Het verkeerskundig
laboratorium
voor studenten

ITS EDU LAB

Gaining new insights regarding traffic congestion, by explicitly considering the variability in traffic

Onno Miete

January 2011

Rijkswaterstaat

TU Delft

Ministerie van Verkeer en Waterstaat

Rijkswaterstaat

Gaining new insights regarding traffic congestion, by explicitly considering the variability in traffic

Onno Miete

January 2011

MSc Thesis Civil Engineering
Delft University of Technology

Colophon

Published by: ITS Edulab, Delft
ITS Edulab is a cooperation between the Rijkswaterstaat centre for Transport and Navigation and the Delft University of Technology

Information: Henk Taale
Email: henk.taale@rws.nl

Author: O.M. Miete
Delft University of Technology
Master student Transport & Planning and Hydraulic Engineering
Student number 1213121

Graduation committee:
Prof.dr.ir. S.P. Hoogendoorn [1] (chair)
Prof.drs.ir. J.K. Vrijling [2]
Dr.ir. P.H.A.J.M. van Gelder [2]
Dr.ir. J.W.C. van Lint [1]
Dr.ir. H. Taale [3]
Ir. P.B.L. Wiggenraad [1]

Date: January 6, 2011

Status: Final

---

[1] Delft University of Technology, Faculty of Civil Engineering and Geosciences, Department of Transport & Planning
[2] Delft University of Technology, Faculty of Civil Engineering and Geosciences, Department of Hydraulic Engineering
[3] Rijkswaterstaat, Centre for Transport and Navigation

Preface

This thesis is the final result of my graduation project. It marks the end of my studies in Civil Engineering at the Delft University of Technology. In my master's, I have taken two different specialization tracks: Transport & Planning and Hydraulic Engineering. One of the elements in the Hydraulic Engineering track which aroused my special interest is the probabilistic design philosophy. Because of this, I decided to devote my graduation project to an application of the probabilistic way of looking at systems to the (motorway) traffic system, focusing on the daily traffic congestion in this system (one of my fields of interest in the Transport & Planning discipline).

I am grateful to have had the opportunity to perform my research at the ITS Edulab, a cooperation between the Rijkswaterstaat Centre for Transport and Navigation and the Delft University of Technology. The Centre of Transport and Navigation offered a very nice workplace, and good facilities for performing the research. Without access to its digital library, an important part of my literature search would not have been possible. At the department of Road Traffic Management I always felt very welcome, for which I would like to thank all my colleagues over there. I would also like to thank my fellow students in the ITS Edulab, for the enjoyable time and good working atmosphere.

Many thanks go to the members of my graduation committee, for reviewing my report and providing valuable advices. Finally, I would also like to express my gratitude to Frank Zuurbier and Chris van Hinsbergen, for providing the opportunity to use their dynamic traffic simulator 'JDSMART' in my model.

Summary

In the past decades, traffic congestion on the Dutch motorway network has developed into a serious problem, causing large costs to society. In this thesis (alleviating) traffic congestion is considered from a probabilistic perspective, meaning that the variability in traffic is explicitly taken into account. Traditionally, in evaluations of the effectiveness of proposed congestion relief measures this variability is taken into account only in a limited or simplified way, or even not at all. Often simply a kind of 'representative' situation is calculated, possibly supplemented with some qualitative considerations or scenario-based analyses regarding the effects on the robustness of the traffic system.

The main objective of this research project was to reveal what kind of new insights can be obtained if we actually *do* explicitly/systematically take into account the inherent variable nature of daily motorway congestion. Two different types of such additional (or revised) insights are distinguished:

- Insights into the relative importance of the various primary sources of traffic congestion.
- Insights into the effectiveness of specific traffic measures proposed to alleviate congestion.

Basically, the mechanism behind traffic congestion can be described as a process of interaction between the traffic demand and supply on the road sections of the network. Both this traffic demand and supply show a significant level of temporal variability, which makes the resulting traffic conditions variable as well. There is a large variety of sources of variability in demand and supply. These include:

- systematic travel behavioral variations as a function of time
  (i.e. time of the day, day of the week and month of the year)
- vacation periods
- special days (like public holidays)
- weather variations
- luminance variations
- road works
- incidents
- events
- traffic control
- variations in vehicle population
- variations in driver population
- intrinsic randomness in people's personal travel choices
- intrinsic randomness in human driving behavior
  (i.e. variations both between and 'within' drivers, which cannot be explained by external influences)

While for *given* demand and supply values for a road section the mechanism behind congestion can be described as a (simple) local demand-supply process, reality is more complicated. This is due to the fact that these demand and supply values actually depend on the traffic conditions on *other* sections of the road network. These spatiotemporal dependencies between the traffic conditions on the different sections of a network are due to a number of 'network effects' of traffic congestion. These mechanisms can only be accounted for by considering the traffic flow dynamics at the network level.

If the variability in traffic is to be explicitly taken into account in evaluations of the performance of the traffic system, it must first be decided which criterion is to be used then for this performance. There is no sharp 'failure boundary' with respect to the amount of traffic congestion (a threshold above which the motorway system can be considered to 'fail', and below which the system can be considered to 'function'). In the end, it is all about the costs that traffic congestion causes to society. Traffic congestion causes costs to society in various ways. Considering these different types of costs, it turns out that they cannot be expressed in one single indicator, since they are related to the traffic conditions in different ways. Because of this, rather a *set* of indicators needs to be considered.

One of the costs of traffic congestion is related to the travel time uncertainty that this congestion creates. It is difficult to find a proper indicator for this. It is clear that the uncertainty costs are reflected in the travel time distribution, but not exactly in what way. The indicators found in practice and international literature all have their limitations, because each of them represents only part of the information contained in the travel time distribution. This problem is dealt with by including multiple statistics of the travel time distribution in the set of selected indicators. Furthermore, an indicator representing the travel time *instability* was added, in view of the fact that this factor plays an important role in the resulting travel time uncertainty as well. This indicator expresses to which extent the traffic conditions experienced by travelers might deviate from the instantaneous traffic conditions at the moment of departure (as disseminated by traffic information).

It was considered that the (potential) gain of new insights could best be shown using a model-based approach. The most obvious choice is then to use a model with a macroscopic traffic simulator as computational core. In practice and in literature, several models specifically designed for addressing the variability in traffic conditions can be found. It was concluded however that none of the considered models was completely adequate for the tasks at hand. Therefore, in this project a new model was developed. The main principle of this model is that a large number of traffic simulations are performed for varying model inputs, reflecting the variabilities in the traffic demand and supply characteristics. Subsequently, the desired performance indicators can be computed from the combined set of all simulation results.

In the model both the demand and supply values are varied per 5-minute interval of the day. Here the demands are varied at the level of origin-destination relations, and the supply characteristics at the level of network cells. The stochastic generation of the demand and supply values proceeds in two steps:

- First, random realizations of the different influencing factors (also indicated as 'sources of variability') are generated. For this the Monte Carlo method is used. (A scenario-based approach is considered inappropriate.) In the Monte Carlo simulation use is made of data on the probabilities/frequencies of occurrence of the different possible conditions. Important interdependencies between the different sources of variability are taken into account by using conditional probability specifications.
- Subsequently, the stochastically generated circumstances are translated into effects on the traffic demand and/or supply characteristics, using tables in which these effects are specified in terms of correction factors. By applying these correction factors on the representative values of the demand and supply characteristics, the stochastic realizations of these demand and supply characteristics are found.

The stochastically generated demand and supply conditions are passed on to the computational core of the model, which then simulates the traffic conditions that would arise from these conditions. This computational core consists of the existing dynamic macroscopic traffic simulator JDSMART (a first order cell transmission model), which was supplemented with some additional functionality to make it suitable for its role in the developed model.

Special care has been taken to make sure that the same pseudo-random numbers are used in different model runs, in order to improve the comparability of the outputs of these runs. This means that the different parts of the model have been programmed in such a way that they always generate the same number of pseudo-random numbers, irrespective of certain model settings.

In order to explore the (potential) new insights obtained by explicitly considering the variability, the developed model has been applied to a reasonably sized real-life motorway network. It should be noted, however, that incidents and road works were omitted from the analyses, because of the inability of the model to deal with these in a sufficiently valid way.

From the results obtained with the model, it is clear that the 'representative' calculation does not give a good impression of the performance of the traffic system. This is not only due to the obvious fact that the (day-to-day) uncertainty aspect of this performance is disregarded (due to the neglect of the day-to-day variability in the traffic conditions). Also, the representative calculation turns out to underestimate the traffic congestion in certain respects. That is, the traffic congestion calculated for the 'representative' situation (i.e. the situation in which all demand and supply characteristics are at their 'representative' level, which for example could be the mean or median value) is not so 'representative' itself. This is related to the predominantly negative influence of the (neglected) variability. This predominantly negative influence arises from:

- the purely negative nature of some of the sources of variability (such as incidents or bad weather events)
- the non-linearity in the traffic system (i.e. the fact that the congestion level is a non-linear function of the difference between demand and supply, causing that the detriments of 'negative occurrences' are often larger than the benefits of 'positive occurrences').

It has been demonstrated that new insights into the relative importance of the different primary congestion sources can be obtained by 'deactivating' them in the model. Although the relative influences of only a limited number of such sources have been compared in this project (by way of illustration), and only one specific test network was considered, it can be concluded that the capacity variations due to the intrinsic randomness in human driving behavior play a central role in peak period-related traffic congestion. The demand variation over the months of the year plays an important role as well, while the ambient conditions (i.e. weather and daylight/darkness), events and the intrinsic randomness in travel behavior seem to have a much smaller (or even negligible) influence. Ignoring the influences of incidents and road works, events seem to be the most important source of weekend day traffic congestion.

This kind of information may be valuable in the following ways:

- It might yield important insights into how traffic congestion can be remedied most effectively.
- Insofar as certain sources of variability are found to be negligible compared to others (as a general rule), these can be omitted in future model evaluations (both in research studies and in practical applications).

By considering the example of a rush-hour lane, the research has shown that new insights can be obtained into the effectiveness of specific measures that are proposed to alleviate traffic congestion. It turned out that the 'traditional' way of evaluating the effectiveness of a measure may actually result in a significant underestimation of the benefits of this measure. This is due to the facts that:

- A 'representative' calculation underestimates the traffic congestion in certain respects (as noted above), and thereby underestimates the beneficial effects of proposed measures (aimed at alleviating this congestion) as well.
- In an evaluation according to the traditional approach potential benefits of a considered measure may remain unnoticed due to nonlinearities and trend breaks in the behavior of the traffic system. This applies particularly to (the prevention of) spillback of congestion to other network elements. If this spillback occurs only in part of the occasions (say less than 50%), it will not be included in the representative analyses. Consequently, the benefits achieved on these other network elements will not be reflected in the evaluation results.

- In an evaluation according to the traditional approach, no information is obtained on the improvements in travel time uncertainty (due to the fact that the day-to-day variability in the traffic conditions is not considered).

The precise nature and extent of the additional/revised insights into the effectiveness of a measure will be highly context and measure specific. Of course, these new insights are not necessarily all positive in nature. Some more negative aspects of a measure could be brought to light as well.

The above implies that in practice more systematic attention should be given to the variability in traffic, when evaluating the effectiveness of measures that are proposed to alleviate congestion. Because of the complexity involved (especially in case of heavily loaded networks in highly urbanized areas), this would have to be done by using a model in which the different sources of variability are explicitly accounted for, such as the model developed in this project.

This model was developed solely for the research task considered in this thesis, however, and thus not directly for practical application in the evaluation of concrete projects. In such practical evaluations, the model can only be used in a qualitative way, to find out whether certain effects (i.e. benefits or detriments) of a measure may be overlooked (or considerably underestimated) in the evaluation according to the traditional approach. The model cannot be used to find the detailed quantitative values of these effects, due to the facts that:

- Its quantitative outputs are affected by a number of deficiencies, related to some modeling issues that require substantial further research.
- It cannot be properly calibrated to the local situation.

Another issue relevant to the practical applicability of the model is its computation time. Currently, the computation time required for one model run is in the order of days or weeks, which is related to the large number of simulations that is to be performed. For practical applications, this computation time would have to be reduced. Such a reduction could be achieved in three different ways:

- Using faster computers (or multiple computers in parallel).
- Reducing the required number of simulation runs, by implementing a more efficient sampling technique (such as Latin Hypercube Sampling or Importance Sampling).
- Increasing the speed of the developed model (i.e. reducing the amount of computation time required per individual simulation run), involving a tradeoff between computation time and model accuracy.

Table of contents

1. Introduction

1.1 The traffic congestion problem and its relation with probabilistic design

In the past decades, traffic congestion on the Dutch main road network has developed into a serious problem. In 2009, a total amount of 62 million hours[4] was lost due to sub optimal traffic operations (DVS, 2010). This corresponds to 8.6% of the total travel time spent on the main road network (723 million hours). In the Randstad, during the morning peak a trip of 30 kilometers on the highway network takes about 30 minutes on average, while the free flow travel time[5] is only about 18 minutes (AVV, 2004).

The main cause of this problematic situation seems obvious. During the past decades, the growth in the amount of traffic using the main road network substantially exceeded the rate of extension of the network capacity. As a result, the relative network loading got more and more heavy. The result is a situation in which during every peak period many traffic jams are formed.

These traffic jams do not only occur at structural bottlenecks in the network, and not only during the peak periods either. Also at other locations and in other periods of the day traffic jams are formed from time to time. Part of these traffic jams is (partly) caused by incidents, road works, events and bad weather conditions. As a result of its heavier loading, the network has become more vulnerable to this kind of circumstances.

The effects of a local overloading of the available infrastructure capacity do not limit themselves to the location concerned. After all, the resulting traffic jam has a certain dimension. Furthermore, this traffic jam can propagate itself over the network. This phenomenon occurs when the head of the queue is dissolving, while the tail of it is still propagating upstream. These characteristics result in traffic jams blocking other traffic streams (consisting of vehicles that do not need to pass the bottleneck location itself). As a consequence, disruptions can rapidly spread out over a substantial part of the network. The high network loading results in the network being more vulnerable to this phenomenon.

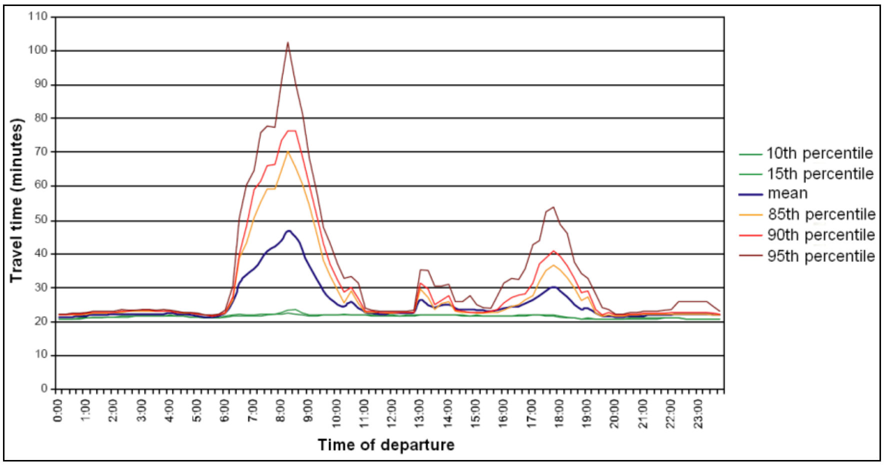
Beside the fact that much time is lost due to delays, an important characteristic of the resulting traffic operations is that travel times are ill

[4] Expressed in lost vehicle hours: the sum of the delay of all vehicles, relative to a norm speed of 100 km/h.

[5] Defined here as the travel time in case of a travel speed of 100 km/h.

predictable. Travel times do not only vary as a function of the time of the day, but show a considerable day-to-day variation as well. This can be illustrated with Figure 1.1, showing the travel time on a route from Amersfoort to Amsterdam, as a function of the time of the day. Comparison of the values of the various percentiles of the travel time provides an indication of the day-to-day variation in the travel time. On this route, the morning peak shows longer travel times than the evening peak. At the busiest moment during the morning peak, the travel time is on average more than twice as long as during free flow conditions (48 versus 22 minutes). However, on 15% of the weekdays, the travel time during the morning peak is hardly longer than the free flow travel time, while on another 15% of the weekdays a travel time of more than 3 times the free flow travel time is reached.

Figure 1.1: Travel times Amersfoort-Amsterdam, weekdays 2002
(Source: AVV, 2004b)



The traffic congestion on the main road network causes large costs to society. For 2008, the total costs were estimated at 2.8 to 3.6 billion euros (KiM, 2009). This corresponds to roughly 0.5% of the Dutch gross national product. In the period 2000–2008, these costs increased by approximately 78%. Because of these large costs, the Dutch government tries to reduce the amount of traffic congestion.

In this graduation project, alleviation of traffic congestion will be considered from a probabilistic design perspective[6]. The essence of probabilistic design is that variability/uncertainty is explicitly taken into account in the design. In deterministic design, this is not the case. In this 'traditional' way of designing, the design is based on a certain 'representative' situation. By neglecting the variabilities/uncertainties, it may well be that the designed product or system finally does not perform as desired, for example in terms of failing (too soon / too often) to perform its intended function. Often, certain (not well-founded) safety factor(s) are applied in the design in order to prevent this. However, one still does not know then how good the product or system actually will be. This is likely to result in an (economically) sub-optimal situation.

[6] Here 'design' should not be taken too literally. This thesis does not deal with the actual design of traffic systems. It is rather the probabilistic way of looking at a system which is applied here to the traffic system, in the context of analyzing (ways to alleviate) traffic congestion.

Therefore, it is better to follow the probabilistic design philosophy, and consider all variabilities/uncertainties explicitly. This way, one is able to optimize the design of the product or system, or at least able to make sure that it will satisfy certain norms. Furthermore, designing the product or system in a probabilistic manner may provide the designer with additional insights into the relative importance of different mechanisms. In combination with information regarding the costs of certain measures, these insights may be used for the identification of the most cost-effective measures to improve the design. Of course, the results of the detailed probabilistic analyses finally may be translated into well-founded safety factors to be incorporated in norms or guidelines, in order to ease the design process.

Probabilistic design is used in various disciplines, like systems engineering, product design, structural engineering and hydraulic engineering. Cleary, it is also applicable to interventions in the traffic system (in this case aimed at alleviating traffic congestion). After all, traffic congestion is a phenomenon characterized by considerable variability and uncertainty. There are several sources of variability and uncertainty. A substantial part of the within-day and day-to-day variation can be explained by fixed social activity patterns. One of the most important uncertainties is the inherent uncertainty stemming from the fact that the traffic flow operations are the result of the human behavior of a heterogeneous collection of individuals (behavior with respect to travel decisions and way of driving). This behavior is predictable only to a limited extent. Another inherent uncertainty is the variation in external conditions, like varying weather circumstances and the occurrence of disasters. In an actual design/evaluation project, the limited data availability is often another important source of uncertainty (an epistemic uncertainty component). When predicting (and assessing) the situation in some future year, much more uncertainty is added. Prediction of future mobility levels and patterns is very difficult. The infrastructure supply available in some future year is uncertain as well. Another source of uncertainty is model uncertainty, related to the uncertainty in the traffic models used to forecast traffic operations (due to incompleteness/simplification).

The relative importance of the various sources of uncertainty is dependent on the type of problem. When dealing with traffic congestion on the main road network, two levels of action can be discerned: the operational level and the strategic level[7]. The operational level refers to the selection of control actions, considering the actual traffic situation and its expected development over the next few minutes/hours, using the available control facilities (like dynamic traffic management measures). At this level, there is of course no uncertainty related to the mobility level and patterns for some future year. The inherent uncertainty in traffic operations on the other hand is an important uncertainty component to be considered, as may be the

[7] Often, *three* levels of action are discerned: the strategic level, the *tactical* level, and the operational level. In this document, the tactical level and strategic level are combined into one level, for the sake of convenience referred to as 'strategic level'.

model uncertainty. The strategic level refers to actions on the longer term, like the realization of new infrastructure or dynamic traffic management facilities, or the implementation of a road pricing system. For the evaluation of this kind of actions, uncertainty related to the use of prognoses for the mobility in some future year is of course very important.

At the operational level, uncertainties are currently not taken into account in calculating the optimal control for traffic networks. Recently some research effort has been devoted to find a new methodology, taking into account the uncertainty in the system dynamics. This is possible by using a stochastic prediction model instead of a deterministic model (as used in traditional optimal control theory) (Hoogendoorn et al, 2008). Such a method enables the consideration of not only the average system performance, but any other statistic of the stochastic system performance as well. This way, when selecting control actions, their robustness can be taken into account too.

At the strategic level, the extent to which uncertainties are taken into account varies. Uncertainty related to the use of prognoses for the mobility and infrastructure supply in some future year is sometimes taken into account by considering a number of scenarios (based on forecasts produced by for example the national planning institute). In other cases, a simple sensitivity assessment is performed. Also the consideration of traffic model uncertainties and uncertainties related to limited data availability often remains limited to carrying out certain sensitivity tests at the most. In 2005 De Jong et al. developed a new methodology to estimate the amount of uncertainty in traffic forecasts for new infrastructure. In this method, model input uncertainties (related to the use of prognoses), uncertainties due to lack of data, and model uncertainties were included. Not all sources of these types of uncertainties were included though. Examples of uncertainties that were not included are the uncertainty in the traffic assignment procedures (procedures assigning the calculated traffic flows to the road network), and the uncertainty in the regional distribution of prognosticated input variables (De Jong et al., 2005). The main purpose of this study was to obtain an indication of the order of magnitude of the uncertainty. For practical use, the methodology is considered to be too complex. That is why usually still only scenario evaluation and sensitivity assessment are used.

When planning new roads or dynamic traffic management facilities, the inherent variability and uncertainty in traffic operations are usually taken into account only in a limited or simplified way. Often a kind of ‘representative’ or ‘average’ situation is calculated. (See for instance Mehran & Nakamura (2009): ‘Evaluation of the efficiency of congestion relief schemes on expressways has generally been based on average travel time analysis’). In case of planning a new road, the effects on the robustness of the traffic system are assessed by qualitative considerations or by an evaluation of the performance in some fictitious disturbance scenarios. In the process of planning new dynamic traffic management facilities such considerations of disturbance scenarios are applied as well.

In current traffic policy and research, much attention is paid to network reliability and robustness. Of course, inherent uncertainties in the variables/processes governing the traffic conditions play a central role in these concepts.

1.2 Main research objective

From the foregoing it can be concluded that:

- traffic congestion has important probabilistic properties, but that
- nevertheless this probabilistic nature often is not (completely) taken into account when dealing with traffic congestion.

Rather than exploring how to deal with all types of variability's and uncertainties, this research project primarily focuses on a subset of these: the *inherent variability in traffic operations.* With this inherent variability, the within-day and (more important) day-to-day variations in the traffic conditions are meant. These variations are caused by known patterns in human activities for one part, and by inherent uncertainty in traffic operations (resulting from the inherent uncertainty in the variables and processes governing these traffic operations) for the other part.

