. Futhermore, it turns out that the collective loss R increases quadratically with the duration of the oversaturation, but the individual loss R_{mean} increases linearly. When q_2 is only slightly smaller than C then

- the duration of the queue will be high
- there will be many vehicles that experience travel time delays
- the collective waiting time R will be large
- the mean delays will not increase

Therefore, it can be concluded that when after congestion in the peak hour, the demand is only slightly smaller than the bottle-neck capacity, the congestion may last for a significant period of time.

7.3 QUAST-method to take account of variability of congestion

Congestion has existed as long as there has been traffic and transportation. Congestion occurs when the demand for the use of a system is larger than its capacity. Delay, often unexpected and unpredictable in size, is a characteristic of congestion. Because of this characteristic, congestion leads to uncertainties in the planning of activities and transportation, the use of sometimes unnecessary time margins, or to unexpected delay. Congestion on motorways is a 'relatively new phenomenon'. During the time that the main part of the motorway system was built, the missing links in the motorway network formed the bottle-necks of the national network. Even then there were days when a great deal of congestion took place on the motorways, for example on Easter Monday and Whitmonday. It was soon decided that the capacity of the network did not need to be so high as to be able to facilitate such exceptional demand patterns.

Nowadays congestion seems to attract a lot of attention. People fear grid-lock; in more objective terms they think that the accessibility of important centres (sea ports, airports, business areas in cities, etc) will degrade too much.

One of the decisions that has to be made in the design of a road is the level of service that should be offered. A commonly used rule is that a road should be able to facilitate the predicted design traffic volumes in a design year, with a certain minimum Level of Service. Since around 1970 a cost-benefit analysis in the Netherlands led to the rule: offer a Level of Service of C on inter-regional roads. This comes down to a capacity of about 4/3 of the design demand and will provide a situation in which travelling speeds are high, and freedom to manoeuver and driving comfort are also at an acceptable level.

After some time it became clear that building new roads did not (by far) keep up with the growth of the demand and congestion on motorways has become a daily returning phenomenon. This resulted in a study by working group QUAST (derived from QUAlity and STructure of the main road network) from Rijkswaterstaat (Part of Ministry of Transport) about what level of capacity should be offered . It was found that a certain amount of congestion led to an optimum in terms of costs and benefits (elements: costs of building roads, delay, accidents, and road-maintenance); see [52].

In this study the 'probability to experience congestion' was chosen as criterion for the quality of traffic operation. This is the fraction of daily road users that experiences congestion on a link. This fraction is determined as an average over all working days of a year. New in this approach is that a year is no longer represented by a single representative value. Instead, fluctuations of all kinds of factors during a year have influence; see [50].

The economic optimum appeared to be at the point of 2% probability of congestion. Later, this standard was relaxed. Whereas the 2% probability of congestion is only valid for hinterland-connections, for the rest of the main road network a 5% probability of congestion is accepted. In

addition, these are national normative values. In fact one should make a cost-benefit analysis for every section and corridor. In areas where the building costs of roads are high, the acceptable level of congestion could be higher.

7.3.1 QUAST-model

Needed is a calculation of the characteristics of congestion as a function of demand and capacity. This has been carried out using a deterministic queueing model with vertical queues.

The task of this model is to calculate:

- the fraction of vehicles that experience congestion (P_c);
- the extra travel time or delay for vehicles that experience congestion, collectively (R_t) as well as the mean value per vehicle (R_{mean});

based on the demand pattern over a year, $D(t)$, and a capacity pattern, $C(t)$.

The queueing model with vertical queue is especially suitable to fulfill this task. It is applied to those network parts that are bottle-necks, i.e. road elements that determine the capacity of a section.

The model confronts the demand pattern and the capacity pattern with each other and calculates the resulting congestion; see Fig. 7.5. The demand pattern is represented by the fluctuations of demand over the 24 hour period of the working days of a year. Over the hours of a day the demand varies; there is a morning peak period and an evening peak period. The demand $D(t)$ also systematically varies with the day of the week (factor F_{day}) and the month of the year (factor F_{month}).

Finally there are random fluctuations $\varepsilon(t)$); given a time of the day, on a given day of the week, in a given month of the year, there is a mean value for the demand intensity. A random variable $\varepsilon_D(t)$ is added to this mean value; it has a mean value of 0 and a normal distribution with a given standard deviation. This term will further be referred to as noise.

$$D(t) = D_0(t) F_{month} F_{day} + \varepsilon_D(t) \tag{7.13}$$

Comparison of the demand patterns generated with formula (7.13), with the results of [3], showed that the demand pattern was fluctuating too wildly. This problem was solved by adding noise with interdependency (correlated noise). It is plausible that if the demand during period i is higher than the average value, it is likely (probability > 0.5) that it is also higher than average in the next period $i+1$. This can be achieved by smoothing independent random variables ε_0 with the model (α = smoothing factor; $0 \leq \alpha \leq 1$):

$$\varepsilon_D(t) = \alpha \varepsilon_D(t-1) + (1-\alpha)\varepsilon_0(t) \tag{7.14}$$

At first glance one might think that the capacity pattern is a constant. However, this is not the case, certainly not under changing weather (factor $F_{weather}$) and light conditions (factor F_{light}). It is known that with rain road capacity reduces by about 10% and in darkness capacity reduces by about 5%. Rain is distributed randomly in time, whereas darkness is a systematic factor.

Similar to the demand, the capacity also has a non-systematic or random component. Under ideal and sustaining circumstances the capacity still varies. This can be explained by the different characteristics of vehicle-driver elements. However, these characteristics can not be observed and their influences are not known, but it is clear that they affect capacity; see [56].

$$C(t) = C_0 f_{weather} f_{light} + \varepsilon_C(t) \tag{7.15}$$

The random component $\varepsilon_C(t)$ has a mean value of 0 and a normal distribution with a standard deviation equal to 6 % of capacity. In contrast to the noise term of the demand, the noise of the capacity is not correlated.

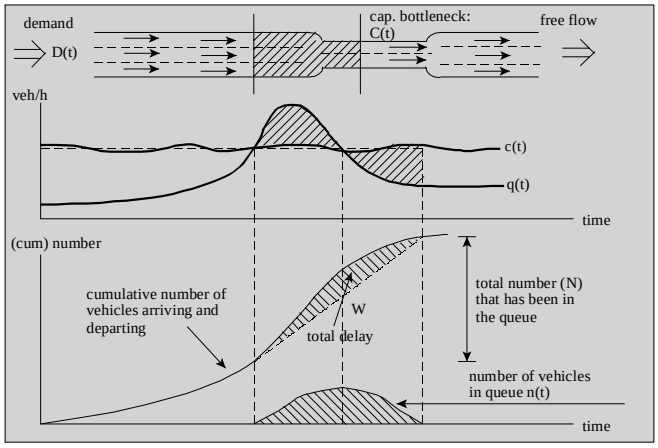

Figure 7.5: Queueing model with vertical queue

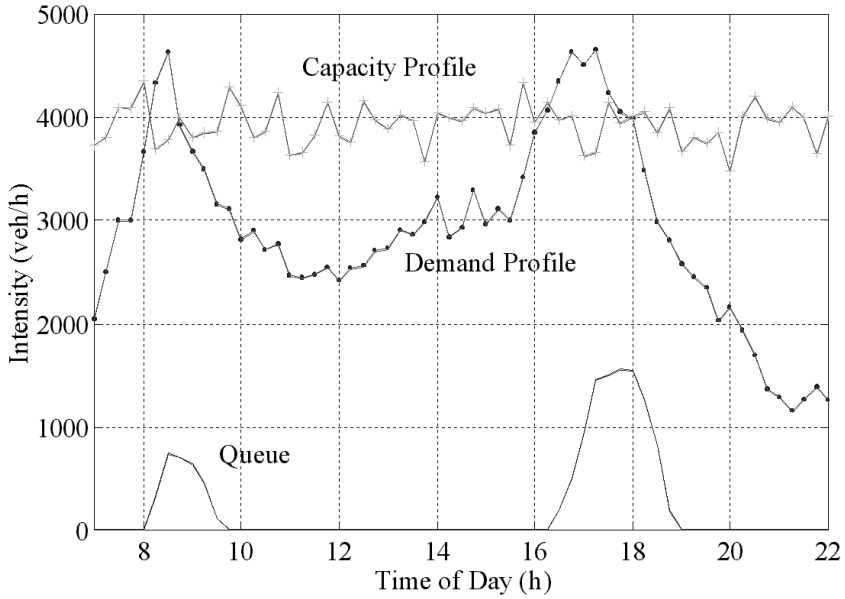

Figure 7.6: Results of a simulation for a day.

Remark 53 *The model (7.15) has not been used literally but in a simplified way. Research in the scope of the QUAST study has led to the following rule of thumb: 'bad weather (including the effect of darkness) occurs during 10% of the peak hours and brings about a reduction of capacity of 12%'. This rule has been interpreted even more deterministically in this model. On the first Tuesday and Wednesday of every month of 20 days bad weather is assumed. Tuesday and Wednesday were chosen because the intensities on these days differ the least from the average daily value. This procedure achieves the fraction of 10% of time bad weather per year exactly. In practice the effect of bad weather and darkness will lead to more varying outcomes than this model produces.*

Fig. 7.6 shows the results of a simulation over a day. The capacity and demand are depicted as a function of the time of a day. There are short periods with demand higher than capacity, resulting in a queue indicated at the bottom line. Note that the demand and capacity are intensities and the queue is a number of vehicles; consequently the scale of the queue in Fig. 7.6 is arbitrary.

Because the pattern of the demand and capacity is now variable, the calculations have to

Factor	Cap. increase (%)
Noise on capacity	2.9
Noise on demand	3.7
Month factor	1.9
Day factor	1.1
Bad weather	2.6
All factors	9.9 (sum = 12.2)

Table 7.1: Effect of different factors on roadway capacity

be carried out per time step (15 minutes), using the equation:

$$n\left(t\right) = \int_{t_1}^{t} \left(D\left(s\right) - C\left(s\right)\right) ds \tag{7.16}$$

The number of vehicles in the queue n is the integral of the over saturation $D(t) - C(t)$ and the integration starts at moment t_1 as $D(t)$ for the first time becomes larger than $C(t)$.

The collective delay R_t is the integral of $n(t)$ over time.

$$R_t\left(t\right) = \int_{t_1}^{t} n\left(s\right) ds \tag{7.17}$$

7.4 QUAST application results

The remainder of this chapter presents some results of application of the QUAST method.

7.4.1 Effect of factors on required capacity

Investigations with the model have begun with a strictly deterministic case: a fixed demand pattern and a constant capacity, chosen at a level such that P_c equals 5%. The 'basic capacity' needed to obtain this P_c value, given the demand pattern, is denoted as C_0. In next steps other sources of variation are added to the model. It turns out that the addition of every factor leads to a higher P_c. Increasing the capacity by a certain amount, P_c can be reduced to its original value of 5 %. It is this increase, expressed as a multiplier, that has been chosen as a yard stick for the influence of the factor. Table 7.1 presents an overview of the effects of the factors investigated. It appears that the sum of the separate effects is larger than the combined effect of all factors together. Further, there are no factors significantly more or less important than others. Consequently, it can not be concluded that one or several factors can be neglected.

The calculated delay in the most extended model at the normative $P_c = 5\%$, is clearly less than the values given in publications on outcomes of QUAST (2 min against 4-5 min). Probably the explanation for this difference is that in the QUAST method of Rijkswaterstaat, the value found for R_{mean} is multiplied by a factor of 2 in order to account for other factors that lead to delay.

7.4.2 Distribution of congestion characteristics

Besides the mean values of P_c, R_{mean} and R_t, their distributions over the days of a year are also of importance. Especially the travel time reliability of the system of main roads is closely related to them. Fig. 7.7 shows the (cumulative) distributions functions where all factors are included in the model and with an overall value for P_c of 5%.

It appears that all three distributions have the same character. There are many zero or nearly zero values and the tail is long. This means high values appear with a low frequency. These results are not convenient, because it means that congestion is hard to forecast. In other

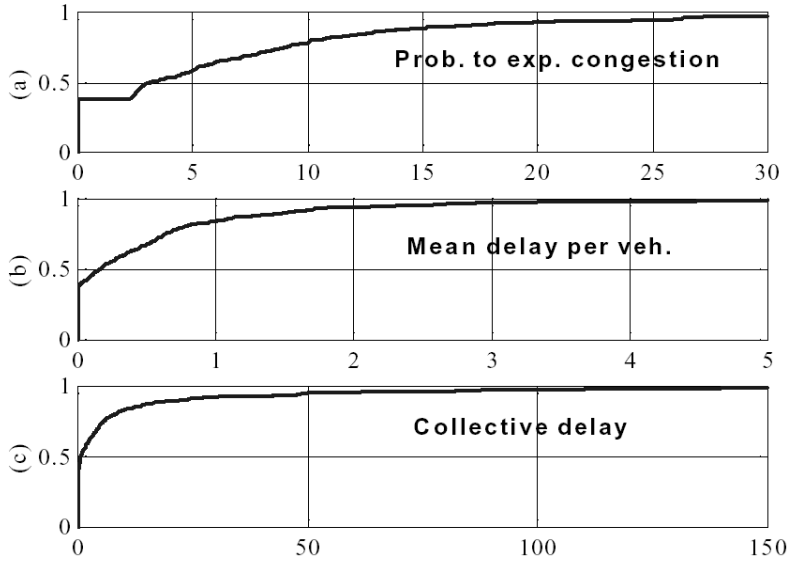

Figure 7.7: Statistical distribution functions over working days of a) P_c (probability of experiencing congestion); b) R_t (collective delay) and c) R_{mean} (average delay per vehicle that experiences congestion).

words its unpredictability is relatively high. P_c and R_{mean} both show about 30% zero-values. R_t even shows 50% zero-values. This means that the collective delay is nearly zero on 50% of the days.

The scale on the axis of R_t requires some explanation. This scale is in minutes per percent of $AADT$; this means that the value should be multiplied by $(1/60)AADT/100$ in order to find vehicle-hours.

Example 54 *$AADT = 50.000$ vehicles. The point on the scale at 100 corresponds to a value of 833 vehicle-hour. If the delay is valued 10 Euro per hour, then it represents an amount of Euro 8333.*

Lastly, the mutual dependence of the characteristics P_c, R_t and R_{mean} is important. The interdependence between P_c and Rmean (see Fig. 7.6a) is nearly linear, but the scatter is large. The dependence between the other two (see Fig. 7.6b) is clearly not linear. R_t is small for $P_c <$ 10% and beyond this point grows more rapidly than linear with P_c. The dependence between R_{mean} and R_t can easily be concluded from the other two dependencies.

7.5 Final remarks

Some remarks are due about the model application:

- Calculations were made with 15-minute time-steps, whereas the Ministry of Transport used 1-hour time steps.
- For the sake of simplicity calculations were made with 12 months of 20 working-days.
- Simulations were made for the time period between 7.00h and 22.00h. Congestion will not occur outside this time period in this model. The nighttime intensity has been taken into account in the 24-hour value.
- A simulation over a period of one year (240 working days) is usually not enough to get stable outcomes. One should average over a few years.

The analysis of the findings obtained with the model is far from complete. Some issues can be brought forward:

- Most analysis has focussed on the probability to experience congestion; more attention should be given to the collective and individual delay.
- The deterministic demand-pattern over the time of a day has not been varied; it is a pattern with two peak periods that differ slightly. Research should be done into the effects of this choice.
- The sensitivity of the results for change in the parameters should be determined.
- As far as the model is concerned, the factors for darkness and rain might be introduced separately.

Finally it should be born in mind that capacity in fact can only be attuned to the demand to a very limited extent. In most cases a road section capacity can only be adjusted by varying the number of lanes. The main goal of the preceding analysis has been to gain more insight into the several factors that contribute to congestion on a yearly basis and to illustrate the fluctuations in congestion. Apart from over saturation there are other factors that can cause congestion, e.g. exceptional weather conditions (dense fog, storms), road works, incidents and accidents.

Drivers can react to congestion experienced in the past by changing their travel behaviour. On a short term the change of departure time is the most likely reaction. This could lead to even more fluctuations in congestion, if the reaction is badly timed, e.g. drivers react to the congestion they experienced yesterday. In fact the process of how drivers respond to congestion is not yet very well understood.

Chapter 8

Shock wave analysis

Summary of the chapter. Flow-speed-density states change over time and space. When these changes of state occur, a boundary is established that demarks the time-space domain of one flow state from another. In rough terms, this boundary is referred to as a shock wave. In some situations, the shock wave can be rather smooth, for instance when a platoon of high-speed vehicles catches up with a sligthly slower moving vehicle. In other situations, the shock wave can be a very significant change in flow states, for instance when high-speed vehicles approach a queue of stopped vehicles.

This chapter presents shock wave analysis. We introduce the concept of kinematic waves and shock waves, and present a number of examples of application of shock wave analysis.

List of symbols

k	veh/m	traffic density
q	veh/s	traffic volume
$Q(k)$	veh/s	equilibrium flow as function of density
v, u	m/s	traffic speed
ω	m/s	shock wave speed
c	m/s	kinematic wave speed
q_c	veh/s	capacity

8.1 Kinematic waves

A useful tool to describe dynamic phenomena in traffic flow is the *conservation of vehicle equation.* This equations reflects the fact that no vehicles are created or lost, i.e. that vehicles are conserved. If for a specific roadway sections, more vehicles are entering than leaving, then the number of vehicles in the section must increase: i.e. the density increases. Let $q = q(x,t)$ and $k = k(x,t)$ denote the traffic volume and density at instant t and location x. We assume that q and k are both smooth functions in time and space, and are defined from the smooth cumulative vehicle count $\tilde{N}(x,t)$. Then, the conservation of vehicle equation equals

$$\frac{\partial k(x,t)}{\partial t} + \frac{\partial q(x,t)}{\partial x} = 0 \tag{8.1}$$

We combine equation (8.1) with the expression of the fundamental diagram $q = Q(k)$, yielding the following partial derivative in the traffic density (see chapter 2)

$$\frac{\partial k(x,t)}{\partial t} + \frac{dQ}{dk}\frac{\partial k(x,t)}{\partial x} = 0 \tag{8.2}$$

Using this partial differential equation, we can describe the dynamics of the traffic flow under the assumption that the speed of the traffic reacts *instantaneously and locally to the traffic*

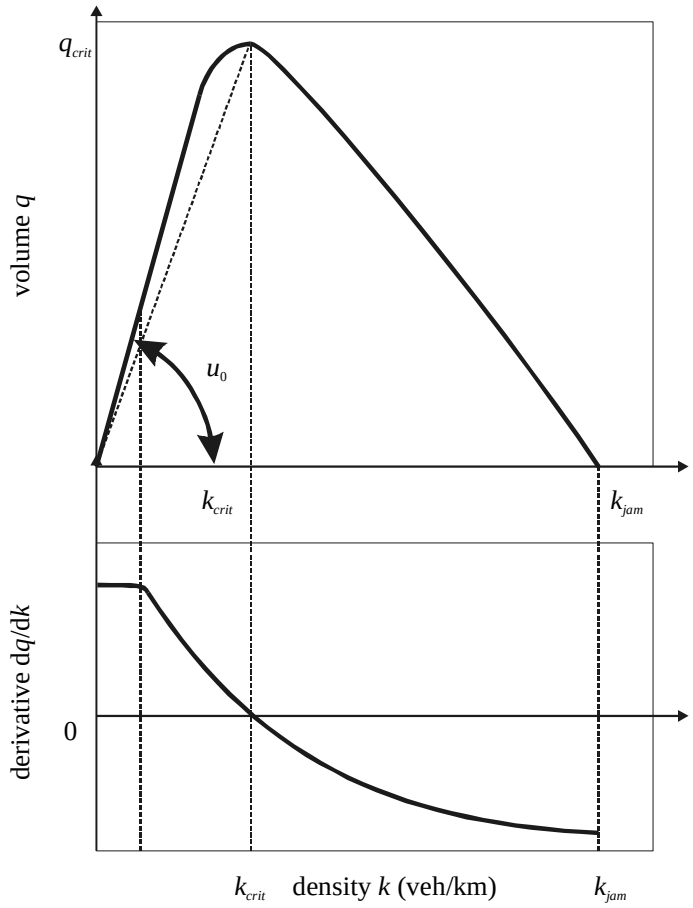

Figure 8.1: Fundamental diagram $Q(k)$ and speed of the kinematic wave $c = dQ/dk$ as a function of the density k

density, that is $u(x,t) = V(k(x,t))$ or $q(x,t) = Q(k(x,t))$[1]. The resulting model (conservation of vehicle equation and fundamental relation) constitutes one of the oldest macroscopic traffic flow models (kinematic wave model, LWR model), and was first studied by [34]. We can show that the partial differential equation (8.2) has the following general solution

$$k(x,t) = F(x - ct, t) \quad \text{where} \quad c = \frac{dQ}{dk} \text{ and } F \text{ an arbitrary function} \tag{8.3}$$

This means that a certain density k_0 propagates with speed $c = \frac{dQ}{dk}(k_0)$. Roughly speaking, c is the speed with which very small disturbances are propagating; these are called kinematic waves.

In chapter 4 several examples of fundamental diagrams have been discussed. Some of these, like Smulder's function or Daganzo's function, $Q(k) = u_0 k$ for small densities (see also Fig. 8.1). Then $\frac{dQ}{dk} = u_0$ for small densities, i.e. the speed of the traffic flow and the speed of the kinematic wave are equal. This is why this density region is referred to as the *stable area*. For larger densities, the difference between the speed of the traffic u and the kinematic wave speed c will increase. This region is referred to as the *unstable area*, since small disturbances in the density will grow into discontinuities. In the congested area, the kinematic wave speed c becomes negative implying that small disturbances in the flow will propagate upstream.

[1] Note that in chapter 5, it was shown how this may not be a realistic assumption.

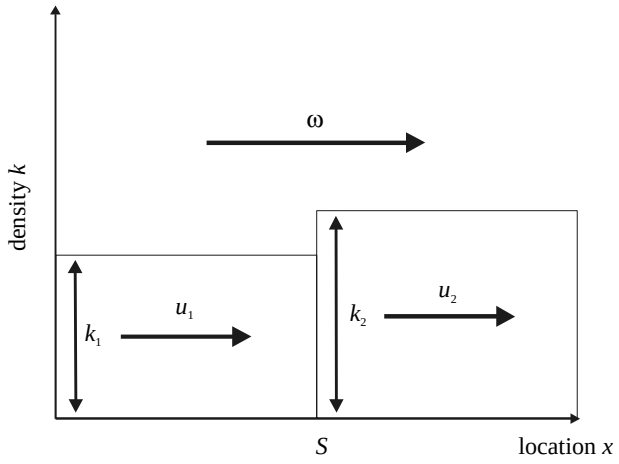

Figure 8.2: Derivation of shock wave speed

8.2 Shock wave speeds

The waves that occur in the traffic flow are categorised as either *kinematic waves* of *shock waves*. Shock waves are waves that originate from a sudden, substantial change in the state of the traffic flow. That is, a shock wave is defined by a discontinuity in the flow-density conditions in the time-space domain.

1. Fig. 8.2 schematises a shock wave. The shock is the boundary between two areas in for which different traffic conditions hold.
2. The state at the left area is (k_1, u_1); the state in the right area is (k_2, u_2). The line S indicates the shock wave with speed ω. The relative speeds with respect to line S for the left and the right area are respectively equal to $u_1 - \omega$ and $u_2 - \omega$.

If we consider the area left of line S, then the volume passing S equals

$$q_1 = k_1 (u_1 - \omega) \tag{8.4}$$

This holds equally for the right area, i.e.

$$q_2 = k_2 (u_2 - \omega) \tag{8.5}$$

Since the flow into the shock must equal the flow out of the shock, we can easily show that the wave speed ω equals

$$\omega = \frac{q_1 - q_2}{k_1 - k_2} \tag{8.6}$$

Note that for small changes in q and k, equation (8.6) becomes $\omega = \frac{dQ}{dk} = c$. In other words, the kinematic wave is the limiting case of a shock wave. Fig. 8.3 shows the following

1. Traffic flow in state (q_1, k_1) with speed $u_1 = \tan \phi_1$.
2. Traffic flow in state (q_2, k_2) with speed $u_2 = \tan \phi_2$.
3. The shock wave which occurs when traffic in state (q_1, k_1) transients into state (q_2, k_2). Between both states a shock wave occurs with speed $\omega = \tan \theta_1 = \frac{q_1 - q_2}{k_1 - k_2}$. Since $q_1 - q_2$ is positive and $k_1 - k_2$ is negative, the shock wave will move upstream.
4. The state (q_1, k_1) itself propagates with the speed of the accompanying kinematic wave, i.e. the tangent of $Q(k)$ at $k = k_1$; this holds equally for the state (q_2, k_2).

The following sections discuss several examples of applications of shock wave analysis, both to different situations, and using different forms of the fundamental diagram.

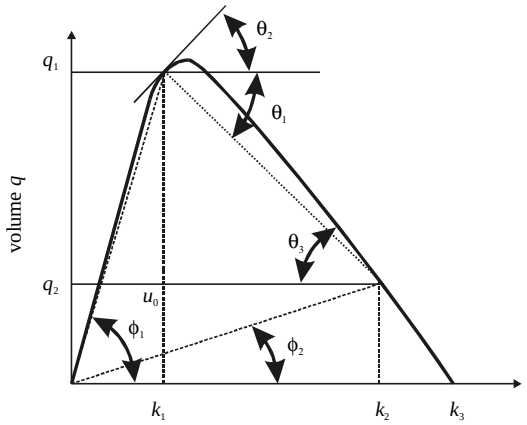

Figure 8.3: Shock waves in the $q - k$ plane

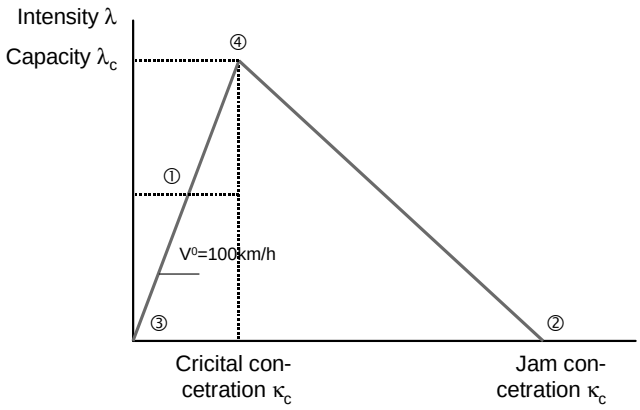

Figure 8.4: Schematised intensity - density (concentration) relation

8.3 Temporary blockade of a roadway

Blockades occur in practice in a number of circumstances, for instance when a bridge is open, during a severe incident blocking all the roadway lanes, or during the red-phase of a controlled intersection. Consider a two-lane road where the volume is equal to 2500 veh/h. The capacity q_c of the road equals 5000 veh/h. The jam-density k_j equals 250 veh/km. The blockade starts at $t = t_0$, and ends at $t = t_1$. At the time the blockade occurs, three shock waves occur (see Fig. 8.5):

1. The transition for freely flowing traffic (state (1)) to standing-still (state (2)). The speed of the shock wave equals $\omega = \frac{q_1}{q_1/u_1 - k_j} = \frac{2500}{25 - 250} = -11.1$. The wave is called the *stop wave*.

2. The head of the queue (state (2)) borders on an empty region (state (3)). Formally, this is a shock wave as well; clearly $\omega = 0 km/h$.

3. Downstream of x_0 a region in which traffic is present that has just passed x_0 before the blockade occurred. This is also a shock wave with speed $w = u_0$.

   When the blockade is removed at $t = t_1$, two more shock waves emerge

4. The transition from queuing (state (2)) to the capacity state (4) (the so-called start-wave) with speed $\omega = \frac{q_4 - q_2}{k_4 - k_2} = \frac{q_c}{k_c - k_j} = -25 km/h$, implying that the head of the queue propagates upstream with a speed of $25 km/h$.

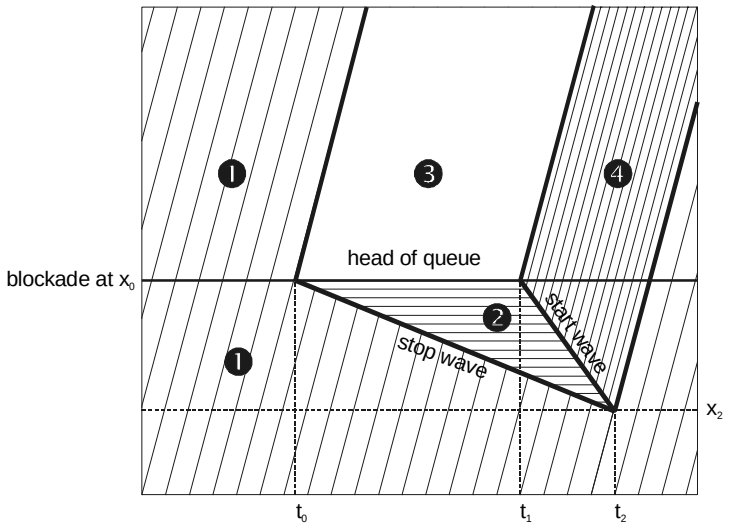

Figure 8.5: Trajectories of shock waves and vehicles. Note that the stop wave is often referred to as the backward forming shockwave; the start wave is also referred to as the backward recovery wave; furthermore the head of the queue is also called the frontal stationary wave

5. Downstream of x_0 exists a boundary between state (4) and the empty road. This shock wave has speed $\omega = u_0$.

Since the start wave has a higher speed that the stop wave, it will eventually catch up, say at time $t = t_2$. At that time, a new shock wave occurs, reflecting the transition from free-flow traffic (state (1)) to the capacity state (4). This wave has speed u_0.

8.4 Shock wave analysis for non-linear intensity-density curves

In the previous section, shock wave analysis was applied to determine traffic flow conditions for piecewise linear intensity-density curves. This section discusses application of shock wave analysis for general $Q(k)$ relations, i.e. relations that are not necessarily piecewise linear. This is done by revisiting the application examples discussed in the previous section.

8.4.1 Temporary blockade revisited

Let us first reconsider the temporary roadway blockade example, using the $Q(k)$ relation depicted in Fig. 8.6. Here, the derivative $Q'(k)$ is a monotonic decreasing function of the density k, with $Q'(k) < 0$. As before, the blockade of the roadway starts at $t = t_0$ and ends at $t = t_1$. At the time the blockade occurs, the shock wave theory will again predict the occurance of three shock waves (see Fig. 8.6), similar to the case of piecewise linear $Q(k)$. In line with [36], the shock wave theory predicts that at time $t = t_1$, traffic will start moving along the roadway, fully using the available capacity q_c. This causes a transition from state 3 (empty roadway) to state 4. Note that the speed at which shock between states 3 and 4 moves equals the capacity speed u_c, which is in this case less than the free flow speed u_0. The start wave and the stop wave propagate in a similar fashion as before. However, at the point where the start wave and the stop wave intersection (i.e. the queue has dissolved), the shock wave separating states 4 and 1 has a different character: its speed is less than the speed of the vehicles in state 1. As a result, the vehicles in state 1 will eventually catch up with the shock wave and driving into the state 4 region.

Clearly, this result is not very realistic: the shock wave theory predicts that the speed of the first vehicle in the queue equals the speed at capacity u_c, while in fact the first vehicle will

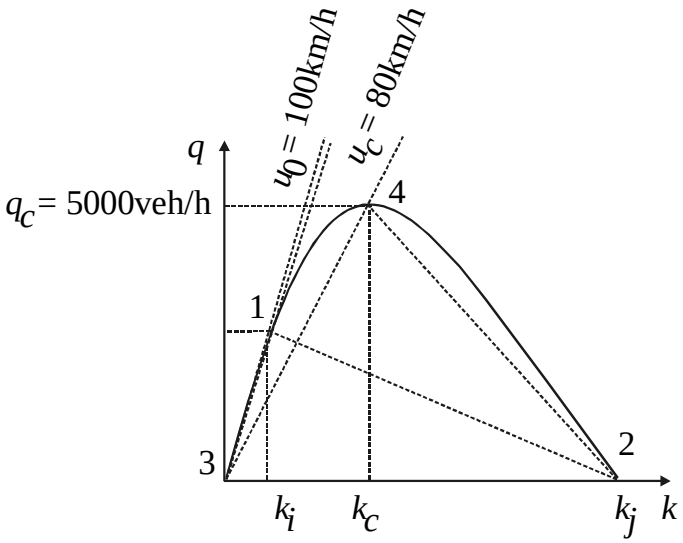

Figure 8.6: Intensity - density relation.

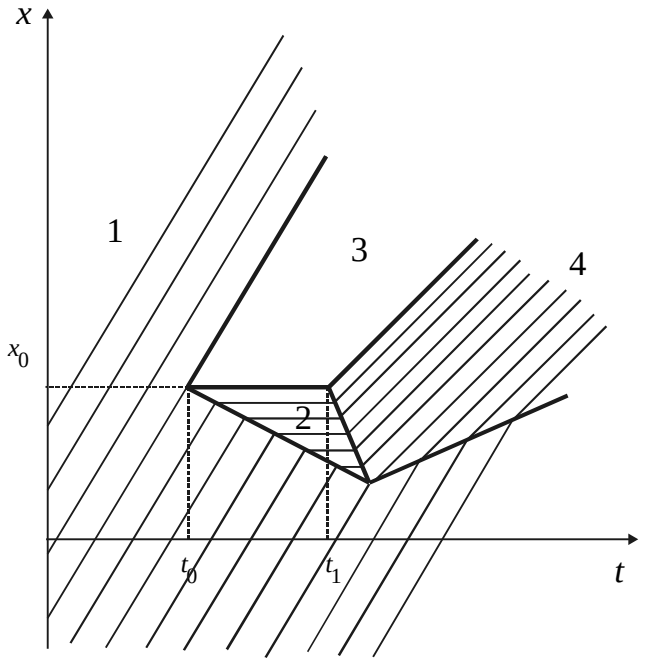

Figure 8.7: Shockwaves for non-linear $Q(k)$ for temporary blockade example

accelerate towards the free speed u_0. This is caused by the fact that shock wave theory does not describe the acceleration - nor the decelaration - of the first vehicle in the queue. As a result of the reduced speed of the first vehicle, the shock that is caused by the temporary blockade will not dissolve over time. Rather, state 4 will disperse further as time goes by.

8.5 Moving bottlenecks

In this section we consider a more complicated example, namely a moving bottle-neck (b-n) that is present on a road over a given distance and/or during a given period. The usefulness of this example is that it implies other relevant cases by assigning values zero to the speed and/or the capacity of the bottle-neck. Practical examples of a moving bottle-neck are:

1. A slow moving vehicle, e.g. an agricultural tractor with a speed of 20 km/h, on a two-lane road. The capacity of the b-n is determined by the overtaking opportunities, hence by the opposing flow and the overtaking sight distance. Those two factors can lead to a more or less constant capacity of the b-n.

2. A platoon of trucks on a long grade on a motorway. In such conditions trucks form a slow platoon and more or less block the right hand lane, causing a substantial capacity reduction.

3. Actions of protesting farmers or truck drivers that from a temporary slow platoon on one or more lanes of a motorway.

To simplify the case three assumptions will be made:

1. Demand intensity is constant and larger than the capacity of the b-n. If the demand is lower, the b-n only leads to a speed reduction with relatively little delay.

2. Daganzo's fundamental diagram is used.

3. The length of the b-n can be neglected; it is set at zero.

8.5.1 Approach 1

Figure 8.8 depicts the phenomena in the $q-k$ plane and the $x-t$ plane. The capacity of the roadway is 4500 veh/h; speed at capacity $u_c = 90km/h$; jam density $k_j = 250veh/km$. The moving b-n has a speed $\hat{v}$ (20 km/h) and covers a distance of 4 km. The b-n has a capacity of 1800 veh/h, a capacity speed of 60 km/h and $k_j = 125veh/km$.

In general terms the b-n will lead to *congestion upstream and free flow downstream with an intensity equal to the capacity of the b-n.* After the b-n is removed the congestion will decay until free flow is restored over the total affected road section. We start looking to the phenomena in the $q-k$ plane. It is obvious that in the b-n the (traffic flow) state is the cap. state of the b-n; state **2**. At the upstream end of the b-n a transition from a congested state **4** to the state **3** must occur and move with speed $\hat{v}$. At the downstream end of the b-n a transition from state **4** to free flow **1** occurs and this transition must also have speed $\hat{v}$. This implies that in the $q-k$ plane the *shock waves are represented by the line going through the capacity point of the b-n with a slope equal to* $\hat{v}$. Consequently the state in the congestion upstream the b-n is state **3** and the state downstream the b-n is state **4**.

Note: Consequently the intensity of the free flow downstream the b-n is not equal to the capacity of the b-n, as was loosely stated earlier, but it is less. Only if the speed of the b-n is zero that is the case.

Free flow state **1** and congested state **3** determine the speed of the shock wave at the tail of the queue. Given this description in the $q-k$ plane the geometrical description in the $x-t$ plane follows easily and the calculations are straightforward.

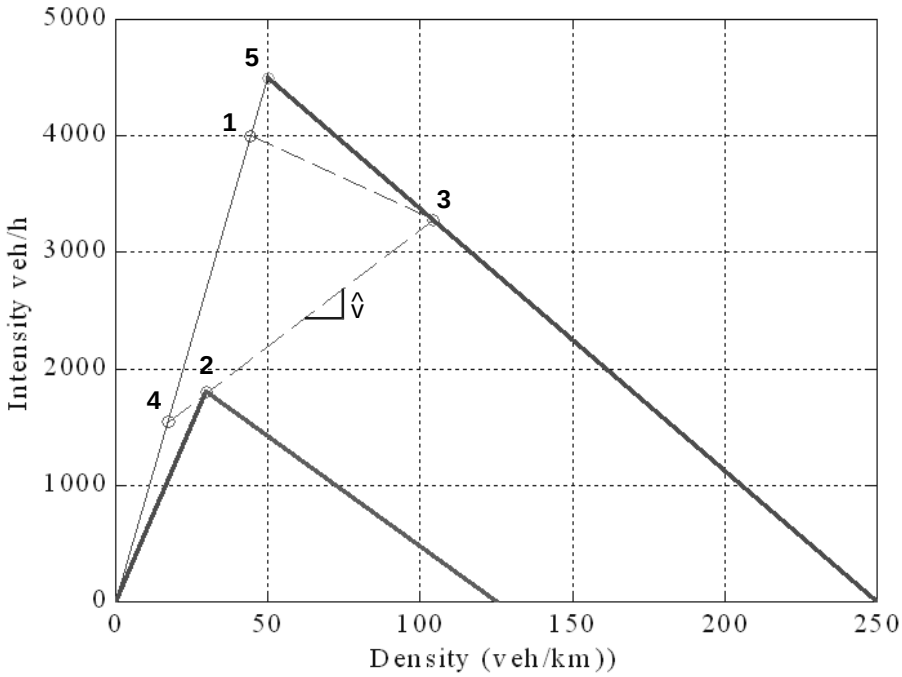

Figure 8.8: Effects of moving bottleneck in the flow-density plane.

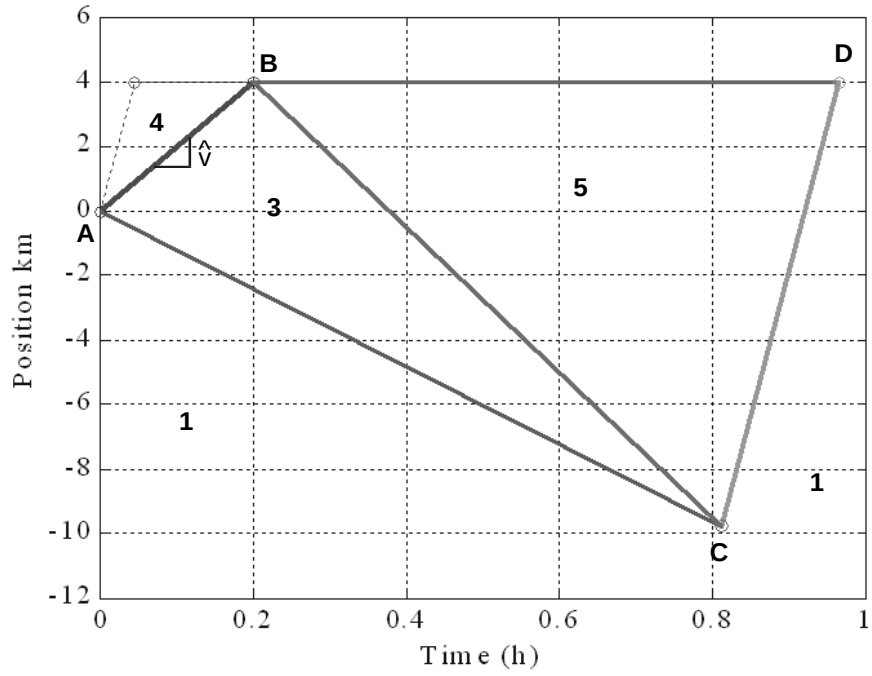

Figure 8.9: Results of moving bottleneck in $x - t$ plane.

The origin in Fig. 8.9 is where the b-n starts. It moves with speed $\hat{v}$ until it leaves the road again. The upstream border of the queue is a shock wave with speed ω_{13}. At a given moment and position the moving b-n leaves the road (point **B**). The congested traffic, state **3**, then transforms to the capacity state **5** and a shock wave between state **3** and state **5** starts and makes the size of congestion shrink. Shock waves S_{13} and S_{53} meet at point **C**, and then starts a shock wave representing the transition from the undisturbed state **1** to capacity state **5** with speed ω_{AE}. When this shock wave reaches the position of point **3** the disturbance is virtually over. Line **AB** is very special, since it represents state **2**; just upstream of a shockwave with speed ω_{32} is present and just downstream a shockwave with the same speed ω_{24}.

Note: downstream point **B**, Daganzo's diagram works out less realistically. The three states 4, 5, and 1 propagate all with the same speed and do not mix. In reality, $u_4 > u_1 > u_5$ and drivers will accelerate to areas with a lower density.

8.5.2 Approach 2

An alternative approach exist to determine the effects of a moving bottleneck. Let us again consider the situation above, where a moving bottleneck enters the roadway and drives with a constant speed $\hat{v}$.The trick is to reformulate the problem with all relevant variables (location, flows, speeds, etc.) defined relative to a frame of reference moving along with the moving bottleneck. This is usefull because the moving observers will notice that the relative flows passing them also obey some $q-k$ relation at each point (albeit different from the origing $q-k$ relation) *and that the bottleneck appears stationary.* Thus the solution in the new frame of reference can be obtained as discussed earlier without any changes, and from this solution one can then retrieve the desired 'static' solution.

The simple mathematical formalism follows. Using primes to denote the variables used in the moving frame of reference we definde

$$x' = x - \hat{v}t, \quad t' = t, \quad k' = k, \text{ and } q' = q - k\hat{v} \tag{8.7}$$

Note that q' is the relative flow measured by the moving observer, which can be either negative or positive. It should be obvious that the relative flows must be conserved and thus

$$\frac{\partial k'}{\partial t'} + \frac{\partial q'}{\partial x'} = 0 \tag{8.8}$$

Of course, this equation can also be obtained by changing variables in $\frac{\partial k}{\partial t} + \frac{\partial q}{\partial x} = 0$. Most importantly, the original relation between q and k is transformed into a relation $Q'(k', x')$ that is independent of t

$$q' = q - k\hat{v} = Q(k, x - \hat{v}t) - k\hat{v} = Q'\left(k', x'\right) \tag{8.9}$$

Thus, the problem has been reduced to the stationary bottleneck case with Eq. (8.8). This new relation essentially lowers the traffic speed and the wave speed for every k by and amount of $\hat{v}$. To retrieve the solution in the original frame of reference, we simply need to adjusr vertically the edges of the polygons in the (x', t') solution using $x = x' + \hat{v}t'$ while keeping the same density inside of each polygon (since $k = k'$).

8.6 Shock wave classification

This chapter has presented shock wave analysis. By means of several examples and exercises we have seen different types of shock waves. Fig. 8.10 classifies the various types of shock waves. The high density area is shown in the centre of the figure, while low densities are located on the outside of the individual shock waves.

A *frontal stationary shock wave* must always be present at a bottle-neck location and indicates the location where traffic demand exceeds capacity. This may be due to recurrent situations where each workday the normal demands exceed the normal capacities, or be due to non

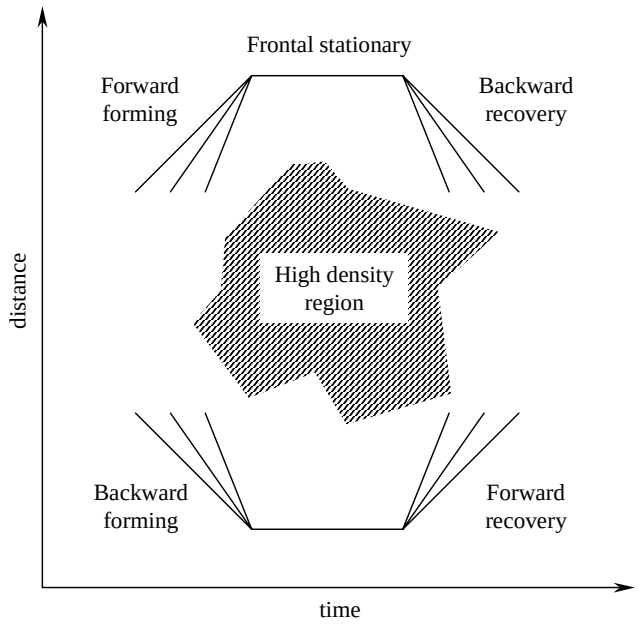

Figure 8.10: Classification of shock waves

recurrent situations where the normal demand exceeds reduced capacity (caused by an acci-dent or an incident) which may occur at any location and at any time. The term 'frontal' im-plies that it is at the front (or downstream edge) of the congested region with lower densities farther downstream and higher densities upstream. The term 'stationary' means that the shock wave is fixed by location.

Backward forming shock waves must always be present if congestion occurs and indicates the area in the time-space domain where excess demands are stored. The term 'backward' implies that over time, the shock wave is moving backward or upstream in the opposite direc-tion of traffic. The term 'forming' implies that over time, the congestion is (gradually) ex-tending to sections farther downstream. The time-space domain to the left of this shock wave has lower densities, and to the right the density levels are higher. There are examples of backward forming shock waves in the examples and exercises presented in the previous sec-tions.

The *forward recovery shock wave* is the next most commonly encountered type of shock-wave and occurs when there has been congestion, but demands are decreasing below the bottle-neck capacity and the length of the congestion is being reduced. The term 'forward' means that over time the shock wave is moving forward or downstream in the same direction as the traffic. The term 'recovery' implies that over time, free-flow conditions are gradually occurring on sections farther downstream.

