PREFACE

One of the biggest societal and economical important transport related problems in industrialised counties nowadays is congestion. In this respect, managing traffic in congested networks, requires a clear understanding of congested flow operations. That is, insights into what causes congestion, what determines the time and location of traffic breakdown, how does the congestion propagate through the network, etc., are essential. During the past fifty years, a wide range of traffic flow theories and models have been developed to answer these research questions. Some of these are currently applied in Dutch application studies to assess among other things roadway geometry design, and the impact of traffic management measures. However, it is unclear if these models are able to correctly describe congested traffic flow operations, and how realistic the outcomes of these assessment studies are.

This is why the Transport Research Centre of the Dutch Ministry of Transportation, Roadworks, and Water Management initiated the research project "Traffic Flow Operations during Congestion", which aims to delineate possible practical problems and theoretical pitfalls encountered when describing, and modelling congested traffic flow operations on both a microscopic level, as well as on a macroscopic level. This report presents the results of this research project, which has been conducted by the AGV Traffic and Transport Consultancy.

SUMMARY

In this report, results of the research project "*Traffic Flow Operations during Congestion*" have been reported, the aim of which was to identify issues relating to current problems encountered in Dutch application studies, current state-of-the-art concerning empirical and theoretical knowledge on traffic flow operations during congestion, and how this knowledge is currently implemented in traffic flow models. Also, the views of experts concerning issues in congested traffic flow operations and modelling were included in the research. Finally, data requirements for additional research is congested traffic flow operations were investigated.

As it turns out, the main problem with respect to congested operations modelling appears to be twofold. For one, little empirical and theoretical knowledge is available about driver behaviour at the onset and dissolution of traffic congestion (traffic breakdown). Secondly, due to the absence of different benchmark situations for which sufficient and sufficiently detailed data has been collected, the inaccuracies in the modelling cannot be contributed to a specific cause or submodel level (e.g. route-choice, traffic operations modelling).

Another issue concerns the current knowledge level regarding both microscopic and macroscopic behaviour of traffic flow. The main question related to this research is whether is possible to build a theory of driver behaviour that can withstand validation at different locations and days and that is strong enough to explain the phoneme mentioned above.

The literature shows observations of traffic hysteresis related flow behaviour, that mostly are not described by traditional flow theories. Another important result from literature is that *different types of congestion* can occur, determining traffic flow behaviour inside the congested area, and the queue tail location. Also on a *microscopic scale*, it appears that driving behaviour is affected considerably by the (type of) congested traffic conditions, as well as on specific parts of the networks, time-of-day, weather conditions, etc. With respect to lane-changing behaviour, important issues are the so-called give-way rule, and differences in lane-changing behaviour upstream of a discontinuity.

After the research, we can conclude that considerable gaps in the current knowledge of congested traffic flow exist. For instance, we do not known what the main factors are in traffic breakdown, or how random breakdown times and locations are. Moreover, we do not know how driving behaviour depends on the flow regime, or how it changes when discontinuities are nearby.

It is clear that to further improve the current state of knowledge and the traffic flow models currently used for Dutch application studies, additional research is required. This research must largely be *data-orientated*. For *microscopic flow analysis*, an extensive and costly data collection program would be needed, yielding detailed information on the behaviour of vehicles compared to vehicles in their direct environment. This type of fundamental research may not be possible without data that consists of complete

vehicle trajectories recorded at a high frequency, and vehicle information. On the contrary, research on the macroscopic level should aim at establishing quantitative relational dependencies of traffic flow variables on internal variables, and external conditions, implying that availability of data is currently not a real problem.

Once the issues above have been clarified, these new insights can be incorporated in the traffic flow models that are currently in use. To operationalise the adapted models, an unambiguous model calibration and validation methodology must be available, and if necessary, developed. Also, the development of advanced, and widely available tools for data analysis is required.

1 INTRODUCTION

These days, one of the main societal and economical issues in transportation is *congestion*. Traffic jams appear to attract much attention from among others commuters, politicians, and especially the media nowadays. This is not surprising, since every day millions of people are trapped in traffic jams as they try to get to work in the morning or going to the beach on a sunny day, while repeating this routine when going home in the evening, blaming other drivers, increasing traffic volumes, or roadworks for the delays they experience.

As a consequence of the negative societal, economical and environmental impact of traffic jams, governments of industrialised countries spend huge amounts of money on solving congestion in their roadway networks. This is done by either building more roads, or by implementing special traffic management measures. However, solving congestion entirely appears to be a lost cause, as congestion is an inherent side-effect of our increasing mobility needs. This insight has resulted in a policy shift from "solving congestion" towards "managing congestion", for instance by Dynamic Traffic Management. However, to be able to manage congestion, insights into the behaviour of traffic jams in essential, notwithstanding the importance of adequate theories and operational models to describe congested traffic flow, and the mechanisms that cause traffic jams realistically.

Over the past fifty years, researchers have developed a wide range of different mathematical traffic flow models to answer the aforementioned research questions. These models must clearly be based on real driver behaviour, while their solutions should reveal phenomena observed in real traffic flow. However, it is unclear if these models are able to correctly describe traffic flow operations prior, during, and after congestion. This research aims to answer these questions.

1.1 Subject of the report

This report presents the results of the research project "*Traffic Flow Operations during Congestion*", initiated by the Transport Research Centre of the Dutch Ministry of Transportation, Roadworks, and Water Management. Aim of this project is to delineate possible theoretical and practical problems that are encountered when describing, and modelling congested traffic flow operations. Particularly both the effects of Dynamic Traffic Management (DTM) measures on these traffic flow operations as well as the problems encountered with models used by the Dutch Ministry are studied. The project aims to determine which problems (if any) are present and what potential solution approaches need to be considered. In addition to among other things a literature survey, interviews with both national and international experts in the field of traffic flow modelling and have been conducted. Finally, data criteria have been established for empirical research, model development, calibration and validation.

1.2 Research questions and objectives

With respect to traffic flow operations during congestion, the following research questions can be posed:

1. Which empirical and theoretical knowledge is currently available concerning traffic flow operations during congestion?
2. How are congested traffic flow operations currently implemented in traffic flow theories and models?
3. How are the attitudes of experts, researchers, and model developers concerning issues in congested traffic flow modelling and the consequent identified problems?
4. To what extent is the currently available data adequate for research of congested traffic flow operations, or what are the criteria that need to be imposed on data to enable corrective analysis of congested traffic flow?

Objective of this research is to answer these questions as far as this is possible given the current state of knowledge in congested traffic flow theory and modelling. Moreover, the research aims to indicate where the available knowledge does not suffice, and future research is needed.

1.3 Reading guide and overview

Figure 3 State-of-the-Art review on Congested Traffic Flow-4 shows the relations between the report chapters. It depicts that in chapter 2, problems relating to the lack of empirical knowledge, traffic flow theories, and accurate models for congested traffic flow and driving behaviour are identified. This demarcation is based upon interviews with experts in the field of traffic flow and driver behaviour modelling, results from Dutch application studies, and finally, results from the literature survey (which can be found in the appendix of the report). From the perspective of the problem identification, chapter 3 provides the state-of-the-art in congested traffic flow empirical and theoretical knowledge. The subsequent chapter discusses how these theories can be incorporated into models.

Based on the lacks in the current knowledge, chapter 5 describes which data is required to gain more insights into traffic flow behaviour during congestion, based on the identified problems (chapter 2) and the unresolved issues following from the state-of-the-art (chapter 3). Finally, chapter 6 summarises the main research findings and provides recommendations regarding both modelling mechanisms to be included in the current models (as far as possible), as well as future research directions.

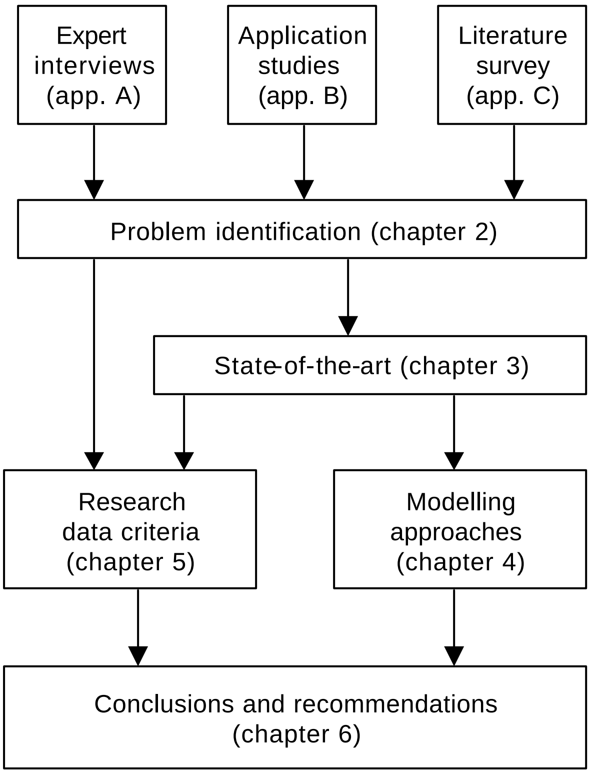

Figure 1 Introduction-1 Overview of contents of report and relations between chapters.

2 RESEARCH PROBLEM AND FOCUS

What are the causes for congestion? Can we predict the time traffic breakdown occurs? How can congested traffic flow operations be described? Are macroscopic theories adequate? Do we need microscopic models, or are macroscopic models sufficient, or even preferable?

These questions relate to the subject of this research, namely traffic flow operations during congestion. This chapter delineates the focus of the research described in this report. To this end, first a theoretical framework is given, in which we describe how and why congestion occurs on different levels in the traffic process. The theoretical framework is followed by a description of the topics in this process on which the research has focussed, based on the interviews with Dutch and international experts, the results of Dutch application studies, and the literature survey.

2.1 Traffic congestion and its causes

Congestion is the result of the (temporarily) imbalance of *traffic demand* and *infrastructure supply.* In its broadest sense, traffic demand relates both to the number of drivers that aim to traverse along specific parts of the transportation network, as well as the amount of the supply given by the roadway that is taken up by the driver. We will refer to the former aspect as the *strategic level*, while the latter aspect is called the *operational level*. More precisely, travel decisions on the strategic level relate to mode choice, departure time choice, and route choice. The operational level relates to behaviour of the driver on the road in relates to other vehicles in the traffic stream. Since traffic demand, and the consequent infrastructural supply shortage, is the net result of both type of demands, in this section, we will consider both aspects in more detail.

2.1.1 Strategic traffic demand

At the strategic level, the prime unit of analysis are (*dynamic*) *route flows*. These flows depict the number of drivers per time unit that traverse a specific route in a transportation network at a given time instant. They are determined by travel choices at the *strategic level* of the drivers decisions hierarchy, and are a result of among other things mode choice, route choice, and departure time choice behaviour. Experience, pre-trip information (such as the internet, CeeFax, etc.), and en-route (congestion) information (for instance Dynamic Route Information Panels, in-vehicle systems, RDS-TMC) affect these travel choices and consequently determine the strategic traffic demand (see Figure 2 Research problem and focus-2).

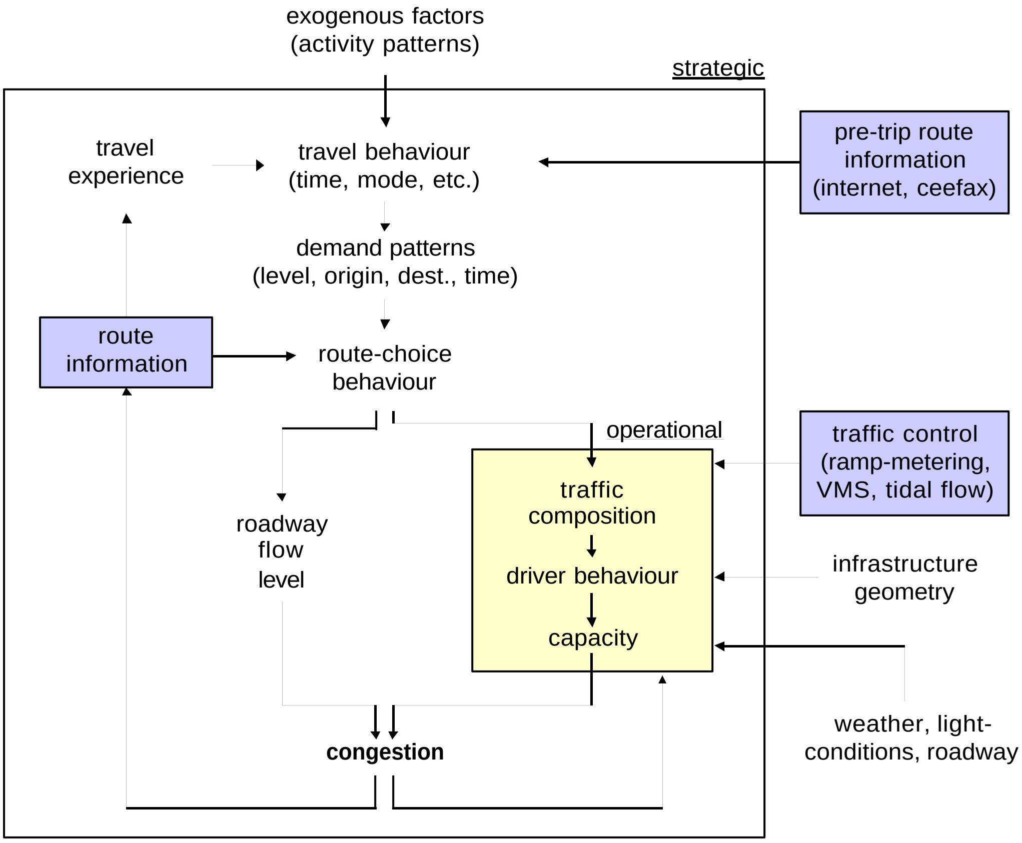

Figure 2 Research problem and focus-2 Strategic and operational levels affecting congestion (adapted from Hoogendoorn and Bovy (2000))

Traffic demand at the strategic level is highly stochastic. This randomness stems from variability in activity patterns and related travel choices (trip making, origin and destination, departure time, route, mode choices) of road users. The variability in daily travel choices also partly is a result of congestion experiences of road users on preceding days (see Figure 2 Research problem and focus-2). In this respect, it is emphasised that the traffic composition (in terms of mix and sequence of vehicle types, driver types, etc.) at any moment varies even much more from day to day. At a given hour of the day, even in the peak, only half of the travellers using a given motorway stretch used that stretch at that hour the day before.

2.1.2 Operational traffic demand and capacity

The capacity is determined by operational factors, and not just the considered infrastructure itself. In this section, we show how the interplay between the supply of infrastructure and the operational traffic demand determines the roadway capacity.

Let us consider a specific location in the transportation network. Recall that the headway of vehicle n is defined by the difference between rear bumper passage times of the leading vehicle n-1 and vehicle n, both traversing the same roadway lane. Roughly speaking, the headway of vehicle n consist of:

]

1. The *empty zone*, reflecting the minimal time headway needed by vehicle *n* for safe, efficient, and comfortable driving, and;
2. The *free headway*, which equals the additional headway when vehicle *n* is free flowing.

The *empty zone* is determined by:
a) the length of the following vehicle, and
b) additional time required by the following driver for safe, efficient, and comfortable driving.

While the former is fixed for each driver, the latter is dependent on among other things the followed vehicle and its type (person-car, truck), road, weather, and ambient conditions, infrastructure geometry, and traffic control measures (see Figure 2 Research problem and focus-3). Note that when the headway too small, the driver will generally decrease its velocity to increase the headway until it is equal to the empty zone, possibly considering some finite reaction time.

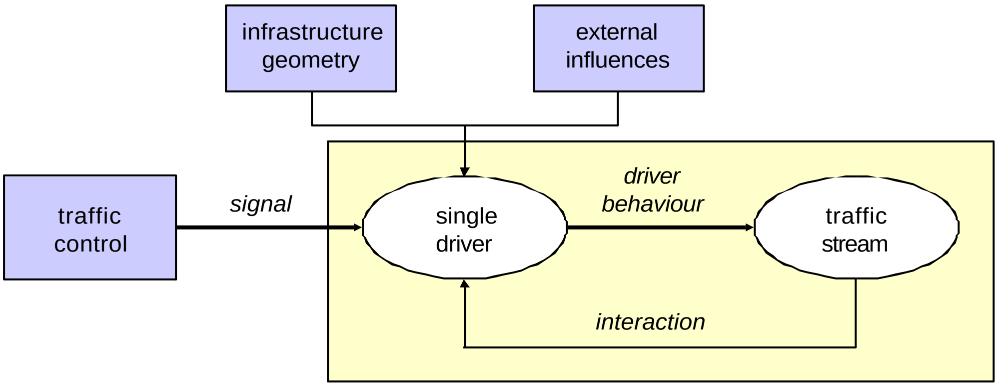

Figure 2 Research problem and focus-3 Factors on the operational level that determine traffic congestion

The *empty zone* reflects the demands that a single driver puts on the infrastructure. The gross sum of all empty zones aiming to pass a specific cross-section (i.e. the strategic demand), determines the *operational traffic demand*. Clearly, when the operational traffic demand exceeds the infrastructure supply, some vehicles can not be served. For instance, consider a one lane road where 3000 vehicles having an average empty zone of 1.5 seconds want to pass a cross-section within an hour. Since the operational traffic demand equals 4500 seconds, while the supply is only 3600 seconds, some vehicle can not be served and traffic congestion occurs. Vice versa, *the maximum operational demand* that can be served by the infrastructure equals 3600 / 1.5 = 2400 vehicles per hour. This is called the *capacity*. Note that the capacity is not only a property of the infrastructure, but is also determined by the traffic on the infrastructure, and all factors influencing its behaviour (driver behaviour, traffic composition, weather and road conditions, and traffic control influence the capacity).

Finally, note that, although vehicle lengths are constant for single vehicles, they vary considerably for different vehicles in the traffic stream. Moreover, what is considered safe, efficient, and comfortable will vary across the drivers in the driver population (differences in travel purpose, experience, vehicle characteristics, driver attitude, etc.), and for a single driver as well (fatigue level, driving error, time of day, activity at destination, etc.). As a consequence, the empty zone is a random variable.

2.1.3 Stochastic nature of congestion

In the previous sections, we have described the causes of congestion for both the strategic level (strategic traffic demand), and the operational level (infrastructure capacity). On both levels, uncertainty plays an important, if not decisive role: on the strategic level[1], the uncertainty is determined by a.o. variability in activity patterns and related travel choices (trip making, origin and destination, departure time, route, mode choices) of road users. The infrastructure capacity is affected by individual driving behaviour (and the randomness herein), driver and vehicle characteristics, interaction with other drivers, geometry of infrastructure, external conditions, and traffic control. Since congestion occurs when demand exceeds supply, traffic breakdown is a stochastic phenomenon.

Moreover, some of the queues manifesting themselves in the traffic networks are random by their very cause. Just consider accidents, exceptional weather conditions (black ice) or exceptional traffic conditions (demonstrations, strikes, etc.). In illustration, from the 16,000 queues on motorways registered in The Netherlands in 1996, about 4,000 were *unpredictable by their very nature*.

The remaining recurrent queues happen from time to time, sometimes almost daily, more or less at the same location. The ill-predictability of recurrent congestion means that the occurrence of congestion in time, place and duration appears to be a highly stochastic phenomenon. Moreover, despite seemingly similar traffic conditions (e.g. traffic flow levels), it appears that location, day, starting time, ending time, duration and length of queues are subject to large and highly stochastic variations. At a number of well-known bottlenecks in the motorway network, congestion is sometimes absent even at working days during rush hours, without any observable cause. Clearly, the ill-predictability differs between locations. At the well-known bottlenecks (bridges, tunnels) or discontinuities (on-ramps) prediction success is higher than elsewhere, but also there precise prediction remains a gambling game with a highly uncertain outcome.

2.2 Definition of congestion

Congestion occurs when traffic demand exceeds infrastructure capacity. This definition appears simple enough. Nevertheless, researchers and practitioners alike have established a very large collection of, sometimes conflicting, criteria to classify traffic flow operations as congested or not. These criteria relate to among others traffic speed

---

[1] In addition to known systematic variations due to hour, day, and month factors, the level of traffic demand on heavy loaded motorways during peak hours is subject to random variation of at least 5% relative standard variation (Bexelius,1995).

(80 km/hr criterion), traffic densities (critical density of 27 veh/km/lane), difference in speeds between person-cars and trucks, or between roadway lanes, etc. Moreover, besides free-flow, and congested traffic flow, different traffic states have been identified (e.g. synchronised flow, start-stop waves). In illustration, Table 2 Research problem and focus-1 consolidates different states in traffic flow, from Kerner (1999) and Helbing *et al.* (1998). The table indicates how these states can be characterised based on the densities, speeds, speed differences with respect to lanes, and classes, and overtaking probabilities. Also the causes for the occurrence of these states is indicated.

Table 2 Research problem and focus-1 Different regimes in traffic flow (Very Low, Low, Medium, High, and Very High; Similar and Different); osc = oscillating, var = varying.

Regime	Characterisation					Cause
	density	**speed**	**lanes**	**classes**	P_{over}	
free flow	L	H	D	D	H	-
synchronised flow	M-H (var.)	L	S	S	L	Spontaneously (metastable and linearly unstable traffic conditions); Bottlenecks ('pinch effect')
single localised clusters	L/H (osc.)	H/L (osc.)	D	D	M	Perturbation at slightly oversaturated bottleneck with low on-ramp flow
start-stop waves	L/H (osc.)	H/L (osc.)	D	D	M	Perturbation at oversaturated bottleneck with low on-ramp flow
oscillatory congested traffic	M/H (osc.)	M/L (osc.)	D	D	M	Perturbation at oversaturated bottleneck with medium on-ramp flow
homogeneous congested traffic	VH	VL	S	S	L	Perturbation at oversaturated bottleneck with high on-ramp flow.

From the perspective of the traffic stream, congestion can also be defined from a physics viewpoint, by considering the direction in which ’flow information’ is transported in the traffic stream. When traffic conditions are determined by upstream conditions only (upstream demand < downstream supply), traffic operations are free flow; when downstream conditions (partially) determine traffic operations (upstream demand > downstream supply), congestion occurs.

Hoogendoorn *et al*. (1998) describe congestion by means of fuzzy variables. In doing so, *levels of congestion* can be assigned to (combinations of) specific traffic conditions (speeds, densities), while crisps thresholds (when congestion occurs when the velocity is smaller than 80 km/hr, 79 km/hr is congested flow, while 81 km/hr is free-flow) can be refrained from. Moreover, fuzzy sets relate closely to the way drivers perceive congestion.

2.3 Aspects of congestion

In this research, the following major aspects of congestion are considered:

1. Mechanisms of traffic breakdown, and the dissolution of congestion: bottleneck modelling (primary aspect);
2. Propagation of congestion: blocking back, blockages and deadlocks / gridlock effects (secondary aspect);
3. Congestion traffic flow behaviour: driving behaviour and traffic stream characteristics within a congested area (e.g. stop-and-go, shockwaves).

The first aspect defines the way bottlenecks are modelled and what the criteria are for the *occurrence of congestion* (e.g. volume-capacity ratio's). The second aspect determines how the congestion propagates to other elements of the road network (via network interconnections) and which other traffic streams are hindered by it. The third aspect has to do with driving behaviour (e.g. distance keeping, lane changing or shockwave propagation), and the consequent traffic flow operations, within the congested area. Especially the second aspect plays an important role in network analyses, since blocking back can result in gridlock.

2.4 Focus and limitations of the research

A dynamical representation of traffic flow is required as the effects of (dynamic) traffic management measures are of interest in model studies. A good reference is available in the "*Leidraad modelstudies verkeersbeheersingsmaatregelen*" of AVV. Models and studies mentioned therein are of first consideration in the VAC project (see Table 2 Research problem and focus-2).

2.4.1 Model classifications

Traffic flow models can be distinguished using a number of criteria, e.g.:

1. Level-of-detail (microscopic, mesoscopic continuum, macroscopic continuum);
2. Network type (interurban, rural, urban, mixed);
3. Scale of application (networks, stretches, links, and intersections);
4. Representation of the processes (deterministic, stochastic);
5. Operationalisation (analytical, simulation);
6. Scale of the independent variables (continuous, discrete, semi-discrete).

In the remainder of this report, the level-of-detail (1) is of main interest. Moreover, with respect to network type (2), only motorway networks are considered.

The level of detail refers to the *distinguished traffic entities* and the *description level* of these entities in the respective flow models. In this project, flow models are classified into three main classes, namely:

a. *Microscopic simulation models*, where both the *space-time behaviour* of the systems' entities (i.e. vehicles and drivers) as well as their interactions at a high level of detail (individually).
b. *Continuum (gas-kinetic and macroscopic) models*, where traffic flow is described in aggregate terms, considered as a continuous media (comparable to a fluid or a gas).
c. *Mixed models*, where features of both microscopic models and continuum models are included.

Table 2 Research problem and focus-2 Modelling techniques for congestion aspects

Level of modelling	**Microscopic**	**Continuum**	**Dynamic assignment**
Models considered	Flexsyt-II, Integration, Aimsun2, MIXIC, FOSIM	METANET, FlowSimulator, STM	Contram, Dyndart, 3DAS
Bottleneck modelling	▪ Following behaviour ▪ Lane change behaviour (incentive and gap acceptance) ▪ Fundamental diagram	▪ Volume to capacity ratio ▪ Fundamental diagram ▪ Relaxation time ▪ Kinematic viscosity ▪ Anticipation coefficient	Volume-capacity ratio's
Propagation of congestion	See bottleneck modelling	▪ Jam density ▪ Speed dynamics	Storage capacity
Congestion behaviour	See bottleneck modelling	See bottleneck modelling	▪ Fundamental diagram ▪ Delay functions

This research is mostly concerned with traffic operations at the operational level. On the contrary, (dynamic) assignment models focus on travel decisions at the strategic level as well (route-choice, departure time-choice). It is clear that the level-of-detail (microscopic, continuum, or assignment) is of major influence on the way the above described congestion aspects can be represented. At these levels the models use different techniques to model the different congestion aspects: an overview is given in Table 2 Research problem and focus-2. Based on representative models of each level used in Dutch studies, these techniques are given.