The main objective of the research project can be defined as:

Main research objective

To reveal what kind of additional or revised insights can be obtained from evaluations of the traffic system's performance (in the context of considering taking strategic[5] measures to alleviate congestion) when the inherent variability in daily motorway congestion is explicitly taken into account. (As opposed to the insights obtained by evaluations according to the more 'traditional' approach, in which only a kind of 'representative' situation is evaluated.)

On purpose, this description of the research objective is specified in modest terms (i.e. using '*to reveal what kind of*' and '*can be obtained*', rather than '*to establish which*' and '*are obtained*'). This is because of the fact that it will not be possible to make *generally valid* quantitative inferences, due to the fact that the exact additional/revised insights will typically be case-specific. That is, they will significantly depend on the spatial configuration (i.e. network layout and spatial traffic demand pattern) and any possible traffic measures considered. Obviously, not all of the possible spatial configurations can be considered here (simply because of the fact that their number is infinitely large). Instead, quantitative observations are made for just one such configuration. Furthermore, in this thesis only one specific measure is considered, by way of example.

As mentioned above, the research project focuses on the inherent variability in traffic operations (partially caused by the uncertainty that is inherent to these traffic operations). All other types of uncertainty identified in the previous section can be neglected here:

- Because of the fact that it is not intended to come up with quantitative conclusions regarding *specific real-life projects*, uncertainty due to a lack of *location-specific* data is not important. The location-specific variables in question can be given any value within the range of values observed in reality.
- Uncertainty related to the use of prognoses (for some future year) is not important either. This type of uncertainty typically only has to be included when dealing with an actual real-life project, in which a situation in a 'design year' needs to be assessed.
- Obviously, quantitative analyses in this project *do* involve model uncertainty (related to incompleteness or simplifications in the traffic model) and uncertainty due to a lack of data. In fact, these uncertainties should be accounted for, by means of a comprehensive sensitivity assessment for example. This would be a very time-consuming and complicated matter, however, because of the very large number of degrees of freedom to be considered, and the difficulties in assessing the model uncertainty. In view of the fact that this project only intends to give an *illustration* of any possible additional/revised insights, and not to come up with firm quantitative inferences, it is considered acceptable to omit such an extensive uncertainty assessment.

The most challenging part of the research project is to find a proper quantification method, enabling to take into account the inherent variability of traffic operations explicitly. This quantification method might also be useful in other contexts (outside the scope of this research project), for instance in actual projects aiming to alleviate traffic congestion in a real-life situation. However, in such situations probably extensive calibration procedures would be needed (related to the lack of location-specific data on the different model parameters), in order to 'fit' the method to the situation at hand. Furthermore, the remaining[8] part of the uncertainty due to a lack of location-specific data would have to be accounted for, as should be the uncertainty related to the use of prognoses.

[8] I.e., after the calibration procedures have been performed.

1.3 Research scope

Daily traffic congestion

This section sets down more formally the scope/boundaries of the research project. As indicated in the main research objective, *daily* traffic congestion is considered. This means that extremely exceptional situations, like disasters, are not accounted for. Note that the research *does* take into account incidents (like traffic accidents and vehicle breakdowns) and road works, as these occur such frequently that they indeed can be considered as everyday disturbances, contributing to daily traffic congestion.

Motorway network

The research project focuses on traffic congestion problems on the *motorway network*, rather than on traffic congestion problems on *lower level networks*. The motorway network is defined here as the network of roads with unidirectional roadways, a design speed of at least 100 km/h and grade separated crossings of traffic flows. This network accounts for more than half of the total yearly amount of kilometers traveled by car (RPB, 2004).

Focusing on congestion problems on the motorway network does not directly imply that all other roads can be left out of account in this project. After all, the secondary road network might provide motorway users with opportunities to get around (exceptional) congestion on their motorway route. In order to account for this, in fact part of the secondary network should be included in the analyses as well. Within the scope of this project, incorporation of this effect of fallback on alternative routes turned out to be unfeasible, however (which will be explained later in this report).

Note that the incorporation of lower level networks would complicate the analyses in several ways:

- On the lower level networks, the frequencies of occurrence and/or effects of many of the sources of variability are clearly different from those on the motorway network. This means that much additional research would be needed into these aspects.
- The traffic propagation over lower level networks is mainly governed by the traffic operations at intersections, which considerably complicates the traffic flow modeling. There are many different types of intersections. In case of *uncontrolled* intersections and intersections with *traffic adaptive control*, crossing traffic flows may affect each other's capacities. At the (grade separated) crossings in motorway networks this is obviously not the case, unless congestion spills back from one road to the other.

- On most of the lower level roads, traffic is able to make U-turns when getting stuck in a heavy traffic jam. This is typically not taken into account in traffic simulation models.
- Especially on urban networks, the finer mesh of the road network offers much more opportunities for traffic to get around heavily congested (or even completely blocked) road segments (for example in incident situations). This means that if the lower level networks would be included in the analyses, the incorporation of the route choice effect of (non-recurrent) congestion would be even more important for obtaining realistic results. This turned out to be unfeasible, however, as noted above.
- If this route choice effect is not accounted for, a finer meshed network will be prone to the occurrence of gridlocks (i.e. situations with a 'ring-shaped' traffic jam, in which car drivers are basically 'waiting for themselves'), which would cause problems in the traffic simulations.

Focus on the Dutch situation

Throughout the whole research project, focus has been on the situation found in the Netherlands. This is an important limitation, because in other countries certain influencing circumstances may well be different from those found in the Netherlands.

For some influencing factors little information is available that specifically relates to the situation in the Netherlands. Out of necessity in such cases use had to be made of research results obtained in other countries, keeping in mind that these might not directly be equally valid for the Dutch situation.

1.4 Research questions

The research necessary to achieve the main research objective has been divided in a number of steps, related to certain research questions, each addressing one specific part of the problem.

In order to be able to evaluate the performance of the traffic system with regard to traffic congestion, first of all a clear view is needed on when the traffic system actually is considered to perform well and when it is not, because this is not something obvious. In fact, the consideration of the inherent variability adds an extra dimension to this. Therefore, the first research question that was dealt with reads:

[Research question 1] *Which criterion(s) should be used to evaluate the performance of the traffic system with regard to traffic congestion (taking into account its variable nature)?*

In order to answer this question, the following two sub questions were considered:

[Sub question 1.1] *What are the societal costs of traffic congestion, and how do these relate to the characteristics of this congestion?*

[Sub question 1.2] *Which indicator(s) to use for traffic congestion on the motorway network (and which limits to set on these)?*

Next, a method had to be found to evaluate the traffic system's performance with respect to the criterion(s) set on traffic congestion. This problem was split up into two parts. First of all, the (probabilistic) mechanisms governing traffic congestion were examined:

[Research question 2] *Which (probabilistic) mechanisms are governing traffic congestion?*

This research question was addressed by considering the following sub questions:

[Sub question 2.1] *What is the basic process governing traffic congestion at a road section?*

[Sub question 2.2] *Which disturbing influences are involved, and how can these be characterized in terms of frequency/probability of occurrence and effects?*

[Sub question 2.3] *Which dependencies between these disturbing influences are involved?*

[Sub question 2.4] *What is the role of network effects?*

When these questions had been addressed, attention was focused on finding an actual quantification method:

[Research question 3] *How to quantify the traffic system's performance with respect to traffic congestion, taking into account the variability in traffic operations?*

This research question was dealt with by addressing the following sub questions:

[Sub question 3.1] *Which requirements does a quantification method need to meet? (taking into account the findings obtained in relation to research questions 1 and 2)*
[Sub question 3.2] *What type of quantification methodology is most appropriate?*
[Sub question 3.3] *Which methods taking into account the variability in traffic operations are currently available, and do these meet the requirements?*
[Sub question 3.4] *Which method to use?*

Regarding sub question 3.4 it has to be mentioned that not necessarily one of the methods considered in sub question 3.3 had to be selected. If none of the methods would meet the requirements to a satisfactory degree (which actually turned out to be the case), a new method would have to be developed.

After all the research questions above had been dealt with (i.e. performance criterion(s) had been defined, and a quantification method had been developed), some quantitative analyses related to the main research objective were performed:

[Research question 4] *What kind of additional or revised insights can be obtained when traffic congestion is approached in a way in which its inherent variable nature is explicitly taken into account, as compared with the insights obtained by a 'traditional approach', in which only a kind of 'representative' situation is evaluated?*[9]

This research question was addressed in two steps (corresponding to the two sub questions below). First of all, it was considered whether the new evaluation approach can provide insights into the relative importance of different primary sources of traffic congestion (like events, special weather conditions, seasonal variations in the traffic

[9] As indicated before, it was not striven for to come up with generally valid quantitative inferences here, because of the strong dependency on the specific case at hand. Rather, an illustration of the additional or revised insights was aimed at.

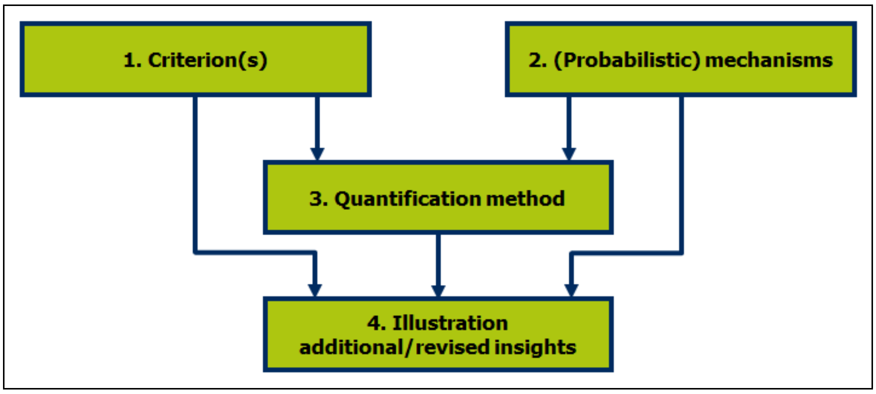
demand level, the intrinsic randomness in capacities, etc.). This 'relative importance' is to be understood as the relative contribution to the traffic system's performance (in terms of the criterion(s) defined in the first part of this research project). Differences in the relative importance of different contributing factors may provide some new insights into the relative effectiveness (or *in*effectiveness) of various categories of measures (aimed at alleviating traffic congestion). This kind of information typically can only be obtained when explicitly considering the inherent variability in the traffic congestion.

Secondly, it was considered what kind of additional or revised insights can be obtained when using the new approach for the evaluation of the effectiveness of some proposed traffic measure. For this, the example of a rush-hour lane was considered.

[Sub question 4.1]	*Can the new evaluation approach provide us with insights into the relative contributions of different primary sources of traffic congestion to the traffic system's performance (in terms of the criterion(s) specified earlier)?*
[Sub question 4.2]	*What kind of additional or revised insights into the effectiveness of proposed measures aimed at alleviating traffic congestion can be gained when explicitly taking into account the variable nature of this congestion?*

In the figure below, the main structure of the research project is summarized.

Figure 1.2: Structure of the research project



1.5 Thesis outline

This report does not one-on-one follow the sequence of the research questions indicated in the previous section. Instead of starting with the issue of selecting appropriate indicators for the level of traffic congestion, it starts with a discussion of the mechanisms governing traffic congestion and its variability (chapter 2). The reason for this is that this discussion might be beneficial for the understanding of the subsequent parts of the thesis. In the concerning chapter, first of all the basic interaction process between traffic demand and supply on a road section is explained (section 2.1). After this, all different sources of variability are discussed (section 2.2). Finally consideration is given to the network effects of traffic congestion (section 2.3).

The next chapter then deals with the selection of appropriate performance indicators for the level of traffic congestion (taking into account the variable nature of this congestion). After a short introduction on this issue (section 3.1), section 3.2 deals with the question which features describing the traffic congestion phenomenon can be identified as being most decisive in bringing about costs to society (which typically are the features to be incorporated in the indicators). Section 3.3 describes which indicators are used in international literature, and which norms were used in the Dutch national traffic policy during the past few decades. In section 3.4 then the final selection of indicators is discussed. Finally a section has been added that discusses the strong relationships that in various empirical studies have been found to exist between the average travel time (or delay) and other indicators based on travel time statistics, and their implications for this research project (section 3.5).

Chapter 4 discusses the selection of a quantification methodology for research questions 4.1 and 4.2. After an introductory section, section 4.2 discusses what type of methodology is most appropriate for illustrating the gain of new insights into the relative importance of different primary sources of traffic congestion (corresponding to research question 4.1). Similarly, section 4.3 then discusses what type of methodology is most appropriate for illustrating the gain of additional or revised insights (if any) into the effectiveness of specific measures (corresponding to research question 4.2). Next, in section 4.4 a list is given of the requirements to be met by a quantification model. Section 4.5 then considers a variety of quantification models that are used in practice or proposed in literature. Here it is also assessed to which extent these models meet the requirements from section 4.4.

Since none of the models was found to be sufficiently adequate for the tasks at hand, a new model was developed. This quantification model is described in chapter 5. The different sections of this chapter successively discuss the general concept of the model (5.1), its approach to the traffic flow modeling (5.2), and its general approach to the modeling of the variations in traffic demand and supply (5.3). A detailed description of the ways in which the various individual sources of variability are modeled is given in section 5.4.

During the development of the model some modeling issues have come to light which require further consideration. These issues are discussed in chapter 6. Since they generally require substantial further research, it was not possible to actually solve them within this project. However, besides explaining the different issues, chapter 6 also tries to suggest some possible strategies to deal with them. This includes a possible strategy for reducing the required number of simulations.

A model is a simplified representation of a part of reality. In order to be able to make sound inferences with such a model, it has to be sufficiently valid for the task at hand. In sections 7.1 - 7.4, the developed model is assessed on three different levels of validity. This is based solely on theoretical considerations. Normally, one would assess the final validity of a model by means of a quantitative validation procedure. In section 7.5 it is argued, however, that the developed model cannot be quantitatively validated in the usual way. Yet, some quantitative considerations are given in this section. These are considerations of a more general nature, relating to the computed congestion levels.

Chapter 8 then discusses the results of the quantitative evaluations with the model. Here it is considered what kind of additional insights are obtained by explicitly taking into account the inherent variable nature of the traffic conditions. After an introductory section, the chapter starts with a description of the network for which the model evaluations have been performed (section 8.2), and a description of some restrictions/simplifications with respect to the indicators considered (section 8.3). After that, section 8.4 discusses the model outputs for the representative situation, which typically are the outputs obtained by a traditional calculation of the traffic conditions in a network. Next, section 8.5 treats the results obtained by a calculation in which the different sources of variability are taken into account, and compares these with the output for the representative situation. Section 8.6 then shows that the relative importance of these different sources can be studied by deactivating them in the model. Section 8.7 considers the effects of a rush-hour lane, as computed with the new model, in which various sources of variability are taken into account, and compares these with the effects that would have been found with a calculation according to the traditional approach (considering a representative situation only). Chapter 8 ends with a section on the practical applicability of the developed model. In this section it is discussed whether/how this model could be used for practical application within the context of real-life evaluations of measures proposed to alleviate traffic congestion.

In chapter 9 finally some conclusions and recommendations are provided, as well as some practical implications of the findings obtained.

2. Mechanisms governing traffic congestion and its variability

In this chapter, the phenomenon of traffic congestion and the causes for its variability are examined in more detail. Section 2.1 describes the basic mechanism governing traffic congestion on a road section. The sources of the variability in this congestion are discussed in section 2.2. An important characteristic of the traffic system is that the traffic conditions on the different road sections may interact with each other. These interaction processes (referred to as 'network effects') are explained in section 2.3.

2.1 Basic mechanism governing traffic congestion on a road section

Traffic flow operations on a road section are governed by the interaction between the traffic demand (the amount of traffic wanting to traverse the section) and the traffic supply (the available capacity). In order to explain this interaction process, an initially empty road section is considered for which the traffic demand gradually rises, starting from zero. At the end of the road section there is a bottleneck: a stretch of road with a lower capacity. Apart from this bottleneck, the road section is homogeneous. There are no connections to other roads.

As long as the traffic demand is smaller than the capacity of the bottleneck, the traffic conditions are referred to as being 'free flow'. The average traffic speed for an (almost) zero traffic volume is referred to as the 'free speed'. The value of this speed usually is governed by the speed limit, the amount of speeding, and the percentage freight traffic. For slightly larger traffic demands, the average traffic speed hardly decreases. For even larger demands however, the average traffic speed gradually decreases, to about 80 km/h if demand reaches capacity (the so-called 'capacity speed', or 'critical speed').

If at a certain moment traffic demand exceeds capacity in a certain cross section of the road section, the traffic flow breaks down. The excess demand is stored in a queue, forming upstream of the bottleneck. In several studies the outflow rate at the head of a queue (referred to a as the 'queue discharge rate') is found to be smaller than the maximum flow rate before the traffic flow breaks down (referred to as the 'free flow capacity'). This difference in capacity is called the 'capacity drop'. It is in the range of 1 to 15 percent (Hoogendoorn, 2007). In a way, the capacity drop makes the occurrence of traffic congestion a self-reinforcing process: once congestion has set in, capacity is reduced, resulting in the traffic conditions deteriorating more rapidly (as compared with the situation without a capacity drop).

The traffic conditions within the queue are referred to as being 'congested' or 'forced'. Within the queue, the average traffic flow (i.e. the average number of vehicles passing a certain cross section per unit of time) is determined by the discharge rate at the head of the queue. This is only true though if this traffic flow is averaged over a sufficient amount of time. Due to the fact that within the congested traffic stop-and-go waves may form, over shorter periods of time the flow might temporarily be larger or smaller.

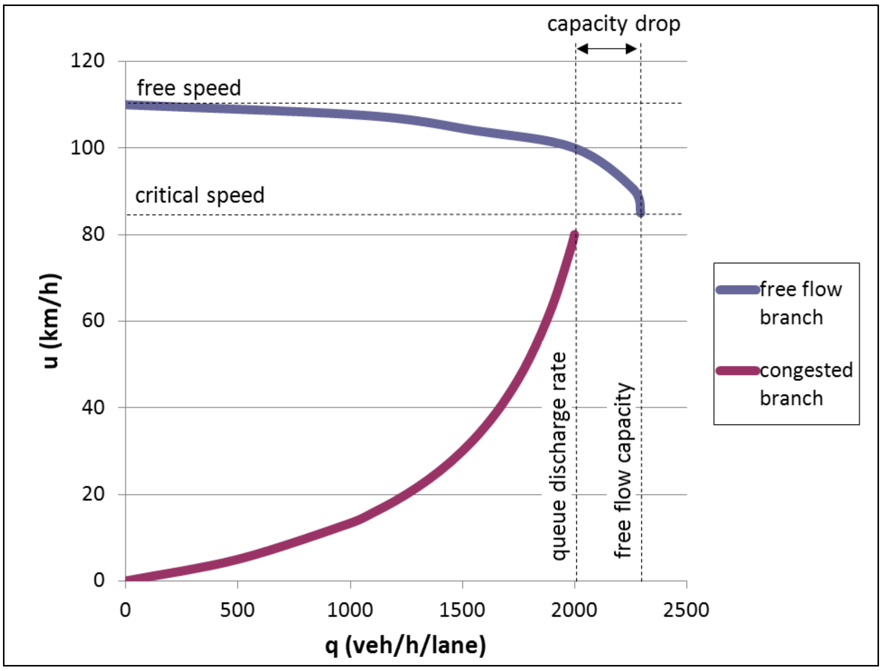
Obviously, the average speed in the queue is dependent on the average traffic flow in this queue. When this average flow is larger, the average speed in the queue is larger as well. If the traffic flow in the queue is zero, the speed is zero too.

All the relationships and notions discussed above are depicted in Figure 2.1. This figure shows an example of the so-called 'fundamental diagram' for a cross section of a road. It describes the average equilibrium relationship between traffic flow (q) and velocity (u). With 'equilibrium' it is meant that this relation is only valid for stationary traffic conditions. Transient traffic states will deviate from this relationship. Also note that it is only an *average* relationship. Real-life data are widely scattered around this average. In particular this is the case for the 'congested branch' of the fundamental diagram.

Finally it should be stressed that a fundamental diagram is in its entirety related to one and the same cross section of a road. If the fundamental diagram depicted in Figure 2.1 for example would belong to a cross section upstream of the bottleneck, then the indicated capacities thus also would concern the capacities of this particular cross section, and not the (lower) ones of the bottleneck.

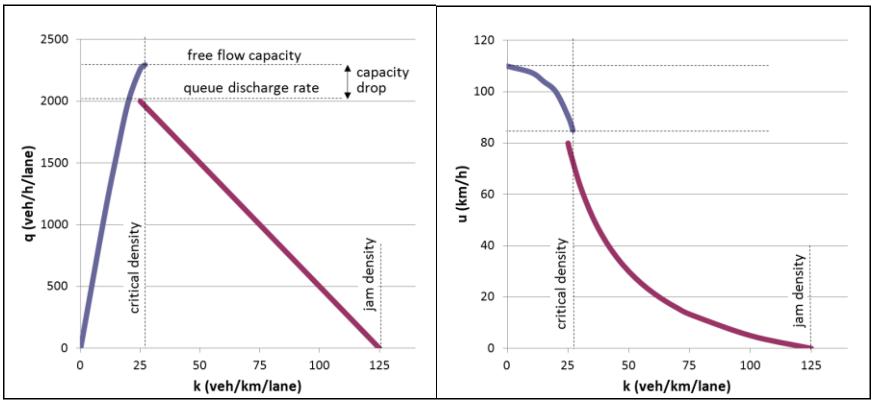
For distinct cross-sections, the fundamental diagram can be rather different (including the values of the free flow capacity and the queue discharge rate). This is not only the case if these are cross sections of different roads. Cross sections of one and the same road may show differences too. This might be due to differences in for example lane width, grade or curvature.

Figure 2.1: Fundamental diagram



The fundamental diagram is also known in two other forms. This is illustrated in Figure 2.2. In the fundamental diagram on the left, the relation between the traffic flow (q) and the traffic density (k) (i.e. the number of vehicles present per unit of distance) is depicted. In the fundamental diagram on the right, the relation between the speed (u) and the density (k) is presented. In fact, these fundamental diagrams represent exactly the same information. They can easily be converted into each other by using the relationship $q=k\cdot u$. This well-known relationship, valid for stationary and homogeneous traffic states, is referred to as the 'fundamental relation'.

Figure 2.2: Other forms of the fundamental diagram



As long as the traffic demand (corresponding to the inflow to the queue) remains higher than the queue discharge rate, the queue keeps on growing. As a consequence, the delay that road users experience keeps on increasing as well. Note the important role of the capacity drop in this respect. Only as soon as the traffic demand decreases to a value below the queue discharge rate, the queue will start to dissolve. Gradually the length of the queue will decrease, until it is completely dissolved (and free flow traffic conditions are restored again).