The *rear stationary shock wave* may be encountered when the arriving traffic demand is equal to the flow in the congested region for some period of time. The *backward recovery shock wave* is encountered when congestion has occurred but then due to increased bottle-neck capacity the discharge rate exceeds the flow rate within the congested region. The last type of shock wave is the *forward forming shock wave*, which is not very common. It may occur when the capacity decreases slowly in the upstream direction, for instance in case trucks are slowed down due to the grade of the roadway.

Chapter 9

Macroscopic traffic flow models

Summary of chapter. The previous chapters of these notes discussed macroscopic traffic flow characteristics. It was show how the main variables density, flow and speed relate via the continuity equation and the conservation of vehicle equation. It was also discussed how these relations could be used to determine spatial-temporal dynamics of the traffic flow using shock wave analysis.

In this chapter, macroscopic traffic flow models are discussed. The aim of these models is to describe the dynamics of the traffic flow. Several macroscopic (or continuum) traffic flow models have proven to be simple enough for the real-time simulation of large traffic networks, while being sufficiently complex to realistically describe de principal aggregate traffic flow variables and their dynamics. These models are applicable due to the following reasons:

1. Macroscopic models describe the most important properties of traffic flows, such as the formation and dissipation of queues, shock waves, etc.

2. Macroscopic models are computationally less demanding. Moreover, the computational demand does not increase with increasing traffic densities, i.e. does not depend on the number of vehicles in the network.

3. They enable determination of average travel times, the mean fuel consumption and emissions in relation to traffic flow operations.

4. These models can be written in closed-form and are generally deterministic and less sensitive to small disturbances in the input.

In sum, macroscopic network traffic flow models are suitable for short-term forecasting in the context of network-wide coordinated traffic management. They are applicable in the development of dynamic traffic management and control systems, designed to optimise the traffic system and can be used to estimate and predict average traffic flow operations.

The underlying assumption of a macroscopic traffic flow model is that traffic can be described as a continuum, similar to a fluid or a gas. Macroscopic equations describing traffic flow dynamics are hence very similar to models for continuum media.

At this point, let us emphasise that the macroscopic models described in this section are in fact macroscopic in their representation of traffic. This does not necessarily imply that the traffic processes or the behaviour of the flow is described in aggregrate terms. On the contrary, several macroscopic models (i.e. the gas-kinetic models) consider the interaction between vehicles (or rather, the average number of interactions between the vehicles and the consequent average changes in the flow dynamics) explicitly. Although the representation of traffic is macroscopic, the behavioural rules describe the dynamics of the flow is in fact microscopic.

List of symbols

k	veh/m	traffic density
q	veh/s	traffic volume
$Q(k)$	veh/s	equilibrium flow as function of density
v	m/s	traffic speed
$V(k)$	m/s	equilibrium speed as function of density
ω	m/s	shock wave speed
c	m/s	kinematic wave speed
q_c	veh/s	capacity
r	veh/s	inflow (at on-ramp)
s	veh/s	outflow (at off-ramp)
$\tilde{N}(x,t)$	veh	cumulative vehicle count (smoothed)
$c(k), \lambda_*$	m/s	kinematic wave speed / characteristic speed
C	-	characteristics
$k_{i,j}$	veh/m	average density for $[x_{i-1}, x_i]$ at instant t_j
$q_{i,j}$	veh/s	average flow during $[t_j, t_{j+1}]$ at interface x_i
T_j	s	time step
l_i	m	cell length
D, S	veh/s	local demand, local supply
τ	s	relaxation time
$\lambda_{1,2}$	m/s	characteristic speeds for higer-order model
κ	veh/m	phase-space density
f	-	speed probability density function
V_0	m/s	desired (free) speed
π	-	immediate overtaking probability
σ	-	transition probability
Θ	m^2/s^2	speed variance

9.1 General traffic flow modelling issues

Traffic operations on roadways can be improved by field research and field experiments of real-life traffic flow. However, apart from the scientific problem of reproducing such experiments, the problem of costs and safety play a role of dominant importance as well. Traffic flow and micro-simulation models designed to characterise the behaviour of the complex traffic flow system have become an essential tool in traffic flow analysis and experimentation. Depending on the type of model, the application areas of these tools are very broad, e.g.:

- Evaluation of alternative treatments in (dynamic) traffic management.
- Design and testing of new transportation facilities (e.g. geometric designs).
- Operational flow models serving as a sub-module in other tools (e.g. traffic state estimation, model-based traffic control and optimisation, and dynamic traffic assignment).
- Training of traffic managers.

Research on the subject of traffic flow modelling started some forty years ago, when Lighthill and Whitham [34] presented a macroscopic modelling approach based on the analogy of vehicles in traffic flow with the dynamics of particles in a fluid. Since then, mathematical description of traffic flow has been a lively subject of research and debate for traffic engineers. This has resulted in a broad scope of models describing different aspects of traffic flow operations, such as microscopic and continuum (or macroscopic) models. The latter macroscopic (or rather

continuum models) consider the traffic flow as a continuum, i.e. like a fluid with specific characteristics, and are the topic of this chapter.

The description of observed phenomena in traffic flow is not self-evident. General mathematical models aimed at describing this behaviour using mathematical equations include the following approaches:

1. Purely deductive approaches whereby known accurate physical laws are applied.
2. Purely inductive approaches where available input/output data from real systems are used to fit generic mathematical structures (ARIMA models, polynomial approximations, neu-ral networks).
3. Intermediate approaches, whereby first basic mathematical model-structures are developed first, after which a specific structure is fitted using real data.

Papageorgiou [41] argues that it is unlikely that any microscopic or macroscopic traffic flow theory will reach the descriptive accuracy attained in other domains of science (e.g. Newtonian physics or thermodynamics). The author states that the only accurate physical law in traffic flow theory is the conservation of vehicles equation; all other model structures reflect either counter-intuitive idealisations or coarse approximations of empirical observations. Consequently, the challenge of traffic flow researchers is to look for useful theories of traffic flow that have sufficient descriptive power, where sufficiency depends on the application purpose of their theories.

9.2 Kinematic wave model and applications

As mentioned before, the most important equation in any macroscopic traffic flow model is the conservation of vehicle equation

$$\frac{\partial k}{\partial t} + \frac{\partial q}{\partial x} = r(x,t) - s(x,t) \tag{9.1}$$

where $k = k(x,t)$ denotes the traffic density, describing the mean number of vehicles per unit roadway length at instant t and location x; $q = q(x,t)$ denotes the traffic volume or flow rate, and describes the mean number of vehicles passing the cross-section x per unit time at time t; r and s respectively denote the inflow and outflow from and to on- and off-ramps.

The *kinematic wave model* (or LWR model, or first-order model) assumes that the traffic volume can be determined from the fundamental relation between the density and the traffic volume (or likewise, the speed), i.e.

$$q = Q(k) \tag{9.2}$$

In the previous chapter, we have illustation how, using shock wave theory, we can construct analytical solutions of the kinematic wave model for simple flow-density relations. However, it turns out that shock wave analysis will not predict the correct behaviour of the flow for any fundamental diagram, since it does not predict the occurence of so-called *acceleration fans* (described in the ensuing of this chapter). These acceleration fans describe the behaviour of traffic flowing out of a congested area, e.g. a queue of stopped vehicles.

9.2.1 Analytical solutions using method of characteristics

This section presents the so-called *method of characteristics* that can be applied to construct analytical solutions to the kinematic wave model. We can rewrite the conservation of vehicle equation (assuming that $r = s = 0$) as follows

$$\frac{\partial k}{\partial t} + Q'(k)\frac{\partial q}{\partial x} = 0 \tag{9.3}$$

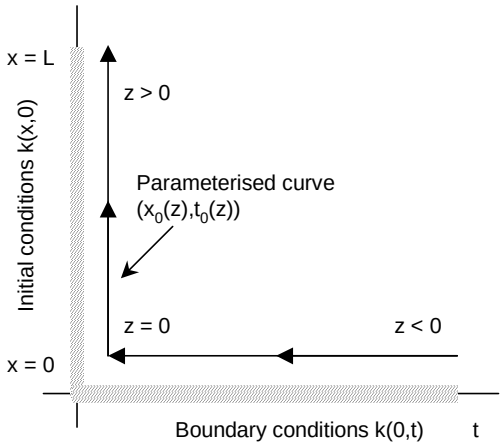

Figure 9.1: Parameterised curve $(x_0(z), t_0(z))$ describing the boundary of the region of interest.

where $Q'(k) = \frac{dQ}{dk}$. The kinematic wave model is a *first-order partial differential equation*. It is non-linear since $Q'(k)$ depends on k. The quantity $c(k) = Q'(k)$ will be referred to as the *characteristic speed*, since it describes the speed of the characteristic curves that play a pivotal role in the solution of hyperbolic equations, such as the kinematic model. For one, it turns out that the density k *is constant along any characteristic curve* of the kinematic wave model, and that the characteristic curves are therefore straight lines. It is also important to note that in the kinematic model the speed of the traffic $u(x,t)$ is always larger than the characteristic speed $c(k(x,t))$, i.e. traffic information never travels faster than the traffic that carries it. This implies that traffic flow is described as an *anisotropic fluid*: what happens behind a vehicle generally does not affect the behaviour of that vehicle.

Method of characteristics overview

Let us now take a closer look at solving equation (9.3). Let us first consider the boundary of the region of interest. Generally, the boundary consist of

1. the initial conditions (at $t = 0$) and
2. boundary conditions (at the start $x = 0$ and end $x = L$ of the roadway section).

We assume that we can construct a parameterised curve $(x_0(z), t_0(z))$ that describes these boundary of the region of interest. Fig. 9.1 shows an example of such a curve. Note that along this curve, the density is known!

For a specific value z, consider another parameterised curve C_z, starting from this boundary at location $(x_0(z), t_0(z))$

$$C_z = (x(s;z), t(s;z)) \quad \text{with} \quad x(0;z) = x_0(z) \quad \text{and} \quad t(0;z) = t_0(z) \tag{9.4}$$

For the example shown in Fig. 9.1, for $z > 0$, the curve C_z emanates from $t_0(z) = 0$ (the initial road conditions), while for $z < 0$, the curve emanates from $x_0(z) = 0$ (the upstream boundary of the region). We will define the curve by the following ordinary differential equations in t

$$\frac{dt}{ds} = 1 \quad \text{with} \quad t(0;z) = t_0(z) \tag{9.5}$$

$$\frac{dx}{ds} = c(k) = Q'(k) \quad \text{with} \quad x(0;z) = x_0(z) \tag{9.6}$$

for $s > 0$. Fig. 9.2 depicts two curves C_z, respectively emanating from the initial road conditions (in this case, for some $z > 0$) and from the boundary (for some $z < 0$). The curves C_z are

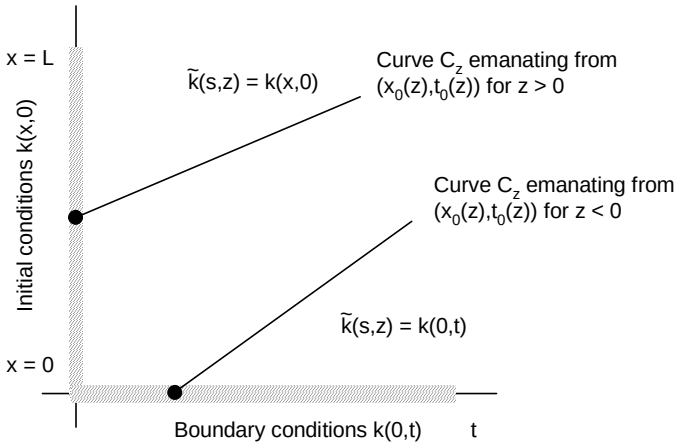

Figure 9.2: Curves C_z emanating from boundary describes by $(x_0(z), t_0(z))$.

called *characteristic curves* or *z-characteristics*.

By defining the curve C_z according to the equations (9.5) and (9.6), it turns out that the density $\tilde{k}(s;z) := k(x(s;z), t(s,z))$ along the curve C_z satisfies

$$\frac{d\tilde{k}}{ds} = \frac{\partial k}{\partial t}\frac{dt}{ds} + \frac{\partial k}{\partial x}\frac{dx}{ds} \tag{9.7}$$

$$= \frac{\partial k}{\partial t}\cdot 1 + \frac{\partial k}{\partial x}\cdot Q'(k) = 0 \tag{9.8}$$

Since $\frac{d\tilde{k}}{ds} = 0$, the density k along C_z is constant. More specifically, the density along C_z is equal to the density at the point where the curve C_z originates, i.e. at $s = 0$. As a result, we have

$$\tilde{k}(s;z) = k\left(x(s;z), t(s;z)\right) = k\left(x_0(z), t_0(z)\right) \tag{9.9}$$

Note that this implies that the curve C_z is a *straight line*, travelling at speed $\tilde{c}(z) = c\left(k\left(x(s;z), t(s;z)\right)\right)$.

To construct the solution of the kinetmatic wave model, initial and boundary conditions need to be established. By considering the curves emanating from the initial conditions as well as the boundaries, we can construct the solution $k(x,t)$ for $t > 0$ and $0 < x < L$.

If we, for the sake of argument, neglect the influence of the boundaries of the considered roadway section, we have seen that the formal solution of Eq. (9.3) is in fact

$$k\left(x + t\cdot ck(x,0), t\right) = k\left(x,0\right) \quad \text{for all } t > 0 \tag{9.10}$$

The speed of the characteristics (also referred to as *waves of constant density*) is equal to the derivative of the equilibrium flow $Q(k) = kV(k)$ and is hence positive as long as $k < k_c$ (unconstrained flow or free flow) and is negative as long as $k > k_c$ (constrained flow or congestion). Furthermore, the speed of the characteristics is bounded by the mean vehicle speed, i.e.

$$c(k) = \frac{dQ}{dk} = \frac{d\left(kV\left(k\right)\right)}{dk} = V(k) + k\frac{dV}{dk} < V(k) \tag{9.11}$$

since $\frac{dV}{dk} \leq 0$ (see Fig. 9.3). Only in the region when the average vehicle speed $V(k)$ is constant (the so-called *stable region*) and thus $\frac{dV}{dk} = 0$, we have $c(k) = V(k)$.

Effect of non-linearity of $Q\left(k\right)$

The objective of this section is to illustrate the effects of non-linearity of the flow-density curve, and particular the differences that will result when applying shock wave analysis and the LWR

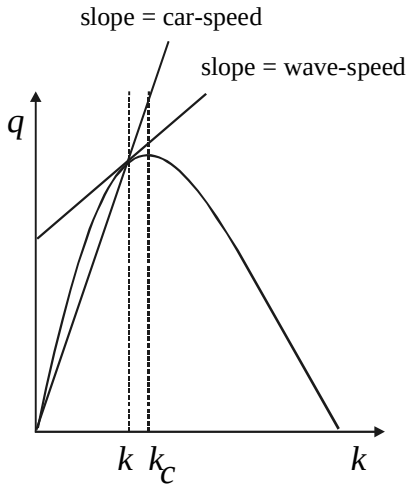

Figure 9.3: Interpretation of the wave speed

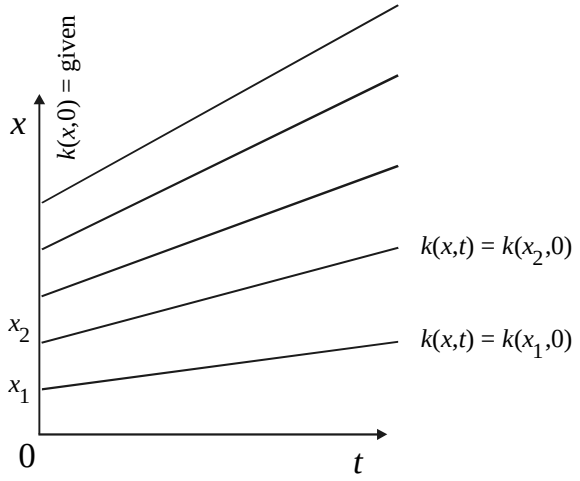

Figure 9.4: Lines of constant density

theory using non-linear flow-density curves. Waves of different density k will travel with different wave speeds $c(k) = Q'(k)$, and thus waves may either *fan out* or *converge and intersect*.

If $k(x,0)$ is a decreasing function of x (or equivalently, if the speed is an increasing function of x, so that cars are accelerating), then from Fig. 9.3 we notice that Q' - assumed to be a decreasing function of k - will be an increasing function of x. This corresponds to the situaiton illustrated in Fig. 9.4 in which the higher speed waves are initially ahead of the slower speed waves and so the region of acceleration tends to expand linearly with time as waves spread further and further apart. Fig. 9.5 shows a decreasing density $k(x,0)$ at time zero. A short time later the high density region has moved forward less than the low density region: a sharp initial change in density thus tends to disperse. At this point, we note that in this situation, we can determine the density $k(x,t)$ for all $t > 0$ is $k(x,0)$ is known. To this end, we only need to determine the slope of the characteristic emanating from $k(x,0)$ which intersects point (x,t), as s shown in Fig. 9.4.

If $k(x,0)$ is an *increasing function of* x, i.e. cars are decelerating, then the higher speed waves at low density are initially behind the slower speed waves for the higher density. A gradually increasing $k(x,0)$ tends to become steeper as shown in Fig. 9.6, and will eventually have a vertical tangent. This occurs as soon as any two waves of neighbouring values of k actually intersect. At this time, the solution (9.10) to the kinematic wave model breaks down

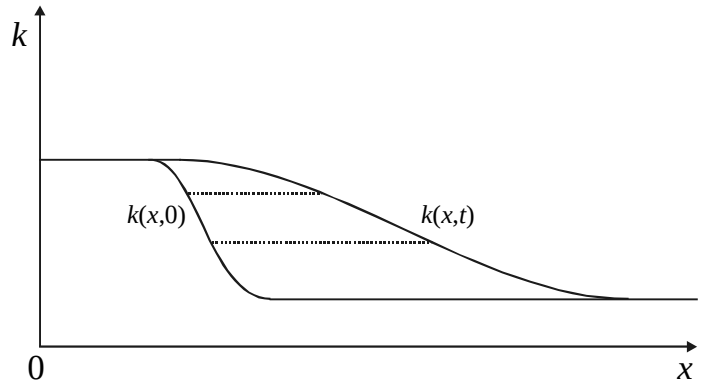

Figure 9.5: Dispersion of an acceleration wave

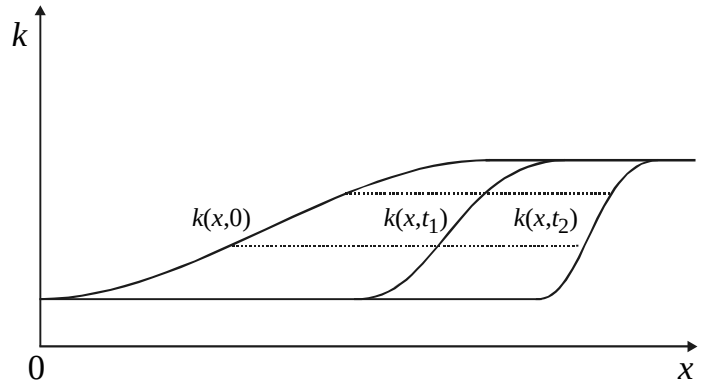

Figure 9.6: Focussing of deceleration waves

and , if it where continued beyond this time, would assign to some (x,t) point more than one value of k, becomes more than one wave passes through this point. The existence of a vertical tangent, however, means that $\frac{\partial k}{\partial x}$ becomes infinite and the differential equation from which the solution (9.10) was derived is no longer valid.

A phenomena such as this is familiar in fluid dynamics. The failure of the differential equation is corrected by allowing discontinuities in $k(x,t)$ which are called shocks. Shock wave theory was described in chapter 8 of these notes and is needed to complement the method of characteristics to find analytical solutions to kinematic wave model. In chapter 8 we have proven that the speed of the shock equals

$$\omega = \frac{q_2 - q_1}{k_2 - k_1} = \frac{Q(k_2) - Q(k_1)}{k_2 - k_1} \tag{9.12}$$

where k_1 and q_1 respectively denote the density and flow upstream of the shock S, and k_2 and q_2 denote the density and flow downstream of the shock. Recall that in the limit of very weak shocks, the shock line converges to the tangent of the Q curve (kinematic wave).

To evaluate $k(x,t)$, Eq. (9.12) is used to determine the path of the shock. Fig. 9.7 illustrates how in a typical problem one can construct the solution $k(x,t)$ graphically. If one is given $k(x,0)$ then one can draw the waves of constant k, or use the solution (9.10), to determine $k(x,t)$ at least until such time t_0 when two waves first intersect at a point (x_0,t_0). At this moment, a shock starts to form. Let us now determine the time t_0 and location x_0 at which this occurs.

Noe that except for very exceptional inital conditions $k(x,0)$ there is generally no *perfect focussing* of waves (waves converge on a single point). Therefore, the shock starts as an 'infinitesimal shock', meaning that the intersecting waves have nearly the same speed. The shock

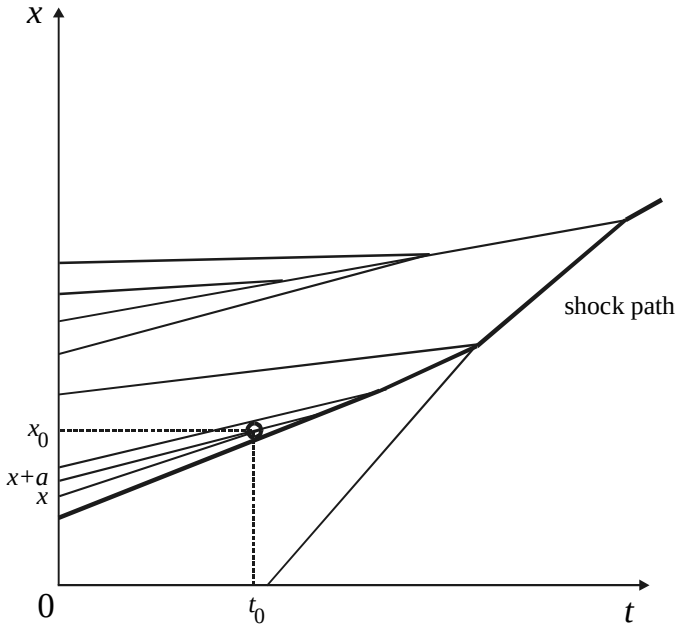

Figure 9.7: Construction of the shock path

will start with a speed approximately equal to those of the two intersecting waves. Futhermore, it is clear that the speed of the shock will have a speed which is between the speeds of the intersecting waves[1].

Let t_0 and x_0 be the first time two waves intersect and the location they intersection respectively. From Fig. 9.7 we observe that two waves starting at the points x and $x+a$ will intersect at this time t_0 and position x_0 given that the following condition is met:

$$x_0 = x + c(x)t_0 = x + a + c(x+a)t_0 \tag{9.13}$$

where $c(x) = Q'(k(x,0))$. Since t_0 is the first time any two waves intersect, it follows that

$$\frac{1}{t_0} = \max_{x,a} \frac{c(x) - c(x+a)}{a} \tag{9.14}$$

If $c(x)$ has a continuous derivative, it follows from the *mean value theorem* that $(c(x) - c(x+a))/a$ is equal to the derivate of c at some point between x and $x+a$. Therefore

$$\frac{1}{t_0} = \max_{x} \frac{dc}{dx} \tag{9.15}$$

The values of x for which the maximum is realised also give the values of x in Eqn. (9.13) that determine the position of x_0 where the shock originates.

We now know where the shock starts as well as its initial speed, thus also the position of the shock a short period later, at time $t_0 + \Delta t$. If no new shocks form between t_0 and $t_0 + \Delta t$, the values of k on either side of the shock at time $t_0 + \Delta t$ are still determined by the waves starting at $t = 0$. The density on either side of the shock is determined by the wave which intersects the schock at time $t_0 + \Delta t$ and approaches the shock from the same side. Another wave will approach from the other side and determine the density of that side. We now know the densities on either side of the shock at time $t_0 + \Delta t$ and we can reevaluate the shock speed at $t_0 + \Delta t$. We can then estimate the shock position at a still later time (e.g. $t_0 + 2\Delta t$), etc. In essesnce, we are performing a graphical integrtion of Eqn. (9.12).

[1] For weak shocks, the shock speed is approximately the average of the wave speeds at either side of the shock.

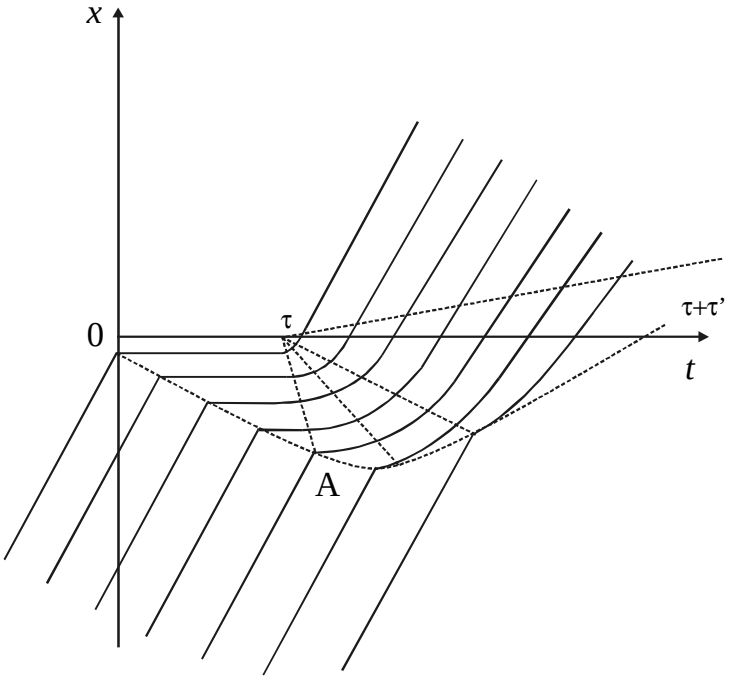

Figure 9.8: Cars stopped at a traffic signal

If some other shock should form elsewhere we proceed similarly to find their paths. If two shocks should intersect, we simple combine them into a single shock as shown in Fig. 9.7. The shock speed of the single shock is determined by the densities in the regions adjacent to the shock not including the region annihilated in the collision.

Example application of the kinematic wave model

As a practical example of how the kinematic wave model can be applied, let us consider what happens when a steady flow of traffic is suddenly stopped, for instance at an intersection, and then released again, as would occur at a traffic signal. The trajectories of cars which the theory should predict are shown in Fig. 9.8. The stopping point is at $x = 0$, and the trajectory of the first car that is stopped is designated by $x_1(t)$. The kinematic wave model *does not describe in detail the trajectory of car 1.* If we are given the density k_i, the intensity q_i or the speed v_i of the initial approach traffic strean, all other quantities are also determined by the $Q(k)$ relation. The speed at which car 1 approaches the stop line can be determined, since the traffic conditions are known. At the stopping line, the speed of the lead car is zero. The kinematic wave model does not describe the details of deceleration and acceleration of car 1, and additional information is required in order to correctly describe its behaviour. However, on a scale of time and distance in which the kinematic wave model aught to be applied, we may expect that deceleration and acceleration of car 1 is of negligible length of time.

The deceleration of the first car, being nearly instantaneous, creates a shock wave immediatly. It is a shock from the initial state (q_i, k_i) to the state $q = 0$ and $k = k_j$. The slope of the shock line in Fig. 9.9 between the states determines the shock speed. In Fig. 9.8, the shock starts at $x = 0$ and $t = 0$ and travels upstream with constant speed (shown by the dashed line). The shock line represents in effect the rear of the queue caused by the traffic signal. The continuum approximation implies that the decelerations occur practically instantaneously.

When the lead car accelerates again, it sends out a *fan of acceleration waves* for all car speeds between 0 and u_0. The slowest wave travels backwards with the wave speed c associatd with car speed $u = 0$. Clearly, this wave speed is given by the slope of $Q(k)$ at $k = k_j$, as shown in Fig. 9.9. When this wave starting at $x = 0$, $t = \tau$ as shown by the dashed line in Fig. 9.8

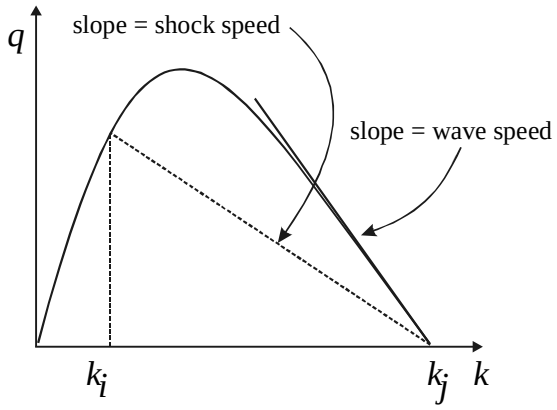

Figure 9.9: Evaluation of flows from a traffic signal

intersects the shock line at point A, this signals the car at the tail at the queue to move. The shock, however, does not apprear: the waves that intersect the shock only assign net values for the density or car speed on the front side of the shock. Since

$$\omega(t) = \frac{q_1 - q_2}{k_1 - k_2} = \frac{k_1 u_1 - k_2 u_2}{k_1 - k_2} \tag{9.16}$$

where ω again denotes the shock speed, and where state 1 and 2 respectively denote the state upstream and the state downstream of the shock. Since the flow q_1 upstream of the shock increases, the shock gains forward speed. It eventually moves with a positive speed and crosses the intersection at $x = 0$.

The time at which this occurs can be determined easily. We konw that the wave with the wave speed zero is the wave corresponding to q_c since the tangent to the $Q(k)$ curve is horizontal at $q = q_c$. The number of cars that cross the intersection before the shock arrives is therefore q_c multiplied by the time τ' for the shock to arrive. This must, however, also be equal to the total number of cars that have arrived by this time, i.e.

$$q_i(\tau + \tau') = q_c \tau' \tag{9.17}$$

in which τ is the length of time the traffic is stopped.

The shock, as it moves forward, becomes weaker and weaker. The car speeds at the rear of the shock remains at v_i but the car speeds at the front keep increasing. Eventually, the shock will overtake all waves for car speeds less thant v_i and the shock will degenerate into a wave or a shock of zero jump. The waves of car speeds larger than v_i however, move forward faster than te shock and the shock never reaches them.

9.2.2 Application of kinematic wave model for bicycle flows at signalised intersection

Consider a road on which a vehicle flow is moving in a certain direction. For the sake of argument, we assume that the flow is constant and equal to q_1. We assume that for the flow, the relation between density and flow can be approximated satisfactory using Greenshield's function

$$q\,(k) = u_0 k \left(1 - \frac{k}{k_j}\right) \tag{9.18}$$

where u_0 is the free speed of the vehicles and k_j is the jam-density. Assume that at the start of our analysis, no queue is present: traffic conditions are stationary and homogeneous (region 1). The density equals k_1 while the flow is equal to $q_1 = q\,(k_1)$. At a certain time (say, $t = -t_r$), the traffic signal that is present at $x = 0$ turns to red, causing the drivers to stop at the stopping

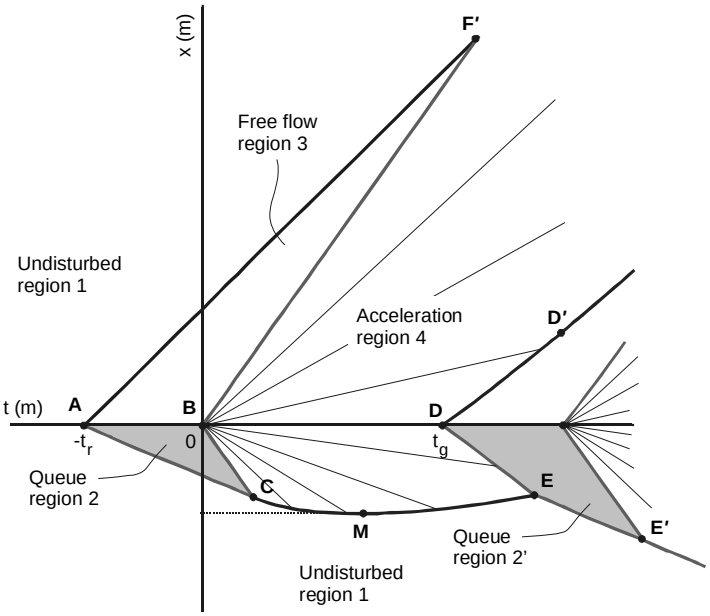

Figure 9.10: Emerging regions for controlled intersection example

line. Downstream of the stopping line, no vehicles will be present (region 3). As a result, a queue forms at the stopping line, which extend upstream (region 2, $x < 0$). At $t = 0$, the green phase starts and the first cyclists drive away from the stopping line. We assume that the drivers at the head of the queue react and accelerate instantaneously (without delay) *to the free speed* u_0. We have seen that the transition from the jam conditions for $x < 0$ to the free flow conditions downstream $x > 0$ causes a so-called *acceleration fan* (region 4). The acceleration fan described how the vehicles inside the congestion drive away from the queue.

At time $t = t_g$, the second red phase starts and cyclists are again held back at the stopping line. This causes the formation of a new queue (region $2'$). Fig. 9.10 shows qualitatively the predicted traffic flow conditions resulting when using the kinematic wave model. It also shows the different emerging regions as well as some points of interest. Furthermore, Fig. 9.11 shows a number of vehicle trajectories.

Let us now use what we have learnt so far to determine a mathematical solution to this problem. Let us first consider the shock wave between region 1 and region 2, described by the line $x_{AC}(t)$. The *speed of the shock wave cannot be determined from the method of characteristics.* Rather, we need to apply shock wave theory to determine the wave speed ω_{AC}. Since the traffic conditions upstream of the shock wave and downstream of the shock wave are constant, the shock speed is also constant. We find

$$\omega_{AC} = \frac{q_2 - q_1}{k_2 - k_1} = -\frac{q_1}{k_j - k_1} \tag{9.19}$$

which, by $(t_A, x_A) = (-t_r, 0)$

$$x_{AC}(t) = \omega_{AC} \cdot (t + t_r) = -\frac{q_1}{k_j - k_1}(t + t_r) \tag{9.20}$$

The next step is to determine the line $x_{AC}(t)$ describing the boundary between the jam region 2 and the acceleration region 4. The latter is described by the characteristics that are emanated from the point $(0, 0)$ (see Fig. 9.10). These characteristics are straight lines described by the kinematic wave speed

$$c(k) = q'(k) = u_0\left(1 - 2\frac{k}{k_j}\right) \tag{9.21}$$

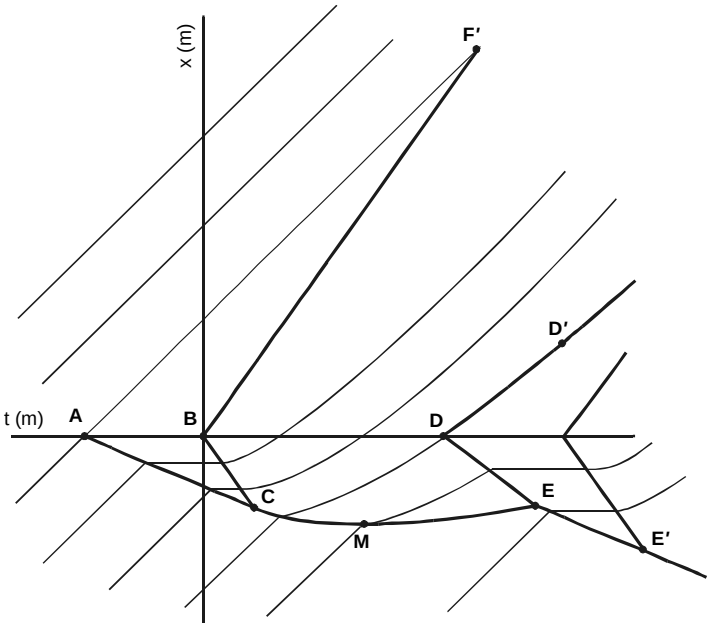

Figure 9.11: Example vehicle trajectories for controlled intersection example

The characteristic 'c' is then described by $x_c(t) = c \cdot t$, where $c \in [-u_0, u_0]$. It is very important to note that we can use this expression to determine the density $k_4(t,x)$ (and the speed and the flow) for all (t,x) in region 4:

$$k_4(t,x) = \frac{k_j}{2u_0}\left(u_0 - \frac{x}{t}\right) \tag{9.22}$$

$$u_4(t,x) = \frac{1}{2}\left(u_0 + \frac{x}{t}\right) \tag{9.23}$$

$$q_4(t,x) = \frac{1}{4}\frac{k_j}{u_0}\left[u_0^2 - \left(\frac{x}{t}\right)^2\right] \tag{9.24}$$

for all (t,x) in region 4.

The characteristic that moves *upstream* (in the direction $x < 0$) at the highest speed moves upstream with speed $-u_0$. This is precisely the characteristic that separates region 2 and region 4. Thus, the line $x_{BC}(t)$ moves with speed $-u_0$ and is thus given by the following expression (since $x_{BC}(0) = 0$)

$$x_{BC}(t) = -u_0 t \tag{9.25}$$

We can then determine the coordinates of point C easily. Some straightforward computations show that

$$t_C = \frac{k_1}{k_j - k_1} t_r \quad \text{and} \quad x_C = -u_0 t_C = -u_0 \frac{k_1}{k_j - k_1} \tag{9.26}$$

The next step is to determine a relation for the shock wave $x_{CE}(t)$ separating undisturbed region 1 and the acceleration region 4. On the contrary to the shock wave $x_{AC}(t)$, the speed of the shock $x_{CE}(t)$ is *not constant* since the downstream conditions (in region 4) are non-stationary (since they satisfy Eqns.(9.22)-(9.23)). We can still however apply shock wave theory on the non-stationary conditions. This provides us with the following expression for the shock wave speed $\omega_{CE}(t)$ describing the slope of line $x_{CE}(t)$

$$\omega_{CE}(t) = \frac{q_4 - q_1}{k_4 - k_1} = \frac{1}{2}\frac{k_j\left[u_0^2 - \left(\frac{x_{CE}(t)}{t}\right)^2\right] - 4q_1u_0}{k_j\left[u_0 - \left(\frac{x_{CE}(t)}{t}\right)\right] - 2k_1u_0} \tag{9.27}$$

Since

$$x_{CE}(t) = x_{CE}(t_C) + \int_{t_C}^{t} \omega_{CE}(s)\, ds \tag{9.28}$$

we can determine $x_{CE}(t)$ by solving the following ordinary differential equation

$$\frac{d}{dt} x_{CE} = \omega_{CE}(t) \quad \text{with} \quad x_{CE}(t_C) = x_C \tag{9.29}$$

where $\omega_{CE}(t)$ depends on $x_{CE}(t)$ according to Eq. (9.27). Solving this differential equation is not straightforward and requires application of the method of *variation of constants*. We can determine that:

$$x_{CE}(t) = c(k_1)\, t - (u_0 + c(k_1)) \sqrt{t_C t} \tag{9.30}$$

is the solution that satisfies Eq. (9.29).

Proof. Consider the ordinary differential equation

$$\frac{dx}{dt} = \frac{1}{2} \frac{k_j \left[u_0^2 - \left(\frac{x}{t}\right)^2\right] - 4 q_1 u_0}{k_j \left[u_0 - \left(\frac{x}{t}\right)\right] - 2 k_1 u_0} \tag{9.31}$$

Define $x = ty$. Then

$$\frac{dx}{dt} = t \frac{dy}{dt} + y \tag{9.32}$$

and thus Eq. (9.31) becomes

$$t \frac{dy}{dt} + y = \frac{1}{2} \frac{k_j \left(u_0^2 - y^2\right) - 4 q_1 u_0}{k_j (u_0 - y) - 2 k_1 u_0} \tag{9.33}$$

Then, we can easily derive that

$$t \frac{dy}{dt} = \frac{k_j \left(u_0^2 - y^2\right) - 4 q_1 u_0 - 2 k_j (u_0 - y)\, y + 4 k_1 u_0 y}{2 k_j (u_0 - y) - 4 k_1 u_0} \tag{9.34}$$

$$= k_j \frac{y^2 - 2 u_0 \left(1 - 2\frac{k_1}{kj}\right) y + u_0^2 - 4 \frac{q_1}{k_j} u_0}{2 k_j (u_0 - y) - 4 k_1 u_0} \tag{9.35}$$

$$= \frac{1}{2} \frac{y^2 - 2 u_0 \left(1 - 2\frac{k_1}{kj}\right) y + u_0^2 - 4 \frac{k_1 u_0 \left(1 - \frac{k_1}{k_j}\right)}{k_j} u_0}{u_0 \left(1 - 2\frac{k_1}{k_j}\right) - y} \tag{9.36}$$

$$= \frac{1}{2} \frac{y^2 - 2 u_0 \left(1 - 2\frac{k_1}{kj}\right) y + u_0^2 \left(1 - 2\frac{k_1}{k_j}\right)^2}{u_0 \left(1 - 2\frac{k_1}{k_j}\right) - y} \tag{9.37}$$

Note that $c(k_1) = u_0 \left(1 - 2\frac{k_1}{k_j}\right)$ and $c^2(k_1) = u_0^2 \left(1 - 2\frac{k_1}{k_j}\right)^2$ and thus

$$t \frac{dy}{dt} = -\frac{1}{2} \frac{y^2 - c(k_1)\, y + c^2(k_1)}{y - c(k_1)} = -\frac{1}{2} \frac{(y - c(k_1))^2}{y - c(k_1)} = -\frac{1}{2} y + \frac{1}{2} c(k_1) \tag{9.38}$$

We can again use Eq. (9.32) to see that the equation above becomes

$$t \frac{dy}{dt} + y - \frac{1}{2} y = \frac{1}{2} c(k_1) \rightarrow \frac{dx}{dt} = \frac{1}{2} \frac{x}{t} + \frac{1}{2} c(k_1) \tag{9.39}$$

This is a linear differential equation, which can be solved by variation of constants. To this end, we first determine the solution to the reduced differential equation

$$\frac{dx}{dt} = \frac{1}{2} \frac{x}{t} \tag{9.40}$$

which equals

$$x = C\sqrt{t} \tag{9.41}$$

The variation of constant approach involves trying the following solution in the non-reduced differential equation (9.39)

$$x = C(t)\sqrt{t} \tag{9.42}$$

which yields

$$\frac{dx}{dt} = C(t)\frac{1}{2}\frac{1}{\sqrt{t}} + \frac{dC}{dt}\sqrt{t} = \frac{1}{2}\frac{C(t)}{\sqrt{t}} + \frac{1}{2}c(k_1) \tag{9.43}$$

$$\frac{dC}{dt}\sqrt{t} = \frac{1}{2}c(k_1) \tag{9.44}$$

and thus $C(t) = c(k_1)\sqrt{t} + C_0$, which in turn yields

$$x(t) = \left(c(k_1)\, t + C_0\sqrt{t}\right) \tag{9.45}$$

We can determine the integration constant C_0 by using the boundary condition $x(t_C) = x_C$

$$C_0 = -\left(u_0 + c(k_1)\right)\sqrt{t_C} \rightarrow x(t) = c(k_1)\, t - \left(u_0 + c(k_1)\right)\sqrt{t_C t} \tag{9.46}$$

■

Let us assume that the next red phase starts at $t = t_g$ before the shock x_{CE} has reached the stopping line. Let us first describe the shock wave between region 4 and region $3'$. Again, this shock will not move at a constant speed, since the traffic conditions upstream of the shock are non-stationary. We have

$$\omega_{DD'}(t) = \frac{q_4 - q_{3'}}{k_4 - k_{3'}} = \frac{q_4}{k_4} = u_4 \tag{9.47}$$

that is, the speed of the shock equals the speed of the last vehicle emanating from the queue before the start of the second red phase. Again we need to solve a differential equation to determine the shock wave $x_{DD'}(t)$

$$\frac{d}{dt}x_{DD'} = \frac{1}{2}\left(u_0 + \frac{x_{DD'}}{t}\right) \quad \text{with} \quad x_{DD'}(t_g) = 0 \tag{9.48}$$

which results in

$$x_{DD'}(t) = u_0\left(t - \sqrt{t \cdot t_g}\right) \tag{9.49}$$

Proof. Eq. (9.48) is an ordinary linear differential equation of the general form $a(t)\frac{dx}{dt} = b(t)\, x + c(t)$. To solve differential equations of this sort, we first consider the reduced differential equation

$$\frac{dx}{dt} = \frac{x}{2t} \tag{9.50}$$

which is easily solved by separating the variables x and t

$$\frac{1}{x}dx = \frac{1}{2t}dt \tag{9.51}$$

which is solved by

$$\ln x = \frac{1}{2}\ln x + C \tag{9.52}$$

for some integration constant C; x thus satisfies

$$x = C\sqrt{t} \tag{9.53}$$

To solve the non-reduced differential equation, the *variation of constant* method is used. That is, we try the following solution

$$x = C(t)\sqrt{t} \tag{9.54}$$

Substituting this solution in Eq. (9.48) yields

$$C(t) = u_0\sqrt{t} + C_0 \tag{9.55}$$

The integration constant C_0 can be determined by considering the initial condition $x(t_g) = 0$ which yields the solution

$$x(t) = u_0\left(t - \sqrt{t \cdot t_g}\right) \tag{9.56}$$

■

In a similar way, we can determine the shock wave $x_{DE}(t)$ separating region 4 and region $2'$. The speed of the shock wave equals

$$\begin{aligned} \omega_{DE}(t) &= \frac{q_{2'} - q_4}{k_{2'} - k_4} = -\frac{q_4}{k_j - k_4} && (9.57)\\ &= -\frac{1}{2}\left[u_0 - \left(\frac{x_{DE}}{t}\right)\right] && (9.58) \end{aligned}$$

and thus

$$\frac{d}{dt}x_{DE} = -\frac{1}{2}\left[u_0 - \left(\frac{x_{DE}}{t}\right)\right] \quad \text{with} \quad x_{DE}(t_g) = 0 \tag{9.59}$$

which has the following solution

$$x_{DE}(t) = -u_0\left(t - \sqrt{t_g t}\right) \tag{9.60}$$

Finally, the shock $x_{EE'}(t)$ separating region 1 and region $2'$ moves at constant speed $\omega_{EE'}$ which equals

$$\omega_{EE'} = \omega_{AC} \tag{9.61}$$

9.2.3 Implications for fundamental diagram

The previous section showed how the method of characteristics can be applied to determine solutions to the kinematic wave model. We have seen how shocks are formed in regions where the speed decreases with space. We have also seen how traffic leaving a queue gives rise to a so-called *acceleration fan*, consisting of characteristics with different speeds $c(k) = \frac{d}{dk}Q(k)$.

Implicitly, during the analysis, we have used special assumptions regarding the shape of the fundamental diagram. More precisely, we assumed that the fundamental diagram $Q(k)$ is a *concave function*, i.e.

$$\frac{d^2Q}{dk^2} \leq 0 \tag{9.62}$$

Let us illustrate why this property is important to represent traffic behaviour correctly. To this end, consider a flow-density curve that is non-concave in the congested branch. Consider the case illustrated by Fig. 9.13. In this case, the initial conditons $k(x,0)$ describe the case of smoothly decreasing densities. Note that when the $Q(k)$ function is concave, this region would smooth out further. Fig. 9.13 shows that, since the congested branch of $Q(k)$ is convex instead of concave, the slopes of the characteristics emanating from the initial conditions at $t = 0$ become more negative as x increases. In the end, the waves even intersect, causing a shock wave. In other words, although dense traffic travels slower, the waves the dense traffic carries travel faster than the waves in light traffic, given rise to the shock. This shock is thus formed by traffic that drives out of the high into the lower density region.