The restriction to occurrence and propagation of congestion is represented by the lightly shaded rows in the table. A further restriction (shaded colums) follows from the observation that dynamic assignment models appear to employ more or less the same congestion modelling techniques as the macroscopic models, which means that this level needs no special attention. Finally, in the microscopic models it appears that all congestion aspects are derived from the driver behaviour (car following and lane changing) descriptions. The remaining darkly shaded cells of Table 2 Research problem and focus-2 indicate the restriction of the VAC project, resulting from the discussion above.

2.5 Demarcation of the research

In the remainder of the report, we will assume that the strategic demand is known. In other words, travel choice mechanisms that describe how traffic demand patterns are caused by route choice, departure time choice, and mode choice, are not considered in this report. Rather, we will focus on traffic congestion issues at the operational level (shortage of roadway capacity), that is, for known traffic demand levels. More specifically, the emphasis in the research presented in this report lies on the description of (individual) driver behaviour, and the interactions with other drivers in the traffic stream (shaded area in Figure 2 Research problem and focus-3**)**.

Although the main focus is not on the way the infrastructure and controls can be or should be modelled, it may be that attention has to be paid to these factors. In some cases (e.g. representation of motorway connections, on ramps and off ramps, etc.) the infrastructure may decisively influence driving behaviour and accuracy of congestion modelling. Where this occurs, attention has to be devoted to these influences.

The focus furthermore is on *congestion modelling principles* and not on congestion inaccuracies that are the result of configuration and calibration problems and that may be reduced without modifying the modelling principles. It is not yet clear how configuration and calibration can be separated from the fundamental modelling principles. It is known that generic model parameters are sometimes adjusted per application in order to achieve a good fit for the specific situation. Although this approach may be practical in a specific study, it is not good practice, as it disguises the (possibly limited) predictability of the model results (when enough parameters are available any model can be made to fit any data-set, a good model is a limited model).

2.6 Problem identification and research issues

In the previous section, we have discussed the focus of the research congested traffic flow theory and modelling. It has been decided that research will focus on modelling of congested traffic flow operations at the operational level (shaded cells of Table 2 Research problem and focus-2).

2.6.1 Results obtained from Dutch application studies

To get a clear insight into the problems that are currently experienced when applying traffic flow models to congested traffic flow, several model descriptions and study reports of models used in the Netherlands were collected and scanned for reported congestion flow modelling problems. A compromising issue when identifying problems with congestion modelling at the operational level is that most reports consider networks, and consequently traffic assignment. This is why it is difficult to address inaccuracies and modelling problems to either the strategic or to the operational level. Moreover, as it turns out none of the considered reports specifically mention congestion modelling problems at all. Sometimes general remarks on the modelled speeds and the level of congestion are made, but in general no causes are identified nor are these causes traceable from the reports. A final, but possibly very important point, is that it still unclear if errors are caused by incorrect model calibration, or flawed modelling principles.

Despite these compromising issues, Table 2 Research problem and focus-3 provides an overview of problems encountered in Dutch application studies. From these, it appears that only a few general results hold with respect to what kind of problems are encountered. In some cases, congestion occurs too soon, while in other cases, congestion does not occur at all. In general, the models results are less random than the measurements. Also, it appears that most of the considered models do not correctly describe the queue head location (based on the comparisons of observed and simulated speed-density curves), and that lane-changing and merging behaviour is not correctly described. Appendix A of this report discusses the relevant research findings with respect to model application reported in the literature.

Concluding, before definitive conclusions can be drawn regarding the incorrect description of congested traffic flow operations, accuracy and reliability of simulation models used by Dutch practitioners must be rigorously tested, *using a simple benchmark situation*. Although it is clear that problems exists, thorough problem identification and consequent cause diagnoses *must still be performed*.

Table 2 Research problem and focus-3 Overview of problems encountered (and possible solutions) in Dutch application studies for different (microscopic) simulation models. For details, see appendix A.

Identified problem (based on comparison with empirical data)	**Models (reported)**	**Possible solution (from reports)**
1. The level of randomness is too low, and simulation results are generally too smooth.	FOSIM FLEXSYT-II AIMSUN2	Introduce more variation between drivers (*inter-driver differences*) Introduce stochasticity in driving behaviour (car-following, lane-changing behaviour; *intra-driver differences*).
2. Transitions from free flow to congestion are generally too sharp or too sudden.	FOSIM AIMSUN2	Improve description of drivers anticipation (drivers do not only consider their direct predecessor, but look further downstream, for instance two-vehicles ahead).
3. Location of the queue head is not correctly described.	FOSIM	Introduce (more) smooth velocity adaptation to attain desired headway.
4. Drivers tend to change lanes too late compared to real data (in case of mandatory lane change).	FOSIM	Reconsider relations describing (mandatory) lane-changing behaviour in case of discontinuity Model calibration.
5. Merge-giveway behaviour is not correctly described (drivers do not provide gaps to merging traffic).	FLEXSYT-II AIMSUN2	Prioritise traffic on on-ramp in merging onto the main-road. Include merge-giveway rules.
6. Differences between free-flow and congested operations too small.	FLEXSYT-II	Extensive model calibration and validation Reduction of capacity values used in model
7. Differences between roadway lanes too small (distribution of trucks).	FLEXSYT-II	Overtaking prohibition for trucks in model Improving class-specific lane changing behaviour.
8. Use of the shoulder-lane is overestimated; in reality, drivers tend to stay on the middle of median lane for some time after overtaking manoeuvre is completed.	FOSIM AIMSUN2	Reconsider lane-changing behaviour mechanisms.
9. In real-life traffic flow, drivers on the target lane of an lane-changing vehicle will generally adapt their behaviour (provide a gap).	AIMSUN2 FOSIM	Inclusion of merge-giveway behaviour
10. Drivers attempt to change lanes too frequently (e.g. leading vehicles is slower than free-speed of follower). This yields unrealistic lane-changing behaviour, especially during congested traffic flow.	AIMSUN2 FOSIM	Include indifference bands for lane-changing incentives; include modelling mechanisms where driver decides whether a lane-change is useful, given anticipated traffic conditions.

Identified problem (based on comparison with empirical data)	Models (reported)	Possible solution (from reports)
11. Driver behaviour is constant for different regimes and network components.	AIMSUN2 FOSIM	multi-regime approach (e.g. regime dependent parameters).
12. Incorrect description of location of congestion in the network in case of traffic control	INTEGRATION	Implementation of DVM and traffic control measures must be improved.

2.6.2 Opinion of Dutch traffic modelling experts

In obtaining insights into the issue of congestion flow modelling as it occurs in the models currently applied in practical studies in the Netherlands, Dutch experts in the field of traffic flow modelling were interviewed. These experts (Henk Taale (Traffic Research Centre of the Dutch Ministry of Transportation, Roadworks, and Water Management), Jaap van Toorenburg (TRANSPUTE), and Bart van Arem (TNO-INRO) were asked about their opinions and experiences in the field of traffic flow simulation and modelling. Let us give a short summary of the results from the expert interviews (for the raw interview material, we refer to appendix A):

1. Little empirical knowledge is available about driver behaviour at the onset and dissolution of traffic congestion (traffic breakdown). For instance, we do not know why, and where drivers change lanes? Which headways are accepted (or chosen) before decelerating or accelerating, and how long these are accepted? What are the determining factors in this behaviour, and how random is this behaviour? This lack of knowledge is fundamental and relates to actual driving behaviour, not just to the modelling of it.
2. An extensive and costly data collection programme would be needed to increase the insight into this behaviour. The data collection would preferably consist of a combination of objective, subjective, infrastructure and vehicle related techniques. Some experts are pessimistic about the chance of success of this kind of research. Does it lead to definitive conclusions and formulations that can be used in practical models? How to incorporate this detailed knowledge in mesoscopic and macroscopic models?
3. Another lack of knowledge concerns the actual quality of the congestion modelling in the models currently used by practitioners. As it turns out, most models have not yet been validated for congested situations, or have not been validated at all. The congested modelling problem as formulated above emerge from general application studies in which the problem causes and the effects of modelling errors and inaccuracies cannot be clearly identified. For instance, it is unclear whether incorrect modelling of queue lengths is due to inaccuracies at the operational modelling level, or that traffic is not correctly assigned to the respective roadway routes. Moreover, it is also unclear whether observed inaccuracies are caused by faulty modelling principles, or by inaccurately calibrated models.
4. There is a clear need for separate validation studies, specifically dedicated to congestion modelling. Preferably, there would be a benchmark situation (single bottleneck in a stretch of motorway sections, no route choice, known traffic

demand) to which all models are applied and validated. Some literature studies use such examples, but currently used models generally are not yet validated in this way.

5. A fundamental question that arises is which level of modelling is actually needed in the effect studies? Is microscopic modelling required? Does the model have to fit a reference situation exactly? Most studies aim to estimate the average effect of traffic controls and infrastructural changes only. This means that small errors in modelling e.g. capacities would be acceptable, as long as the average situation is correctly represented. This holds in particular if one realises that the same type of errors originate from day-to-day variations in level of demand, origin-destination distribution, route choice and the occurrence of small unknown incidents or disturbances.
6. The previous observation leads to the idea or proposal to introduce sensitivity analyses in the studies involved. In most cases the sensitivity of the simulation results to input or parameter variations is not investigated (this may be as a result of trying to fit the model as close to an existing situation by calibrating all available parameters). One should identify the most critical model elements in congestion modelling and determine the level of predictability that can be attained and expected per model (and situation).
7. No consensus exists on the importance of the respective aspects that cause traffic congestion. Some experts consider the behaviour within the congested area to be of importance for the dissolution and propagation of it. Similarly, the experts appear to disagree on the relevance of traffic hysteresis and the so-called capacity drop: while some consider it an importance mechanism causing traffic breakdown, others see it as a theoretical discussion, as in practice it turns out that congested traffic flows at principally the same capacity as free flowing traffic.
8. In macroscopic modelling, accurately describing bottlenecks and modelling the propagation of queues are the main research issues.

From the points above, it is clear that there is a lack of knowledge on both the actual congestion flow processes in real-life traffic flow operations, as well as on the quality (accuracy, and reliability) of current traffic flow models. To heighten the knowledge state, an extensive data collection and research programme may be needed. However, the success of such a program in uncertain. Nevertheless, improving the quality of modelling requires a clear statement of the level of modelling needed, which can be obtained from the formulation of the problem at hand, and requires reliable model calibration and validation procedures.

2.6.3 Opinions of international experts

In the second phase of the research, interviews with international experts were conducted. These interviews resulted in the following viewpoint with respect to problems encountered when modelling congested traffic flow operations (see Hoogendoorn and Alkim (1999)):

1. Current traffic flow modelling principles contain no fundamental theoretical flaws, within the context and scope of application of the models. The lack of empirical knowledge is the main problem.

2. Most models do not describe self-organising phenomena (i.e. from free flow to synchronised flow, from synchronised flow to traffic jams, from free-flow to traffic jams), which have been observed in real-life traffic flow. Some models (i.e. second order macroscopic models) can describe self-formation (and stability) of wide jams (free-flow to jam).
3. While first-order macroscopic models do not correctly describe how vehicles accelerate *from* congestion, higher-order macroscopic models do not correctly describe congestion spillback.
4. None of the models appears to correctly describe lane-changing in general. In particular, the mechanisms of traffic flow near a merge or diverge is flawed. For instance, a problem is the description of a congested exit with spillback and the consequent effects on the lane-changing behaviour of drivers.
5. Current understanding of the relation between data and models is insufficient, especially on the level of interactions between vehicles and the driver itself. The aim of such research should be to determine how people react on information. The impact of information should consequently be used as an input of the system.

2.6.4 Literature survey

During the literature survey, problems in describing congested traffic flow have been have been identified. In this case, issues concern the following aspects:

- Empirical data reveals the existence of specific phenomena in congestion, relevant for a correct description of traffic flow, given the foreseen model application, but is not correctly predicted by the currently used models (e.g. hysteresis, self-organisation, merge-giveway behaviour);
- Complications with application of the models currently used (for instance with respect to model calibration, increasing model complexity, etc).

The following presents an overview of the problems identified from the literature survey (for a detailed analysis of the different contributions, we refer to appendix C of this report).

1. Empirical data supports a difference in high prior-congestion flow-rate and queue discharge flow, not incorporated in most models.
2. Observations suggest a (predictable or random) displacement of the queue head with respect to physical bottleneck (500-1000m downstream). Current models do not describe this displacement correctly.
3. In traffic flow, wide jams propagating upstream are observed. Most simulation models do not incorporate mechanisms describing propagation of these (stable) wide jams.
4. Traffic data reveals that driving behaviour (car-following, lane-changing behaviour) depends on traffic regime (free-flow or congested flow), and location in the network (homogeneous stretch, merge, on-ramp, off-ramp, etc.). Most models do not differentiate between either of these.
5. There are distinctive differences in driving behaviour for different classes during the respective flow-regimes. Models should correctly describe differences in driving

characteristics of the respective classes, as this influences both microscopic and macroscopic traffic flow behaviour.

6. Increasing complexity microscopic models prevents straightforward calibration of the models.

2.7 Conclusions with respect to problem identification

Based on results of expert interviews, Dutch application studies, and the literature survey, several key issues needing further attention have been posed:

1. Need for a *benchmark case* (single bottleneck, no route-choice, sufficient and detailed calibration and validation data) to calibrate, validate and compare different traffic flow models. This is an essential step in precisely identifying current problems in practical model applications, diagnoses, and determining solution approaches.
2. Driver and traffic stream behaviour at the time and location of traffic breakdown (congestion onset):
   - What are important factors in traffic breakdown, determining among others its time and location? How (ill-) predictable are these breakdown times and locations?
   - Is traffic hysteresis important (from a practical viewpoint)?
   - What factors influence the difference is prior- and post-breakdown traffic flow rates, and are these differences significant?
3. Propagation of congestion (standing queues as well as upstream moving wide jams) in models.
4. Relation of the points made above with *macroscopic modelling*.
5. Dependence of driving and flow characteristics on both the regime as well as the location in the network:
   - To what extent does driving behaviour depend on the flow regime (car-following behaviour, lane-changing behaviour), and what are the implications for modelling?
   - How does driving behaviour change in the presence of discontinuities?
6. Importance of vehicle-type / user-class distinction in practical applications (control, cost-benefit analysis, modelling accuracy).
7. Increasingly complex model calibration due to complexity of microscopic models, and the importance of sensitivity analysis in model calibration:
   - How do parameter changes influence driving and traffic stream behaviour?
   - How sensitive are these changes?

In the remainder of this report, we aim to answer some of these questions by considering the current state of the art (both empirical studies and traffic flow theories) in congested traffic flow theory. Secondly, chapter 4 discusses how the different relevant issues have been incorporated in traffic flow models. For the unresolved issues of chapter 3, chapter 5 provides data criteria that need to be posed in order to assure that empirical analysis will resolve the research questions. Also, data criteria for model calibration and validation is discussed.

3 STATE-OF-THE-ART REVIEW ON CONGESTED TRAFFIC FLOW

What is the current state of knowledge concerning traffic congestion? What do we know about its causes? Can we predict traffic congestion, or is traffic breakdown mostly a stochastic phenomenon (at the operational level)? Do phantom jams exists, and are they relevant in the description of traffic flow?

In this chapter, we aim to answer some of the questions above by considering results from both the literature survey and the expert interviews, with respect to the issues summarised in section Problem identification and research issues. For convenience sake, Table 3 State-of-the-Art review on Congested Traffic Flow-4 summarises the main results of the survey and the expert interviews. In the remainder of this chapter, these issues are discussed in detail.

Observed phenomena	Available flow theory	Relevance	Research questions
1. Difference in high prior-congestion flow-rate and queue discharge flow (capacity drop); differences in range of 5-10% upto 50%.	Hysteresis-related: traffic flows into congested area with higher velocity than when leaving the congested area. Difference between *ancitipation dominant stage* and *relaxation dominant stage (Zhang,1999).* Catastrophe theory and utility maximisation of drivers (Van Toorenburg (1983)).	DTM measures postponing traffic breakdown Relevance for Dutch networks stems from a.o. peak-shaving due to ramp-metering	How common is the capacity drop, and to which extent does it influence correct description of congested traffic flow? What are the dominant factors influencing it (e.g. geometric design of discontinuity, traffic composition)?
2. Phase-transitions in presence of discontinuities; observations of high-prior congestion flows for considerable period.	See Difference in high prior-congestion flow-rate and queue discharge flow (capacity drop); differences in range of 5-10% upto 50%..	Describing queue dynamics. DTM measures postponing traffic breakdown (metering)	What causes the precise instant of traffic breakdown? Is it at all predictable? What are the dominant factors determining the duration of the high-prior flow period, and how does it relate to these factors?
3. Traffic hysteresis	See Difference in high prior-congestion flow-rate and queue discharge flow (capacity drop); differences in range of 5-10% upto 50%..	Fundamental behaviour, underlying capacity-drop and high-prior congestion flows.	What are the factors influencing hysteresis, and how do these related to hysteresis quantitatively?

Observed phenomena	Available flow theory	Relevance	Research questions
4. Occurrence of spontaneous transitions (e.g. free-flow to synchronised flow).	Distinction of stable, metastable, and unstable traffic flow. Non-linear / chaos-like behaviour of traffic flow.	DTM measures increasing stability (e.g. traffic calming and speed homogenising). Relevance for Dutch motorways is unclear due to network topology.	How frequently does spontaneous breakdown of traffic flow occur in the Dutch case (if at all)? Is the phenomenon predictable?
5. Phase-transitions in presence of discontinuities; displacement of queue head with respect to physical bottleneck (500-1000m downstream)	Drivers smoothly adapt their headway and velocity, e.g. in case of merging drivers.	Spillback of congestion; blocking back effects in traffic networks. Correct testing geometric design.	What are the factors determining the precise location of traffic breakdown? Can this location be at all predicted?
6. Location of the tail of the queue.	Dynamics of the queue tail depends on the type of congestion that emerges (homogeneous congested traffic, oscillatory congested traffic); Shockwave and queue theories are applicable, depending on the congestion type.	Spillback of congestion; blocking back effects in traffic networks.	Under which conditions, and for which congestion regimes, can the location of the queue tail be predicted? Are shockwave and queue theories applicable?
7. Correct description of propagation of wide jams	Wide jams either result from bottlenecks or are self-formed. Stability requires jam inflow to be *at least as large* as jam outflow for a considerable period (slow-to-start rules; see hysteresis).	Congestion spillback Traffic safety (congestion warning systems and traffic calming)	What are the characteristic properties of wide jams, and on what do these characteristics depend? What are their most common causes?

	Observed phenomena	Available flow theory	Relevance	Research questions
8.	Observation of different types of congested traffic flow	Kerner (1999): free-flow, synchronised flow, and traffic jams. Helbing *et al*. (1998): homogeneous congested traffic, localised clusters, start-stop waves, oscillatory congested traffic, free-flow	Fundamental behaviour to improve description and understanding of congested traffic flow	Which states can be distinguished? What are their characteristics and causes? Which of these are most common in (Dutch) motorway networks?
9.	Differences in car-following behaviour free-flow / congested flow	See hysteresis. Driving behaviour is dependent on a.o. prevailing traffic regime.	Corrective description of traffic flow in congested region	What are the factors that determine structural behaviour of drivers?
10.	Observed differences in lane-changing behaviour, dependent on location of drivers.	Drivers tend to change lanes to the median and middle lane, upstream of an on-ramp, sometimes causing congestion to appear there first. Drivers on the main road give-way to drivers from the on-ramp.	Corrective description of traffic flow in congested region; instant (and location) of traffic breakdown. Correct description of where congestion occurs (on main-road rather than at on-ramp).	Can structural relations between lane changing behaviour and factors like distance to discontinuity, on-ramp traffic demand, time-of-day, etc. be determined? Are these general, or do they pertain to a specific situation?
11.	Observed differences between user-classes	During free-flow differences between classes are considerable. During homogeneous synchronised flow / homogeneous congestion, differences are diminished; differences are large during oscillatory congestion.	Corrective description of traffic flow in congested region. Scattered observations during synchronised flow operations	To what extent can class-distinction improve the description of congested traffic flow? Under which circumstances are the differences between user-classes considerable?

Table 3 State-of-the-Art review on Congested Traffic Flow-4 Overview of literature survey and expert interviews concerning empirical studies and traffic flow theories explaining the observations of traffic flow behaviour during congestion.

3.1 Capacity related issues

From the literature survey, it appears that the most common cause for traffic breakdown is that traffic demand exceeds the capacity. Consequently, the 'capacity' is the most important parameter in describing congested traffic flow operations, designing, planning, etc. In this section, several issues concerning the roadway capacity found in the literature are summarised.

3.1.1 Capacity definition

The most common definition of capacity is, is the maximal flow-rate that can be expected to sustain for a reasonable time period. Depending on what one calls reasonable, capacity can either reflect 1) the queue discharge rate, or 2) the high prior-congestion flow-rate, which is known to maintain for relatively short periods only (from 0 to 40 minutes)[2]. Depending on ones viewpoint, either definition is applicable, and useful. For instance, considering a congested network, in which prediction of congestion propagation is essential to predict whether congestion spillback occurs. In this case, it is useful to define capacity by the *queue discharge rate*. However, in controlling traffic flows e.g. by ramp-metering, to postpone or even prevent traffic breakdown, the capacity equals the higher prior-congestion flow rate.

Another definition of capacity is based on a probabilistic description of traffic flow breakdown, by the maximum flow rate at which the probability that traffic breakdown occurs within an hour equals 50%.

In the remainder of this report, the *capacity* generally refers to the *queue discharge rate*, whereas the term (*maximum*) *prior-congestion flow-rate* is reserved for the high flow-rates that are observed before congestion has set in.

3.1.2 Differences in prior-congestion flow-rate and queue discharge rates

Various contributions discuss the issue of the so-called *capacity drop*, that is, the phenomenon that before traffic breakdown, very high flow-rates are observed. These flow-rates that are *significantly higher than the queue discharge rate* observed after the on-set of congestion. Although the capacity drop and the time flow breakdown occurs are clearly related, in this section, prediction of time and location of the onset of congestion is not discussed (see section Traffic breakdown and bottlenecks).

On a microscopic scale, traffic breakdown occurs when the headway of drivers is smaller than optimal (from a efficiency and safety perspective). This can occur under different conditions, but the most common is near on-ramps and merges. In this case, vehicles from the on-ramp enter the main road, causing average headways of drivers to reduce. Although (some) drivers may temporarily accept these smaller headways (high flow-rates), drivers will increasing their headway 0after some time (reduction of the flow-rate).

---

[2] Another viewpoint is given by Hertel (1994), who states that the capacity (distribution) of a facility can only be defined in relation to the corresponding quality of traffic flow. This quality notion includes aspects such as reliability. If a high level of quality is required, that is a low probability of disfunctioning of the facility, the corresponding capacity is low. Hertel states that the maximum capacity (or rather, limiting capacity) of a facility is defined as its ability to achieve the maximum throughput under the full utilisation of personal capabilities, means of transportation and available infrastructure.

Note that the concept of a capacity drop (differences in pre-congestion capacity and queue discharge rate) implies that following behaviour *before*, *during*, and *after* congestion is different (hysteresis). Let us further note that the issues discussed in this section hold only for motorway traffic roadways and discontinuities herein. For highways, or urban roads, where the bottleneck is caused by traffic lights, the existence of the capacity drop has never been established empirically, nor is the existence of the capacity drop to be expected. On the contrary, drivers upstream of a traffic light will mostly try to catch the first available green phase, and will speed up to do so, causing the headways of vehicles leaving the jam to be relatively small, rather than comparably high.

3.1.2.1 Empirical results

Observed differences in maximum prior-congestion flow rates and queue discharge rates are 5-10% (Cassidy and Bertini (1999a)), upto 50% (Kerner (1999)). For the Dutch situation (A9 motorway), Dijker (1997) discusses similar results. The high pre-congestion flow rates are generally only measured for a relatively small time duration (say, 10-30 minutes), and can as such not regarded as the *capacity* of the bottleneck. Rather, the capacity equals the queue discharge rate. Among others, Cassidy and Bertini (1999a) observed that queue discharge rates are nearly stationary, and are very similar over the different days.

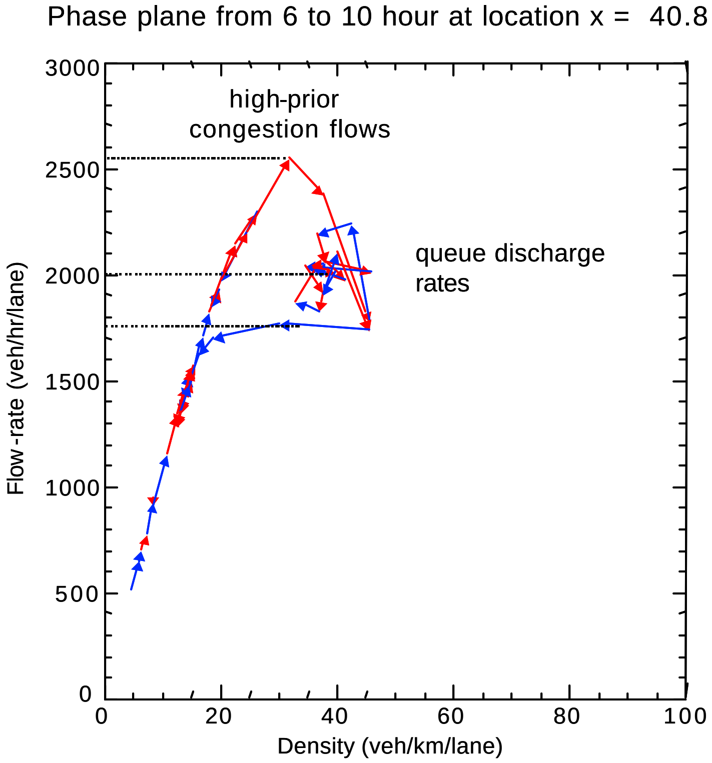

Figure 3 State-of-the-Art review on Congested Traffic Flow-4 Measurement results from A9 motorway (downstream of on-ramp) based on 4 minute periods. In this case, the high prior flows last for approximately 20 minutes.