Note that the physical length of a queue does not directly follow from the number of vehicles in this queue and the number of lanes on which they are 'stored'. Obviously, the physical length is dependent on the (average) traffic density within the queue as well. This density can be derived from the fundamental diagram, using the flow rate in the queue. This flow rate follows from the capacity of the downstream bottleneck (responsible for the queue) and possible flows leaving or entering the queue somewhere in between its head and tail.

A complicating factor in the process governing traffic congestion is that, especially in the somewhat longer term, the traffic demand is not independent of the traffic supply. If the quality of the traffic supply is improved (for example by adding some extra capacity) traffic demand will increase.

2.2 Sources of temporal variation in traffic congestion

2.2.1 Introduction

In chapter 1 it was pointed out that the traffic conditions on the main road network show a significant degree of variability, not only over the course of the day, but also between days. This variability is due to a significant variation in both traffic demand and traffic supply (capacity). Illustrations of this variability in traffic demand and supply can be found in (Tu, 2008) en (Brilon, 2005), respectively. Tu gives some data on the variability in the total number of car trips that are made in the Netherlands during the peak hours of a working day. On average, this number amounts to 5.5 million trips. However, on the 5% quietest working days this number is less than 4 million (a difference of more than 25%), whereas on the 5% busiest days there are more than 6.9 million trips by car (again a difference of more than 25%). For the number of trips during the off-peak part of the day a similar bandwidth is found.

Brilon gives some data on the variability of the capacity of a number of German freeway sections. For the capacities of these freeway sections (calculated from 5-minute counts) coefficients of variation of about 9% have been found[10] (considering each freeway section individually). In view of this significant variability, it is in fact quite striking that the capacity traditionally is treated as a constant value in traffic engineering guidelines around the world.

For both the temporal variability in traffic demand and the temporal variability in traffic supply a large number of causes can be discerned.

[10] Note that these are all freeway sections without a distinct bottleneck. For freeway sections *with* a distinct bottleneck of course different values might be obtained.

As for the *traffic supply*, the following sources of variability are identified:

- variations in weather conditions
- variations in luminance
- incidents
- road works
- traffic control actions
- variations in vehicle population
- variations in driver population
- 'intrinsic' variations in human behavior[11]
- demonstrations
- emergencies

Regarding the *demand* fluctuations, the following factors are distinguished:

- regular pattern of variation in human travel behavior over the day
- regular pattern of variation in human travel behavior over the days of the week
- regular pattern of variation in human travel behavior over the periods of the year
- public holidays
- events
- strike actions
- weather conditions
- road works
- emergencies
- other variations in human travel behavior
  (i.e. those not explained by the aforementioned factors)

Two other factors influencing the traffic demands are:

- travel behavioral changes in response to traffic information
- travel behavioral changes in response to one's recent travel experiences

In spite of the fact that these two factors affect the traffic demands in a variable way as well, they actually cannot really be considered *sources* of fluctuations. After all, in fact they only exist because of variations in traffic conditions that are caused by other (i.e. 'real') sources of fluctuations in traffic demand and supply. If these latter sources would not exist, the two factors mentioned above would not exist either. After all, if there would not be any variations in traffic conditions at all, traffic conditions would be fully predictable in advance. Consequently there would be no question of changing one's travel behavior in response to information on the actual traffic conditions or one's own recent travel experiences.

[11] Note that most of the other sources of variability have a behavioral component as well. For example, weather conditions affect the traffic conditions by affecting the driving behavior. The item 'intrinsic variations in human behavior' rather refers to the fact that in spite of finding himself in similar circumstances, one and the same person may still behave differently. Furthermore, there are obviously variations in behavior between different individuals (even if these belong to one and the same driver population). These variations are categorized under this heading 'intrinsic variations in human behavior' as well.

It should be noted that at the level of an individual road section, the variability in the traffic conditions on *other* sections of the road network might be a source of variability as well. After all, due to the occurrence of network effects (i.e. interaction effects between the different sections of a network), traffic conditions on other road sections might affect the traffic demand and discharge capacity of the road section under consideration. In this section these network effects are not considered however. These effects are the topic of section 2.3. Note that these network effects cannot really be seen as sources of variability. Considered at the network level, they are merely part of the processes that determine how variations in demand and supply conditions finally affect the traffic conditions on the network.

2.2.2 Classification

The various sources of variations are different in nature. They can be classified in different ways. First of all a distinction can be made between *regular* (or *systematic*) variations, occurring according to some regular pattern over time, and *irregular* variations. Another distinction that can be made is between sources of variability that have a *continuous* (though variable) influence on the traffic demand and/or supply, and sources of variations that affect the traffic demand and/or supply *only during well-defined spaces of time*. For the rest of the time, the effects on demand and supply are zero. These latter sources of variability can be referred to as 'events' or 'disturbances'.

Finally, yet another way to classify the various sources of fluctuations is according to their spatial scope. Some sources of variations affect the traffic demand and/or supply network wide, while others affect the demand and/or supply only locally. Some sources of variability have a spatial scope that is in between the network level and the local level. In such cases for example a certain subarea of the network might be affected, or a specific group of origin-destination relations.

It should be noted that if the traffic demand and/or supply are affected at an above-local level, this of course not necessarily means that the effect is homogeneous in magnitude. This means that effects occurring at an *above-local* level in fact might have some *local* component as well. Actually, some effects might even differ in 'direction' among different locations. Consider for example the demand effect of the summer vacation period. On some routes, daily traffic demands might be lower during this period (especially on routes with relatively much commuting traffic), while there might also be routes on which the traffic demand is larger (especially on routes with relatively much recreational traffic).

Further it should be noted that a substantial part of the sources of variability can take place at different spatial levels. Adverse weather conditions for example might be local in nature (like in case of a small, yet possibly heavy shower), cover the whole network (like in case of an extensive rain belt), or somewhere in between (like in case of a small rain front). Also note that if the primary *effect* on traffic demand/supply is only local in nature, the *consequence* for the traffic conditions actually might have a much larger spatial scope (due to the occurrence of network effects, which will be discussed in section 2.3).

In the two tables below, all sources of variations in traffic demand (Table 2.1) and supply (Table 2.2) are assigned to classes formed by combining the three classification systems discussed above. The columns and rows of the tables represent two of these classification systems. Symbols are used to indicate the third classification. Inevitable, in some cases this assignment to categories is debatable to some extent. The variation in weather conditions for example is assigned to the class 'events/disturbances', whereas in reality variations in weather conditions are practically continuously present. However, bad weather conditions (or just the opposite: summery weather conditions) can be distinguished reasonably well from the more 'everyday' weather conditions. Therefore it was decided to assign the variation in weather conditions to the category 'events/disturbances', in spite of its continuous element. Another example of a classification that is debatable to some extent is the assignment of the influence factor 'events' to the category 'sources of *irregular* variations'. In fact, a lot of events are organized every year again, at the same moment of the year (similar to public holidays). This subset of events therefore rather belongs to the category 'sources of *regular* fluctuations'.

Table 2.1: Classification of the various sources of variations in the traffic demand according to time span (horizontal), degree of regularity (vertical) and spatial scope (N=network-wide, L=local, B='in between') of their effects

Time span / Degree of regularity	Continuously present	Event / Disturbance
Sources of regular variations	Regular pattern of travel behavior over the day *(N)* Regular pattern of travel behavior over the days of the week *(N)* Regular pattern of travel behavior over the periods of the year *(N)*	Public holidays *(N)*
Sources of irregular variations	Unexplained variations in human travel behavior *(B)*	Varying weather *(N/B)* Road works *(B)* Events *(B)* Strike actions *(N/B)* Emergencies *(N/B)*

Table 2.2: Classification of the various sources of variations in the traffic supply according to time span (horizontal), degree of regularity (vertical) and spatial scope (N=network-wide, L=local, B='in between') of their effects

Time span / Degree of regularity	Continuously present	Event / Disturbance
Sources of regular variations	Var. in vehicle population *(N)* Var. in driver population *(N)*	Darkness *(N)*
Sources of irregular variations	Var. in vehicle population *(B)* Var. in driver population *(B)* Var. in human behavior *(L)*	Incidents *(L)* Demonstrations *(L)* Emergencies *(N/B/L)* Varying weather *(N/B/L)* Road works *(L/B)*

Note that in Table 2.2 the influence factors 'variations in vehicle population' and 'variations in driver population' are assigned to both the categories 'sources of *regular* variations' and 'sources of *irregular* variations'. This is because of the fact that these sources of variations to a large extent can be described by regular patterns over time (representing their systematic parts), but still with a part of the variations remaining unexplained (representing their random/irregular parts).

In Table 2.2, the influence factor 'traffic control actions' is omitted. This is because of the fact that it cannot really be assigned to one of the categories. Some traffic control actions influence the supply conditions on a continuous basis, while others are active only during specific periods in time. Furthermore, some types of traffic control are regular in nature, while others act in a traffic responsive (and thus partially irregular) way.

Note that the category of irregular events/disturbances can be further divided into circumstances that are *planned* (road works and events) and circumstances that are *unplanned* (incidents and emergencies). Strike actions, demonstrations and varying weather conditions cannot be unambiguously assigned to one of these two categories.

2.2.3 Relevant characteristics of the various sources of fluctuations

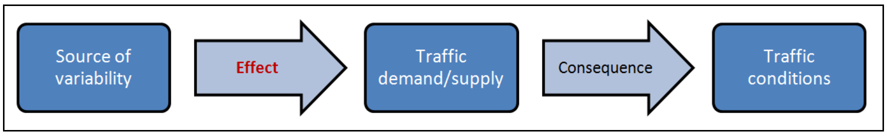
In the following subsections the various sources of fluctuations listed above are discussed in more detail. Table 2.3 shows the important aspects to be considered for the different categories of sources of variability. For the sources of variability that are continuously present, consideration is given to their patterns over time, or to the magnitude of their stochastic fluctuations (depending on whether it concerns a source of *regular* (i.e. systematic) variation or a source of *irregular* variation). For the sources of variability that can be referred to as 'events/disturbances', consideration is given to both their frequencies of occurrence (deterministic or stochastic) and the magnitude of their effects.

Table 2.3: Relevant aspects for the different categories of sources of variability in traffic demand and supply

Time span / Degree of regularity	Continuously present	Event / Disturbance
Sources of regular variations	(deterministic) pattern over time	- frequency of occurrence (deterministic) - effect
Sources of irregular variations	magnitude of the stochastic fluctuations	- frequency of occurrence (stochastic) - effect

It should be noted that in this context 'effect' refers to the influence on the demand or supply, and not to the final impact on the traffic conditions, created through the interaction between demand and supply. To distinguish this final impact on the traffic conditions from the influence on the demand or supply, the latter is called 'effect', while the former is termed 'consequence'. This is illustrated in Figure 2.3.

Figure 2.3: Distinction between effect and consequence



First, in sub section 2.2.4, the sources of demand fluctuations are considered. In subsection 2.2.5 the sources of supply variations are discussed. It should be noted that the various sources of variability are not all independent from each other. In fact there are a lot of non-linear, dynamic dependencies involved between these fluctuations. This topic is shortly returned to in subsection 2.2.6.

2.2.4 Sources of the temporal fluctuations in traffic demand

1) Regular pattern of variation in human travel behavior over the day

Due to the fact that the traffic demand is strongly related to people's activity patterns, there is a clear relation between the time of the day and the size of the traffic demand. During the nighttime, traffic demand is typically very small, while during the daytime traffic demand is much larger. On working days two clear peaks can be observed in the traffic demand pattern. These are related to commuters traveling from home to work and back. Weekend days are clearly different in this respect. On these days clear peaks in the traffic demand are typically lacking. This can be illustrated with Figure 2.4, showing the traffic demand pattern over the course of the day for the Dutch A12 motorway, separately for working days and weekend days. It should be noted, however, that the patterns shown in this figure in fact are traffic *flow* patterns. These actually might be different from *traffic demand* patterns, since they might be affected by the occurrence of traffic congestion. After all, due to this traffic congestion, the peaks in the *measured traffic flow* might be less pronounced than the peaks in the *traffic demand.*

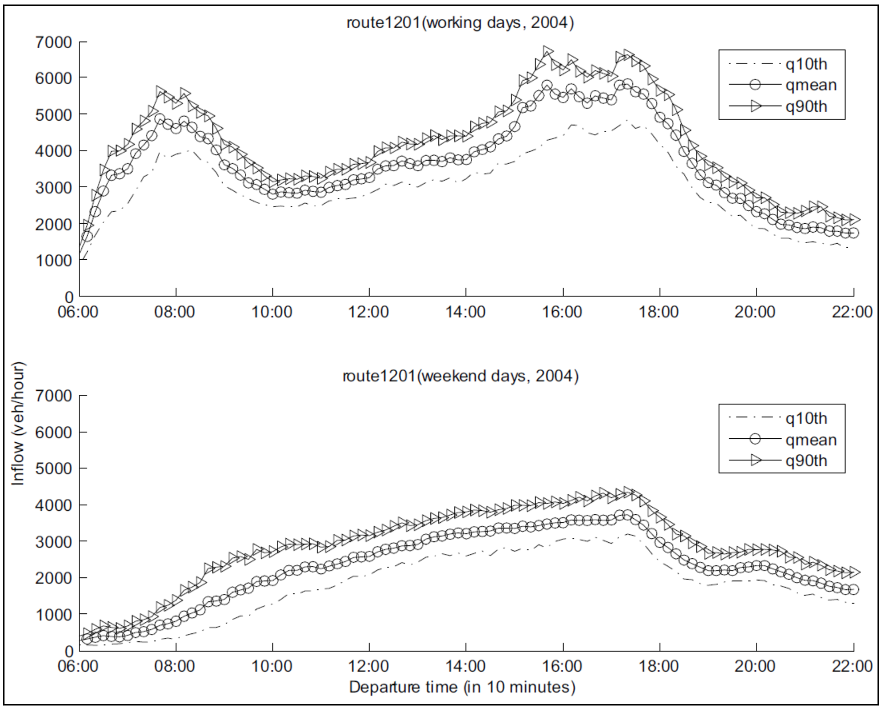


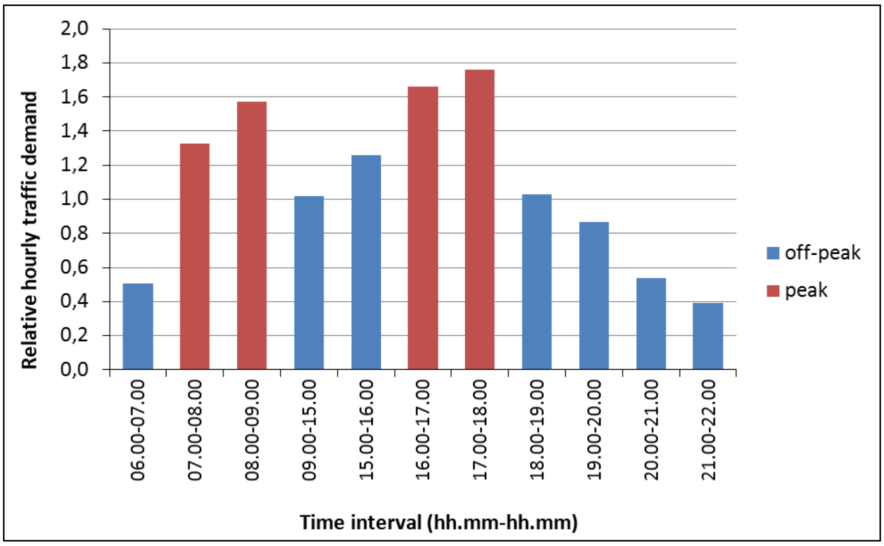
Figure 2.4: Time dependent variations in the traffic flow on the Dutch A12 motorway (southern part) on working days and weekend days, based on data from the year 2004 (Source: Tu, 2008)

Often the two peaks in the working day demand pattern are different in size. This then is due to the commuting traffic being unequally distributed over the two directions. In Figure 2.4 such a difference between the two peaks can be observed as well.

Obviously, the validity of the patterns in Figure 2.4 is limited to only one specific measurement location (on the A12 motorway). For other road sections the course over time might be rather different. In (Hilbers et al, 2004) some general (i.e. average) data on the relative traffic demands for various periods of the day is given. Using these data, the diagram shown in Figure 2.5 has been constructed. This diagram shows the relative hourly traffic demands for various time intervals of the day, normalized with the average hourly traffic demand over the period between 6:00 and 24:00. Note that one time interval in this diagram has a length that deviates from the length of the other intervals: the largest part of the daytime off-peak period is put together in one time interval (between 9:00 and 15:00).

It should be noted here that there are two limitations associated with these data, which are relevant in the context of this discussion. First of all, the data are most likely related to 'private' trips (including home-work trips) only. This means that 'professional' trips are not taken into account. The latter are typically relatively evenly distributed over the daytime period. Secondly, the data are related to all car trips together, so not specifically to the traffic on the motorway network. Obviously, the trips over the motorway network might be distributed over time (somewhat) differently than the other car trips.

Figure 2.5: Relative hourly traffic demands for different periods of the day, normalized with the average hourly traffic demand over the period between 6:00 and 24:00
(Based on data from Hilbers et al, 2004)



From Figure 2.5 the hourly traffic demand during the evening peak appears to be somewhat larger than the one during the morning peak. Of course, though, this is just an *average* pattern. For individual road sections the course over time might be very different, as discussed already in relation to Figure 2.4. From the figure it also appears that for both peak periods in the second hour the traffic demand is somewhat larger than in the first hour.

2) Regular pattern of variation in human travel behavior over the days of the week

As was already indicated above, the traffic demand on working days is rather different from the traffic demand on weekend days (which is due to significantly differing human activity patterns). On weekend days clear peaks in the traffic demand are typically lacking and the traffic demand sets in much later than on working days. All in all the total daily traffic demand generally is significantly smaller on weekend days. However, the actual size of this difference shows a relatively large variation among different locations on the motorway network (BGC, 1986).

While working days are different from weekend days, to some degree the different working days (Monday to Friday) are *mutually* divergent as well. Based on data from 1984 and 1985, BGC (1986) found that the total daily traffic demand on Dutch motorways gradually increases in the course of the week, meaning that the daily traffic demand is lowest on Mondays and highest on Fridays. More recently (based on data from 1995 and 2000), Harms (2003) found a similar result, as shown in Table 2.4. These results however should be interpreted with care, since they are related to the mobility as a whole (including all transport modes) and thus not specifically to the traffic demand on motorways. Another important limitation is that the results are based on an analysis of the 'private' trips only. Professional trips were not considered.

Table 2.4: Relative level of the total mobility on the different weekdays
(Based on data from Harms, 2003[12])

Weekday	Mon	Tue	Wed	Thu	Fri
Index of total mobility (Mon=100)	100	107	107	110	116

Hilbers et al (2004) provide data specifically for the car trips (but still for all road networks together, rather than for the motorway network alone). These are shown in Table 2.5. According to these results, the peak period demands on Tuesday through Thursday are more or less equal, while those on Monday and Friday are slightly lower (based on fixed peak time windows of 7:00 - 9:00 and 16:00 - 18:00). For the off-peak periods (defined as the remaining parts of the period 6:00 - 24:00), traffic demand increases from Monday to Friday. From the separate indices for the peaks and off-peaks one can also compute the indices for the daily totals, by taking the average of these indices, weighted according to the average numbers of trips during the peak and off-peak (5.7 million and 10.5 million, respectively). Based on the resulting indices (included in Table 2.5) one can conclude that the daily traffic demand gradually increases in the course of the week. This is in line with the results of BGC mentioned above.

Table 2.5: Relative level of traffic demand (car trips only) on the different weekdays
(Based on data from Hilbers et al, 2004)

Weekday	Mon	Tue	Wed	Thu	Fri
Relative traffic demand peak (mean peak = 1)	0.96	1.03	1.02	1.02	0.98
Relative traffic demand off-peak (mean off-peak = 1)	0.90	0.97	1.01	1.04	1.08
Relative traffic demand peak and off-peak combined (mean weekday = 1)	**0.92**	**0.99**	**1.01**	**1.03**	**1.04**

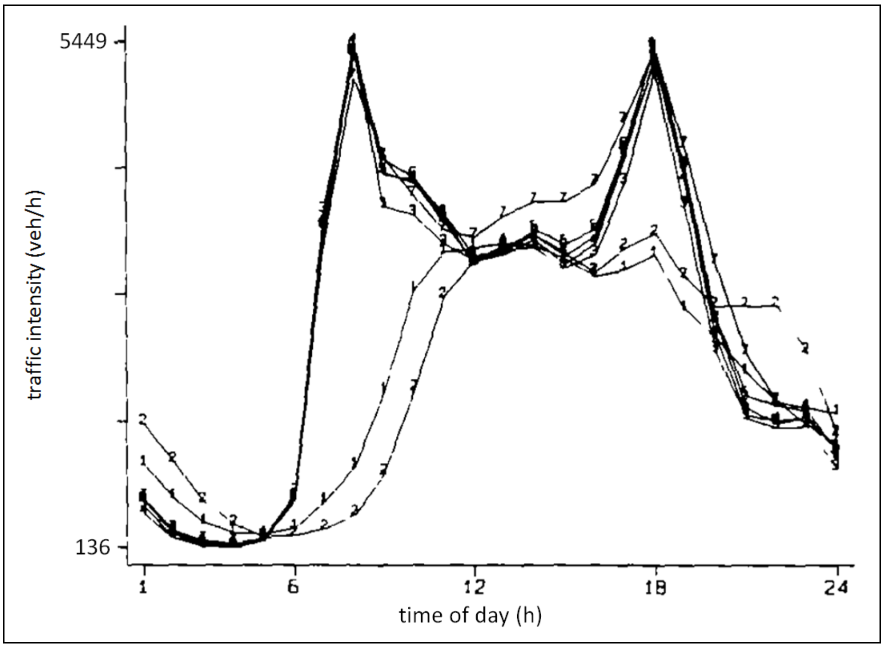
Friday is the day that distinguishes itself most from the other weekdays. On this day, the evening peak starts much earlier than on the other days, and lasts longer. This results in the Friday evening peak being significantly longer than the evening peaks on the other days (though not significantly *higher*). This is clearly observable in Figure 2.6, showing the hourly traffic intensity patterns for the various days of the week, obtained for a measurement location on the Dutch A2 motorway (near Maarssen).

In fact not only the *evening peak* is different on Fridays, but the *morning peak* and *daytime off-peak period* as well. On Fridays the morning peak traffic demand is typically smaller than on the other days. During the daytime off-peak period there is more traffic than on the other working days. While Monday through Thursday show an almost symmetrical off-peak period between the morning and evening peak, on Friday the traffic demand monotonically increases from about 10 a.m. until the (gradual) transition to the evening peak (BGC, 1986). To a more limited extent than Friday, Monday is different from the other working days as well. On this day the morning peak is smaller than on the other working days.

[12] Data relating to the numbers of trips longer than 7.5 minutes, per quarter of an hour (where one trip might be counted in multiple quarters), excluding professional trips and trips of persons younger than twelve years.