Exercise 55 *Describe yourself what will happen in the convex region of the fundamental diagram in case the initial $k(x,0)$ describes conditions in which the density $k(x,0)$ increases with increasing x. Also discuss why this is not realistic.*

Remark 56 *Appendix A discusses an alternative solution approach for the kinematic wave model that is based on application of Green's theorem.*

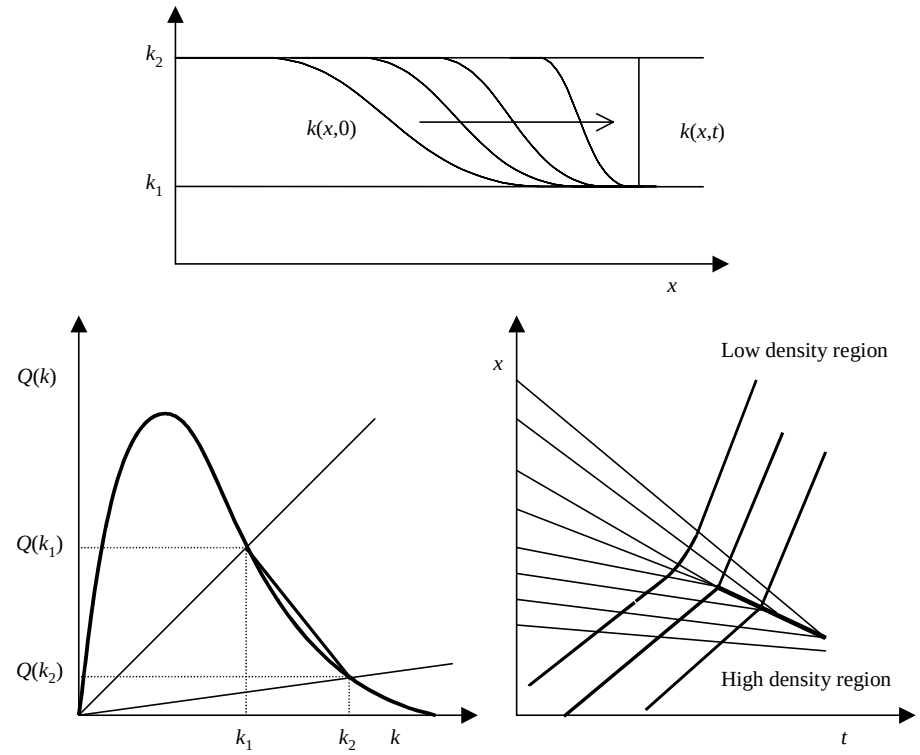

Figure 9.12: Effect of convex congested branch of $Q(k)$.

9.3 Numerical solutions to the kinematic wave model

The advantage of numerical results presented thus far is that they visually depict the effects of downstream disturbances on upstream flow. Thus they provide a good insight into the formation and dissipation of queues and congestion in time and space. The *major disadvantage* of analytical approaches is that they are generally only applicable to simple situations, e.g. cases where no on-ramps or off-ramps are present, where initial conditions and arrival patterns are simple, cases in which there is no interaction between traffic lights, etc. However, in practise, such ideal situations are seldom encountered and other approaches are needed, namely using numerical approximations.

9.3.1 Approach to numerical approximation

Numerical solution approaches come in all sizes and shapes. In most cases, however, the approaches are based on discretisation in time and space. For one, this means that the roadway is divided in small *cells* i. In most cases, these cells have a fixed length l_i (see Fig. 9.13). Cell i has interfaces at x_{i-1} and x_i, i.e. $l_i = x_i - x_{i-1}$. Furthermore, let t_j denote a discrete time instant.

Considering time instant t_j, the number of vehicles in cell i at t_j can be determined from using the cumulative flow function $\tilde{N}(x,t)$, or by integration of the density along the x. The latter yields that the number of vehicles in cell i at instant t_j equals

$$\int_{x_{i-1}}^{x_i} k\left(x, t_j\right) dx = l_i \cdot k_{i,j} \tag{9.63}$$

where $k_{i,j}$ denotes the average vehicle density (*space-averaged density*) in cell i at time instant t_j, with $T_j = t_{j+1} - t_j$. At the same time, the number of vehicles passing the interface at x_i during a period $[t_j, t_{j+1})$ can be determined directly from the cumulative flow function $\tilde{N}(x,t)$,

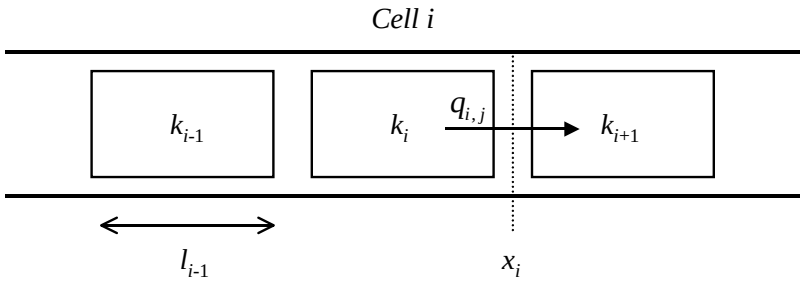

Figure 9.13: Discretisation of roadway into cells i of length l_i

or by integration of the flow rate during this period, i.e.

$$\int_{t_j}^{t_{j+1}} q(x_i, t)dt = T_j \cdot q_{i,j} \tag{9.64}$$

where $q_{i,j}$ denotes the average flow (*time-averaged flow*) across the interface at x_i during the period $[t_j, t_{j+1})$. By conservation of vehicles (or by integration of the conservation of vehicle equation (9.1) over the region $[x_{i-1}, x_i) \times [t_j, t_{j+1})$), we have

$$l_i \left(k_{i,j+1} - k_{i,j}\right) + T_j \left(q_{i,j} - q_{i-1,j}\right) = 0 \tag{9.65}$$

(assuming that the on-ramp inflows and off-ramp outflows are zero, i.e. $r = s = 0$) or

$$\frac{k_{i,j+1} - k_{i,j}}{T_j} + \frac{q_{i,j} - q_{i-1,j}}{l_i} = 0 \tag{9.66}$$

Note that the discrete representation (9.66) is still exact, i.e. no numerical approximation has been applied so far. The remaining problem is to find appropriate expressions for the time-averaged flow $q_{i,j}$. This will require two types of approximation, namely *space averaging* and *time evolution.*

9.3.2 Fluxes at cell-interfaces

In a numerical approximation scheme, in general the spaced-average densities $k_{i,j}$ are computed for all cells i and all time instants t_j. Hence, when establishing approximations for the time-averaged flow $q_{i,j}$, we need to express the flow $q_{i,j}$ as a function of the densities in the upstream cell i, the downstream cell $i+1$ at time instants t_j and t_{j+1}, i.e.

$$q_{i,j} = f\left(k_{i,j}, k_{\,i+1,j}, k_{i,j+1}, k_{i+1,j+1}\right) \tag{9.67}$$

Clearly, when the flows $q_{i,j}$ have been determined, eqn. (9.66) can be solved, either rightaway (when $q_{i,j} = f\left(k_{i,j}, k_{i+1,j}\right)$) or using an iterative approach (generally, when $q_{i,j} = f\left(k_{i,j}, k_{\,i+1,j}, k_{i,j+1}, k_{i+1,j+}\right.$ Schemes where the function f only depends on the 'current time' t_j are called *explicit schemes*; schemes where the function f depends also on the future time t_{j+1} are called *implicit schemes.* In general, explicit schemes are computationally more efficient, but less stable. For now, only explicit schemes will be considered.

9.3.3 Simple explicit schemes

The most simple scheme imaginable describes the flow $q_{i,j}$ out of cell i as a function of the conditions in that cell only, i.e.

$$q_{i,j} = f\left(k_{i,j}\right) = Q\left(k_{i,j}\right) \tag{9.68}$$

It is obvious that using this expression will not provide a realistic representation of upstream moving shock waves. Image a situation where the downstream cell $i+1$ is congested (e.g. due to a blockade upstream of the cell). The analytical solution of the kinematic model states that the flow out of cell i will be equal to zero, and that the region of jam-density will move further upstream. However, using Eq. (9.68) shows that traffic will keep flowing out of cell i. As a result, the density in the *recieving* cell $i+1$ will become unrealistically high.

An easy remedy to this problem, would be to assume that the flow out of cell $i+1$ is only a function of the conditions in the receiving cell $i+1$, i.e.

$$q_{i,j} = f\left(k_{i+1,j}\right) = Q\left(k_{i+1,j}\right) \tag{9.69}$$

Although this would clearly remedy the problem of traffic flowing into congested cells, the problem now is that the *transmitting cell* i will also transmit vehicles when its empty. Some researchers have proposed combining both approaches using weighted averages, i.e.

$$q_{i,j} = \alpha Q\left(k_{i,j}\right) + (1-\alpha)\, Q\left(k_{i+1,j}\right) \tag{9.70}$$

The problem is then to determine the correct values for α.

9.3.4 Flux-splitting schemes

A class of well-known schemes are the so-called flux-splitting schemes. Recall that the conservation of vehicle equation can be written as follows

$$\frac{\partial k}{\partial t} + c\left(k\right)\frac{\partial k}{\partial x} = 0 \tag{9.71}$$

where $c\left(k\right) = Q'\left(k\right)$. The wave speeds $c\left(k\right)$ in effect describe how information is transmitted in the flow: when $c\left(k\right) > 0$, information is transmitted downstream, i.e. from cell i to cell $i+1$. When $c\left(k\right) < 0$, the waves propagate in the upstream direction. The flux-splitting approach is aimed at splitting the flux (flow) into contributions into the downstream direction and contributions into the upstream directions. These are respectively described by

$$c^{+}\left(k\right) = \max\left\{0, c\left(k\right)\right\} = \frac{1}{2}\left(c\left(k\right) + \left|c\left(k\right)\right|\right) \tag{9.72}$$

$$c^{-}\left(k\right) = \min\left\{0, c\left(k\right)\right\} = \frac{1}{2}\left(c\left(k\right) - \left|c\left(k\right)\right|\right) \tag{9.73}$$

Note that

$$c^{+}(k) = \begin{cases} c(k) & k < k_c \\ 0 & k \geq k_c \end{cases} \quad \text{and} \quad c^{-}(k) = \begin{cases} 0 & k < k_c \\ c\left(k\right) & k \geq k_c \end{cases} \tag{9.74}$$

Note that $Q\left(k\right) = f\left(k\right) = \int c\left(\kappa\right) d\kappa$. Now, let

$$f^{\pm}\left(k\right) = \int c^{\pm}\left(\kappa\right) d\kappa \tag{9.75}$$

which in turn yields

$$f^{+}\left(k\right) = \int_0^k c^{+}\left(\kappa\right) d\kappa = \begin{cases} Q(k) & k < k_c \\ Q\left(k_c\right) & k \geq k_c \end{cases} \tag{9.76}$$

and

$$f^{-}\left(k\right) = \int_0^k c^{-}\left(\kappa\right) d\kappa = \begin{cases} 0 & k < k_c \\ Q\left(k\right) - Q\left(k_c\right) & k \geq k_c \end{cases} \tag{9.77}$$

Note that $f^{+}(k)$ is always positive, while $f^{-}\left(k\right)$ is always negative. Furthermore, note that

$$f\left(k\right) = f^{+}\left(k\right) + f^{-}\left(k\right) = Q\left(k\right) \tag{9.78}$$

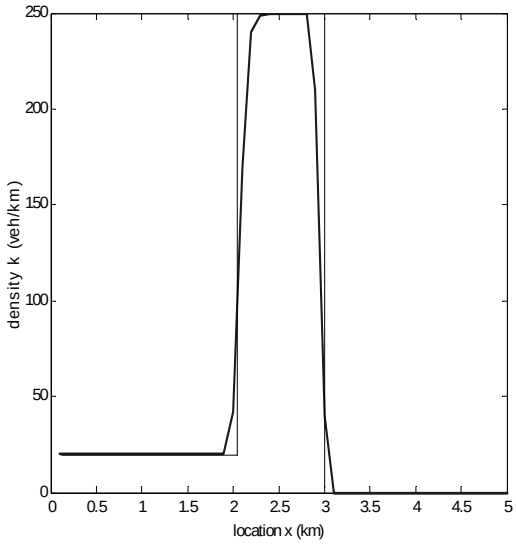

Figure 9.14: Numerical solution using flux-splitting approach and exact solution.

The flux-splitting scheme infers that the numerical flux $f^*_{i,j}$ at the cross-section x_i is approximated by considering the downstream moving contribution $f^+(k_{i,j})$ of cell i and the upstream moving contribution $f^*(k_{i+1,j})$ of cell $i+1$ as follows

$$f^*_{ij} = f^+(k_{i,j}) + f^-(k_{i+1,j}) \tag{9.79}$$

It turns out that this relatively simple scheme provides excellent results, which can be applied easily in any practical application of the kinematic wave model. From a theoretical viewpoint, the method has some undesirable properties. It is well known that the approach tends to smooth shocks. An example of this is shown in Fig. 9.14. The extent to which this occurs depends on the time-step and the cell lengths.

9.3.5 Godunov schemes

The flux-splitting approach has some practical and theoretical problems, one of which is the fact that in theory, the flux may become negative. The Godunov scheme, which closely resembles the flux-splitting scheme, does not have these drawbacks. In fact, it turns out that this scheme is both simple, and can be interpreted from the viewpoint of traffic flow theory.

Riemann problem[2]

The Godunov solution scheme is based on the solution of a specific problem, which is know as the so-called Riemann problem. The Riemann problem is defined by an intial value problem with the following starting conditions

$$u(x,0) = \begin{cases} u_L & x < 0 \\ u_R & x > 0 \end{cases} \tag{9.80}$$

Note that the LWR model assumes that the flows are in equilibrium, implying that the speeds determine the densties and the flows.

We can now define four different cases and determine the exact solution for these case. In the fist case 1, depicted by Fig. (9.15), both the downstream and upstream flow conditions

[2] In this section, we assume that the flow-density curve is concave.

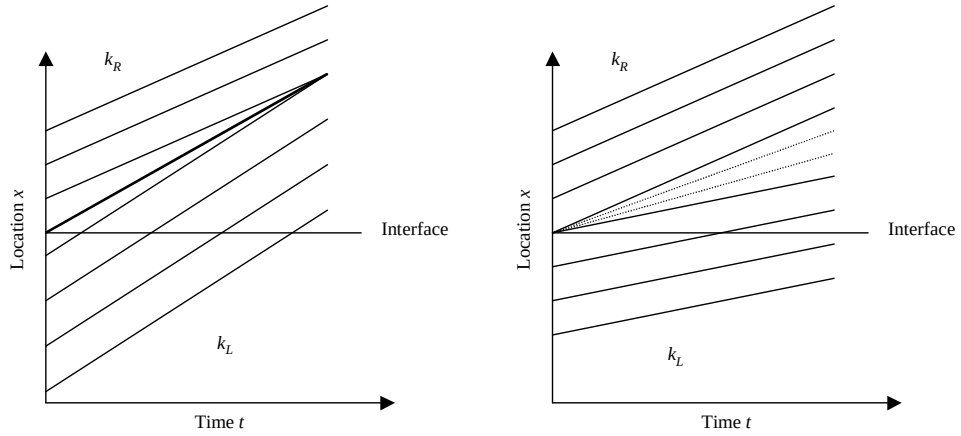

Figure 9.15: Case 1: downstream flows (k_R, u_R) and upstream flows (k_L, u_L) undercritical.

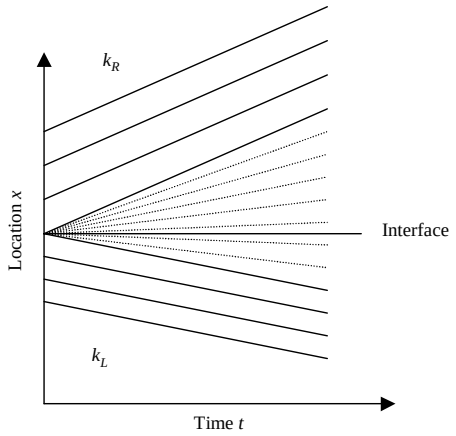

Figure 9.16: Case 2: downstream flows (k_R, u_R) undercritical and upstream flows (k_L, u_L) overcritical.

are undercritical. When $k_R > k_L$ (and thus $q_R > q_L$), an upstream moving shock wave will occur; when $k_R < k_L$, an acceleration fan will occur that describes the interface between the two regions. In either case, the flow f^* at the interface at $x = 0$ equals the downstream flow, i.e $f^* = q_L$.

Case 2 depicts the situation where the downsstream conditions k_R are undercritical, while the upstream conditions are overcritical. In this situation, shown in Fig. 9.16, an acceleration fan is transmitted from $x = 0$ at $t = 0$. Since along the interface at $x = 0$, the acceleration fan transmits a characteristic with slope 0, we can use the fact that the slope of a characteristic satisfies $c\,(k) = dQ/dk$ to see that the (constant) density along this characteristic equals the critical density k_c, and hence the flow f^* at the cross-section at $x = 0$ equals the capacity q_c.

Case 3 represents the situation where the downstream flow conditions are overcritical, while the upstream flows are undercritical. In this situation, shown in Fig. 9.17, a shock wave will emerge at $x = 0$ at time $t = 0$. The flow f^* at the cross-section at $x = 0$ depends on the direction of the shock: if the shock moves downstream, i.e. when $q_R < q_L$, the flow f^* at the interface equals the upstream flow q_L. When the shock moves upstream, which happens when $q_R > q_L$, the flow f^* at the interface equals the downstream flow q_R. Clearly, the following relation is valid in this situation: $f^* = \min\{q_L, q_R\}$.

The final case 4 (Fig. 9.18) is the case where the downstream conditions k_R and the upstream conditions k_L are overcritical. Similar to case 1, either an acceleration fan or a shock wave

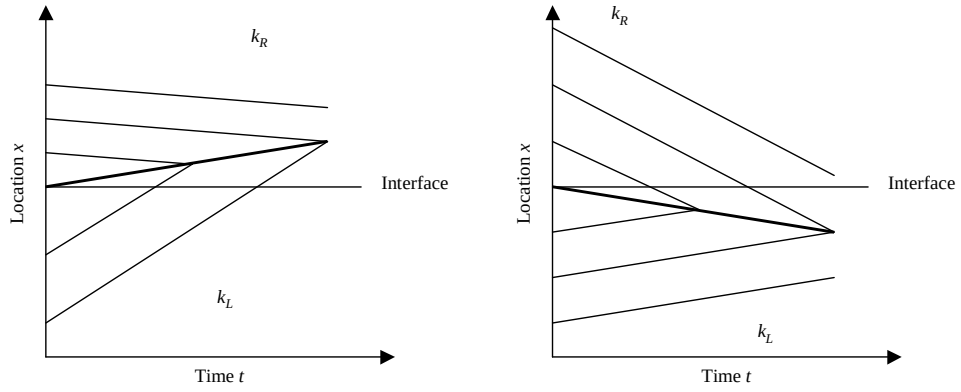

Figure 9.17: Case 2: downstream flows (k_R, u_R) overcritical and upstream flows (k_L, u_L) undercritical.

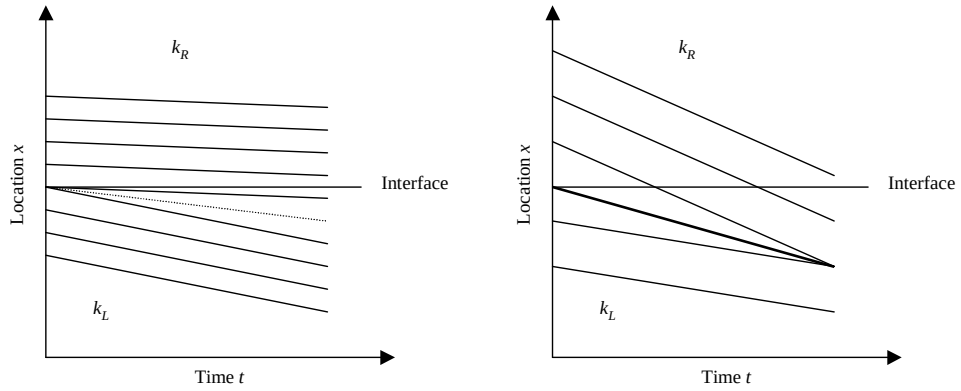

Figure 9.18: Case 4: downstream flows (k_R, u_R) and upstream flows (k_L, u_L) overcritical.

emerge. In either case, the flow f^* across the cross-section at the interface at $x = 0$ equals the flow q_R in the downstream region.

Table 9.1 summarises the results for cases 1 to 4.

Let us now defined the *traffic demand* D_L of the upstream region and the traffic supply S_R of the downstream region as follows

$$D_L = \begin{cases} q_L & k_L < k_c \\ q_c & k_L > k_c \end{cases} \tag{9.81}$$

and

$$S_R = \begin{cases} q_c & k_R < k_c \\ q_R & k_R > k_c \end{cases} \tag{9.82}$$

The traffic demand describes the traffic flow aiming to move out of the upstream region L (assuming that the flow conditions in the downstream region R allows this flow). For undercritical conditions, the demand D_L equals the flow q_L; for overcritical conditions the demand is limited to the capacity q_c.

Flow f^* at interface	$k_R < k_c$	$k_R > k_c$
$k_L < k_c$	q_L	$\min\{q_L, q_R\}$
$k_L > k_c$	q_c	q_R

Table 9.1: Flow at interface between upstream and downstream region

The traffic supply describes the maximum traffic flow the downstream region R can admit. When the conditions in the downstream region are undercritical, this flow is limited by the capacity of the road in this region; when the conditions in the downstream region are overcritical, the flow is restricted by the flow q_R in the downstream region. In a way, this flow describes the space becoming available per unit time.

Using these definitions of demand and supply, we can use the results depicted in table 9.1 to show that the realised flow f^* at the cross-section equals

$$f^* = \min\{D_L, S_R\} \tag{9.83}$$

That is, the flow f^* at the interface equals the minimum between demand and supply.

Godunov's scheme

The Godunov scheme is based on the solution to the Riemann problem. To this end, it is assumed that the conditions in the (flow) transmitting cell i and the recieving cell $i+1$ are homogeneous with the cell, i.e.

$$k(x, t_j) = k_{ij} \text{ for all } x_{i-1} \leq x \leq x_i \tag{9.84}$$

It needs no further clarification that to determine the flow at the cell interface x_i between cell i and cell $i+1$, one in fact has to solve the Riemann problem. Thus, for cell i we define the demand D_i by

$$D_i = \begin{cases} q_i = Q(k_i) & k_i < k_c \\ q_c = Q(k_c) & k_i > k_c \end{cases} \tag{9.85}$$

while for cell $i+1$ we define the supply

$$S_{i+1} = \begin{cases} q_c = Q(k_c) & k_{i+1} < k_c \\ q_{i+1} = Q(k_{i+1}) & k_{i+1} > k_c \end{cases} \tag{9.86}$$

It the time period is short enough (i.e. assuming that the conditions in cells $i-1$ and $i+2$ will not reach the interface at x_i, the flow $f^*_{i,j}$ during the period $[t_j, t_{j+1})$ equals

$$f^*_{i,j}(k_{i,j}, k_{i+1,j}) = \min\{D_i, S_{i+1}\} \tag{9.87}$$

Note that, given the assumption that the conditions in both cells i and $i+1$ are homogeneous, the flow $f^*_{i,j}$ between the cells is an exact solution of the LWR model; the approximation lies in the assumption that the homogeinity assumption holds for each time step t_j (which is obviously not true).

It turns out that the Godunov scheme has excellent properties, such as accurate representation of shocks, ability to incorporate flow-density relations that are a function of x, etc.

9.4 Higher-order models

The kinematic wave model is a simple yet sufficient theory of traffic if one only cares about the size and the end of a queue, i.e. the time-space trajectory of a shock. However, traffic flow phenomena are very complex and some rather important phenomena elude the kinematic wave model, as we have seen in chapter 5. The kinematic wave model described in the preceding chapter has been criticised by many authors, mainly because of the following reasons:

- Speeds in the kinematic wave model are described by stationary speed-concentration relations implying that the speed reacts instantaneously to the traffic concentration without any delay. Dynamic fluctuations around the equilibrium speed are observed in real-life traffic flow (hysteresis; recall the acceleration and deceleration curves from chapter 5) but are not described by the kinematic wave model.

- The kinematic wave theory shows (speed-) shock formation by steeping speed jumps to infinite sharp discontinuities in the concentration. That is, it produces discontinuous solutions irrespective of the smoothness of the initial conditions. These are in contradiction with smooth shocks observed in real-life traffic flow.
- The kinematic wave theory is not able to describe regular start-stop waves with amplitude dependent oscillation times which have been observed in real-life traffic (recall transition of synchronised flow into wide moving jams). Besides being a source of stress and irritation for drivers, stop-and-go traffic also produces higher amounts of pollutants, while leading to a higher fuel consumption due to the frequency of acceleration and deceleration.
- In real-life traffic flow, hysteresis phenomena have been observed, showing that the average headways of vehicles approaching a jam are smaller that those of vehicles flowing out of a jam (in case of relaxation dominant phase; see chapter 5). These phenomena are not described by the kinematic wave theory.
- Kinematic wave models do not address the issue of traffic instability. It has been often observed that in a certain traffic concentration range, small perturbations can induce violent transitions.We again refer to chapter 5 where the issue of traffic instability was discussed.

To resolve these problems, different approaches have been considered in the past.

9.4.1 Derivation of the model of Payne

The model of [42] consist of the conservation of vehicle equation, and a dynamic equation describing the dynamics of the speed $u(x,t)$. The model can be derived in several ways, but the most intuitive one is by considering the following, simple car-following model

$$v_\alpha(t+\tau) = \tilde{V}(s_\alpha(t)) = V\left(\frac{1}{s_\alpha(t)}\right) \tag{9.88}$$

where $s_\alpha(t) = x_{\alpha-1}(t) - x_\alpha(t)$ denotes the distance headway with between vehicles $\alpha-1$ and α. Eq. (9.88) describes how vehicle α adapts a speed $\tilde{V}$ that is a function of the distance between him and his predecessor. However, the reaction of α will not be instantaneous, but will be delayed by the reaction time τ. Let us define

$$x_\alpha(t) = x \tag{9.89}$$

and

$$v_\alpha(t) = u(x_\alpha(t), t) = u(x,t) \tag{9.90}$$

Let us first consider the left-hand-side of Eq. (9.88). We have

$$v_\alpha(t+\tau) = u(x_\alpha(t+\tau), t+\tau) \tag{9.91}$$

Since, by the theorem of Taylor, we have

$$x_\alpha(t+\tau) = x_\alpha(t) + \tau\frac{dx_\alpha}{dt} + O\left(\tau^2\right) \tag{9.92}$$

Moreover

$$\begin{aligned} u(x_\alpha(t+\tau), t+\tau) &= u\left(x_\alpha(t) + \tau\frac{dx_\alpha}{dt} + O\left(\tau^2\right), t+\tau\right) && (9.93)\\ &= u(x,t) + \tau\frac{dx_\alpha}{dt}\frac{\partial u(x,t)}{\partial x} + \tau\frac{\partial u(x,t)}{\partial t} + O\left(\tau^2\right) && (9.94)\\ &= u(x,t) + \tau u(x,t)\frac{\partial u}{\partial x} + \tau\frac{\partial u}{\partial t} + O\left(\tau^2\right) && (9.95) \end{aligned}$$

Neglecting the higher-order terms thus yields

$$v_\alpha(t+\tau) \approx u(x,t) + \tau u(x,t)\frac{\partial u}{\partial x} + \tau\frac{\partial u}{\partial t} \tag{9.96}$$

Let us now consider the right-hand-side of Eq. (9.88). First, note that the inverse of the distance headway equals the density (evaluated at the location between vehicle α and vehicle $\alpha-1$), i.e.

$$\frac{1}{s_\alpha(t)} = \frac{1}{x_{\alpha-1}(t) - x_\alpha(t)} = k\left(x_\alpha(t) + \frac{1}{2}s_\alpha(t), t\right) \tag{9.97}$$

Again, the use of Taylor's theorem yields

$$k\left(x_\alpha(t) + \frac{1}{2}\left(x_{\alpha-1}(t) - x_\alpha(t)\right), t\right) \tag{9.98}$$

$$= k\left(x_\alpha(t), t\right) + \frac{1}{2}s_\alpha(t)\frac{\partial k}{\partial x} + O\left(s_\alpha(t)^2\right) \tag{9.99}$$

$$= k(x,t) + \frac{1}{2k}\frac{\partial k}{\partial x} + O\left(\frac{1}{k^2}\right) \tag{9.100}$$

Then

$$V\left(\frac{1}{s_\alpha(t)}\right) = V\left(k\left(x_\alpha(t) + \frac{1}{2}s_\alpha(t), t\right)\right) \tag{9.101}$$

$$= V\left(k(x,t) + \frac{1}{2k}\frac{\partial k}{\partial x} + O\left(\frac{1}{k^2}\right)\right) \tag{9.102}$$

$$= V\left(k(x,t)\right) + \left(\frac{1}{2k}\frac{\partial k}{\partial x} + O\left(\frac{1}{k^2}\right)\right)\frac{dV}{dk} \tag{9.103}$$

Neglecting the higher-order terms yields

$$\tilde{V}\left(s_\alpha(t)\right) \approx V\left(k(x,t)\right) + \frac{1}{2k}\frac{dV}{dk}\frac{\partial k}{\partial x} \tag{9.104}$$

Combining Eqn. (9.96) and (9.104) yields the Payne model, describing the dynamics of the speed $u = u(x,t)$

$$u + \tau u\frac{\partial u}{\partial x} + \tau\frac{\partial u}{\partial t} = V(k) + \frac{1}{2k}\frac{dV}{dk}\frac{\partial k}{\partial x} \tag{9.105}$$

or, in the more familiar form

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \frac{V(k) - u}{\tau} + \frac{1}{2\tau}\frac{dV}{dk}\frac{1}{k}\frac{\partial k}{\partial x} \tag{9.106}$$

$$= \frac{V(k) - u}{\tau} - \frac{D(k)}{\tau}\frac{1}{k}\frac{\partial k}{\partial x} \tag{9.107}$$

$$= \frac{V(k) - u}{\tau} - c^2(k)\frac{1}{k}\frac{\partial k}{\partial x} \tag{9.108}$$

where $D(k) := -\frac{1}{2}\frac{dV}{dk} = \frac{1}{2}\left|\frac{dV}{dk}\right| > 0$, and where $c^2(k) = D(k)/\tau$. For reasons which will become clear later on, $c(k)$ is generally referred to as the wave speed. Furthermore, if many applications of the Payne model, the wave speed $c(k)$ is chosen constant, i.e. $c(k) = c_0$.

Let us briefly describe the different terms that are present in the Payne model, and how these terms can be interpreted from a traffic flow point-of-view:

- In hydrodynamics, the term $u\frac{\partial u}{\partial x}$ is referred to as the *transport* or *convection term*. It describes changes in the traffic speed u due to the inflow and outflow of vehicles with different speeds.

- The term $-\frac{D(k)}{k\tau}\frac{\partial k}{\partial x}$ is the *anticipation term*, since it can describe the reaction of drivers to downstream traffic conditions.
- The relaxation term $\frac{V(k)-u}{\tau}$ describes the exponential adaptation of the speed to the concentration-dependent equilibrium speed $V(k)$.

Payne [42] has shown that his model predicts instability of the traffic flow in a certain density range. This means that small disturbances in the flow can lead to the formation of a local traffic jam. This will be illustrated in the remainder. Note that a number of high-order models fall under the same model family.

9.4.2 Mathematical properties of the Payne-type models

This section discusses some of the important properties of the Payne-type models; the derivation of these properties from the model equations (B.1) and (B.2) is explained in the subsequent sections.

We have seen that the LWR model has a single family of characteristic curves, which are described by the lines $\frac{dx}{dt} = \lambda_*$, where $\lambda_* = \lambda_*(k) = \frac{dQ}{dk}$. The Payne model has *two families of characteristics* along which the properties of the flow are transported. These characteristic curves are defined by the following ordinary differential equations

$$C_+ : \frac{dx}{dt} = \lambda_1 = u + c(k) \tag{9.109}$$

$$C_- : \frac{dx}{dt} = \lambda_2 = u - c(k) \tag{9.110}$$

It was shown that for the LWR model, the density was constant (conserved) along the characteristic. This is not the case in the Payne model; in the Payne model, the dynamics of the so-called *Riemann* or *characteristic* variables $z_{1,2}$ are described. Assuming that $c(k) = c_0$, these are defined by

$$z_1 = u + c_0 \ln k \quad \text{and} \quad z_2 = u - c_0 \ln k \tag{9.111}$$

The characterstics C_+ and C_- are sometimes refered to as the *second-order waves*, whereas the waves C_* described by $\frac{dx}{dt} = \lambda_*$ are called the *first-order waves*.

It can be shown that solutions to the Payne-model are unstable, if they satisfy the following condition.

$$k\left|\frac{dV}{dk}\right| = -k\frac{dV}{dk} \geq c(k) \tag{9.112}$$

For small densities k, condition (9.112) *will generally not hold*, implying that the Payne model would predict stable traffic flow. In other words, small perturbations $\delta k(x,t)$ and $\delta u(x,t)$ will dissipate over time. This holds equally for regions in which $dV/dk = 0$. When the concentration and the rate of decrease dV/dk is large enough, the second-order models will predict that traffic flow becomes unstable. Under unstable conditions, small perturbations will grow over time and will eventually turn into a traffic jam.

For the Payne mode, we have

$$c(k) = \sqrt{\frac{1}{2\tau}\left|\frac{dV}{dk}\right|} \tag{9.113}$$

implying that traffic flow conditions are unstable when

$$k\left|\frac{dV}{dk}\right| \geq \frac{1}{2k\tau} \tag{9.114}$$

Consider, for instance, Greenshields' fundamental diagram, we get

$$\frac{dV}{dk} = -\frac{V_0}{k_j} \tag{9.115}$$

yielding that the Payne model would predict unstable traffic conditions when

$$k \geq \sqrt{\frac{1}{2\tau}\frac{k_j}{V_0}} \tag{9.116}$$

Note that the stability of the traffic conditions increase when the relaxation time τ of the flow decreases. That is, the more timely drivers react to the prevailing traffic conditions, the more stable the resulting traffic flow conditions are. For the Payne model, traffic flow conditions are always stable in case $\tau \to 0$. For details on the mathematical analysis, we refer to appendix B of these notes.

9.4.3 Numerical simulation and high-order models

Since the mathematical properties of higher-order models are still not well understood, reliable numerical analysis is also still a problem. In appendix B, we propose a numerical scheme for general high-order models. Another numerical approximation is implemented in the METANET model, which is discussed in detail in Chapter 15.

9.5 Alternative modelling approaches and generalisations

Appendix C discusses a specific type of continuum model, namely gas-kinetic models. Appendix D discusses generalisations of macroscopic models.

Chapter 10

Human factors

Summary of chapter. The driving task is a comprehensive term that comprises all tasks which must be executed by the driver to rearch his/her destination safely, comfortably, and timely. It is performed by execution of *vehicle control actions*. For an accurate performance of the driving taks, it can be assumed that the traffic state is continuously monitored and estimated, expectations about future states are continually being made, and control actions are initiated according to driver's objectives and preferences. The control actions affecting the vehicle's status and thus the status of the traffic system itself, by for instance braking, accelerating or decelerating by adjusting the throttle position, changing gear, change foot from one pedal to another, change course by adjusting steering wheel, apply turn indicator, etc.

In this intermezzo chapter, we discuss the exectution of driving tasks and human performance aspects in the context of person-machine control systems with the aim to provide a theoretical rationel for the microscopic models discussed in the ensuing chapters. First, we provide an classification of the different (sub)tasks of driving. After this classification has been presented, we will discuss a conceptual driver model. We will describe the discrete components of performance, largely centered around time lags which are fundamental parameters in human performance. These include perception-reaction time, control movement time, responses the the presentation of traffic control devices, responses to the movements of other vehicles, etc. Next, the kind of control performance that underlies steering, braking and speed control (primary control functions) will be described briefly.

List of symbols

$\mathbf{x}(t)$	-	state vector at time t
$\mathbf{u}(t)$	-	control vector at time t
$\mathbf{y}(t)$	-	observation vector at time t
τ	s	reaction time
$\mathbf{r}_q$	m	location of vehicle q
$\mathbf{v}_q$	m/s	velocity vector of vehicle q
RT	s	total reaction time
MT	s	movement time

10.1 The driving task

Driving requires the driver to possess many perceptual and cognitive skills for steering, changing gears, operating the pedals, prediction, fault analysis, making control decisions, etc. Driver experience will play an important role in controlling the vehicle. For example, controlling the vehicle will occupy much of the conscious attention of the novice driving, leaving little reserves for attending other tasks, such as hazard detection. More experienced drivers will execute more driving tasks automatically and subconsciously, leaving more resources for other tasks.

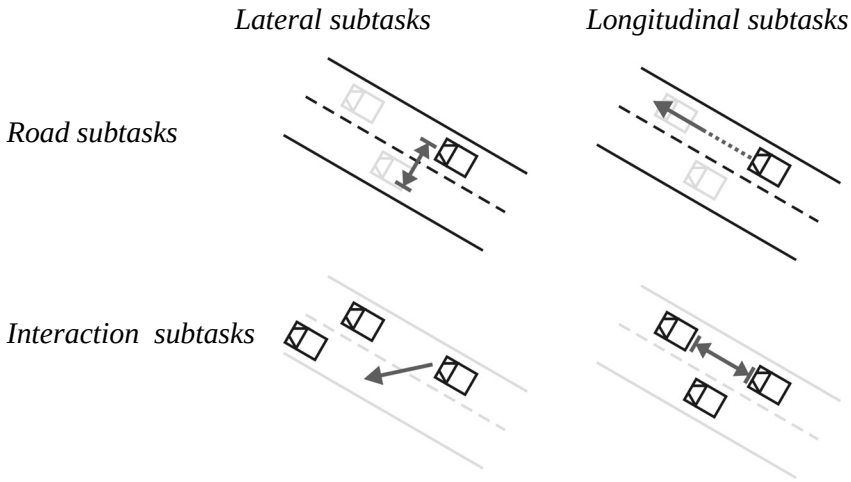

Figure 10.1: Schematisation of driver subtasks

Several researchers have considered the driving task as a hierarchical process, where three levels can be distinguished [45]:

1. Control (operational level)
2. Guidance (manoeuvre or tactical level)
3. Navigation (strategic level)

The *control* level of performance comprises all activities that involve split-second-to-second exchange of information and control inputs between the driver and the vehicle. This level of performance is at the control interface. Most control activities are performed "automatically", that is, with little concious effort. In short, the control level of performance is *skill-based*. Since the vehicle control task itself is a skill-based activity, humans are generally considered to perform quite well and errors are minimal. Accidents hence stem mainly from risk-taking behaviour, influence of alcohol, and violations of traffic rules.

Once a person has learned the rudiments of control of the vehicle, the next level is the rule-based *guidance level*. The driver's main activities "involve the maintenance of a safe speed and a proper path relative to the roadway and traffic elements". Guidance level inputs to the system are dynamic speed and path responses to roadway geometrics, hazards, traffic and the physical environment. Information presented to the driver-vehicle system is from traffic control devices, other traffic and other features of the environment, continually changing as the vehicle moves along the roadway.

The former two levels of vehicle control are of great importance to modelling for instance a roadway stretch. The third level is *navigation*, comprising for instance route planning and en route guidance. This behaviour is *knowledge based* will be of lesser importance in the ensuing of this chapter.

Minderhoud [37] proposes a more elaborate description of the driving task, explicitly distinguishing between subtasks dealing with the interaction of the driver-vehicle unit and the roadway, and the interaction between the driver-vehicle unit and other traffic participants. In doing so, the following five subtasks are identified (see Fig. 10.1):

1. The *strategic route navigation* is a subtask in which drivers choose their destination, travel direction, route and roads. The choice for a destination or route depends on specific goals and preferences of the drivers, and on the traffic conditions.

2. The *lateral roadway subtask* is defined as a collection of decisions of the driver which are needed to guide the vehicle properly and comfortably over the available infrastructure and its elements such as driving lanes, curves and on- and off-ramps, while dealing with the lateral direction. Examples of this task are: maintaining an adequate lateral position on the lanes or roadway, given the prevailing traffic rules.

3. The *longitudinal roadway subtask* pertains to the decisions needed to guide the vehicle properly and comfortably over the available infrastructure and its elements while dealing with the longitudinal direction. Examples are maintaining an adequate speed to avoid unsafe situations, e.g. in curves or with bad road conditions, given the prevailing traffic rules (e.g. speed-limits).

4. The *lateral vehicle interaction subtask* describes the collection of decisions needed to guide the driver-vehicle unit properly and comfortably around obstacles, vehicles, and possible other traffic participants, while focussing on the lateral direction. Examples are lane-change decisions and gap acceptance behaviour.

5. Finally, the *longitudinal vehicle interaction subtask* pertains to all decisions needed to guide the driver-vehicle unit around obstacles, vehicles, and possibly other traffic participants, while focussing on the longitudinal direction. Examples are speed and distance keeping with respect to the vehicles in front.

In the remainder, the above classification will be used. In particular, chapter 11 discusses car-following models (longitudinal vehicle interaction subtaks); chapter 12 discusses lane changing behaviour and gap acceptance modelling (lateral vehicle interaction subtasks).

10.2 Sharing attentional resources

Drivers frequently have to deal with several sources of information concurrently. For example, the manoeuvres of other traffic participants, traffic signals, and other local traffic conditions and traffic control measures, as well as information from inside the vehicle. The concept of limited pools of attentional resources states that drivers share their attention among the tasks that need their attention. A direct result of this theory is that the allocation of attentional resources to one task will leave fewer resources in performing the other tasks.

Under certain conditions, the driver may find the demands of multiple concurrent tasks overwhelming, and may decision not to perform certain tasks. In illustration, drivers will not frequently check their rear-view mirror when following another vehicle under bad weather conditions.

10.3 Driver objectives

It can be concluded from several studies that drivers differ in driving style. The differences can be explained by different preferences (such as desired speed, desired minimum gap distance), and driving control objectives (e.g. safe, comfortable, or efficient driving) of the drivers, given the possibilities of their vehicles and their own driving capabilities. The distribution of preferences over the driver population and car types also explains the stochastic nature of the maximum traffic volumes and other macroscopic flow characteristics.

Driving objectives are generally a combination of the following options:

1. Maximise safety and minimise risks;

2. Minimise lane-changing manoeuvres;

3. Maximise travel efficiency (restricting speed deviations with the desired driving speed);

4. Maximise smoothness and comfort;

5. Minimise stress, inconvenience, fuel consumption, accelerations, decelerations, etc.

The importance of each objective will also vary among the individuals. Experienced, trip purpose, age, driver's conditions, time-of-day, etc., are factors affecting the subjective valuation of these objectives. Vehicle characteristics, such as length, manoeuvrability, maximum speed, acceleration and deceleration capabilities, are also influencing factors, as are the prevailing weather and environmental conditions.

10.4 Modelling driving tasks

Traditionally, two modelling approaches have been followed to describe human control behaviour. The first approach is the conventional servo-system approach. The second approach is rooted in the optimal estimation and control theory, resulting in the widely used *optimal control model*. The latter approach is based on the fundamental hypothesis that the human operator, in this case the driver, behaves optimally, given his inherent limitations and constraints. The approach generally assumes that the driver first estimates the state of the system, based on his knowledge of the system dynamics and the inaccurate observations. Next, the driver applies the control strategy that optimises some performance criterion, which is usually a trade-off between the different driver objectives that have been discussed in the previous section. The approach optimal control approach has been used to provide a theoretical framework to describe driving behaviour as well as to derive mathematical models.

10.4.1 Car-driver-roadway system diagram

This section discusses the driver-vehicle system as a block diagram, first developed by Weir (1976). Fig. 10.2 shows the resulting block diagram describing the car-driver-roadway system. Let us briefly discuss the diagram. The *inputs* enter the driver-vehicle system from other vehicles, the roadway, and the driver him/herself (activing at the navigation level of performance). The main source of information for the driver is the visual field as seen through the windshield, and the dynamic changes to that field generated by the motion of the vehicle. The driver attends only to selected parts of this input. Factors such as the experience of the driver, his state of mind, and stressors (e.g. being on a crowded facility when being late for a meeting) all impringe on the supervisory or monitoring level of performance, and directly or indirectly affect the control level of performance. The actual control movements made by the driver couple with the vehicle control and the interface of throttle (T), brake (B) and steering (S). The vehicle in turn, as a dynamic physical process in its own right, is subject to inputs from the road and the environments. The resolution of control dynamics and vehicle disturbance dynamics is the vehicle path.

As will be discussed, a considerable amount of knowledge is available for some of the lower blocks in Fig. 10.2, i.e. the ones associated with braking reactions, steering inputs, and vehicle control dynamics. Before discussing some research results pertaining to perception-response time and control movement times, we present an model adopted from [37] discribing driving as a feedback-oriented control task. In fact, this conceptual model is simplification of the model discussed in this section, which can be used to derive mathematical models.