In illustration, Figure 3 State-of-the-Art review on Congested Traffic Flow-4 shows a typical morning peak hour (from 6:00AM to 9:30AM) on the A9 motorway (21st of

October 1994). The data were collected downstream of a bottleneck (x = 40.8km). In this case, the difference in the high prior flows and the queue discharge rate is very clear, and significant. Moreover, it appears that the high prior flows last for a considerable time, namely 20 minutes. The length of this period varies across the different days (from 0 to 30 minutes). The height of the prior-flows is also not constant, as it varies from day to day. For more details, we refer to Dijker (1997).

Another interesting issue raised by Cassidy and Bertini (1999a) is the drop following the high prior-to-congestion flow rate to the *minimum queue discharge rate*, lasting up to 25 minutes, then taking on a higher, stationary pattern. However, other studies (and the empirical investigations performed during this research) did not reveal these dynamics at all. Cassidy and Bertini (1999a) it appears that the day-to-day variation in the queue discharge rate is relatively small. Since the queue-discharge rate is most important in describing the propagation of congestion (queue tail), it can be argued that there is no need to strive towards predicting the detailed short-time variations in this capacity.

In analysis car-following behaviour prior and during congestion, Dijker (1997) shows that differences exist between the respective roadway lanes, and user-classes. Using data on a two-lane motorway, it appears that *person-cars on the left lane* indeed maintain smaller headways before that during and (some time) after congestion, i.e. the headway is a discontinuous, regime-dependent function of the density. This generally does not hold for the right lane.

As a final point, Zhang (1999) argues that the differences reported in different studies are mostly due to the alternative treatment of data. Without a clear definition of traffic equilibrium and a traffic theory to guide the proper processing of raw data, many previous studies lumped together every bit of data collected, leading to two serious problems: 1) data contain inherently different traffic phenomena; 2) aggregation of data over longer time intervals than the acceleration / deceleration 'wave-length' adds considerable distortion to obtained flow relationships.

3.1.2.2 Traffic flow theories explaining capacity drops and hysteresis

Although the capacity drop and traffic hysteresis are not the same, they are caused by the same driving behavioural mechanisms, and stems from the interplay of *anticipation behaviour* of drivers and their *acceleration characteristics*. Zhang (1999) shows that depending on the anticipation distance, the relaxation time (reflecting acceleration and deceleration characteristics), and the traffic density, the velocity is either above or below the equilibrium velocity. He argues that, when going through strong disturbances, drivers responses go through the following three stages:

1. Anticipation dominant phase.
2. Balanced anticipation and relaxation phase.
3. Relaxation dominant phase.

It is hypothesised that, for dilute traffic conditions, drivers are aware of downstream traffic conditions, such that they adapt their speeds before reaching the disturbance (stage 1). In this case, the anticipation effect is dominant. When traffic is heavy, a driver cannot see that a disturbance is coming until it reaches him. Therefore, the response to the disturbance is retarded (stage 3). Stage 2 is a intermediate stage between the two extremes.

Drivers in congested flow are generally constrained (following their leading vehicle), that is, they are in stage 3. When emerging from the congested zone, the velocity (flow rate) is retarded and below the equilibrium velocity (equilibrium flow rate), contrary to drivers that do not emerge from congestion (stage 1 or 2). Clearly, this theory can explain the differences between high prior flows and queue discharge rates plausibly.

Another interesting concept in describing the capacity drop, is *catastrophe theory*. Catastrophe theory is a relatively new research area in mathematics, having started in the early sixties. In oversimplified terms, catastrophe theory describes systems in which specific variables can show irregular behaviour, while focussing on this discontinuous behaviour (see Zeeman (1977)). With respect to traffic flow, both applications with respect to macroscopic modelling as well as microscopic modelling are known.

On a *macroscopic level*, applications of catastrophe theory in traffic flow focuses on describing the transitions from non-congested traffic flow to congested traffic flow, and vice versa, and the hypothesised drop in the capacity value.

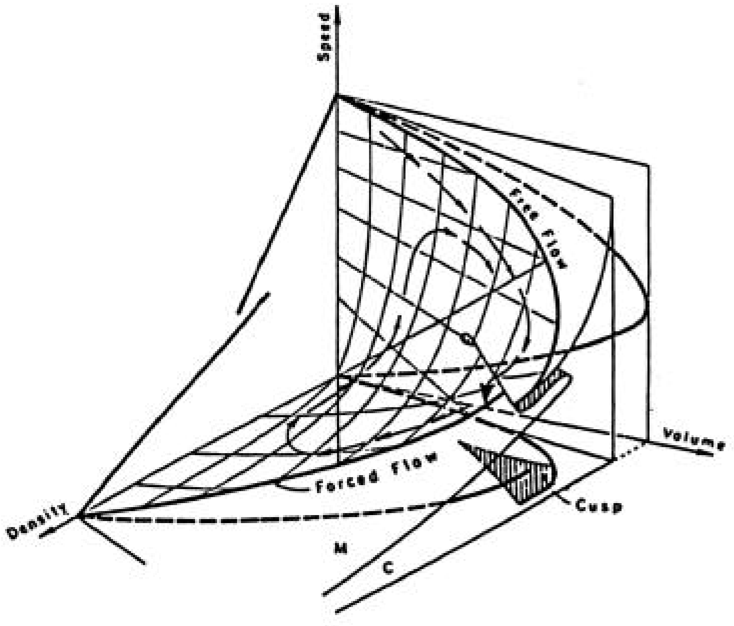

Figure 3 State-of-the-Art review on Congested Traffic Flow-5 Fundamental diagram in three dimensions according to Navin (1986),

Figure 3 State-of-the-Art review on Congested Traffic Flow-5 shows the fundamental diagram in three dimensions, according to Navin (1986). Key to the catastrophe theory is the existence of a *cusp* (or *fold*), resulting in a ‘boundary’ from which the traffic state can drop (non-congested flow to congested flow). Recuperation (congested flow to non-congested flow) can only occur via a detour with moves besides the fold. In this way, hysteresis results, i.e. the traffic state dynamics describe closed curves in the phase plane. Applications of catastrophe theory are, also using empirical data, are found in Forbes and Hall (1990), and Gilchrist and Hall (1989); more recent publications have not been found.

For *microscopic applications*, preliminary to the application of catastrophe theory on a microscopic scale is that drivers optimise their behaviour on some level. To operationalise this concept, it is assumed that costs and benefits of driving can be described by a function of a specific behavioural parameter, say *a*. The function describing the profit, is dependent on the traffic flow *q*. Consequently, the optimum of the function is different for different flow-rates. Toorenburg (1983) describes the costs and benefits by functions of the behavioural parameter *a*, reflecting the amount of effort a driver invests to optimise his travel time. Figure 3 State-of-the-Art review on Congested Traffic Flow-6 depicts the dependence of the *benefits* on the behavioural parameter *a* and the flow-rate *q*. For a fixed flow-rate *q*, the benefits (travel time gains) will increase with increasing *a*, first steeply, later more moderately. Since for higher flow rates *q*, driving is more constrained, a driver cannot increase his benefits by putting in more effort, to the same extent as for low flow-rates. Also the costs (lane changing effort, fuel, safety) are as a function of *a* and *q* are shown in Figure 3 State-of-the-Art review on Congested Traffic Flow-6. Both low values as well as high values of *a* lead to additional costs. Their is a clear ‘middle area’, reflecting adaptation of the driver to the surrounding conditions, where the costs are minimal. The *utility* (benefits minus costs) are also shown in Figure 3 State-of-the-Art review on Congested Traffic Flow-6. Clearly, the optimum is located a lower values of *a* for increasing flow-rates *q*. For low flow-rates, these optimal values are on the curve *A*.

From a specific intensity value, the utility function shows a second maximum (curve 4), which becomes the global maximum for increasing flows (curve 5, and beyond). This implies that the optimal values of *a* are a discontinuous function of *q*. Clearly, the theory depends on the choice for the curves of the benefit and cost functions. Main use of these models is a provide a thinking model, explaining how the capacity drop can be explained from the behaviour of the driver.

The current state of affairs is as follows:

- It is plausible that transitions in traffic flow can be described using catastrophe theory.
- The most appropriate variant has not yet been determined.
- Benefits from application of the theory stem mostly in gaining insights into the mechanisms of transitions in traffic flow operations (thinking models), rather than providing applicable mechanisms for operational flow theories.

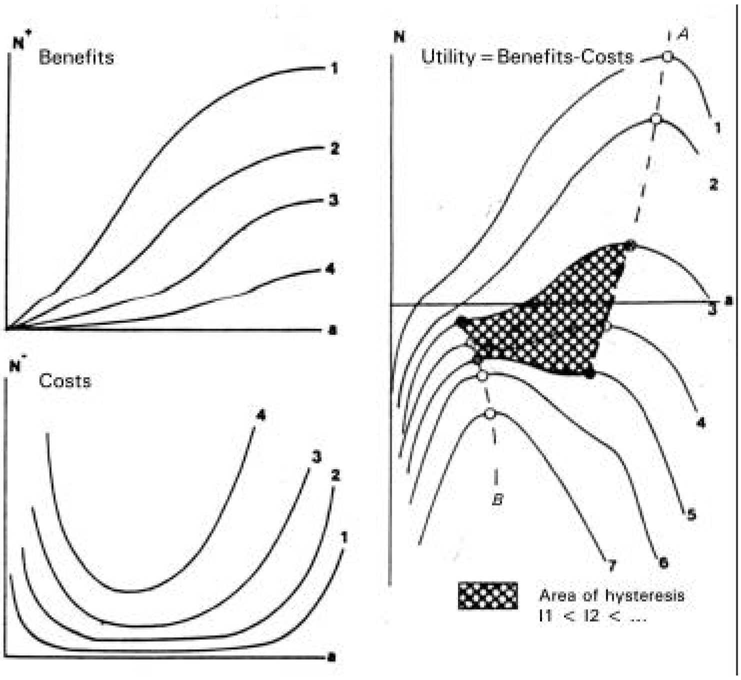

Figure 3 State-of-the-Art review on Congested Traffic Flow-6 Utility of driving according to Van Toorenburg (1983). Benefits, costs, and utility as a function of the behavioural parameters 'a'; the index besides the curves indicate the flow-rate classes.

3.1.2.3 Factors influencing the capacity (drop)

Based on the discussion in the previous sections, we argue that the extent to which the characteristics (duration, magnitude) of the capacity drop is among other things depended on the following:

1. Factors influencing anticipation behaviour:
   - Driver-related factors (experience, awareness, fatigue)
   - Roadway-related factors (geometry, grade, curvature)
   - External conditions (visibility: weather, and ambient conditions)
   - Traffic control (speed homogenising, congestion warning)
2. Factors affecting acceleration / deceleration times:
   - Vehicle and driver characteristics (acceleration behaviour, braking ability)
   - Flow composition (truck fraction)

The extent to which these factors are of influence is however unclear. Additional research is required to answer this question.

3.1.2.4 Relevance in modelling

Since the high prior flow rate state appears unstable, it is not relevant for description of correct congestion propagation, nor for the description of traffic behaviour inside the congested area. This will mainly be determined by the queue discharge rate. Nevertheless, high prior flows are relevant in *ramp-metering control* (peak-shaving), which, once correctly tuned, can increase the throughput of a bottleneck by postponing the on-set of congestion.

Moreover, in predicting the time traffic breakdown occurs, clear insights into the mechanisms causing the capacity drop are of importance. Nevertheless, it appears that traffic breakdown is a stochastic phenomenon, impairing its predictability and consequently the need in precisely describing its dynamics.

3.1.2.5 Conclusions with respect to modelling

- The capacity drop is important for correctly describing traffic breakdown instant.
- The queue discharge rate is relevant for modelling congestion propagation and behaviour inside congestion.
- Precisely describing its causes and the influence of different factors is currently impossible, and mainly a theoretically interesting issue.
- Capacity drop should be modelled phonemically (models should somehow distinguish high prior flows and queue discharge rates).

3.2 Causes for congestion

This section discusses phase-transitions in traffic flow: how and when does free-flowing traffic become congested? Do these transitions appear spontaneously? Are the mechanisms behind these transitions understood (microscopically and macroscopically)?

From the interviews conducted among national and international experts in the field of traffic flow theory and modelling, it appeared that no common ground existed on the mechanisms causing traffic breakdown. This holds equally for the results found in the literature. Although all researcher acknowledge that the main reason for traffic breakdown are bottlenecks, the existence of phantom jams (spontaneous traffic breakdown due to congested traffic flow) has been an issue for debate among the researchers.

Nevertheless, the answer to the question what causes traffic breakdown seems univocal: *traffic demand* temporarily *exceeds the capacity*. While traffic demand is determined by among other things activity patterns, and travel choices, the capacity is determined by the interplay of infrastructure, driving behaviour, and interaction between drivers. Both traffic demand as well as capacity can be described by random variables, yielding a traffic breakdown *probability*, rather than a deterministic parameter indicating that, when, and where traffic breakdown will occur.

3.2.1 Traffic breakdown and bottlenecks

Congestion is mostly caused by (or appears near) bottlenecks (on-ramps, merges, or traffic lights). In case of congestion spillback caused by supply shortage on the underlying network, off-ramps and diverges can also yield an active bottleneck (if one or more lanes on the roadway are blocked). In either case, traffic demand is larger than the bottleneck capacity. As a consequence, traffic breakdown occurs with a very high probability.

3.2.1.1 Instant of traffic breakdown due to bottlenecks

Congestion is frequently postponed *temporarily* and *locally*. In other words, for particular traffic conditions, roadway configurations, etc., *prior congestion flow* is substantially higher than the *queue discharge rate* (e.g. Cassidy and Bertini (1999a), Dijker (1997), Kerner (1999)). As was discussed in section Differences in prior-congestion flow-rate and queue discharge rates, this situation is unstable, and will mostly result in traffic breakdown. What exactly triggers this transition is not yet understood, nor is it clear *if* a specific cause can be appointed (randomness).

3.2.1.2 Theories of traffic breakdown due to bottlenecks

Various theories describe congestion caused by bottlenecks. Well known theories are (first-order) *shockwave theory* and *queuing theory*, which are founded on the premises traffic breakdown occurs when traffic demand exceeds available supply, i.e. the infrastructure demand due to vehicles aiming to use the roadway exceeds the bottleneck supply.

Conversely, Kerner (1999) argues that traffic breakdown due to bottlenecks in *not fundamentally different* from spontaneous phase-transitions in traffic flow, albeit near a bottleneck, traffic breakdown *is more likely to occur*. In other words, downstream of the bottleneck, demand attains a critical value in which the breakdown probability is very high, and traffic breakdown is likely to occur.

3.2.1.3 Factors influencing traffic breakdown

Given that traffic breakdown can best be described stochastically, various factors will influence this breakdown probability. Among these factors are:

- Total traffic demand upstream of bottleneck (e.g. main-road traffic demand, on-ramp traffic demand).
- Bottleneck traffic supply (number of lanes); roadway geometry and bottleneck design (e.g. design speed and length of on-ramp, merging section)
- External factors (road, weather, and ambient conditions)

3.2.1.4 Relevance in modelling

Bottlenecks are the main cause of congestion in traffic network, and should consequently be described correctly by any traffic flow model. It appears that the precise time traffic breakdown occurs is difficult to predict precisely, and that only coarse estimations can be given. Although these predictions can be of importance for traffic control (e.g. congestion warning, ramp metering), once traffic breakdown has occurred, the propagation of the queue is mostly determined by the queue discharge rate, traffic demands at the on-ramp, bottleneck geometry, etc. See section Differences in prior-congestion flow-rate and queue discharge rates.

3.2.1.5 Summary

- Bottlenecks are the main cause for congestion in the Dutch motorway networks.
- Bottleneck related traffic breakdown is not fundamentally different from a self-organised traffic jam, as it can be described in terms of breakdown probabilities (or better: the time-to-breakdown is a random variate), which depends on different factors, such as main-road, and on-ramp traffic (if applicable), etc.

3.2.2 Spontaneous phase-transitions

Spontaneous phase transitions are probably not a common cause of traffic congestion on Dutch motorways. Nevertheless, the notion of (seemingly) spontaneous occurrence of congestion is very interesting from a theoretical viewpoint. Let us first discuss the hypothesised mechanisms causing these spontaneous phase-transitions.

3.2.2.1 Empirical results

Kerner (1999) uses traffic data collected at German motorways to illustrate how traffic jams are spontaneously formed. That is, under critical conditions, phase transitions from free-flow to traffic jams can occur spontaneously. Krauss *et al.* (1999) show that the probability traffic breakdown occurs is among other things dependent on the traffic conditions (flow-rate, velocity variance), and the length of the considered (homogeneous) roadway stretch. Given the short homogeneous sections in the Dutch motorway networks (relative large number of on-ramps, off-ramps, merges, and diverges), it is doubtful whether spontaneous breakdown occurs frequently in Dutch traffic flow).

Daganzo *et al.* (1998) is not convinced that the traffic jam formations are spontaneous; they argue that the traffic jams are caused by bottlenecks only. It is shown that all the phase transitions in and out of freely flowing traffic reported earlier for a German site could be caused by a bottlenecks, as are all the transitions observed at two other sites examined in their paper. The evidence suggests that bottlenecks cause these transitions in a predictable way, and does not suggest that jams appear spontaneously in free flow traffic for no apparent reason. It is also shown that many of the complicated instability phenomena observed at all locations can be explained qualitatively in terms of a simple Markovian theory specific to traffic that does not necessarily include spontaneous transitions into the queued state as a feature. Neither the German nor the North American data support the conclusion that free flowing traffic will spontaneously break down randomly, without obvious reasons, and then remain in that state due to traffic’s tendency to self-maintain congestion. Rather, the evidence indicates that traffic breaks down (queues form) at locations of freeway inhomogeneities (bottlenecks) due to reproducible exogenous reasons, and that, following breakdown, the bottleneck flow behaves in a predictable way.

However, their theories cannot explain several phenomena for empirical studies, such as the fact that the prior-congestion flows can exceed the bottleneck capacity for a significant, but unpredictable period. Similar results have been established on empirical analysis using Dutch A9 motorway data. Moreover, in the meantime, Kerner has illustrated his theories under *different circumstances and different data* (see Kerner (1999b)).

3.2.2.2 Stable, unstable and metastable traffic flow

On a macroscopic scale, transitions from free-flow traffic to synchronised traffic flow, or congested traffic flow can be described by considering the traffic density in (approximately) homogeneous traffic flow. Mostly three regions are considered (see Figure 3 State-of-the-Art review on Congested Traffic Flow-7), namely stable conditions ($\rho < \rho_{c1}$, or $\rho > \rho_{c4}$), linearly unstable conditions ($\rho_{c2} \leq \rho \leq \rho_{c3}$)[3], and finally nonlinearly metastable (elsewhere); see Kerner (1999). When conditions are stable, small (possibly stochastic) perturbations is the traffic density (for instance caused by a braking vehicle, drivers not paying attention, "rubber necking") are "damped", and will not yield a transition of the traffic state. When the density is in the region $\rho_{c2} \leq \rho \leq \rho_{c3}$, any small perturbation in the traffic density will result in traffic breakdown. In the metastable region, perturbations that are large enough will invoke a state transition. In other words, this theory states that under critical traffic conditions, small perturbations can result in spontaneous congestion, stop-and-go waves, etc; see among others Kerner (1999), Krauss *et al*. (1999), Helbing and Huberman (1998).

---

[3] The system is called *linearly unstable* since this region is found when performing a linear stability analysis.

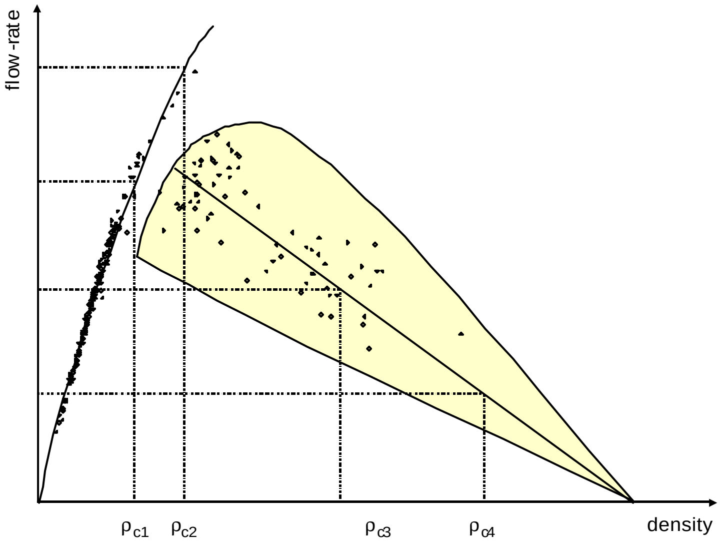

Figure 3 State-of-the-Art review on Congested Traffic Flow-7 Density regions for stable, metastable, and linearly stable traffic conditions. Observations are 2 minute average from consecutive locations on the A9 motorway.

The theory of (spontaneous) phase-transitions is mostly adhered to by German physicists. On the contrary, other researches (University of Berkeley) try to explain the observed chaotic like phenomena using more straightforward theories of traffic flow, where congestion is mostly caused by bottlenecks.

Helbing *et al.* (1998b) discuss different types of congested traffic states (see Table 2 Research problem and focus-1), depending on the combined values of the flow-rate on the main road and the on-ramp respectively. In other words, the main road and on-ramp flow rates to a large extent determine if, and what kind of traffic breakdown occurs. The paper shows good agreement of the model results with Dutch motorway data. The paper also reveals clearly how both the self-formation of spatial clusters and the congestion due to active bottlenecks are linked (caused by similar mechanisms).

3.2.2.3 Theories of spontaneous phase-transitions

Among the causes for spontaneous traffic breakdown is the highly non-linear nature of traffic flow, which may cause of spontaneous formation of congestion due to small, but critical perturbations in the traffic flow: when traffic flows into the disturbance, at a higher flow rate than upon leaving the disturbance, the disturbance will under specific circumstance (depending on drivers anticipation and relaxation behaviour) increase. This will generally only occur when drivers are in relaxation dominant mode (reacting on the vehicle which they are following), that is, when anticipation is of less influence of driver behaviour. Once the density in the disturbance grows, the probability of traffic breakdown increases.

Depending on its density, the disturbance will move downstream (low density), upstream (high density), or be stationary. Thus, initially, the perturbation moves along with the traffic flow. When it is small, and has not yet grown into a jam, it will probably dissolve once reaching a bottleneck. Due to the high bottleneck frequency, spontaneously formed congestion is seldom observed on Dutch motorways.

Another approach in describing spontaneous traffic breakdown phenomena is using chaos theory (for a popular introduction see Gleich (1987) and Crutchfield (1986)). Chaos theory has been applied in various fields of technical engineering, such as meteorology, thermodynamics, electrical engineering, construction engineering, and hydraulic systems. A system is called *chaotic* is it can be described by the following characteristcs:

- The dynamics are *deterministic*.
- The solutions to the dynamics ”cannot be predicted completely”, and are extremely sensitive to the initial conditions.
- The state of the system evolves to a so-called *attractor*, which is a special set of system states.

Generally, the analysis of an attractor provides more insight into the behaviour of the chaotic system. In illustration, Figure 3 State-of-the-Art review on Congested Traffic Flow-8 shows the so-called *bifurcation curve* of an equations expressing a population of rabbits:

$$\theta_{k+1} = c \cdot \theta_k \cdot (1 - \theta_k) \qquad \text{(3 State-of-the-Art review on Congested Traffic Flow.1)}$$

where θ_k denotes the size of the population after year k, which as been scaled to ensure that it is between zero and one.

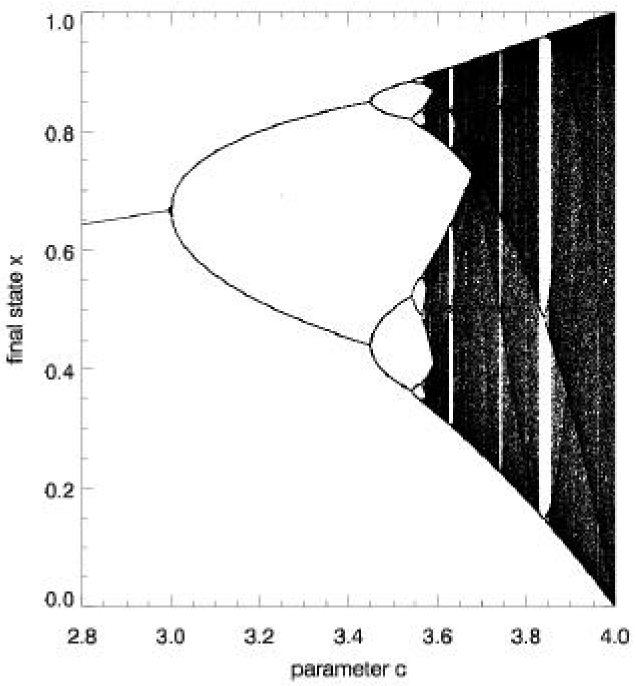

Figure 3 State-of-the-Art review on Congested Traffic Flow-8 Bifurcation curve for rabbit population dynamics for different values of parameter 'c'

Figure 3 State-of-the-Art review on Congested Traffic Flow-8 shows the final states of the rabbit population, after a long period. From *c* > 3.0, the final state is not longer unique. Rather, two final states can be identified between which the system's state will fluctuate. As *c* increases. so does the number of states. For *c* > 3.57, the number of states becomes infinite.

Disbro and Frame (1989) studied whether or not the well-known GM-car following models are chaotic. From their research, it appears that this model indeed has a chaotic character. This implies that the solution is extremely dependent on the initial conditions of the system, and that therefore no reliable outcomes can be expected. Let us however emphasise that only the *final system state* cannot be determined. For shorter term predictions, the state of the system *can be predicted* reasonably accurate.

Pozybill (1998) argues that traffic is an chaotic system. His motivation for this statement is that, on the one hand, traffic flow is a highly non-linear system, while on the other hand, strong *feedback mechanisms* exists between the drivers in the flow. Moreover, for certain density ranges, the system is highly dependent on small perturbations in the initial conditions. This implies that seemingly identical initial situations can self-organise into different, *unpredictable* final traffic states. At this point, let us emphasise that most models analysed in the literature convey *closed system* where external influences, such as on- and off-ramps, are not considered.