The two different weekend days, Saturday and Sunday, show some mutual differences as well. On Sundays the traffic demand can be expected to set in later than on Saturdays. This is confirmed by the patterns shown in Figure 2.6. In this figure it can also be observed that on Sunday the traffic demand is relatively high during the early night (i.e. the night from Saturday on Sunday). Considering the total daily traffic demands, there is no general rule for the difference between Saturdays and Sundays. On some locations the daily traffic demand is higher on Saturdays, while on the other locations it is higher on Sundays (BGC, 1986).

Figure 2.6: Hourly traffic intensity patterns for the different days of the week (excluding public holidays), as obtained for a measurement location on the Dutch A2 motorway. (1=Saturday, 2=Sunday, ..., 7=Friday).
(Source: De Vries & Praagman, 1995)



Weijermars and Van Berkum (2007) grouped daily traffic profiles on a Dutch highway location by applying clustering analysis. A pre-classification in working days and non-working days was applied to successfully improve the clustering results. For the subsequent clustering of the *working days* a satisfactory result was obtained (i.e. a clustering with a small remaining variation within all individual clusters). In general, Mondays and Fridays were classified to separate clusters (i.e. one with Mondays and one with Fridays). Tuesdays, Wednesdays and Thursdays were classified to the same cluster. This confirms what was discussed above.

The results of the mutual clustering of *non-working days* were less satisfactory. There remained a substantial variation within some of the clusters, in spite of the fact that the clusters were already rather small. Therefore, Weijermars and Van Berkum concluded that it is not possible to distinguish recurrent traffic patterns for non-working days, probably caused by the fact that these days show less fixed activity patterns. This conclusion is also supported by the fact that several other studies have shown that weekend days show a larger variability in traffic demand than weekdays (Weijermars, 2007).

3) **Regular pattern of variation in human travel behavior over the periods of the year**

Over the different periods of the year, traffic demands are not constant. Vacation periods play an important part in this. During these vacation periods traffic demand is lower. Probably this effect is most apparent in the morning and evening peaks of the traffic demand pattern. This hypothesis seems to be confirmed by the results of the clustering analysis performed by Weijermars and Van Berkum (2007). In this clustering analysis days within a vacation period were classified to a separate cluster that showed a flatter daily flow profile.

At the start and end of vacation periods however just the opposite effect may occur: higher (peaks in) traffic demand due to a large amount of vacation traffic. In the Netherlands the government tries to limit this effect though, by staggering vacations. This seems to be successful: Bexelius and Kengen (1993) conclude from a study on the characteristics and predictability of the days with the highest traffic demands that the contribution of the summer vacation months (i.e. July and August) to the list with the busiest days is remarkably modest (at least for the roads that were included in their analysis).

However, from one of the tables in their article (showing the days with the busiest afternoons in 1991) it can be derived that the start and end of the *autumn* vacation week might give significant peaks in the traffic demand. On the last workday before the vacation (a Friday) the traffic demand during the late afternoon was higher than usual. On the Wednesday halfway the vacation and the Friday and Sunday at the end of the vacation period, this was the case as well. It should be noted that the vacation traffic at the *end* of the vacation (i.e. the *returning* traffic) might have a dominant orientation that is opposite to the one of the vacation traffic at the *start* of the vacation period (i.e. the *leaving* traffic). This of course will be dependent on the geographical situation of the road section considered.

Hilbers et al. (2004) provide some quantitative data on the vacation effects and other (remaining) seasonal effects on the *working day* traffic demand. These data are plotted in Figure 2.7 and Figure 2.8, separately for the peak periods and off-peak periods respectively. Again it should be noted that the data relate to the total population of car trips, rather than to the subpopulation of car trips over the motorway network. Furthermore, most likely the data are limited to private trips only.

Figure 2.7 shows that in vacations, the average peak period traffic demand is more than 20% lower than on non-vacation working days. According to Figure 2.8, during the off-peak periods this effect is only about half as large. From the figures also a seasonal variation over the various months of the year can be observed (even though the monthly data have been corrected for the vacation effect). During the summer months the traffic demand is typically lower than during the winter months. This might (partially) be explained by the fact that some trips are made by bike in the summer, while these are made by car in the winter. Since this latter phenomenon will be restricted to the relatively short distance trips, the seasonal variation in the *motorway network* traffic demand might be more limited than the seasonal variation in the *total* traffic demand shown in the figures.

December shows a remarkably high off-peak traffic demand (Figure 2.8). This probably is related to special activities during the public holidays in this month (like family visits) or other social or recreational activities during the Christmas vacation period.

Figure 2.7: Vacation effects and other seasonal effects on the working day peak period traffic demand (average working day peak period traffic demand = 1)
(Data from Hilbers et al, 2004)

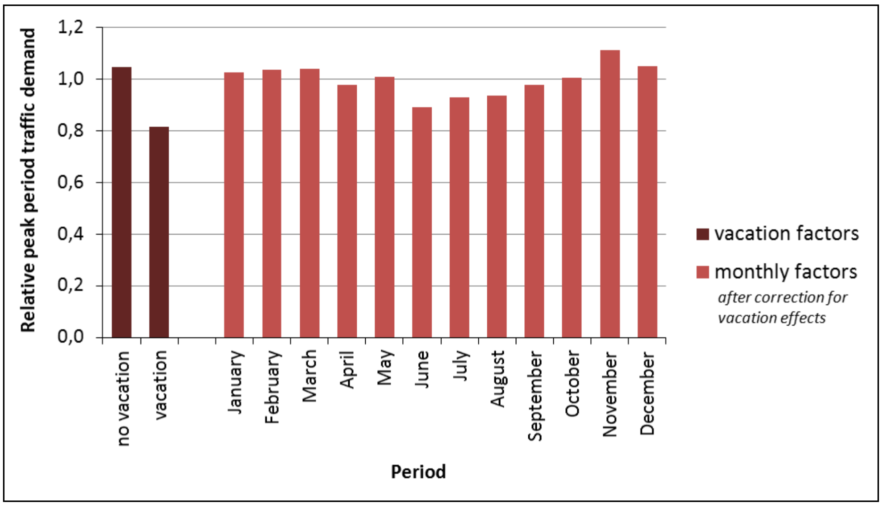
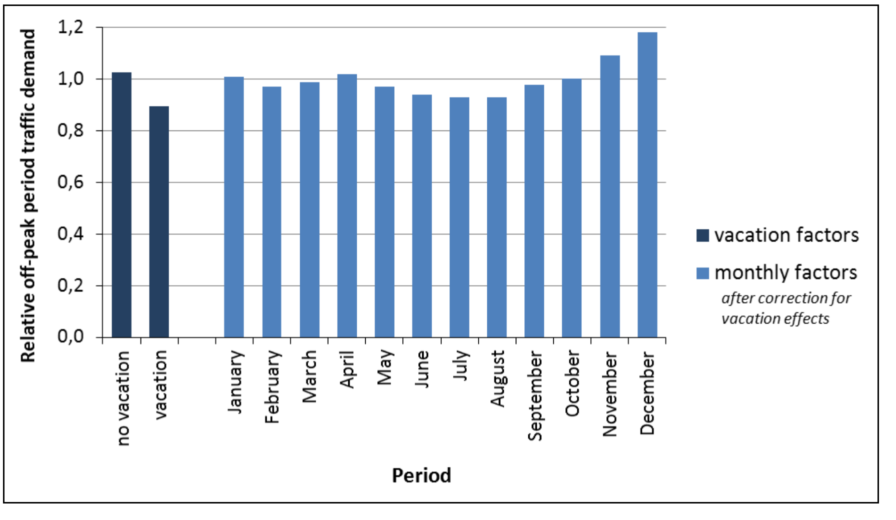


Figure 2.8: Vacation effects and other seasonal effects on the working day off-peak period traffic demand (average working day off-peak period traffic demand = 1)
(Data from Hilbers et al, 2004)



An empirical study of BGC (1986) for a large set of measurement locations on the Dutch motorway network (based on a dataset for the period 1975-1984) revealed that the pattern of variation over the months of the year is in fact quite variable among various locations on the motorway network, especially as far as the summer months are concerned. On part of the locations a dip in the traffic demand can be observed during these summer months (related to the summer vacation period), while on another part of the locations a (substantial) peak can be observed during these months. Most likely this peak can be attributed to additional (recreational) traffic during the off-peak periods (and not so much to additional traffic during the peak periods), although BGC did not examine this.

Remarkably, BGC (1986) found that the traffic demands during the winter months (especially during December and January) are relatively low as compared with the yearly average, which is in contradiction to the results of Hilbers et al (2004), as can be seen from the figures above. One possible explanation for this might be that the extra winter trips found by Hilbers et al. are to a large extent shorter distance trips (possibly trips that are made by bike in the summer), for which relatively little use is made of the motorway network. Another partial explanation might be found in the Christmas vacation period, for which in the study of Hilbers et al. a general vacation correction is applied. Yet another possible explanation could be that in the twenty years between the two studies some behavioral changes have taken place. However, since De Vries and Praagman (1995) in a later study (based on data from 1993) found something similar for the winter months as BGC (except for December, for which the average traffic demand was found to be only slightly below the yearly average), this latter explanation seems not very plausible. Finally, a possible explanation might be that the winter period traffic demand in the study of Hilbers et al. was relatively high 'by coincidence' (which might be the case because of the fact that data from only one single year were considered).

4) Public holidays

Just like during vacation periods, on public holidays the usual (working day) peaks in traffic demand are less pronounced or even absent. In the Netherlands these public holidays are: New Year's Day, (Good Friday), Easter Monday, Queen's Birthday, (Liberation Day), Ascension Day, Whit Monday, Christmas Day and Boxing Day. If such a public holiday falls on a Tuesday or Thursday, many people take the preceding Monday or following Friday off as well, resulting in the peak period traffic demands to be lower on these days too.

In spite of the demand reduction mentioned above, on certain parts of the road network the traffic demand might be *higher* on public holidays, related to events or other recreational destinations attracting a lot of traffic. In most cases this increase in traffic demand (as compared with normal working day conditions) will be limited to the off-peak periods. Events are discussed in more detail below. In this thesis public holidays and events are considered separately. Of course, any possible relationships between these two should be taken into account however (see section 2.2.6).

Most public holidays fall in the first or last part of the (working) week, in particular those in the spring. (As their names imply, Easter Monday and Whit Monday are always on a Monday, and Ascension Day is always on a Thursday.) In these situations, many people leave for a long weekend, which can be expected to result in peaks in the traffic demand at the start and end of these long weekends. This is confirmed by Bexelius and Kengen (1993), whose research shows that systematically recurring 'peak days' are mainly found near the start and end of long weekends around public holidays in the spring (and thus certainly not necessarily on the public holidays *themselves*). The peak due to the leaving traffic typically starts in the afternoon of the last working day before the long weekend, and remains apparent until the next morning. The peak due to the returning traffic starts on the last day of the long weekend and ends on the next morning.

For east-west orientated motorways in the Netherlands (especially the A12) also public holidays in Germany appear to have a clear influence on the traffic demand (Bexelius and Kengen, 1993). A special situation occurs when a German public holiday falls on a day that is a normal working day in the Netherlands. In this situation day-trippers from Germany intensify the Dutch commuting peaks. Especially Corpus Christi (on the Thursday ten days after Whit Monday) causes a very large traffic demand on the A12 motorway.

Intermezzo: systematic (or regular) variability vs. non-systematic (or irregular) variability

Now that all sources of *systematic* variations in the traffic demand have been discussed (see Table 2.1), we can move on to the sources of *non-systematic (*or *irregular)* variations, which together make up the remaining part of the total variation in the traffic demand. An interesting question, however, is what we can say now already about this remaining part of the variations, by studying the difference between the total variability of the traffic demand and the variability made up by the combination of all systematic variations. This question is dealt with in this intermezzo, again by considering results from previous research.

Figure 2.9 and Figure 2.10 show the overall day-to-day spreading in the number of car trips over the period of a whole year (for the peak period and the off-peak period respectively). Clearly this spreading is rather substantial. For the peak periods the coefficient of variation amounts to about 0.17 and for the off-peak periods about 0.15. Hilbers et al (2004) performed a variance analysis in order to find out which part of this total day-to-day variation in the traffic demand could be explained by the factors 'day of the week', 'vacation/non-vacation', and 'month of the year'. This part turned out to be rather limited: only 23% and 22%, for the peak periods and off-peak periods respectively. Although the systematic influences of public holidays were not considered, this means that the largest part of the (considerable) variation between the working days seems to be non-systematic.

Figure 2.9: Empirical distribution of the number of car trips in the peak periods
(Source: Hilbers et al, 2004)

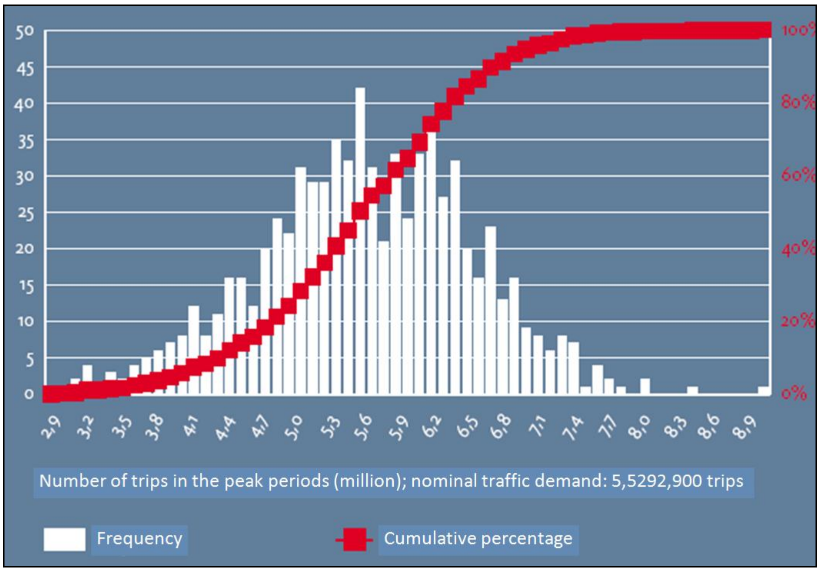
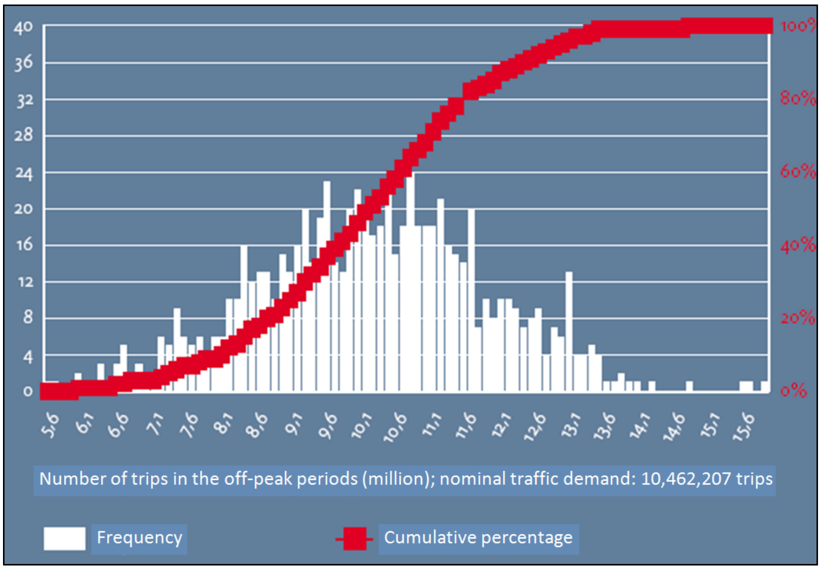


Figure 2.10: Empirical distribution of the number of car trips in the off-peak periods
(Source: Hilbers et al, 2004)



For a large set of measurement locations on the Dutch motorway network, De Vries and Praagman (1995) subtracted the systematic variations (i.e. the patterns over the day, over the days of the week, and over the weeks of the year) from the hourly traffic volumes that had been measured over the whole of 1993, and analyzed the residual variation. They found a relation between this residual variation and the average hourly traffic volume: for periods of the day with a larger average hourly traffic volume the residual variation appeared to be larger. However on weekend days this residual variation turned out to be relatively larger than on weekdays. That is, for the *same* average hourly traffic volume, on weekend days a larger residual variation was found. This is in agreement with the finding of Weijermars and Van Berkum (2007) that non-working days (being predominantly weekend days) show rather variable traffic demand patterns (see the earlier subsection on the systematic influence of the day of the week).

De Vries and Praagman found that the residual traffic volumes are correlated over time: if during a certain hour the traffic demand is larger than the average traffic demand for that hour (and for the concerning day of the week and week of the year), the traffic demand during the next hour is likely to be above average as well. Interestingly, they also found a substantial *spatial* correlation (correlation coefficients around 0.5) to exist between the residual traffic volumes on the various measurement locations that they considered, while these locations were not all closely together. (For measurement locations located very closely together, like two measurement locations on one and the same road, a strong correlation in traffic volumes of course would be rather obvious.) As De Vries and Praagman indicate, this suggests that aside of the patterns already eliminated (i.e. the patterns over the day, over the days of the week, and over the weeks of the year), there are still other common factors that result in correspondence between the traffic volume variations on the various measurement locations. One can think of weather conditions or other circumstances that affect the traffic volumes on a more or less network-wide basis. This kind of (network-wide) non-systematic influence factors will be discussed in more detail below, as well as some non-systematic influence factors that affect the traffic volumes on a more local basis.

5) Events

For a limited number of days a year, events like exhibitions, shows, festivals and sporting events may attract a lot of traffic. These peaks in traffic demand however are not network wide (as all variations discussed above), but limited to the surroundings of the locations of the events.

In fact, 'events' should not be understood too literally. After all, also recreation areas and tourist attractions might have peaks in traffic attraction on a certain number of days a year (for example on public holidays or on days with summery weather conditions).

Meeuwissen et al (2004) constructed a 'top 45' of destination zones (in the Netherlands) in which events take place. In this top 45 the destination zones with events are ranked according to the average number of cars per hour that are attracted by the events. This number ranges from 46 vehicles per hour (rank 45) to 11,786 vehicles per hour (rank 1). The frequencies with which the events occur during 'typical' peak hours and 'typical' off-peak hours have been estimated as well (for workdays only). These widely varying frequencies range from 0.40% (events covering only one weekday a year) to about 30% (destination zones with one or more tourist attractions and/or festivals). These are the frequencies estimated for the 'typical' off-peak hours. For the 'typical' peak hours the frequencies are assumed to be half of those.

6) Variability in weather conditions

- Effects -

Adverse weather conditions may affect the general level of traffic demand in two opposing ways. On the one hand, there might be a shift in the modal split from transportation modes that are characterized by a larger exposure to the weather (like the bicycle or the public transport[13]) to the car, *increasing* the traffic demand. For short distance trips this shift will be larger than for long distance trips. On the other hand, road users might refrain from making trips, resulting in a *decrease* of the traffic demand. This will especially be the case in situations in which the traffic safety might be reduced (e.g. if there is black ice on the roads), or if a traffic chaos might be expected (e.g. in case of heavy snowfall)[14]. However, also in case of rainy weather conditions traffic demand may be reduced, for example due to people abandoning outdoor activities.

Besides effects on the modal split and the total number of trips, there might be effects on the destination choices, route choices, and departure time choices as well. However, these effects are more diffuse (i.e. increasing the traffic demand for some locations and moments in time, while decreasing it for other ones), so that they cannot really be characterized in general terms. In the following they are discussed in slightly more detail.

The destination choices might be affected by a shift to activities that are less sensitive to the weather, insofar as these activities are performed at other locations. This might lead to the number of trips being increased on part of the origin-destination relations, while being decreased on another part of these relations. In general, these changes probably are rather limited however, except for typical leisure destinations like seaside resorts. In case of exceptionally adverse weather conditions, part of the travelers might opt for destinations closer to home. For many trips (like those between home and work) the destination obviously cannot be freely chosen however.

In general, the *direct* effect of adverse weather conditions on route choices is likely to be rather limited. (In really extreme weather conditions drivers obviously may be more inclined to divert from their normal route than in more normal adverse weather conditions, because routes might have become impassable, or might be considered too unsafe under the circumstances at hand. This will reduce the traffic demand on these routes, while increasing the demand on other ones.) The *indirect* effect of adverse weather conditions on the route choices probably is larger. By affecting the traffic supply (see section 2.2.5),

[13] In the case of public transport, this larger exposure is not related to the means of transportation themselves (as for the bicycle), but rather to the access and egress parts of the trip, which are at least partially to be covered on foot (or by bike).

[14] Actually, in these situations there also might be travelers switching from the car to the public transport, instead of vice versa.

adverse weather conditions affect the traffic conditions on the network (reflected in the levels of traffic congestion). Since road users might be informed on these 'nonstandard' traffic conditions (see also the intermezzo further below), this in turn may affect the route choices. Of course road users also might *anticipate* non-standard traffic conditions, based on the (forecasted) weather circumstances and their own experiences regarding the influence of such circumstances on the traffic conditions. It is difficult, though, to say something about the extent of these effects. Since the traffic conditions on alternative routes will often be affected by the adverse weather conditions as well, probably the route choice effects are limited.

As far as the departure time choices are concerned, adverse weather conditions might have a direct effect by making travelers postponing outdoor activities, resulting in the associated trips to be postponed as well. Another reason why trips might be postponed (or brought forward) is that people want to avoid driving in bad weather conditions. However, in particular during the peak periods this effect is not likely to be very large, unless (forecasted) weather conditions are really extreme (significantly reducing the safety, or causing a traffic chaos). Just like the route choices, the departure time choices might be affected by the adverse weather conditions in an *indirect* way as well: in view of the effects of the adverse weather on the traffic conditions, travelers might decide to depart earlier or later than originally planned. This decision is not necessarily based on information on the actual traffic conditions, but might be motivated by past experiences as well.

From a survey among commuters, conducted in Brussels, Khattak & De Palma (1997) found that road users have a higher propensity to change their departure time in response to adverse weather conditions, than to change their route. Possible explanations identified for this are that changing departure time might be easier, that the effect of changing departure time is better predictable, and that alternative routes are affected by adverse weather as well (so that route switching may not be beneficial). In another study it was found that commuters' propensity to change departure time or route is much greater during the a.m. peak than during the p.m. peak (Hranac et al, 2006). A possible explanation for this might be that the urge to arrive at one's destination in time is larger for the home-to-work trips than for the return trips.

So far, this section has concentrated on *adverse* weather conditions. However, *beautiful* weather conditions (i.e. sunny skies and high temperatures) may make traffic demands deviate from their 'representative' values as well. After all, recreational destinations will attract more visitors in such situations, resulting in extra traffic on the relations to these destinations. This effect is confirmed by a study conducted by Cools et al. (2008). They found relatively high correlations between the daily temperature maxima and the daily traffic intensities, in particular for a highway that is one of the access roads to the Belgian seashore (and consequently is typified by a large fraction of leisure traffic). For this road, correlation with the sunshine duration is relatively high as well.

Good weather conditions also might stimulate people to take the bicycle instead of the car (resulting in a decrease of the traffic demand), but this will only concern the shorter distance trips. For given car trips, route choices and departure time choices are not very likely to be significantly affected by summery weather conditions.