10.4.2 Optimal control model

This section discusses a conceptual model to describe driving behavior in terms of a closed feedback control system, where the driver predicts the behavior of him/herself and of other drivers (referred to as opponents in game-theoretic terms), including the presumed opponents'

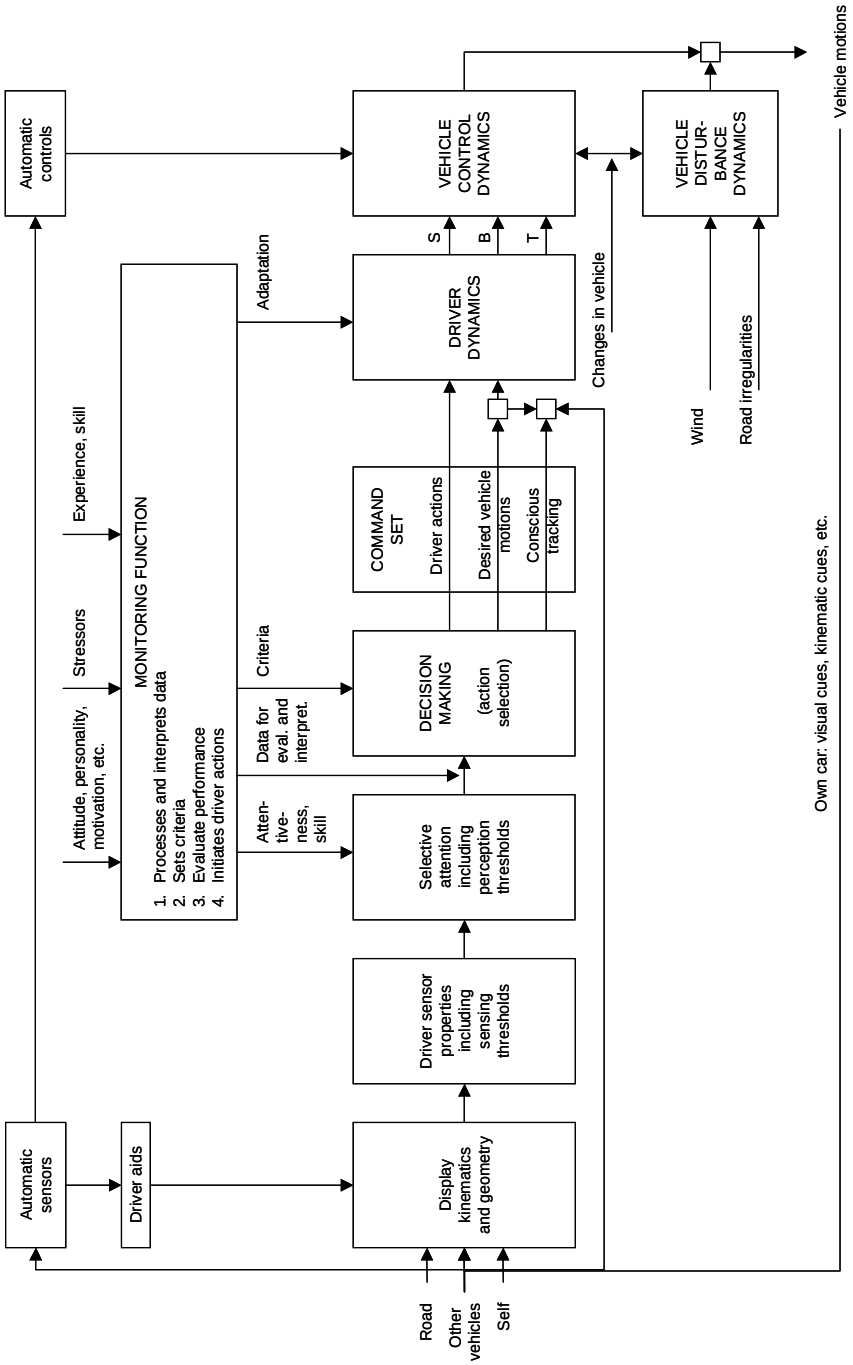

Figure 10.2: Generalised block diagram of the car-driver-roadway system.

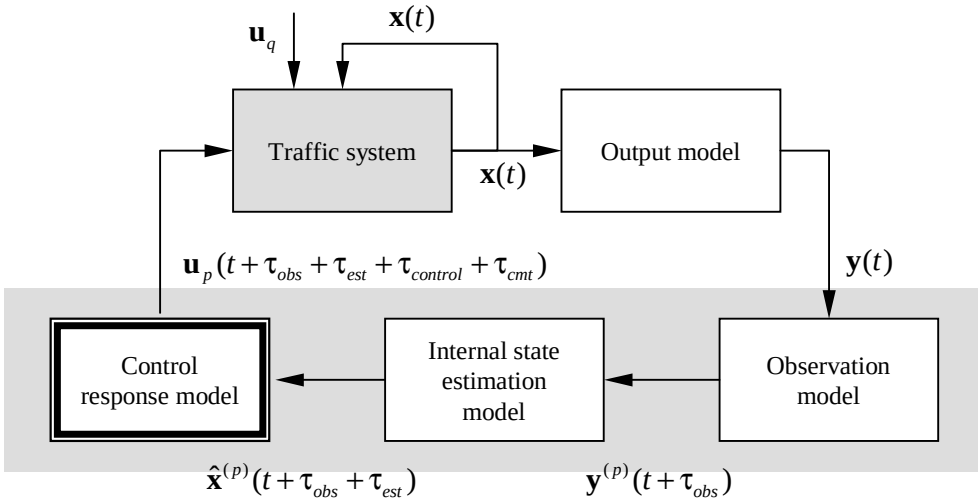

Figure 10.3: Driving as feedback-oriented control.

reactions to the control decisions of the driver. As mentioned in the introduction, these driver task comprise both control ad guidance of the vehicle.

To describe the driving task, often the concept of utility maximisation is put to use. A good example pertaining to execution of the driving task is the *Risk Homeastasis Theory* proposed by Wilde, which is based on the hypothesis that drivers seek to maintain a target level of risk despite changes in the environmental risk. If the environment becomes safer, drivers will engage in riskier behaviour and conversely. Application of the utility optimisation approach entails that drivers are assumed to optimise their control actions taking into account their individually estimated risk, comfort, and preferred safety levels.

Minderhoud [37] applies the utility optimisation principle to conceptually describe the performance of the driving task. Hoogendoorn and Bovy [29] us a similar approach to describe the walking behaviour of pedestrians, as well as to derive mathematical walker models.

The driver is assumed to use an internal model for determining an appropriate guidance control decision, thereby aiming to optimise his/hers task performance. Experience and knowledge have skilled the driver in performing these tasks, and hence the assumption of optimal behaviour is justified. The optimisation includes the limitations of the driver in terms of observation, information processing and internal state estimation, as well as processing times and reaction times. Moreover, the optimisation includes the driving strategy.

A driver is constantly monitoring his position and velocity relative to the other vehicles in the flow, as well as relative to the road. This monitoring allows the driver to perform corrective actions, implying that driving is *feedback-oriented*. A continuous feedback control system consists of a comparison between the input (positions, velocities) and the controlled output (acceleration, deceleration and direction changes).

Fig. 10.3 depicts the continuous control loop for a certain driver. The traffic system describes the *actual traffic operations*. The state $\mathbf{x}(t)$ of the system encompasses all information required to summarize the system's history, typically including the locations $\mathbf{r}_q(t)$ and the velocities $\mathbf{v}_q(t)$ of all vehicles q in the traffic flow

$$\mathbf{x}\,(t) := (\mathbf{r}, \mathbf{v}) \text{ where } \mathbf{r} := (\mathbf{r}_1, ..., \mathbf{r}_n) \text{ and } \mathbf{v} := (\mathbf{v}_1, ..., \mathbf{v}_n) \tag{10.1}$$

The location consists of a longitudinal component r_1, describing the locations of the vehicles along the roadway, and a component r_2 reflecting the lateral position. In principle, the latter is described in terms of a continuous variable. In most applications, the lateral position of vehicles only indicates the lane in which vehicles drive, and hence is discrete. It is emphasized that the state of the system will also include other issues which are relevant to describe the

future behavior of the system, for instance including driving mode, driver fatigue, awareness, etc.

The dynamics of the traffic system are described by kinematic equals described the changes in the system state over time. These are reflected by a coupled system of ordinary differential equations

$$\dot{\mathbf{x}}(t) = \mathbf{f}\left(t, \mathbf{x}(t), \mathbf{u}(t)\right) \tag{10.2}$$

where $\mathbf{u} = (\mathbf{u}_1, ..., \mathbf{u}_n)$ denotes the control vector, reflecting how the drivers influence the state $\mathbf{x}$ (e.g. by accelerating and changing directions). The state dynamics Eq. (10.2) describe the dynamics of the traffic system and its constituent entities. The output model maps the state $\mathbf{x}(t)$ onto the observable system output $\mathbf{y}(t)$.

In Fig. 10.3, the shaded block containing the three boxes describes the driving task model for a driver p, consisting of

1. the observation model
2. the state estimation model, and
3. the control response model.

The *observation model* of driver p describes which of the elements of the traffic system's output $\mathbf{y}\left(t\right)$ can be observed by p. The output $\mathbf{y}^{(p)}(t)$ is generally a function of the system's output $\mathbf{y}\left(t\right)$, state $\mathbf{x}(t)$, the observation time delay τ_{obs} (also referred to as the perception-response time), and observation errors $\boldsymbol{\varepsilon}_{obs}$. While observing, the driver will collect and process various types of information, such as

- own speed and (longitudinal) acceleration (own visuals and driving aids)
- distances to other vehicles and relative speeds with respect to other vehicles (radial motion detection, driving aids / automatic sensors)
- lane-changing states (turn indicator) of other vehicles
- traffic signs, road geometrics, etc.

The output $\mathbf{y}^{(p)}(t)$ typically describes the locations and velocities of vehicles that p can observe, but will thus also include other types of information on the state of the traffic system. Conceptually, the output can be expressed as follows

$$\mathbf{y}^{(p)}\left(t\right) = \mathbf{g}\left(\mathbf{x}\left(t - \tau_{obs}\right), \boldsymbol{\varepsilon}_{obs}\right) \tag{10.3}$$

Driver p will use his observation $\mathbf{y}^{(p)}$ as well as his experiences to estimate the state of the traffic system, which is described by the *internal state estimation model* of Fig. 10.3. Without going into details, we assume that the internal state estimate can be described by a function of the previous internal state estimate $\mathbf{x}^{(p)}\left(t\right)$, the observation $\mathbf{y}^{(p)}\left(t\right)$, and the estimation delay τ_{est} and error $\boldsymbol{\varepsilon}_{est}$. Here, $\hat{\mathbf{x}}^{(p)}\left(t\right)$ describes estimates of the locations and velocities of driver q. Note that these estimates generally contains more information than the observations $\mathbf{y}^{(p)}(t)$, for instance because driver p uses also information collected at previous time periods. A conceptual model describing the relation between the estimate at time t, the observation, and the previous state estimate is

$$\hat{\mathbf{x}}^{(p)}\left(t\right) = \mathbf{h}\left(y^{(p)}\left(t - \tau_{est}\right), \hat{\mathbf{x}}\left(t - \tau_{est}\right), \boldsymbol{\varepsilon}_{est}\right) \tag{10.4}$$

The *control response model* in Fig. 10.3 reflects the process of determining the control actions of driver p. These control actions are described by the control vector $\mathbf{u}_p(t)$ applied by p, indicating acceleration / deceleration, direction changing, applying turining indicators, etc. The control response model comprises three submodels, namely

1. prediction of the state of the system
2. making the control decision, and
3. taking the control action.

We hypothesize that the control decision depends on the internal state estimate, the driving strategy, reflecting the driving objectives, and the control delay $\tau_{control}$. The control delay $\tau_{control}$ describes the time needed to determine the control decision. On top of the control delay, an additional lag τ_{cmt} is incurred to perform the control movement (referred to control movement time; see section 10.5.2).

In determining the control actions, we assume that p will predict and optimize the driving cost $J^{(p)}$ incurred during some time interval $[t, t+T)$. The prediction of the internal state $\mathbf{x}^{(p)}$ will be described by an internal state dynamics model, i.e.

$$\dot{\mathbf{x}}^{(p)}(s) = \mathbf{f}^{(p)}\left(s, \mathbf{x}^{(p)}(s), \mathbf{u}^{(p)}(s)\right) \quad \text{s.t.} \quad x^{(p)}(t) = \hat{\mathbf{x}}^{(p)}(t) \tag{10.5}$$

for $t \leq s \leq t+T$. These state dynamics include the reactions of the drivers q as well. For any control path applied during $[t, t+T)$, p predicts the state $\mathbf{x}^{(p)}_{(t,t+H)}$ using eqn. (10.5), and can thus determine the predicted driving cost $J^{(p)} = J^{(p)}\left(u^{(p)}_{[t,t+T)} | t, \hat{\mathbf{x}}^{(p)}(t)\right)$. These driving cost will generally reflect the driver objectives. The utility optimisation concept then asserts that the driver chooses the optimal control path minimizing this cost $J^{(p)}$. The cost optimization process, its parameters, and its resulting control decisions differ between the drivers, due to different objectives, preferences, and abilities.

In the remainder of this chapter, we will briefly discuss the notions of perception-response times, and control-response times. Within the context of the model presented here, the former summarises the observation delay τ_{obs}, the state estimation delay τ_{est}, and the control delay $\tau_{control}$.

10.5 Discrete driver performance

10.5.1 Perception-response times

If compared to physical or chemical processes, the simplest human reaction to incoming information is very slow. The early 1950's say *Information Theory* take a dominant role in experimental phychology. The linear equation

$$PRT = a + bH \tag{10.6}$$

where a is the minimum reaction time for a particular type of decision, and $b \approx 0.13$ is an empirically derived slope, expresses the relationship between the number of alternatives N represented via the transmitted information content $H = \log_2 N$ that must be sorted out to decide on a response and the total reaction time RT. The reaction time is defined by the time of detection of the input and the start of the response. If the time for the response itself is also included, then the total lag is generally termed "response time". However, often the term are interchanged. Interestingly, part of the reaction time depends on the choice variables, while another part is common to all reactions. Most models that have been proposed prepresent chaines of the model described above, where different decision components are linked together. As an example, Tab. 10.1 shows empirical estimates of the different components of the Perception-Response-Time (PRT) for a chained model for braking response time.

Each of the elements in Tab. 10.1 are established from empirical data, and represents the 85th percentile for that aspect of time lag. However, since no driver is likely to produce 85th percentile values for each of the individual elements, a more likely *upper limit* for a driver's PRT is 1.50 s.

Component	Time(s)	Cumulative time (s)
Perception		
Latency	0.31	0.31
Eye movement	0.09	0.40
Fixation	0.20	1.00
Recognition	0.50	1.50
Initial brake application	1.24	2.74

Table 10.1: Hooper-McGee chaining model of perception-response time.

	surprise	expected
mean	1.31	0.54
standard dev.	0.61	0.10

Table 10.2: Mean and variance of PRT for surprise and expected situation.

It has been shown that the distribution of the PRT is non-normal, i.e. it is skewed. In most applications known from literature, the probability distrtibution of the PRT is assumed to follow the log-normal distribution

$$f\left(PRT\right) = \frac{1}{\sqrt{2\pi}\sigma^*} \exp\left[-\frac{1}{2}\left(\frac{\ln PRT - \mu^*}{\sigma^*}\right)^2\right] \tag{10.7}$$

where the two parameters that define the shape of the distribution are μ^* and σ^*, which are related to the mean and the standard deviation of the sample data of the PRT. Other studies have considered differences between the following situations: 1) the driver does not know when or even if the stimulus for braking will occur (surprise), and 2) the driver is aware that the signal to brake will occur, and the only question is when. Tab. 10.2 shows the results from this study. It is left to the reader to determine the 0.5 and 0.85 percentiles of the resulting log-normal distribution. Also, differences between older and younger drivers in different driving environments have been studied. For details, we refer to [21].

10.5.2 Control movement times

Control movement times reflect overt movements of the human body, with attendant inertia and muscle fiber latencies tgat come into plane once the efferent nervous implulses arrive from the central nervous system. A very simple relation can be used to describe the relation between the movement time MT, the range or amplitude of movement, size of the required control movement (Fitts' law)

$$MT = a + b\log_2\left(2\frac{A}{W}\right) \tag{10.8}$$

where

a = minimum response time lag, no movement
b = slope, empirically determined, different for each limb
A = amplitude of movement
W = width of control device (in direction of movement)

This law does not apply to all movements, especially those that are short. Such movements, usually not involving visual feedback, can be modelled by the following alternative to Fitts' law

$$MT = a + b\sqrt{A} \tag{10.9}$$

Application of these models to for instance modelling braking movements yielded control movement times between 0.15 s and 0.17 s.

A different study concerned the driver responses to the sudden opening of a car door. It turned out that drivers responded with a "ballistic yerk" of the steering wheel. The mean response latency was 1.5 s, and reached the half-way point of maximum displacement form the original path in about 2.5 s.

10.5.3 Response times to traffic control devices

From the standpoint of traffic flow theory and modelling, one of the major concerns is driver response or lag to changing traffic signals. Studies into the PRT's of traffic signals have indicated that the average PRT (from the traffic signal change to the response) equals 1.3 s; the 85-percentile is approximately 1.5. It appears that there is little relation between the distance from the traffic signal and the PRT.

10.5.4 Response to other vehicle dynamics

Consideration of the vehicle ahead has its basis in thresholds for detection of so-called *radial motion*, referring to the change in the apparant size of the target. When a vehicle approaches another vehicle with constant speed, the magnification turns from linear to geometric. When this occurs, the human visual system triggers a warning that an object is going to collide with the observer (or is pulling away from the observer). If the incoming information is irregular, than the human perceptual system knows that the relative speed to the object is changing. The visual angle of an observed object relates to the visible width of the object. The *time-to-go* (time-to-impact) is roughly equal to the reciprocal of this visual angle.

With respect to detection of distance changes, it has been found by Mortimer that drivers can detect a change in the distance to the vehicle in from when it is varied by approximately 12 percent: if a driver were following a car ahead at a distance of 30 m, at a change of 3.7 m the driver would become aware that the distance is decreasing or increasing, i.e. a change in the relative velocity. Mortimer concludes that "unless the relative speed between two vehicles becomes quite high, the drivers will respond to changes in their headway, or the change in the angular size of the vehicle ahead, and use that as a cue to determine the speed that they should adopt when following another vehicle".

With respect to the car-following task, let us finally state that the human visual perception of acceleration of an object is very gross and inaccurate. It is very difficult for a driver to discriminate acceleration from constant velocities unles the object is observed for a relative long period of time (10 to 15 seconds).

With respect to the vehicle alongside, it is noted that the motion detection is in this respect (peripheral vision) less acurate than in foveal vision (straight ahead), implying that drivers only respond to larger changes in the relative speeds. On top of this, stationary objects in the periphery (such as a neighbouring vehicle exactly keeping pace with the driver's vehicle) tends to dissappear for all intents and purposes *unless it moves with respect to the viewer* against a patterned background. Relative motion in the periphery also tends to look slower than the same movement as seen using fovea vision. Radial motion detection presumably would follow the same pattern as the vehicle ahead case, although no direct evidence is avaiable to justify this presumption.

10.6 Continuous driver performance

The previous sections have sketched the relevant discrete performance characteristics of the driver in a traffic stream. Driving, however, is primarily a continuous dynamic process of managing the present heading and the future path of the vehicle through the steering function. The first and second derivatives of location on the roadway in time, speed and acceleration, are also continuous control processes.

Steering can be modelled by a tracking model, with the following inputs:

1. the desired path as perceived by the driver
2. the perceived present course

The resulting differentia equations can be (and have been) calibrated using experimental data. We refer to [21] for details.

Chapter 11

Longitudinal driving taks models

Summary of chapter - The description of a traffic stream at microscopic level is about the movements of individual driver-vehicle combinations or driver-vehicle elements (DVE's). Microscopic models describe the interactions between the DVE's and sometimes are called manoeuver models.

An important aspect ot these microscopic model is how they describe the longitudinal driving task, both with respect to the road, and with respect to the other vehicles in the traffic stream. These are discussed in this chapter. Other manoeuver models concern e.g. overtaking on a road with oncoming vehicles; entering the roadway of a motorway from an on-ramp; crossing a road; weaving, etc. These will in part be discussed in the following chapter.

We have seen that the longitudinal driving task pertains both to the interaction with the roadway and the interaction with the other vehicles in the traffic stream. With respect to the former, the longitindal roadway subtask will rougly describe how vehicles will accelerate towards their free speed (or desired speed), when no other vehicles are directly in front of them. The longitudinal interaction task rougly describes how vehicles interact with other, generally slower vehicles. In particular, it describes the car-following behaviour. This section in particular describes the interaction subtask.

List of symbols

s_i	m	distance headway of vehicle i
T_r	s	reaction time
v_i	-	speed of vehicle i
τ	s	reaction time
a_i	m/s^2	acceleration of vehicle i
x_i	m	position of vehicle i
κ	$1/s$	sensitivity
u	-	mean speed

11.1 Model classification

In presenting the different modelling approaches that exist to describe the longitudinal driving task, the following approaches are considered[1]:

1. Safe-distance models
2. CA-models

[1]This list is certainly not exhaustive, but merely identifies the main modelling approaches to describe the longitudinal driving tasks.

3. Stimulus-response models
4. Psycho-spacing models
5. Optimal velocity models / optimal control models
6. Fuzzy-logic / rule-based models

Safe distance models are static and therefore do not descibe the dynamics of vehicle interactions. CA-models are discrete (in time and space) dynamic models, which provide a very coarse (but fast) description of traffic flow operations. Stimulus-response models, Pshycho-spacing models, and optimal control models provide a continuous time description of traffic flow.

Headway distribution models can also be considered microscopic models. These statistical models describe the headway distributions, and are generally based on very simple assumptions regarding the behaviour of drivers. Section 3.2 presented several distribution models as well as several applications. For a discussion on headway distribution models, the reader is referred to section

11.2 Safe-distance models

The first car-following models were developed in [44]. He stated that a good rule for following another vehicle at a safe distance S is to allow at least the length of a car between your vehicle and the vehicle ahead for every ten miles per hour of speed at which the vehicle is travelling.

$$s_i = S(v_i) = S_0 + T_r v_i \tag{11.1}$$

where S_0 is the effective length of a stopped vehicle (including additional distance in front), and T_r denotes a parameter (comparable to the reaction time). A similar approach was proposed by [20]. Both Pipes' and Forbes' theory were compared to field measurements. It was concluded that according to Pipes' theory, the minimum headways are slightly less at low and high velocities than observed in empirical data. However, considering the models' simplicity, agreement with real-life observations was amazing (cf. [43]).

Leutzbach [33] discusses a more refined model describing the spacing of constrained vehi-cles in the traffic flow. He states that the overall reaction time T_r consists of:

- perception time (time needed by the driver to recognise that there is an obstacle);
- decision time (time needed to make decision to decelerate), and;
- braking time (needed to apply the brakes).

In line with the terminology presented in chapter 10, the *perception-response time PRT* consists of te perception time and the decision time; the *control movement time* is in this case equal to the braking time.

The braking distance is defined by the distance needed by a vehicle to come to a full stop. It thus includes the reaction time of the driver, and the maximal deceleration. The latter is a function of the weight and the road surface friction μ, and eventual acceleration due to gravity g (driving up or down a hill). The total safety distance model assumes that drivers consider braking distances large enough to permit them to stop without causing a rear-end collision with the preceding vehicle if the latter come to a stop instantaneously. The safe distance headway equals

$$S(v_i) = S_0 + T_r' v_i + \frac{v_i^2}{2\mu g} \tag{11.2}$$

A similar model was proposed by [30]. Consider two successive vehicles with approximately equal braking distances. We assume that the spacing between the vehicles must suffice to avoid

a collision when the first vehicle comes to a full stop (the so-called reaction time distance model). That is, if the first vehicle stops, the second vehicle only needs the distance it covers during the overall reaction time T_r' with unreduced speed, yielding Forbes' model. Jepsen [30] proposes that the gross-distance headway S effectively occupied by vehicle i driving with velocity v_i is a function of the vehicle's length L_i, a constant minimal distance between the vehicles d_{min}, the reaction time T_r' and a speed risk factor F

$$S(v_i) = (L_i + d_{\min}) + v_i \left(T_r' + v_i F\right) \tag{11.3}$$

Experienced drivers have a fairly precise knowledge of their reaction time T_r. For novice drivers, rules of thumb apply ("stay two seconds behind the vehicle ahead", "keep a distance of half your velocity to the vehicle ahead"). From field studies, it is found that the delay of an unexpected event to a remedial action (perception-response time + control movement time; see chapter 10) is in the order of 0.6 to 1.5 seconds. The speed-risk factor F stems from the observation that experienced drivers do not only aim to prevent rear-end collisions. Rather, they also aim to minimise the potential damage or injuries of a collision, and are aware that in this respect their velocity is an important factor. This is modelled by assuming that drivers increase their time headway by some factor – the speed-risk factor – linear to v. Finally, the minimal distance headway d_{min}describes the minimal amount of spacing between motionless vehicles, observed at jam density.

Note that this occupied space equals the gross distance headway only if the following vehicle is constrained. In the remainder of this thesis, this property is used. Otherwise, the car-following distance is larger than the safe distance needed. Dijker et al. [16] discuss some empirical findings on user-class specific car-following behaviour in congested traffic flow conditions.

11.3 Stimulus-response models

Stimulus response models (ofter referred to as *car-following models*) are dynamic models that describe the reaction of drivers as a function of the changes in distance, speeds, etc., relative to the vehicle in front. Generally speaking, these models are applicable to relatively busy traffic flows, where the overtaking possibilities are small and drivers are obliged to follow the vehicle in front of them.

One can considerer as an example the cars on the left (or fast) lane of a motorway. They do not want to maintain a distance headway that is so large, that it invites other drivers to enter it. At the same time, most drivers are inclined to keep a safe distance with respect to their leader. As a consequence, the drivers must find a compromise between safety and the encouragement of lane changes.

On a two-lane road (for two directions) drivers that can not make an overtaking, due to the presence of oncoming vehicles or a lack of sight distance, are obliged to follow the vehicle in front. In that case the intruding of other vehicles in the gap in front is less frequent but nevertheless it appears that drivers maintain a rather short distance headway. Also in dense traffic on urban roads drivers are often obliged to follow the vehicle in front.

As traffic grows faster than the road network is expanding, drivers will more and more be engaged in car-following. Especially as capacity is reached, nearly all driver-vehicle units are in a car-following state.

How does a driver carry out his car-following task? He/she must keep a sufficient large distance headway and respond, or at least be prepared to respond, to speed changes (especially speed reductions) of the vehicle in front. This boils down to the fact that a driver must maintain a certain minimum distance headway that will be dependent on the speed. In the sequel we will assume that this is the case and discuss the consequences for dynamic situations.

Which variable can a driver control? At first sight that is the speed, but at second thought he/she operates the gas pedal and the break pedal, in other words the acceleration is controlled.

Measured value	T_r (s)	κ $(1/s)$
minimum	1.00	0.17
average	1.55	0.37
maximum	2.20	0.74

Table 11.1: Parameter estimates for stimulus-response model

Measured value	T_r (s)	κ_0 $(1/s)$
minimum	1.5	40.3
average	1.4	26.8
maximum	1.2	29.8

Table 11.2: Parameter estimates for stimulus-response model

It is evident that drivers have a certain response time, they can not respond immediately but have to go through the cycle of 'observation, processing, deciding, action'. Taking this into account the following *linear car-following model* has been proposed:

$$a_i\left(t+T_r\right)=\gamma\left(x_{i-1}\left(t\right)-x_i\left(t\right)-s_0\right) \tag{11.4}$$

Eq. (11.4) is an example of a so-called stimulus-response model. These stimulus response models assume that drivers control their acceleration, given some response time T_r. This finite response time stems from the observation time (perception and information collection), processing time and determining the control action (decision making), and applying the action (operating the gas-pedal, braking, shifting). In general terms, the stimulus response model is given by the following equation

$$response\left(t+T_r\right)=sensitivity\times stimulus \tag{11.5}$$

The response typically equals the vehicle acceleration. Various definitions for the stimulus, other than the stimulus in Eq. (11.4), have been put forward in the past. As an example, a well known model is the model of [12], using the relative speed $v_{i-1}\left(t\right)-v_i\left(t\right)$ for the stimulus, i.e.

$$\frac{d}{dt}v_i\left(t+T_r\right)=\kappa\left(v_{i-1}\left(t\right)-v_i\left(t\right)\right) \tag{11.6}$$

In eqn. (11.6), κ denotes the sensitivity. Field experiments were conducted to quantify the parameter values for the reaction time T_r and the sensitivity κ. The experiment consisted of two vehicles with a cable on a pulley attached between them. The leading driver was instructed to follow a pre-specified speed pattern, of which the following driver was unaware. Tab. 11.1 shows the results of these experiments.

It turned out that the sensitivity depended mainly on the distances between the vehicles: when the vehicles were close together, the sensitivity was high. When the vehicles were far apart, the sensitivity was small. Hence, the following specification for the sensitivity κ was proposed

$$\kappa=\frac{\kappa_0}{x_{i-1}\left(t\right)-x_i\left(t\right)} \tag{11.7}$$

Tab. 11.2 depicts the resulting parameter estimates.For this simple model, the steady-state will occur when the speed of the preceding vehicle $i-1$ is constant, and the acceleration of i is zero.

11.3.1 Model stability

One of the main points in any control model is its stability, i.e. will small disturbances damp out or will they be amplified. Whether the control model is stable will depend on the charac-ter and the parameters of the control model (11.6). For car-following models, two types of stability can be distinguished:

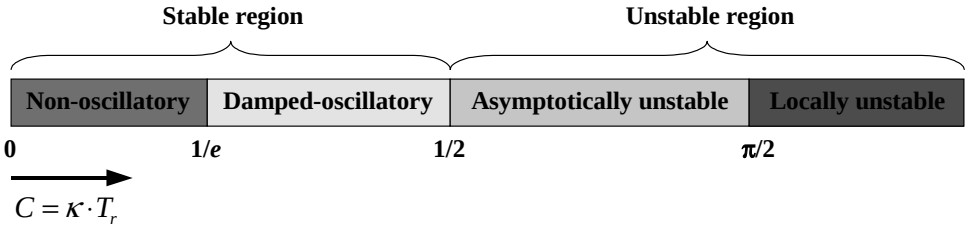

Figure 11.1: Relation between $C = \kappa T_r$ and stability

1. *Local stability*, concerning only the response of a driver on the leading vehicle $i-1$.
2. *Asymptotic stability*, concerning the propagation of a disturbance along a platoon of vehicles that are (car-) following each other.

The second type of stability is of much more practical importance than the local stability. If a platoon of vehicles in asymptotically unstable, a small disturbance of the first vehicle is amplified as it is passed over to the next vehicle, which in turn can lead to dangerous situations. Fig. 11.1 shows the relation between the local and asymptotic stability of the car-following model, and the parameter $C = \kappa T_r$ Let us briefly consider both.

Application of Laplace transforms to local stability analysis

There are various approaches to investigate stability. The most straightforward approach is to apply techniques from conventional system's control theory. The local and asymptotic stability of the model depends on the sensitivity κ and the reaction time T_r. It can be proven that the model is *locally stable* if

$$C = \kappa T_r \leq \frac{\pi}{2} \tag{11.8}$$

Proof. Let us consider the simple car-following model eqn. (11.6) and apply the Laplace transform to it. Recall that the Laplace transform is defined as follows

$$F(s) = \mathcal{L}(f(t)) = \int_0^\infty e^{-st} f(t)\, dt \tag{11.9}$$

Let us also recall the following important properties of the Laplace tranform

$$\mathcal{L}(f(t-t_0)) = e^{-t_0 s}\mathcal{L}(f(t)) \tag{11.10}$$

$$\mathcal{L}\left(\frac{d}{dt}f(t)\right) = s\mathcal{L}(f(t)) \tag{11.11}$$

Let $V_i(s) = \mathcal{L}(v_i(t))$. In applying the Laplace transformation to both the left-hand-side and the right-hand-side of Eq. (11.6), we get

$$se^{T_r s}V_i(s) = \kappa\left(V_{i-1}(s) - V_i(s)\right) \tag{11.12}$$

This expression enables us to rewrite the Laplace transform of the speed $V_i(s)$ of vehicle i as a function of the Laplace transform $V_{i-1}(s)$ of the leading vehicle i, i.e.

$$V_i(s) = H(s)\, V_{i-1}(s) = \frac{\kappa e^{-T_r s}}{s + \kappa e^{-T_r s}} V_i(s) \tag{11.13}$$

In control theory, the function $H(s) = \frac{\kappa e^{-T_r s}}{s+\kappa e^{-T_r s}}$ is often referred to as the unit-response (since it describes how the system would respond to a unitary signal ($V_{i-1}(s) = 1$)) or *transfer function.*

The transfer function describes how the output $V_i(s)$of the control system relates to the input function $V_{i-1}(s)$.

The stability of the control system can be studied by studying the properties of the transfer function $H(s)$. To ensure stability of the system the so-called *poles* of the transfer function $H(s)$ must be on the left hand side of the imaginary plane. The poles are defined by the points where the denominator $D\,(s)$ of $H(s)$ are equal to zero, i.e.

$$D\,(s) = s + \kappa e^{-T_r s} = 0 \tag{11.14}$$

To study stability, let $s = x + jy$, where $j^2 = -1$ defines the unit imaginary number. The denominator $D(s)$ then equals

$$D\,(s) = x + jy + \kappa e^{-T_r(x+jy)} = x + \kappa e^{-T_r}\cos\,(T_r y) + j\left(y - \kappa e^{-T_r x}\sin\,(T_r y)\right) \tag{11.15}$$

where we have used that $e^{-jT_r y} = \cos\,(T_r y) - j\sin\,(T_r y)$. To determine the boundary of local instability, consider the limiting case $x = 0$. Then, the necessary condition for the pole is given by the following expression

$$\kappa\cos\,(T_r y) + j\,(y - \kappa\sin\,(T_r y)) = 0 \tag{11.16}$$

Since both the real part of eqn. (11.16) and the imaginary part of eqn. (11.16) must be equal to zero, we have for the real part

$$\kappa\cos\,(T_r y) = 0 \rightarrow y_k = \frac{1}{T_r}\left(\frac{\pi}{2} + k\pi\right) \tag{11.17}$$

For the imaginary part, we then have

$$\begin{aligned} y_k - \kappa\sin\,(T_r y_k) &= \frac{1}{T_r}\left(\frac{\pi}{2} + k\pi\right) - \kappa\left(\frac{\pi}{2} + k\pi\right) && (11.18)\\ &= \frac{1}{T_r}\left(\frac{\pi}{2} + k\pi\right) - \kappa\,(-1)^k = 0 && (11.19)\end{aligned}$$

implying

$$\left(\frac{\pi}{2} + k\pi\right) = \kappa T_r\,(-1)^k \tag{11.20}$$

Consider $k = 0$. Eqn. (11.20) will only have a solution when

$$\kappa T_r = \frac{\pi}{2} \tag{11.21}$$

This implies that eqn. (11.21) defines the boundary between stable and unstable parameter values. It can also be shown that for

$$\kappa T_r > \frac{\pi}{2} \tag{11.22}$$

the model is both locally and (thus also) asymptotically unstable. ■

In illustration, Fig. 11.2 shows the behaviour of the car-following model in case of local unstable parameter settings. Clearly, the amplitude of the oscillations grows over time, eventu-ally leading to a collision of the lead car and the following car.

The local stability region is further divided into two regions: non-oscilatory ($C < 1/e$) and damped oscillatory ($C > 1/e$). In the latter case, the response of the following vehicle oscillates, but these oscillations damp out over time.

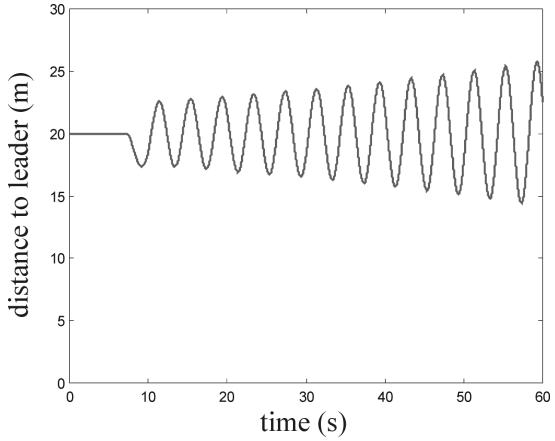

Figure 11.2: Gap between vehicle $i-1$ and vehicle i, assuming that leader $i-1$ brakes suddenly at $t = 10s$ for $T_r = 1s$ and $\kappa = 1.6s^{-1}$

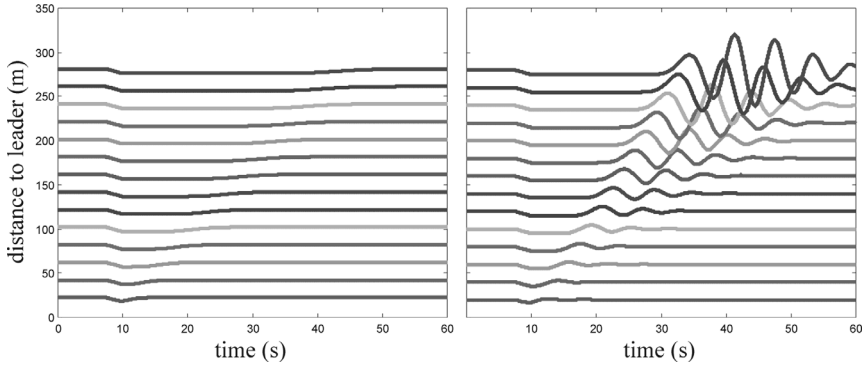

Figure 11.3: Position of vehicles with respect to platoon leader, when leader brakes suddenly at $t = 10s$ in case: $T_r = 1.00s$ and $\kappa = 0.4s^{-1}$ (left); $T_r = 1.00s$ and $\kappa = 0.7s^{-1}$ (right)

Asymptotic stability

A platoon of vehicles is asymptotically unstable if

$$C = \kappa T_r > \frac{1}{2} \tag{11.23}$$

Fig. 11.3 (left) shows a simulation example of a platoon of 15 vehicles, of which the leader brakes abruptly at $t = 10$ s. In this case, the parameters yield a locally and asymptotically stable system. Fig. 11.3 (right) shows the same case, but for parameter settings yielding a locally stable but asymptotically unstable system. See how the amplitude of the disturbance grows when it passes from one vehicle to the next.

Note that the local stability is less critical than the asymptotic stability. If the controller is locally stable, it can still be asymptotically unstable. Furthermore, observe that eqn. (11.23) shows that the car-following model becomes unstable for large response times as well as large sensitivity values. Reconsidering the empirical values of the response time T_r and the sensitivity κ shows that the average values satisfy the local stability condition. However, the conditions for asymptotic stability are not met. Moreover, the minimum values are neither asymptotically, nor locally stable.

Proof. The study of asymptotic stability of the car-following model is based on the analysis of periodic functions of the form $A_i e^{j\omega}$. In this respect, let us first note that we can approximate any periodic function $f(t)$ of time by a series of such signals, i.e. we can write $f(t) = \sum_k a_k e^{j\omega_k t}$.

This thus also holds for the behaviour of the leading vehicle $i-1$, i.e.

$$v_{i-1}(t) = \sum_k a_k^{(i-1)} e^{j\omega_k t} \tag{11.24}$$

(Fourier series). Here, $\omega_k = \frac{2\pi k}{T}$ denotes the frequency of the signal, T denotes the period of the function $f(t)$ (defined by $f(t) = f(t+T)$), and a_k denotes the amplitude of the contribution $e^{j\omega_k t}$. To study asymptotic stability, we consider one of these terms $a_k^{(i-1)} e^{j\omega_k t}$. The speed of the following $v_i(t)$ is the solution of a delayed differential equation (11.6). This solution can be written in the following form

$$v_i(t) = \sum_k a_k^{(i)} e^{j\omega_k t} \tag{11.25}$$

Let us now substitute If we now substitute Eq. (11.25) into Eq. (11.6):

$$\begin{aligned} \frac{d}{dt} v_i(t+T_r) &= \kappa\left(v_{i-1}(t) - v_i(t)\right) && (11.26)\\ &\Leftrightarrow & \\ \sum_k j\omega_k a_k^{(i)} e^{j\omega_k(t+T_r)} &= \kappa\left(\sum_k \left(a_k^{(i-1)} - a_k^{(i)}\right) e^{j\omega_k t}\right) && (11.27) \end{aligned}$$

If this equation holds for all t, then Eq. (11.26) implicates

$$j\omega_k a_k^{(i)} e^{j\omega_k(t+T_r)} = \kappa\left(\left(a_k^{(i-1)} - a_k^{(i)}\right) e^{j\omega_k t}\right) \text{ for all } k \geq 0 \tag{11.28}$$

Let us now determine what happens to the amplitude $a_k^{(i)}$ as it passes from vehicle $i-1$ to vehicle i. By rewriting Eq. (11.28), we can easily show that

$$\frac{a_k^{(i)}}{a_k^{(i-1)}} = \frac{\kappa}{j\omega_k e^{j\omega_k T_r} + \kappa} \tag{11.29}$$

To ensure asymptotic stability, it is required that the amplitude of the speed profile *decreases* as it passes from vehicle $i-1$ to vehicle i (and from vehicle i to vehicle $i+1$, etc.), implying that

$$\left|\frac{a_k^{(i)}}{a_k^{(i-1)}}\right| = \left|\frac{\kappa}{j\omega_k e^{j\omega_k T_r} + \kappa}\right| < 1 \Leftrightarrow \kappa < \left|j\omega_k e^{j\omega_k T_r} + \kappa\right| \tag{11.30}$$

Rewriting

$$\begin{aligned} j\omega_k e^{j\omega_k T_r} + \kappa &= j\omega_k\left(\cos(\omega_k T_r) + j\sin(\omega_k T_r)\right) && (11.31)\\ &= j\omega_k \cos(\omega_k T_r) - \omega_k \sin(\omega_k T_r) && (11.32) \end{aligned}$$

and substituting the result in eqn. (11.30), we get

$$\begin{aligned} \left|j\omega_k e^{j\omega_k T_r} + \kappa\right| &= \left|j\omega_k \cos(\omega_k T_r) - \omega_k \sin(\omega_k T_r)\right| && (11.33)\\ &= \sqrt{\left(\kappa - \omega_k \sin(\omega_k T_r)\right)^2 + \omega^2 \cos^2(\omega_k T_r)} && (11.34)\\ &= \sqrt{\kappa^2 + \omega_k^2 - 2\kappa\omega_k \sin(\omega_k T_r)} && (11.35) \end{aligned}$$

We can show that this the condition becomes critical for very small ω_k. Hence, we may use the approximation

$$\sin(\omega_k T_r) \approx \omega_k T_r \tag{11.36}$$

yielding

$$\omega_k^2 > 2\kappa\omega_k^2 T_r \rightarrow \kappa T_r < \frac{1}{2} \tag{11.37}$$

Situation	Reponse time (s/100)
Cue free, i.e. left to driver	50
Driver has to concentrate on brake lights vehicle in front	63
Cue free; first car is only one with functioning brake lights	101

Table 11.3: Effect of 'cue' (stimulus) on response timel

Asymptotic stability thus requires $C = \kappa T_r < \frac{1}{2}$. ■

On the test track of General Motors around 1960 much research has been carried out concerning the car-following model has been investigated. It was found that the response time T_r varied between 1.0 to 2.2 s and the sensitivity κ_0 from 0.2 to 0.8 s^{-1}. It appeared that often platoons were nearly asymptotic unstable. Two examples of experiments:

- Effect on response time

The situation concerns a platoon of 11 cars that follow each other closely. The first car brakes strongly at an arbitrary moment.

The conclusion is that a driver uses not only the brake lights to anticipate a needed speed reduction ($0.63 > 0.50$) but that brake lights as such are useful ($1.01 > 0.50$).

- Effect of extra information at rear side of car

The brake lights were modified as follows:

- acceleration $\geq 0 \rightarrow$ blue;
- no break and no gas (coasting) $\rightarrow$ yellow;
- braking $\rightarrow$ red.

The effect of this modification, that offers more information, was: response time smaller; sensitivity larger; product of both hardly changed, which means the same stability; a smaller distance headway at the same speed ($\rightarrow$ smaller time headway too). An interpretation of this is a gain in capacity but not in safety.

This simple model has several undesirable and unrealistic properties. For one, vehicles tend to get dragged along when the vehicle in front is moving at a higher speed. Furthermore, when the distance $s_i(t)$ is very large, the speeds can become unrealistically high. To remedy this deficiency, the sensitivity κ can be defined as a decreasing function of the distance. In more general terms, the sensitivity can be defined by the following relation

$$\kappa = \kappa_0 \frac{v_i\,(t + T_r)^m}{[x_{i-1}\,(t) - x_i\,(t)]^l} \tag{11.38}$$

where κ_0 and T_r denote positive constants. Equation (11.38) implies that, the following vehicle adjusts its velocity $v_i(t)$ proportionally to both distances and speed differences with delay T_r. The extent to which this occurs depends on the values of κ_0, l and m. In combining eqns. (11.6) and (11.38), and integrating the result, relations between the velocity $v_i\,(t + T_r)$ and the distance headway $x_{i-1}(t) - x_i(t)$ can be determined. Assuming stationary traffic conditions, the following relation between the equilibrium velocity $u\,(k)$ and the density k results

$$u\,(k) = u_f \left(1 - \left(\frac{k}{k_j}\right)^{l-1}\right)^{\frac{1}{1-m}} \tag{11.39}$$

for $m \neq 1$ and $l \neq 1$.

Proof. In combining eqns. (11.6) and (11.38), we get

$$\frac{d}{dt}v_i\left(t+T_r\right)=\kappa_0\frac{v_i\left(t+T_r\right)^m}{\left[x_{i-1}\left(t\right)-x_i\left(t\right)\right]^l}\left(v_{i-1}\left(t\right)-v_i\left(t\right)\right) \tag{11.40}$$

which can be, by re-arranging the different terms, and noticing that $v_i\left(t\right)=\frac{d}{dt}x_i\left(t\right)$, written as follows

$$\frac{\frac{d}{dt}v_i\left(t+T_r\right)}{v_i\left(t+T_r\right)^m}=\kappa_0\frac{\frac{d}{dt}\left(x_{i-1}\left(t\right)-x_i\left(t\right)\right)}{\left[x_{i-1}\left(t\right)-x_i\left(t\right)\right]^l} \tag{11.41}$$

Assuming $m\neq 1$ and $l\neq 1$, we have

$$\frac{1}{m-1}\frac{d}{dt}\left(\frac{1}{v_i\left(t+T_r\right)^{m-1}}\right)=\frac{\frac{d}{dt}v_i\left(t+T_r\right)}{v_i\left(t+T_r\right)^m} \tag{11.42}$$

and

$$\frac{\kappa_0}{l-1}\frac{d}{dt}\left(\frac{1}{\left[x_{i-1}\left(t\right)-x_i\left(t\right)\right]^{l-1}}\right)=\kappa_0\frac{\frac{d}{dt}\left(x_{i-1}\left(t\right)-x_i\left(t\right)\right)}{\left[x_{i-1}\left(t\right)-x_i\left(t\right)\right]^l} \tag{11.43}$$

yielding

$$v_i\left(t+T_r\right)^{1-m}=C+\kappa_0\frac{1-m}{1-l}\left[x_{i-1}\left(t\right)-x_i\left(t\right)\right]^{1-l} \tag{11.44}$$

i.e.

$$v_i\left(t+T_r\right)=\left(C+\kappa_0\frac{1-m}{1-l}\left[x_{i-1}\left(t\right)-x_i\left(t\right)\right]^{1-l}\right)^{\frac{1}{1-m}} \tag{11.45}$$

where C is an integration constant. Under stationary conditions, both $s_i\left(t\right)=x_{i-1}\left(t\right)-x_i\left(t\right)$ and $v_i\left(t\right)$ will be time-independent. Furthermore, under stationary conditions, the speeds and the distances headways of all vehicles i must be equal to u and $s=1/k$. We then have

$$u=u\left(k\right)=\left(C+\kappa_0\frac{1-m}{1-l}k^{l-1}\right)^{\frac{1}{1-m}} \tag{11.46}$$

For $k=0$, we have $u\left(0\right)=u_f$ (mean free speed), and thus

$$C^{m-1}=u_f \tag{11.47}$$

For $k=k_j$, we have $u\left(k_j\right)=0$ (speed equals zero under jam-density conditions), and thus

$$\frac{\kappa_0}{C}\frac{m-1}{l-1}k_j^{l-1}=-1\rightarrow k_j=\left(\frac{C}{\kappa_0}\frac{l-1}{1-m}\right)^{\frac{1}{l-1}} \tag{11.48}$$

In sum, we have determined the following relation between the mean speed u and the density

$$u\left(k\right)=u_0\left(1-\left(\frac{k}{k_j}\right)^{l-1}\right)^{\frac{1}{1-m}} \tag{11.49}$$

■

Car-following models have been mainly applied to single lane traffic (e.g. tunnels, cf. [40]) and traffic stability analysis ([25], [36]; chapter 6). That is, using car-following models the limits of local and asymptotic stability of the stream can be analysed.