The implications of these research results are not yet clear. Does the chaotic characteristics of traffic flow imply that we cannot longer use microscopic and continuum models to describe and predict traffic flow, or is this only the case for long-term predictions? What is the value of these models? What happens when we consider stochastic models rather than deterministic models? Since these questions have not yet been answered, the practical use of chaos theory in describing traffic flows reduces to conceptual frameworks in which these chaotic phenomena can be viewed.

Let us finally notice that the national and international experts have all clearly expressed their serious doubts concerning applications of chaos theory. This is why these issues will not be considered in the remainder of this report.

3.2.2.4 Influencing factors

Whether traffic jams are self-organised depends on among others:

1. Factors influencing anticipation behaviour:
    - Driver-related factors (experience, awareness, fatigue)
    - Roadway-related factors (geometry, grade, curvature)
    - External conditions (visibility: weather, and ambient conditions)
    - Traffic control (speed homogenising, congestion warning)
2. Factors affecting acceleration / deceleration times:
    - Vehicle and driver characteristics (acceleration behaviour, braking ability)
    - Flow composition (truck fraction)
3. Traffic conditions and characteristics disturbance

3.2.2.5 Relevance of modelling

Due to the topology of Dutch motorway networks, spontaneously formed traffic jams are seldom observed. Therefore, its relevance with respect to traffic flow modelling is doubtful. Nevertheless, since the mechanisms of spontaneous phase-transitions are related to traffic hysteresis and the capacity drop, correct description of the former phenomena may imply that traffic flow models *do* inhibit emergent self-organised traffic jams.

Since the probability that traffic breaks down spontaneously is dependent on among other things the reaction time of drivers, and the speed variability, traffic calming control measures may postpone or even prevent the occurrence of self-organised traffic jams.

3.2.2.6 Summary

- Self-organising traffic jams have been observed in traffic flow (German motorways).
- Probability of spontaneous breakdown relates among other things to homogeneous roadway length, and speed variability, which are characteristic of German motorways.
- Relevance for the Dutch case is unclear, mostly due to network topology.
- Regarding traffic control, traffic calming measures (speed homogenising control) may reduce the frequency and severity of self-organising congestion.
- Underlying mechanisms relate to traffic hysteresis and capacity drop.

3.2.3 Traffic hysteresis

Traffic hysteresis is characterised by the phenomenon that drivers show different behaviour when emerging from a disturbance compared to their behaviour when approaching and entering the disturbance (see Zhang (1999) for a theoretical description). Empirical evidence of hysteresis is observed by among others Verweij (1989), Ferrari (1989), and Leutzbach (1991). Hysteresis is the main factor causing capacity drops and congestion self-formation.

For a brief description and theories describing hysteresis, we refer to see section Traffic flow theories explaining capacity drops. Recall from this section that for dilute traffic conditions, drivers are aware of downstream traffic conditions, such that they adapt their speeds before reaching the disturbance. In this case, the anticipation effect is dominant. When traffic is heavy, a driver cannot see that a disturbance is coming until it reaches him. Therefore, the response to the disturbance is retarded (relaxation dominant stage). In relation to this theory, Zhang (1999) argues:

1. There is more than one non-equilibrium curves (transients).
2. Acceleration and deceleration are asymmetric.
3. Phase-trajectories form hysteresis loops.
4. Both acceleration and deceleration curves are nearly smooth.
5. Mixing acceleration and deceleration flow generates discontinuities.

Figure 3 State-of-the-Art review on Congested Traffic Flow-9 shows so-called hysteresis-loops determined from Dutch A9 observations during the morning peak at October 17, 18, 19, and 21. Clearly, unmistakable differences exist between drivers moving into congestion and drivers emerging from congestion, for each of the days considered (and all other days in the dataset as well). It appears that traffic hysteresis is a very common feature of real-life traffic flow.

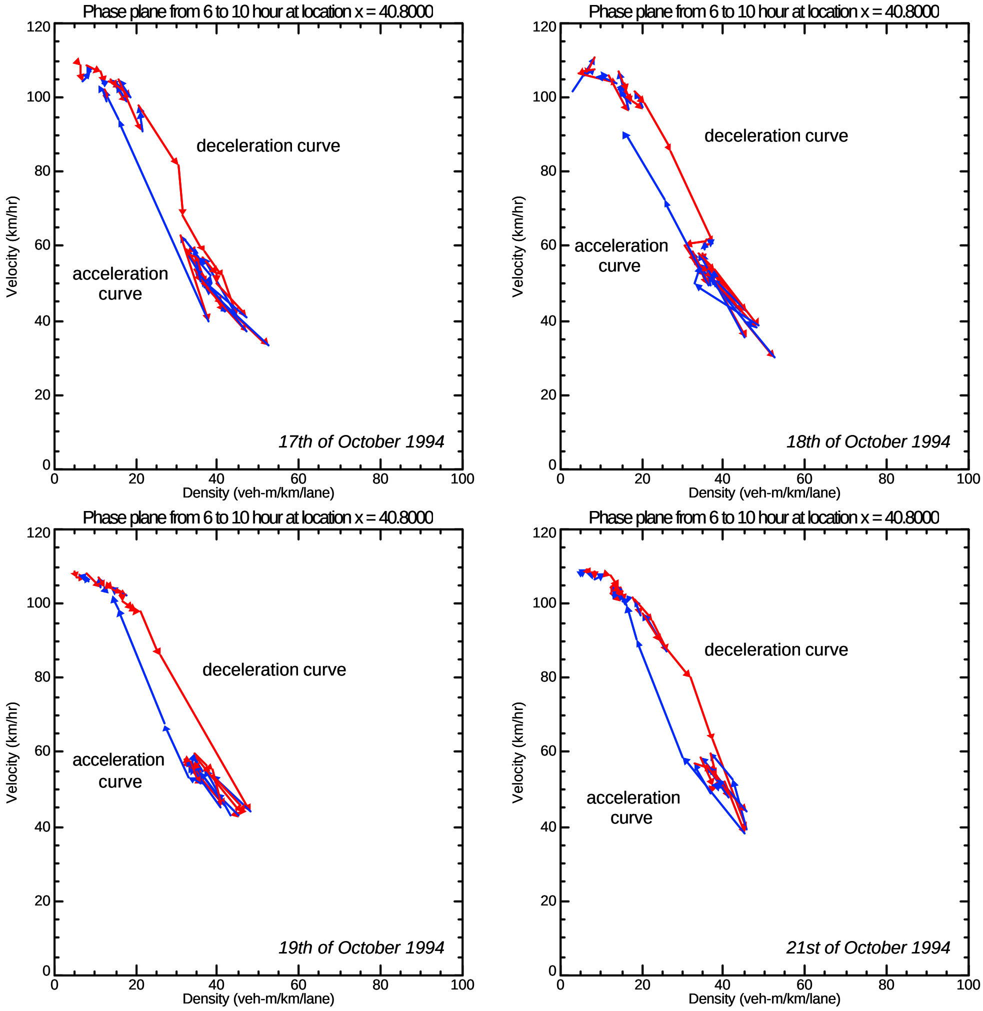

Figure 3 State-of-the-Art review on Congested Traffic Flow-9 Hysteresis loops for different days in October 1994 collected at location x = 40.8km on the A9 motorway.

3.2.3.1 Relevance for modelling

Is correct description of traffic hysteresis an prerequisite for a correct description of congested traffic flow operations, or is it merely a theoretical issue, which is irrelevant in practical model applications?

Traffic hysteresis is fundamental in traffic flow, and is the underlying cause for a number of phenomena that are characteristic properties of traffic flow, which are frequently observed in real-life traffic flow. *Whether* and *how* it can be described sufficiently accurate to be of practical use in traffic flow modelling is however unclear. Moreover, in describing the congestion propagation, mainly the queue discharge rates are of importance. In other words, once congestion has occurred, correct description of hysteresis *may not be important*. Whether traffic hysteresis plays an important role in describing the within queue traffic operations in unclear.

As with self-organising traffic jams, the hysteresis effect can be reduced by speed homogenising control and congestion warning, since these measures improve the awareness of drivers (shift from the relaxation dominant stage towards the anticipation dominant stage; see Zhang (1999)).

3.2.3.2 Summary

- Traffic hysteresis is a fundamental property of traffic flow, underlying various characteristic phenomena in traffic flow (self-organising traffic jams, capacity drop).
- Hysteresis is important in describing and predicting the instant of traffic breakdown. It is of lesser importance for the description of congestion propagation, since this is mainly determined by the queue discharge flows.
- With respect to traffic control, hysteresis is relevant for traffic calming and congestion warning control.
- Qualitative theories of traffic hysteresis exists.

3.3 Congestion propagation

One of the important issues that it addressed in this research project, is to determine how traffic jams propagate, that is, how the locations of the head and the tail of the queue change over time. In this section, we address these issues.

3.3.1 Location of traffic breakdown and head of the queue

Several studies have shown that traffic breakdown is not only delayed in time (section Traffic breakdown and bottlenecks), but breakdown can only *displaced*, that is, located downstream of the physical bottleneck. In this section, we the location traffic breaks down and the consequent location of the queue head (not necessarily the same).

3.3.1.1 Empirical observations

Cassidy and Bertini (1999a,b) describe empirical observations of this displacement (both based on the same data). Their contributions show how the head of the queue is located at 500 *m* to 1000 *m* downstream from the physical bottleneck. From their observations it appears that the head of the queue *always appears at the same location* (within a fairly large region of approximately 500 *m*), and is therefore predictable.

On the contrary, preliminary data investigations have revealed different results: traffic breakdown occurs at random locations, either *at* or *downstream* of the discontinuity. In illustration, Figure 3 State-of-the-Art review on Congested Traffic Flow-10, Figure 3 State-of-the-Art review on Congested Traffic Flow-11, and Figure 3 State-of-the-Art review on Congested Traffic Flow-12 show congestion levels (between 0 and 1, dependent on the average speed) during 7AM-8AM at the A9 motorway. These figures clearly show the differences in traffic breakdown and congestion dynamics of the different days from the same week. The location (and time) of traffic breakdown varies considerably for the respective days. In this particular case, neither the moment and location of traffic breakdown, nor the location of the queue head appears to be the same, the latter being either located directly at the on-ramp location, or 1200*m* downstream of the bottleneck.

This implies either that these traffic flow parameters are highly stochastic (and therefore unpredictable), or that different factors (e.g. main road and on-ramp traffic demand, traffic composition, etc.) play an important, and yet to be determined role in traffic breakdown.

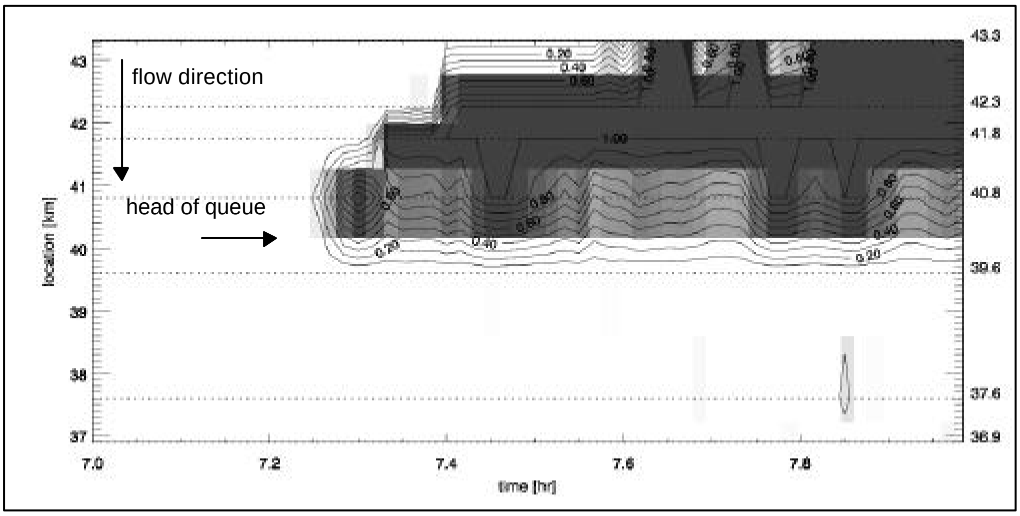

Figure 3 State-of-the-Art review on Congested Traffic Flow-10 Congestion of A9 motorway on 19th of October 1994. The congestion occurs directly at on-ramp location (x = 40.8km). From the on-ramp, congestion moves upstream (that is, from location x = 40.8km to x = 41.8km, etc.)

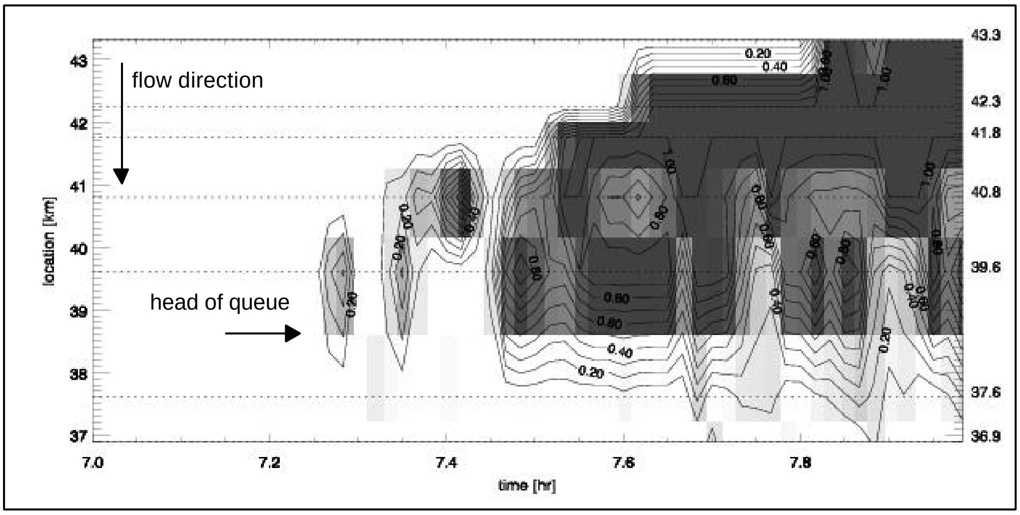

Figure 3 State-of-the-Art review on Congested Traffic Flow-11 Congestion of A9 motorway on the 20th of October 1994. Congestion occurs and stays downstream of bottleneck location (at x = 39.6km), more than one kilometre downstream from on-ramp at x = 40.8km

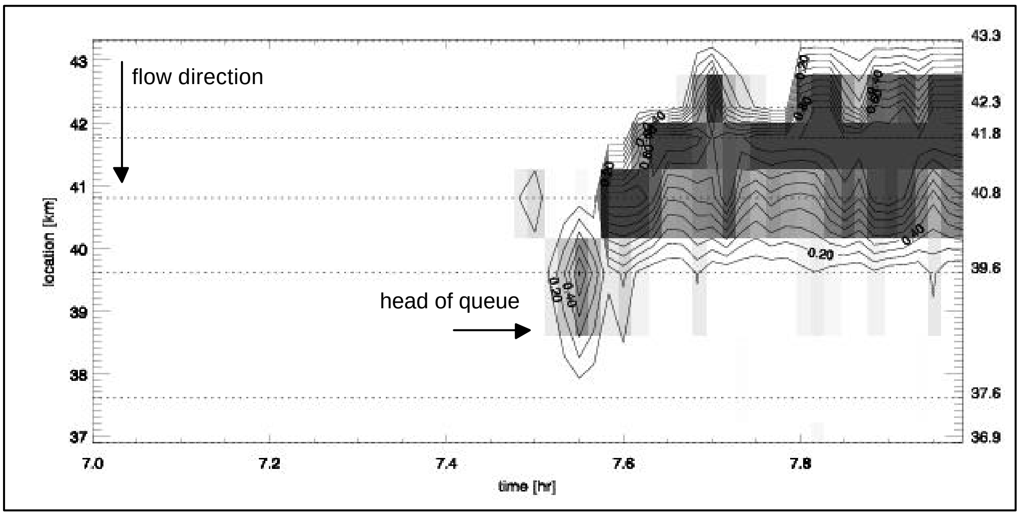

Figure 3 State-of-the-Art review on Congested Traffic Flow-12 Congestion of A9 motorway on the 21th of October 1994. Traffic breaks down downstream of the on-ramp (at x = 39.6), after which it moves upstream until it reaches the bottleneck at x = 40.8km.

3.3.1.2 Traffic flow theories describing displacement

Similar to traffic hysteresis, queue head displacement can be explained by considering the anticipation abilities and relaxation behaviour of drivers. When drivers are able to anticipate on downstream traffic (e.g. drivers merging onto the lane on which they are driving), drivers will anticipate and adapt their speeds, lane, and gaps. However, when drivers are constrained and mostly occupied with the behaviour of their direct predecessor, they are not able to timely anticipate on changes in the flow (for instance due to a bottleneck). The adaptation is therefore postponed in space and time.

Moreover, the location of the queue head appears to be stochastic, implying that the extent to which its exact location can be predicted in bounded. The precise limits to the predictability are however unclear; additional research is needed.

3.3.1.3 Factors influencing queue head displacement

The factors that influence the location of the queue head are similar to the factors that influence the instant of traffic breakdown (see section Factors influencing traffic breakdown). Among other things, this implies that the downstream traffic conditions (determining the stage of the driver) are determinant with respect to the location of the head of the queue. It is expected that other important factors are the traffic demand on the on-ramp (if applicable), traffic composition, and experience of drivers.

3.3.1.4 Relevance with respect to modelling

Depending on the considered network, modelling of the location of the head of the queue can be essential, especially when on-ramps and off-ramps are located very close to each other (for instance the A9 motorway near Rottepolderplein). In these cases, the ability to correctly describe the location of traffic breakdown is essential in correctly modelling congested traffic flow operations. Insights into the location traffic breakdown occurs can be relevant for design of infrastructure.

3.3.1.5 Summary

Studies on America motorways have shown that the head of the queue is displaced (500m-1000m downstream) with respect to the location of the on-ramp. Preliminary empirical studies for the A9 motorway has shown similar displacements, although these do not appear as regular as in case of the American studies (more variability in the location of the queue head).
Among the factors influencing the queue head location are the main-road traffic conditions, and on-ramp traffic conditions (if applicable) downstream of the discontinuity.
Displacement of the queue head may be related to the mechanisms causing traffic hysteresis.

3.3.2 Queue tail location

Prediction of the exact location of the queue tail is important, especially when spillback of congestion plays a critical role. However, judging from the results in the literature, determining this location *is not easily accomplished*. Fortunately, the precise location of the queue tail is not essential to correctly predict important parameters such as travel time.

3.3.2.1 Empirical studies and theories

Smilowitz and Daganzo (1999) show that, albeit some macroscopic features of the traffic flow could be predicted reasonably accurate, higher detail features cannot be predicted easily. Two potential reasons for this "unpredictability" spring to mind. For one, the microscopic processes causing these high-detail features are uncertain. This uncertainty can be divided into *randomness* (stochastic variables) and *vagueness* (fuzzy variables). While the latter is predictable, the former is not. For instance, the gap a driver maintains with respect to his predecessor is a random variable, with a significant variance. Subsequently considering the sum of these random variable (constituting the length of the queue), yields a new random variable whose variance equals the sum of the aforementioned variances.

Another reason is discussed in Helbing and Huberman (1999): the scattering that is observed in the fundamental diagram is caused by the *incorrect description* of differences between user-classes (person-cars and trucks), or even *different leader-follower pairs consisting of different user-classes*. They show that by correctly accounting for the composition of traffic, the scattering can be reduced significantly.

3.3.2.2 *Factors determining location of queue tail*

Several factors determine the location of the tail of the congested area, the most important of which are the upstream traffic demand, and the queue discharge rate combined with the traffic demand (and its dynamics) at the on-ramp (if applicable). The capacity that remains for the vehicles on the main road determines the flow-rates upstream of the bottleneck, and consequently yields both the location of the queue tail as well as the traffic conditions inside the congested area (in terms of flows, densities, and speeds).

In some cases (when the conditions inside the congested area are stable), behaviour inside the congested area is homogeneous ('homogeneous congested traffic', and 'oscillatory congested traffic' (see Table 2 Research problem and focus-1)) and the queue tail can be estimated using first-order theories with reasonable accuracy. However, when traffic conditions inside the congested area are unstable, different patterns may emerge from the head of the queue that are amplified with the upstream region ('triggered stop-and-go waves', 'moving localised clusters'; (see Table 2 Research problem and focus-1)). Due to the highly non-linear behaviour inside these regions, precisely describing how congestion propagates is not feasible.

3.3.2.3 *Relevance with respect to modelling*

Homogeneous congested traffic occurs when traffic demand downstream of the bottleneck (main-road plus on-ramp (if applicable)) is significantly larger than the bottleneck capacity, and is the most common form of congestion. The other congested states are less common, mostly due to the narrow conditions that are required upon their occurrence. Thus, describing the tail of the queue given homogeneous congestion is most relevant and should therefore be prioritised. Other congested states, such as 'oscillatory congested traffic' and stop-and-go traffic (see Table 2 Research problem and focus-1 and Figure 3 State-of-the-Art review on Congested Traffic Flow-13) are also interest, but will probably only amount to a smaller improvement in describing the location of the queue tail.

3.3.2.4 *Summary*

The location of the queue tail can be predicted to a certain extent, depending on the flow conditions inside the congested area.

Homogeneous congested traffic is most common in busy traffic networks.

Homogeneous congested traffic flow is predictable, and adequate theories are available for its description.

3.3.3 Propagation of wide jams

Wide jams are jams in which the velocity is very low, and the density is very high. These jams usually propagate upstream with a fixed velocity, which is characteristic in traffic flow, and can pertain (without an active bottleneck) for a long time (several hours). Note that the notions of traffic hysteresis and the capacity drop are linked to the existence of wide jams: wide jams can only pertain for a significant time duration if the queue inflow is *at least as high* as the queue discharge rate.

3.3.3.1 Empirical studies

Kerner (1999) shows presents the dynamics of wide jams using German highway data. In the Netherlands, wide jams have among others been observed by TRANSPUTE (2000) on the A13 motorway.

3.3.3.2 Theories of wide jams

Kerner (1999) describes wide jams as phenomena in traffic flow that inhibit characteristic parameters which are independent on the initial conditions in the network. He defines a wide jam as a congested region of which the homogeneous region is much larger than the deceleration and acceleration regions. The characteristic parameters of the wide jams are its density, and the outflow from the wide jam. Wide jams can be self-formed in traffic flow, when the traffic density of the flow exceeds some critical density value (metastable traffic flow). The notion of traffic hysteresis is however essential in the formation of wide jams, since otherwise, critical disturbances do not self-organise into a wide jam, and, once formed, will dissipate within a finite time duration.

3.3.3.3 Factors determining wide jams

Wide jams emerge either from bottlenecks, or are self-formed. Since the occurrence of these jams is directly related to traffic hysteresis, factors influencing its occurrence and behaviour can be deduced from section Traffic hysteresis.

3.3.3.4 Relevance with respect to modelling

Wide jams have been observed frequently in empirical studies, which justifies attempting to correctly model these jams in traffic flow. Moreover, prediction of wide jams can be useful in traffic control, especially for improving congestion warning systems.

3.3.4 Propagation of shocks and stop-and-go traffic

One of the reasons for rejecting first-order continuum models is their inability to describe so-called *start-stop waves* (or *stop-and-go traffic*). That is, it has been observed that unstable traffic flow is characterised (under appropriate conditions) by regular start-stop waves with amplitude-dependent oscillation time, which cannot be derived from kinematic wave solutions. Data on these exist for Germany (Leutzbach (1991)), the Netherlands (Verweij (1985)), and Italy (Ferrari (1989)). Verweij (1985) reports how the erratic speed development in the downstream congested area (x=4.35km) is amplified from measurement site to measurement site upstream, and becomes a regular stop-start wave at $x=3.3km$, with a specific amplitude and oscillation time. Further upstream, the traffic flow becomes erratic again ($x=1.1km$).

3.4 Traffic operations inside congestion

In this section, we focus on the properties of traffic flow inside the congested area, and the behaviour of the drivers in this region. It appears that different states of congested traffic flow can be identified, all characterised by specific flow and driver behaviour (Kerner (1999), Helbing *et al*. (1999b)). Moreover, results from empirical studies are discussed as well (e.g. Smilowitz and Daganzo (1999), Cassidy and Bertini (1999b), Dijker (1997), Daganzo *et al*. (1999)).

3.4.1 Macroscopic phases in traffic flow

Traditional traffic flow theories and models mainly consider two different types of traffic flow, namely free-flow and congested flow. Although other flow-types have been described, e.g. based on the volume-to-capacity ratio, these are mostly not fundamentally different. The recent work of Kerner and Helbing has however resulted in the discovery of different phases in traffic flow, inhibiting fundamentally different flow behaviour. Table 2 Research problem and focus-1 depicts these different phases.

3.4.1.1 Theory of phases in traffic flow

The traffic flow theory of Kerner (1999) has probably attained the most attention in the recent years. His theory is based upon three main types of states in traffic flow, namely free-flow, traffic jams, and synchronised flow (characterised by small overtaking probabilities); see Figure 3 State-of-the-Art review on Congested Traffic Flow-7. The latter is subdivided into three types of synchronised flow, namely complex synchronised flow, homogeneous synchronised flow (speed and density are uniform over space and time), and homogeneous-in-speed synchronised flow (speed is uniform of space and time).

Helbing *et al.* (1998b) identifies different, partially overlapping, phases, arguing that the occurrence of the different phases in traffic flow depends on the interplay of main road traffic demands, on-ramp traffic demands, and the geometry. Figure 3 State-of-the-Art review on Congested Traffic Flow-13 shows different emerging phases in for different combinations of main road and on-ramp flows. As can be seen from the figure, in severely overloaded bottlenecks, mostly homogeneously congested traffic occurs, which can be characterised as a equilibrium solution to the high-density branch in of the flow-density diagram. The others are mostly transient states, since they only exist for typical combinations of main road flows and on-ramp flows.