In literature, the quantitative effects of weather conditions on traffic demand are generally studied in terms of changes in daily traffic volumes on some selected road sections. Chung et al. (2005) studied the effect of rain on traffic demand measured on the Tokyo Metropolitan Expressway, and found that on rainy days traffic demand was lower. For weekdays an average reduction in the daily number of trips of about 3 percent was found. For weekend days, a much larger reduction was found. Chung et al. note that this difference can easily be explained by a difference in travel purposes of the road users. On weekdays, most of the trips are work-related. These trips are typically not very flexible. On weekend days however there are less work-related trips and more leisure-related trips. The latter are typically more flexible. Hanbali and Kuemmel found a similar difference in the reductions in hourly traffic volumes due to snowstorms[15]: on weekdays and especially during peak hours smaller reductions in hourly traffic volumes were observed (see Hranac et al, 2006).

Hogema (1996) investigated the effects of rain on traffic volume for a location on the Dutch A16 motorway. However, no significant effect on the total daily traffic volume was found. Therefore, Hogema concluded that, apparently, rain does not cause a major modality shift towards the private car. Changnon (1996) studied the effects of rain on daily traffic volume using data from toll highways in the Chicago area (daily numbers of vehicles entering the toll roads at each of three interchanges). The results revealed that rain on weekdays had no measurable effect, while for rainy weekends a 9% decrease was found.

Keay and Simmons (1995) analyzed data for two freeways in the Melbourne metropolitan area. In general, traffic volume appeared to be reduced on wet days. The larger the amount of rainfall, the larger this reduction tended to be. However, the effects were significant only for the winter and the spring (as well as for the year as a whole). For the winter period, a reduction in the daily volume of 1.35% was found, and for the spring period a reduction of 2.11%. Keay and Simmons also analyzed the effects separately for daytime and nighttime periods. For daytime rainfall reductions of 1.86% (winter) and 2.16% (spring) were obtained, while for nighttime rainfall reductions of 0.87% (winter) and 2.91% (spring) were found.

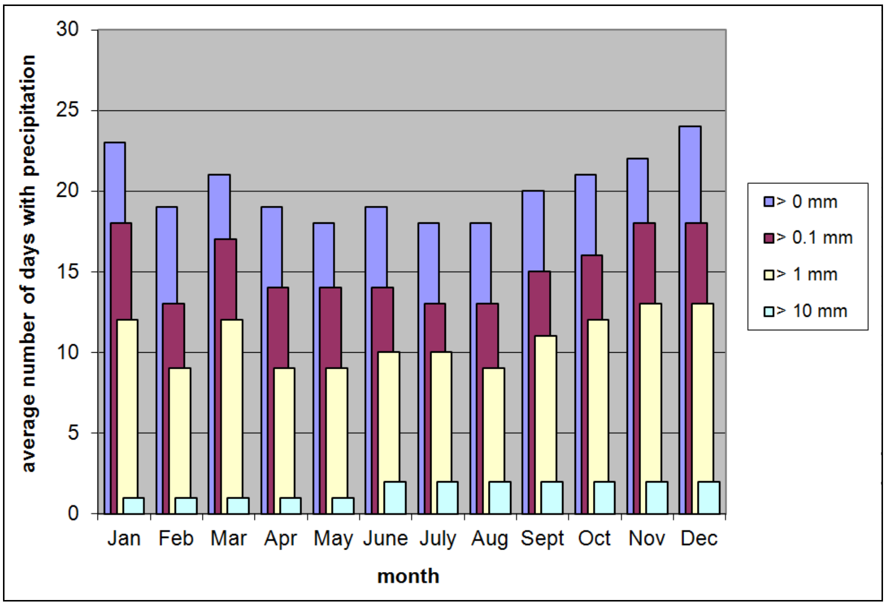
[15] Using traffic data from highways and freeways outside of urban areas in Illinois, Minnesota, New York, and Wisconsin (USA).

Cools et al. (2008) studied the effects of weather conditions on daily traffic intensities measured on four highway traffic count locations on the Belgium highway network. It was found that snowfall, rainfall and wind result in smaller traffic volumes, while high temperatures result in larger traffic volumes. The effects turned out to be rather different for different count locations. Cools et al. indicate that these differences can be (partially) explained by differences in the travel purposes of the road users. As mentioned previously, leisure activities are typically more flexible than work activities. Consequently, on roads with a lot of leisure traffic, traffic volumes will react more strongly to the weather conditions than on roads with a lot of commuter traffic. (Analogous to the previously discussed finding that in time periods with a relatively large share of leisure traffic larger effects on the traffic volumes are observed than in time periods with a lot of commuter traffic.)

- Frequencies of occurrence -

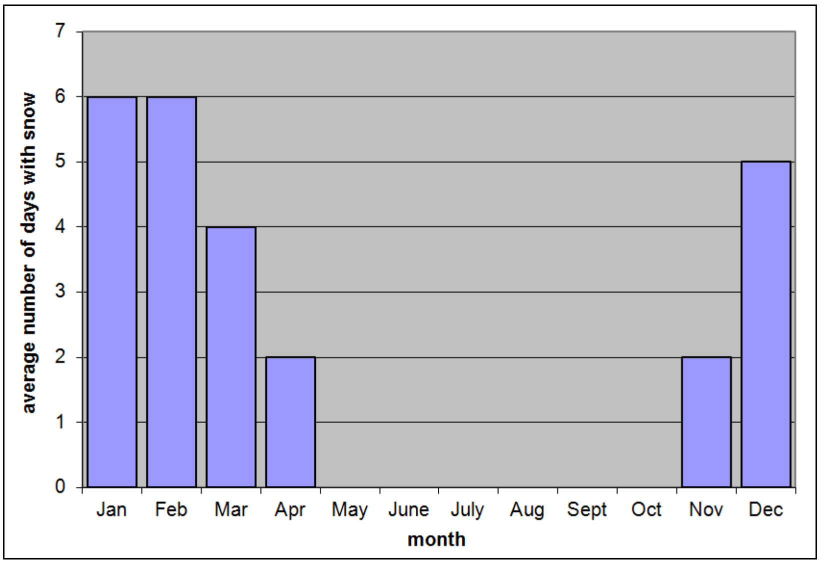
Besides the *effects* of different weather conditions on the traffic demands, of course their *frequencies* should be addressed as well. In the Netherlands, precipitation occurs on average about 7% of the time (KNMI, 2002). Since precipitation usually lasts only for a limited part of the day, the average percentage of *days* on which precipitation occurs is much larger though. Figure 2.11 shows the average numbers of days with precipitation for the various months of the year (for different threshold values). In autumn and winter days with precipitation clearly occur more frequently than in spring and summer. It should be noted however that the average intensity of precipitation in summer is relatively high (due to the occurrence of heavy showers). In terms of the total amount of precipitation, this partially compensates for the shorter precipitation duration.

Figure 2.11: Average numbers of days with precipitation per month in the Netherlands
(Based on data from KNMI, 2002)



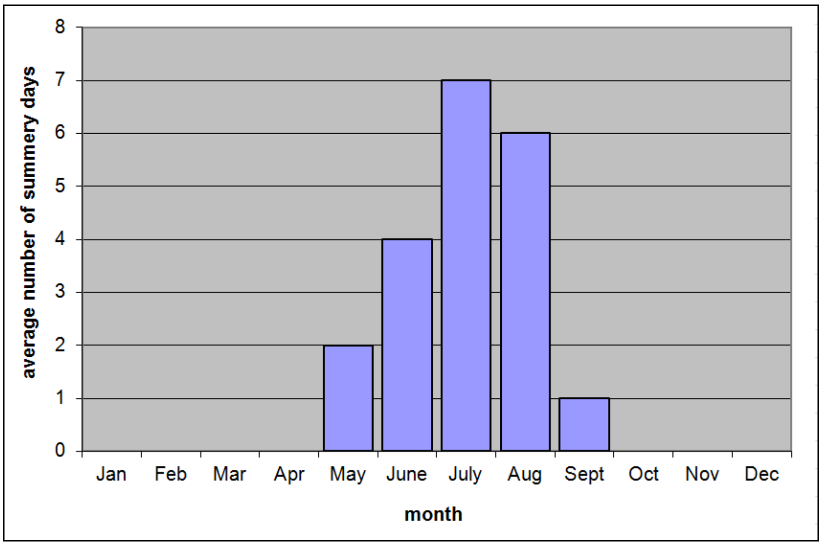
In the Netherlands, the largest part of the precipitation is rain. Only about 3% of it is snow (Buishand and Velds, 1980). On average, on 25 days a year snowfall is reported (see Figure 2.12). Mostly the amount of snow is limited however, and often it falls down in the form of sleet (i.e. snow that is melting away). On average, 19 days a year a (closed) snow cover is observed in the Netherlands. This number shows a large variation from year to year however (ranging from zero to a number much larger than 19) (KNMI, 2009b). Black ice occurs nearly every winter (November - mid March) one or more times, on average on about 2-5 days (KNMI, 2009b; Huiskamp, 2010).

Figure 2.12: Average monthly numbers of days with snow
(Based on data from KNMI, 2002)



The frequency of occurrence of beautiful weather conditions is dealt with by considering the number of summery days. According to the definition used by the KNMI (the Dutch national meteorological institute) these are days with a maximum temperature higher than 25°C. Figure 2.13 shows the average number of summery days for the different months of the year.

Figure 2.13: Average numbers of summery days per month in the Netherlands
(Based on data from KNMI, 2002)



7) Road works

- Effects -

The presence of road works might result in a reduction of the traffic demand for the road sections in question. This reduction can be effected in three ways:

- If the road works have been announced, part of the road users will be informed on them in advanced. In view of possible traffic congestion problems, these road users might decide to avoid using the road sections in question during the peak periods (by opting for another route, departure time, transport mode or destination, or even canceling their trip altogether), resulting in a decrease of the traffic demand.

  The road authority might also go one step further than just providing information in order to prevent traffic chaos from occurring. Examples of more active forms of demand management are rewarding systems (systems in which road users are rewarded for avoiding the affected road sections during the peak periods) and temporary reductions on public transport fares. The most drastic demand reducing measure applied in practice (except for the complete closure of the road) is the closure of on- and off-ramps in the work zone.

- If the road works last for several days (or even longer), road users might adapt their travel behavior based on their own experiences: if their travel times are increased due to the road works, they may switch to another route, departure time, transport mode or destination, or even decide to stay at home.

- Based on traffic information received before departure or during the trip (possibly reporting on traffic congestion caused by the road works), travelers might decide to avoid using the road sections in question. (See also the intermezzo below.)

Of course on other routes (or on the same route, but at other moments in time) just the opposite effect might occur: on these routes (or moments in time) traffic demand might be increased, due to diverting traffic.

The magnitude of the demand effects of road works obviously will be dependent on the specific situation at hand (i.e. the remaining capacity, relative to the traffic demand under normal circumstances; the duration of the road works; the availability of alternative travel options; the traffic and demand management measures taken by the road authority; etc.). An example of the magnitude of the effects can be found in (AVV, 2002c), which discusses the traffic effects of a large-scale road maintenance project on a part of the beltway of Amsterdam (the A10-West), carried out in the summer of 2001. Beforehand, it was expected that the road works would cause a traffic chaos. However, this traffic chaos did not materialize, largely due to the fact that the traffic volume on the A10-West was reduced by 38%.

This reduction in traffic demand can mainly be attributed to the closure of on- and off-ramps in the work zone, and the intensive information campaign. Especially local traffic (i.e. with origin and destination within the area around the A10-West) changed its behavior. The most important change that was observed in the travel behavior was the 'forced' choice for another route. This resulted in increased traffic volumes on surrounding motorways, and particularly in a sharp increase of the traffic volumes on the surrounding on- and off-ramps and the urban network. On the surrounding motorways the usual congestion was not substantially increased, but on the urban network a lot more traffic congestion did occur. Only 10% of the road users switched to another transport mode (one half of which to the public transport, and the other half to the bike), in spite of the measures taken to stimulate this. It turned out that the road works did not result in less trips being made. (The total amount of trips was reduced by a few percent, but this reduction was due to the vacation period and public holidays.)

A large part of the road users changed their departure time during the road works (in the morning peak 60% and in the evening peak 70%). However, the average changes were limited to a few minutes only. In the morning peak the average departure time was 5 minutes earlier, and in the evening peak 8 minutes. According to Scholtens (2001) the peak in the traffic demand was somewhat leveled off in the period of the road works.

Another example of the magnitude of the effects of road works on the traffic demand can be found in (AVV, 2006), which gives an evaluation of the effectiveness of the mobility management measures that were taken to reduce the (peak period) traffic demand during the large-scale road maintenance project on another part of the motorway network around Amsterdam (the A4 and A10-South), that took place in the summer of 2006. These measures included the provision of a public transport card to people working in the region, and the implementation of an extensive media campaign.

In this case the traffic demand in the morning peak was reduced by 8 to 15%, and the traffic demand in the evening peak by 5 to 9%. The contribution of the mobility management measures to this reduction was substantial (5%). Among the commuters that had received a public transport card, the modal share of the public transport was increased from 23% to 43%. Among the cardholders that continued to use their car during the road works, a substantial shift in route choice was observed. The group of commuters that uses the A4 and/or A10-South four or more times a week was reduced by a quarter in the period of the road works. The road works and/or provision of the public transport card did not result in unambiguous changes in the departure times of the travelers.

In (Scholtens, 2001) it is stated that a rule of thumb for road works on major roads is that the traffic volume is reduced by half. One half of this reduction is due to road users diverting to parallel roads. Usually it remains unclear where the other half went.

Note that all figures mentioned in the above are related to *large-scale* road works. In case of smaller road works (with a shorter duration and/or a smaller effect on the available capacity) obviously the demand effects will be more limited. In case of emergency repairs, road users will not be able to anticipate the road works (by adjusting their travel choices), unless they receive some traffic information on these before departure or during the trip.

- <u>Frequency of occurrence</u> -

Without information on their frequency of occurrence (and duration), the effects of road works on the traffic demand of course do not say much about the actual impact on the traffic conditions. For these aspects (i.e. frequency of occurrence and duration), the reader is referred to the subsection on the effects of road works on the traffic *supply* conditions (section 2.2.5).

▪ ▪

▸ **Intermezzo: the influence of traffic information and past travel experiences**

Although they are not really sources of variability themselves (as explained in section 2.2.1), traffic information and past travel experiences certainly play a role in the demand effects of 'real' sources of variability, as was already shortly mentioned in the discussions on the demand effects of varying weather conditions and road works. Therefore, in this intermezzo these factors are considered in some more detail.

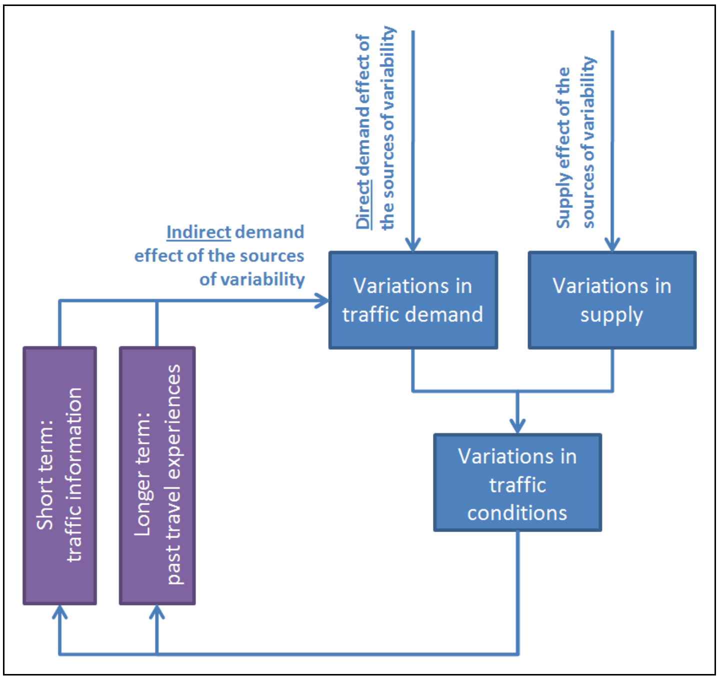
First of all the influence of traffic information is considered. Based on information regarding the actual traffic conditions (obtained from for example the internet, the radio, a real-time navigation system or a roadside variable message sign), travelers can adapt their travel choices to these conditions, which will affect the traffic demand. In the Netherlands, about 45% of the users of the main road network regularly use traffic information before departing. During the trip this percentage even amounts to about 55% (AVV, 2006b). Of course, the *use* of traffic information in itself certainly does not necessarily mean that one adapts one's travel choices. This will only be the case if the traffic information gives sufficient reason for this (that is, if the actual traffic conditions differ sufficiently from those initially expected). Furthermore, the purpose of using traffic information might also be to just reduce one's uncertainty regarding the traffic conditions that one will face, rather than enabling a reconsideration of one's travel choices.

It should be noted that in subsection 2.3.3 the effect of traffic information is shortly returned to, insofar as the influence on route choice is concerned. That subsection deals with the route choice effect of traffic congestion (one of the network effects of traffic congestion, which are discussed in section 2.3), in which traffic information obviously plays an important role.

Besides traffic information, road users' past travel experiences may make them adapt their travel choices as well (again affecting the traffic demand). If a road user experiences an above average travel time on a certain day, he might for instance decide to take another route on the next day. In view of the fact that in travel behavior habit formation plays an important role, it is doubtful however whether this process is really significant. If the experienced traffic conditions have a *specific cause* though, adjustments in the travel behavior are much more likely. For example, if a road user experiences traffic congestion due to a newly started road maintenance project, it is not unlikely at all that this road user will adapt his habitual travel behavior, until the road works have finished.

As illustrated in Figure 2.14, the effects of traffic information and past travel experiences in fact form a feedback mechanism from the variability in the traffic conditions to the variations in traffic demand (at two different timescales). Due to this mechanism, the final effect of the sources of demand variability actually may consist of two components: a direct component and an indirect one.

Figure 2.14: The influence of traffic information and past travel experiences



8) Strike actions

On days with large-scale strike actions, traffic demands in the peak periods might be substantial lower (due to a reduction in the amount of commuters). The size of the effect on the demand is strongly dependent on the scale of the strike actions and the professional groups involved. Furthermore, the size of the effect may vary from region to region.

A special situation occurs if the strike is in public transport. In this case, traffic demands are likely to be higher instead of lower, because of travelers switching from a public transport mode to the car.

The frequency of occurrence of large-scale strike actions however is rather low. Therefore, they cannot really be considered a source of daily variability in traffic congestion. For this reason, the influence of strike actions will not be given any further consideration in this thesis.

9) Emergencies

(Imminent) emergencies or disasters, like floods, terrorist attacks or chemical accidents, might lead to large streams of refugees. In situations in which an evacuation plan is put into effect, this process will take place in a more orderly fashion than in other situations. In both cases however traffic demands will be very large. In the rest of this thesis no further consideration will be given to this kind of situations. The focus of this research project after all is on *daily* traffic congestion. Because of the very small frequencies/probabilities of occurrence, traffic congestion due to emergencies falls outside this scope.

10) Other variations in human travel behavior

The factors discussed above do not fully explain the variations in traffic demand. Ultimately the variations in traffic demand are to a large extent the result of variations in peoples' activity patterns. These variations typically cannot completely be explained by a limited set of factors (no matter how large this set is chosen). As a result, the variations in traffic demand cannot be fully explained either.

From some literature referred to in (Weijermars, 2007) it can be concluded that, *relatively speaking*, the remaining (i.e. unexplained) variation is probably not very large:

- in a study on the relation between weather conditions and daily traffic volumes, it turned out to be possible to explain 95% of the variation in daily traffic volumes on two Australian freeways using a linear regression model incorporating trend, day of the week, holidays, and weather effects.
- in a study on the classification of traffic data time series, it was found that 75% of the variation in hourly traffic volumes in Vienna could be explained by differences between measurement locations, time of day, type of day, season, and weather conditions, while the residual 25% of the variation can be partly attributed to road works.

As was already noted before, various studies show that the (day-to-day) demand variability is larger for weekend days than for weekdays. This larger variability can be explained by the fact that people have less fixed activity patterns on weekend days. Due to the larger flexibility in the activity patterns, there is also a larger sensitivity to variations in the weather conditions, as discussed before.

While the unexplained variation from a *relative* perspective may be limited, from an *absolute* perspective it might still be an important factor however. This is due to the fact that the relationship between traffic demand and traffic congestion (expressed in for instance travel times) is not linear. A small difference in traffic demand may just make the difference between the traffic remaining free flowing and the traffic flow breaking down to a congested state (resulting in a queue that might continue to grow for some time).

2.2.5 Sources of temporal fluctuations in the supply

1) Adverse weather conditions

- Effects -

Adverse weather conditions (rain, snow, black ice, fog) may seriously affect the traffic supply, by reducing driving speeds and road capacity. Snow and black ice can result in a slippery road surface. To a more limited extent, rainy weather conditions can have a similar effect. Furthermore, fog and precipitation (especially snow) can significantly reduce visibility. Low sun can affect the visibility as well. Drivers deal with these effects by reducing their speeds and keeping larger headways. This results in road capacities being lower in this kind of circumstances. The capacity might also be reduced due to a reduction of the number of available lanes: snow, black ice or flooding may make one or more lanes (or even an entire road section) impassable.

On roads with a rush-hour lane (i.e. an extra lane which is only used during periods with a high traffic demand), weather conditions might also affect the capacity by prohibiting the opening of this additional lane. If visibility is bad (like during fog), it cannot be verified – using the cameras installed for this purpose – whether there are no obstacles on this lane. As a consequence the rush-hour lane has to remain closed for traffic, resulting in the capacity to be lower than usual for the traffic demand level at hand. Also in situations with snow or black ice on the road surface it might be necessary to keep the rush-hour lane closed.

Of course the actual size of the reductions in speed and capacity is dependent on the type of weather conditions and their intensity. In addition, it might be dependent on the type of road surface. The more porous the asphalt is, the better it drains. This results in a reduction of splash and spray (improving visibility) and a reduction of the road slipperiness (reducing the risk of hydroplaning). This might result in the speed and capacity reductions during rainy weather to be smaller.

Since weather conditions might have various spatial scales (ranging from a large rain belt to a local shower) their effects on the traffic supply can be either network-wide or local.

There is quite a lot of literature on the effects of weather conditions on speed and capacity, based on empirical investigations. The results of these studies are difficult to compare however, since:

- The investigations are performed for different regions across the world. Consequently there might very well be differences in the characteristics of the adverse weather conditions and the road users' response to these conditions. Here factors like the familiarity with the weather conditions, the drainage quality of the road surface, and the quality of snow clearing activities play a role (Hranac et al, 2006).
- There are methodological differences between the various studies, like differences in the time interval that is chosen for the analysis and differences in the classification of the weather conditions (i.e. the distinction between 'good' and 'bad' weather conditions and a possible further breakdown of the latter).