The discovery of the possibility of asymptotic instability in traffic streams was considered to explain the occurring of congestion without a clear cause, 'Stau aus dem Nichts' (*phantom jams*); Fig. 11.4 (from [53]) shows a famous measurement from those years in which the first generation of car-following models was developed. The vehicle trajectories have been determined from aerial photos, taken from a helicopter, and exhibit a shock wave for which there seems to be no reason. It can also be noted that the shock wave more or less fades out, al least becomes less severe at the right hand side of the plot. From the plot can be deduced a shock wave speed of approximately -20 km/h.

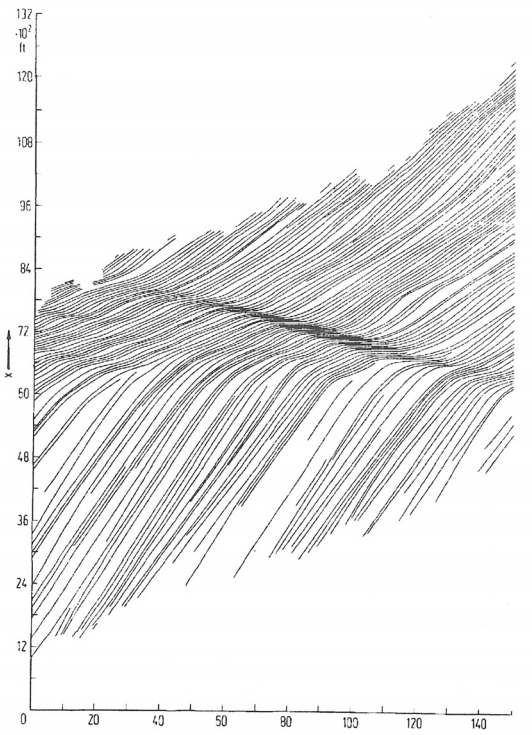

Figure 11.4: Traffic stream with shock wave at lane of motorway

11.4 Psycho-spacing models / action point models

The car-following models discussed so far have a rather mechanistic character. The only human element in the models so far is the presence of a finite response time, but for the rest a more or less perfect driver is assumed. In reality a driver is not able to:

1. observe a stimulus lower than a given values (perception threshold);
2. evaluate a situation and determine the required response precisely, for instance due to observation errors resulting from radial motion observation;
3. manipulate the gas and brake pedal precisely;

and does not want to be permanently occupied with the car-following task.

This type of considerations has inspired a different class of car-following models, see e.g. [33]. In these models the car-following behaviour is described in a plane with relative speed and headway distance as axis. The model is illustrated with Fig. 11.5. It is assumed that the vehicle in front has a constant speed and that the potential car-following driver catches up with a constant relative speed v'_r. As long as the headway distance is larger than s_g, there is no response.

Moreover, if the absolute value of the relative speed is smaller than a boundary value v_{rg}, then there is also no response because the driver can not perceive the relative speed. The boundary value is not a constant but depends on the relative speed. If the vehicle crosses the boundary, it responds with a constant positive or negative acceleration. This happens in Fig. 11.5 first at point A, then at point C, then point B, etc.

Leutzbach [33] has introduced the term 'pendeling' (the pendulum of a clock) for the fact that the distance headway varies around a constant value, even if the vehicle in front has a constant speed. In this action-point model the size of the acceleration is arbitrary in the first instance, whereas it was the main point of the earlier discussed car-following models.

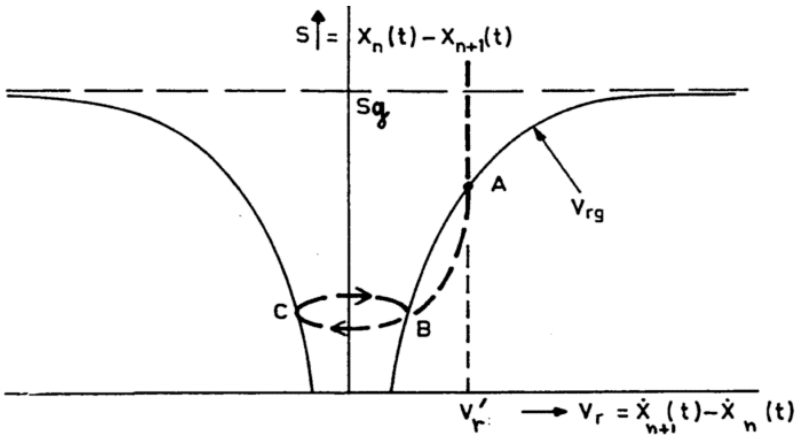

Figure 11.5: Basic action-point car-folllowing model

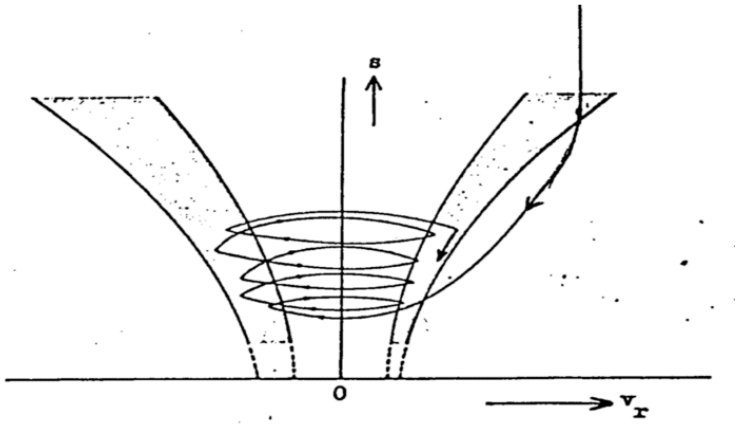

Figure 11.6: Action-point model with variable perception thresholds and a response time

The first version of the basic action point model has been extended with the following

1. a separate threshold for catching up (= approaching from a large distance);
2. different perception thresholds for negative and positive v_r;
3. a zone in stead of a function for the perception threshold;
4. an extra response if the vehicle in front shows its brake lights;

These extensions are illustrated in Fig. 11.6, and Fig. 11.7 where the following points are indicated:

- d_1: distance at speed 0
- d_2: minimum desired distance at small v_r
- d_3: maximum s at pendeling
- d_4: threshold if catching up
- d_5: threshold at pendeling

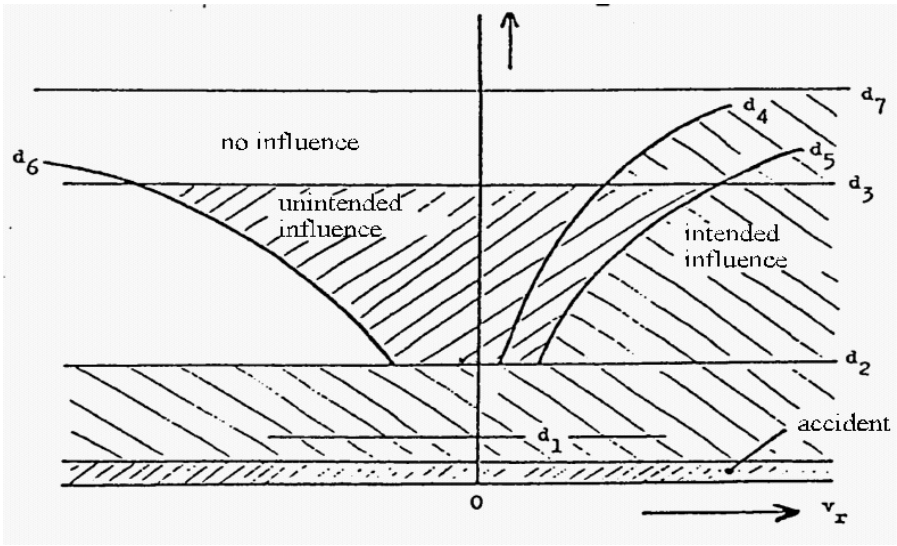

Figure 11.7: Thresholds in the $v_r - s$ plane used in a simulation model

- d_6: threshold at pendeling if $v_r < 0$
- d_7: no response at al, if $v_r > 0$.

The action point models form the basis for a large number of contemporary microscopic traffic flow models, such as FOSIM, AIMSUN2 and VISSIM.

11.5 Optimal control models

The conceptual model discussed in chapter 10 can also be applied to establish a mathematical model. Since we are only considering the longitudinal driving task, the state $\mathbf{x}(t)$ of the system can be described by the locations $r_i(t)$[2] and the speeds $v_i(t)$ of the drivers i,i.e.

$$\mathbf{x} = (r_1, ..., r_n, v_1, ..., v_n) \tag{11.50}$$

where n is the number of vehicles in the traffic system. Let us consider driver i. The prediction model used by i to predict the conditions of the traffic system are very simple

$$\frac{d}{dt} r_j = v_j \quad \text{and} \quad \frac{d}{dt} v_j = a_j \quad \text{for} \quad j = 1, ..., n \tag{11.51}$$

where a_j denotes the acceleration of vehicle j. In fact, the accelerations can be considered the controls u of the traffic system (i.e. accelerating and braking). Note that driver i can only directly influence the acceleration a_i, i.e. $u = a_i$; in principle, he/she can *indirectly* influence the control behaviour of the other drivers $j \neq i$. For the sake of simplicity, we assume that driver i predicts that the other drivers will not accelerate/decelerate during the considered time period, i.e. $a_j = 0$ for $j \neq i$.

Having specified the prediction model Eq. (10.5), we need to specifying the driving cost function $J^{(i)}$ incurred by driver i. We assume that the driving cost are of the following form

$$J^{(i)}\left(u_{[t,\infty)} | t, \hat{x}(t)\right) = \int_t^\infty e^{-\eta s} L(\mathbf{x}(s), u(s))\, ds \tag{11.52}$$

subject to the prediction model

$$\dot{\mathbf{x}}(s) = \mathbf{f}(\mathbf{x}(s), u(s)) \quad \text{for } s > t \quad \text{with} \quad \mathbf{x}(t) = \hat{\mathbf{x}}(t) \tag{11.53}$$

[2] We have used the notation $r_i(t)$ to describe the location of driver i to avoid confusion with the state $\mathbf{x}(t)$.

where $\eta > 0$ is the so-called *temporal discount factor*, describing that drivers will discount the running cost $L(x, u)$ over time (i.e. the cost incurred in the near future are more important than the cost incurred in the far future).

The running cost reflect the cost that driver i will incurr during a very short interval $[t, t+dt)$. The total cost $J^{(i)}$ are thus determined by integrating the running cost L with respect to time. In this section, we will assume that the running cost will consist of three different factors, namely

1. running cost incurred due to not driver at the free speed
2. running cost incurred when driving too close to the vehicle $i-1$ directly in front
3. running cost incurred due to accelerating/braking

Let $v_f^{(i)}$ denote the free speed of driver i. The running cost due to not driving at the free speed is modelled as follows

$$\frac{c_1}{2}\left(v_f^{(i)} - v\right)^2 \tag{11.54}$$

where $c_1 > 0$ is a parameter expressing the relative importance of this cost component.

Similarly, the running cost incurred due to driving too close to the vehicle $i-1$ directly in front (so-called proximity cost) equals

$$c_2\Phi\left(r_{i-1} - r_i, v_{i-1} - v_i\right) = c_2 e^{-(r_{i-1}-r_i)/S_0} \tag{11.55}$$

where $c_2 > 0$ again denotes the relative importance of this cost component, and where S_0 is some scaling parameter. Eq. (11.55) shows that as the distance between vehicle $i-1$ and vehicle i increases, the running cost decreases and vice versa.

Finally, we propose using the following expression to describe the running cost incurred due to acceleration/deceleration

$$\frac{c_3}{2}u^2 \tag{11.56}$$

where $c_3 > 0$ is the relative importance of this cost component.

Now, all the components are in place to derive the model. To mathematicall derive the model, we define the so-called *Hamilton function* H as follows

$$H = e^{-\eta t}L + \boldsymbol{\lambda}'\mathbf{f} \tag{11.57}$$

The vector $\boldsymbol{\lambda}$ denotes the so-called shadow-costs. The shadow cost describe the marginal changes in the total cost $J^{(i)}$ due to small changes in the state $\mathbf{x}$. We can show that for the control u to be optimal, it has to satisfy the so-called optimality condition

$$\frac{\partial H}{\partial u} = 0 \rightarrow u^* = -\frac{1}{c_3}e^{\eta t}\lambda_{v_1}(t, \mathbf{x}) \tag{11.58}$$

where $\lambda_{v_i}(t, \mathbf{x})$ denotes the marginal cost of the speed v_i of driver i. Eq. (11.58) shows that when the marginal cost $\lambda_{v_i}(t, \mathbf{x})$ of the speed v_i is positive, the control is negative, meaning that the driver will decelerate ($u^* < 0$). On the contrary, when the marginal cost of the speed is negative, it makes sense to further increase the speed and thus to acceletate ($u^* > 0$). When the marginal cost $\lambda_{v_i}(t, \mathbf{x}) = 0$, driver i will keep driving at constant speed.

The remaining problem is to determine the shadow cost. From optimal control theory, it is well known that the marginal cost satisfy

$$\dot{\boldsymbol{\lambda}} = -\frac{\partial H}{\partial \mathbf{x}} \tag{11.59}$$

The fact that we are considering a time-independent discounted cost problem implies furhtermore that $\dot{\boldsymbol{\lambda}} = -\eta\boldsymbol{\lambda}$. Let us first consider the marginal cost of the location r_i. We can easily determine that

$$\eta\lambda_{r_i} = \frac{\partial H}{\partial r_i} = e^{-\eta t}\frac{\partial L}{\partial r_i} = e^{-\eta t}\frac{\partial \Phi}{\partial r_i} \tag{11.60}$$

This equation shows that the marginal cost of the location of driver i only depends on the proximity cost function. Note that generally, $\frac{\partial \Phi}{\partial r_i} > 0$, that is, the proximity cost will increase with increasing r_i. For $\Phi = c_2 e^{-(r_{i-1}-r_i)/S_0}$, we would have

$$\frac{\partial \Phi}{\partial r_i} = \frac{c_2}{S_0}e^{-(r_{i-1}-r_i)/S_0} \tag{11.61}$$

We can then easily determine the marginal cost of the speed v_i of driver i. We find

$$\begin{aligned}
\eta\lambda_{v_i} &= \frac{\partial H}{\partial v_i} = e^{-\eta t}\frac{\partial L}{\partial v_i} + \lambda_{r_i} && (11.62)\\
&= e^{-\eta t}\left(c_1\left(v_f^{(i)} - v_i\right) + \frac{\partial \Phi}{\partial v_i}\right) + \lambda_{r_i} && (11.63)\\
&= e^{-\eta t}\left(c_1\left(v_f^{(i)} - v_i\right) + \frac{\partial \Phi}{\partial v_i} + \frac{1}{\eta}\frac{\partial \Phi}{\partial r_i}\right) && (11.64)
\end{aligned}$$

This equation shows how the marginal cost of the speed v_i of driver i depends on the difference in the free speed of driver i, the function Φ (proximity cost) and the marginal cost of the position of driver i.

We can now very easily determine that for the optimal acceleraton, the following equation holds

$$\begin{aligned}
\frac{d}{dt}v_i &= u^* = \frac{v_f^{(i)} - v_i}{\tau} - A_0\left(\frac{\partial \Phi}{\partial r_i} + \eta\frac{\partial \Phi}{\partial v_i}\right) && (11.65)\\
&= \frac{v_f^{(i)} - v_i}{\tau} - A_0 e^{-(r_{i-1}-r_i)/S_0} && (11.66)
\end{aligned}$$

where the acceleration time τ and the interaction factor A_0 are respectively defined by

$$\frac{1}{\tau} = \frac{c_1}{\eta c_3} \text{ and } A_0 = \frac{c_2}{\eta^2 c_3} \tag{11.67}$$

Eq. (11.65) shows how the acceleration of driver i is determined by the driver's aim to driver at the free speed $v_f^{(i)}$, and a factor that describes effect of driving behind another vehicle at a certain distance $s_i = r_{i-1} - r_i$. In appendix E we present the NOMAD pedestrian flow model, which has been derived using the optimal control paradigm discussed in this chapter.

Let us consider the stationary case, where the speed $v_i(t)$ is time-independent. Stationarity implies that the acceleration $\frac{d}{dt}v_i = 0$. This implies

$$v_i = v_f^{(i)} - \tau A_0 e^{-s_i/S_0} \tag{11.68}$$

This equation shows the relation between the stationary speed v_i, the free speed $v_f^{(i)}$ and the distance s_i between vehicle $i-1$ and vehicle i.

The following model was developed by [2]

$$\frac{d}{dt}v_i(t) = \kappa_0\left[V'(s_i(t)) - v_i(t)\right] \tag{11.69}$$

where κ_0 describes the sensitivity of the driver reaction to the stimuli, in this case the difference between the speed $V(s_i(t))$ - a function of the distance between vehicle i and its leader $i-1$ -

and the current speed $v_i(t)$. This model is equal to the optimal control model describes by Eq. (11.65), assuming that

$$V'(s_i(t)) = v_f^{(i)} - A_0\tau\left(\frac{\partial}{\partial r_i} + \eta\frac{\partial}{\partial v_i}\right)\Phi \tag{11.70}$$

and $\kappa_0 = 1/\tau$. In [2], the following relation is used

$$V'(s) = \tanh(s-2) + \tanh 2 \tag{11.71}$$

For this particular choice, it turns out that traffic flow becomes unstable at critical densities and that small distrurbances will cause stop-and-waves. For small *and* large vehicle densities, the flow is however stable. More precisely, the traffic flow is unstable when

$$\frac{d}{ds}V' > \frac{\kappa_0}{2} \tag{11.72}$$

Proof. To determine the instability criterion (11.72), we first linearise Eq. (11.69) around the spatially homogeneous solution

$$x_i^e(t) = x_0 - is + V'(s)\,t \tag{11.73}$$

For the variable

$$\delta x_i(t) = x_i(t) - x_i^e(t) \tag{11.74}$$

with $|\delta x_i(t)| \ll s$, we can then derive the following dynamic equation

$$\frac{d^2(\delta x_i(t))}{dt^2} = \kappa_0\left\{\frac{dV'}{ds}\left[\delta x_{i-1}(t) - \delta x_i(t)\right] - \frac{d}{dt}(\delta x_i(t))\right\} \tag{11.75}$$

$\delta x_i(t)$ can then be written as a Fourier-series

$$\delta x_i(t) = \frac{1}{N}\sum_{k=1}^{N} c_{k-1}\exp\left(2\pi j\frac{i-1}{N}(k-1) + (\lambda - j\omega)t\right) \tag{11.76}$$

again with $j^2 = -1$, which can in turn be substituted into the dynamic equation, yielding

$$(\lambda - j\omega)^2 + \kappa_0(\lambda - j\omega) - \kappa_0\frac{dV'}{ds}\left[\exp\left(-2\pi j\frac{k-1}{N}\right)\right] = 0 \tag{11.77}$$

It can subsequently be shown that when we determine solutions for $k = 1, 2, ..., N$, these solutions $\tilde{\lambda}(k) = \lambda(k) - j\omega(k)$ will have a positive damping rate $\lambda(k) > 0$ when Eq. (11.72) holds. Hence, the solutions are not stable. ■

11.6 Fuzzy Logic Models

Some researchers recognized that the reactions of the following vehicle to the lead vehicle might be based on a set of approximate driving rules developed through experience. Their approach to modeling these rules consisted of a fuzzy inference system with membership sets that could be used to describe and quantify the behavior of following vehicles. However, the logic to define the membership sets is subjective and depends totally on the judgment and approximation of the researchers. Furthermore, no field experiments were conducted to calibrate and validate these fuzzy membership sets under real driving conditions. While we agree with the premise of the paper, and seek to resolve similar issues, the methodology employed does not warrant any further explanation. Some researchers may argue over semantics, but there are no quantitative problems in fuzzy logic that cannot be solved in an equivalent manner using classical methods.

11.7 Cellular automata models

The CA models (*Cellular Automata*) divide the roadways into small cells. For instance, in the model of [38], these cells have a constant length of $\Delta x = 7.5$m. The cells can either be occupied by one vehicle or not. Besides the location of the vehicles, also their speeds v is discretised, and can only attain discrete values

$$v = \hat{v}\frac{\Delta x}{\Delta t} \quad \text{with} \quad \hat{v} = 0, 1, ..., \hat{v}_{\max} \tag{11.78}$$

With respect to the discretisation in time, the time-step is chosen such that a vehicle with speed $\hat{v} = 1$ precisely moves 1 cell ahead during one time step. Thus, in case of a time-step $\Delta t = 1$s, a maximum speed of 135 km/h holds. Despite this rather coarse represation, the CA-model describes the dynamics of traffic flow fairly well.

The updating of the vehicle dynamics is achieved using the following rules

1. *Acceleration.* If a vehicle has not yet reached its maximum speed $\hat{v}_{\max}$, and if the leading vehicle is more that $\hat{v} + 1$ cells away, then the speed of the vehicle is raised by one, i.e. $\hat{v} := \hat{v} + 1$.

2. *Braking.* If the vehicle driving with speed $\hat{v}$ has a distance headway Δj with $\Delta j \leq \hat{v}$, then the speed of the vehicle is reduced to $(\Delta j - 1)$. As a result, the minimum safe distance of $(\hat{v} + 1)\,\Delta x$ is maintained.

3. *Randomisation.* With a probability $\hat{p}$ will the speed be reduced with 1, i.e. $\hat{v} := \hat{v} - 1$. This describes the fact that the vehicle is not able to perfectly follow its predecessor.

4. *Convection.* The vehicle will move ahead with $\hat{v}$ cells during a single time step.

The updating of the vehicles can be achieved in a number of ways, e.g. in a random order, in the direction of travel, or just in the opposite direction. For the particular model here, the updating order does not affect the model behaviour.

Due to the simplicity of the model, complex networks with a large number of vehicles can be simulated in real-time. By varying the parameters $\hat{v}_{\max}$ and $\hat{p}$, different fundamental diagrams can be established, approximating real-life traffic flow at different levels of accuracy. The CA model describes the spontaneous formation of traffic congestion (unstable traffic conditions) and stop-and-go waves.

A large number of modifications to the basic CA model of [38] have been established, e.g. including lane changing. Also analytical results were obtained for simplified model dynamics.

11.7.1 Deterministic CA model

Let us briefly consider some of the properties of the CA model by excluding the randomisation rule. In that situation, all initial conditions eventually lead to one of the following two regimes (depending on the overall density of the system):

1. *Free-flow traffic.* All vehicles move with speed $v_{\max}$ and the gap between the vehicles is either $v_{\max}$ or larger. As a result, the flow rate in this regime equals

$$q = kv_{\max} \tag{11.79}$$

2. *Congested traffic.* If the density k is larger than the critical density k_c, not all vehicles will be able to move at maximum speed. In that case, the average speed of the drivers equals

$$u = \frac{1}{k} - 1 \tag{11.80}$$

and the average flow thus equals

$$q = 1 - k \tag{11.81}$$

The two regimes meet at the critical density k_c, which equals $k_c = \frac{1}{k_j+1}$. The capacity equals $q_c = \frac{v_{\max}}{v_{\max}+1}$.

Chapter 12

Gap-acceptance theory and models

Summary of chapter - In traffic it happens rather frequently that a participant (a car driver, a pedestrian, a cyclist) has to 'use' a gap in an other traffic stream to carry out a manoeuvre. Examples are: crossing a street as a pedestrian; overtaking at a road with oncoming vehicles; entering a roundabout where the circulating vehicles have priority; entering a motorway from an on-ramp; a lane change on a motorway; etc.

In this chapter the process of gap-acceptance will be discussed for an overtaking manoeuvre on a two-lane road. It will be assumed that the sight distance along the road is sufficient for an overtaking. Consequently only the oncoming vehicles are impeding a desired overtaking manoeuvre. As an exercise the reader is invited to investigate which modifications are needed when the discussion is concerning a pedestrian that has to cross a road with vehicles having priority.

List of symbols

s_i	m	distance headway of vehicle i
T_r	s	reaction time
v_i	-	speed of vehicle i
τ	s	reaction time
a_i	m/s^2	acceleration of vehicle i
x_i	m	position of vehicle i
κ	$1/s$	sensitivity
u	-	mean speed

12.1 Gap acceptance at overtaking

In general terms the gap acceptance process preceding an overtaking can be described as follows: drivers that want to make an overtaking estimate the 'space' they need and estimate the available 'space'. Based on the comparison between required and available space, they decide to start the manoeuvre or to postpone it. The term space is deliberately somewhat vague; it can be expressed either in time or in distance.

Definition 57 *The* required space *is dependent on characteristics of the driver, the vehicle and the road.*

Definition 58 *The* available space *is dependent on the characteristics of the on-coming vehicles and the vehicle to be overtaken (the passive vehicle).*

Drivers have to perceive all these characteristics, process them and come to a decision. Humans differ a lot in perception capabilities, e.g. the ability to estimate distances can vary

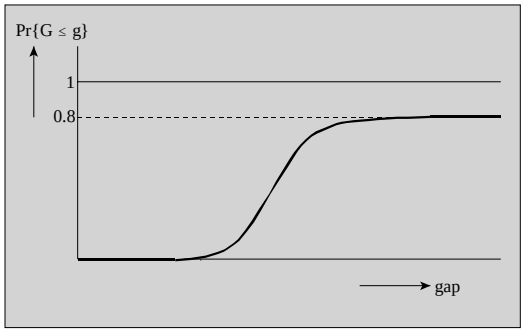

Figure 12.1: Example of a critical gap distribution with a part of the population that never carries out an overtaking, irrespective of how large the gap is.

substantially between persons, and they differ in the acceptation of risk. The total acceptance process is dependent on many factors of which only a subset is observable. This has lead to the introduction of stochastic models.

12.2 Model with a fixed critical gap for each driver

A much used model is the following: it is assumed that each driver has his/her own critical gap, i.e. gaps smaller than the critical gap are rejected and on the other hand gaps larger than the critical value are accepted. The critical gap of a driver is a constant value in a given situation and not dependent on e.g. the time a driver has already been waiting for an opportunity to carry out the overtaking (impatience is not taken into account).

Different drivers have different critical gaps and in a given situation it is assumed these critical gaps have a specific distribution. Probability densities that might be suitable should have a long tail to the right and should be limited to the left. The tail represents the careful drivers that need a long gap and the left limit is caused by more objective and mechanical factors. Suitable distributions are the *Log-normal distribution* and the *Logit distribution*. The Logit distribution of gap g is:

$$\Pr\{G \leq g\} = F(g) = \frac{1}{1 + \exp[c_1 - c_2 g]} \quad \text{with parameter } c_2 > 0 \tag{12.1}$$

Especially for overtaking on two-lane roads a distribution as sketched in Fig. 12.1 might be appropriate. In the example around 20% of the drivers have an infinite critical gap, in other words they never carry out an overtaking.From observation of rejected and accepted gaps it is possible (in principle) to derive the *gap acceptance function*. This function describes the probability that an arbitrary driver accepts an *offered gap*, i.e. starts the overtaking manoeuvre. From this gap acceptance function can be derived the distribution of critical gaps. In the special case that one uses only one observation per driver, i.e. only one accepted or rejected gap, the gap acceptance function is equal to the distribution of critical gaps.

$$\begin{aligned} Pr\{\text{acceptance gap} &= g\} && (12.2)\\ &= Pr\{\text{an arbitrary driver has a criticalgap} < g\} && (12.3)\\ &= \text{Distribution of critical gaps} && (12.4) \end{aligned}$$

If one uses more than one observation per driver, e.g. 2 rejected gaps and one accepted, then *this equality is no longer valid*. The observed gaps are far from independent and this necessitates special procedures.

(1)	(2)	(3)	(4)	(5)	(6)	(7)
class (mean gap)	distr. crit. gaps	# offered gap	# acc. gaps	distr. acc. gaps	# rej. gaps	distr. rej. gaps
5	0.00	0	0	0.00	0	0.00
7	0.00	30	0	0.00	30	0.50
9	0.33	30	10	0.11	20	0.83
11	0.67	30	20	0.33	10	1.00
13	1.00	30	30	0.66	0	1.00
15	1.00	30	30	1.00	0	1.00
Sum		150	90		60	

Table 12.1: Main characteristics of the gap acceptance process

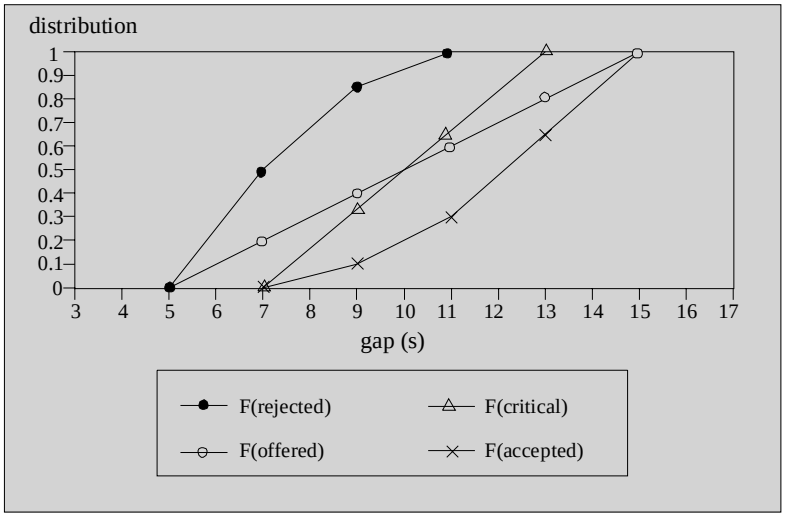

Figure 12.2: Illustration of several distribution functions that plan a role in the gap acceptance process

12.3 Example to illustrate the concepts

Table 12.1 and Fig. 12.2 illustrate the concept of the distributions of critical gaps, rejected gaps and accepted gaps. We have assumed a uniform distribution for the critical gaps; the distribution function is a straight line from 0 to 1 between 7 and 13 s. The offered gaps also have a uniform distribution, but the range is different, between 5 and 15 s.

Column (4), the number of accepted gaps, equals the product of the number of offered gaps, column (3), and the distribution function of the critical gaps, column (2). The total number of accepted gaps is 90. The distribution of the accepted gaps, column (5), is the quotient of the cumulative number of accepted gaps and the total of 90.

The number of rejected gaps, column (6), is the number of offered gaps, column (3) minus the number of accepted gaps, column (4). The distribution of the rejected gaps follows from column (6), just as column (5) from column (4).

The table and the graph show clearly there is a difference between the distribution of the accepted gaps and the distribution of the critical gaps. However, in this simple example the fraction of accepted gaps is equal to the distribution of critical gaps.

Remark 59 *In the American handbook Highway Capacity Manual (HCM) (edition 1985, Ch.*

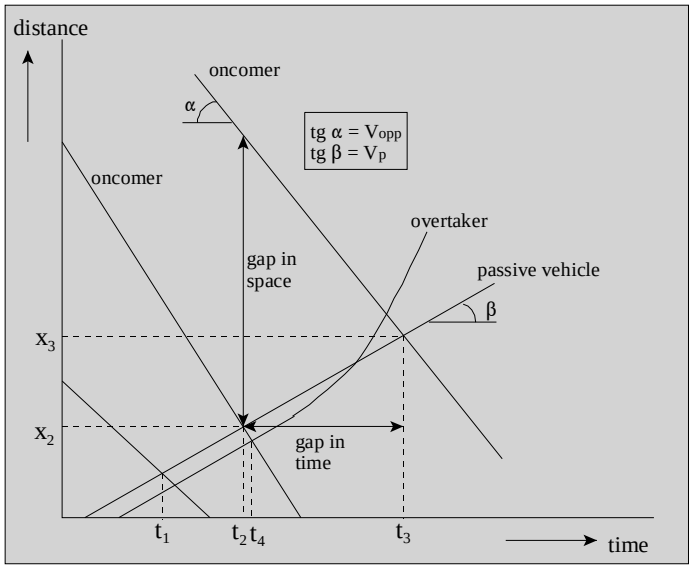

Figure 12.3: Accelerating overtaking manoeuvre in time-space plane

5 page 10; [4]) the critical gap is defined as: "the median time headway between two successive vehicles in the major street traffic stream that is accepted by drivers ...".
It is obvious that in the example presented here, this definition would lead to a biased estimate. The median (50 percentile) of the accepted gaps is around 12 s and the median of the critical gaps is 10 s. The crucial point is that characteristics of accepted gaps, depend on the gaps offered, that is on the traffic situation. A good method to determine critical gaps should always be based on both accepted and rejected gaps. In the HCM of 2000 this point has been improved.

12.4 Overtaking manoeuvre in time-space plane

After the acceptance of a gap follows the overtaking manoeuvre itself. During the manoeuvre the driver can deviate from his/her intended behaviour. He/she can for instance detect that the oncoming vehicle approaches faster than anticipated or that the acceleration of his/her car is not as good as expected. In a critical case the driver can decide to abort the manoeuvre and go back to his/her original (relative) position. Of overtakings carried out the characteristics 'used distance', 'used time', and the 'margin' that is left at the end of the manoeuvre are important.

Fig. 12.3 exhibits an overtaking manoeuvre in the time-space plane by using schematised vehicle trajectories. It is a so-called accelerative overtaking manoeuvre at which the overtaker has the same speed as the vehicle being overtaken at the start of the manoeuvre.

The gap expressed in time is usually defined as the duration of the interval between two successive meetings of the passive vehicle and oncomers. This interval equals the relative headway of oncomers as observed from the passive vehicle. It is the absolute value of the headway of the oncoming stream divided by $1 + v_p/v_{opp}$; index $_p$ of passive and *opp* of oncoming or opposing vehicle. The derivation of this relation is left to the reader.

In Fig. 12.3 $t_2 - t_1$ is a rejected gap and $t_3 - t_2$ an accepted gap. These gaps can not be observed directly by the overtaking driver. He/she has to estimate the distance to the oncoming vehicle and its speed and process this to get an estimation of the available gap, expressed in time.

If the driver can not estimate the speed of the oncomer, it is likely he/she bases his/her decision whether to overtake on the distance to the oncomer. Consequently this distance can

also be considered to be the gap that is either rejected or accepted.

Remark 60 *Sometimes the distance covered by the passive vehicle during the gap in time, that is x_3 minus x_2, is taken as the gap in distance.*

Remark 61 *The gap in time, as defined earlier, is not completely usable for the overtaking. The nett gap starts at moment t_4, when the oncomer meets the overtaker. Moreover, the overtaker has to be back on the right lane before moment t_3 in order to prevent a collision. In reality the difference between the nett and the gross gap is not very large.*

Remark 62 *Besides the accelerative overtaking also flying overtaking manoeuvres are possible, at which the overtaking vehicle (hardly) changes its speed. Flying overtakings are difficult to investigate and to observe, e.g. the definition of the gap is not obvious.*

12.5 Studies into overtaking behaviour

Because the overtaking manoeuvre on two-lane roads is important, as well regarding traffic operation as safety, much research has been carried out. These studies can be roughly divided into:

- Experiments in laboratories concerning the perception abilities of drivers. The traffic situation is sometimes strongly schematised, e.g. two moving points on a tv screen represent the rear lights of a car at darkness.
- Experiments with vehicle simulators in a laboratory; an example is shown in Fig. 12.4.
- Experiments on laboratory roads or test tracks. In that case the situation is more realistic than in a vehicle simulator and it is still possible to control the traffic situation and offer drivers specific overtaking tasks.
- Observations in real traffic, sometimes with an instrumented vehicle and sometimes only observed from 'outside'. In those field studies one just has to wait how many interesting events do occur. Moreover the measurement of many relevant variables can be rather problematical.

The main results of the perception studies are:

- Drivers have a rather precise feeling for the speed they drive and the possibilities of their vehicles in terms of accelerations and decelerations they are prepared to apply or normally use.
- Drivers can estimate distances fairly accurate.
- In contrast the estimation of the speed of an oncoming vehicle is hardly possible for human beings. This can be understood: an acceptable time headway is around 20 s; an oncomer with a speed of 90 km/h then is at a distance of 500 m. ($20 \times 90 / 3.6$)

It is assumed that drivers estimate the distance of the oncoming vehicle and assume a value of the speed (e.g. based on the character of the road, their experience, the type of vehicle). They somehow deduce an available time from these perceived variables. Knowing this, it is not a surprise that drivers do differ so much in their gap acceptance.

The Daimler-Benz simulator in Berlin is an advanced driving simulator and uses computer-generated images. It consists of the following elements:

- A cylindrical-shaped projection dome with a diameter of 7.40 m.

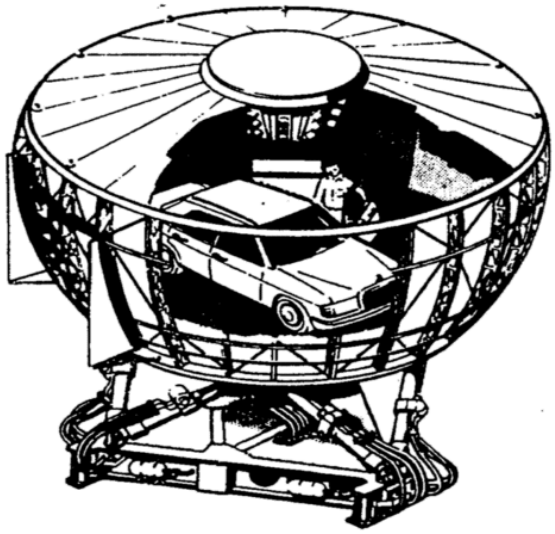

Figure 12.4: The Daimler-Benz vehicle simulator

- A complete car, a Mercedes 19O E, is positioned inside the dome in order to simulate the driver's immediate environments as realistically as possible.
- A moving base with six degrees of freedom and a maximum permissible acceleration in excess of 1g. The system allows for translatory motions of ±1.50 m and rotations of : ±33 to 45°.
- A projection system giving a visual field 33° vertically and 180° horizontally. The system comprises six video projectors.
- A digital colour picture unit (256 colours) with a library of up to 3000 basic shapes. The picture can be updated 50 times a second.
- A sound system capable of simulating noise from car engines, tires, wind, etc.
- A dynamic vehicle model that is powerful enough to be used in vehicle design.

12.5.1 Calculated required space

Based on simple mechanical rules and some empirical facts Brilon [8] has calculated the required distance to the oncoming vehicle for an accelerative overtaking manoeuvre. Important assumptions are:

- drivers use the full acceleration capabilities of their vehicles. These are dependent on the vehicle type and the speed, which in this case is the speed of the passive vehicle, $v_{p(assive)}$.
- overtakers keep accelerating until their rear side is at the same position as the front of the passive vehicle.

The following cases have been distinguished:
A: $v_{opp} = 120$ km/h and the overtaker respects no speed limit;
B: $v_{opp} = 100$ km/h and the overtaker limits its speed to 120 km/h.
It appeared that in reality most drivers use longer distances than are required according to the 'mechanical' model. This is representative for a general fact: very few drivers use the full accelerative and braking possibilities of their vehicles.

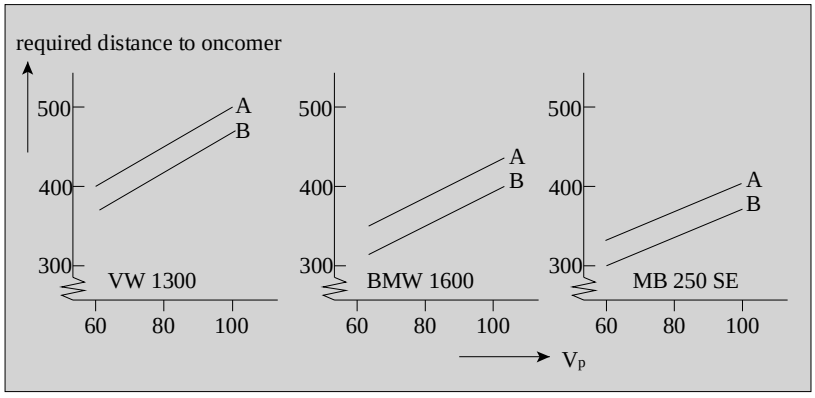

Figure 12.5: Required distance oncomer for three vehicle types and two situations

12.5.2 Studies with vehicle simulators

In the Netherlands both TNO 'Human Factors Research Institute' ('Technische Menskunde') in Soesterberg and the 'Verkeerskundig Studiecentrum' in Groningen have a modern vehicle simulator at their disposal. Advantages of studies with such a tool are:

- conditions can be controlled; e.g. one can expose different test persons to exactly the same conditions and in that way assess differences between drivers; on the other hand one can offer the same person the same conditions on different times and investigate consistency of driver behaviour.
- nearly all variables and conditions can be measured precisely.
- it is possible to investigate dangerous situations.
- it is possible to investigate new driver support systems, e.g. how will drivers react to intelligent cruise control (what distance do they prefer; when do they overrule the automatic system; etc.).

However, there are also disadvantages of using vehicle simulators and these are focussing on the 'validity' of the results. The main point always is: will test persons behave as they do in reality; see e.g. [31]. Especially when investigating dangerous situations, drivers might always be aware of the fact that the danger is not real.

The Mecedes Benz vehicle simulator has been used to investigate overtaking behaviour. The test drivers complained about:

- there were no mirrors on the vehicle to look backward;
- the scenery surrounding the car was very dreary;
- soon after each overtaking had been carried out, there appeared a new slow vehicle in front. This was done because it increased the efficiency of the experiment, i.e. it took less time. Using the simulator is very expensive.

12.6 Estimation of critical gap distributions

Many different methods to estimate the distribution of critical gaps, from observations of the gap acceptance process in reality, can be found in the literature. In a recent study [9] the most

used methods have been compared using simulated data (only in this situation one knows the real outcome). It appeared that many methods with a good reputation failed the test. However, the Maximum Likelihood (ML) estimation method did not fail and will be discussed.

Suppose a driver successively rejects gaps of 3, 9, 12 and 7 s and accepts a gap of 19 s. The only thing one can conclude from these observations is that this driver has a critical gap between 12 and 19 s. Stated in other words: the critical gap can not be observed directly. Secondly it can be concluded that only the maximum of the rejected gaps is informative for the critical gap; the smaller gaps are rejected by definition (we assume a consistent driver).

The unknown distribution of the critical gaps is denoted as $F(g;\theta)$ with θ is a vector of parameters. (e.g. for a Normal distribution $\theta = [\mu, \sigma]$). Denote the maximum rejected gap by g_1 and the accepted gap by g_2. The probability that the pair (g_1, g_2) occurs, is proportional to the probability that the critical gap is between g_1 and g_2 ; that is: $F(g_2) - F(g_1)$[1]. For each observed driver i we have available a pair $[g_{1,i}, g_{2,i}]$. The combined probability of occurrence, i.e. the likelihood function, is:

$$L(g_{1,i}, g_{2,i}; \theta) = \prod_{i=1}^{n} (F(g_{2,i}; \theta) - F(g_{1,i}; \theta)) \tag{12.5}$$

The maximum likelihood procedure implies that the function L (or the logarithm of L which is usually much easier to maximize) must be maximized by varying the parameter θ; the value of θ at the maximum is the likelihood estimator.

12.7 Practical results

12.7.1 Case 1. Overtaking of long trucks

Fig. 12.6 shows the result of an Australian study [55] of overtaking behaviour of cars that overtake a long or even extra long truck. The passive vehicle (the long truck) was equipped with video camera's and other apparatus to record the overtaking behaviour and the gaps offered.

From the figure it is clear that the distribution of the critical gaps is situated more to the left than the gap acceptance function. Besides the distribution of the accepted gaps, the distribution of the overtaking times are depicted. The differences between the gaps used and the critical gaps are the margins drivers have who carry out an overtaking at their critical gap.

12.7.2 Case 2. Left turn movement at intersection, taking into account many factors

At such a manoeuvre the critical gap is much smaller (order of 5 s) than at an overtaking and the driver is able to take more factors into account in his/her decision process. A recent study in Canada [1] is a good example of such a study; see Fig. 12.7 and Table 12.2.

12.8 Conclusions and main points

- Gap acceptance is an important element of the traffic processes; it is relevant for operational aspects and safety.
- Differences between drivers are relatively large.

[1] The probability of pair (g_1, g_2) also depends on the distribution of the offered gaps, but that distribution is not dependent on the unknown parameters θ and consequently is not relevant for the estimation procedure.

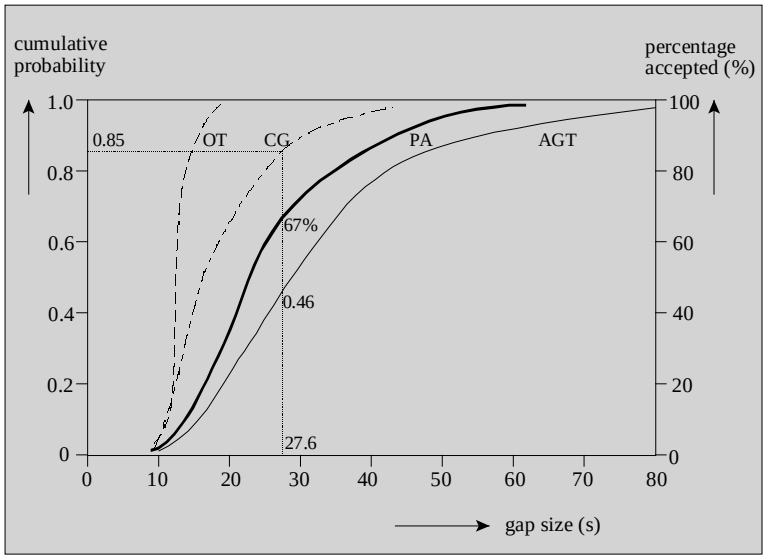

Figure 12.6: Results of overtaking study with a truck of 20 m length at a speed of 70 km/h as passive vehicle; AGT = accepted gaps; PA = percentage accepted = gap acceptance function; CG = critical gaps; OT = overtaking times

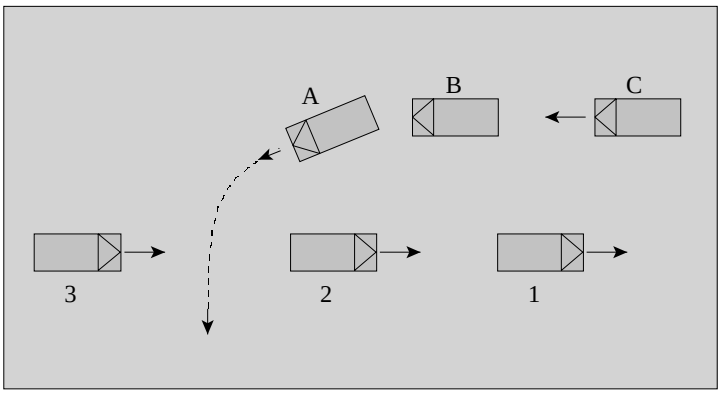

Figure 12.7: Gap acceptance process at left turn

Factor	Effect on gap acceptance
Gap in time	The most explaining factor if considering only one factor
Gap in distance and speed of oncomer	The most explaining combination of the set of explaining factors
Type of oncomer	In sequence: Motorbike-Truck-Car-Bicycle, the gap decreases
Presence vehicle behind (traffic pressure)	If present, then accepted gap $-0.2s$
Delay in queue (not being in front yet)	$10s$ more $\rightarrow$ accepted gap $-0.2s$
Delay in front position	No monotone effect; gaps increase for delay $< 30s$ and decrease for larger delay
Accelerative power	No effect
Gender driver	If woman, then gap $+0.5s$
Age driver	No effect
Presence of passengers	If present, then gap + 0.3 s

Table 12.2: Main characteristics of the gap acceptance process

- A relatively simple assumption - each driver has his/her own constant critical gap in a given situation - implies already complex models and data analyses. Simulation models offer the opportunity to investigate more realistic assumptions and their implications.
- Data from gap acceptance in real traffic are difficult to collect. Vehicle simulators in laboratories are becoming more realistic and consequently can partially replace field studies.