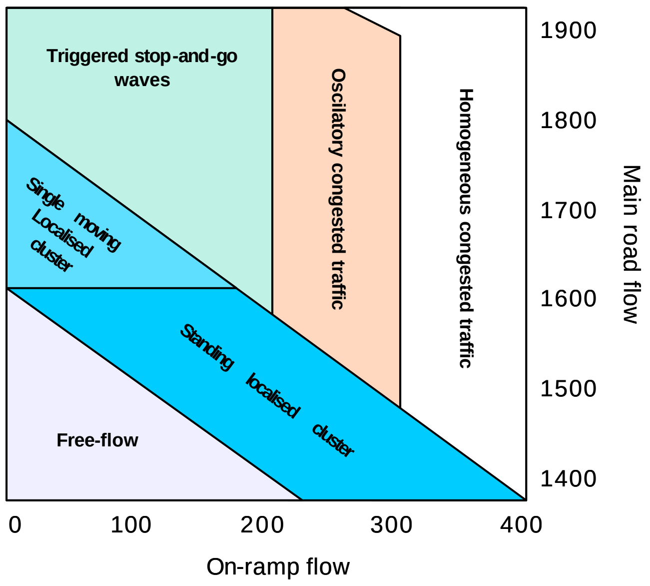

Figure 3 State-of-the-Art review on Congested Traffic Flow-13 Phase diagram of traffic states.

Finally, Helbing and Huberman (1998) shows the existence of co-operative, coherent states arising from competitive interactions, that lead to a new phenomenon in *heterogeneous motorway traffic*. As the density of vehicles increases, their interactions cause a transition into a highly correlated state in which all vehicles move at approximately the same speed. This state is associated with a reduced lane rate, and a safe, high and stable flow. It disappears as the vehicle density exceeds a critical value. The phenomenon is observed in Dutch motorway data. The results are studied using a Cellular Automata-like model for multilane traffic flow consisting of person-cars and trucks, that are characterised by different free speeds, where it is assumed that the lane changing is *symmetrical* (American regulations).

When density increases, the average speed of cars decays rapidly towards the speeds of the trucks, the latter still being close to the their maximum velocity. At this point, the freeway space is almost used up by the safe vehicle headways, implying that sufficiently large gaps for lane changing only occur seldom. The solid-like flow does not change by adding vehicles, until the whole motorway is saturated by the vehicular space requirements at the speed of the lorries. When density increases further, the vehicles speeds decrease significantly to maintain safe headways. Moreover, the onset of stop-and-go traffic at this density produces largely varying gaps, implying that overtaking is again possible and the coherent state is destroyed. Lane-changing cars begin interfere with the trucks, so that the average velocities of trucks start to decrease with growing density *before* the average car velocities.

3.4.1.2 Relevance of phases in modelling

From a theoretical perspective, the identification of different congestion states is interesting. However, the practical relevance in traffic flow modelling is yet unclear, mainly because the frequency at which these states occur (especially in the Dutch case) is unknown. Should specific states be common in real-life traffic flow, microscopic and macroscopic models that inhibit the aforementioned emergent behaviour are to be considered and, if needed, developed. Moreover, data collection and analysis should incorporate these different traffic states. this especially holds for congested traffic flow modelling, since flow and traffic behaviour appears to be very different for the different phases. Nevertheless, no indication of this necessity is yet discovered for the Dutch case.

3.4.1.3 Summary

> Kerner (1999) and Helbing *et al.* (1999b) have identified different phases of (congested) traffic flow, in which flow and driver behaviour are substantially different. Both empirical, experimental, and theoretical traffic flow research should consider these traffic phases.
> It is unclear whether for practical applications of traffic flow, identification of traffic states (besides free-flow and congestion) is essential due to the transient nature of these states.

3.4.2 Car-following behaviour

Dijker (1997) reports clear differences in minimum distance functions during free-flow and congested traffic flow operations, in other words different following distances occur before and during congestion (hysteresis). The variance in choice of minimum driving distance at a specific (congested) speed is high. Depending on the speed level the standard deviation is 50% of the mean distance or higher. A possible explanation is due to Zhang (1999), explicitly considering drivers in different stages (anticipation dominant stage, relaxation dominant stage, and combinations of these). For details, see section Traffic flow theories explaining capacity drops.

Different car-following regimes for free-flow, congested traffic, and possibly transient states (e.g. acceleration, deceleration regimes) are essential for the correct description of congested traffic flow, relating directly to emergent traffic flow behaviour (hysteresis, capacity drop, propagation of wide jams). Based on the results of this study, multi-regime microscopic approaches are therefore recommended.

3.4.3 Lane-changing behaviour

Lane changing behaviour can play a dominantly important role in describing congested traffic flow.

3.4.3.1 Empirical studies

In illustration, in Cassidy and Bertini (1999b) show how bottleneck-related lane changing appears to take place on a large scale, resulting that large numbers of vehicles move to the median, and the middle lane when approaching and passing through the bottleneck. This results in very high flow rates on these lanes (prior congestion), which can last to up to 40 minutes before traffic breakdown occurs. When traffic breaks down, it does so first on the median and middle lane, and a few minutes later on the shoulder lane. The location of traffic breakdown is approximately 500*m* downstream of the discontinuity.

Cassidy and Bertini (1999b) also observe that traffic entering the main road using the on-ramp force themselves onto the main road, at the expense of the through traffic. That is, vehicles enter the motorway from the upstream on-ramp at very high flow-rates, even after the bottleneck's queue capacity propagated beyond the ramp, thereby obstructing the flow. Clearly, traffic on the main-road demonstrates merging giveway behaviour (see Kita and Fukuyama (1999)).

Description of this behaviour is important for different reasons. Firstly, traffic breakdown mechanisms may relate to the aforementioned lane-changing behaviour. Secondly, in case of on-ramps, congestion will mostly occur on the main-road, and to a lesser extent on the on-ramp (unless ramp-metering is active).

3.4.3.2 Relevance in modelling

Prediction of congestion propagation *does not require discerning different lanes*. Congestion propagates simultaneously at all lanes and the lane discharge rates are fairly constant over time and days. This holds both for microscopic and macroscopic traffic flow models.

Contrary to uninterrupted roadway stretches, lane changing at bottlenecks appears to be a relevant phenomenon however, which might play an important role in the precise mechanisms that lead to the onset of queuing. In illustration, congestion appears to occur first on the median lane and middle lane, albeit only a few minutes before congestion appears on the shoulder lane. It can be argued, that congestion is caused by the large number of vehicles changing lanes towards the median and middle lane, in order to facilitate traffic entering the roadway using the shoulder lane. This implies that the active bottleneck is formed by the two roadway lanes, and that the imbalance in traffic demand and supply of these lanes causes congestion to occur there. However, when congestion has set in, queued vehicles will quickly change lanes to the shoulders, resulting in queued conditions on this lane as well (but only after several minutes).

The behavioural findings with respect to pre-bottleneck behaviour and behaviour of traffic entering the main road should be incorporated in microscopic models to correctly describe the observed lane changing behaviour. Especially the latter, since the traffic demand on the on-ramp dictates the queue discharge rate of the queue on the main road.

3.4.4 Differences in vehicle classes

Helbing and Huberman (1999) show the existence of co-operative, coherent states arising from competitive interactions, that lead to a new phenomenon in *heterogeneous motorway traffic*. As the density of vehicles increases, their interactions cause a transition into a highly correlated state in which all vehicles move at approximately the same speed. This state is associated with a reduced lane rate, and a safe, high and stable flow. It disappears as the vehicle density exceeds a critical value. The phenomenon is observed in Dutch motorway data.

Starting from free-flow traffic conditions, as the density increases the average speed of cars decays rapidly towards the speeds of the trucks, the latter still being close to the their maximum velocity. At this point, the freeway space is almost used up by the safe vehicle headways, implying that sufficiently large gaps for lane changing only occur seldom. The solid-like flow does not change by adding vehicles, until the whole motorway is saturated by the vehicular space requirements at the speed of the lorries. When density increases further, the vehicles speeds decrease significantly to maintain safe headways. Moreover, the onset of stop-and-go traffic at this density produces largely varying gaps, implying that overtaking is again possible and the coherent state is destroyed. Lane-changing cars begin interfere with the trucks, so that the average velocities of trucks start to decrease with growing density *before* the average car velocities.

Additionally, the distinction of user-classes is motivated by among other things:
1. Different control regimes may exist for different user-classes. Examples of such class-specific control options are pay-lanes, overtaking prohibition for trucks, class-dedicated lanes (HOV-lanes, truck-lanes, tidal flow-lanes). Class distinction is necessary to correctly describe resulting traffic operations.
2. Only if the model distinguishes different classes, control laws for the (automated) generation of class-specific dynamic control regimes can be determined.
3. We assume that by class-distinction, model potential for synthesis, analysis, and control of roadway traffic improves significantly. In other words, we have the strong belief that model performance improves by distinguishing user-classes and their distinct (driver-) characteristics (macroscopic models).

3.5 Summary

This chapter discussed the results of the literature survey, concerning the current state of empirical and theoretical knowledge on congested traffic flow modelling. Summarising, the literature discusses observations of ‘non-linear flow behaviour’ that mostly cannot be described by traditional flow theories.

Most of these can be contributed to (some sort of) traffic hysteresis, in that prior-congestion behaviour is somehow different than in-congestion behaviour (observed differences in prior congestion flows rates and queue discharge rates, spontaneous traffic breakdown, stability of wide jams). Traffic hysteresis is therefore a theoretically important issue, since it can explain various types of congestion flow phenomena. Different theories were recalled, among which are the distinction of anticipation and relaxation dominant phases in traffic flow, and applications of catastrophe theory.

Another issue is that different types of congestion can occur, depending on among other things the traffic density, and the relation between main road and on-ramp traffic volumes. It is argued that the type of congestion that occurs to a large extent determines the traffic flow behaviour inside the congested area, and consequently the location of the queue tail. For instance, in case of homogeneous congested traffic, behaviour inside the congested area, and the location of the queue tail, can be described by first-order theories.

On a microscopic scale, it appears that driving behaviour is affected considerably by the congested traffic conditions. This holds for both car-following behaviour, as well as for lane-changing behaviour. With respect to the latter, empirical research has revealed that drivers anticipate on downstream on-ramps by changing to the middle and median lanes well before arriving at the bottleneck. Also, congestion does not appear at the same time for all roadway lanes. Finally, during congested flow, traffic on the main road appears to give-way to traffic from the on-ramp, causing queues to appear on the main road mostly.

Several research question remain however unresolved, most important of which are the following:

1. What are the main factors in traffic breakdown, determining among others its exact time and location? When are traffic conditions critical? How random are breakdown times and locations?
2. How important is traffic hysteresis (from a practical viewpoint)?
3. What factors influence the difference is prior- and post-breakdown traffic flow rates, and are these differences significant? Can the high prior-traffic flows be exploited in practise (e.g. in control applications)?
4. How does driving behaviour depend on the flow regime (car-following behaviour, lane-changing behaviour)?
5. How does driving behaviour change when discontinuities are nearby? Can general relations be established, or is this behaviour highly site-specific?

4 CONGESTED TRAFFIC FLOW MODELLING

Research on the subject of traffic flow modelling started some forty years ago, when Lighthill and Whitham (1955) presented a macroscopic model, based on the analogy of traffic flow with the dynamics of a fluid. Since then, mathematical description of traffic 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. In addition to the controversy between microscopic, and macroscopic modelling streams, several researchers have joined the debate on the macroscopic modelling approach most suitable for a correct description of traffic flow. Moreover, recent theoretical and empirical findings of Boris Kerner and co-workers resulted in increased public attention for the subject of macroscopic flow modelling (see Kerner *et al.* (1996), Kerner (1999)). Also, due to improved techniques and increased computational capacity to solve large-scale control problems, applications of realistic flow models in model-based control approaches have become feasible.

This chapter describes if and how traffic flow models do or can describe traffic flow and driver behaviour discussed in the previous chapter, by outlining the emerging problems observed when applying the traffic flow models mentioned earlier. Table 4 Congested traffic flow modelling-5 summarises the results from this stage of the research. It describes the observed driver or flow behaviour during congestion, its relevance in traffic flow modelling, and finally, modelling approaches to describe this behaviour using either microscopic models or continuum models. In the remainder of this chapter, these issues are discussed in more detail.

Considered congestion traffic flow behaviour	**Relevance**	**Potential solution approach**	
		Micro models	**Continuum models**
1. Difference in high prior-congestion flow-rate and queue discharge flow (capacity drop).	DTM measures postponing traffic breakdown; relevance for Dutch networks stems from a.o. peak-shaving due to ramp-metering.	Slow-to-start traffic rules *Multi-regime models* differentiating between prior- and during congestion car-following behaviour: *anticipation dominant, relaxation dominant, and balance.*	Higher-order models with finite relaxation times (non-instantaneous relaxation to equilibrium state). Traffic condition dependent relaxation time and anticipation behaviour, e.g. using dynamic equations for velocity variance.

Considered congestion traffic flow behaviour		Relevance	Potential solution approach	
			Micro models	**Continuum models**
2.	Occurrence of spontaneous transitions (e.g. free-flow to synchronised flow)	DTM measures increasing stability (e.g. traffic calming and speed homogenising); relevance for Dutch motorways is unclear.	Slow-to-start traffic rules. Incorporation of randomness Multi-regime models.	Higher-order models with finite reaction times. Non-local interaction terms. Description of breakdown in terms of probabilities.
3.	Phase-transitions in presence of discontinuities	Describing postponement of traffic breakdown. DTM measures postponing traffic breakdown (metering)	See Occurrence of spontaneous transitions (e.g. free-flow to synchronised flow).	See Occurrence of spontaneous transitions (e.g. free-flow to synchronised flow).
4.	Displacement of queue head with respect to physical bottleneck (500-1000*m* downstream)	Spillback of congestion; blocking back effects in traffic networks. Correct testing geometric design.	Delayed or smooth adaptation (increase) of time headways to safe time headways.	Higher-order models with finite relaxation times. *First-order models are unsuitable*.
5.	Location of the tail of the queue.	Spillback of congestion; blocking back effects.	Improved distinction of classes and respective behaviour. Dependent on randomness in car-following behaviour	Higher-order models with ability to correctly describe stop-start waves. Description of tail queue as random variable. First-order models unsuitable.
6.	Correct description of propagation of wide jams.	Congestion spillback; traffic safety (congestion warning).	Slow-to-start traffic rules. Multi-regime models.	First-order models unsuitable Higher-order models recommended.
7.	Differences in car-following behaviour free-flow / congested flow.	Corrective description of traffic flow in congested region.	Multi-regime car-following models, *GM models unsuitable.*	NA.
8.	Observed differences in lane-changing behaviour.	Corrective description of traffic flow in congested region. Improved ability to predict instant of traffic breakdown.	Multi-regime car-following and lane-changing models. Incorporation of different (*giveway*) driving rules.	Incorporation of behavioural regions. Precedence of traffic entering roadway.

Considered congestion traffic flow behaviour		Relevance	Potential solution approach	
			Micro models	Continuum models
9.	Observed differences between user-classes.	Corrective description of traffic flow in congested region. Scattered observations during synchronised flow operations.	Detailed description of class-specific behaviour (acceleration, braking, reaction times). Distinction of vehicle type pairs.	Generalisation of continuum models to multiclass flow. Macro models based on microscopic principles.
10.	Increasing complexity microscopic models	Model calibration. Conflicting behavioural rules.	Fuzzy logic with isolated calibration of rules (if possible)	NA.
11.	Model calibration.	Unrealistic (*non-interpretable*) model parameters; underspecification of optimisation problem (too many unknown variables, too few equations).	Step-by-step calibration; new data sources (instrumented vehicles) yielding microscopic data. Avoid using macroscopic data for calibration purposes. Use of turing tests for model validation.	Calibration of continuum models usually not an issue;

Table 4 Congested traffic flow modelling-5 Overview of results from literature survey concerning modelling approaches to describe observed behaviour of congested traffic flow operations.

4.1 Capacity related issues

Let us first discuss how capacity play a role is microscopic and macroscopic traffic flow modelling. Then, we will discuss how issues such as the capacity drop can be incorporated into microscopic and macroscopic traffic flow models.

4.1.1 Describing capacity

In several microscopic simulation models, such as Flexsyt2, Aimsun2, Fosim, and Mixic, the capacity results from car-following behaviour and lane-changing behaviour near the bottleneck, and is therefore *endogenous*. Conversely, the microscopic model Integration uses speed-density curves for the microscopic behaviour, implying that the capacity is *exogenous*. This holds equally for most macroscopic models (e.g. Metanet, Corsim, FlowSim, and STM). However, some recently developed macroscopic models (see Hoogendoorn (1999), and Helbing *et al.* (1998)) are founded on microscopic principles, in which case the capacity follows from driver's behaviour.

In most macroscopic models, capacity is deterministic, although possibly dependent on prevailing traffic conditions (vehicle composition, traffic regulation and control).

Stochastic microscopic models result in random capacity values, implying that congestion is a random variate as well.

4.1.2 Differences in prior-congestion flow-rate and queue discharge rates

Section Differences in prior-congestion flow-rate and queue discharge rates discussed differences between prior-congestion traffic flow, and queue discharge flow was discussed. Several traffic flow models proposed in the literature contain mechanisms describing these differences.

4.1.2.1 Microscopic models

With respect to *microscopic simulation models*, the drop in the maximum flow rate can be described by distinguishing the *minimal headway* with which drivers enter a traffic jam ("reaction time"), and the average headway with which drivers leave the jam ("jam escape time"). On way to achieve this, is to differentiate the *acceleration of cars leaving a jam*, and the *deceleration upon entering a jam*. This is discussed in Krauss *et al.* (1999), and Dijker *et al.* (1997). To achieve this, we can consider a *multi-regime* model of driver behaviour, where it is hypothesised that drivers react differently to other vehicles depending on the prevailing traffic situation. This enables the inclusion of more complex behavioural mechanisms.

4.1.2.2 Macroscopic models

Daganzo *et al.* (1999) shows how *first-order models* can be modified by considering different equilibrium flow relations; see also Newell (1965), who differentiates between acceleration and deceleration curves. However, for practical applications, determination of the aforementioned acceleration and deceleration curves in difficult (see Zhang (1999)). Moreover, the precise moment of traffic breakdown needs to be indicated by the user of the model; the model does not predict that the high prior-congestion flow rates will exists for, say 13 minutes, after which traffic breakdown occurs (the instant congestion occurs must be provided by the user of the model).

Higher-order continuum models are able to (qualitatively) describe the capacity drop, by admitting dynamic deviations from the equilibrium solution. Results of the calibrated model MASTER (Helbing *et al.* (1999b)) shows generally good agreement with Dutch motorway data, as does the model of Hoogendoorn (1999), and Hoogendoorn and Bovy (2000). Since the capacity drop results from the interplay of *driver anticipation* and *relaxation*, a prerequisite is that the latter processes are correctly described. This implies among other things that the so-called *anticipation coefficient* c_0, that is assumed constant in most models (Metanet; the models of Payne (1971), Kerner *et al.* (1996)), must be dependent on the regime (e.g. c_0 is a monotonically decreasing function of the density, $c_0 = c_0(r)$). This holds equally for the relaxation time τ (e.g. τ is a monotonically increasing function of the traffic density $\tau = \tau(r)$). For details, we refer to Zhang (1998), Helbing (1997), and Hoogendoorn (1999). The latter describes dependence of traffic conditions by introducing dynamic equations describe the dynamics of the

4.2 Causes for congestion

Recall that on a microscopic scale, traffic breakdown occurs when the net time headway of drivers is smaller than is preferred (for a comfort and safety perspective) given prevailing traffic conditions (density, speeds). This can occur under different conditions, but the most common is near an on-ramp. In this case, vehicles from the on-ramp enter the mean road, causing the average headways of drivers to reduce both near the on-ramp. Although (some) drivers may temporarily accept these smaller headways, yielding a very high flow-rate, after some time drivers will reduce their velocity thereby increasing their headway, which subsequently yield a reduction in the flow-rate.

Section Causes for congestion discussed that congestion is mostly caused by bottlenecks. However, from empirical studies it results that congestion is frequently *postponed*, that is, although traffic demand exceeds the capacity, traffic does not break-down (immediately). Moreover, in some cases, traffic breaks down *spontaneously*, i.e. without the presence of an apparent bottleneck. This section discusses approaches to include both phenomena into microscopic and macroscopic models.

4.2.1 Probabilistic description of traffic breakdown

In section Causes for congestion it was argued that traffic breakdown can best be described in terms of probabilities. For stochastic microscopic simulation models, this is already the case. Macroscopic simulation models are deterministic, implying that either congestion occurs or not. However, due to the strong non-linear (even chaos-like) behaviour of the higher-order models, slight perturbations in the initial or boundary conditions may yield insights into the breakdown probabilities.

4.2.2 Traffic breakdown and bottlenecks

How can the difference in high prior-congestion flow and queue discharge rate be described? In this section, we aim to provide some answers to this question for both microscopic and macroscopic models.

4.2.2.1 Microscopic models

Krauss *et al*. (1999) show that in order for microscopic simulation models to exhibit the correct description of the capacity drop (and thus of wide jams), these models need to inhibit mechanisms that yield higher queue inflows than queue discharge rates. This holds for any microscopic model for which the reaction time (or mean headways during free-flow conditions) is smaller than the "jam escape time". Simple (General-Motors) car following models do not have this feature, and are therefore *not suited for correct description of congested traffic flow*.

A possible approach is to incorporate multi-regime flow models (see Microscopic models), where driving behaviour is dependent on the traffic conditions or the driver (e.g. anticipation dominant stage or relaxation dominant stage, see section Differences in prior-congestion flow-rate and queue discharge rates).

4.2.2.2 Macroscopic models

Helbing *et al.* (1998b) shows how the non-local macroscopic model yields the occurrence of different traffic states (types of congestion; see Table 2 Research problem and focus-1), depending on the combined values of the flow-rate on the main road and the on-ramp respectively. The paper also shows that the model is generally in good agreement with Dutch motorway data. The paper also reveals clearly how both the self-formation of spatial clusters and the congestion due to active bottlenecks are correlated, that is, caused by similar mechanisms. However, due to the bottleneck, traffic breakdown under metastable or linearly unstable conditions will occur frequently (that is, with a very high probability).

4.2.3 Spontaneous phase-transitions, without any apparent bottleneck

Section Spontaneous phase-transitions discussed empirical observations of congestion self-formation, mainly on German motorways. This section also discussed theories for congestion self-formations. Although self-organising traffic jams are probably not common on Dutch motorway networks, in this section we will discuss approaches to incorporate spontaneous phase-transitions in microscopic and macroscopic traffic flow models.

Most of the current traffic flow models do not describe observed (spontaneous) phase transition from free-flow to synchronised or congested flow. For specific model applications (e.g. traffic calming measures), clear insights into the underlying mechanisms causing these transitions are required, possibly even if self-organisation does not occur frequently in the considered roadways.

4.2.3.1 Microscopic models

For microscopic models, spontaneous traffic breakdown is linked to among other things differences between driving behaviour upon entering and leaving congestion, the finite reaction time of drivers, and the stochastic nature of driving behaviour. Krauss *et al.* (1999) shows how spontaneous occurrence of traffic breakdown can be described phenomenologically correct by and adapted (*stochastic*) Cellular Automata model. The model is modified by inclusion of the so-called *slow-to-start* driver behaviour rule.

4.2.3.2 Macroscopic models

A specific class of higher-order macroscopic traffic flow models can qualitative describe transition from (perturbed) free-flow conditions to congested traffic flow (stable or wide jams). A good description of these spontaneous transitions using macroscopic models is found in Kerner (1996). However, only the models including non-local interactions (such as MASTER (Helbing *et al.* (1998)) and the model of Hoogendoorn and Bovy (2000)) can describe spontaneous transitions to *synchronised flow*. Since these models are deterministic, jams are not formed spontaneously at all: they are a result of an avalanche-like process that cause small perturbations in the flow to grow into localised clusters. The first-order models cannot describe spontaneous formation of traffic jams, nor the mechanisms behind them.

4.2.4 Traffic hysteresis

We have already discusses issues concerning traffic hysteresis (see section Traffic hysteresis). Zhang (1999) discusses a traffic flow theory for the description of traffic hysteresis (see section Traffic flow theories explaining capacity drops and hysteresis), which can be easily used to make operational models. This section describes possible approaches to include hysteresis into macroscopic and microscopic modelling.

4.2.4.1 Microscopic models

For microscopic models, incorporating traffic hysteresis can be included in a similar fashion as the capacity drop, for instance using a multi-regime car-following approach, where the headway of a driver is dependent on the regime (and history) of the driver (prior-congestion, in congestion, post-congestion); see sections Microscopic models and Microscopic models.

4.2.4.2 Macroscopic models

Generally, first-order models can not describe traffic hysteresis. To include traffic hysteresis, different speed-density relations, and mechanisms to describe transients from one speed-density curve to another are to be included in the first-order model. Examples of these approaches are described in Newell (1965) and Daganzo *et al.* (1999). To this end, Newell exploits the fact that acceleration flows and deceleration flow follow distinctively different paths in the speed-density plane.

Higher-order models include mechanisms that describe the traffic hysteresis, due to the inclusion of traffic relaxation and anticipation. Improving this behaviour can be achieved by considering traffic conditions relaxation times and anticipation coefficients (see Helbing (1997), Zhang (1998), and Hoogendoorn (1999)).

4.3 Describing queue location and congestion propagation

This section discusses if and how microscopic and macroscopic simulation models can describe the location of the head and the tail of the queues.

4.3.1 Queue head location in case of discontinuities

Section Location of traffic breakdown and head of the queue discussed empirical findings and theories concerning the issue of queue head displacement in case of on-ramps. To incorporate this behaviour into the modelling, the corresponding behavioural mechanisms must be included in the modelling as well (i.e. drivers tendency to temporarily accept small headways). Let us briefly discuss the research findings concerning this modelling issue.

4.3.1.1 Microscopic models

None of the interviewed experts or literature specifically address displacement of the queue head with respect to the discontinuity. However, based on the speed-density curves established using the microscopic models, it can be concluded that the location of the head of the queue is not correctly described. For instance, from Dijker (1997) we can conclude that the real-life observations and the simulated measurements are located at different points in the phase-plane, from which the location of the active bottleneck (the queue head) can be determined: this implies that either the queue is located elsewhere or traffic demand is not comparable to measured traffic demands.