Furthermore, not all results are very reliable/accurate, due to:

- the use of rather limited data sets (i.e. data for a limited number of time intervals and/or a limited number of locations),
- the use of weather data measured at a location at a quite a distance from the road (possibly resulting in errors in the weather conditions used in the analyses), and
- the use of too long time intervals (for which the weather conditions cannot accurately be characterized anymore by single values for the time interval as a whole[16])

Quantitative effects of rain

On the basis of other literature, in (AVV, 2002) it is proposed to use the following reduction percentages for the (free flow) capacity during rainy weather (as compared with the capacity under 'ideal' conditions): 8% in case of closed asphalt, and 5% in case of very porous asphalt. More recently, Brilon et al. (2005) derived the capacity distribution for a large set of German freeway sections (using empirical data), separately for dry road surface conditions and wet road surface conditions. For all sections a very clear reduction in capacity of around 11% was found for wet conditions (as compared with the capacity under dry conditions).

Smith et al (2003) made a distinction between light rain (0.25 to 6.35 mm/h) and heavy rain (> 6.35 mm/h). Based on empirical data for two freeway links in Virginia (United States), they found capacity reduction factors of 4% to 10% for light rain and 25% to 30% for heavy rain. Agarwal et al (2005) conducted an extensive study on the freeway network of Minneapolis and St. Paul (including several interstates and trunk highways built to freeway design standards). For light rain (0.25 to 6.35 mm/h) they found more or less the same result as Smith et al: a reduction of 5% to 10%. For heavy rain however they found a smaller reduction than Smith et al: 10% to 17%. Hranac et al (2006)

[16] Consider for example the fact that a short but very heavy shower and a prolonged rainfall with a modest precipitation rate might both yield the same precipitation measurement over a time interval of one hour, which implies that a time interval of one hour is in fact too long.

conducted an even more extensive study on freeways (or highways) in three major metropolitan areas in the United States (Seattle, Baltimore, and Minneapolis–St. Paul). They found a constant capacity reduction of 10% to 11% for rainy weather, independent of the rainfall rate (within the range of 0 to 17 mm/h). Chung et al (2005) analyzed the impact of rainfall of various intensities on the capacity of a number of sections of the Tokyo Metropolitan Expressway. They found a capacity reduction of 4% to 7% for light rain (1 mm/h), increasing up to 8% to 14% for heavy rain (10-20 mm/h).

For the (uncongested) speed reductions due to wet road conditions, Brilon and Ponzlet (1996) found values of about 9.5 km/h on two-lane roadways and 12 km/h on three-lane roadways[17]. It should be noted however that these values were established for a set of German autobahns without a speed limit. In situations with a speed limit the reductions can be expected to be smaller. Nevertheless, Hogema (1996) found a rather similar result for a motorway *with* a speed limit (the Dutch A16): in rainy weather the mean speed turned out to be 11 km/h lower than in dry conditions[18].

In the same research project as referred to above, Smith et al (2003) found reductions of 5 to 6 km/h (corresponding to 5% to 6.5%) for both light and heavy rain. Argawal et al (2005) found reductions of the same order of magnitude. Hranac et al (2006) made a distinction between the effect on the *free speed* and the effect on the *speed-at-capacity*. The effects were found to generally increase with rain intensity. For the *free speed* the reduction factor was found to increase from 2%-3.6% for light rain (< 0.1 mm/h) up to 6%-9% at a rain intensity of approximately 16 mm/h. For the speed-at-capacity the reduction factor was found to increase from 8%-10% for light rain up to 8%-14% at a rain intensity of approximately 16 mm/h. In their investigation of the weather effects on the performance of the Tokyo Metropolitan Expressway, Chung et al (2005) calculated the reduction in the median free flow speed (for flow rates ≤ 500 veh/h/lane) for various rainfall rates. This resulted in a reduction factor of 4.5% for a rainfall rate of 0-1 mm/h, increasing up to 8.2% for a rainfall rate of 5-10 mm/h (relative to a median free flow speed of 77.7 km/h under dry conditions).

Quantitative effects of snow

Obviously, snow can have very detrimental effects on the (uncongested) speeds and the capacity. The magnitude of the effects, however, will strongly depend on the snowfall rate, on whether it falls in the form of sleet or in the form of 'dry' snow, and on the extent to

[17] Using these reductions in (uncongested) speeds, Brilon and Ponzlet also estimated the reductions in capacity. For two-lane roadways a capacity reduction of 350 veh/h was found, and for three-lane roadways a capacity reduction of more than 500 veh/h, both corresponding to roughly 10% of the capacity under dry conditions.

[18] It should be noted that in his analysis, Hogema only considered situations with relatively low traffic volumes (well below capacity).

which the road authority succeeds in keeping the road clear from a snow cover. The Highway Capacity Manual (TRB, 2000) points at the fact that if snow-clearing operations cannot keep the road reasonably clear, the snow accumulation will obscure the lane markings. According to the HCM, observation suggests that drivers often seek not only larger headways in that situation, but also larger lateral clearances. This results in a three-lane roadway being used as if it had only two (widely separated) lanes. Obviously, this effect alone already results in a significant reduction of the capacity.

Table 2.6 shows the reductions in capacity and uncongested speeds for various snowfall rates, as found by Agarwal et al (2005). Larger snowfall rates clearly have a larger impact. Remarkably, Hranac et al (2006) found that the capacity reduction is *not* dependent on the snow intensity [19]. They found a value of 12% to 20% for this capacity reduction. In contrast to this capacity reduction, the reductions in speed were found to typically increase with snow intensity. For both the free speed and the speed-at-capacity the reduction factor was found to range from 5%-16% (light snow: 0.1 mm/h [20]) up to 5%-19% (at a snow intensity of approximately 3 mm/h [20]). As reflected by the large bandwidths, the results for the different metropolitan areas were rather different however.

Table 2.6: Reductions in capacity and uncongested speeds for various snowfall rates
(Source: Agarwal et al, 2005)

Snowfall rate (mm/h)	Capacity reduction	Uncongested speed reduction[21]
0 – 1.3	3%-5%	3%-5%
1.5 – 12.7	6%-13%	7%-10%
> 12.7	19%-27%	11%-15%

Quantitative effects of other weather conditions

Agarwal et al (2005) also considered the effects of temperature, wind (in the opposite direction of travel) and reduced visibility (due to fog events) on capacity and (uncongested) operating speeds. No significant effects were found, however, for wind and temperature (except for temperatures below -20°C). For reduced visibility conditions (corresponding to a visibility below one mile[22]) a capacity reduction of 10% to 12% was found. The (uncongested) operating speeds were found to be reduced by about 7% for visibilities between 0.4 and 1.6 km, and by about 12% for visibilities below 0.4 km.

It is conceivable that low sun conditions can affect the capacity and operating speeds as well. After all, depending on the orientation of the road, the visibility can be seriously reduced if the sun is low in the sky. Nevertheless, there is little literature on the influence of this phenomenon.

[19] This however might be due to the fact that in their analyses, Hranac et al. considered visibility as a separate factor, concurrently with the snowfall rate (or rainfall rate). Maximum capacity reductions in the range of 10% were observed for situations with reduced visibility.

[20] expressed in the *liquid-equivalent* precipitation rate

[21] average reference speed: 107 km/h

[22] equivalent to 1.6 km

Accounting for the quantitative effects in the fundamental diagram

Based on the quantitative impacts on capacity and speeds (and some additional information, such as the observation that the jam density is not affected by the weather conditions: see Hranac et al, 2006) one can construct a fundamental diagram for specific adverse weather conditions. In Figure 2.15 two examples of this are shown (relating to the Minneapolis–St. Paul area, USA), taken from (Hranac et al, 2006). In the left half of the picture a comparison of the fundamental diagrams for clear weather conditions and the 'worst' rainy conditions (i.e. for the highest rainfall rate considered by Hranac et al.) is shown. Similarly, in the right half of the picture the fundamental diagram for clear weather conditions is compared with the one for the 'worst' snowy conditions.

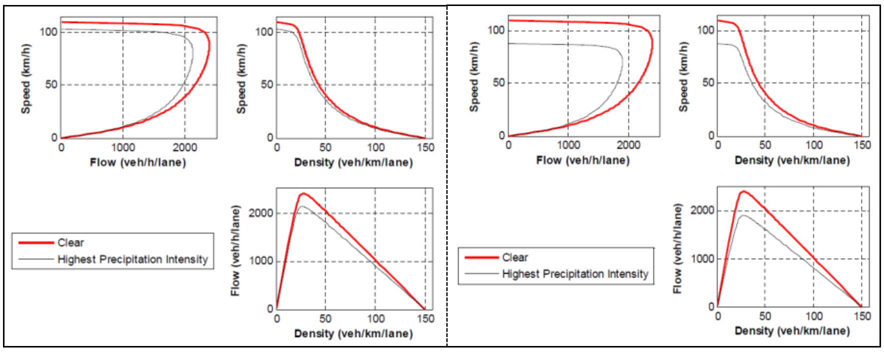


Figure 2.15: Comparison of the fundamental diagram for clear weather conditions with its counterpart for rainy weather (left) or snowy weather (right)
(Source: Hranac et al, 2006)

- <u>Frequencies of occurrence</u> -

Without data on the frequencies of occurrence of the bad weather conditions, the reductions in speed and capacity in these bad weather conditions of course do not say much about their actual impact on the traffic conditions. However, in the discussion on the *demand* effects of variations in weather conditions (see section 2.2.4) already a lot of information has been given on these frequencies, which will not be repeated here. Some additional information relating to precipitation rates is given however, because of their influence on the magnitude of the effects on capacity and (uncongested) speeds. Also, some information on the frequency of occurrence of fog is added.

In the Netherlands, the average yearly amount of precipitation is nearly 800 mm. Given the fact that, on average, it rains about 7% of the time, one can compute that the average precipitation rate amounts to about 1.3 mm/h. Considering time intervals of 15 minutes, on average 10 times a year a precipitation rate of more than 12 mm/h occurs (KNMI, 2009d). If time intervals of 60 minutes are considered, the precipitation rate that (on average) is exceeded 10 times a year equals 5 mm/h.

There are some differences in rainfall between the various periods of the day, in particular in the inland parts of the country. During the late afternoon and the early evening the amount of precipitation might be up to several tens of percents higher than during the morning or night, in these parts of the country (see Buishand and Velds, 1980).

The probability of fog is largest in the period just before sunrise (KNMI, 2008). After the sunrise the probability gradually decreases, and during the nighttime it gradually increases again. Over a whole year, there are about 40 'fog days' in the Netherlands (KNMI, 2009a). This however does not mean that traffic is impeded on all of these days. A day is considered to be a 'fog day' already if in one of the 24 hours of this day visibility is below 1 km. Traffic is only hindered if visibility is reduced to below 400 m. Since there are about 15 days in a year on which a visibility below 200 m is reported, and about 35 days on which a visibility below 500 m is registered (see KNMI, 2009), the yearly number of days with a visibility below 400 m will be somewhere in between. In the Netherlands, the frequency of fog is largest in the period between October and January (KNMI, 2008). In the spring and summer there are fewer fog days. Furthermore, the fog dissolves faster in the summer. In the autumn and winter sometimes the fog remains all day.

2) Luminance

- Effects -

Besides the weather conditions, luminance conditions affect the traffic supply as well. Darkness may result in a (limited) reduction of driving speeds and capacity, as compared with the situation during daylight. AVV (2002) states that in research studies an average capacity reduction of 5% is found (in otherwise ideal circumstances). It should be noted, however, that this reduction is strongly dependent on location. In case of street lighting the decrease in capacity might be lower.

More recently, in the same study as referred to earlier, Brilon et al. (2005) found contrasting results. They clearly found that darkness does not shift the capacity distribution of German freeway sections. On the other hand, Chung et al (2006) found the capacity during daybreak (i.e. in poor natural lighting conditions) to be 12.8% lower than during daylight. This effect cannot be attributed to drivers being blinded by the sun in the daybreak period, since they were not facing the sun. This result was obtained from data from only one location however, making it impossible to draw firm conclusions. Moreover this was a location on the Tokyo Metropolitan Expressway. In all likelihood the situation on German autobahns is more representative for the situation on the Dutch motorways than the situation on a Japanese expressway.

As far as the effect on the uncongested speeds is concerned, Brilon and Ponzlet (1996) found that darkness causes an average reduction of about 5 km/h. It should be noted however that this result was obtained for German autobahns without a speed limit. Obviously, in situations *with* a speed limit the reduction could be smaller.

- <u>Frequency of occurrence</u> -

Based on the sunset and sunrise times for all days of the year, Meeuwissen et al (2004) calculated the relative frequencies of darkness for the peak periods and the off-peak periods in the Netherlands. For the peak periods (7:00 - 9:00 plus 16:00 - 18:00) a value of 22% was found, and for the off-peak periods (6:00 - 7:00, 9:00 - 16:00, and 18:00 - 24:00) a value of 35%.

3) Road works

- <u>Effects</u> -

Road works create variability in the traffic supply characteristics in two ways:

- During road works the speed limit is temporarily lowered over the length of the work zone. Usually the speed limit is set at 70 km/h. Sometimes speed limits of 90 km/h or 50 km/h are imposed, however.
- During road works, the capacity of the road section in question is reduced. This can be due to a reduction in the number of available lanes and a less efficient use of the (remaining) lanes. This less efficient use of the (remaining) lanes, typically reducing their capacity by roughly 20 to 40%, might be due to a reduction in the width of the lanes, shifts in the course of the roadway, the lowered speed limit (if lowered to 50 or 70 km/h), and the distracting effect of the ongoing activities. As a result of this effect, the capacity is reduced even when only the hard shoulder is closed.

Obviously, road works only affect the traffic supply characteristics on the road sections on which the road works take place. This is in contrast with the earlier discussed effects of darkness and adverse weather conditions, which are (or might be) network-wide. The road works might be limited to one single location, or cover an entire road stretch. The speed effect will only affect road users' travel times to a significant degree in the latter case. The capacity effect however might have very detrimental consequences already if the road works are limited to only one single location.

A general value for the capacity effect of road works cannot be given. This is due to the fact that work zones are found in many different configurations, ranging from a simple closure of the hard shoulder to settings in which traffic in one of both directions is diverted over the roadway in the opposite direction. Occasionally also full road closures occur (mostly during weekends or during the night). The capacity reduction is not only dependent on the configuration of the work zone though. Other factors involved are the width of the (remaining) lanes, the speed limit, the duration of the road works (because of the process of habituation), and the type and intensity of the activities in the work zone (in connection with attention distraction). In (AVV, 2002) for a large set of lane configurations and width reductions, values for the reduced capacities are given, measured or estimated for work zones on

Dutch motorways. In combination with a database containing information on all road works on the Dutch motorway network (insofar as these have been registered), it would in principle be possible to construct an approximate [23] relative frequency distribution of the capacity effects of motorway road works.

- Frequency of occurrence -

The Dutch road authority strives to minimize the impacts of road works on the traffic conditions. Therefore the vast majority of the road works are carried out during the evening and night periods. Only a very limited part is executed during the peak periods, and even then mostly on days or road stretches on which the traffic demands are not very high (Rand Europe, 2004). However, it is of course not always possible to avoid carrying out road works during busy periods or on busy locations. Some types of (major) road works simply take a rather long period, which almost inevitably will contain certain periods with high levels of traffic demand as well. Furthermore, sometimes emergency repairs need to be carried out, which cannot wait until the traffic demand is lower.

Figure 2.16 shows the frequency distribution of the starting times of all reported road works on the Dutch main road network in 2002. In Figure 2.17, the frequency distribution of their duration is shown. If information on the starting time of the road works is combined with information on their duration, a histogram can be constructed of the numbers of road works that have been going on during the different hours of the day. This histogram is shown in Figure 2.18. It should be noted that the road works that took longer than 24 hours are excluded from the histogram. Expressed in *numbers*, these road works form only a few percent of the total amount of road works. Because of their long durations, they form a rather large part of the total duration of road works though. Note that including these road works in the histogram would not alter the absolute differences in the histogram, however. For all time intervals (i.e. hours of the day) the frequency would be raised by about the same amount (which is roughly estimated at around 10,000).

Besides the start time and duration of the road works, of course also the length of the work zones is of importance. Information on the length of work zones can be found in the database mentioned above. Probably there is a rough relationship between the length and the duration of road works. On average, a longer length will correspond to a longer duration.

[23] Note that this distribution would only be an approximation of the 'true' distribution, since it would not be possible to fully account for the influences of the duration of the road works, the speed limit, and the intensity and type of the ongoing activities. Moreover, the unregistered part of the road works might introduce a bias in the calculated distribution.

Figure 2.16: Frequency distribution of the starting time of the road works reported over 2002
(Source: AVV, 2003)

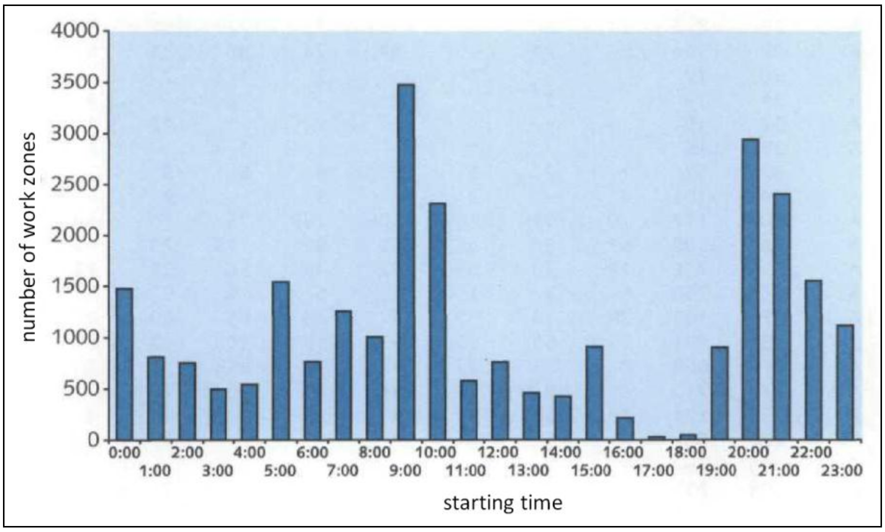


Figure 2.17: Frequency distribution of the duration of the road works reported over 2002
(Source: AVV, 2003)

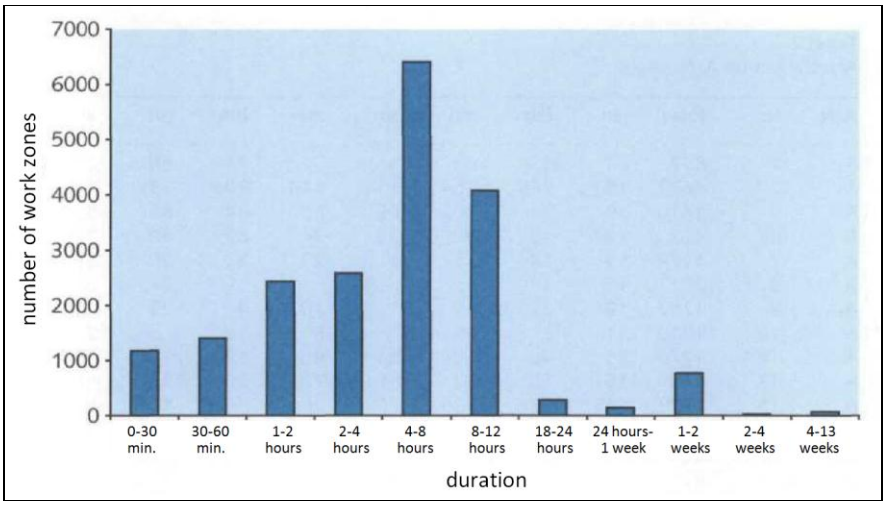
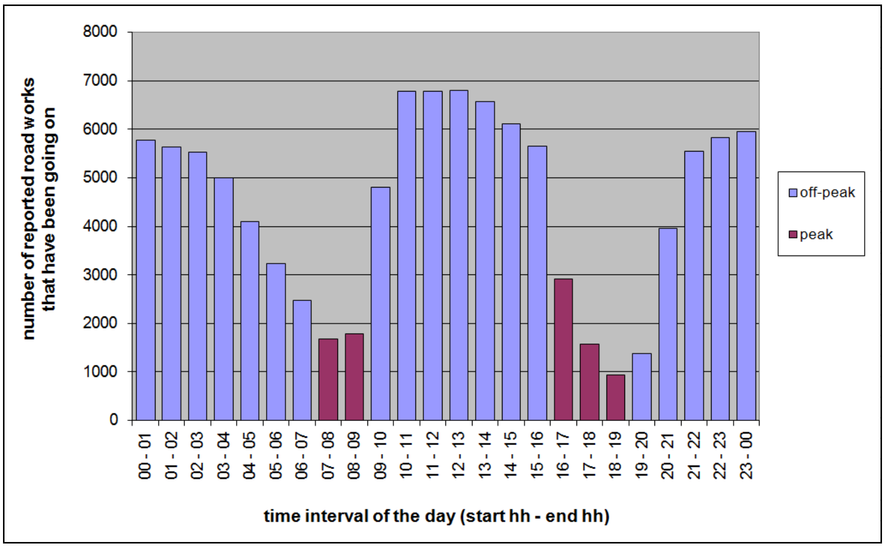


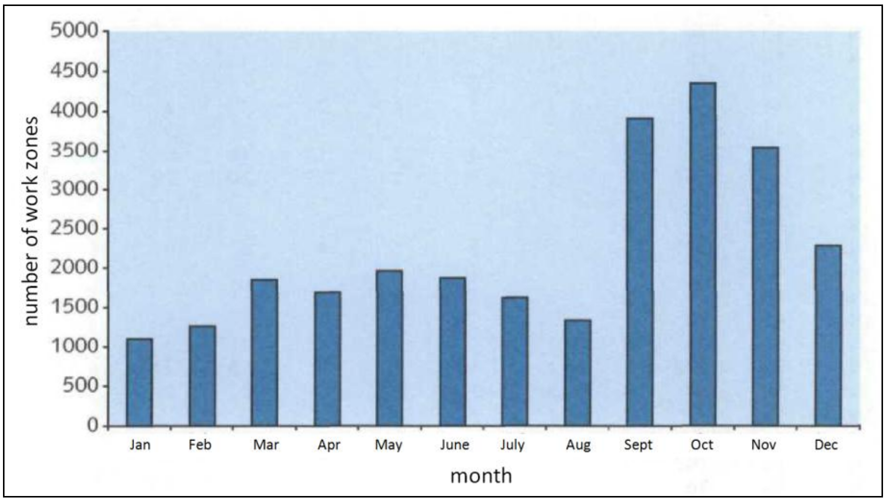
Figure 2.18: Numbers of road works reported over 2002 that have been going on during the different hours of the day
(Approximation, based on data from AVV, 2003)[24]



[24] Only including the road works that started between Monday 0:00 hours and Friday 19:00 hours, and excluding the road works that took longer than 24 hours.