Part IV

Traffic Simulation

Simulation models come in many forms and shapes. They can be either macroscopic or microscopic, deterministic or stochastic, or even static or dynamic. In this part of the reader, we will discuss simulation models in general, as well as focus on a particular macroscopic model and a particular microscopic model. Before doing so, we will adress the different steps that are important when doing an simulation study. In particular, we will focus on applying simulation models to perform ex-ante studies.

Before doing so however, let us emphasise that most traffic simulators do not only contain models predicting the traffic flow operations, but in many cases also route choice and even departure time choice. The macroscopic model METANET for instance has different algorithms that can be used to predict how drivers choose routes. It should be clear that in the remainder of this reader, these navigation processes are not discussed. It is difficult to give a short and complete definition of a simulation model. The words simulation and model have different meanings. [36] has given a very broad definition of simulation.

Definition 63 *Simulation is a numerical technique for conducting experiments on a digital computer, which may include stochastic characteristics, be microscopic or macroscopic in nature, and involve mathematical models that describe the behaviour of a transportation system over extended periods of real time.*

Another very general definition is the following:

Definition 64 *A simulation model is a representation of reality and it is possible to do observations on the model like one can do observations in real traffic.*

From this definition, the main difference between a model and a simulation model should be clear: a model is a mathematical representation of traffic; a simulation model is the numerical process, generally based on a mathematical model, that is used to simulate traffic flow on a digital computer. In illustration, the Payne model is a traffic model; a the computer implementation of a numerical approximation scheme of the Payne model is the associated simulation program.

Simulation models can satify a wide range of requirements:

1. *Evaluation of alternative treatments.* With simulation, the engineer can control the experimental environment and the range of conditions to be explored. Historically, traffic simulation models were used to initially evaluate signal control strategies, and are currently applied for this purpose as an integral element of traffic manegement research. This aspect will be discussed in detail in the sequel.

2. *Testing new designs.* Simular to (1): simulation can also be applied to quantify traffic performance responding to different geometrci designs before actually building the transportation facility.

3. *As an element of the design process.* The classical iterative design paradigm of conceptual design followed by the feedback process and design refinement - which formed the backbone of this reader - will benefit from the application of simulation as well. Here, the simulation model can be used for the evaluation: the detailed statistics provided can form the basis for identifying design flaws and limitations. The statistics augmented with animation displays can provide insights guiding the engineer to improve the design and continue the process.

4. *Embed in other tools.* In addition to its use as a stand-alone programm, simulation submodels can be integrated within software tools designed to perform other functions.

5. *Training personnel.* Simulation can be used in the context of a real-time laboratory to train operators of Traffic Management Centers (TMC).

This inconclusive list indicates the variety and scope of traffic simulation models. The remainder of this part of the reader will discuss general guidelines for applying a simulation model. The first step, however, should be the decision to actually choose a simulation model as the main tool for analysis. This choice is by no means obvious, and care should be taken to decide whether simulation is indeed the best way of solving the problem at hand.

Chapter 13

Simulation study guidelines

Summary of this chapter. This chapter focusses on general guidelines to perform a simulation study. It is important to keep in mind that a simulation study will generally form a part of a larger study, that it, it seldom stands on its own. In a lot of cases, traffic simulation will be used to test alternative treatments of problems in traffic systems. The Traffic Research Centre (AVV) of the Dutch Ministry of Transportation (RWS) has developed an architecture for traffic management (AVB) which is aimed at providing a framework to determine traffic management scenarios. In this architecture, traffic simulation plays an important role.

Fig. 13.1 shows an overview of this approach and the position of models studies within this context. The different processes in which traffic simulation plays an imporant role are:

1. 'Referentiekader' (objective situation - not to confuse with reference situation). The objective situation describes the traffic situation which the traffic analysist aims at. The objective situation reflects among other things the goals and constraints imposed by the policy makers. It for instance reflects that some relations are more important than others (e.g. an important motorway road with speeds higher than 80 km/h; less important motorways have a speed of 60 km/h or higher). The objective situation will be input to the model study for cross-comparing with the null situation and the situations after deploying the alternative treatments.

2. 'Feitelijke situatie' (null situation). The null situation describes the traffic situation in the network in case no additional measures are taken. This can be either the current situation or a future situation. In this phase of the AVB, simulation models are used for problem diagnosis (will problems occur and what are the cuases for these problems).

3. 'Knelpunten' (bottle-necks). In this phase of the AVB, the traffic analyst determines which part os the problem problems occur. Ofter, a distinction is made between *traffic bottle-necks* and *policy bottle-necks*. The former problems generally refer to congested roads, i.e. roads with low speeds. A policy bottle-neck is a problem where the traffic conditions are worse that sketched in the objective situation.

4. 'Services'. This phase of the AVB aims at identifying the traffic services that can be deployed in order to solve the policy bottle-necks in the network. Examples of traffic services are reducing inflow, reducing speeds, increasing capacity, etc. The services may be an input of the simulation model, which in turn can be used to predict the effect of the services. Note that it may not be easy to implement traffic services is all simulation models.

5. 'Maatregelen' (measures). From the services, the traffic (management) measures are determined. Once this is achieved, one (or more) alternative scenarios (consisting of sets of measures) will be implemented in a simulation model. By doing so, the effects of the

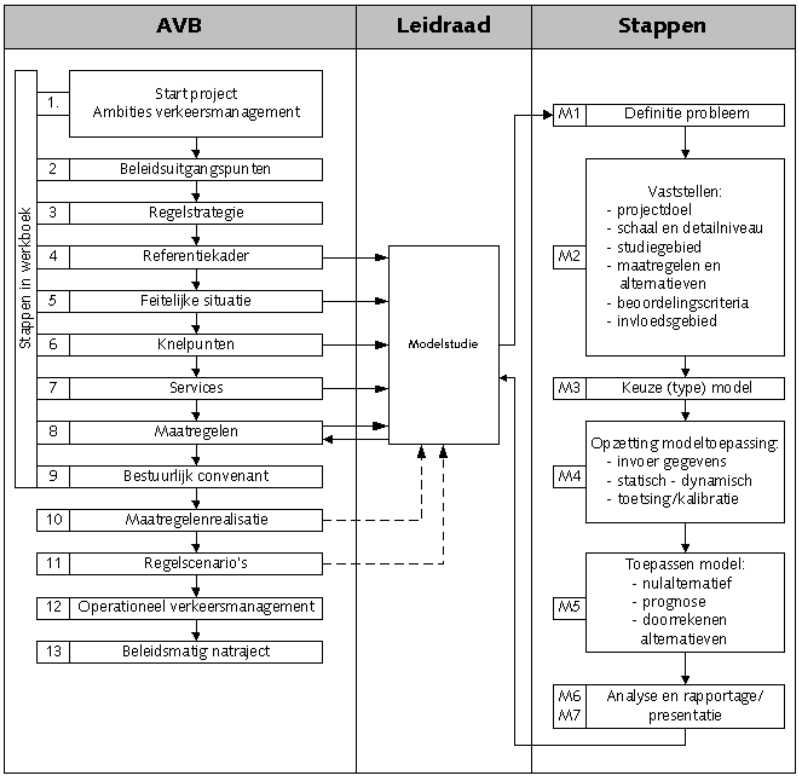

Figure 13.1: Overview of AVB approach

measures can be predicted and thus it can be determined whether the (set of) measures solve the policy bottle-necks in the system.

In the remainder of this chapter, we will discuss the different steps that are important when doing a simulation study in line with the steps distinguished in the Leidraad.

13.1 Overview of simulation study steps

Let us first briefly consider the different steps that are important when doing a simulation study. These steps are roughly in line with the approach in the 'Leidraad'.

1. Summary of problem definition
2. Determine goal of the project, i.e.
   (a) Scale and level-of-detail
   (b) Study area (problem domain), the area of influence, and the area of dependence.
   (c) Performance criteria
   (d) Current management measures and alternatives
   (e) Identify other tools required to answer research questions
3. Model choice (is a simulation model required? Which simulation model is most appropriate? Should a new model be developed?)
4. Setting-simulation study

   (a) Determine and acquire required input parameters and variables
   (b) Determine and acquire required data (static / dynamic)
   (c) Model calibration and validation

5. Model application

   (a) Simulate traffic conditions in reference situation
   (b) Predict traffic conditions for future situation (prognoses), i.e. for the null-alternative
   (c) Predict conditions for alternatives

6. Analysis and presentation of results (not discussed here).

The remainder of this section focusses on these different steps in some detail.

13.2 Summary of problem definition

As mentioned earlier, a simulation study always is a part of a larger study. In the first step of performing a simulation study, the following issues show be made clear:

- What are the main characteristics of the traffic system that are relevant for the study? What are the most important problems? What are the main phenomena that cause the problems? How can these problems be characterised and measured?
- What are the different alternative treatments that have been developed and need to be tested? Which management measures are considered (and thus need to be modelled?)
- Which performance indicators can be used to assess whether the alternative treatments will be successful?

An example may be an oversaturated bottle-neck, where congestion spills back onto an off-ramp further upstream. Alternative treatments for this problem are ramp-metering and dynamic inclusion of the shoulder-lane (so-called "spitsstroken"). Important characteristics of the problem are thus the dynamics of congestion and shock waves, differences in traffic behaviour when including the shoulder lane, impacts on capacity when ramp-metering is active (difference between pre-queue capacity and queue discharge rate). Important performance critieria are flows and capacities, speeds and travel times, delay times, fuel consumption, and emmissions.

13.3 Determining objectives of simulation study

It should be clear where the simulation study will be used for: will it be used to predict the effect of a certain treatment or will it be used to cross-compare alternative treatments. Furhtermore, in this step the abstact aims of the simulation study have to be translated into research questions.

To this end, the following questions need to be aswered:

1. *What is the level-of-detail / scale of the study?* Does the problem address the individual driver / vehicle? Are we considering a single discontinuity or an entire network? Do we need to model an entire peak period? Do we need to distinguish between different types of users? Do we need to include different types of infrastructure facilities (motorways, urban roads)

2. *Determination of the study area, the area of influence, and the area of dependence.* The study area describes the area of interest. In an ex-ante assessment study, this area consist of the area in which the measures are deployed. The area of influence includes also the parts of the network that will be of influence to the traffic operations in the study area; the area of dependence considers all parts of the network that are likely to be affected by the measure. For instance, when it is believed that a measure will have a significant effect or route choice, a larger part of the network will need to be included as well.

3. *What information is needed from the simulation study (performance indicators)? And, what information is available / needed by the model for input, calibration and validation (static and dynamic model input, data).* Determination of input and output; direct and indirect performance indicators. We can distinguish between direct performance indicators, following directly from the simulation study results (flow, speed, etc.), and indirect indicators that need to be computed from these outcomes.

4. *Idenfitication of current and future management measures.* What are the settings of the traffic control signals? Is route guidance installed and relevant within the context of the assessment study?

5. *What are other tools that may be involved in the study?* We can think of capacity estimation, demand modelling, signal optimisation, etc.

13.4 Model choice

In this step, the engineer has to decide *if* a simulation will be the most appropriate tool to answer the research questions and if so, *which* simulation model will be most appropriate. These choices are largely based on the results of the previous step. A simulation model is not a tool that can solve every problem, neither is it always the most efficient way to solve problems. One should look for alternatives that might be more appropriate or cheaper. [36] mentions some positive points for simulation:

- Other techniques may not be appropriate.
- It can yield insight into what variables are important and how they interrelate.
- Dynamic aspects can be handled; e.g. demand can be varied over time and space; and information can be produced as a function of time and space; not just average values and standard deviations.
- With a simulation model one can experiment off-line without using on-line trial-and-error approach.
- One can experiment with new situations that do not exist today (how about the validity of the model in such a case?).
- Potentially unsafe situations can be simulated without real risk (validity?).

He also mentions some reservations:

- There may be easier ways to solve the problem. Consider all possible alternatives.
- Simulation is time consuming and expensive. Do not underestimate time and cost.
- Simulation models can require considerable input characteristics and data, which may be difficult or impossible to obtain.

- Simulation models may be difficult to use for non-developers because of lack of documentation or the requirement of unique computer facilities
- Some users may apply simulation models and not know or appreciate model limitations and assumptions; and even more: Users may apply simulation models and treat them as black boxes and really do not understand what they represent.
- Development of simulation models requires knowledge in a variety of disciplines, including traffic flow theory, computer programming and operation, probability, decision making, and statistical analysis. The developer of the model must fully understand the system.
- Simulation models require verification, calibration, and validation, which, if overlooked makes the model useless.

Regarding the choice to use a simulation model, the engineer should first ask him/herself if other (analytical) tools are available to answer the research questions that are for instance less costly and time consuming. An additional problem may be caused by the data availability. The choice for a particular model will depend on the purpose of the study, and its characteristics on the one hand (see previous section), and the properties and limitations of the candidate simulation models. What are the assumptions of the model? What are the computational requirements? Etc.

To aid the model choice, classification of the models based on their functionality and characteristics is important. Amongst the most important classification dimensions are

1. Level-of-detail in traffic representation (microscopic and macroscopic) and behavior modelling (microscopic and macroscopic); length of time-step and simulation period
2. Discrete time or discrete event systems
3. Deterministic or stochastic models
4. Urban and / or inter-urban traffic flow
5. Section, intersecton, bottle-neck, or network models; structure of the network (simple, complex)

Fig. 13.2 shows several simulation models classified based on the time-scale and spatial scale.

13.5 Setting-up a simulation study

In general, the data needs for a simulation study are large, and in most cases, not all data is available right away. Since the outcomes of the calculations depent on the data available, the quality of the data needs to be high.

The data needed by the model includes both traffic-related, infrastructure-related and measure-related data. Infrastructure related data generally describes the roadway and bottleneck geometry and the network topology. Measure-related data describes the characteristics of the traffic measure, e.g. signal settings of a traffic controller. Note that these can either be very simple (and static) or traffic response. In the latter case, matters may become rather complicated.

Data can be used to calibrate the model. In a macroscopic model, model calibration generally pertains to choosing the appropriate fundamental diagram and other parameters from data (or chose the parameters such that the model output resembles the calibration data with sufficient accuracy). In a microscopic model, the parameters defining driving behaviour will need to be determined from available data.

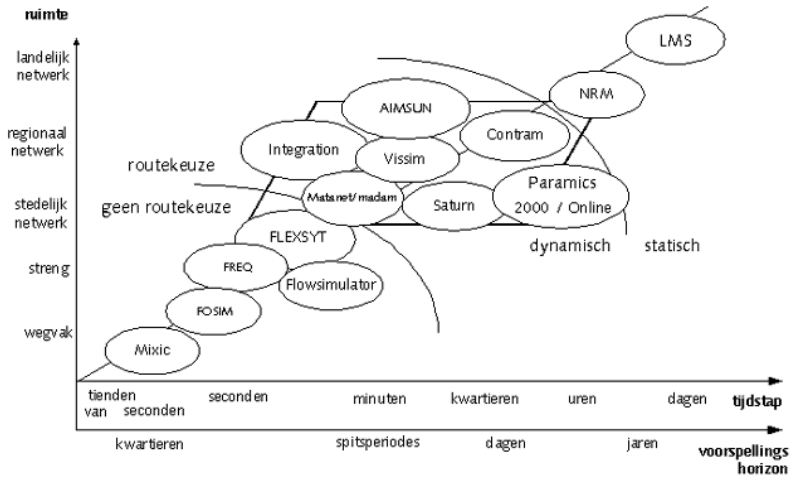

Figure 13.2: Overview of models based on classification into time and space

Also data regarding the traffic demands is needed. In general, these data concist of dynamic origin-destination patterns, describing the traffic demands from one origing (e.g. an on-ramp) to another (e.g. an off-ramp).

When the model has been calibration and all the input required to run the simulation model has been collected, the model needs to be verified and validated for the *reference situation*. Generally, verification pertains to the functioning of the software. For most commercial simulation software, this step will not be involved. Validation establishes that the model behaviour accurately represents the real-world system that is being simulated. Validation of the model involves the following activities:

- Acquiring real-world data which, to the exent possible, extends *over* the domain of the model
- Reducing and sturcturing these empirical data so that they are in the same format as the data generated by the model
- Establishing validation criteria, stating the underlying hypotheses and selecting the statistical tests to be applied
- Developing the experimental design of the validation study, including a variety of "scenarios" to be examined
- Performing the validation study:
  - Executing the model using input data and calibration data representing real-world conditions
  - Performing the hypothesis testing
- Identification of the causes and repairing the model accoridngly

13.6 Application of the simulation model

When the model has been calibrated and validated for the reference situation, the model is applied to predict the flow conditions for the different alternatives. Frequently, amongth these alternatives will be the null-alternative, i.e. the situations where the current situation is simply

extrapolated to the future. In most cases, traffic demands are changed (increased?) and the model is used to predict the problems that will occur in the future. In some cases, the null-alternative is equal to the reference case used for validation, for instance when the measures are to be determined that solve the current problems in the network. In either case, in an ex-ante study, the null-alternative will serve as a reference to which the alternative treatments will be compared.

When the traffic conditions in case of the null-alternative have been established, the effects of the alternative treatments will be determined. Different methods exist to cross-compare the different alternatives (multicriteria analysis, sensitivity analysis, cost-effectiveness, cost-benefit analysis). These fall out of the scope of this reader.

Chapter 14

Microscopic simulation models

This chapter consists of a general part about simulation (definition, general description, developing and use of a model) and a specific part concerning the microscopic model FOSIM for traffic operation at motorways.

14.1 General aspects of simulation models

14.1.1 General description of simulation

We will restrict ourselves to microscopic simulation models; however, this does not mean that macroscopic simulation models are not important. A microscopic simulation model of a traffic stream can be defined as:

It is a description of the movements of individual vehicles that are considered to be a result of:

- characteristics of the drivers and the vehicles
- interactions between the driver-vehicle elements
- interactions between driver-vehicle elements and the road characteristics
- conditions (weather, light)
- control (static, such as a fixed speed limit; and dynamic, such as: traffic lights; motorway signalling systems; on-ramp control; and, policemen)

In general this process is so complex and the number of factors so large that it can not be handled by analytical methods. The only method that remains is to imitate the process with a computer. The program calculates the position, speed and all other 'state variables' of a vehicle every time step Δt. An often used value for this time step is 1 second. However, in a good model this time step should be an input parameter that makes it possible to use other values. It's obvious that a smaller time step will lead to a longer computer time and that it should lead to a more precise description of behaviour.

14.2 Applications

Many simulation models have been developed in the transportation field. They are used for analysis of existing systems that need modifications and for the design of new systems. In general the use of a simulation model as a tool to solve problems is indicated when the system considered is complex and its components interact.

Examples:

- Traffic operation on motorways with a lot of on-ramps and off- ramps.
- Traffic operation on weaving sections of a motorway.
- Pedestrian flows at airports and railway stations.
- Design of control of isolated signalized intersections.
- Design of control for a network with controlled intersections.
- Two-lane roads with varying road characteristics and extra overtaking lanes.

14.2.1 Parts of the model

- Submodels for the interactions between Driver Vehicle Elements (DVE's).

Examples are submodels for overtaking, lane changing, car-following, submodels for interactions between DVE's and other factors. Examples: Effect of road width, curves, grades, weather, speed limit

- Input

For one, the characteristics of DVE's need to be established. Examples are type of car, free speed of vehicle, maximum acceleration used on grade, acceleration used when accelerating to free speed, critical gap when overtaking, preferred distance headway at car-following, type of driver (calm, aggressive), etc.

Other input consist of road characteristics (number of lanes, width, radius and length of curve, grade, etc.)

- Output

The most complete output consists of all characteristics of all vehicles on the road at every moment of time. This consists of an enormous quantity of numbers that are difficult to handle. It is mostly only useful when developing the model, looking whether it behaves as intended. A graph of the vehicle trajectories can be a good help. A sequence of 'views from the air' (snapshots) of the road with the vehicles moving over the computer screen (animation) is a good tool for demonstrations and also for checking the correctness of the model.

- Random processes

Most models have chance processes to model the fluctuations that are part of reality. The input very often is stochastic; e.g. the type of vehicle and its desired speed are drawings from a probability distribution of which the parameters are given. Also during the process on the road probabilities can play a role. Differences between drivers or even within one driver at different moments can be modelled in this way.

- Generation of random numbers

Most computers have built-in generators of pseudo random numbers that are uniformly distributed between 0 and 1. From these one can generate all types of drawings from other probability distributions. In the past this element of the models was considered to be the most important, leading to the name 'Monte Carlo' simulation.

Remark 65 *The random number generator of a PC can be controlled by a so called seed, usually an integer number. If the seed of a run is the same, the sequence of random numbers is exactly the same. This is a useful property, one can change a deterministic element of the model and by keeping the seed the same, the random elements of the model are the same.*

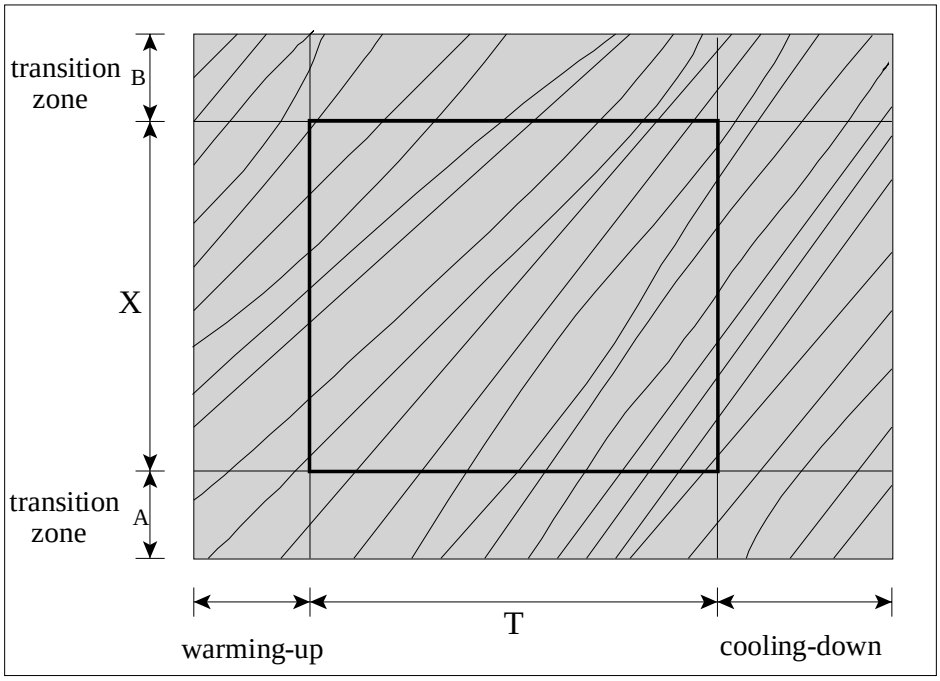

Figure 14.1: Simulation over an area in the road-time plane of size X times T with extra strips to ascertain correct functioning (not to scale)

Using a different seed and keeping all other input the same, is like carrying out observations on two comparable days (same weather, same demand, same truck percentage, etc). The differences in results can then be attributed to the inherent randomness of the traffic process.

The general procedure to generate stochastic variates with distribution F is shown in Fig. **??**. The drawing X, from a uniform probability density between 0 and 1, is transformed by $F^{-1}(x)$ and the outcome Y is the required variate. When the calculation of the transformation is not easy, it might take too much computer time. It should be realized that in some models over 1 million of random drawings can be needed for a single run and then it is worth the trouble to use a more efficient method for generation. There are a lot of books discussing methods to generate stochastic variates. A standard trick, that is cheap in terms of computer time, to generate random variables from a Normal Distribution, is the following: Take the sum of a number (say 12) of uniform distributed random numbers. Due to the 'Central-limit Theory' this approximates the required Normal variates rather well.

- Warming-up time

In most models the simulation starts with an empty road. This makes the behaviour in the beginning of the simulation not realistic. One should simulate until at least the vehicles have reached the end of the section that is modelled.

The required so called warming-up time can be estimated from the length of the road and a rough indication of the travel speed.

- 'Cooling-down time'

Some models simulate longer than the period T that is used as an input. This makes it possible to collect characteristics such as the travel time for the total road length for all vehicles that entered the road between time = 0 and the duration of the simulation T; see Fig. 14.1.

- Transition sections

Many models have an extra length of road at the beginning and sometimes also at the end of the road section considered. This is needed because the entering vehicles are not immediately in a realistic state and need a certain distance to reach more realistic behaviour. At the end of the section, the sudden disappearance of vehicles can lead to unrealistic effects. For both reasons it is useful to have a transition section, at both ends of the section on which the investigation is carried out.

14.2.2 Development of a simulation model

Suppose the decision has been taken that for a given type of problem(s) a simulation model is needed. This does not imply that one should develop a new model. It is worthwhile to consider the alternative of buying an existing model, even when it does not fulfil all requirements. The development of a model is very time consuming.

Suppose one chooses to develop a model, then there are a lot of choices to be made. An important point is: Should one develop a special purpose model, that is as simple as possible and just able to solve the problem at hand; or should one develop a more comprehensive model that can be used for a wider class of problems. All those decisions have implications for costs and benefits, but this point will not be discussed here further because it depends so much on the case at hand.

Now follows a list of points that are important in developing a model.

- Define the aim(s) of the model
- Derive from that the type of model that is needed.
- Design the model in broad outline.
- Draw up a general description of the modules of the model
- Work out the modules in flow diagrams
- Choose a computer language or a simulation package
- Program the modules
- Verification
  - Test the modules separately to see that they work as intended.
  - Combine modules in larger parts and verify them.
  - Combine the parts in a complete model.
  - Start with making simple test runs.
  - (A good tactic when verifying and testing can be to eliminate temporary the stochastic element because a completely deterministic model is more easy to check)
- Criterion variables
  - Determine which outcomes of the model are crucial for the application; e.g. for a certain application travel time might be an important characteristic and the number of overtakings not that much.
- Sensitivity analysis
  - Determine which parameters are most important for the outcomes of the model.

  - Determine how variations in important parameters influence the outcomes. This is a so called sensitivity analysis. The results indicate how precise one should determine the parameters.
  - A comprehensive model very often has too many parameters (say > 25) to calibrate and validate them all. Most of them have to be based on existing knowledge and engineering judgement.

- Calibration
  - Compare outcomes from the model with results of observations in reality.
  - Change parameters to make the model reproduce reality as close as possible.

- Validation

Compare model outcomes with fresh real data using the parameters as found in the calibration. When the result of this is positive, the model is considered to be valid. However, one should keep in mind that validity is always with respect to a certain field of application. When more cases are used for validation, and the outcome is favourable, the validity of the model increases.

14.2.3 Use of a model

The carrying out of investigations with a simulation model as a tool resembles the process of collecting observations in real traffic. The task of designing appropriate runs with the model and analysing the results is often underestimated.

When investigating the effect of (say) 5 factors that can have 4 values, and being interested in 5 different traffic intensities with 2 different percentages of trucks, one has to do $5 \times 4 \times 5 \times 2 = 200$ runs. If the effect of randomness is large one might wish to do each simulation 10 times, which increases the number of runs to 2000. Even if the simulations are cheap, think about analysing the results of 2000 runs.

Consequently there is a need for using techniques from the field of 'design of experiments' to reduce the number of runs and the size of the analysis. With those techniques it is possible to choose a specific subset of runs (say 200) and almost get the same information as from 2000 runs.

An other solution, to process the outcomes of many runs, can be to carry out the analysis of the results directly by a computer program of which the simulation model is a module.

14.3 Description of simulation model FOSIM

14.3.1 General

The Transportation Laboratory of TU Delft has developed (since around 1988) - and is continuously extending and improving - a microscopic simulation model with the name FOSIM (Freeway Operations SIMulation). This model simulates the traffic operation on motorways and has been successfully applied at several types of road configurations: weaving areas; sections with on-ramps; and, sections with road works. See: [39], [47], [46], [48]. FOSIM is a microscopic simulation model, i.e. the behaviour of a traffic stream is imitated on the level of individual driver-vehicle elements. The simulation period is divided in time steps of 1 s (other time steps can also be used but in general 1 s is a suitable value). Each time step the positions and other dynamic characteristics of the vehicle are updated, starting with the most downstream vehicle. The new position of the vehicle depends on its previous position, the properties of the driver and the vehicle; the interactions with other road users; and, the characteristics

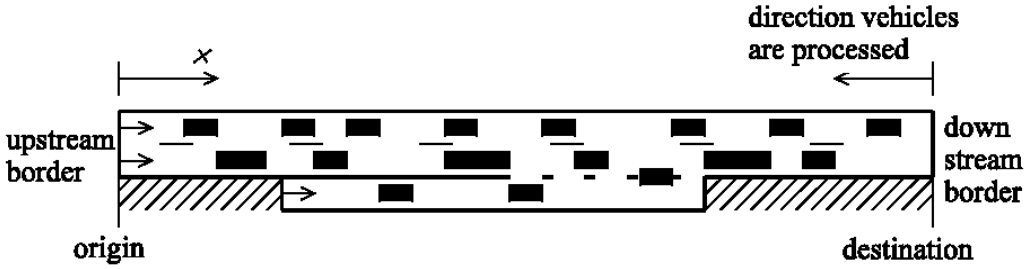

Figure 14.2: Schematisation of a roadway section with an on-ramp

(mainly geometrical properties) of the road. The new position is calculated taking into account the expected position after a few time steps, representing the anticipation of the drivers.

Using FOSIM it is possible to define a large variety of road configurations. A simple example is the roadway section with an on-ramp, depicted in Fig. 14.2. It is assumed that the road sections have no grades and the horizontal curves are so smooth that they do not influence vehicle behaviour. These assumptions are not very restrictive for Dutch motorways.

Several driver-vehicle elements can be distinguished. Their main characteristics are:

- vehicle length
- accelerative power
- declarative power
- free speed (or desired speed)
- parameters determining the car-following behaviour
- origin lane and destination lane
- When a vehicle enters the model, it is assigned a destination lane in a random way, taking into account the origin-destination matrix. Also the properties that are assigned to a driver-vehicle element (e.g. the desired speed) have some random elements. Consequently the repetition of a run, with a different random seed will produce a different outcome.

In general one has to repeat the simulation several times in order the get a reliable mean outcome and an impression of its variability (standard error). Intensities at the entrances can be varied per lane. The results of the simulation (intensities, speeds and densities) can be collected at detectors at the sections at arbitrary positions.

14.3.2 The vehicle process

The decision proses that each driver has to go through at each time step, is called the vehicle process; see Fig. 8.4. The first step is to compare the actual lane with the destination lane. If the destination lane and the actual lane are different, this can lead to a desired lane change, dependent on the position of the vehicle (if the vehicle is still far from its destination, the destination lane has little influence).

If a driver does not have to change lane to reach the destination, he/she can decide to overtake a slower vehicle in front. If a lane change is desired, FOSIM checks the possibilities to carry it out in the actual time step. If the outcome is positive, then the vehicle is placed on the new lane and its new position is calculated. If the lane change is not possible, then the vehicle becomes a car-follower.

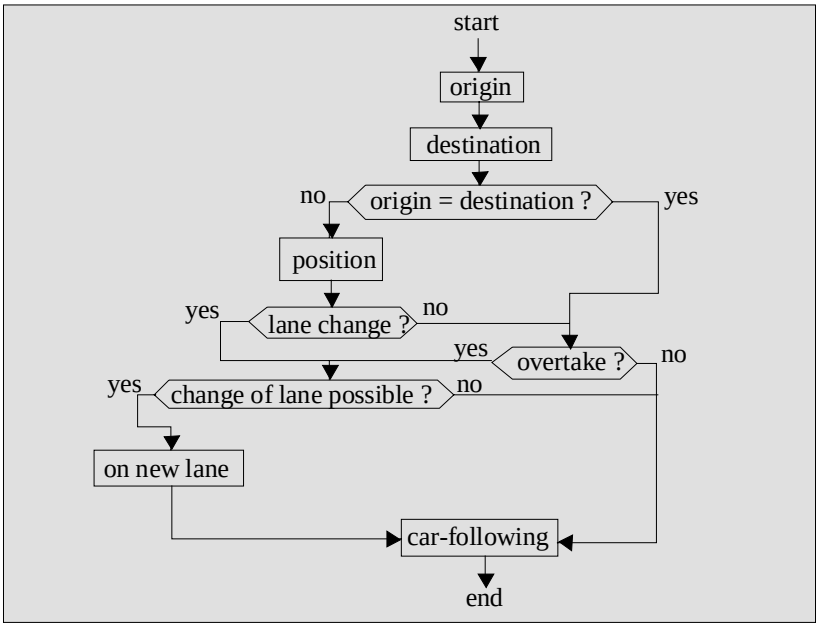

Figure 14.3: The vehicle process in FOSIM

14.3.3 The car-following model

In FOSIM each driver will always try to maintain or go back to his/her desired speed. If a vehicle in front is present (at a distance headway that is relevant), FOSIM calculates an acceleration that aims at reaching an acceptable distance headway.

Fig. 14.4 depicts the car-following situation; vehicle a is the leader and vehicle b the car-follower

The headway distance the vehicle is aiming at, is given by:

$$d_i = L_j + z_1 + z_2 f_j v_i + z_3 v_i^2 \tag{14.1}$$

with:

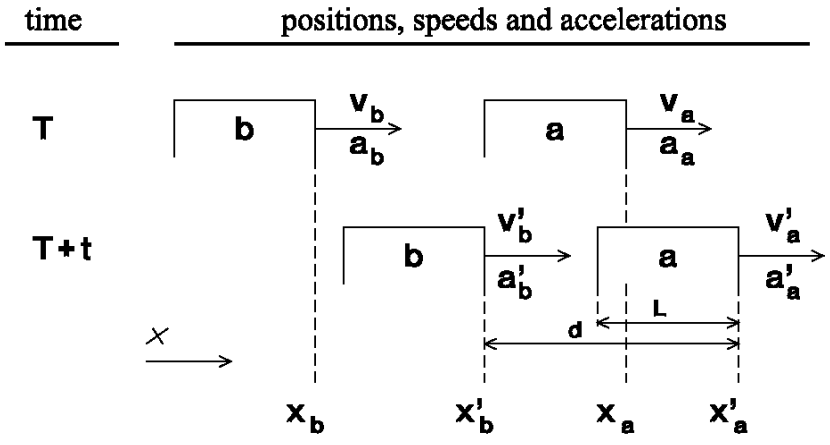

Figure 14.4: Two vehicles at succesive moments

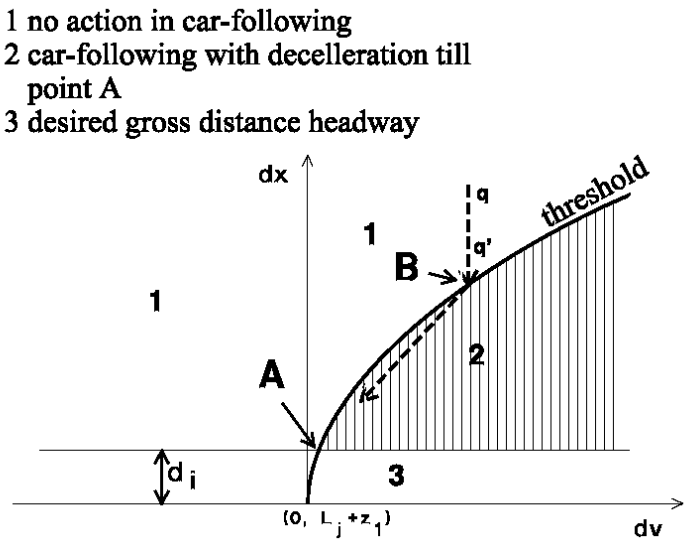

Figure 14.5: Anticipation threshold at approach of vehicle in front

i, j	$-$	index of vehicle i, j
d_i	m	target gross distance headway
L_j	m	length of leader
f_j	s	follower specific car-following parameter
v_i	m/s	speed of car-follower
z_1	m	general car-following parameter
z_2	m/m	general car-following parameter
z_3	s^2/s	general car-following parameter

If a vehicle catches up a vehicle in front along line $q-q\prime$ (see Fig. 14.5), then it will decelerate if the combination of relative speed and distance headway requires it.

The threshold function of FOSIM is given by:

$$dV = \left((dx - L_j - z_1)/50\right)^2 \tag{14.2}$$

where dV = speed of car-follower - speed leader, and dx = distance headway of car-follower. After crossing the threshold in point B, an acceleration is determined that will lead to the situation in point A. As an extra check the possibility of a collision is investigated and the acceleration will possibly be adapted to prevent that.

14.3.4 Lane change model

In FOSIM are distinguished: mandatory lane changes; desired lane changes; and, overtaking movements; see Fig. 14.6 and Fig. 14.7. The desired lane change between cross-sections a and b is carried out in order to get into a better starting position for the mandatory lane change between positions b and c. The driver is willing to take a greater risk at a mandatory lane change than at a desired lane change. This risk is translated in the deceleration a driver accepts to make the lane change possible.

In some cases, e.g. at weaving, FOSIM allows extremely short headways (these do occur in reality too) by halving the anticipation time.

In the case a *vehicle blocks an intended lane change*, the driver can change his speed a little to increase the possibility for the lane change in the next time step. Dependent on the driver type, this will be a speed increase (tactic 1) or a speed decrease (tactic 2). In case of tactic 1, the vehicle that blocks the lane change can possibly cooperate by decreasing its speed and, so to speak, give priority to the lane changer.

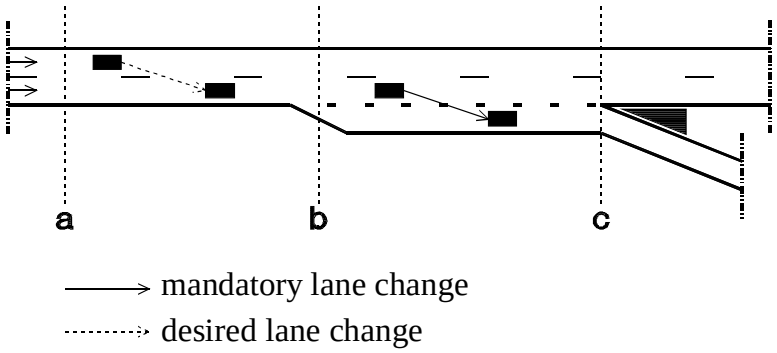

Figure 14.6: Mandatory and desired lane change

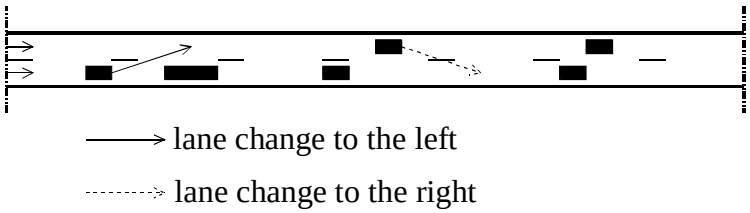

Figure 14.7: Lane changes at an overtaking

An *overtaking*, see Fig. 14.7, will primarily be carried out to maintain the desired speed or deviate less from it. The overtaking consists of two parts: a lane change to the left and one to the right. These are more or less independent manoeuvres.

A *lane change to the right* is carried out if there is no speed advantage in staying on the actual (non-right) lane. This will in general be the case if the vehicle drives at its desired speed.

14.3.5 Calibration and validation

Before a simulation model can be applied, one should be reasonable sure that the model resembles reality sufficiently for the relevant aspects. FOSIM has been calibrated for several configurations: weaving sections; sections with on-ramps; roadways with temporary lane reductions due to road works. The main aim was to use the model for capacity analyses and estimations. For that case the car-following behaviour is most decisive and consequently the focus of the calibration were the parameters of that process. Criteria for the calibration were:

- capacity
- distribution of roadway intensity over the lanes
- location of the start of congestion
- location of the lane changes
- mean speeds per lane

The outcome is that FOSIM is considered to be (sufficiently) valid for:

- weaving sections of the type 2+2 and 2+1 lanes;
- road works of the type 2→1, i.e. a two-lane roadway has one lane dropped over a given distance;

- roadway section with on-ramp.

In fact the calibration of a model such as FOSIM is an ongoing process. Recently the car-following behaviour module has been modified: the parameters now used are different for non-congested and congested traffic; see [16].

14.4 Application of FOSIM

In the Second Transport Structure Plan (SVV-II) several measures are proposed to make a Better Use of the Existing Infrastructure. Several of these measures are new, their effects are unknown and at least an estimation of the potential benefits is needed before they will be applied. The effects of these measures can be estimated by applying: general theory about traffic behaviour; studies with instrumented vehicles; studies with vehicle simulators in laboratories; and, last but not least simulation models.

Two measures have been studied: (with truck percentages of 5% en 10%)

- Introduction of no overtaking periods for trucks on motorways (truck overtaking ban);
- Prohibition for trucks to use motorway sections at periods congestion is frequent (truck ban).

Four scenarios have been investigated:

- One-lane on-ramp at a two-lane roadway
- Permanent lane reduction at a roadway from 3 to 2 lanes with the most left lane falling off
- Weaving section of the type 2+1 lanes
- Weaving section of the type 3+1 lanes

For each scenario the effect of the measure on the capacity has been investigated and per alternative situation 50 simulations were carried out; see [48] for more details.

14.4.1 Truck overtaking ban

Relevant differences between trucks and cars are:

- larger dimensions of the truck (vehicle width and length unto 18 m);
- lower speed (the official speed limit for trucks on motorways = 80 km/h, in reality truck speeds at free flow are 85 to 100 km/h);
- lower accelerative power.

Because differences of desired speed of trucks can be small, overtakings between trucks can last long and lead to temporary blocks for other vehicles. Large speed differences between cars and trucks can lead to sudden and unpredictable speed reductions of cars. To reduce the negative impact of trucks at busy hours the measure considered was to introduce no overtaking periods for trucks. This will result in platoons of trucks on the right lane of the roadway behind a slow vehicle (which could also be a slow car). The longer this situation persists, the longer will become the platoon; see Fig. 14.8.

A possible negative effect of a platoon of trucks is the temporary reduction of gaps for other vehicles to make a lane change. This could lead to a deterioration of traffic operation at on-ramps and weaving sections. Consequently the measure could have positive and a negative

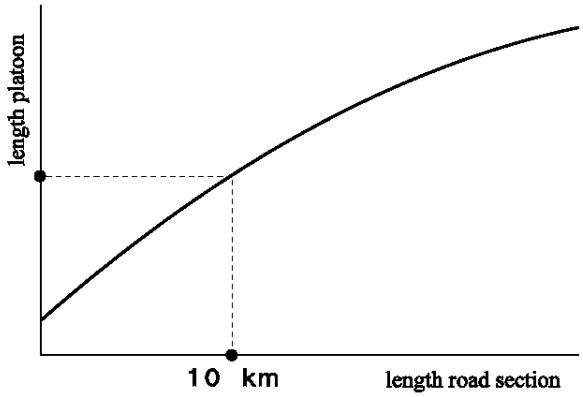

Figure 14.8: Hypothetical growth of platoon lengths as a function of the road section length

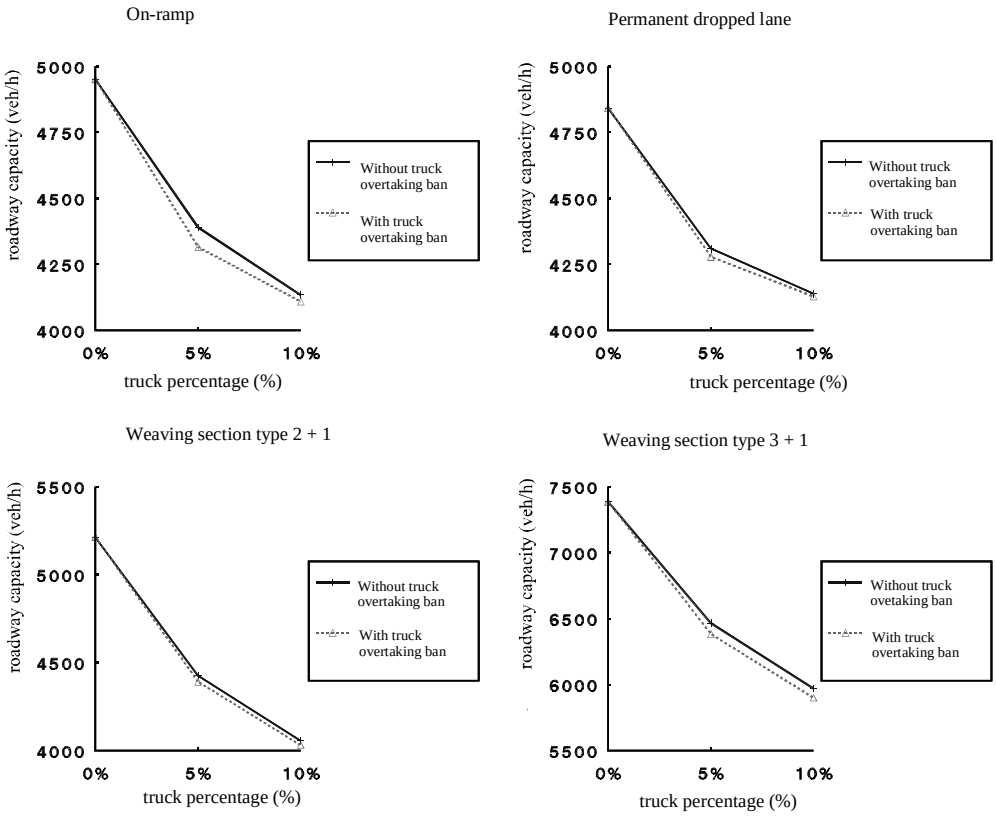

Figure 14.9: Capacity of roadway with and without truck overtaking prohibition

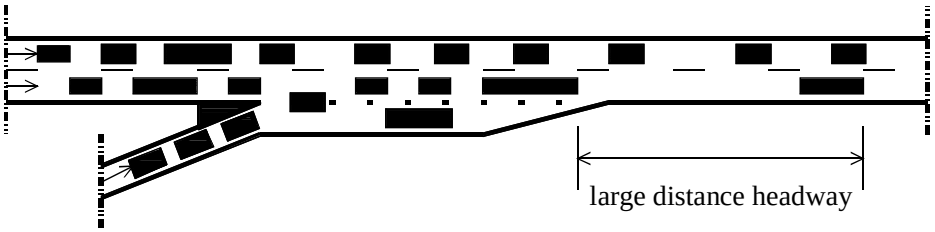

Figure 14.10: Accelerating from a congested region at an on-ramp

effects, both of unknown size. On theoretical grounds it is not possible to determine the overall effect. This is especially a situation in which a simulation model can be a useful tool.The results of the simulations are depicted in Fig. 14.9 . From this figure can be concluded that the effect on capacity of the truck overtaking ban is negative but rather small.