To correctly describe the location of the bottleneck, the corresponding behavioural mechanisms must be included in the *microscopic modelling*, i.e. drivers tendency to temporarily accept small headways / not immediately braking when the headway becomes to small for safe and comfortable driving. This can be achieved by including mechanisms that cause the time headway to increase *smoothly* (or delayed) after being too small due to merging vehicles. When the time headway smoothly increases, traffic flow rate decreases equally smoothly until the capacity of the roadway (queue discharge rate) is reached, that is, at the head of the queue.

4.3.1.2 Macroscopic models

Hoogendoorn (1999) shows that the higher-order platoon-based multilane multiclass model has spatially displaced traffic breakdown, with approximately 500 *m*. This spatial displacement of the queue head is linked to the fact that in higher-order macroscopic models, the velocity does not decrease instantaneously when traffic density changes, but rather takes some time (namely, the relaxation time). From the viewpoint of an observer moving along with the traffic flow, the fact that relaxation time reflects the time needed to adapt the velocity towards the equilibrium velocity, yields that the point at which capacity is reached is displaced downstream. The precise location congestion occurs depends among other things on the finite relaxation time. Since *first-order models* can be deduced from higher-order models by setting the relaxation time to zero, no displacement of the queue head is and can be predicted by first-order macroscopic models.

4.3.2 Queue tail location

Prediction of the exact location of the queue tail can be important, especially when spillback of congestion can play a critical role. Most traffic flow models are however not able to correctly describe the location of the tail of the queue. On the one hand, this is inherent to the randomness of the process we aim to describe. In this respect, a possible approach is to represent the queue end *as a random variable*. To improve the applicability of the models, regular updating of the system state with measurements is essential.

4.3.2.1 Microscopic models

Another improvement is the distinction of user-classes: microscopic models can be improved by differentiating according to the leader-type of the following vehicle, which can be achieved in a multi-regime approach.

4.3.2.2 Macroscopic models

Given that the bottleneck conditions are such that *homogeneous congested traffic flow* results (see Figure 3 State-of-the-Art review on Congested Traffic Flow-13), first-order models can accurately describe the location of the tail of the queue. Compared to regular higher-order models (such as METANET), this description is much better. In the latter case, the finite time vehicles need to adapt their velocity to the prevailing traffic conditions may result in queues in which traffic density exceeds the jam density (approx. 175 *veh*/*km*), while the length of the queue is too short; see a.o. Helbing (1996). In Hoogendoorn (1999) this problem is resolved by including the finite space requirements of vehicles. Another approach is to include non-local interaction (i.e. drivers react on *downstream*, rather than on *local* traffic conditions). The inclusion of *non-locality* in the interaction equation (i.e. the relaxation term), can be easily achieved[4] (see Helbing *et al.* (1998a)), and Hoogendoorn and Bovy (2000c).

Another approach to improve description of the location of the queue tail is to distinguish different user-classes, such as person-cars and trucks; see Hoogendoorn and Bovy (2000b).

4.3.3 Propagation of wide jams

Wide jams are jams in which the velocity is very low, and the density is very high. These jams usually propagate upstream with a fixed velocity, which is characteristic in traffic flow, and can pertain (without an active bottleneck) for a long time (several hours). Note that the notions of traffic hysteresis and the capacity drop are linked to the existence of wide jams: wide jams can only pertain for a significant time duration if the queue inflow is *at least as high* as the queue discharge rate. Kerner (1999) shows presents the

---

[4] This non-locality also improves the stability of numerical solution approaches.

dynamics of wide jams using German highway data. In the Netherlands, wide jams have among others been observed by TRANSPUTE (2000).

4.3.3.1 Microscopic models

Krauss *et al.* (1999) considers the existence of *stable* (or wide) *jams*. Although the mechanisms describing the creation of traffic jams is different for the respective models, the mechanisms causing the stability of the jams is universal, and in their opinion very simple. The stability is caused by driver behaviour that allows the queue discharge rate (the inverse of the so-called "jam escape time") to be lower than the queue inflow (the inverse of the reaction time). On way to achieve this, is to reduce the acceleration of cars leaving a jam, compared to the deceleration upon entering a jam, for instance using a multi-regime approach. To this end, a variant of the Gipps' car-following model is formulated, for which a stability analysis has been carried out and which is flexible with respect to modelling different types of behaviour (depending on the parameter choices), a.o. the hysteresis (lower queue discharge rate). In comparising, Krauss *et al.* (1999) show how traditional General-motors-like car following models do not provide a realistic description of traffic flow. That is, these traditional car-following models inhibit structural shortcomings which cannot be overcome by adaptation of the models.

4.3.3.2 Macroscopic models

First-order continuum models are not able to correctly describe the propagation of wide jams. Traffic jams are only stable (i.e. they pertain for a considerable amount of time) in case the conditions surrounding the wide jam are precisely at capacity. Otherwise, the wide jams will (eventually) dissolve. Inclusion of different speed-density relations (e.g. distinguishing between acceleration and deceleration curves; see Newell (1965)) can remedy these problems.

Conversely, higher-order models can generally predict the dynamics of wide jams reasonably accurate (see Kerner (1999)), due to the inclusion of mechanisms describing hysteresis.

4.3.4 Propagation of shocks; stop-and-go traffic

In this section, we discuss modelling approaches to include the propagation of shocks in traffic flow models. For microscopic models, none of the considered manuscripts, nor any of the interviewed experts have considered this issue. Therefore, we restrict the discussion below to macroscopic models.

4.3.4.1 Macroscopic models

One of the reasons for rejecting first-order continuum models is their inability to describe so-called *start-stop waves* (or *stop-and-go traffic*).

Although the LWR model makes sensible predictions of propagation and dissipation of traffic jams, it fails to model these stop-start waves caused by the instability in traffic flow (see Edie and Bavarez (1967)). When traffic is *not in equilibrium*, acceleration flows and deceleration flows follow distinctively different paths in the (ρ, v) phase plane, and these paths usually form one or more hysteresis loops. In closure, consider the possibility that the unsuitability of the LWR theory to describe the dynamics inside the queue *can be explained* (at least qualitatively) by the inability of the LWR model to describe start-stop waves in unstable traffic flow (see Newell (1965)).

4.4 Traffic operations inside the congested region

A number of the publications considered in this survey discuss features of the traffic flow at the bottleneck and inside the congested region (see section Traffic operations inside congestion).

4.4.1 Car-following behaviour in microscopic models

Zhang *et al.* (1999) discuss how most microscopic simulation models only consider *single regime car-following* (only to a certain extent by pshycho-spacing models). The main limitation of such these single-regime models is that they do not accurately describe the diverse driver control actions and responses to different traffic situations (e.g. free-flow, congestion).

Dijker *et al* (1997) incorporate the empirical findings by redefining the following functions for free-flow and congested regimes. Zhang *et al*. (1999) propose a more comprehensive *multi-regime approach*, that determines the behaviour of a driver, based on the speed, position, and acceleration of the subject compared to its leader, for various traffic regimes. Examples of different regimes are normal car-following (GM-model), uncomfortable regime (speed differences are zero, but gap is too small), free-flow regime (acceleration towards free speed), emergency regime, intersection arrival regime, and queue discharge regime.

4.4.2 Lane-changing behaviour in microscopic models

Lane changing behaviour can play a dominantly important role in describing congested traffic flow. In illustration, in Cassidy and Bertini (1999) it is shown how the lane use changes in the presence of a bottleneck (section Lane-changing behaviour). That is, bottleneck related lane changing appears to take place on a large scale, resulting that large numbers of vehicles move to the median, and the middle lane when approaching and passing through the bottleneck. This results in very high flow rates on these lanes (pre-congestion). When traffic breakdown occurs, it occurs first on the median and middle lane, and later on the shoulder lane after a few minutes. Secondly, traffic entering the main road using the on-ramp force themselves onto the main road, at the expense of the through traffic.

Although the *prediction of congestion propagation* does not require the correct description of lane changing behaviour nearby bottlenecks, this *may be essential* to correctly describe the mechanisms that cause traffic breakdown. Regarding microscopic models, this implies that the lane changing behaviour must correctly described, for instance by adhering to a multi-regime approach that prescribes different driving behaviour in the vicinity of an on-ramp or a merge. The few macroscopic models distinguishing roadway lanes can be adapted in a similar manner.

With respect to the precedence of traffic entering the roadway using the on-ramp, such behaviour can be easily included in both microscopic as well macroscopic simulation models, that is, if this behaviour is observed in the Dutch case as well. Note that giveway rules, such as described by Kita and Fukuyama (1999), may be needed to correctly describe the way traffic from the on-ramp merges onto the main-road.

4.4.3 Macroscopic models

With respect to describing traffic flow behaviour inside the congested area, first-order models are mostly inadequate for correct description (see a.o. Smilowitz and Daganzo (1999)). Higher-order models may improve upon the performance of first-order models, given their ability to describe stop-and-go waves, although the extent in which this holds has to be determined.

4.4.4 Differences in vehicle classes (person-cars and trucks)

In Helbing and Hubermann (1999), the existence of co-operative, coherent states arising from competitive interactions, that lead to a new phenomenon in *heterogeneous motorway traffic* is discussed (see section Differences in vehicle classes). They show how the distinction of different user-classes is essential for correct description of traffic flow.

Moreover, in a recent study, Hoogendoorn and Bovy (2000d) have shown differences in car-following behaviour between vehicle pairs (e.g. person-cars following person-cars, person-cars following trucks). It appears that this following behaviour depends on the type of vehicle that is followed. This can be included easily in microscopic models by considering lead-type dependent following relations and parameters.

4.5 General modelling issues

Hoogendoorn and Alkim (1999) argue that modelling approaches should be chosen to best suit the application that they are intended for. When macroscopic quantities, such as queue-lengths, are of interest, a macroscopic approach should be considered. While microscopic quantities are analysed, microscopic approaches are required. This implies that for network analysis where the formation of queues determines the travel times, macroscopic approaches are favourable. Analysis of geometric designs requires microscopic simulation.

The statement above is justified by the following fundamental issue, which is relevant when considering a specific modelling approach. For instance, the high non-linearity and the inherently chaotic behaviour of traffic flow implies that small disturbances or errors in parameters may yield severe impacts on a macroscopic scale. Therefore, small errors in the microscopic modelling of traffic may have a large impact on the predicted lengths of queues, the exact location of the occurrence of congestion, and so on. This also complicates microscopic model calibration.

4.5.1 Microscopic models

In this section, issues of increased *model complexity* and *model calibration* are discussed, especially considering microscopic simulation models.

4.5.1.1 Model complexity

In this respect, it is clear that driving behaviour is much more complex and possibly more random and complex than currently modelled in most microscopic simulation models (FOSIM). Lane and driver type differences occur as well as effects of leader-follower combinations (not shown in this study, but proven by Hoogendoorn in another study). On the one hand, this calls for even more detail in the model, on the other hand it poses the question of how detailed a "final" model is to be and what the relevance of each of these factors is for the required outputs. *Fuzzy logic* can remedy these problems, since fuzzy logic enables consideration of different, partially conflicting objectives of drivers (see Hoogendoorn *et al.* (1997)).

4.5.1.2 Model calibration

There is apparently a clear interaction between several driver behaviour rules in the model, which stresses the complexity and means that parameter modifications cannot be carried out in a isolated way. This also complicates possible calibration activities and validation. Regarding calibration efforts, the lack of data necessitates to macroscopic or mesoscopic calibration that cannot produce the "perfect" data set, as the number of degrees of freedom is way too large. Brackstone and MacDonald (1998) recommend that new data sources should be used (instrumented vehicles), and that models should be dis-assembled and tested in a step-by-step fashion. That is, whenever new behavioural rules are added to the model, they should be tested extensively, and preferably in *isolation* (if possible). Moreover, calibration of microscopic simulation models using mesoscopic or macroscopic quantities is advised against, since it appears that this potentially yields incorrect estimations of the microscopic parameters at hand.

Another powerful instrument in model calibration is the use of sensitivity analysis to study the impacts of changes in the model parameters. This is especially useful for complex microscopic models, in which the effects of the values of the parameters on the flow behaviour are hard to analyse mathematically.

Finally, let us stress that an importance tool for model validation are so-called *turing tests*, where results from model application are presented to practitioners in the field. Simply stated, a model passes a turing test, if the practitioner is unable to detect whether the simulation results stem from a model, or are measurements collected from real-life situations.

4.5.2 Continuum models; first- and higher-order model controversies

The question which model type should be considered is not self-evident, given the ongoing debate between the LWR-model followers and the Payne-type model (and Helbing-model) followers. In this section we will discuss the main arguments presented by both streams to show the relative superiority of their respective modelling approaches. Let us remark that Lebaque and Lesort (1999) propose a methodology for theoretically comparing LWR-type models and Payne-type models. They propose a set of problems and situations that can be used as a test-case for model comparison. Although their approach has obvious limits, their work contributes to constitute an exhaustive model comparison framework.

4.5.2.1 Critique on LWR-type models

The simple continuum model has some shortcomings, given in the following list (see Liu *et al*. (1998), Zhang (1998), and Papageorgiou (1998)):

- *Steady-state speed-density relationship*. The LWR-models contain *stationary speed-density relations*, implying that the mean velocity adapts *instantaneously* to the traffic density rather than considering some delay. That is, the kinematic theory does not allow fluctuations around the equilibrium speed-density relationship.
- *Discontinuities in the density.* The kinematic wave-theory shows shock wave formation by steeping velocity jumps to infinite sharp discontinuities in the density. These are in contradiction with smooth shocks observed in real-life traffic that can be described by higher-order flow models.
- *Regular start-stop waves.* The LWR-theory is not able to describe regular start-stop waves with amplitude-dependent oscillation times that are observed in real-life traffic (e.g. Verweij (1989)).
- *Traffic hysteresis.* In real-life traffic flow, hysteresis phenomena have been observed (cf. Treiterer and Myers (1974)), showing that the average headways of vehicles approaching a jam are smaller than vehicles flowing out of a jam. These hysteresis phenomena are not described by the LWR models. Conversely, the Payne-type models are able to describe traffic hysteresis (see Zhang (1999)).
- *Localised structures and phantom-jams*. Similarly, the LWR-models are not able to predict the occurrence of localised structures and phantom-jams, i.e. the LWR-theory does not describe the amplification of small disturbances in heavy traffic.

4.5.2.2 *Critique on Payne-type models*

The most fundamental criticisms regarding the Payne-type models have been formulated in Daganzo (1995). The author's criticisms stem from the dissimilarity between vehicular flow and the flow of molecules in a fluid or a gas:

- *Anisotropy*. In opposition to fluid particles responding to both stimuli from upstream and downstream, a driver-vehicle combination is an anisotropic particle. In other words, a driver will primarily react to traffic conditions downstream. In opposition to vehicular flow, particles in a fluid or a gas 'react' to stimuli from all directions.
- *Unaffected slow-vehicles*. The speed of slow vehicles should be virtually unaffected by faster vehicles. Conversely, the slow particles in a fluidic flow or gas flow are affected by faster particles.
- *Personality*. Unlike particles in a fluid or gas, driver-vehicle combinations have their own personalities that remain largely unaffected by traffic conditions. For instance, drivers are aggressive or timid, or drivers aim to drive at their desired velocities.

Daganzo (1995) shows that existing Payne-type models can result in negative speeds at the tails of congested regions. This is caused by the second order dissipation term. However, Liu *et al*. (1998) unconvincingly argue that the violation of the anisotropy condition is a result of the pressure term. They state that by observing the *inviscid flow equations* under congested conditions one characteristic curve moves *upstream*, while the other characteristic moves downstream (with a velocity which is larger than the average velocity of the flow). However, these characteristics describe the upstream moving influence of congestion downstream, and the fact that some vehicles drive faster than others do. They *do not reflect physical movements of vehicles in traffic flow*. In other words, the fact that a characteristic curve is directed *upstream* does not imply that vehicles move in that direction. Neither is it necessary that all vehicles have the same velocity. In fact, c_0^2 can be interpreted as the *variance* in the velocities of the different vehicles.

4.6 Summary

This chapter discussed how the observed behaviour in congested traffic flow can be incorporated into microscopic and macroscopic traffic flow models. Let us summarise the main findings first for *microscopic models*:

1. Multi-regime approaches, where behavioural mechanisms and parameters depend on past and prevailing traffic conditions, are most likely to enable correct description of driving behaviour during congestion, and during transitions between free-flowing operations. Moreover, irregular lane-changing behaviour near discontinuities can be included as well.
2. An important issue is the so-called *merge give-way rule*, dictating that drivers on the main-road provide sufficient space to drivers from the on-ramp to merge into the main stream.
3. Due to the increasing complexity of microscopic models, calibration must be performed in a step-by-step approach, using car-based data sources (e.g.

instrumented vehicles). Moreover, fuzzy logic can be used to handle possible conflicting rules.

4. Stability of wide jams can be described by considering so-called slow-to-start driving rules. General Motors car-following models are unsuitable for describing traffic flow operations in general and the stability of wide jams in particular, due to their inability to limit occurring oscillations to realistic values, yielding unreasonably high acceleration and deceleration rates.

For macroscopic models, the following issues have been discussed:

5. Similar to multi-regime approaches is microscopic models, relaxation and anticipation behavioural parameters in macroscopic models must depend on the traffic conditions. This can be achieved by explicitly considering dependence on the relevant traffic parameters (e.g. $\tau = \tau(r)$, $c_0 = c_0(r)$), or introduction of dynamic equations for the different flow parameters, such as the local wave velocity c_0.
6. Non-local interaction terms improve description of congested traffic flow.
7. First-order macroscopic models cannot capture transient behaviour, queue head displacements, etc. The accuracy of the within-jam operations description depends on the type of congestion in the jam.
8. An increasingly important role is played by macroscopic models that are build upon microscopic principles, for one since these enable the straightforward implementation of traffic control affecting the behavioural driver parameters. These models appear to close the gap between microscopic and traditional macroscopic models.

Most of these modelling solutions will create additional model complexity, and complicate model calibration. This is why the model should only contain parameters that relate to directly observable driving behaviour, instead of including *non-observable explaining factors*, such as state-of-mind of a driver, or fatigue level, since it will not be possible to validate such mechanisms in practise.

Preferably, the different behavioural rules of these models are calibrated in isolation as much as possible, using microscopic data. Moreover, a power method in model calibration and validation is sensitivity analysis.

5 DATA-ANALYSIS

Chapter Research problem and focus aimed to identify current lacks in empirical and theoretical knowledge concerning traffic flow modelling during congestion. In chapter State-of-the-Art review on Congested Traffic Flow, the current state of knowledge w.r.t congested traffic flow operations was sketched, thereby answering some of the questions raised in chapter Research problem and focus, while some issues are still unresolved. Moreover, it appears that the flow theories are seldom general, i.e. can be applied directly to any situation. Most of the observed phenomena and flow theories only pertain to specific situations for particular locations, characterised by legislation, driving rules, and drivers' attitudes.

To gain more insight into macroscopic and microscopic behaviour of congested traffic flow operations, an extensive and costly data collection programme is needed, especially in case of the microscopic behaviour. This chapter discusses the kind of data that would be required for this research.

5.1 General approach to model development

Generally, development of operational traffic flow models can be divided into the following steps:

1. Establishing and testing theories on observed microscopic and macroscopic traffic flow behaviour, either using inductive or deductive reasoning. In this phase, qualitative relations and mathematical dependencies between the flow variables and parameters of interest are established, frequently based on empirical knowledge and qualitative data analysis.
2. Once the structure of the driver and traffic behavioural process in congestion is understood and models have been formulated to represent these processes in an approximate fashion, it is likely that calibration of the models for a specific application using empirical data, is necessary.
3. Validation of the calibrated models using different data.

For corrective modelling of traffic flow in case of traffic congestion, development and rigorous model testing is necessary in all the aforementioned phases. For instance, predicting the precise location of traffic breakdown may require the development of new traffic flow theories, for which detailed data analysis is needed. This empirical analysis may result in new models describing the (qualitative) mechanisms of displaced traffic breakdown. To apply this newly developed model (e.g. on a different location), the model parameters are calibrated using empirical data. Furthermore, the calibrated model is validated using different data. The calibration data-set and the validation data-set can be of the same, or of a different type.

This chapter addresses requirements for data and the consequent data collection methods for all phases of traffic flow theory and model development, testing, calibration, and validation. Given the scope of the research project, emphasis will be on research data criteria. The problem of model calibration and subsequent validation is tackled in another project. Nevertheless, some relevant issues are discussed.

5.2 Criteria for data-analysis for research purposes

Based on the results of both the literature survey, as well as the interview with national and international experts in the field of traffic flow modelling, the general conclusion holds, that both microscopic driver behaviour as well as macroscopic traffic flow operations during congestion are not well understood yet. With respect to congested flow modelling, the current lack of knowledge is two-fold, namely:

1. *Empirical knowledge* of real-life traffic flow operations prior, during, and after congested traffic flow, as well as upstream, inside, and downstream of the congested area, is still lacking. This holds for both levels considered in this research, i.e. microscopic behaviour and continuum traffic flow characteristics.
2. Partly due to the lack of empirical knowledge, it is unknown whether the *traffic flow models* currently used in the Netherlands (Flexsyt, Fosim, Mixic, Aimsun2, etc.) are, once correctly calibrated, able to correctly describe traffic flow operations prior, during, and after congestion. Currently used models are seldom exhaustively calibrated and validated.

To gain more insight into both of these issues, measurement data is required. In this respect, this note discusses the following issues:

a. How to determine what type of data is required for these purposes.
b. How these data can be obtained (measurement or observation methodologies).

In this chapter, these issues are addressed based on the results obtained in the previous phases of this research, that is, based on the results of the interviews and the literature survey.

5.2.1 Focus of data analysis for research purposes

Based on the results from the previous project phases, it can be concluded that data analysis with respect to traffic flow modelling during congestion must focus on (at least for the Dutch case, where self-formation of traffic jams is generally not an issue due to the lack of long, homogeneous roadway stretches):

- Empirical analysis of traffic flow at the location of traffic breakdown (mostly at discontinuities);
- Modelling of congestion propagation and traffic flow behaviour inside the congested area.

The first point relates mainly to the question why, where, when and how congestion occurs, while the second question relates to among other things the precise workings of blocking back (spillback) mechanisms, especially due to discontinuities (weaving areas etc), and the behaviour of traffic inside the congested area.

These two questions lead to different detailed subquestions, depending on the level of modelling (micro- or macroscopic). In this respect, note that it is still not clear *if and how* traffic flow operations *inside* the congested area influences the location of the queue tail. If the latter location is not influenced by traffic operations inside the queue, that is, if traffic operations inside a traffic jam are homogeneous, and predictable, on a *macroscopic description-level*, capacity (queue discharge rates) and a flow-density relation (fundamental diagram) for the entire congested area suffice to predict bottleneck congestion and propagation, in combination with a factor that determines the distribution of capacity over merging or dividing streams at discontinuities. Otherwise, macroscopic characteristics of traffic flow (i.e. stop-start waves) must be considered.

Similar requirements hold for data analysis on a *microscopic level*. That is, when the behaviour of drivers inside the congested area does not influence the location of the queue tail, driving behaviour (car-following and lane changing behaviour) at the bottleneck and at the tail of the congested area are of relevance, as well as at the discontinuities.

On both levels, the importance of the dynamic nature of the variables and processes is to be emphasised. That is, relevant research questions are: does capacity drop as a result of congestion? How variable is the congestion capacity? Can this variability be explained, for instance by considering traffic composition, or it is random and therefore unpredictable? Do drivers show a different following behaviour when accelerating from congestion in comparison to decelerating when entering congestion?
Most analysis techniques (capacity estimation methods, fundamental diagrams, headway distribution estimators etc) until now do not address the mentioned dynamics or randomness in driver behaviour and traffic flow processes, nor are these correctly incorporate in the currently used models. With respect to this uncertainty, let us emphasise that the level of randomness to a large extent determines the predictability of the congestion processes and the degree of accuracy that can be expected from (or even attained by) traffic flow models.

In summary, the indicators that are required in the research of traffic and driver congestion behaviour relate to *several locations on the roadway*. Their precise specification depends critically on the level of detail to be considered. Table 1 gives an overview of the points made in the previous paragraphs.

Table 1 Overview of relevant issues for data analysis on microscopic and macroscopic levels.

Location	**Micro level**	**Macro level**
Bottleneck	Dynamics of car-following behaviour; dynamic lane-change behaviour.	Dynamics of capacity (capacity drop); dependence on traffic composition; location of traffic breakdown; randomness.
Congestion tail	Dynamics of car-following behaviour; dynamic lane-change behaviour; dependence on dynamic characteristics inside congested area.	Characteristics of congestion propagation; dependence on dynamic characteristics inside congested area.
Discontinuities	Dynamics of merging behaviour for traffic on main road and merging traffic.	Capacity distribution rule (dynamic).
Congested area	Dynamics of car-following behaviour; dynamic lane-change behaviour.	Dynamic flow-density relation; relaxation of traffic flow.

In the remainder of this note, we will discuss the consequences of the above issues for both data collection at the macroscopic level as well as the microscopic level. To this end, several issues are addressed:

- location of data collection (where must measurement equipment be located?);
- level of aggregation / temporal density (are individual vehicle data required? What is the required period for time averaging?);
- spatial density (what is the ideal distance between detectors?;
- periods of data collection;
- variables to be measured;
- external variables, and finally;
- indicators to be derived.

5.2.2 Consequences for macroscopic level

In general, macroscopic traffic flow modelling is not concerned with *explaining* the behaviour of individual vehicle behaviour, nor with explaining the *resulting effects* on macroscopic characteristics of traffic flow. In illustration, although the net effect of traffic hysteresis can be included in continuum models, the underlying microscopic behaviour inducing the differences in queue inflow and outflow is generally not included in most macroscopic models. Exceptions are the *gas-kinetic models* (and the macroscopic model derived from these gas-kinetic models): in this case, aggregate flow behaviour is derived directly from microscopic behaviour of traffic.

This type of research should therefore aim at establishing *quantitative relational dependencies* of traffic flow variables on *internal variables* (such as traffic demands, mean speeds, traffic composition), and / or *external conditions* (such as road, weather, and ambient conditions, time-of-day). Moreover, the level-of-randomness (to what extent can we predict the magnitude of the considered parameter?) is of interest as well.