In Figure 2.19 the distribution of the reported road works over the months of the year is shown. Apparently, both in the summer period and in the winter period there is a dip in the number of road works carried out. The dip in the summer period probably is related to the vacation period. The dip in the winter period might be related to the weather conditions. In the autumn the number of road works appears to be relatively large. In 2001 the monthly numbers of road works were highest in the autumn as well, although the difference with the other months was smaller (AVV, 2002b).

Figure 2.19: Number of reported road works per month in 2002
(Source: AVV, 2003)



4) Incidents

- Definition -

In this thesis incidents are defined as unpredictable short-term events (with durations ranging from a few minutes to a few hours) that affect traffic operations by strongly reducing the available capacity and/or the (uncongested) traffic speed. There a many different types of incidents: accidents, vehicle breakdowns, cargo spills, technical malfunctioning of roadside equipment (like traffic control devices, bridges, and tunnel safety systems), oil spills, roadside fires, tunnel closures triggered by vehicles exceeding the height restriction, events alongside the road that distract drivers' attention from the driving task, etc. Please note the difference between *incidents* and *accidents*. Accidents are just one of many categories of incidents.

Incidents occur such frequently in practice that they certainly can be considered to be contributing to daily traffic congestion. On routes of 30 kilometers or more, on average once every five days the travel time is influenced by an incident (Transpute, 2003). In 2000, around 21% of the total number of lost vehicle hours was caused by accidents (Kouwenhoven et al, 2006).

- Effects -

Incidents affect the traffic operations by locally reducing the available capacity and/or reducing the traffic speed, during a certain amount of time. Since the reduction in traffic speed occurs only locally (not counting the speed reductions due to the emergence of traffic congestion, caused by the reduction in capacity), its effect on (origin-destination) travel times is very limited. The reduction of the available capacity might have a much larger effect. Therefore, this reduction in capacity is discussed in more detail below.

Obviously, incidents can physically affect the available capacity by blocking one or more of the lanes (or even the complete roadway). However, the capacity is reduced as well due to the (remaining) lanes being used less efficiently during incident situations. This is the result of the driving behavior being different than during normal (i.e. non-incident) conditions. Probably to a large extent this difference in driving behavior can be attributed to the fact that incidents divert drivers' attention away from the driving task. This results in drivers reducing speeds and increasing headways, by which the capacity is reduced.

Incidents do not only affect the capacity of the roadway in question. The capacity of the roadway in the opposite direction is affected as well. Since there is no *physical* capacity reduction in this direction, in this case the capacity reduction can be fully attributed to diverted attention (i.e. drivers trying to see what is going on).

On both roadways most likely both the *free flow capacity* and the *queue discharge rate* are affected. Knoop (2009) states that to the best of his knowledge there has been no research into the effect on the free flow capacity. The reduction in queue discharge rate, however, has been investigated in several studies.

In the US 'Traffic Incident Management Handbook' (PB Farradyne, 2000) a table is included that shows the percentage remaining capacity under a variety of incident conditions (ranging from shoulder disablement to three lanes blocked) for freeways with various numbers of lanes (Table 2.7). It should be noted, however, that this table is based on a more encompassing definition of 'incidents'. Here also unplanned work zone activities are considered to be incidents, while in this thesis these are considered separately. For unplanned work zone activities different capacity reduction factors may apply than for other incidents. Furthermore, it should be noted that the capacity reduction factors in the Netherlands might have different values than those found in the United States. The table clearly illustrates that an incident reduces the roadway capacity by an amount far greater than the physical reduction in the number of available lanes.

Table 2.7: Fraction of freeway capacity available under incident conditions
(Source: PB Farradyne, 2000)

Number of Freeway Lanes in Each Direction	Shoulder Disablement	Shoulder Accident	Lanes Blocked		
			One	Two	Three
2	0.95	0.81	0.35	0.00	N/A
3	0.99	0.83	0.49	0.17	0.00
4	0.99	0.85	0.58	0.25	0.13
5	0.99	0.87	0.65	0.40	0.20
6	0.99	0.89	0.71	0.50	0.25
7	0.99	0.91	0.75	0.57	0.36
8	0.99	0.93	0.78	0.63	0.41

Based on empirical data, Knoop (2009) determined the reduction in queue discharge rate for 90 incidents on a few Dutch motorway stretches (all of them with 3 lanes in each direction). The analysis was limited to accidents that resulted in one or more (driving) lanes being closed and vehicle break downs that resulted in the hard shoulder being occupied. For each incident, Knoop calculated the ratio of the queue discharge rate during that incident and the queue discharge rate at the same location in normal conditions (referred to as the 'capacity factor'). Subsequently, for all four incident types the mean and the standard deviation of this ratio were calculated. Table 2.8 shows the results that were obtained. As explained before, the reductions in queue discharge rate are the combined result of a reduction in the number of available lanes and a less efficient use of the remaining lanes. Factors expressing this reduced efficiency are included in the table as well.

Table 2.8: Remaining capacities under incident conditions
(Data from Knoop, 2009)

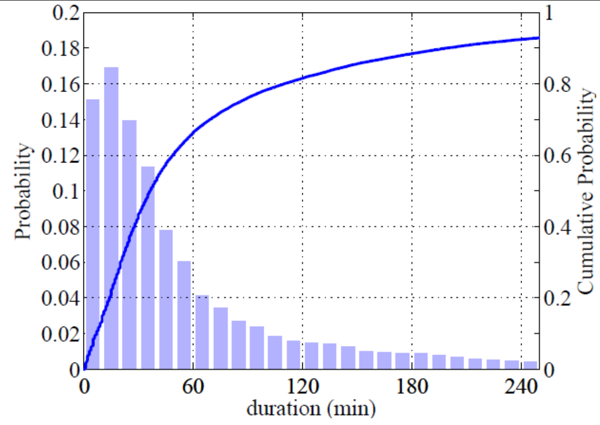
Incident type	Broken down vehicle on hard shoulder	1 out of 3 lanes blocked	2 out of 3 lanes blocked	Incident on roadway in opposite direction
Mean value of the capacity factor *(between brackets: standard deviation)*	0.72 *(0.09)*	0.36 *(0.14)*	0.18 *(0.12)*	0.69 *(0.08)*
Efficiency use of remaining lanes	0.72	0.54	0.54	0.69

The reduction in queue discharge capacity in situations with an incident on the roadway in the opposite direction is remarkably high (-31%). This reduction is entirely due to a change in driving behavior, since there are no lanes blocked on the roadway in question. The same applies to the capacity reduction in situations with a vehicle on the hard shoulder (-28%). It is remarkable that for both situations in which one or two lanes are blocked the same value is found for the efficiency of the use of the remaining lanes (54%). According to Knoop (2009) this means that both situations lead to the same behavioral effects.

When comparing these values with those found in the United States (Table 2.7), the capacity reductions found in the Netherlands turn out to be relatively high, except for situations in which two of the three lanes are blocked. This difference might be due to differences in the definition used for 'incidents', or due to behavioral differences between Dutch drivers and drivers in the United States.

A special situation is found for roads with a rush-hour lane (i.e. a hard shoulder which is used as additional lane during peak hour conditions). For such roads, incidents on the hard shoulder actually might also have a *physical* effect on the capacity (while they normally would have an *efficiency* effect only), by prohibiting the opening of the rush-hour lane. If the hard shoulder lane is blocked, it cannot be opened for traffic, resulting in the capacity to be lower than usual for the traffic demand level at hand.

Besides the magnitude of the capacity reduction, the incident duration is a very important factor for the impact on the traffic conditions as well. In Figure 2.20 the empirical probability distribution of the duration of incidents on the Dutch freeway network is shown.



Property	value (*min.*)
Mean	77.0
Median	36.8
Standard deviation	106

Figure 2.20: Distribution of incident duration
(Source: Knoop, 2009)

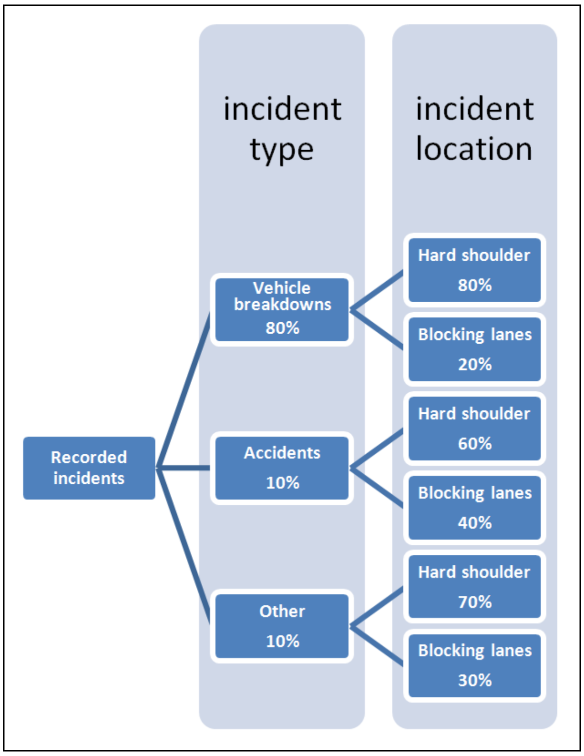
Fifty percent of all incidents are cleared in 37 minutes or less. However, the distribution is very skewed: it has a very long tail to the right. This means that some of the incidents have a very long duration. About 8% of the incidents even take longer than 4 hours! In view of the fact that an incident causes a total delay that is at least proportional to the square of its duration, this characteristic is rather detrimental for the performance of the traffic system.

- Frequency of occurrence -

As far as the incident frequency is concerned, Immers et al. (2005) refer to a study conducted in the United States, in which an incident probability of 0.171 incidents per 100,000 car kilometers was found. For a route with a length of 30 km and a daily traffic volume of 40,000 vehicles, this corresponds to a daily incident probability (i.e. the probability of one or more incidents occurring on an arbitrary day) of: $1\text{-}(1\text{-}0.171/100{,}000)^{40{,}000}=0.07$ (assuming independence between the individual vehicle kilometers).

Most incidents concern vehicle breakdowns, usually only blocking the hard shoulder. Also in case of accidents the drivers involved often try to move to the side of the road, resulting in only the hard shoulder being blocked. However, in case of an accident the probability of a lane blocking is larger than in case of a vehicle breakdown. Figure 2.21 shows an overview of the relative frequencies of occurrence of these different types of incidents, taken from the US ‘Traffic Incident Management Handbook’ (PB Farradyne, 2000). It should be noted, however, that these are values obtained for the United States. Values for the Netherlands might deviate from these. Moreover, there might be differences related to the scope of the category ‘Other’.

Figure 2.21: **Relative frequencies of occurrence of incident categories**
(Based on: PB Farradyne, 2000)



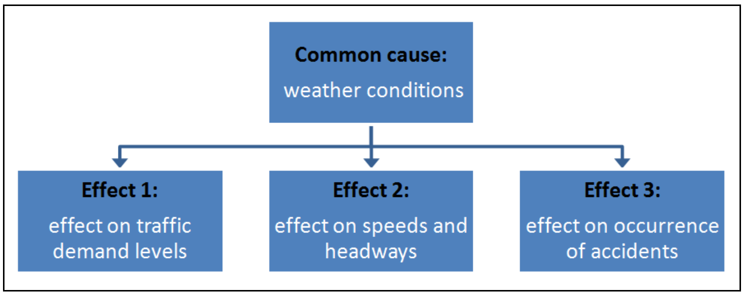
As far as *accidents* are concerned, it is found that large differences in rate of occurrence can be observed for different road sections. The geometrical properties of the road sections turn out to be an important factor in this respect. Accidents mainly occur at road sections with an on or off ramp, at weaving sections, and at road sections on which one of the lanes ends (Kraaijeveld, 2008). For vehicle breakdowns, differences between different road sections[25] obviously are likely to be smaller.

To a certain extent, the occurrence of incidents is linked to the other sources of variations in traffic demand and supply. In particular this is the case for *accidents*. One of the dependencies between the occurrence of accidents and other supply or demand variations originates from the weather conditions (Figure 2.22). Earlier in this chapter it was already discussed that weather conditions might have an effect on traffic demand and traffic supply, the latter by affecting the driving speeds and the headways. However, weather conditions may also affect traffic supply by influencing the occurrence of accidents. During certain adverse weather conditions the accident rate is significantly higher than during good weather conditions (in spite of the road users adapting their driving behavior). This can be attributed to a reduced visibility (due to precipitation or fog), road surface slipperiness (due to wet, icy or snowy road surfaces, and resulting in longer breaking distances and increased probabilities of losing control), and reduced vehicle stability (vehicles may be blown sideways or blown over by strong wind gusts). Although not directly associated with ‘bad weather’, low sun (shortly after sunrise and shortly before sunset) may

[25] Expressed in terms of the number of vehicle breakdowns *per vehicle-kilometer*, in order to correct for differences in road section length and traffic volume.

also increase the accident rate, by affecting the visibility (depending on the orientation of the roadway).

Figure 2.22: Relationships between the occurrence of accidents and other supply or demand variations, due to the common cause 'weather conditions'

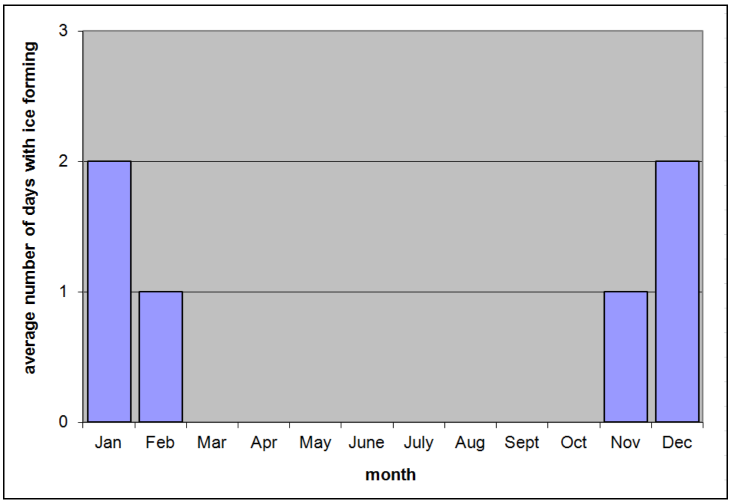


Based on international literature, the SWOV (2009) concludes that the crash rate approximately doubles during rain. During nighttime the crash rate during rainfall appears to be even larger than during daytime. For Dutch state highways rainy conditions are reported to result in an increase in the number of accidents of between 25% and 182%. Ice forming on the road surface is reported to result in an even larger increase: between 77% and 245%. Nevertheless, its impact on the total number of accidents is smaller, due to the fact that ice forming occurs far less frequently than rain (SWOV, 2009).

For the effect of snow on the accident rate contrasting results are found in international literature. The SWOV refers to a study in which it was concluded that snow seems to lower the crash rate, because it makes road users drive more carefully. Maze et al (2005) on the other hand conclude that snowy weather greatly increases the crash rate.

For the frequencies of occurrence of rain, snow, fog and black ice, the reader is referred to the earlier subsections on the weather effects on traffic demand and supply. On average, ice forming occurs on about 6 days a year, as illustrated in Figure 2.23[26].

Figure 2.23: Average monthly numbers of days with ice forming (Based on data from KNMI, 2002)



[26] It should be noted that this figure (as well as the average frequency of 6 days a year) does not specifically relate to ice forming on the *road surface*, meaning that for the road surface somewhat different values might apply. Furthermore, ice forming on the road surface is of course combatted by the road authority, which reduces the slipperiness.

Besides the weather conditions, also the luminance conditions might be a 'common cause' of the occurrence of accidents (affecting the traffic supply) and other supply fluctuations, resulting in these influence factors to be interrelated to some extent. In this case 'other supply fluctuations' refers to the effect that darkness might reduce the capacity and/or speeds by a few percent, as was discussed earlier in this section. As far as the occurrence of accidents is concerned, it can be assumed that in darkness the accident rate is higher, especially on locations without street lightning. It should be noted, however, that more (empirical or literature-based) research would be needed to find out whether this indeed is the case.

The presence of road works might influence the occurrence of accidents as well. In view of the associated discontinuities in the road geometry, possible reductions of lane widths, and attention distraction due to ongoing activities, the presence of road works can be expected to result in a higher accident rate. Since road works have other effects on traffic supply and demand as well (discussed in other parts of this chapter), the occurrence of accidents and these other effects of road works are interrelated to some extent. In this case these interrelationships can be attributed to the common cause 'road works'.

Besides on the weather conditions, the luminance conditions, and the presence of road works, the accident rate is also dependent on the traffic conditions. Actually, this might well be the most important influence factor of all (see Mehran & Nakamura, 2009). In several studies (for freeways in the United States, freeways in Korea, and expressways in Japan, respectively) it was found that the relationship between the volume-to-capacity ratio and the accident rate follows a general U-shape pattern (Mehran & Nakamura, 2009). This means that the accident rate is relatively high for both low and high traffic volumes, and lower in between.

That the accident rate is relatively high for high, free flow traffic volumes is of course quite logical, since larger traffic volumes give rise to more vehicle-vehicle interactions (each of which potentially might result in the occurrence of an accident), and result in smaller vehicle headways. If traffic is still free flowing, speeds are still rather high, resulting in little time to react on vehicles in front of one.

That the accident rate is relatively high as well for very *low* traffic volumes, might be more surprising. A possible explanation for this might be found in the fact that mutual speed differences are relatively large for low traffic volumes. Furthermore, it should be noted that very low traffic volumes typically occur during nighttime, and therefore are automatically connected to darkness and a driver population that is different from the one at daytime[27]. Therefore, while being attributed to the very low traffic volumes, in fact the higher accident rate might (partially or completely) be due to differences in luminance and driver population.

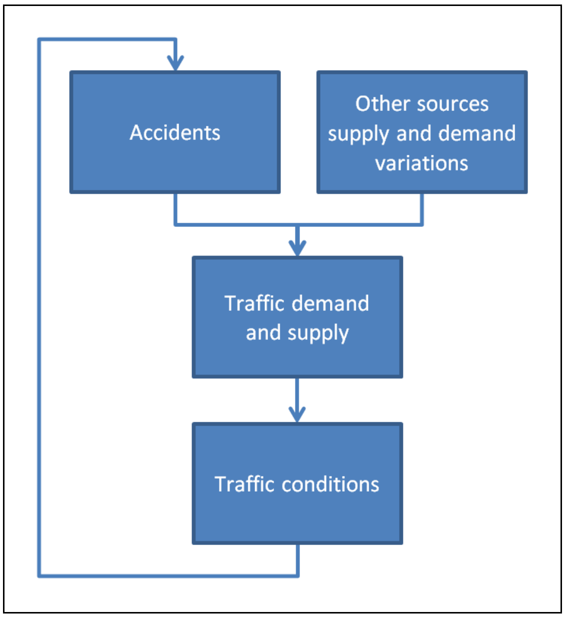
[27] On average, the nighttime drivers might be more 'reckless' than the daytime drivers, especially in the weekends.

Likely, the occurrence of traffic congestion affects the accident rate in two opposing ways. On the one hand, the lower speeds might result in the accident rate being lower. On the other hand, the shock waves in congested traffic flow cause extra traffic accidents (mainly head-tail collisions). It is not unlikely that this latter effect is dominant. More research would be needed, however, to find out whether this is indeed the case.

Since the rate of occurrence of accidents is dependent on the traffic conditions, and the traffic conditions on their turn are dependent on all sources of variations in traffic demand and supply, in fact the rate of occurrence of accidents is related not only to the weather conditions, luminance conditions and the presence of road works (as discussed above), but to all other sources of demand and supply variations as well (although in a more indirect way). This is illustrated in Figure 2.24. Since the occurrence of traffic accidents is one of the sources of variations in the traffic supply itself as well, in fact the accident rate is also dependent on itself. Of course there is a dependency on the rate of occurrence of the other incident types as well.

Please note that accidents are not the only type of incidents for which the probability of occurrence is dependent on the traffic conditions. Some other types of incidents (like vehicle breakdowns and cargo spills) for example are clearly dependent on the traffic volume: for larger traffic volumes these types of incidents are more likely to occur. Expressed in terms of the probability of occurrence *per vehicle-kilometer*, these dependencies on traffic volume are smaller (or even absent), however. Besides on the traffic volume, the probability of occurrence of *vehicle breakdowns* is also dependent on the *presence of traffic congestion*: if the traffic state is severely congested, vehicle breakdowns are more likely to occur than in free flowing traffic.

Figure 2.24: Indirect dependency of the occurrence of traffic accidents on all other sources of supply and demand variations, as well as on itself



5) Demonstrations

During demonstrations in some cases the road traffic is impeded by roadblocks or platoons of slow moving vehicles. Note that this kind of actions actually might affect the traffic demand as well (especially if they are announced in advance, or if they last for multiple days).

The frequency of occurrence of this type of actions is rather low, however. Therefore, they cannot really be considered a source of daily variability in traffic congestion. For this reason, the influence of demonstrations will not be given any further consideration in this thesis.

6) Emergencies

Emergencies or disasters (like flooding) obviously may significantly affect the traffic supply characteristics (i.e. the capacity and (uncongested) operating speeds). Not only the physical conditions on the road might be seriously deteriorated, but the driving behavior might be significantly affected as well. After all, in (imminent) emergency situations people might very well behave differently than under normal conditions. In section 2.2.4 it was however already noted that in the rest of this thesis no further consideration will be given to this kind of situations, since they clearly cannot be considered to be contributing to the *daily* variation in traffic congestion (because of their very small frequencies/probabilities of occurrence).

7) Variations in vehicle population

The composition of the vehicle population on a road section is constantly changing. Different vehicles have different characteristics like length, acceleration and deceleration capabilities, and applicable speed limit. The result of this is that the traffic supply characteristics (i.e. the capacity and (uncongested) operating speeds) are constantly changing as well.

A large part of this variation can be attributed to the variation in the (relative) amount of trucks. One truck consumes a larger part of the available capacity than one car. Expressed in the number of vehicles per unit of time the capacity therefore is lower if the percentage of freight traffic is larger. Besides considerable random short-term fluctuations, the percentage of freight traffic often shows a systematic variation over time as well. This is due to the fact that this freight traffic is quite evenly distributed over the day, while the other traffic is not. During the peak periods there is a lot of commuting traffic on the road, resulting in the percentage of freight traffic to be lower (and consequentially in the capacity – expressed in the number of vehicles per unit of time – to be higher).

This variation in capacity can largely be 'removed' (artificially) by expressing both the traffic demand and the capacity in 'passenger car equivalents' (per unit of time). In this case trucks are converted into an equivalent number of passenger cars. In the Netherlands, a value of 1.5 appears to be a reasonable conversion factor for this (AVV, 2002).

Geistefeldt (2009) estimated the conversion factor by determining for which value of this factor the coefficient of variation of the (free flow) capacity distribution becomes minimal, for a set of German freeways. (Note that normally, if capacity distribution functions are estimated in terms of *vehicles* per hour, the influence of the variations in the percentage of heavy vehicles is not explicitly considered. This influence is just part of the total stochastic variance of the capacity distribution function then.) With this method, Geistefeldt found conversion factors ranging from 1.3 to 2.6, for the different freeways considered. For an increasing number of lanes, the estimated conversion factor tended to decrease, indicating a smaller impact of heavy vehicles on freeway capacity. For the estimated values of the conversion factor, the coefficient of variation of the capacity distribution was reduced by up to 10% (as compared with the coefficient of variation of the capacity distribution in terms of *vehicles* per hour[28]). This means that apparently only a relatively small fraction of the capacity variation can be attributed to the varying percentage of heavy vehicles.