14.4.2 Truck ban

Because trucks have a much smaller accelerative power than cars, they can substantially decrease capacity in the case of congestion, when vehicles have to accelerate (compare with trucks after a stop at a traffic light). Fig. 14.10 illustrates this situation; trucks can have a large empty space in front in such conditions, which represents a loss.To circumvent this problem it has been proposed to forbid trucks to use busy road sections during peak periods.

Remark 66 *On the other hand special facilities for the benefit of truck traffic are under construction, e.g. the truck traffic lane at the Rotterdam orbital motorway.*

Simulations with FOSIM have resulted in effects, depicted in Fig. 14.11.

The general effect of the banning of trucks is substantial. At 5% trucks the capacity gain amounts to 7.0%-13.4% and at 10% trucks the increase is 14.9%-22.1%. The effect of the truck ban is smallest at the case with the lane reduction. This makes sense because the left lane, without truck traffic, is disappearing. For other cases the gains are approximately the same.

It can also be concluded that the capacity gain is relatively smaller at higher truck percentages; the law of diminishing returns seem applicable here. Inspecting the effects per lane, it is evident that the effect is largest at the right lane, mostly used by trucks.

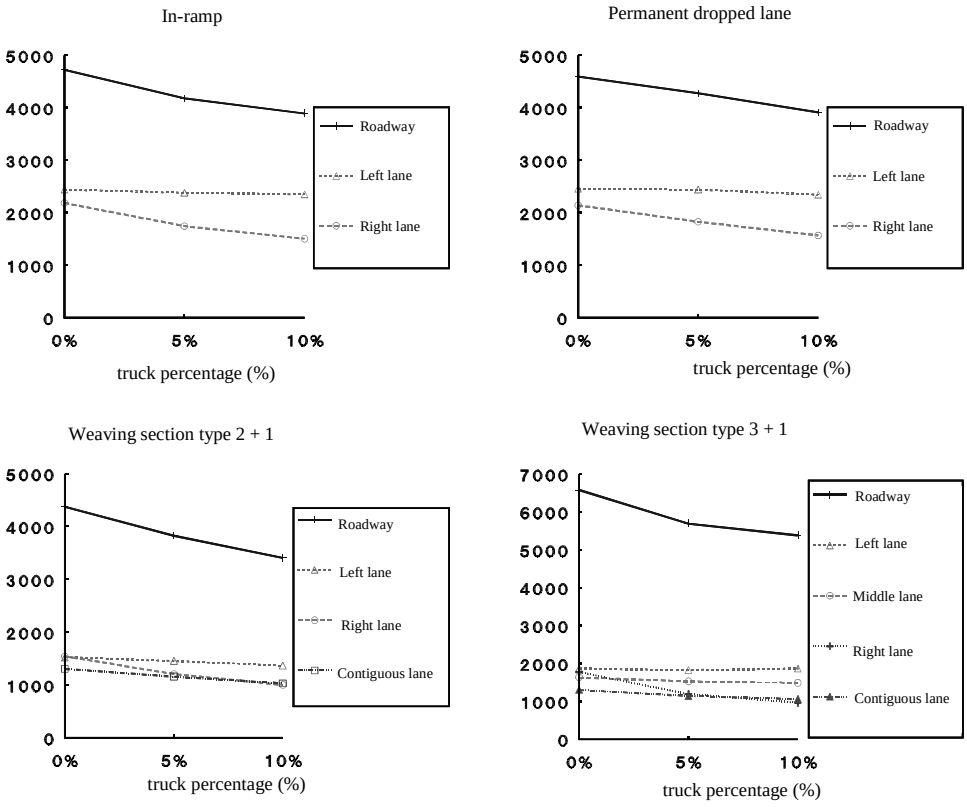

Figure 14.11: Queue discharge rate of roadway per type of roadway section

Chapter 15

Macroscopic simulation

The final chapter of this reader discusses macroscopic simulation of traffic in general, and in particular the METANET model. METANET is a model for predicting traffic conditions on a motorway network. Besides the macroscopic traffic flow model describing what happens on the motorway links, METANET also describes route choice behaviour as well as the influence of dynamic traffic management measures.

It is emphasized that, on the contrary to microscopic models such as FOSIM and AIMSUN2, METANET is not a stochastic model but a deterministic model. As a result, it is of no use to rerun the same simulation more than one time. Instead, the robustness of the simulation results with respect to the simulation input should be considered. This particularly to the traffic demands on the network, but also on the parameters that are used to predict the traffic flows through the network,

15.1 Traffic processes in METANET

In METANET, the motorway network is represented by a directed graph that is described in the filename.nwd file. More precisely, bifurcations, junctions and on/off-ramps are represented by the nodes of the graph whereas the motorway stretches between these locations are represented by the links. The two directions of a motorway stretch are modelled as separate links with opposite directions. Inside each link we suppose homogeneous geometric characteristics. An inhomogeneous motorway stretch may be represented by two or more consecutive links separated by nodes at the locations where the change of geometry occurs. At the bounds of the network, origin or destination links are added where traffic enters or leaves, respectively, the simulated network part.

METANET considers five types of links: normal motorway links, dummy links, store-and-forward links, origin links, and destination links. In this short manual, we will only consider motorway links, origin links and destination links.

15.1.1 Normal motorway links

The simulation of traffic behaviour in the motorway links is based on a macroscopic modelling approach with the traffic variables density k (veh/km/lane), mean speed u (km/h), and traffic volume (or flow) q (veh/h). Each link is subdivided in segments with typical lengths of 300 to 800 meters. Starting with some initial values, the time evolution of every traffic variable in every segment is calculated by means of difference equations and algebraic relationships, which are calculated for every simulation time step (typically every 10s).

The following equation describes the dynamic of traffic on a motorway link. The outflow $q_{m,i}(j)$ of each segment i of link m during time period j is equal to the density multiplied by the mean speed and the number of lanes λ_m

$$q_{m,i}(j) = \lambda_m k_{m,i}(j)\, u_{m,i}(j) \tag{15.1}$$

The dynamics of the density on a segment (cell) is defined by the conservation of vehicle equation, given by

$$k_{m,i}(j+1) = k_{m,i}(j) + \frac{T}{L_m \lambda_m}\left[q_{m-1,i}(j) - q_{m,i}(j)\right] \tag{15.2}$$

where T is the time step. Note that the discretisation of the flow is a rather rough approximation of the actual flow between two cells, since it only depends on the transmitting segment i.

The speed dynamics are based on the Payne model. METANET uses

$$\begin{aligned} u_{m,i}(j+1) &= u_{m,i}(i) + \frac{T}{\tau}\left[V\left(k_{m,i}(j)\right) - u_{m,i}(i)\right] && (15.3)\\ &+\frac{T}{L_m} u_{m,i}(k)\left[u_{m,i-1}(j) - u_{m,i}(j)\right] && (15.4)\\ &-\frac{\vartheta T}{\tau L_m}\frac{k_{m,i+1}(j) - k_{m,i}(j)}{k_{m,i}(j) + \kappa} && (15.5) \end{aligned}$$

where τ, ϑ and κ are model parameters and where the fundamental diagram is given by the following expression

$$V(k) = V_m^0 \cdot \exp\left[-\frac{1}{a_m}\left(\frac{k}{k_m^c}\right)^{a_m}\right] \tag{15.6}$$

where a_m is a link-specific model parameter and where V_m^0 denotes the free speed of the link; k_m^c is the link specific critical density.

For steady-state and space-homogeneous conditions, this value becomes identical with the current mean speed of the segment. The shape of the fundamental diagram is determined by the parameters V_m^0, k_m^c and the parameter a_m. The exponent a_m has no direct physical significance and is therefore substituted by other variables in the user input.

15.1.2 Origin links and destination links

Origin links receive the corresponding user-specified demand. Their outflow may be limited by the link capacity and/or downstream congestion and/or traffic lights (in case of ramp metering). If for any of these reasons, the link demand exceeds the link outflow, a queue will be formed. Partial queues and composition rates are also present in this kind of links to represent traffic with different destinations. A simple queuing model is used to describe the traffic operations at the origin links.

The destination links represent the locations from which traffic leaves the simulated network part, in other words the exits of the network. Occasionally, traffic flow at the exits of the network may be influenced by the traffic conditions in downstream stretches (e.g. spill back of congestion). In such cases, if traffic densities at the destination links are available, they may be used to limit the outflow according to the downstream traffic conditions.

15.1.3 Discontinuities

METANET includes different models describing the effect of on ramps and lane drops. In order to account for the speed drop caused by merging (in case of on on-ramp), the following term is added to the speed-dynamic equation

$$-\frac{\delta T q_{on-ramp}(j)\, u_{m,1}(j)}{L_m \lambda_m \left(k_{m,1}(j) + \kappa\right)} \tag{15.7}$$

assuming that the on-ramp 'attaches' to the first segment of link m.

Where there is a lane drop, the speed reduction due to the lane drop is described by the following term which is added to the speed dynamic equation

$$-\frac{\phi T \Delta\lambda k_{i,N_m}(j)\, u_{m,N_m}^2(j)}{L_m \lambda_m k_m^c} \tag{15.8}$$

where $\Delta\lambda = \lambda_m - \lambda_{m+1}$ denotes the lane drop amd where $\phi > 0$ is a model parameter.

15.1.4 Modelling of the Network Nodes

Motorway bifurcations and junctions (including on- and off-ramps), are, as already mentioned, represented by the nodes of the model. Generally speaking, traffic enters a node n through a number of input links and is distributed to a number of output links. For proper calculation of the distribution, the destinations of the inflows must be considered. The partial flow with a certain destination in an output link is calculated according to the total incoming traffic bound to this destination and according to the portion of users who choose the corresponding output link in order to travel to their destination. This portion of users is the splitting rate and describes the average route choice behaviour of the drivers at the network nodes. As already mentioned, splitting rates may be provided in several ways including the possibility to simulate the impact of variable message signs and/or route guidance measures.

Before describing the different input needed to run a METANET simulation study, let us first consider the different phases that are relevant when performing a simulation. In this short manual, we assume that METANET will be used for control scenario assessment purposes. That is, for a certain network, we aim to cross-compare different ITS scenarios. We will distinguish two main phases: preparation METANET and performing the study. The remainder provides a coarse checklist, but does not go into details.

15.2 Preparing a METANET simulation study

In chapter IV we have already outlined the different steps that are important when doing a simulation study. This section discusses some specific issues that are important when doing a METANET study, namely setting up of the simulation study and application of the model. However, another important issue is the decision to use METANET rather than some other (microscopic) simulation model. This depends on the required scale and level-of-detail: METANET can handle large network and do so with relatively low computation times. Furthermore, the data requirements are relatively low. Being a macroscopic model, the level-of-detail is only moderate. On top of this, the METANET model is macroscopic both in the representation of the traffic flow and in its description of te traffic processes. Requiring the speed-density curve as an input (and a way to describe the effects of incidents, roadworks, and Dynamic Traffic Managament measures), effects of changes in (bottle-neck) geometry, traffic control, etc. need to be determined either from empirical data or using another traffic flow model.

15.2.1 Determine and acquire required input parameters and variables

METANET has no true Graphical User-Interface. Hence, to do a simulation study, different input files need to be prepared using a text editor. Furthermore, METANET has two simulation modes:

1. non-destination oriented mode (no route choice) and
2. destination oriented mode (route choice).

In destination oriented mode, METANET explicitly distinguished traffic with different destinations and their route choice. In non-destination oriented mode, no distinction is made, and hence route choice cannot be modelled.

It is beyond the scope of this reader to discuss in detail all these input files. We refer to the METANET manual for more information.

Filename	Destination	Non-destination	Description
Filename.nwd	x	x	Description of network topology (origins, destinations, motorway links, nodes)
Filename.ctr	x	x	Simulation control file, describing run-time parameters
Filename.msd	x	x	Traffic measurement file, describing the data collected at various points in the network
Filename.odm	x	-	Dynamic Origin-Destination information
Filename.spl	x	-	Splitting fractions (not required when dynamic assignment is used)
Filename.trn	-	x	Turning rates
Filename.cpa	x	x	Control Parameters file, describing the parameter settings of the ITS measures (not required)
Filename.evt	x	x	Event and incident file, describing the time and duration of incidents and ITS operation.

Table 15.1: Input files needed for doing METANET simulation study

Appendix A

Green's theorem and its applications

In section 2.6 we have introducted the cumulative vehicle count (also local stream function) $\tilde{N}(x,t)$ can be thought of as labelling the vehicles on the roadway that pass cross-section x until time instant t. Fig. A.1 illustrates this in case there is passing (vehicle 7 passes vehicle 8, causing their indices to interchange).

Theorem 67 *(Green's Theorem) - The line integral of the projection of the gradient along any curve joining two points is independent of the curve, and that the result equals the difference in $\tilde{N}$ at the extremes of the curve.*

In our case, we consider a curve C in the (x,t) plane between two points (x_1,t_1) and (x_2,t_2). According to Green's theorem, we have

$$\tilde{N}(x_2,t_2) - \tilde{N}(x_1,t_1) = \int_C \frac{\partial \tilde{N}}{\partial t}dt + \frac{\partial \tilde{N}}{\partial x}dx = \int_C qdt - kdx \quad \text{for all curves } C \tag{A.1}$$

The validity of Eq. (A.1) is understood easily by parameterizing the curve C, i.e.

$$C = \{(x(s), t(s)) \mid \text{with } s_1 \le s \le s_2 \text{ s.t. } (x(s_i), t(s_i)) = (x_i, t_i)\} \tag{A.2}$$

Along the curve, we have

$$\frac{d\tilde{N}}{ds} = \frac{\partial \tilde{N}}{\partial t}\frac{dt}{ds} + \frac{\partial \tilde{N}}{\partial x}\frac{dx}{ds} \tag{A.3}$$

and thus, we find that

$$\tilde{N}(s_2) - \tilde{N}(s_1) = \int_{s_1}^{s_2} \frac{d\tilde{N}}{ds}ds = \int_C \frac{\partial \tilde{N}}{\partial t}dt + \frac{\partial \tilde{N}}{\partial x}dx \tag{A.4}$$

By considering a *closed curve* C, we can determine the following form of the conservation equation:

$$\int_C qdt - kdx = 0 \quad \text{for all curves } C \tag{A.5}$$

The main benefit of Eq. (A.5) rather than the p.d.e. is that the equation is also valid if q and k are discontinuous.

Eqns. (A.1) and (A.5) have many applications. Consider for instance an observer travelling a distance L in a period of length T. Consider the number of cars that the observer passes or that pass the observer, assuming that the traffic conditions are stationary with flow q and density k. To determine this number, we can apply eqn. (A.1) by choosing a curve from (x_1,t_1) to (x_1+L, t_1+T). For the sake of simplicity, we choose a curve that travels from (x_1,t_1) to (x_1,t_1+T) and then to (x_1+L,t_1+T). We then have

$$n_{obs} = \int_C qdt - kdx = \int_{t_1}^{t_1+T} qdt - \int_{x_1}^{x_1+L} kdx = qT - kL \tag{A.6}$$

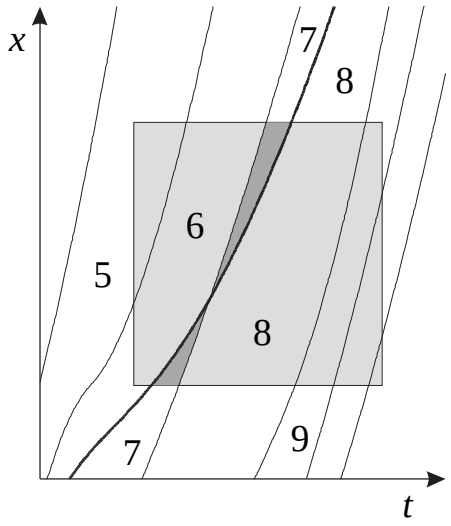

Figure A.1: Values of the cumulative flow function (adopted from [14])

which thus equals the number of vehicles passed by the moving observer. By rearranging the terms in this expression, we can show easily that the relative flow q_{obs} measured by the moving observed equals

$$q_{obs} = \frac{n_{obs}}{T} = q - k\frac{L}{T} = q - kv_{obs} \tag{A.7}$$

where v_{obs} is the speed of the moving observer.

A.1 Alternative analytical solution scheme to the kinematic wave model

The approach discussed in this section is still rather qualitative, and actually determining an analytical solution will be quite painstaking. This section recalls the procedure from [14] to construct analytical solutions of the kinemaic wave model. The approach is based on first determining the vehicle trajectories and determining the cumulative flow function $\tilde{N}(x,t)$ from these trajectories.

The thin lines of Fig. A.2 depict a family of characteristics (emanating from the line $t = 0$) corresponding to a hypothetical infinite roadway problem. We again assume that the traffic state $k(x,0)$ is known at $t = 0$. Let $k\left(x_0(z), t_0(z)\right) = k_0(z)$. In the previous section we have already discussed the equations of the z-characteristic (i.e. the characteristic emanating from the $(x_0(z), t_0(z))$). We showed that the wave had slope $c(k) = Q'(k)$.

Since the intensity and the density are constant on the characteristic, we can use eq. (1.6) to evaluate the change in vehicle number seen by an observer travelling along the characteristic, i.e. with velocity $v_o = \tilde{c}\left(z\right)$, during a time interval Δt.

$$\Delta\tilde{N} = \left(q - kv_o\right)\Delta t = \left[Q\left(k_0(z)\right) - k_0(z)\tilde{c}(z)\right] = \Delta\left(z\right)\Delta t \tag{A.8}$$

where the last equality simplie introduces $-\Delta\left(z\right)$ as shorthand notation for the factor in brackets; this symbol represents the *known* fixed rate at which the vehicle number increases with time along the z-characteristic.

If the boundary data are specified in terms of density, $k_0\left(z\right)$, we first need to find the vehicle number $\tilde{N}_0\left(z\right)$ along the boundary for all z. This can be done by rewriting the conservation law eq. (1.3) in parametric form, using for C the portion of the boundary from $z = 0$ to $z = z$

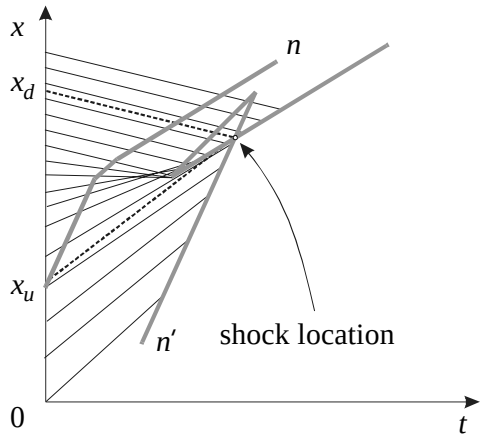

Figure A.2: Map of characteristic displaying the trajectories of vehichles n and n' $(n' > n)$

(end of the region of interest). If we define $N_0(0) = 0$, then the result is

$$\tilde{N}_0(z) - \tilde{N}_0(0) = \int_0^z \left[Q\left(k_0\left(\zeta\right)\right) \frac{dt}{d\zeta} - k_0\left(\zeta\right) \frac{dx}{d\zeta} \right] d\zeta \tag{A.9}$$

which reduces to

$$\tilde{N}_0\left(z\right) = \int_0^z -k_0\left(\zeta\right) d\zeta \tag{A.10}$$

for the initial value problem (because the boundary satisfies $dt/d\zeta = 0$, and $dx/d\zeta = 1$ in this case).

In any cae, once $\tilde{N}_0\left(z\right)$ and $\Delta\left(z\right)$ have been determined, one can express the vehicle number at time s on the z-characteristic, denotes by $S(s;z)$ as follows

$$S\left(s;z\right) = \tilde{N}_0\left(z\right) + \Delta\left(z\right)\left[s - t_0\left(z\right)\right] \tag{A.11}$$

Then, on setting $S\left(s;z\right)$ equal to an arbitrary n and solving it for s, one obtains the time at which vehicle n crosses the z-characteristic

$$s = t_0\left(z\right) + \left[n - \tilde{N}_0\left(z\right)\right] / \Delta\left(z\right) \tag{A.12}$$

This equation can be evaluated for every characteristic shown in a sketch such as Fig. A.2. If the point on each characteristic corresponds to the calculated time is darkend, then the smooth darked line connecting the pints will be the trajectory of the n-th vehicle. Thus, we can view characteristics, not just as waves that tell every car what spacing and speed to adopt, but also when and where they should be.

If the n-th vehicle traverse a region in the (x,t) plane where the characteristics do not intersect as occurs for vehicle n in Fig. A.2, then the result of the procedure identifies the actual vehicle trajectory. As n is increased however, characteristics may begin to intersect and it may no longer be clear to which signal a vehicle should respond as it traverses such a region. In this case the purely mathematical recipe [14] describes returns a parametric trajectory that intersects all the characteristics but also exhibits some loops, as shown by curve n' of Fig. A.2. Although such a curve is mathematically valid, it cannot be a real vehicle trajectory and something must be done to fix the problem.

Note that if we remove the loop from Fig. A.2 then the remaining portion of the curve is a valid and physically meaningful vehicle trajectory in that the vehicle speed (and the flow and density along its trajectory) is at all times equal to that demanded by the pertinent characteristic. Thus, points on the physically relevant trajectory still satisfy the conditions of the

kinematic wave model. To obtain this result, we had to avoid multi-signals, i.e. crossings of the characteristics. This was achieved by ignoring all signals that made the loop (i.e. those issued between x_u and x_d). The solution then includes a sudden vehicular deceleration at the point P of contact with the loop.

Note as well that the characteristics issued at x_u and x_d cannot effect the following vehicles, since they cross at point P, and must therefore terminate there. In the complete solution with all the vehicle trajectories, the collection of all points such as P (where trajectories bend and characteristics terminate) defined curves which have been called *shocks* in the previous sections.

Without further proof, we note that when $Q(k)$ is concave, the family of *active (truncated) characteristics* covers the whole region of the (x,t) plane wihout crossing shocks or one another, i.e. *there is one and only one characteristic passing through every point that is not on a shock.* This fundamental property can be expressed in two equivalent ways:

1. Every point in the solution can be connected in a unique way to the boundary though a truncated characteristic that does not cross any shocks or other truncated characteristics.

2. 'Entropy' condition: truncated characteristics run into a shock.

From (1) and (2) it follows that shocks can merge into a combined shock but cannot diverge. Furhtermore, once a shock has formed, e.g., at time $t_s > 0$ the solution must containa shock at any future time $t_f > t_s$.

Daganzo [14] notes that the collection of all the real (loopless) trajectories identifies the physically relevant solution to our probkem in its entirety in that they give a 'topographical map' of $\tilde{N}(x,t)$. For more details on the approach, as well as yet other alternative approaches to determine analytical solutions, we refer to [14].

In practise, the approach can be implemented very easily in a spreadsheet. This requires the following steps:

1. Determine the x-locations from which the z-characteristics emanate at $t = 0$, i.e. $z_i = i\Delta z$, for $i = 0, ..., m$ where $\Delta z = L/m$. m denotes the number of characteristics used to determine the solutions.

2. For all i, determine the initial density $k_0(z_i)$ at $x_i = z_i$ and $t = 0$.

3. For each characteristic, determine the rate $\Delta(z_i)$ at which the vehicle count increase along the z_i-characteristic. This is done using the equation $\Delta(z_i) = Q(k_0(z_i)) - c(k_0(z_i)) \cdot k_0(z_i)$.

4. Determine the number of vehicles at $t = 0$ for each characteristic z_i by numerically solving $\tilde{N}_0(z_i) = -\int_0^{z_i} k_0(\zeta)\, d\zeta$.

5. The trajectory of vehicle n is given by the points (x_i, t_i) where $t_i = \left[n - \tilde{N}_0(z_i)\right] / \Delta(z_i)$ and $x_i = z_i + c(k_i)\, t_i$

The remaining problem is to remove the loops. How to solve this problem is left to the reader. Fig. A.3 shows application of the procedure to describe the acceleration of vehicles from a queue of $200m$. The results resulting from the approach of Daganzo are similar to the approach described in the previous section.

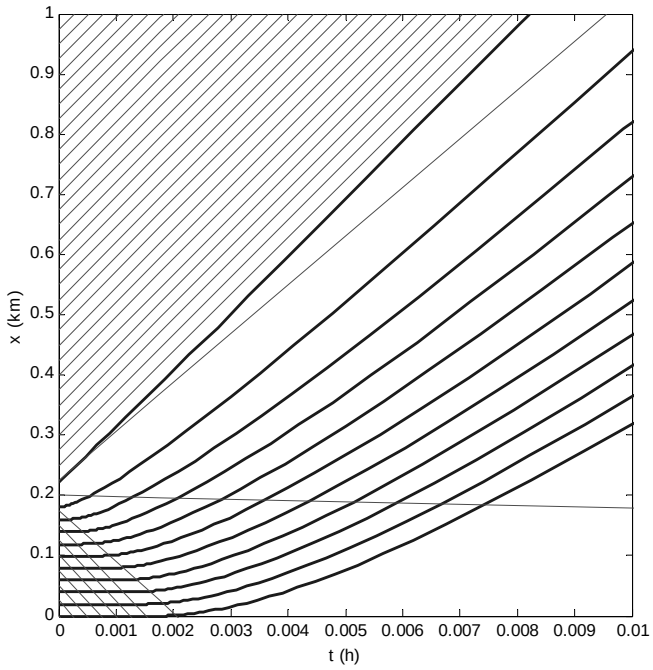

Figure A.3: Vehicles driving away from standing queue using Greenshields intensity-density curve. Thick lines reflect trajectories; thin lines are kinematic waves. Notice how the procedure predicts that the instantaneously accelerating first vehicle sends out an acceleration fan, causing the following vehicles to smoothly accelerate from the queue.

Appendix B

Mathematical and numerical analysis of higher-order models

This appendix discusses the mathematical properties as well as a numerical solution approach for the Payne model, which consist of the conservation of vehicle equation

$$\frac{\partial k}{\partial t} + \frac{\partial (ku)}{\partial x} = 0 \tag{B.1}$$

and the speed dynamics equation

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \frac{V(k) - u}{\tau} - \frac{c^2(k)}{k}\frac{\partial k}{\partial x} \tag{B.2}$$

where $V(k) = Q(k)/k$ denotes the equilibrium speed.

B.1 Riemann variables

To study the behaviour of the Payne model, this form of the model is not convenient: the conservation equation depends on the speed, while the speed dynamics depend on the density. To mathematically analyse the model, a different form of the model is prefered, which is referred to as the *characteristic form*. To derive this characteristic form, we first the equations in their *quasi-linear form*. Let the vector $w = (k, u)$; then the following equation is equivalent to Eqns. (B.1) and (B.2):

$$\frac{\partial}{\partial t}\begin{pmatrix} k \\ u \end{pmatrix} + \begin{pmatrix} u & k \\ \frac{c^2(k)}{k} & u \end{pmatrix}\frac{\partial}{\partial x}\begin{pmatrix} k \\ u \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{V(k)-u}{\tau} \end{pmatrix} \tag{B.3}$$

$$\frac{\partial w}{\partial t} + A(w)\frac{\partial w}{\partial x} = B(w) \tag{B.4}$$

The aim is to *diagonalise* the matrix $A(w)$. This can be achieved easily by determining the eigenvalues[1] and the eigenvectors[2] of $A(w)$. Some straightforward computations show that the eigenvalues $\lambda_{1,2}$ of the matrix $A(w)$ are equal to

$$\lambda_1 = u + c(k) \tag{B.5}$$

and

$$\lambda_2 = u - c(k) \tag{B.6}$$

[1] The eigenvalues λ of the matrix A are defined by those values satisfying $\det(A - \lambda I) = 0$.

[2] An eigenvector e of the matrix A stemming from the eigenvalue λ is *any vector* that is in the null-space $\mathcal{N}(A - \lambda I)$ of the matrix $A - \lambda I$, i.e. e satisfies $(A - \lambda I)e = 0$.

The corresponding eigenvalues are $e_1 = \left(\frac{k}{2c(k)}, \frac{1}{2}\right)$ and $e_2 = \left(-\frac{k}{2c(k)}, \frac{1}{2}\right)$. Let us now define the matrix of eigenvectors S

$$S = \begin{pmatrix} \frac{k}{2c(k)} & -\frac{k}{2c(k)} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \quad \text{with} \quad S^{-1} = \begin{pmatrix} \frac{c(k)}{k} & 1 \\ -\frac{c(k)}{k} & 1 \end{pmatrix} \tag{B.7}$$

Using these matrices S and S^{-1}, we can determine the so-called *compability equations* by left-multiplying the eqn. (B.4), we get

$$S^{-1}\frac{\partial w}{\partial t} + S^{-1}A(w)\frac{\partial w}{\partial x} = S^{-1}B(w) \tag{B.8}$$

Let us now consider the so-called Riemann variables z as follows. Consider any variation δw (either $\partial w/\partial t$ or $\partial w/\partial x$) in w and defined

$$\delta z = S^{-1}\delta w \tag{B.9}$$

Inversely, one has

$$\delta w = S\delta z \tag{B.10}$$

Note that these definition imply

$$S\frac{\partial z}{\partial t} = \frac{\partial w}{\partial t} \quad \text{and} \quad S\frac{\partial z}{\partial x} = \frac{\partial w}{\partial x} \tag{B.11}$$

which means that the compability equation (B.8) can be written as follows:

$$S^{-1}S\frac{\partial z}{\partial t} + S^{-1}A(w)S\frac{\partial z}{\partial x} = \frac{\partial z}{\partial t} + \Lambda(w)\frac{\partial z}{\partial x} = \tilde{B}(z) \tag{B.12}$$

where $\Lambda(w) = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} = \begin{pmatrix} u + c(k) & 0 \\ 0 & u - c(k) \end{pmatrix}$ and $\tilde{B}(z) = S^{-1}B(w)$.

If we consider the definition of the Riemann (or characteristic) variables, we see that

$$\delta z_1 = \delta v + \frac{c(k)}{k}\delta k \tag{B.13}$$

and

$$\delta z_2 = \delta v - \frac{c(k)}{k}\delta k \tag{B.14}$$

If we assume that the wave speed $c^2(k) = c_0^2$, when we can determine by simple integration that the Riemann variables are equal to

$$z_1 = \int \delta v + c_0\frac{\delta k}{k} = v + c_0 \ln k \tag{B.15}$$

$$z_2 = \int \delta v - c_0\frac{\delta k}{k} = v - c_0 \ln k \tag{B.16}$$

For the Payne model, were $c^2(k) = -\frac{1}{2\tau}\frac{dV}{dk}$, the Riemann variables depend on the precise shape of the fundamental diagram. For instance, for Greenshield's function, we have $c^2(k) = \frac{1}{2\tau}a$, where a is the slope of the linear speed density curve.

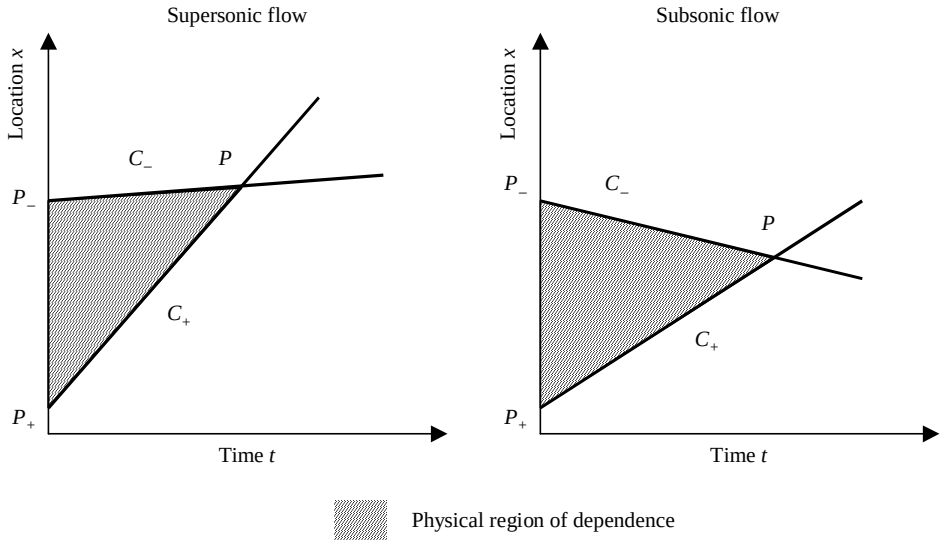

Figure B.1: Propagation of flow quantities.

B.2 Propagation of disturbances

The decoupling applied in the previous section, yielded the following decoupled partial differential equation for the Riemann variables $z_{1,2}$

$$\frac{\partial z_1}{\partial t} + \lambda_1 \frac{\partial z_1}{\partial x} = \tilde{B}_1(z) \tag{B.17}$$

$$\frac{\partial z_2}{\partial t} + \lambda_2 \frac{\partial z_2}{\partial x} = \tilde{B}_2(z) \tag{B.18}$$

These equations show that the quantities $z_{1,2}$ propagate along the corresponding characteristics with speed $\lambda_{1,2} = (u \pm c(k))$. Thus, $\delta z_1 = \delta v + \frac{c(k)}{k}\delta k$ propagates along the characteristic C_+ defined by $\frac{dx}{dt} = (u + c(k))$; $\delta z_2 = \delta v - \frac{c(k)}{k}\delta k$ propagates along the characteristic C_- defined by $\frac{dx}{dt} = (u - c(k))$.

Similar to applying the method of characteristics to the LWR model, the method of characteristics can be applied to the higher-order model discussed here. If we consider any point $P(x,t)$, we know that the density and speed at this point are determined from the quantities $z_{1,2}$ transported with speed $(u \pm c(k))$ along the characteristics, that is

$$(z_1)_P = (u + c_0 \ln k)_P = (u + c_0 \ln k)_{P_+} = (z_1)_{P_+} \tag{B.19}$$

$$(z_2)_P = (u - c_0 \ln k)_{P_-} = (u - c_0 \ln k)_{P_-} = (z_2)_{P_-} \tag{B.20}$$

The left side of Fig. B.1 is drawn for the case of supersonic flow, while for a subsonic flow, the C_- characteristic has a negative slope and one has the situation shown on the right side of the figure. Each point P in the (x,t) plane is reached by one characteristic of each type.

For the LWR model, we have seen that when $\frac{\partial u}{\partial x} < 0$ (speed decreases over space), that shocks will occur. This will generally also in the higher-order models: when $\frac{\partial (u+c)}{\partial x} < 0$ one can have the situation depicted in Fig. B.2, where the C_+ characteristic from P_{1+} intersects the C_+ characteristic from P_+, and hence multi-valued quantities would occur in P_1; that is, one would have

$$(u + c_0 \ln k)_{P_1} = (u + c_0 \ln k)_{P_+} \tag{B.21}$$

and

$$(u + c_0 \ln k)_{P_1} = (u + c_0 \ln k)_{P_{+1}} \tag{B.22}$$

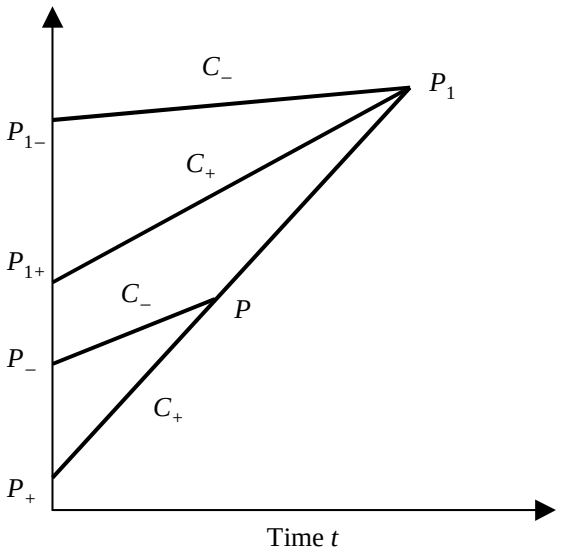

Figure B.2: Intersecting characteristics with $\frac{\partial}{\partial x}(u+c) < 0$.

where

$$(u + c_0 \ln k)_{P_+} \neq (u + c_0 \ln k)_{P_{+1}} \tag{B.23}$$

This impossible situation leads to a discontinuity in the flow that we refer to as a shock wave.

As mentioned earlier, the Payne model has two families of characteristics, while the LWR model only has one. As a result, the Payne model has *two families of shock and expansion waves*: the first family is associated with the first characterstic C_+, while the second is associated with the second characteristic C_-.

B.3 Instabilities and jam formation

It is clear that the Payne model allows homogeneous and stationary equilibrium solutions (k_e, u_e), which satisfy $u_e = V(k_e)$. We will now examine the stability of this solution. This is achieved by considering small disturbances $\delta k(x,t), \delta u(t,x)$ to the solution. Let us furthermore assume that

$$k(x,t) = k_e(x,t) + \delta k(x,t) \text{ and } u(x,t) = u_e(x,t) + \delta u(x,t) \tag{B.24}$$

is a solution of the Payne model. If we assume that $\delta k(x,t)$ and $\delta u(x,t)$ are small, we can linearise Eqns. (B.1) and (B.2) resulting in linear partial differential equations for $\delta k(x,t)$ and $\delta u(x,t)$

$$\frac{\partial(\delta k)}{\partial t} + u_e \frac{\partial(\delta k)}{\partial x} = k_e \frac{\partial(\delta u)}{\partial x} \tag{B.25}$$

and

$$\frac{\partial(\delta u)}{\partial t} + u_e \frac{\partial(\delta u)}{\partial x} = -\frac{c^2(k_e)}{k_e}\frac{\partial(\delta k)}{\partial x} + \frac{1}{\tau}\left[\frac{dV}{dk}\delta k - \delta u\right] \tag{B.26}$$

Using simple wave analysis, we can write the solutions $\delta k(x,t)$ and $\delta u(x,t)$ in terms of the following series

$$\delta k(x,t) = \int_s \sum_j k_j^*(s) \exp\left(i \cdot sx + \left[\lambda_j(k_e,s) - i \cdot \omega_j(k_e,s)\right] t\right) ds \tag{B.27}$$

and

$$\delta u(x,t) = \int_s \sum_j u_j^*(s) \exp\left(i \cdot sx + \left[\lambda_j(k_e,s) - i \cdot \omega_j(k_e,s)\right] t\right) ds \tag{B.28}$$

where $i = \sqrt{-1}$. In fact, we assume that the disturbances δk and δu can be decomposed into simple waves, traveling with a certain speed ω, while its amplitude increases / decreases according to λ_j. Let us now consider one simple wave solution

$$\delta k\left(x,t\right) = k^* \exp\left(i \cdot sx + \left[\lambda - i \cdot \omega\right]\right) \tag{B.29}$$

$$\delta u\left(x,t\right) = u^* \exp\left(i \cdot sx + \left[\lambda - i \cdot \omega\right]\right) \tag{B.30}$$

Note that s denotes the wave length, ω denotes the wave's frequency, and λ the wave's damping rate; k^* and u^* are amplitudes. If we substitute Eqns. (B.29) and (B.30) into Eqns. (B.25) and (B.26), we find the following set of algabraic equations

$$M \left(\begin{array}{c} k^* \\ u^* \end{array} \right) = 0 \tag{B.31}$$

with

$$M = \left(\begin{array}{cc} -\tilde{\lambda} & -i \cdot sk_e \\ -\frac{i \cdot s}{k_e} c^2\left(k_e\right) + \frac{1}{\tau}\frac{dV}{dk} & -\tilde{\lambda} - \frac{1}{\tau} \end{array} \right) \tag{B.32}$$

where we have used

$$\tilde{\lambda} = \lambda - i \cdot \tilde{\omega} \quad \text{with} \quad \tilde{\omega} = \omega - sV\left(k_e\right) \tag{B.33}$$

Eqn. (B.31) will only have a non-zero solution for certain values of $\tilde{\lambda}$, namely the eigenvalues[3] of

$$\left(\begin{array}{cc} 0 & -i \cdot sk_e \\ -\frac{i \cdot s}{k_e} c^2\left(k_e\right) + \frac{1}{\tau}\frac{dV}{dk} & -\frac{1}{\tau} \end{array} \right) \tag{B.34}$$

To compute these eigenvalues, we determine the characteristic polynomial $\zeta\left(\tilde{\lambda}\right) = \det\left(M\right)$[4], which equals

$$\zeta\left(\tilde{\lambda}\right) = \tilde{\lambda}^2 + \frac{1}{\tau}\tilde{\lambda} + i \cdot sk_e\left[-\frac{i \cdot s}{k_e} c^2\left(k_e\right) + \frac{1}{\tau}\frac{dV}{dk}\right] = 0 \tag{B.35}$$

The two solutions to this equation are

$$\begin{aligned} \tilde{\lambda}_{1,2} &= \tilde{\lambda}_{1,2}\left(k_e, s\right) = -\frac{1}{2\tau} \pm \sqrt{\frac{1}{4\tau} - \left(C_r + i \cdot C_i\right)} && \text{(B.36)} \\ &= -\frac{1}{2\tau} \pm \left[\sqrt{\mathcal{R} + i \cdot \left|\mathcal{I}\right|}\right] && \text{(B.37)} \\ &= -\frac{1}{2\tau} \pm \left[\sqrt{\frac{1}{2}\left(\sqrt{\mathcal{R}^2 + \mathcal{I}^2} + \mathcal{R}\right)} \pm i \cdot \sqrt{\frac{1}{2}\left(\sqrt{\mathcal{R}^2 + \mathcal{I}^2} - \mathcal{R}\right)}\right] && \text{(B.38)} \end{aligned}$$

where

$$C_r = s^2 c^2\left(k_e\right) \geq 0 \quad \text{and} \quad C_i = \frac{sk_e}{\tau}\frac{dV}{dk} \leq 0 \tag{B.39}$$

and

$$\mathcal{R} = \frac{1}{4\tau^2} - C_r \quad \text{and} \quad \pm\left|\mathcal{I}\right| = -C_i = \frac{sk_e}{\tau}\left|\frac{dV}{dk}\right| \tag{B.40}$$

In sum: when the wave damping factors $\tilde{\lambda}\left(k_e, s\right)$ are equal to Eqn. (B.38), we can find amplitudes k_*, u_* that satisfy Eqn. (B.31), which means that the simple waves (B.29) and (B.30) are solutions to the partial differential equations (B.25) and (B.26). This means that the simple waves provide a *non-zero contribution* to the solution given by Eqns. (B.3) and (B.4).

[3] Recall that by definition, the eigenvalues λ of a matrix M are defined by those values for which $\det\left(M - \lambda I\right) = 0$, which in turn implies that the matrix $M - \lambda I$ is singular and has a non-empty null-space.

[4] Recall that for any 2x2 matrix $A = \left(\begin{array}{cc} a & b \\ c & d \end{array} \right)$, we have $\det\left(A\right) = ad - bc$.

Let us now consider the conditions under which the contributions of the simple waves (B.29) and (B.30) will cause the solution (B.3) and (B.4) to become unstable. This will be the case when either of the damping factors $\tilde{\lambda}_{1,2}$ has a real part which is larger than zero, which would mean that the perturbation would grow, rather than decrease. Since the real-part of $\tilde{\lambda}_1$ is per definition larger than the real-part of $\tilde{\lambda}_2$, we only have to consider $\tilde{\lambda}_1$. Let us now study under which circumstances $\text{Re}\left(\tilde{\lambda}_1\right) \geq 0$. We then have

$$\text{Re}\left(\tilde{\lambda}_1\right) \geq -\frac{1}{2\tau} + \sqrt{\frac{1}{2}\left(\sqrt{\mathcal{R}^2+\mathcal{I}^2}+\mathcal{R}\right)} = 0 \tag{B.41}$$

which implies that

$$\left[\frac{1}{4\tau^2} - \frac{\mathcal{R}}{2}\right]^2 \leq \frac{1}{4}\left(\mathcal{R}^2+\mathcal{I}^2\right) \tag{B.42}$$

By simple substitution, we can conclude that the Payne model is unstable, under the following conditions

$$k_e \left|\frac{dV}{dk}\right| \geq c\left(k\right) \tag{B.43}$$

B.4 Numerical solutions to the Payne model

Given that the progress regarding the mathematical analysis of the Payne model lacks far behind the LWR model, the schemes that are applicable to the Payne model are less far developed. In this section, we will extent the flux-splitting scheme developed for the LWR model. To this end, we first need to rewrite the Payne model - which is given in its so-called primitive form, i.e. describing the dynamics of density and speed - in its conservative form, that is describing the dynamics of the density $k(x,t)$ and the flow $q(x,t)$. When this is achieved, we apply a flux-splitting approach to the model in an equivalent way that was used to determine the flux-splitting scheme for the LWR model

B.4.1 Payne model in conservative form

The form is called conservative, since the density and the flow are conserved (i.e. change only by inflow minus outflow) in case only convective processes are present. Let us consider the speed dynamics equation (B.2) and multiply it by the density $k\left(x,t\right)$. Subsequently we multiply the conservation of vehicle equation (B.1) by the speed $u(x,t)$ and add it to the left hand side of the former result. We get the following result

$$k\left(\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x}\right) + u\left(\frac{\partial k}{\partial t} + \frac{\partial\left(ku\right)}{\partial x}\right) = k\frac{V\left(k\right)-u}{\tau} - k\frac{c^2\left(k\right)}{k}\frac{\partial k}{\partial x} \tag{B.44}$$

which can be written as follows

$$\begin{aligned} k\frac{\partial u}{\partial t} + u\frac{\partial k}{\partial t} + ku\frac{\partial u}{\partial x} + u\frac{\partial\left(ku\right)}{\partial x} &= \frac{\partial q}{\partial t} + \frac{\partial\left(uq\right)}{\partial x} && \text{(B.45)} \\ &= \frac{Q\left(k\right)-q}{\tau} - c^2\left(k\right)\frac{\partial k}{\partial x} && \text{(B.46)} \end{aligned}$$

Let us define the function $P\left(k\right)$ as follows

$$\frac{dP}{dk} = c^2\left(k\right) \tag{B.47}$$

thus

$$c^2\left(k\right)\frac{\partial k}{\partial x} = \frac{\partial P\left(k\right)}{\partial x} \tag{B.48}$$

Then, we speed dynamics equation can be transformed into the following equation describing the dynamics of the flow

$$\frac{\partial q}{\partial t} + \frac{\partial \left(uq + P\left(k \right) \right)}{\partial x} = \frac{Q\left(k \right) - q}{\tau} \tag{B.49}$$

In defining the vector $r = (k, q)$, we can rewrite the conservation of vehicle equation (B.1) and the flow dynamic equation (B.49) in the following form

$$\frac{\partial r}{\partial t} + \frac{\partial F\left(r \right)}{\partial x} = D\left(r \right) \tag{B.50}$$

where

$$F\left(r \right) = F\left(k, q \right) = \begin{pmatrix} q \\ \frac{q^2}{k} + P\left(k \right) \end{pmatrix} \tag{B.51}$$

and

$$D\left(r \right) = D\left(k, q \right) = \begin{pmatrix} 0 \\ \frac{Q(k) - q}{\tau} \end{pmatrix} \tag{B.52}$$

B.4.2 Flux-splitting numerical solution approach

This sectio discusses one of the more refined approaches to numerically approximating solutions to the Payne model. While having its drawbacks, the approach is generally applicable and robust, and provides accurate solutions.