Another new development in macroscopic traffic flow modelling is the distinction of roadway lanes, and more importantly, user-classes (person-cars and trucks). Not only will distinction of classes (and lanes) yield improved insights into the traffic behaviour prior, during and after congestion (e.g. synchronised flow is characterised by the occurrence of small differences in person-car and truck velocities, while the speeds on the roadway lanes are also nearly identical), the significant differences in traffic flow behaviour due to varying traffic compositions can only be correctly explained by including class distinction, and consequent data that supports this distinction. In illustration, on of the results of the literature survey was the following: the wide scattering in the phase-plane during synchronised flow can to a large extent be contributed to the incorrect handling of different classes.

In order to be able to carry out research into the congestion process at a *macroscopic level* the following data criteria hold (based on the previous discussion).

5.2.2.1 Data collection locations

At the bottleneck (or better: the head of the congested area), within the congested itself, at the discontinuities, and finally up- and downstream of the congested area. Different traffic situations / bottleneck types (number of lanes on the main-road, light/moderate/heavy on-ramp, spill-back from off-ramp, etc.).

5.2.2.2 Level of aggregation and temporal density

Preferred: individual vehicle measurements (cross-section passage instants of both loops per lane, enabling among other things the *lane-specific* determination of rear-bumper passage instants, vehicle lengths, and velocities of individual vehicles); class specific measurements (person-cars and trucks) can be determined from measurements.
Alternative: one-minute time-averages (or shorter), per lane and per vehicle category (trucks and person-cars), with as little pre-processing or filtering as possible. Class-specific and lane-specific measurements are preferred.

5.2.2.3 Spatial density of data collection

- *At the bottleneck*: a few data collection stations, approximately 500 m apart to determine the bottleneck location.
- *In the congested area*: between all discontinuities at least one data collection station to determine the flow-density relation
- *At the discontinuities*: at least one data collection station per stream to determine the capacity distribution.
- *Upstream and downstream of congested area:* one data collection station upstream and one downstream of the congested area to determine traffic demand and to check for spillback of downstream bottlenecks

5.2.2.4 *Periods*

Periods containing entire range of traffic conditions (pre-traffic breakdown (short), traffic breakdown, congested flow, dissolution of congestion, post-congestion (short)), for several days to determine day-to-day variability

5.2.2.5 *Measurement variables*

Individual vehicle data: vehicle passage times for both loops of double loop detector.
Aggregate measurements: flows (number of vehicles passed during minute) and speeds/occupancies; speed variance (optional); class-specific and lane-specific measurements.

5.2.2.6 *External information*

Detailed geometry information
Data collection location positions
Road-, weather-, and ambient conditions

5.2.2.7 *Indicators to be derived*

Spatial and temporal distribution of flow, density, speeds, and speed variations; contour and surface plots indicating how congestion propagates.
Dynamics of traffic flow operations at different measurement locations, reflected in the phase-planes.
Capacity dynamics; capacity probability distribution functions.

From the above it is clear that in principle the provision of this type of data is currently not a real problem. Several data collection systems (based on loop detection) are operational or available.

Potential data-sources: inductive loops (MONICA, MARE, RE SI, Peek), video, infrared and radar, …

5.2.3 Consequences for microscopic level

Research at the microscopic level requires a much higher detail of data collection than at macro level in order to be useful. Brackstone and MacDonald (1995) convincingly argue that for the calibration of true microscopic simulation models[5], requires more detailed information on the *pair-wise behaviour of vehicles*. That is, to improve modelling of car-following and lane-changing behaviour during congested traffic flow operations

---

[5] Here, *true microscopic models* refer to microscopic simulation models in which the behaviour of the vehicles is determined directly by the individual vehicles in their direct environment. In control, semi-microscopic models (such as Integration) describe the dynamics of vehicles as relations with *aggregate characteristics* of traffic flow (e.g. densities, and speeds).

significantly, the microscopic behaviour of vehicles *compared to vehicles in their direct environment* to which they react and on which the anticipate, is needed.

The main question related to this research is whether is possible to build a theory of driver behaviour that can withstand validation at different locations and days and that is strong enough to explain the phoneme mentioned above. The fundamental character of this question expresses itself in the *high level of detail required* in the research data.

We here state that the type of fundamental research meant here is not possible without data that consists of complete vehicle trajectories recorded at a high frequency, and (detailed) information on the respective vehicles (e.g. vehicle-type, length). Longitudinal and lateral positions are to be collected very precisely in order to determine the exact lane changing and acceleration/deceleration behaviour of the vehicles.

At this point, let us mention that a fundamental requirement is further assumed in the sense that only variables are considered, that can (in principle) be directly measured or observed. Models could of course include *non-observable explaining factors* (for instance, the state-of-mind of a driver, fatigue level). As it will not be possible to validate such mechanisms in practice, we consider only models that do not aim for explanation of these influences, but treat these as (non-modelled) internal processes, noise, or variation around the 'average' behaviours. Whether this is at all possible, is one of the main questions to be answered in modelling at a microscopic level.

5.2.3.1 Location (similar to macroscopic)

At the bottleneck (head of the congested area), within the congested itself, at the discontinuities, and finally up- and downstream of the congested area.
Different traffic situations / bottleneck types (number of lanes on the main-road, light/moderate/heavy on-ramp, spill-back from off-ramp, etc.).

5.2.3.2 Level of aggregation / temporal density

Measurements each 0.1 sec (or less);
Vehicle location per one tenth of a meter (or less).

5.2.3.3 Density of data collection

Complete vehicle trajectories, preferably also of vehicles in direct environment (pair-wise analysis of driver behaviour).

5.2.3.4 Periods

Complete congested periods for several days under varying circumstances with respect to road, weather, and ambient conditions, to determine day-to-day dynamics and influence of external conditions.

5.2.3.5 Measurement variables

Lateral and longitudinal location and time per vehicle.
Relative distance and relative velocity with respect to other vehicles in the direct environment;
Macro attributes in the vicinity of the vehicle (traffic composition per lane, velocity per lane);
Distance with respect to specific locations (bottlenecks, diverges, etc.)

5.2.3.6 External information

Detailed geometry information
Data collection location positions
Road-, weather-, and ambient conditions

5.2.3.7 Indicators to be derived

Time and space headways and gaps with respect to other vehicles in the direct environment (pair-wise analysis);
Acceleration, and deceleration behaviour.
Response / reaction times (time differences between stimulus (e.g. car in front braking or acceleration and respective reaction of driver; slow to start rules?)
Regression analysis of driver behaviour and micro/macro traffic flow attributes (e.g. influence of heavy vehicle factor, average velocities, etc.)

It is clear that this type of data is not yet readily available, as it has a distinct spatial and individual vehicle related character, that standard road-side collection methodologies cannot provide.

Data sources: extended floating car data (instrumented vehicles); video data.

5.3 Criteria for data for model calibration and validation

Once the structure of the driver and traffic behavioural process in congestion is understood and models have been formulated to represent these processes in an approximate fashion, it is likely that calibration of the models for specific applications is necessary. It is assumed that at this time, model verification has already been performed successfully.

5.3.1 Definition of calibration and validation

By definition, model calibration is the numerical estimation process to determine correct values of the model parameters, based on traffic measurements. In other words, calibration roughly to adapting the unspecified model parameters and coefficients

optimally (in a yet to be specific manner), such that the model is suitable for a specific situation. Since noise plays a role of dominant importance, statistical procedures are likely to be used in the calibration phase. Moreover, in some cases, it may be beneficial to calibrate specific submodels in an isolated fashion.
Validation goes one step beyond model calibration: the objective is to test whether the model is able to describe selected phoneme sufficiently accurate, both from a qualitative and quantitative point of view. In this phase, the *whole* model (i.e. not only submodels) need to be tested.

5.3.2 Required detail level

Generally, the level of detail of the data used for this exercise is much less than for the research. Whether this is indeed the case, depends on whether *univocal relations* exist between model parameters θ and mesoscopic or macroscopic traffic flow variables (density, flow, distribution of vehicles across roadway lanes). However, in some cases changes is model parameters do *not* lead to unique changes in the relevant mesoscopic and macroscopic variables. In this case, the latter variables can not be used to determine the model parameters, unless additional information is available. However, the practical use of models which performance depends critically on parameters that are hard or even impossible to calibrate for practical applications is unclear. This leads to the question, whether only models should be considered, whose unknown variables can be determined using generally available data (e.g. time-averages of macroscopic variables).

In this respect, an important tool to determine the influence of certain parameters is sensitivity analysis, where the relation between model input and output is analysed. However, it is emphasised that (local) methods should be applied with case, since the relation between model parameters and outcomes is generally non-linear and shows features of chaotic processes.

In the remainder of this chapter, some general issues on calibration and validation of respectively macroscopic and microscopic flow models are made. Finally, data requirements for either are discussed.

5.3.3 Calibration and validation of macroscopic models

This section discusses calibration and validation of macroscopic models. These models are divided into two classes, namely the (traditional) first-order / higher-order models (established using the analogies between vehicular flow and fluidic flow), and macroscopic models determined from microscopic principles.

5.3.3.1 First-order and higher-order models

On of the main advantages of macroscopic traffic flow models is their straightforward calibration. Usually, model relations and parameters, typically speed density relations, or speed variance as a function of the density, are determined directly from the

calibration data and applied to the model. Since most macroscopic models only contain directly observable quantities (e.g. speed-density relations), model calibration is relatively straightforward.
In case traffic parameters are not directly observable from available data (e.g. relaxation time and kinematic viscosity), the dynamic behaviour of the considered macroscopic model could be compared to available data, and .

The fundamental question is then, whether emphasis in the calibration phase should be adapting the model parameters such to achieve correct *reproduction of the phenomena of interest* (for instance, describing the downstream location of traffic breakdown in case of a bottleneck, or the location of the queue tail). If so, care should be taken in the (macroscopic) meaning and implications of the parameter values. That is, the value of the model parameters should be plausible as well, and not just yield the best fit with available data.

Consequent validation of the model is then achieved by comparing modelling results with validation data (e.g. different days, different conditions, etc.). In principle, validation data can (but does not have to) be less detailed (time-average data instead of individual vehicle data) than the research data and the calibration data.

With respect to calibration and validation effort, compared to microscopic models, the number of unknown model relations and parameters in continuum models is relatively low (and are mostly directly observable from the currently available data). For *first-order models*, only the *speed-density* curves need to be determined, while higher-order models require additional parameters to be determined (*relaxation times*, *kinematic viscosity*, *velocity variances*, etc.).
Note that in improving the performance of continuum models, increasing model complexity may reduce the ease with which these models can be calibrated, due to the addition of model parameters and complex relations.

5.3.3.2 Macroscopic models based on microscopic principles

Recently, macroscopic model have been established based on *microscopic principles*. These models represent the relations between speed and density as outcomes competing processos:

- Vehicles aiming to accelerate towards the desired velocity;
- Interaction between fast and slow vehicles, and the probability fast vehicles are able to (immediately) changes lanes.

To *ensure the correct description of traffic flow at high densities*, vehicle spacing requirements are explicitly considered in these models (either by redefining traffic density, or introducing non-local interaction terms). To this end, vehicle spatial use, depending on a.o. the vehicle length, and the reaction time, are considered explicitly. These macroscopic models require determination of free-speeds, (equilibrium) velocity variances, immediate overtaking probabilities, acceleration times, reaction times, and vehicle lengths. Most of these parameters can be determined from empirical speed-density and speed variance-density curves (see Helbing (1997), Hoogendoorn (1999)). Moreover, these new developments have led to new insights in the interpretation of the different macroscopic parameters and relations. For instance, the *anticipation constant* c_0 reflects the spread in velocities, while the viscosity coefficient can be interpreted as the transition rate from brisk to careful driving. This enables application of different techniques for model calibration.

5.3.3.3 Model calibration and validation

With respect to model calibration, *first-order models* require construction of speed-density (or flow-density) curves. These can mostly be determined from currently widely available traffic data (1 minute averages of free-flow, and congested flow, at various locations, to ensure a complete fundamental diagram).

For *higher-order models* and *microscopically-based models*, in addition to the speed-density curves and the respective data requirements, other model parameters can be estimated using different techniques. For instance, the *anticipation factor* c_0 can be determined either a) from the velocity with which shockwaves propagate in the flow, or b) using speed variances (considering the meaning of the parameter from a microscopic point-of-view). In the first case, requirements with respect to the spatial density is required. In the second case, specific analyses techniques are required to establish correct speed variance-density relations.
Similarly, the *relaxation time* can be determined by comparing the time in which the traffic flow adapts to the equilibrium flow. Alternatively, the microscopic meaning of the relaxation time can be used, while it can be determined from the time needed for a vehicle to accelerate towards its free speed, or the free-speed f the platoon-leader. The kinematic viscosity is mostly determined using trial-and-error (considering whether model results are similar to empirical observations).

As with the first-order models, these parameters can mostly be determined from *currently available traffic data* (1 minute averages of free-flow, and congested flow, at various, densely spaced locations), using dedicated model calibration techniques (i.e. theoretical considerations; see Helbing (1997)). Alternatively, other data-types are useful (and in some cases preferable) as well (individual vehicle data). Newly developed lane-specific and class-specific flow models require data to accommodate lane and vehicle-type distinction.

Potential data-sources: inductive loops (MONICA, MARE, RE SI, Peek), video, infrared and radar, more detailed data (individual vehicle measurements, floating-car data).

5.3.4 Calibration and validation of microscopic models

During data analysis, traffic theories and models of driving behaviour are developed or improved, using observed (qualitative and quantitative) driver behaviour (e.g. car-following, lane-changing behaviour). Having established these, the resulting microscopic models must be *calibrated*[6].

5.3.4.1 *Approaches to calibration and validation*

Due to the relative large number of parameters, calibration of model parameters is not necessarily straightforward, especially in case of unobservable model parameters, which relation with observable model behaviour and dynamics is unclear. In this respect, it is noted that it is preferable to have parameters that can be (in principle) directly observed from (available) observations.
For corrective model calibration, complex microscopic models and submodels should be dis-assembled, calibrated and tested in a step-by-step fashion, whenever possible.

After calibration, validation must ensure that the model behaviour matches readily observable features of traffic flow, given a number of selected situations and phenomena of interest.

Note that the lack of detailed data has led to mesoscopic calibration (using distributions of variables at locations instead of time series of dynamic and spatial variables), that cannot produce the "perfect" data set, as the number of degrees of freedom in the microscopic model is way too large. In other words, multiple parameter value combinations yield the same model behaviour with respect to the relevant observed model behaviour. Although these mesoscopic data is generally adequate for validation purposes, calibration data relates to *dynamic quantities*, that are relative to a moving vehicle and its surroundings, and these are exceptionally difficult to obtain[7]. Even if available, detailed data alone does not guarantee that good microscopic models result: calibration and validation should be carried out with care.

To aid the model calibration, sensitivity analysis can be used to gain insights into the dependence of the dynamic model results on the microscopic traffic flow parameters. This is especially useful in case of manual calibration (*trial-and-error*). However, the highly non-linear relations and chaotic-like dependence of the model parameters on the model outcomes requires that one should be very careful in applying indirect calibration techniques.

5.3.4.2 *Data requirements for microscopic calibration and validation*

What type of data is required for calibration of microscopic models depends critically on ones point-of-view, and the consequent approach to model calibration. On the one hand, it can be convincingly argued that *only directly observable model parameters* must be included in the microscopic model. The consequence of this is that data requirements

---

[6] Calibration of *microscopic* model is the process that concerns the tuning of parameters that determine the vehicle-vehicle interactions.

[7] When relations between microscopic (behavioural) parameters and macroscopic parameters are univocal, macroscopic data will generally suffice in the calibration phase as well.

are quite stringent, in that *individual vehicle data are preferred*, possibly covering entire vehicle trajectories, and the respective operational dependencies with other vehicles in the direct surroundings (pair-wise observations; see section Consequences for microscopic level).

Alternatively, (unobservable) model parameters can be determined such that the microscopic model yields specific macroscopic behaviour (e.g. correct speed-density curves; correct headway distributions; number of lane-changes, etc.). In this case, care should be taken that the model parameters are plausible given their meaning in the model (note aforementioned non-linear dependence and chaotic-like behaviour). In illustration, reaction times should be in the area of 0.5*s* and 2.5*s*. For the microscopic model on which the Payne models were based, reaction times of approximately 30*s* were found to yield the best model results.

For model validation, data criteria are less stringent. Generally, validation data reflects the specific traffic flow behaviours that the model is expected to describe. Most of these characteristic behaviours have explicitly put into the model during model development (see criteria for data analysis). In the validation phase, emphasise should be on quantitative analysis and accuracy. When the model is expected to describe macroscopic characteristics of the traffic flow, macroscopic data (1 minute average of speeds, flows, and densities, collected using inductive loops) can be used. To validate if the model correctly describes congested flow behaviour, the data must again be spatially dense (i.e. 500m between detector loops), to enable description of the queue head and tail behaviour, and the (average) behaviour inside the queue (see section Consequences for macroscopic level).

6 CONCLUSIONS AND RECOMMENDATIONS

In this report, results of the research project "*Traffic Flow Operations during Congestion*" have been presented, the aim of which was to identify issues relating to:

- Views of experts concerning issues in congested traffic flow operations and modelling (chapter 2).
- Current state-of-the-art concerning empirical and theoretical knowledge on traffic flow operations during congestion (chapter 3)
- How can this knowledge be implemented in traffic flow models, and how it can be applied in the models used for Dutch research studies (chapter 4).
- Data requirements for additional research (theory building, and validation, model calibration) in congested traffic flow operations (chapter 5).

This closing chapter aims at consolidation of the results of the different chapters of this report, aiming at providing both general and specific recommendations for enriching the current empirical and theoretical knowledge on congested traffic flow operations, and improving modelling of congested traffic flow both by including fundamentally new modelling approaches, as well as ny improved model improved model calibration approaches.

6.1 General approach to model enhancement and development

Before discussing tangible research results, conclusions, and recommendations, note that improving congested operations modelling requires performing the following general model development phases:

1. Knowledge acquisition from empirical data and theory building.
2. Validation of theories using empirical data.
3. Translating theories into modelling mechanisms.
4. Model calibration and testing.

From this perspective, the main problems when modelling congested flow operations are twofold. The first problem relates to points 1 and 2, in that *little empirical and theoretical knowledge is available about driver behaviour at both the onset and dissolution of traffic congestion, as well as during congestion*. For instance, we do not know why, and where drivers change lanes? Which headways are accepted (or chosen) before decelerating or accelerating, and how long these are accepted? What are the determining factors in this behaviour, and how random is this behaviour? We believe that this *lack of knowledge is fundamental* and relates to actual driving behaviour, not just to the modelling of it (despite the opinion of the international experts).

The second problem relates to points 3 and 4: it appears that the considered models are seldom calibrated thoroughly for the situation to which they are applied. This is why no definitive conclusions can be drawn with respect to problems experienced in model application, although it is clear that problems exists. This issue also relates to the need for a unambiguous model calibration methodology (see Development of comprehensive calibration methodology). Moreover, in most application studies, complex roadway geometries and networks are considered, implying that inaccuracies in modelling cannot be contributed to a single cause or submodel (e.g. route-choice behaviour, or traffic operations modelling).

To establish practical recommendations to improve the description of traffic flow operations during congestion, the following steps have been carried out:

a. Identification and demarcation of observed phenomena, traffic flow theories, data supporting theories, and traffic flow models operationalising (or at least phenomenally describing) these theories.
b. Drawing conclusions based on the identified gaps in
c. Recommendations for future research.

In combining these steps with the different stages in traffic research, the structure in Table 6 Conclusions and recommendations-6 has been established to provide recommendations of future research. This structure does not exhaustively consider each of the conclusions that can be drawn for the research, since some of them are much more general. These will be discussed separately in one of the following sections.

Table 6 Conclusions and recommendations-6 Structure in identifying recommendations

Observed phenomenon	**Identification and demarcation**	**Conclusions**	**Recommendations**
Knowledge acquisition and theory building	Is theory available? If so, describe theory	Approach to theory development	*Develop theory; verify theory using Dutch data; operationalise theory in models, calibration of model*
Supporting data available	Availability of data supporting theory? If so, which data	Finding supporting data	
Modelling	Do models feature this theory? If so, how?	Research into theory operationalisation	
Calibration / validation	Are these rules calibrated?	Approach to calibration of submodels	
Relevance	Is issue at all relevant for application studies?		

6.2 Results from identification and demarcation

Let us first summarise the results with concerning identification and demarcation of phenomena observed in traffic flow, the (availability of) respective traffic flow theories, and the subsequent modelling approaches to include these problems in operational traffic models (if available). These results are split up into the results from the Dutch application studies, and the results stemming from the literature survey.

6.2.1 Dutch application studies

The following table summarises the *general problems in congested traffic flow description* for Dutch application studies (see section 2.6.1):

Table 6 Conclusions and recommendations-7 Discrepancy between phenomena observed in real-life traffic flow and description in traffic models. Numbers do not reflect importance of issues. Note that in this case, no distinction is made between 'observation' and 'theory'.

Identification and demarcation			
Observation / theory	**Supporting data?**	**Modelling available (only microscopic models) and correct?**	**Calibrated / validated?**
1. Traffic behaviour is non-smooth and random due to inter- and intra driver differences	Application studies using data collected at (mostly) Dutch motorways	No, model results are too smooth; there is too little randomness.	No
2. Transitions between traffic regimes is relatively smooth		No, modelled transitions are too sharp.	No
3. Drivers accept (some) difference in free speed and speed of leading vehicle		No, drivers try to overtake when speed below free speed.	No
4. Drivers tend to stay on the median or middle lane for some time, after the overtaking manoeuvre is completed		No, drivers return to their lane as soon as manoeuvre is completed.	No
5. Drivers on the main road 'provide gaps' to merging traffic		Models have been developed, but are not applied in models applied in Dutch case studies.	No
6. High merging traffic flow-rates, even if congestion on the main road has occurred			No
7. Driving and traffic behaviour depends on flow regime, network component, and external factors		Limited models are available, and included in applied models; multi-regime modelling frameworks have been developed.	Partially

6.2.2 Literature survey

Table 6 Conclusions and recommendations-8 summarises the results from the literature survey with respect to observations and theories of congested traffic flow. For details, we refer to chapter 3 of this report.

Table 6 Conclusions and recommendations-8 Discrepancy between phenomena observed in real-life traffic flow and description in traffic models. Numbers do not reflect importance of issues. In this case, observations and theories are distinguished.

Identification and demarcation				
Observation	**Theories**	**Supporting data?**	**Modelling**	**Calibrated / validated?**
8. Difference in high prior-congestion flow-rate and queue discharge flow (capacity drop); differences in range of 5-10% upto 50%.	a) Car-following behaviour is different before and after traffic breakdown. b) Difference between *ancitipation dominant stage* and *relaxation dominant stage* (Zhang,1999). c) Catastrophe theory and utility maximisation of drivers (Van Toorenburg (1983)).	a) Dutch data. b) US data. c) No data (theoretical model)	a) *Microscopic*: Extended FOSIM differentiates between behaviour before and after congestion has occurred; models are limited b) *Microscopic models*: multi-regime approaches; *Macroscopic*: higher-order models (METANET) include required mechanisms, at a crude level. c) Not operationalised in applied models.	Partially.
9. Observations of high-prior congestion flows for considerable period (upto 40 minutes)				
10. Traffic hysteresis				
11. Occurrence of *spontaneous transitions* between regimes.	Distinction of stable, *metastable*, and unstable traffic flow; transition occurs in case of critical perturbation	German data	Microscopic: CA-models, stochastic models, and multi-regime models with slow-to-start rules. *Macroscopic*: higher-order models (Kerner *et* al (1996)); non-local interaction terms (describing synchronised flow).	No

Identification and demarcation				
Observation	**Theories**	**Supporting data?**	**Modelling**	**Calibrated / validated?**
12. Displacement of queue head 500m to 1000m downstream of discontinuity	Drivers smoothly adapt their headway and velocity, e.g. in case of merging drivers.	US data, Dutch data.	*Microscopic*: FOSIM describes displaced location of queue head *Macroscopic*: higher order models include required mechanisms; first-order models unsuitable.	No
13. Observation of different types of congested traffic flow	a) Kerner (1999): free-flow, synchronised flow, and traffic jams. b) Helbing *et al*. (1998): homogeneous congested traffic, localised clusters, start-stop waves, oscillatory congested traffic, free-flow	a) German data. b) No (results from model, calibrated using Dutch data).	a) Traffic jams: microscopic models with slow-to-start rule / multi-regime models a+b) synchronised flow *and* traffic jams: non-local macroscopic models; macroscopic models with condition-dependent parameters	No
14. Location of the tail of the queue.	a) Dynamics of the queue tail depends on the type of congestion that emerges (homogeneous congested traffic, oscillatory congested traffic); b) Shockwave and queue theories.	No (results from model, calibrated using Dutch data).	Non-local *macroscopic* model based on microscopic principles describes occurring congestion types.	Partially
15. Correct description of propagation of wide jams, existing upto several hours	a) Wide jams result from bottlenecks *or* are self-formed. b) Stability requires jam inflow to be *at least as large* as jam outflow for a considerable period (slow-to-start rules; see hysteresis).	German data, Dutch data	b) *Microscopic*: slow-to-start rule and multi-regime approaches a+b) *Macroscopic*: models include necessary mechanisms	No

Identification and demarcation				
Observation	**Theories**	**Supporting data?**	**Modelling**	**Calibrated / validated?**
16. Differences in car-following behaviour for different regimes	Driving behaviour is dependent on prevailing traffic conditions; see also Traffic hysteresis.	Dutch data, US data	a) *Microscopic*: multi-regime (location-dependent) car-following models a) *Macroscopic*: condition-dependent parameters; platoon-based description of traffic flow.	No
17. Observed differences between user-classes	During free-flow differences between classes are considerable, while during homogeneous synchronised flow / homogeneous congestion, differences are diminished; differences are large during oscillatory congestion.	Dutch data	Most microscopic models describe differences. Multiclass macroscopic models based on microscopic principles	Yes

6.3 Conclusions

In the previous section, we have summarised the results of this report as far it considers the description of the current state of empirical knowledge, traffic flow theories and supporting data, and the models or submodels operationalising these theories. Let us now reconsider these phenomena and
summarise the results with respect to identification and demarcation of phenomena observed in traffic flow, the (availability of) respective traffic flow theories, and the subsequent modelling approaches to include these problems in operational traffic models (if available).

The following table summarises the *general problems in congested traffic flow description* for Dutch application studies (see section 2.6.1), and the literature survey (chapters 3 and 4).

Table 6 Conclusions and recommendations-9 Conclusions from Dutch application studies and literature survey

Observation / theory	Conclusion

Observation / theory	Conclusion
1. Traffic behaviour is non-smooth / random due to inter- and intra driver differences	Comparison of model results with empirical data shows that the former are too smooth. No validated theories are available (suggestion from literature: driver behaviour is stochastic; difference between drivers is random.)
2. Transitions between traffic states are relatively smooth	Comparison between model results and empirical data shows that simulated transitions are too sharp. No validated theories are available (suggestion from literature: drivers anticipate on traffic conditions that are further downstream, e.g. two-vehicles ahead).
3. Drivers accept difference in free speed and speed of leading vehicle	From application studies it was concluded that models overestimate the number of drivers attempting to change lanes. No validated theories are available (suggestion: indifference bands exist in which drivers do not consider lane-changes. driver will not attempt lane-change if his conditions do not improve.)
4. Drivers tend to stay on the median or middle lane for some time, after the overtaking manoeuvre is completed	Dutch application studies reveal that drivers tend to return to the right lane too soon compared to empirical observations. No validated theories are available (suggestion: drivers change back to right lane, only if they have enough incentive to do so (e.g. vehicles moving in from behind), or with some (random) delay)
5. Drivers on the main road 'provide gaps' to merging traffic	Available theory: drives inhibit merge give-way behaviour. Modelling approaches have been developed, but are not calibrated for nor applied to the Dutch case.
6. High merging traffic flow-rates, even if congestion on the main road has occurred	Traffic on the main road has priority on traffic on main road, which inhibit merge-giveway behaviour. Definitive data supporting this for the Dutch case is not available, neither are calibrated models.
7. Driving and traffic behaviour depends on flow regime, network component, and external factors	Dutch and US data reveal differences in prior-breakdown and post-breakdown traffic flow operations (for which flow theories are available), and dependence of behaviour on location in network, e.g. near an on-ramp.
8. Difference in high prior-congestion flow-rate and queue discharge flow (capacity drop); differences in range of 5-10% upto 50%.	Capacity drop and traffic hysteresis are observed on different locations (a.o. Dutch, German, and US data; although properties are clearly site dependent). Different theories have been developed, explaining these observations. Among these theories is the distinction of driver stages (anticipation dominant, relaxation dominant, and combination), and the use of catastrophe theory in utility optimisation. The former appears most practically applicable, although no supporting data has been studied. Nevertheless, the theories underlies some operational flow models, which have been (partially) calibrated (FOSIM).
9. Observations of high-prior congestion flows for considerable period (upto 40 minutes)	

Observation / theory	**Conclusion**
10. Traffic hysteresis	
11. Occurrence of *spontaneous transitions* between regimes.	German observations show self-organisation of congestion. Self-organisation has not been reported in Dutch studies. The underlying theory has not been validated on other than German data. Models are available that describe the different phases of traffic flow, and the consequent (spontaneous) phase-transitions. These models have been calibrated using Dutch data, although in these cases, no self-organisation occurred.
12. Displacement of queue head 500m to 1000m downstream of discontinuity	US, and Dutch data shows that the queue head is displaced. No theory has been validated that explains this displacement, although several models show the same displaced behaviour.
13. Observation of different types of congested traffic flow	Different kinds of congestion has been studied extensively using German data (limited studies using Dutch data). Models have been developed, but not validated, describing these different congestion types.
14. Location of the tail of the queue.	It is hypothesised that the tail of the queue can be determined using rather crude flow theories, such as shockwave analysis and queuing theory. However, little research has been done on the validity of these theories, and the circumstances under which they are applicable. Models describe the location of the queue tail. The validity of this description has not been determined, at least not in the considered application studies.
15. Correct description of propagation of wide jams, existing upto several hours	Wide jams have been observed in Dutch and German data, although self-formed wide jams only have been observed in Germany. Data supports that these wide jams are indeed stable. Models have been developed, although not calibrated and validated.
16. Differences in car-following behaviour for different regimes	It has hypothesised that traffic behaviour depends on among other things (combinations of) traffic regime, and the location of the vehicles in the network. However, data supporting this hypothesis has not yet been considered. Multi-regime models have been developed, and tested (limited calibration and validation).
17. Observed differences between user-classes	Dutch data studies clearly show the differences in driving behaviour of the respective user-classes. These differences stem from differences in vehicle characteristics and driving behaviour. Validated microscopic models mostly consider different classes and their respective behaviours. Only a few macroscopic models do so. With respect to the latter, limited calibration and validation of the latter has been performed.

6.4 Recommendations for future research

Research recommendations can be deduced directly from the conclusions above, by identifying gaps in the current state of knowledge in congestion traffic flow theory and modelling. These research recommendations can be categorised in either of the phases discussed in section General approach to model enhancement and development:

1. Knowledge acquisition from empirical data and theory building, and
2. Validation of theories using empirical data.
3. Translating theories into modelling mechanisms.
4. Model calibration and testing.

This section discusses recommendations for future research based on the conclusions of the previous section, classified according to the categories 1-4 above. Due to their dependence, categories 1 and 2, and categories 3 and 4 are considered jointly. Moreover, the relevance of the observed phenomena is considered (see Table 2 Research problem and focus-3, Table 3 State-of-the-Art review on Congested Traffic Flow-4, and Table 4 Congested traffic flow modelling-5) **[to be done]**. Table 6 Conclusions and recommendations-10 provides an overview of the recommendations based on the conclusions delineated in **Fout! Verwijzingsbron niet gevonden.**.

6.4.1 Recommendations concerning knowledge acquisition and theory validation

Following the conclusions for this research, one prime result is the need to gain insights into the dynamics of congested traffic flow. Increasing empirical insights and into both microscopic and macroscopic behaviour of traffic flow, and consequent theory development requires a *data-orientated approach*. The main question related to this research is whether is possible to build theories of driver and traffic behaviour that can withstand validation at different locations and days, while having sufficient explanatory ability. The fundamental character of this question expresses itself by the *high level of detail required* in the research data, especially when driver behaviour is of interest.

This is why for *driver-based (microscopic) flow analysis*, an extensive and costly data collection program would be needed, and may not be possible without data that consists of complete vehicle trajectories recorded at a high frequency, and (detailed) vehicle information (see chapter 5). Longitudinal and lateral positions are to be collected very precisely in order to determine the exact lane changing and longitudinal behaviour.

On the contrary, due to the nature of macroscopic models, research on this level should aim at establishing *quantitative relational dependencies* of traffic flow variables on *internal variables* (such as traffic demands, mean speeds, traffic composition), and / or *external conditions* (such as road, weather, and ambient conditions, time-of-day). Moreover, the level-of-randomness (to what extent can we predict the magnitude of the considered parameter?) is of interest as well. This implies that for establishing empirical macroscopic knowledge and traffic flow theories, the availability of data is currently not a real problem. Several data collection systems (based on loop detection) are operational or available. The main issue is the spatial density of the data collection.

Keeping this in mind, not only is a proficient data analysis program of importance, also advanced tools for data visualisation and analysis are required (see section Development of advanced tools for data visualisation and analysis). Moreover, theory building by considering high quality data using advanced visualisation and data analysis tools may provide incentives for new flow theories concerning congested traffic flow operations.

Let us not discuss the issues that are recommended as subjects for data-orientated research aimed at establishing general theories of driver behaviour and traffic flow (depending on envisaged model application) are depicted in Table 6 Conclusions and recommendations-10.

Table 6 Conclusions and recommendations-10 Recommendations from Dutch application studies and literature survey.

Observation / theory	**Recommendations, perform:**
1. Traffic behaviour is non-smooth and random due to inter- and intra driver differences	▪ Data-orientated research aimed at establishing general theories of driver behaviour and traffic flow explaining observation. Research can in principle be flow-orientated or car-orientated, depending on the envisaged application of the theories. ▪ Validation of developed theories using (different) data sets. ▪ Mathematical operationalisation of theories, towards model development (e.g. establish dependence on key flow variables).
2. Transitions between traffic states are relatively smooth	▪
3. Drivers accept difference in free speed and speed of leading vehicle	▪
4. Drivers tend to stay on the median or middle lane for some time, after the overtaking manoeuvre is completed	▪
5. Drivers on the main road 'provide gaps' to merging traffic	▪ Validation of merge-giveway theory using empirical data collected at Dutch motorways, for different flow regimes and roadway geometries. ▪ Theory refinement (or even development of new theories), based on validation outcomes. ▪ Mathematical operationalisation of theories aimed at model development (e.g. establish dependence on key flow variables), depending on validation outcomes.
6. High merging traffic flow-rates, even if congestion on the main road has occurred	▪

Observation / theory	**Recommendations, perform:**
7. Driving and traffic behaviour depends on flow regime, network component, and external factors	▪ Data-orientated research aimed at establishing general theories of driver behaviour and traffic flow explaining observation, and its dependencies on key flow variables. Research can in principle be flow-orientated or car-orientated, depending on the envisaged application of the theories. ▪ Validation of developed theories using (different) data sets. ▪ Mathematical operationalisation of theories, towards model development (e.g. establish dependence on key flow variables).
8. Difference in high prior-congestion flow-rate and queue discharge flow (capacity drop); differences in range of 5-10% upto 50%.	▪ Validation of available theories (see **Fout! Verwijzingsbron niet gevonden.**) using empirical data collected at Dutch motorways. ▪ Theory refinement (or even development of new theories), based on validation outcomes. ▪ Mathematical operationalisation of available theories aimed at model development (e.g. establish dependence on key flow variables).
9. Observations of high-prior congestion flows for considerable period (upto 40 minutes)	▪
10. Traffic hysteresis	▪
11. Occurrence of *spontaneous transitions* between regimes.	▪
12. Displacement of queue head 500m to 1000m downstream of discontinuity	▪ Data-orientated research aimed at establishing general theories of driver behaviour and traffic flow explaining observation. Research can in principle be flow-orientated or car-orientated, depending on the envisaged application of the theories. ▪ Validation of developed theories using (different) data sets. ▪ Mathematical operationalisation of theories, towards model development (e.g. establish dependence on key flow variables).
13. Observation of different types of congested traffic flow	▪ Validation of available theories (see **Fout! Verwijzingsbron niet gevonden.** for description of these theories) using empirical data collected at Dutch motorways. ▪ Theory refinement (or even development of new theories), based on validation outcomes. ▪ Mathematical operationalisation of available theories aimed at model development (e.g. establish dependence on key flow variables).
14. Location of the tail of the queue.	

Observation / theory	Recommendations, perform:
15. Correct description of propagation of wide jams, existing upto several hours	
16. Differences in car-following behaviour for different regimes	

6.4.2 Recommendations regarding development of modelling mechanisms

Judging the results of the literature survey, a number of flow theories have been developed and tested (although not on Dutch data). Some of these theories are likely to be valid in the Dutch case. Once the validity of these theories has been established, consequent modelling structures can be included in the simulation models. In some cases, this requires translating the considered theory into an operation traffic flow or driver behaviour model. In other cases however, readily applicable modelling mechanisms are available.

Table 6 Conclusions and recommendations-11 Recommendations from model development.

Observation / theory	Recommendations, perform:
1. Traffic behaviour is non-smooth and random due to inter- and intra driver differences	▪ Model development based on mathematical operationalisation of theories to be developed. ▪ Model calibration and validation.
2. Transitions between traffic states are relatively smooth	▪
3. Drivers accept difference in free speed and speed of leading vehicle	▪
4. Drivers tend to stay on the median or middle lane for some time, after the overtaking manoeuvre is completed	▪
5. Drivers on the main road 'provide gaps' to merging traffic	▪ Merge-giveway behavioural models are available, but need to be adapted to the Dutch case, once behavioural theory has been established. ▪ Model calibration and validation.
6. High merging traffic flow-rates, even if congestion on the main road has occurred	▪
7. Driving and traffic behaviour depends on flow regime, network component, and external factors	▪ Multi-regime modelling structures have been developed, but may have to be adapted for the Dutch case. ▪ Models are to be calibrated and validated.

Observation / theory	**Recommendations, perform:**
8. Difference in high prior-congestion flow-rate and queue discharge flow (capacity drop); differences in range of 5-10% upto 50%.	▪ Models describing hysteresis effects and capacity drops have been developed. Some have been applied to Dutch application studies (differences in car-following behaviour), while others have not ('slow to start rules'). Different models are to be established using validate theory of traffic hysteresis. ▪ Models are to be calibrated and validated.
9. Observations of high-prior congestion flows for considerable period (upto 40 minutes)	
10. Traffic hysteresis	
11. Occurrence of *spontaneous transitions* between regimes.	
12. Displacement of queue head 500m to 1000m downstream of discontinuity	▪ Model development based on mathematical operationalisation of theories to be developed. ▪ Model calibration and validation.
13. Observation of different types of congested traffic flow	▪ Available modelling approaches (e.g. non-local behaviour) must be adapted based on results of theory validation, and included in operation models. ▪ Models are to be calibrated and validated.
14. Location of the tail of the queue.	▪ Model development based on mathematical operationalisation of theories to be developed. ▪ Model calibration and validation.
15. Correct description of propagation of wide jams, existing upto several hours	▪
16. Differences in car-following behaviour for different regimes	▪ Available modelling approaches must be adapted based on results of theory validation, and included in operation models. ▪ Models are to be calibrated and validated.
17. Observed differences between user-classes	▪ Available *multiclass macroscopic models* are to be calibrated and validated.

6.5 General recommendations

In this final section, some general conclusions which do not pertain to specific issues observed in traffic flow modelling of congested traffic operations are discussed.

6.5.1 Development of advanced tools for data visualisation and analysis

Notwithstanding the importance of collecting data with sufficient detail for successful empirical data analysis, application of dedicated tools may prove to be of decisive importance in understanding congested traffic flow operations. Besides qualitative data analysis (visual studies of video observations, vehicle trajectories, etc.), good data representation techniques are very important. In this respect, it is important to consider both the time dimension, as well as the spatial dimension, for instance using contour plots, or by including the time in the fundamental diagram.

It is stressed that the availability of correct data processing algorithms is of dominant importance in traffic flow analysis. This is motivated by the results from the literature survey, which indicate that the differences reported in studies are frequently due to *alternative treatments of data*. Moreover, without a clear definition of traffic equilibrium and a traffic theory to guide the proper processing of raw data, many previous studies lumped together every bit of data collected, leading to two serious problems:

1. Data contain inherently different traffic phenomena and regimes.
2. Aggregation of data is performed over longer time intervals than the acceleration / deceleration 'wave-length', thereby adding considerable distortion to obtained flow relationships.

As a consequence, the need exists to develop dedicated tools for traffic data analysis. The main objective of these tools would be to separate data collected when traffic conditions were different (e.g. free-flow, platoon-driving, and standing still inside a queue). Moreover, in averaging traffic data, care should be taken that structural changes in traffic conditions are not rejected due to large data averaging periods. Various approaches can be considered, among which the following are likely to be very successful:

- Low-pass frequency filters and wavelet analysis techiques (first steps have been taken by Hoogedoorn (1999), and AVV (1999));
- Model-based filter approaches. Note that these Kalman-like filtering approaches also provide insights into the explanatory ability of the models used (see Hoogendoorn (2000));
- Combination of flow theories, models, and data analysis techniques (e.g. consider traffic flow as a mixture of traffic driving in different states (free-flow, platooning, and standing in the queue); consideration of information origin by shockwave analysis.
- Application of fuzzy techniques for data clustering.

Another issue is that it is still unclear to what extent congestion and its properties are predictable (e.g. time to breakdown, queue head location). The use of advanced statistical methods (neural networks, data-mining) may improve insights into this predictability.

6.5.2 Development of comprehensive calibration methodology

During the study, it has been concluded that models are seldom exhaustively calibrated and validated for the considered application, or note at all. This holds especially for situations in which congestion plays an important role.

In this respect, accurate model calibration requires the detail in the data should be on par with the detail level of the considered model (see chapter 5). In illustration, the lack of sufficiently detailed data has frequently resulted in microscopic simulation models being calibrated using macroscopic or mesoscopic traffic data. Due to the considerable degrees of freedom in microscopic models, these data are generally unsuitable for microscopic models. Moreover, behavioural rules that are not calibrated and validated should be considered in isolation as much as possible, especially in case of complex microscopic models.

As it turns out, calibration and validation of models may be of key importance in efficient and correct model application studies. This is why we recommend to develop an unambiguous, clear, structured, and most of all, operational calibration methodology (-standard), which would ideally include the following, interdependent stages:

1. Identification of model application area.
2. Determination of assessment objectives, given model application area, divided into different assessment categories (technical assessment, user acceptance).
3. Determination of performance indicators, based on assessment objectives: definition of the indicator, concept of calculation, nature of the indicator (which effect does it aim to quantify), scope of the indicator (cross-section, link, network), and data collection needs.
4. Description of general tools for calibration or validation, e.g. sensitivity analysis, step-by-step calibration, t-tests, analysis of variance (ANOVA).
5. Overall definition of success;
6. Experimental set-up.

Among the main benefits of such a methodology is the guarantee that model calibration and validation is a structured affair (rather than ad hoc), and the ability for cross-comparison between models. PLATOS provides an excellent platform to develop such a methodology. For model *validation*, first steps have already been taken in the PLATOS project "Leidraad validatiestudies toedelings- en simulatiemodellen".

6.5.3 Model verification and comparison for simple benchmark situations

As a final point of concern, we recommend to test the models' accuracy and reliability using simple benchmark situations, before definitive and general conclusions can be drawn on the problem causes for incorrect description of congested traffic flow operations. We stress that these insights are of dominant importance, if one aims to apply simulations models for decisive evaluation studies.

The proposed research would consist of the following steps, namely:

- Consideration of an useful location (simple bottleneck, or weavings section, where dynamic OD-relations are known, sufficient data collection possibilities, etc.).
- Determination of evaluation methodology by an independent evaluation manager.
- Model calibration and validation.
- Cross-comparison of simulation results by the evaluation manager
- Synthesis of results and recommendations for model adaptation.

6.5.4 Congestion in urban and rural networks

This report has focussed on congestion traffic flow in motorway networks. However, contemporary traffic and congestion management is directed towards the integrated management of both the motorway networks, as well as the underlying urban and rural networks. This is why it is important to consider dynamics of congestion (e.g. blocking-back effects in urban networks, and from metering on-ramps) in non-motorway networks as well. The main question is whether the same problems that are encountered when describing congestion on motorway networks are found when considering non-motorway networks.

7 REFERENCES

[1] Adamo, V., V. Astarita, M. Florian, M. Mahut, J.H. Wu, (1999) *Modelling the spill-back of congestion in link based dynamic network loading models: A simulation model with application* - ISTTT - 1999

[2] Bexelius, S. (1995), *Fluctuations In Traffic Intensities. How Normal Is Normal*? Verkeerskundige Werkdagen, CROW (in Dutch).

[3] Brackstone, M. and M. McDonald (1995). *The microscopic modelling of traffic flow: weaknesses and potential developments* - Traffic and Granular Flow.

[4] Brackstone, M. and, M. McDonald (1998). *Modeling of motorway operations* – Transportation Research Records 1485*.*

[5] Cassidy, M.J., and R.L. Bertini (1999a). *Some traffic features at freeway bottlenecks*. Transportation Research B 33

[6] Cassidy, M.J., and R.L. Bertini (1999b) *Observations at a freeway bottleneck*. Proceedings ISTTT - 1999

[7] Cheu, R.L., W.W. Recker, S.G. Ritchie (1997). *Calibration of INTRAS for simulation of 30s loop detector output*. Transportation Research Records 457

[8] Crutchfield (1986), *Chaos.* Scientific American, December issue, 46-57.

[9] Daganzo, C.F., M.J. Cassidy, R.L. Bertini (1999). *Possible explanations of phase transitions in highway traffic* - Transportation Research A 33 pp. 365-379

[10] DHV (1998). *Gedoseerd naar een oplossing?*

[11] Dijker, T. (1997). *Verkeersafwikkeling bij congestie*. Graduation Thesis, Delft University of Technology.

[12] Dijker, T. and B-J. Boekholt (1998) *Driver behaviour at tapered motorway exits*. Research Report, Delft University of Technology

[13] Ferrari, P. (1989). The Effect of Driver Behaviour on Motorway Reliability. *Transportation Research B* **23**, 139-150.

[14] Forbes and Hall (1990). *The Applicability of Catastrophe Theory in modelling Freeway Traffic Operation*. Transportation Research 24A, 335-344.

[15] G.F. Newell (1999) *Flows Upstream of a Highway Bottleneck*. Proceedings ISTTT - 1999.

[16] Gilchrist and Hall (1989), *Three-dimensional relationships among traffic flow variables*. Transportation Research Board Annul Meeting.

[17] Gleich, J. (1987), *Chaos: making a new science*. Viking, New York.

[18] Helbing, D. (1996), *Traffic modelling by means of physical concepts*, Proceedings of the workshop on Traffic and Granular flow, Forschungszentrum Jülich, World Scientific Publishing Co. Pte. Ltd., Singapore.

[19] Helbing, D., A. Hennecke, and M. Treiber (1998b) *Phase Diagram in the Presence of Inhomogeneities*. Physical Review Letters.

[20] Helbing, D., A. Hennecke, V. Shvetsov, and M. Treiber (1998a). *MASTER: Macroscopic Traffic Simulation based on a Gas-Kinetic, Non-Local Traffic Model*. Institute of Theoretical Physics, University of Stuttgart.

[21] Helbing, D., and B.A. Huberman (1999) *Coherent Moving States in Highway Traffic*.. Nature (in print).

[22] Hertel (1994).

[23] Hoogendoorn and Bovy (2000a), The Ill-Predictability Of Road Traffic Congestion

[24] Hoogendoorn S.P., Bovy P.H.L. (1999) *Macroscopic multiple user-class traffic flow modelling: a multilane generalisation using gas-kinetic theory*, proceedings ISTTT 1999.

[25] Hoogendoorn, S.P., and P.H.L. Bovy (2000c). Non-Local Multiclass Gas-Kinetic Modeling of Multilane Traffic Flow (2000). *Submitted to Networks and Spatial Theory (invited contribution).*

[26] Hoogendoorn, S.P., and P.H.L. Bovy. (2000b). Modelling Multiple User-Class Traffic Flow. *Transportation Research B* (34)2, 123-146.

[27] Hoogendoorn, S.P., G. Copinga, and U. Kaymak (1998). *Perspectives of Fuzzy Logic in Traffic Engineering*. Research Report Dutch Ministry of Transport, Public Works and Water Management, Rotterdam, The Netherlands.

[28] Hoogendoorn, S.P., G. Copinga, and U. Kaymak (1998). *Perspectives of Fuzzy Logic in Traffic Engineering*. Research Report Dutch Ministry of Transport, Public Works and Water Management, Rotterdam, The Netherlands.

[29] Hoogendoorn, S.P., and T.P. Alkim (1999). *Expert views on traffic flow operations during congestion*. Research Report on behalf of the Traffic Research Centre of the Dutch Ministry of Transport.

[30] Kerner, B.S. (1999). Theory of Congested Traffic Flow: Self-Organization without Bottlenecks. *Proceedings of the 14th International Symposium of Transportation and Traffic Theory*, 147-172.

[31] Kerner, B.S., Konhäuser, P., and Schilke, M. (1996), A new approach to problems of traffic flow theory, *Proceedings of the 13th International Symposium of Transportation and Traffic Theory,* INRETS, Lyon, 119-145.

[32] Kerner, B.S., Konhäuser, P., and Schilke, M. (1996), *A new approach to problems of traffic flow theory*, Proceedings of the 13th International Symposium of Transportation and Traffic Theory, INRETS, Lyon, 119-145.

[33] Kita, H., and K. Fukuyama (1999). *A merging-giveway behaviour model considering interactions at expressway on-ramps* - Proceedings ISTTT - 1999

[34] Krauss, S., K. Nagel, and P. Wagner (1999) *The mechanism of flow breakdown in traffic flow models*. Proceedings ISTTT 1999.

[35] Leutzbach, W. (1988), *An introduction to the theory of traffic flow*, Springer-Verlag, Berlin.

[36] Lighthill, M.H., and G.B. Whitham (1955). 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.

[37] Lyrintzis, A.S. , G. Liu, and P.G. Michalopoulos (1997) *Development and comparative evaluation of high-order traffic flow models*, Transportation Research Record 1457

[38] Minderhoud, M. (1999) *Supported Driving: Impacts on Motorway Traffic Flow* - TRAIL Thesis series T4/99.

[39] Navin (1986), *Traffic Congestion Catastrophes*. Transportation Planning and Technology 11

[40] Newell, G.F. (1965). *Instability in Dense Highway Traffic, a Review.* In J. Almond, ed. Proceedings of the Second International Symposium on the Theory of Traffic Flow, 73-83.

[41] Pozybill, M. (1998). *Ist Verkehr Chaotisch?* Straßenverkehrtechnik 10/98, 538-545

[42] Smilowitz, K.R., and C.F. Daganzo (1999). *Predictability of time-dependent traffic backups and other reproducible traits in experimental highway data* - - PATH Working Paper

[43] Van Toorenburg, J. (1983), *De mens in het autosnelwegverkeer*. Verkeerskunde 34, nr. 4 en 5.

[44] Verweij, H.D. (1985). *Congestion Warning and Traffic Flow Operations* (in Dutch). Dutch Ministry of Traffic and Transportation, Delft, The Netherlands.

[45] Zeeman (1972), *Catastrophe Theory*. Addison-Wesley, Reading MA.

[46] Zhang, M, (1999), *A Mathematical Theory of Traffic Hysteresis.* Transportation Research B, Vol. 33(1), 1-23.

[47] Zhang, M. (1998), *Analyses of Non-Equilibrium Traffic Flow*. Transportation Research B, Vol. 32(7), 485-498.