Similar to the flux-splitting approach applied to the LWR model, the considered roadway is split-up into cells i of length l_i. In cell i, we consider space-averaged *densities* and *flows*[5]

$$k_{i,j} = \int_{x_{i-1}}^{x_i} k\left(x, t_j \right) dx \quad \text{and} \quad q_{i,j} = \int_{x_{i-1}}^{x_i} q\left(x, t_j \right) dx \tag{B.53}$$

The densities and flows change due to convection (inflow and outflow), as well as relaxation processes. The latter reflect traffic accelerating or decelerating to adapt to the equilibrium speed $V\left(k \right)$. To model the convective processes, the density-fluxes and the flow-fluxes at the cell interface (described by the vector F in the quasi-linear formulation (B.50)) need to be numerically approximated. Similar to the LWR model, the flux-splitting scheme will split the flux (density and flow flux) F into contributions of the sending cell i and the receiving cell $i+1$, i.e.

$$F\left(r_{ij}, r_{i+1j} \right) = F^{+}\left(r_{i,j} \right) + F^{-}\left(r_{i+1,j} \right) \tag{B.54}$$

subject to $F = F^{+} + F^{-}$.

To determine appropriate definitions for F^{+} and F^{-}, we, we need to see how the 'information' is transported in the flow. In the LWR model, we saw that the densities are transported along the characteristics. For the Payne model, we have seen that the flow properties (in terms of the Riemann variables or characteristic variables) are transported along the characteristics C_{+} and C_{-} defined by (9.109) and (9.110). The characteristic C_{+}, defined by $\frac{dx}{dt} = u + c\left(k \right) = u + \sqrt{\frac{dP}{dk}}$ always has a positive slope. As a result, the Riemann variable propagates along C_{+} in the downstream direction. Thus, the part of the equation that approximates the dynamics of the Riemann variable z_1 depends on the cell i only. However, the characteristic C_{-} may have a negative slope, i.e. when $u - c(k) = u - \sqrt{\frac{dP}{dk}} < 0$. In this case, the waves carrying the Riemann variable z_2 move upstream (i.e. from cell $i+1$ to cell i), which should be reflected in the definition of F^{-}.

[5] We will see that we will explicitly distinguish between spaced-averaged flows and flows at the cross-section x_i, defining the interface between two cells.

B.4.3 Quasi-linear formulation of Payne model

The approach to determine the correct description for F^+ and F^- is relatively straightforward. First, we rewrite the Payne model in its quasi-linear form

$$\frac{\partial r}{\partial t} + C(r)\frac{\partial r}{\partial x} = D(r) \tag{B.55}$$

where

$$C(r) = \begin{pmatrix} 0 & 1 \\ \frac{dP}{dk} - u^2 & 2u \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ c^2(k) - u^2 & 2u \end{pmatrix} \tag{B.56}$$

where $u = q/k$. Note that this form is equivalent to the quasi-linear form of Eqns. (B.3) and (B.4). We can use the same approach to diagonalise the matrix $C(r)$, namely by determining the eigenvalues $\lambda_{1,2}$ and the eigenvectors $e_{1,2}$ of the matrix $C(r)$. It turns out that

$$\lambda_{1,2} = u \pm c(k) \tag{B.57}$$

and

$$e_{1,2} = \begin{pmatrix} 1 \\ u \pm c \end{pmatrix} \tag{B.58}$$

Let R denote the matrix of eigenvectors (similar to the matrix S of eigenvectors of $A(w)$). As we have seen before, the Riemann variables can be defined by considering small variations δr and defining

$$\delta z = R^{-1}\delta r \quad \left(= S^{-1}\delta v\right) \tag{B.59}$$

The Riemann variables z satisfy the so-called *compability equations*

$$\frac{\partial z}{\partial t} + \Lambda(w)\frac{\partial z}{\partial x} = \tilde{D}(z) \tag{B.60}$$

where $\Lambda = R^{-1}C(r)R = S^{-1}A(w)S$, and where $\tilde{D}(z) = R^{-1}B(z)$. It is left to the reader to determine that

$$R = \begin{pmatrix} 1 & 1 \\ u+c & u-c \end{pmatrix} \text{ and } R^{-1} = \begin{pmatrix} -\frac{u-c}{2c} & \frac{1}{2c} \\ \frac{u+c}{2c} & -\frac{1}{2c} \end{pmatrix} \tag{B.61}$$

To determine the appropriate flux-splitting, the approach presented in section 9.3.4 is applied to each of the Riemann variables z_1 and z_2 separately. Then, with the defintions

$$\lambda^+_{1,2} = \frac{\lambda_{1,2} + |\lambda_{1,2}|}{2} \tag{B.62}$$

and

$$\lambda^-_{1,2} = \frac{\lambda_{1,2} - |\lambda_{1,2}|}{2} \tag{B.63}$$

we can define the matrices

$$\Lambda^\pm = \begin{pmatrix} \lambda_1^\pm & 0 \\ 0 & \lambda_2^\pm \end{pmatrix} = \begin{pmatrix} \frac{(u+c)\pm|u+c|}{2} & 0 \\ 0 & \frac{(u-c)\pm|u-c|}{2} \end{pmatrix} \tag{B.64}$$

Clearly, the eigenvalues of Λ^+ are all positive; Λ^- only has negative eigenvalues. That is, Λ^+ describes the contributions of all waves that propagate downstream, while Λ^- describes the contributions of the waves that move upstream. Note that, when $u \geq c$, $\Lambda^+ = \Lambda$ and $\Lambda^- = 0$, implying that only downstream propagating waves are present. When $u < c$

$$\Lambda^+ = \begin{pmatrix} u+c & 0 \\ 0 & 0 \end{pmatrix} \text{ and } \Lambda^- = \begin{pmatrix} 0 & 0 \\ 0 & u-c \end{pmatrix} \tag{B.65}$$

Since $C(r) = R\Lambda R^{-1}$, it makes sense to define

$$C^{+}(r) = R\Lambda^{+}R^{-1} \quad \text{and} \quad C^{-}(r) = R\Lambda^{-}R^{-1} \tag{B.66}$$

Then, $C^{+}(r)$ describes the contributions of the downstream moving waves, while $C^{-}(r)$ describes the contributions of the upstream moving waves. Note again that $C^{-}(r) \neq 0$, only if $u < c(k)$. That is, when $u > c(k)$, we have

$$C^{+}(r) = C(r) \quad \text{and} \quad C^{-}(r) = 0 \tag{B.67}$$

For $u < c(k)$, we find

$$C^{+}(r) = \frac{1}{2c}(u+c)\begin{pmatrix} -(u-c) & 1 \\ -(u+c)(u-c) & (u+c) \end{pmatrix} \tag{B.68}$$

$$C^{-}(r) = \frac{1}{2c}(u-c)\begin{pmatrix} (u+c) & -1 \\ (u-c)(u+c) & -(u-c) \end{pmatrix} \tag{B.69}$$

Unlike for the LWR model, it is in general not possible to find flux functions $F^{+}(r)$ and $F^{-}(r)$ that satisfy $\frac{\partial}{\partial r}F^{\pm}(r) = C^{\pm}(r)$. Therefore, we adopt an approximation known as the Steger-Warming scheme, where

$$F^{+}(r) = C^{+}(r) \cdot r \quad \text{and} \quad F^{-}(r) = C^{-}(r) \cdot r \tag{B.70}$$

It is left to the reader to determine that for $u > c(k)$

$$F^{+}(r) = \begin{pmatrix} q \\ \frac{q^2}{k} + P(k) \end{pmatrix} \quad \text{and} \quad F^{-}(r) = 0 \tag{B.71}$$

while for $u < c(k)$, we have

$$F^{+}(r) = \frac{k}{2}\begin{pmatrix} u+c \\ (u+c)^2 \end{pmatrix} \quad \text{and} \quad F^{-}(r) = F^{+}(r) = \frac{k}{2}\begin{pmatrix} u-c \\ (u-c)^2 \end{pmatrix} \tag{B.72}$$

Appendix C

Gas-kinetic modelling

Remark 68 *Gas-kinetic modelling of traffic flow is an advanced topic that can be skipped on a first reading.*

This section has discussed models that bear a strong resemblence to the models used to describe the dynamics of a fluid or a gas, or other continuous media. Therefore, these models are often referred to as *continuum models*. There is however, one specific type of continuum model that has not yet been discussed in this chapter, namely the so-called *gas-kinetic models*. These models offer a more refined and complicated description of traffic, and are deduced from theories applied in theoretical physics.

C.1 Prigogine and Herman model

The starting point of the gas-kinetic models, is the so-called *phase-space density (PSD)*

$$\kappa\left(x,t,v\right) \tag{C.1}$$

The PSD is a function of the location x, the time t, and the speed v. Simply stating, it describes not only the mean number of vehicles per unit roadway length (such as the density), but also includes information on the speed distribution at that location. First, recall that by definition, the number

$$k(x,t)dx \tag{C.2}$$

denotes the probability that at time-instant t in the small roadway segment $[x, x+dx)$ a vehicle is present. Equivalently, the number

$$\kappa\left(x,t,v\right)dxdv \tag{C.3}$$

denotes the probability that at time-instant t in a small roadway segment $[x, x+dx)$ a vehicle is present *with a speed in the region* $[v, v+dv)$. Note that

$$\kappa\left(x,t,v\right) = f(x,t,v)\cdot k(x,t) \tag{C.4}$$

where $f\left(x,t,v\right)$ denotes the probability density function describing the speed distribution of the vehicles at x.

Prigogine and Herman were the first to use the notion of the PSD to derive a model describing the behaviour of traffic flow. They achieved this by assuming that the PSD changes according to the following processes:

1. *Convection* $\frac{\partial(\kappa v)}{\partial x}$. Vehicles with a speed v flow into and out of the roadway segment $[x, x+dx)$, causing a change in the PSD $\kappa(x,t,v)$. This term is equivalent to the term $\frac{\partial(ku)}{\partial x}$ in the conservation of vehicle equation.

2. *Acceleration towards the desired speed* $\frac{V_0(x,t,v)-v}{\tau}$. Vehicles not driving at their desired speed will accelerate; $V_0(x,t,v)$ denotes the expected desired speed of vehicles driving with speed v; τ denotes the acceleration time.

3. *Deceleration when catching up with a slower vehicle* $(1-\pi)\,\kappa(x,t,v)\int(w-v)\,\kappa(x,t,w)$. A vehicle that encounters a slower vehicle will need to reduce its speed if it cannot immediately overtake.

Their deliberations yielded the following partial differential equation (PH-model):

$$\frac{\partial \kappa}{\partial t}+v\frac{\partial \kappa}{\partial x}=\frac{\partial}{\partial v}\left(\kappa\frac{V_0(x,t,v)-v}{\tau}\right)+(1-\pi)\,\kappa(x,t,v)\int(w-v)\,\kappa(x,t,w)\,dw \tag{C.5}$$

The most complex process here is probably the deceleration process reflected by the term $(1-\pi)\,\kappa(x,t,v)\int(w-v)\,\kappa(x,t,w)\,dw$. Let us briefly discuss how this term is determined from the following, simple behavioural assumptions:

1. The "slowing down event" has a probability of $(1-\pi)$, where π denotes the so-called *immediate overtaking probabity*. This probability reflects the event that a fast car catching up with a slow car can immediatly overtake to another lane, without needing to reduce its speed.

2. The speed of the slow car is not affected by the encounter with the fast car, whether the latter is able to overtake or not.

3. The lengths of the vehicles can be neglected.

4. The "slowing down event" has an instantaneous duration.

5. Only two vehicle encounters are to be considered, multivehicle encounters are excluded.

To derive the deceleration term, we recall the result of Hoogendoorn and Bovy (2001). The argue that event-based (event = encounter) changes in the PSD can be divided into changes that decrease the PSD $\kappa(x,t,v)$ and those that increase the PSD $\kappa(x,t,v)$. With respect to the former, they show that the rate of decrease of the PSD due to encounters equals

$$\left(\frac{\partial \kappa(x,t,v)}{\partial t}\right)^{-}_{encounter}=\int_w\int_{v'}\sigma_w(v'|v)\kappa(x,t,v)\,\Pi_w(v)\,dv'dw \tag{C.6}$$

Here, $\Pi_w(v)$ denotes the *event-rate*, i.e. the number of encounters per unit time that a vehicle driving with speed v has with another vehicle driving with speed w; $\sigma_w(v'|v)$ denotes the *transition probability* describing the probability that in case a vehicle with speed v encounters another vehicle with speed w, the former vehicle changes its speed to v'.

The event-rate can be determined by recalling that the number of *active overtakings* that a vehicle driving with speed v makes with slower vehicle driving with speed w in a small period $[t,t+dt)$ equals (see section 2.10.1)

$$(v-w)\,\kappa(x,t,w)dt \tag{C.7}$$

Is it clear that when $v<w$ (a *passive encounter*), the vehicle driving with speed v will not change its speed. This can be described by the transition probabilities via the following expression

$$\sigma_w\left(v'|v\right)=\delta\left(v'-v\right)\quad\text{for}\quad v<w \tag{C.8}$$

where $\delta(v)$ denotes the *delta-dirac function*[1]. When $v>w$, the vehicle can either immediately change lanes, without reducing its speed (implying $v'=v$), or it needs to decelerate to the speed

[1] Recall that the δ-dirac function by definition satisfies $\int f(x)\,\delta(x-x_0)\,dx=f(x_0)$

w of the preceding vehicle (i.e. $v' = w$). Since π denotes the immedate overtaking probality, we get

$$\sigma_w\left(v'|v\right) = \pi\delta\left(v' - v\right) + \left(1 - \pi\right)\delta\left(v' - w\right) \tag{C.9}$$

At the same time, encounters cause the PSD to increase. For instance, $\kappa\left(x,t,v\right)$ is increased when vehicles are slowed down *to the speed* v when encountering another vehicle. These increases can be modelled via the following term

$$\left(\frac{\partial\kappa(x,t,v)}{\partial t}\right)^{+}_{encounter} = \int_w\int_{v'}\sigma_w\left(v|v'\right)\kappa\left(x,t,v'\right)\Pi_w\left(v'\right)dv'dw \tag{C.10}$$

The transition probabilities are again given by the Eqns. (**??**) and (**??**). It is left to the reader to verify that the gross-effect of encounters equals

$$\left(\frac{\partial\kappa(x,t,v)}{\partial t}\right)^{+}_{encounter} - \left(\frac{\partial\kappa(x,t,v)}{\partial t}\right)^{-}_{encounter} = \left(1-\pi\right)\kappa\left(x,t,v\right)\int\left(w-v\right)\kappa\left(x,t,w\right)dw \tag{C.11}$$

C.2 Paveri-Fontana model

The model of Prigogine and Herman has been critised and improved by Paveri-Fontana (1975), who extended the PSD $\kappa\left(x,t,v\right)$ by also including the distribution of the desired speeds. That is, he considered $\hat{\kappa}\left(x,t,v,v_0\right)$, defined according to

$$\hat{\kappa}\left(x,t,v,v_0\right) = f\left(v,v_0\right)\cdot k(x,t) \tag{C.12}$$

where $f(v,v_0)$ is the joint speed-desired speed probability density function. The equation derived by Paveri-Fontana reads (PF-model):

$$\frac{\partial\hat{\kappa}}{\partial t} + \frac{\partial}{\partial x}\left(\hat{\kappa}\frac{dx}{dt}\right) + \frac{\partial}{\partial v}\left(\hat{\kappa}\frac{dv}{dt}\right) + \frac{\partial}{\partial v_0}\left(\hat{\kappa}\frac{dv_0}{dt}\right) = \left(\frac{\partial\kappa}{\partial t}\right)_{encounter} \tag{C.13}$$

Clearly, we have

$$\frac{dx}{dt} = v \tag{C.14}$$

The acceleration-term dv/dt reflects the acceleration of vehicles towards their desired speed v_0. Paveri-Fontana assumed the following exponential acceleration law

$$\frac{dv}{dt} = \frac{v_0 - v}{\tau} \tag{C.15}$$

Furthermore, it was assumed that the desired speed did not change over time, i.e. $dv_0/dt = 0$. The term reflecting the impacts of fast vehicles catching up with a slow vehicle equals

$$\left(\frac{\partial\kappa}{\partial t}\right)_{encounter} = \left(1-\pi\right)\int_{w>v}\int_{w_0}\left|w-v\right|\hat{\kappa}\left(x,t,v,w_0\right)\hat{\kappa}\left(x,t,w,v_0\right)dw_0dw \tag{C.16}$$

$$-\left(1-\pi\right)\int_{w<v}\int_{w_0}\left|w-v\right|\hat{\kappa}\left(x,t,w,w_0\right)\hat{\kappa}\left(x,t,v,v_0\right)dw_0dw \tag{C.17}$$

Note that the PH model can be derived from the PF model by integrating both the left-hand-side and the right-hand-side of equation (C.12) with respect to the desired speed v_0. We find

$$\frac{\partial\kappa}{\partial t} + \frac{\partial}{\partial x}\left(v\kappa\right) + \frac{\partial}{\partial v}\left(\kappa\frac{V_0\left(x,t,v\right) - v}{\tau}\right) = \left(1-\pi\right)\kappa\int\left(w-v\right)\kappa\left(x,t,w\right)dw \tag{C.18}$$

where

$$V_0\left(x,t,v\right) = \int_{v_0}\frac{\hat{\kappa}(x,t,v,v_0)}{\kappa\left(x,t,v\right)}dv_0 \tag{C.19}$$

denotes the expected desired speed of vehicles driving with speed v at (x,t).

C.3 Applications of gas-kinetic models

Gas-kinetic models are much more complicated than macroscopis models. It is not surprising to see that the mathematical analysis of these models lacks far behind the analysis of the LWR and even the Payne mode. Although it is possible to derive appropriate numerical solution, these scheme tend to be rather complicated. It would thus seem that the applicability of the gas-kinetic models is rather limited.

Amongst the most important and interesting applications of gas-kinetic models is its use to derive macroscopic models. Let us briefly show how this is achieved using the Paveri-Fontana equation (C.13). The method in question is called the *method of moments*. In rough terms, the approach consists of multiplying the Paveri-Fontana equation with powers of v and v_0, and subsequently integrating the result over v and v_0. Let us illustrate this approach by multiplying with $v^l v_0^m$. Then, we find

$$\frac{\partial}{\partial t}\left(k\left\langle v^l\right\rangle\right)+\frac{\partial}{\partial x}\left(k\left\langle v^{l+1}\right\rangle\right)-\frac{l}{\tau}k\left(\left\langle v^{l-1}v_0\right\rangle-\left\langle v^l\right\rangle\right)=\left(1-\pi\right)k^2\left(\left\langle v\right\rangle\left\langle v^l\right\rangle-\left\langle v^{l+1}\right\rangle\right) \tag{C.20}$$

where

$$\left\langle v^l v_0^m\right\rangle=\int_v\int_{v_0}v^l v_0^m f\left(v,v_0\right)dvdv_0 \tag{C.21}$$

where $f\left(v,v_0\right)$ denotes the joint probability distribution of the speed v and the desired speed v_0. Note that the following relations hold; for the mean speed $u\left(x,t\right)$ we have

$$u\left(x,t\right)=\left\langle v\right\rangle \tag{C.22}$$

For the speed variance $\theta\left(x,t\right)$ (denote the variance of speed around the mean speed $u\left(x,t\right)$), we have

$$\theta\left(x,t\right)=\left\langle\left(v-u\left(x,t\right)\right)^2\right\rangle \tag{C.23}$$

If we consider $l=m=0$, we find

$$\frac{\partial k}{\partial t}+\frac{\partial\left(k\,u\right)}{\partial x}=0 \tag{C.24}$$

which is obviously the conservation of vehicle equation.

For $l=1$ and $m=0$, after dividing Eq. (C.20) by k, we find

$$\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}+\frac{1}{k}\frac{\partial}{\partial x}\left(k\theta\right)-\frac{1}{\tau}\left(V_0-u\right)=-\left(1-\pi\right)k^2\theta \tag{C.25}$$

where V_0 denotes the mean desired speed of the entire traffic flow. This expression can be written as follows

$$\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}=\frac{V-u}{\tau}-\frac{1}{k}\frac{\partial}{\partial x}\left(k\theta\right) \tag{C.26}$$

where $V=V\left(k,\theta\right)$ denotes the relaxation or equilibrium speed, defined by

$$V\left(k,\theta\right)=V_0-\tau\left(1-\pi\right)k\theta \tag{C.27}$$

Note that Eq. (C.27) is equivalent to Eq. (B.2), assuming that

$$c^2\left(k\right)=\frac{dP}{dk}=\frac{d}{dk}\left(k\theta\right) \tag{C.28}$$

The latter expression provides a different interpretation of the wave speed $c\left(k\right)$, namely that is it related to the *standard speed deviations*. For instance, assuming that $d\theta/dk=0$, we have $c=\sqrt{\theta}$. Besides this alternative interpretation of c, we have also derived an expression for the equilibrium speed Eq. (C.27). This expression shows among other things that the slope of the

speed-density curve depends on the relaxation time τ, the immediate lane-changing probability π, and the speed variance θ. Also note that under the assumption that all these are constant values, the speed-density relation equals Greenshield's function.

It is clear that for the model to be applicable, the speed variance θ must either be explicitly determined (either as a function of the density, the speed, etc., or simply as a constant), *or* we can consider Eq. (C.20) for $l = 1$ and $m = 0$. After some complicated mathematics, the latter option yields the following, dynamic equation for the speed variance

$$\frac{\partial \theta}{\partial t} + u\frac{\partial \theta}{\partial x} = -2\theta\frac{\partial u}{\partial x} - \frac{1}{k}\frac{\partial}{\partial x}(kJ) + 2\frac{\Theta - \theta}{\tau} \tag{C.29}$$

where J denotes the *skewness of the speed distribution*, defined by

$$J = \left\langle (v - u(x,t))^3 \right\rangle \tag{C.30}$$

and where the *equilibrium speed variance* equals

$$\Theta = C - \frac{\tau}{2}(1 - \pi)\, kJ \tag{C.31}$$

where C denotes the *covariance between the speed and the desired speed*, i.e.

$$C(x,t) = \langle (v - u(x,t))(v_0 - V_0) \rangle \tag{C.32}$$

The problem now remains to appropriately choose the skewness of the speed distribution, and the covariance between the speed and the desired speed. Note that, under the assumption that the speed distributions are Gaussian, we have $J = 0$

Appendix D

Higher-order model extensions and generalisations

This last section discusses some further model extensions and generalisations of the macroscopic flow models discussed in this section. The most important generalisation, is the multiclass extension of the model, due to [26] and [24].

D.1 Multiclass generalisation of PF-model

Here, let a denote the user-class. The classes can be distinguished based on a number of characteristics, such as vehicle-driver characteristics, socio-economic importance of the user-class, destination of the traffic, etc. Let $\hat{\kappa}_a(t, x, v, v_0)$ denote the PSD of class a, denoting for instance the mean number of trucks that is driving at x with speed v, while having desired speed v_0. The PF-equation (C.13) clearly also holds for each class a, i.e.

$$\frac{\partial \hat{\kappa}_a}{\partial t} + \frac{\partial}{\partial x}\left(\hat{\kappa}_a \frac{dx}{dt}\right) + \frac{\partial}{\partial v}\left(\hat{\kappa}_a \frac{dv}{dt}\right) + \frac{\partial}{\partial v_0}\left(\hat{\kappa}_a \frac{dv_0}{dt}\right) = \left(\frac{\partial \kappa_a}{\partial t}\right)_{encounter} \tag{D.1}$$

where $\frac{dx}{dt} = v$, and $\frac{dv}{dt} = \frac{v_0 - v}{\tau_a}$, where τ_a is the class-specific acceleration time. The latter reflects the differences in acceleration capabilities of the different vehicle classes. Similar to the PH-model and the PF-model, we can show that for the multiclass generalisation, we have

$$\left(\frac{\partial \hat{\kappa}_a}{\partial t}\right)^{-}_{encounter} = \sum_{a',b} \int \sigma_{(w,w_0,b)}\left(v', v_0', a' | v, v_0, a\right) \Pi_{(w,w_0,b)}(v, v_0, a)\, dv' dv_0' dw dw_0 \tag{D.2}$$

and

$$\left(\frac{\partial \hat{\kappa}_a}{\partial t}\right)^{+}_{encounter} = \sum_{a,b} \int \sigma_{(w,w_0,b)}\left(v, v_0, a | v', v_0', a'\right) \Pi_{(w,w_0,b)}\left(v', v_0', a'\right) dv' dv_0' dw dw_0 \tag{D.3}$$

Eq. (D.2) denotes the reductions in $\hat{\kappa}_a(t, x, v, v_0)$ due to interactions with vehicles of class b, driving with speed w having desired speed w_0; Eq. (D.3) denotes the increases in $\hat{\kappa}_a(t, x, v, v_0)$ caused by vehicles of class a' driving with speed v' having desired speed v_0' changing their class / speed / desired speed to a, v and v_0 respectively after an interaction.

Thus, $\Pi_{(w,w_0,b)}(v, v_0, a)$ denotes the *rate* at which drivers of class a driving with speed v while having desired speed v_0 (referred to as vehicle (v, v_0, a)) *interact* with drivers of class b driving with speed w while having desired speed w_0. Clearly, we have

$$\Pi_{(w,w_0,b)}(v, v_0, a) = |w - v|\, \hat{\kappa}_b(t, x, w, w_0) \tag{D.4}$$

The *transition probabilities*

$$\sigma_{(w,w_0,b)}\left(v', v_0', a' | v, v_0, a\right)$$

denote the probability that after an interaction / encounter of vehicle (v, v_0, a) with vehicle (w, w_0, b), the speed / desired speed / class of vehicle (v, v_0, a) changes to (v', v_0', a').

Let us assume that the vehicle classes remain preserved upon interaction. Furthermore, assume that drivers will not change their desired speed when interacting with another vehicle. Then, using the same behavioural assumptions used to derive the PH-model, we can show that

$$\left(\frac{\partial \hat{\kappa}_a}{\partial t}\right)_{encounter} = \sum_b (1-\pi_a) \int_{w>v} \int_{w_0} |w-v| \, \hat{\kappa}_b(x,t,v,w_0) \, \hat{\kappa}_a(x,t,w,v_0) \, dw_0 dw \quad \text{(D.5)}$$

$$- \sum_b (1-\pi_a) \int_{w<v} \int_{w_0} |w-v| \, \hat{\kappa}_b(x,t,w,w_0) \, \hat{\kappa}_a(x,t,v,v_0) \, dw_0 dw \quad \text{(D.6)}$$

This expression shows how driver-vehicle combination of class a are decelerated by driver-vehicle combinations of their own class $b = a$ and by driver-vehicle combinations of other classes $b \neq a$.

D.2 Multiclass higher-order model

Multiclass PF-model Eq. (D.1) can be used in a similar fashion as the mixed-class PF-model Eq. (C.13) to derive a multiclass macroscopic model. The main benefit of using this approach is that the non-symmetrical interactions between fast and slow user-classes is contained within the model. In applying the method-of-moments to the multiclass PF-model, we find for $l = m = 0$

$$\frac{\partial k_a}{\partial t} + \frac{\partial (k_a \, u_a)}{\partial x} = 0 \quad \text{(D.7)}$$

Here, $k_a(x,t)$ and $u_a(x,t)$ respectively denote the class-specific density and mean speed. This equation describes the multiclass conservation of vehicle equation. It shows that the class-specific density $k_a(x,t)$ is conserved. Clearly, conservation will only hold when driver-vehicle combinations do not changes classes during the considered time period.

Considering $l = 1$ and $m = 0$, we can determine the speed-dynamic equations for class a

$$\frac{\partial u_a}{\partial t} + u_a \frac{\partial u_a}{\partial x} = \frac{V_a - u_a}{\tau_a} - \frac{1}{k_a} \frac{\partial}{\partial x}(k_a \theta_a) \quad \text{(D.8)}$$

where V_a denotes the equilibrium speed of class a stemming from interactions with slower vehicles from classes $b = a$ (within-class interactions) and $b \neq a$ (between class interactions). Although the general form of the speed dynamics equation is conserved, for the multiclass model it is not possible to derive a close-form expression similar to Eq. (**??**). The reason for this is the assymmetry of the speed distributions of the different classes a and b. Hoogendoorn and Bovy [29] propose using the following approximation. First, let

$$S_{ab} = \theta_a + \theta_b - 2\xi\sqrt{\theta_a \theta_b} \quad \text{(D.9)}$$

where θ_a denotes the speed variance of class a and $0 \leq \xi \leq 1$ denotes the correlation between the speed distributions of class a and b. Furthermore, let

$$\delta u_{ab} = \frac{u_a - u_b}{\sqrt{S_{ab}}} \quad \text{(D.10)}$$

denote the normalised mean speed difference between class a and b. Then, Hoogendoorn and Bovy [29] propose

$$V_a = V_a^0 - (1-\pi_a)\tau_a P_a \quad \text{(D.11)}$$

where V_a^0 denotes the mean desired speed of class a and where

$$P_a(x,t) = \sum_b P_{ab}(x,t) \quad \text{(D.12)}$$

with

$$P_{ab} = k_b S_{ab} \left[\delta u_{ab} f\left(\delta u_{ab}\right) + \left(1 + \delta u_{ab}^2\right) F_{ab}\left(\delta u_{ab}\right) \right] \tag{D.13}$$

describes the effects of interactions between vehicles on the average vehicle speed. Note that for $b = a$ we have

$$P_{aa} = k_a \left(1 - \xi\right) \theta_a \tag{D.14}$$

For $\xi = 0$ (no correlation between speeds), we have $P_{aa} = k_a \theta_a$ (compare to Eq. (**??**)). This implies that the model discussed here is a true generalisation of the mixed-class model.

Hoogendoorn and Bovy [29] provide a further generalisation of their model by distinguishing roadway lanes. The resulting model is can be used to explain and predict observed behavior on multilane facilities. It is beyond the scope of this course to explain the details of the resulting model.

Appendix E

Pedestrian flow model NOMAD

At this point, it would be interesting to note that the optimal control model paradigm has also been applied to the microscopic modeling of pedestrian flows (see [28]). In this model overview, it was show that the walker model describes the individual walking behavior of pedestrians, given the decisions made at the destination choice and route choice level. In the following sections, we will discuss the workings of the latter model. For now, it is sufficient to note that for each pedestrian p the optimal velocity $\mathbf{v}^*(t, x)$ (i.e. speed and direction) has been determined.

Hoogendoorn and Bovy [28] present the conceptual model, where pedestrians are described in terms of a closed feedback control system. This model leads to an optimal control formulation, where pedestrians are assumed to minimize the predicted costs of walking. These costs are determined by among other things the extent to which pedestrians can walk at their desired velocity, the proximity of other pedes-trians, walking near to obstacles, and the need to accelerate or decelerate, Furthermore, it is hypothesized that the costs are discounted over time, implying that events in the near future are deemed more important than events that are expected to happen much later. Application of Pontryagin's Maximum Principle establishes the mathematical model describing walking behavior. Besides a control model, predicted the optimal walking behavior of a pedestrian, a physical model describes the effects of physical contact between pedestrians and obstacles.

The walker model) is governed by the following ordinary differential equation, describing the behavior of pedestrian p

$$\dot{\mathbf{r}}_p = \mathbf{v}_p \quad \text{and} \quad \dot{\mathbf{v}}_p = \mathbf{a}_p \tag{E.1}$$

with

$$\mathbf{a}_p = \left(\frac{\mathbf{v}_p^* - \mathbf{v}_p}{\tau_p} - \nabla L \right) \tag{E.2}$$

where L denotes the proximity cost incurred by pedestrian p walking close to other pedestrians q, e.g.

$$L = \sum_p A_0 \exp\left[-\frac{1}{R_0} \frac{\mathbf{r}_p - \mathbf{r}_q}{\|\mathbf{r}_p - \mathbf{r}_q\|} \right] \tag{E.3}$$

where $A_0 > 0$ and $R_0 > 0$ pedestrian-type specific parameters. This specification shows that the cost L increase when pedestrians walk closer to each other. The walker model used in the actual simulation model NOMAD is more complex, and features pedestrian anisotropy, grouping behavior, and impatience. The user can set the parameters of these processes. A similar expression is used to describe the way pedestrians walk around obstacles.

NOMAD predicts the effects of special infrastructure on the walking speeds of pedestrians. In case of stairs or grades, the desired speed of the pedestrians is reduced in accordance with the grade of the walking surface or the stairs.. The behavior on an escalator is described easily by assuming that the pedestrian will move at the same speed as the escalator.

Bibliography

[1] M. Abou-Heniady, S. Teply, and J. Hunt. Gap acceptance investigations in canada. In "Proc. II Int. Symp. On Highway Capacity. Ed. Akçelik". ARRB and TRB (1994).

[2] M. Bando, K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama. Dynamical model of traffic congestion and numerical simulation. *Physical Review* **E(51)**, 1035–1042 (1995).

[3] Bexelius. Variatie in verkeersintensiteiten. hoe normaal is normaal. In "Proceedings Verkeerskundige Werkdagen", Ede (1995). CROW.

[4] H. R. Board. Highway capacity manual 1985 (1985). Special Report 209.

[5] H. R. Board". Highway capacity manual 2000 (2000). Special Report.

[6] H. Botma, C. V. Goeverden, and P. Bovy. Effect of road lighting on capacity of freeways. In "3rd Int. Symp. On Highway Capacity" (1998).

[7] D. Branston. Models of single lane time headway distributions. *Transportation Science* **10(2)**, 125–148 (1976).

[8] W. Brilon. Warteschlangenmodell des verkehrsablaufs auf zweispurigen landstra en (in german). Technical Report, Karlsruhe (1974).

[9] W. Brilon, R. Koenig, and R. Troutbeck. Useful estimation procedures for critical gaps. In M. Kyte, editor, "Proc. III Int. Symp. On Intersections Without Traffic Signals" (1997).

[10] D. Buckley. A semi-poisson model of traffic flow. *Transportation Science* **2(2)**, 107–132 (1968).

[11] D. Buckley. A semi-poisson model of traffic flow. *Transportation Science* **2(2)**, 107–132 (1968).

[12] R. E. Chandler, R. Herman, and E. W. Montroll. Traffic dynamics: Studies in car-following. *Operations Research* **6(2)**, 165–184 (1958).

[13] R. J. Cowan. Useful headway models. *Transportation Research* **9**, 371–375 (1975).

[14] C. Daganzo. "Fundamentals of Transportation and Traffic Operations". Pergamon (1997).

[15] E. de Romph. "A Dynamic Traffic Assignment Model". PhD thesis, Delft University of Technology (1994).

[16] T. Dijker, P. Bovy, and R. Vermijs. Car-following under congested conditions: Empirical findings. *Transportation Research Record* **1644**, 20–28 (1998).

[17] J. Drake. A statistical analysis of speed-density hypothesis. In "Proc III Int. Symp. On the Theory of Traffic Flow". Elsevier (1965).

[18] D. Drew. Deterministic aspects of freeway operation and control. In "Highway Research Record", vol. 99, Washington D.C. (1965). Transportation Research Board.

[19] L. Edie. Discussion of traffic stream measurements and definitions. In "2nd Int. Symp. On the Theory of Traffic Flow", Paris (1965). OECD.

[20] T. Forbes, H. Zagorski, E. Holshouser, and W. Deterline. Measurement of driver reactions to tunnel conditions. *Highway Research Board Proceedings* **37**, 345–357 (1958).

[21] N. Gartner, H. Mahmassani, C. Messer, H. Lieu, R. Cunard, and A. Rathi. "Traffic Flow Theory". Transportation Research Board (1998).

[22] B. Greenshields. A study of traffic capacity. In "Highway Research Board Proc.", vol. 14, pp. 468–477 (1934).

[23] P. Groeneboom. Onderzoek volgtijden op 2x2-strooks autosnelweg. Technical Report, Centrum voor Wiskunde en Informatica, Amsterdam (1984).

[24] D. Helbing. "Verkehrsdynamik". Springer-Verlag (1997).

[25] R. Herman. Traffic dynamics: Analysis of stability in car-following. *Operation Research* **1(7)**, 86–106 (1959).

[26] S. Hoogendoorn. Non-local multiclass gas-kinetic modeling of multilane traffic flow. *Networks and Spatial Economics* **1**, 137–166 (2001).

[27] S. Hoogendoorn and P. Bovy. New estimation technique for vehicle type-specific headway distributions. *Transportation Research Record* **1646**, 18–28. (1998).

[28] S. Hoogendoorn and P. Bovy. Normative pedestrian behaviour theory and modelling. *15th International Symposium on Transportation and Traffic Theory, Adelaide, Australia* (2002).

[29] S. Hoogendoorn, P. Bovy, and H. V. Lint. Short-term prediction of traffic flow conditions in a multilane multi-class network. In M. Taylor, editor, "Transportation and Traffic Theory in the 21st Century, Pergamon 625-652.", pp. 625–652. Pergamon (2002).

[30] M. Jepsen. On the speed-flow relationships in road traffic: A model of driver behaviour. *Proceedings of the Third International Symposium on Highway Capacity* pp. 297–319 (1998).

[31] N. Kaptein, J. Theeuwes, and R. van der Horst. Driving simulator validity: Some considerations. In "Transportation Research Board Annual Meeting", pp. paper nr. 96–1338. (1996).

[32] B. Kerner. Complexity of synchronized flow. *Networks and Spatial Economics* **1(1/2)**, 35–76 (2001).

[33] W. Leutzbach. "An Introduction Into the Theory of Traffic Flow". Springer-Verlag, Berlin (1988).

[34] M. Lighthill and G. Whitham. On kinematic waves II: A theory of traffic flow on long, crowded roads. *Proceedings of the Royal Society of London series A* **229**, 317–345 (1955).

[35] W. Maes. Traffic data collection system for the belgian motorway network - measures of effectiveness aspects. *Proceedings of the International Symposium on Traffic Control Systems* **2D - Analysis and Evaluation**, 45–73 (1979).

[36] A. May. "Traffic Flow Fundamentals". Prentice Hall, Englewood Cliffs (1990).

[37] M. M. Minderhoud. "Supported Driving: Impacts on Motorway Traffic Flow". PhD thesis, Delft University of Technology (1999).

[38] K. Nagel and M. Schreckenberg. A cellular automaton model for freeway traffic. *Journal de Physique I* **2**, 2221–2229 (1002).

[39] J. Nelis, H. Schuurman, R. Vermijs, and D. Westland. Werk in uitvoering op autosnelwegen; microsimulatie van de verkeersafwikkeling. Technical Report VK 2207.305, Laboratorium voor Verkeerskunde, Fac. der Civiele Techniek, TU Delft (1991).

[40] G. F. Newell. A theory of traffic flow in tunnels. In R. Herman, editor, "Theory of Traffic Flow", pp. 193–206 (1961).

[41] M. Papageorgiou. Some remarks on macroscopic traffic flow modelling. *Transportation Research A* **32(5)**, 323–329 (1998).

[42] H. Payne. Models for freeway traffic and control. *Mathematical Models of Public Systems* **1**, 51–61. (1971).

[43] L. Pignataro. "Traffic Engineering - Theory and Practice". Prentice-Hall Inc., Englewood Cliffs (1973).

[44] L. A. Pipes. An operational analysis of traffic dynamics. *Journal of Applied Physics* **24**, 274–287 (1953).

[45] J. Reason. "Human Error". Cambridge University Press., New-York (1990).

[46] H. Schuurman and R. Vermijs. Ontwikkeling van het mikrosimulatiemodel FOSIM voor weefvakken en invoegingen. Technical Report VK 2205.307, Laboratorium voor Verkeerskunde, Fac.der Civiele Techniek, TU Delft. (1993).

[47] H. Schuurman and R. Vermijs. Verkeersafwikkeling bij discontinuïteiten; analyse van de verkeersafwikkeling m.b.v. microsimulatie bij weefvakken en invoegingen op autosnelwegen. Technical Report VK 2205.305, Laboratorium voor Verkeerskunde, Fac. der Civiele Techniek, TU Delft (1993).

[48] H. Schuurman and R. Vermijs. Simulatie benuttingsmaatregelen; effektenstudie m.b.v. het mikrosimulatiemodel FOSIM. Technical Report VK 2205.309, Laboratorium voor Verkeerskunde, Fac. der Civiele Techniek, TU Delft (1994).

[49] S. Smulders. "Control of Freeway Traffic Control". PhD thesis, TU Twente (1989).

[50] H. Stembord. Quality of service on the main road network in the netherlands. In B. Brannolte, editor, "Highway Capacity and Level of Service", Rotterdam (1991).

[51] H. Stipdonk. Time headway decomposition, a powerful tool in the study of following behaviour. Technical Report, RWS, Transportation and Traffic Engineering Division, Den Haag (1986).

[52] Transpute. Wachttijden voor hoofdwegen. Technical Report, (1988).

[53] J. Treiterer. Investigations of traffic dynamics by aerial photogrammetry techniques. Technical Report Interim report EES 278-3, Ohio State University (1970).

[54] J. Treiterer and J. Myers. The hysteresis phenomena in traffic flow. In D. Buckley, editor, "Proceedings of the Sixt Symposium on Transportation and Traffic Flow Theory", pp. 13–38. Elsevier (1974).

[55] R. Troutbeck. Overtaking behaviour on australian two-lane rural highways. Technical Report Special Report no. 20, ARRB (1981).

[56] J. van Toorenburg. Praktijkwaarden voor de capaciteit. *Verkeerskunde* **37 (5,6).** (1986).

[57] M. Westerman. "Real-Time Traffic Data Collection for Transportation Telematic". PhD thesis, TRAIL (1995).

[58] N. Wu. A new approach for modelling of fundamental diagrams and its applications institute for transportation and traffic engineering. Technical Report, Ruhr University Bochum (2000).

[59] H. Zhang. A mathematical theory of traffic hysteresis. *Transportation Research B* **33**, 1–23 (1999).

Index

acceleration fan, 153
arrival process, 55
  application, 58
  Binomial, 56
  Negative binomial, 56
  parameter estimation, 58
  Poisson, 55

capacity, 83, 107
  definition, 107
  estimation, 120
    Product-Limit method, 121
  ramps, 111
  stochastic nature of, 119
  weaving section, 114
capacity density, *see* critical density
capacity drop, 120
capacity speed, *see* critical speed
car-following model, *see* stimulus-response model
cellular automata, 203
congestion probability, 132
conservation of vehicles, 139
critical density, 83
critical speed, 83
cumulative vehicle count, *see* cumulative vehicle plot
cumulative vehicle plot, 32
  conservation of vehicle equation, 39
  delay, 34
  travel time, 34

density, 23
  definition, 24
  generalized definition, 39
distance headway distribution, 74
driving task, 175

equilibrium, 99

free speed, 83
fundamental diagram, 81
  causality, 82
  factors influencing, 84
  special points, 83
fundamental relation, 27
fundemantal diagram
  capacity estimation, 96
  discontinuous, 88
  field location, 94
  model, 85
  roadway lane, 90
  weather conditions, 97
fuzzy logic model, 202

gap, 206
  critical, 206
  distribution of critical, 211
gap acceptance, 205
gap acceptance function, 206
Godunov, 170

headway, 21
  definition, 22
  distance, 23
    definition, 23
  gross
    definition, 22
  nett
    definition, 22
headway distribution, 59
  capacity estimation, 72
  composite model, 69
  Erlang, 66
  exponential, 60
  exponential tail model, 68
  log normal, 67
  shifted exponential, 66
higher-order model, 170
  numerical solution, 174
  Payne model, 171
  properties of, 173
  waves, 173
homogeneous flow, 26
hysteresis, 99

instability, 102
intensity, 23
  definition, 23
  generalized definition, 39

jam density, 83

kinematic equations, 20
kinematic wave, 139
kinematic wave model, 151
  method of characteristics, 152
  numerical solution, 164

level of service, 83, 107
  definition, 108

measuring method, 43
  density, 45
  distance headway, 45
  double induction loop, 45
  headway, 43
  individual speed, 45
  induction loop, 44
  infrared detector, 44
  instantaneous speed, 45
  instrumented vehicle, 44
  intensity, 43
  licence plate identification, 45
  local mean speed, 45
  passage time, 43
  pneumatic tube, 44
  probe vehicle, 46
  radar, 45
  remote sensing, 46
metastability, 102
moving bottleneck, 145
moving observer, 47
  speed distribution observed by, 51

observation method, *see* measurement method
occupancy, 47
occupancy rate, *see* occupancy
optimal control model, 199
overtaking frequency, 75

psycho-spacing model, 197

queuing model, 127
  deterministic, 127
  QUAST, 132
  stochastic, 132
queuing theory, *see* queuing model

rest gap, 62

safe-distance model, 188
service volume, 108
shock wave, 141
  classification, 147
  formation of, 153
speed
  distribution, 75
  generalized mean definition, 39
  individual, 75
  instantaneous mean
    definition, 25
  local mean
    definition, 25
  mean, 23
    relation between local and instantaneous, 28, 31
speed distribution
  relation between instantaneous and local, 27
stability, 102
stationary flow, 26
stimulus-response model, 189
  stability, 190
survival function, 59

time headway, *see* headway
traffic simulation
  calibration, 235
  definition, 217
  FOSIM, 231
  macroscopic, 241
  METANET, 241
  microscopic, 227
  validation, 235
traffic simulation model
  definition, 217
traffic state
  free flow, 102
  moving jam, 103
  synchronized flow, 103
trajectory, 16
  definition, 16
transient state, 99
travel time function, 83