Al-Kaisy et al (2002) found that the capacity effect of heavy vehicles is larger under congested conditions than under free flow conditions. Just like Geistefeldt, Al-Kaisy et al. estimated the conversion factor by determining its value for which the variation of the capacity distribution becomes minimal. However, instead of the free flow capacity, they considered the queue discharge rate.

Mean conversion factors of 2.4 to 3.2 were found (with the highest value for a sloping road section). In addition, it was found that the conversion factor in fact is a random variable, which generally follows a normal distribution. This can be explained by variations in the weight-to-power ratios among the trucks. It was found that the conversion factor was not dependent on the weather conditions or roadside maintenance work. The fact that the effect on the queue discharge rate turns out to be larger than the effect on the free flow capacity can be explained by the limited acceleration performance of heavy vehicles, hampering the traffic flow out of the front end of the queue (which reduces the queue discharge rate). The difference in the effect on the free flow capacity and the queue discharge rate may explain the capacity drop phenomenon (discussed in section 2.1). Furthermore, differences in the percentage of heavy vehicles between different locations might explain the differences in the capacity drop observed in practice (Al-Kaisy et al, 2002).

Variations in the (relative) amount of freight traffic do not only affect the available capacity (expressed in the number of vehicles per unit of time), but also the average (uncongested) operating speed. This is due

[28] Note that in both cases, the influences of a large number of other factors introducing variability in the freeway capacity (viz. weather conditions, luminance, work zones, incidents, and driver population) had been excluded.

to the fact that the speed limit for trucks (in the Netherlands 80 km/h) is lower than the speed limit for passenger cars (generally 100 or 120 km/h in the Netherlands). As a consequence, the average speed gets lower as the proportion of freight traffic gets larger. This is not only due to the lower speeds of the trucks themselves, but also due to the fact that these slower trucks make the passenger cars slow down as well. The latter effect gets stronger for larger traffic densities (because of decreasing overtaking possibilities).

8) Variations in driver population

Another source of the variability in the supply characteristics (i.e. capacity and uncongested speeds) is the variability in the driver population. Just like the composition of the vehicle population, the composition of the driver population on a road section is constantly changing. Since different drivers behave differently (in terms of following the vehicles in front of them, changing lanes, desired driving speed, etc.), and this driving behavior is directly governing the capacity and (uncongested) operating speeds, this results in these supply characteristics to be constantly varying as well.

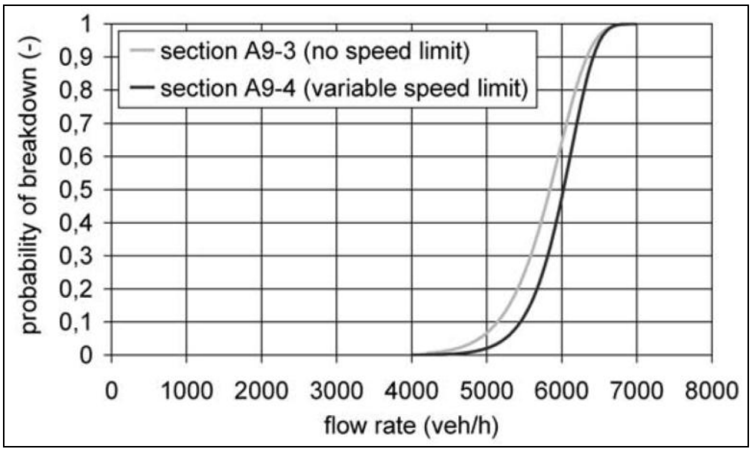
Mutual differences in driving behavior among drivers (and consequently also differences in capacity and speed) can be attributed to both personal characteristics and trip characteristics. Examples are skills, experience, age, gender, risk-taking propensity, and travel purpose. Two clearly different driver populations are the one in the peak periods and the one in the off-peak periods. During the peak periods a large part of the traffic consists of commuters and other road users with a profession-related travel purpose (like commercial traffic). These typically are experienced drivers. As a consequence, during the peak periods the capacity can be expected to be relatively high. Outside the peak periods and especially during weekends, public holidays and vacation periods the share of social and recreational traffic is relatively high. These drivers are typically less experienced (not necessarily regarding driving itself, but also regarding the traffic situations they are faced with). This can be expected to result in a relatively lower capacity during these periods.

For German autobahns (without a speed limit) Brilon and Ponzlet (1996) found that the uncongested average speeds (corrected for the effects of differences in traffic densities and proportions of freight traffic) generally are lower during periods with predominantly leisure traffic, such as Sundays or the summer vacation season. For four-lane metropolitan autobahns (two lanes per direction) differences between the various days of the week or months of the year of up to 3.5 km/h were found. Of course for Dutch motorways different values might be found, for example due to the presence of a speed limit. Note that the lower speeds during periods with predominantly leisure traffic might be an indication for the capacity to be lower during these periods as well (confirming what was suggested above), although this not necessarily has to be the case.

9) Traffic control actions

Probably rush-hour lanes and dynamic speed limits are the clearest examples of traffic control being a source of variation in the traffic supply. The opening or closure of rush-hour lanes clearly has a direct effect on the total available capacity, by adding or removing a certain amount of capacity. Dynamic speed limits obviously create variation in the traffic supply characteristic by dynamically affecting the speeds at which the road users drive from A to B. It should be noted however that as far as the other traffic supply characteristic, i.e. the road capacity, is concerned, dynamic speed limits might just be the *opposite* of a source of variability. That is, they seem to *reduce* the variability in the capacity. This can be illustrated with Figure 2.25. This figure shows a comparison of the empirically derived (free flow) capacity distributions for two freeway sections, one with and one without a traffic adaptive variable speed limit[29]. On the section *with* a variable speed limit the standard deviation of the capacity distribution is clearly significantly lower than on the section *without* a variable speed limit.

Figure 2.25: Capacity distributions (based on 5-minute intervals) for 3-lane freeway sections with and without a variable speed limit (Source: Brilon, 2005)



Temporary closures of motorway segments for bridge openings can be considered traffic control actions as well. In the Netherlands, the motorway network contains several movable bridges, which are regularly opened. In such situations, all traffic is temporarily halted (i.e. the capacity is temporarily zero). In total, the whole process of opening the bridge, the passing of the ships, and closing the bridge takes several minutes. It should be noted that if the bridge opens according to a fixed schedule, and/or information is provided on the opening of the bridge, road users might anticipate it, by adapting their travel choices. This way the traffic demand may be affected as well.

[29] These two sections are two opposing roadways of *one and the same* freeway, and therefore have similar geometric and traffic characteristics.

10) Intrinsic randomness in human behavior

In the end, the capacity is dependent on the combined behavior of all individual drivers involved. To a large extent, the variations in this behavior can be explained by the factors discussed above (related to the external conditions and the composition of the traffic). Another part of these variations however cannot be explained by these factors. This is due to the fact that human behavior is characterized by a certain 'intrinsic randomness'. In spite of finding himself in similar circumstances, one and the same person may still behave differently. This might be the result of differences in this person's mental constitution (i.e. differences in mood, level of fatigue, concentration, etc.) or have other, more inscrutable causes.

In addition, while the other part of the variations in the behavior *theoretically* can be explained by the factors discussed above, this does not mean that this part of the variations also *in practice* can be completely explained by these factors. This is due to the facts that not all of these factors are equally well observable and that not all of these influences are fully understood. The result of both these practical problems and the 'intrinsic randomness' in human behavior is that when all observable and understood influences are taken into account, still a certain 'residual randomness' will remain.

An indication of the magnitude of this residual randomness might be obtained by considering results from Brilon et al. (2005). They analyzed the (free flow) capacity distributions for a large set of 3-lane freeway sections (without a distinct bottleneck), based on empirical data for 5-minute intervals. Unfortunately it is not clear from their article to which extent they have tried to exclude the explainable part of the variability (for example by only including data that are measured during certain specific conditions). It is only stated that periods of work zones were excluded from the data. As a consequence, the actual residual randomness might be smaller than the variability apparent from the distribution functions found by Brilon et al.

Brilon et al. (2005) found that the empirical distribution of the (free flow) capacity is best represented by a Weibull distribution. In mathematics, this type of probability distribution is known as an asymptotical extreme value distribution for minimums. This characteristic of the Weibull distribution actually might play a role here as well. After all, assuming a constant traffic demand during the 5-minute interval (the time interval of analysis), the decisive value of the capacity equals its *minimum* over the interval.

Brilon et al. found that a Weibull shape parameter of about 13 seems to be characteristic for 3-lane freeways. For the Weibull scale parameter on the other hand widely varying values were found (from about 6000 to about 7900 veh/h). Differences in geometric conditions, control conditions, driver populations and vehicle populations are suggested as possible explanations for this. For the expected value of the distribution values were found ranging from about 5800 to 7500 veh/h, and for the coefficient of variation values ranging from 0.07 to 0.13.

Based on the assumption that traffic breakdowns in succeeding time intervals (or rather the *capacities* in succeeding time intervals) are independent of each other (which is a realistic assumption according to Brilon et al., since there is no imaginable reason why the opposite should be true), the 5-minute capacity distribution function can easily be transformed into the corresponding distribution function for a longer (or shorter) time interval. In this transformation the shape parameter remains unchanged, while the scale parameter is decreased (in case of a longer time interval). This corresponds to a decrease in both the expected value and the standard deviation of the capacity, as would be expected.

Less research seems to be conducted regarding the random variation in the *queue discharge rate* (i.e. the capacity after congestion has set in). This may be due to the fact that the free flow capacity is considered more important. After all, together with the traffic demand it is the free flow capacity that determines whether the transition from free flow conditions to congested conditions occurs or not. Therefore, fluctuations in this capacity might have an important impact on the resulting traffic conditions. Fluctuations in the queue discharge rate, on the other hand, might to a certain degree average out over the duration of the congested traffic state, resulting in their impact on the traffic conditions to be smaller.

2.2.6 Interdependencies between the various sources of temporal fluctuations

As noted already in the introductory subsection of this section, the various sources of variability are not all independent from each other. There are a lot of non-linear, dynamic dependencies involved between these fluctuations. These interdependencies are shortly discussed in this subsection. It would be taking things too far to discuss the individual interdependencies in detail. The purpose of this subsection is rather to give an overview of all interdependencies involved. Some important interdependencies have already been discussed in more detail in the previous subsections on the sources of variability concerned.

First of all, there are many links between different sources of variability because of the fact that these are all related to the common factor 'time' in some way. As a result of this common time-dependency, different sources of variation are linked even without necessarily having a causal relationship. If their time-dependencies do reasonably 'match', they will 'coincide' more than would be the case if these time-dependencies would not exist. If their time-dependencies instead are rather 'divergent', the situation is just opposite. In this case they will 'coincide' less than if these time-dependencies would not exist.

In Table 2.9, the dependencies in the frequency of occurrence (or patterns of occurrence) of all sources of variability on the factor 'time' are indicated, distinguishing between 'time of day', 'day of week' and 'period of year'. If different sources of variability have a ✓-sign in one and the same column (or even in more than one of the columns), this means that these sources are linked somehow due to their common time-dependency.

Table 2.9: Common time-dependency of sources of variability, resulting in mutual interdependencies

Source of variability	affecting demand or supply	Dependency in frequency of occurrence or pattern of occurrence on:		
		Time of day	Day of week	Period of year
Regular pattern of variation in travel behavior over the day	d	✓		
Regular pattern of variation in travel behavior over the days of the week	d		✓	
Regular pattern of variation in travel behavior over the periods of the year	d			✓
Public holidays	d		✓	✓
Events	d	✓	✓	✓
Varying weather conditions	d + s			✓
Road works	d + s	✓	✓	✓
Randomness in travel behavior (i.e. unexplained variations)	d		✓	
Variations in vehicle population[30]	s	✓	✓	✓
Variations in driver population[30]	s	✓	✓	✓
Darkness	s	✓		✓
Incidents	s			
Intrinsic randomness in driving behavior	s			

In addition to these mutual interdependencies due to common time-dependency, there are other interdependencies as well, which are more causal in nature. These interdependencies are shown in the tables below. A distinction is made between dependencies in the *frequencies/patterns* of occurrence of the sources of variability (Table 2.10), and dependencies in the *effects* of the sources of variability (Table 2.11).

As an example of a dependency of the first type, consider the dependency of the frequency of occurrence of accidents (a subcategory of incidents) on the weather conditions. As was already discussed in section 2.2.5, in adverse weather conditions the accident rate is observed to be significantly higher than under favorable weather conditions. This therefore is one of the dependencies included in Table 2.10.

As an example of a dependency of the second type, consider the interaction between the effects of adverse weather and darkness. Although the combination of adverse weather and darkness has a larger effect on the roadway capacity than each of these two conditions individually, it is known that this combined effect in general is *smaller* than the sum of both individual effects (AVV, 2002). Therefore, the

[30] In fact, these sources of variability in traffic supply are strongly connected to the regular patterns in travel behavior, which are included in the table as sources of demand variations.

effect of the occurrence of one of these two sources of variability is dependent on the occurrence of the other, which is indicated in Table 2.11.

Note that in both tables 'weather conditions' and 'road works' are included twice, due to the fact that these are both a source of variability in traffic *demand* and a source of variability in *supply*. Strictly speaking, in Table 2.10 this double inclusion is not necessary, since as far as the *frequencies of occurrence* are concerned the distinction between demand variations on the one hand and supply variations on the other is in fact not relevant. In Table 2.11, however, this double inclusion is absolutely necessary, since the dependencies in the *demand effects* of these sources of variability are different from those in their *supply effects*. For example, the *demand* effect of weather conditions may be dependent on whether it is a public holiday or not (since on public holidays a larger part of the trips is non-discretionary in nature, resulting in a larger sensitivity to weather conditions), while the *supply* effect of weather conditions is not dependent on this (apart from any possible effects via the effect on the driver/vehicle population, which is separately accounted for in the tables).

It should also be noted that the interdependencies indicated in the tables of course are not all equal in strength. Some of them are likely to be much stronger than others. In some cases it is not even certain if the indicated interdependency is really significant or not. More research would be needed to get clarity on this.

From Table 2.10 it is clear that especially the occurrence of incidents is characterized by many dependencies on other sources of fluctuations. Obviously, these dependencies are primarily related to one specific subcategory of incidents, namely accidents. Note that, although this is not indicated in the table, the probability of occurrence of incidents is also clearly dependent on the traffic conditions (as discussed in section 2.2.5), which actually makes the frequency of occurrence of incidents dependent on *all* sources of variations, including the occurrence of incidents itself.

Table 2.10: Dependencies of the frequencies/patterns of occurrence of the different sources of variability on the occurrence/level of other sources of variability

Dependencies of the ***frequencies*** of occurrence (or ***patterns*** of occurrence) of the different sources of variability (horizontal) on the occurrence (or level) of other sources of variability (vertical) ↥		Demand								Supply						
		Regular pattern in t.b. over the day	Regular pattern in t.b. over the d.o.w.	Regular pattern in t.b. over the p.o.y.	Public holidays	Events	Weather conditions	Road works	Unexplained variations	Vehicle population	Driver population	Darkness	Weather conditions	Road works	Incidents	Randomness driving behavior
Demand	Regular pattern in t.b. over the day		-	-	-	-	-	-	-	↥	↥	-	-	-	-	-
	Regular pattern in t.b. over the d.o.w.	-		-	-	-	-	-	-	↥	↥	-	-	-	-	-
	Regular pattern in t.b. over the p.o.y.	-	-		-	-	-	-	-	↥	↥	-	-	-	-	-
	Public holidays	-	-	-		↥	-	↥	-	↥	↥	-	-	↥	-	-
	Events	-	-	-	-		-	-	-	↥	↥	-	-	-	-	-
	Weather conditions	-	-	-	-	-		↥	-	-	-	-	=	↥	↥	-
	Road works	-	-	-	-	-	-		-	-	-	-	-	=	↥	-
	Unexplained variations	-	-	-	-	-	-	-		-	-	-	-	-	-	-
Supply	Vehicle population	-	-	-	-	-	-	-	-		↥	-	-	-	↥	-
	Driver population	-	-	-	-	-	-	-	-	↥		-	-	-	↥	-
	Darkness	-	-	-	-	-	-	-	-	-	-		-	-	↥	-
	Weather conditions	-	-	-	-	-	=	↥	-	-	-	-		↥	↥	-
	Road works	-	-	-	-	-	-	=	-	-	-	-	-		↥	-
	Incidents	-	-	-	-	-	-	-	-	-	-	-	-	-		-
	Randomness driving behavior	-	-	-	-	-	-	-	-	-	-	-	-	-	↥	

t.b. = travel behavior
d.ow. = days of the week
p.o.y. = periods of the year
= = demand and supply variations with a common cause
- = no significant influence
↥ = significant influence obvious / not unlikely

Table 2.11: Dependencies of the effects of the different sources of variability on the occurrence/level of other sources of variability

Dependencies of the ***effects*** of the different sources of variability (horizontal) on the occurrence (or level) of other sources of variability (vertical) ↥		Demand								Supply						
		Regular pattern in t.b. over the day	Regular pattern in t.b. over the d.o.w.	Regular pattern in t.b. over the p.o.y.	Public holidays	Events	Weather conditions	Road works	Unexplained variations	Vehicle population	Driver population	Darkness	Weather conditions	Road works	Incidents	Randomness driving behavior
Demand	Regular pattern in t.b. over the day		↥	↥	-	-	↥	↥	-	-	-	-	-	-	-	-
	Regular pattern in t.b. over the d.o.w.	↥		↥	-	-	↥	↥	-	-	-	-	-	-	-	-
	Regular pattern in t.b. over the p.o.y.	↥	↥		-	-	↥	↥	-	-	-	-	-	-	-	-
	Public holidays	-	-	-		↥	↥	↥	-	-	-	-	-	-	-	-
	Events	-	-	-	-		↥	↥	-	-	-	-	-	-	-	-
	Weather conditions	-	-	-	↥	↥		-	-	↥	↥	↥	=	↥	↥	-
	Road works	-	-	-	↥	↥	-		-	↥	↥	↥	↥	=	↥	-
	Unexplained variations	-	-	-	-	-	-	-		-	-	-	-	-	-	-
Supply	Vehicle population	-	-	-	-	-	-	-	-		↥	↥	↥	↥	-	-
	Driver population	-	-	-	-	-	-	-	-	↥		↥	↥	↥	↥	-
	Darkness	-	-	-	-	-	-	-	-	↥	↥		↥	↥	↥	-
	Weather conditions	-	-	-	↥	↥	=	-	-	↥	↥	↥		↥	↥	-
	Road works	-	-	-	↥	↥	-	=	-	↥	↥	↥	↥		↥	-
	Incidents	-	-	-	-	-	-	-	-	-	↥	↥	↥	↥		-
	Randomness driving behavior	-	-	-	-	-	-	-	-	-	-	-	-	-	-	

t.b. = travel behavior
d.ow. = days of the week
p.o.y. = periods of the year
= = demand and supply variations with a common cause
- = no significant influence
↥ = influence obvious / not unlikely

Note the interaction effects between the regular patterns of variation in travel behavior at different time scales, indicated in the upper left corner of Table 2.11. As an example, consider the interaction effect between the regular pattern of variation in travel behavior over the day, and the regular pattern of variation in travel behavior over the days of the week. The difference between the regular amounts of traffic demand on different days of the week is dependent on the time of the day, or formulated the other way around, the difference between the regular amounts of traffic demand on different times of the day is dependent on the day of the week. See also section 2.2.4 for an illustration of this interdependency (Figure 2.6). Between the regular pattern of variation in travel behavior over the day and the regular pattern of variation over the periods of the year a similar interaction exists. For the regular pattern of variation over the days of the week and the regular pattern of variation over the periods of the year this is the case as well.

2.3 Network effects

So far, the description of the traffic congestion mechanisms has been limited to individual road sections. However, between the traffic conditions on the various sections of a road (or road network) there are some strong spatiotemporal dependencies:

- Due to its physical dimension, a traffic jam created on a certain road section might block the traffic flow on other road sections (of the same road or even of other roads) as well. (Referred to as the 'blocking back effect' of traffic congestion.)
- By restricting the traffic throughput, and thereby delaying the traffic, the occurrence of traffic congestion on a certain road section influences the traffic demand on the downstream road sections. (Referred to as the 'temporal redistribution effect' or the 'filtering and releasing effects' of traffic congestion.)
- In situations with non-recurring congestion on a certain route, road users might deviate from their 'standard' route, in order to get around this traffic congestion. This will reduce the traffic demand on the congested route, but increase the traffic demand on the alternative routes. (Referred to as the 'route choice effect' of traffic congestion.)

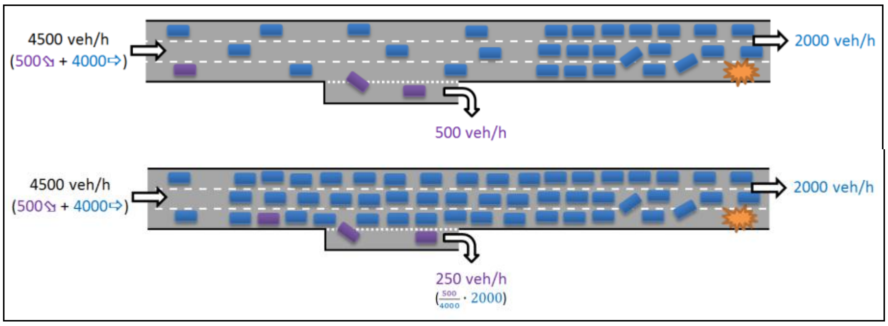
In the following three subsections, these three types of 'network effects' are discussed in more detail.

2.3.1 Blocking back

Due to its physical dimension, a traffic jam created by a certain bottleneck might block other traffic streams (consisting of travellers that do not even want to pass the bottleneck location) as well. During a certain period directly following on the onset of the congestion (i.e. the start of the formation of a queue), only road users with a destination downstream of the bottleneck are delayed by the congestion. However, as soon as the queue grows longer than the distance between the bottleneck location and the closest upstream off-ramp, travellers wanting to leave the motorway at this off-ramp (i.e. before the bottleneck location) will be delayed as well. From this moment onwards the queue will grow more rapidly, due to the effective capacity of the off-ramp being decreased (since it is blocked by the queue for the bottleneck).

This can be illustrated with the example given in Figure 2.26. In this case the outflow of the off-ramp is reduced by 50% due to the queue for the bottleneck blocking back to this off-ramp. Considering the system as a whole, the total outflow is reduced by (2500-2250)/2500*100% = 10% due to the blocking back effect. This results in the queue growing (2500-2250)/(4500-2500)*100% = 12.5% faster.

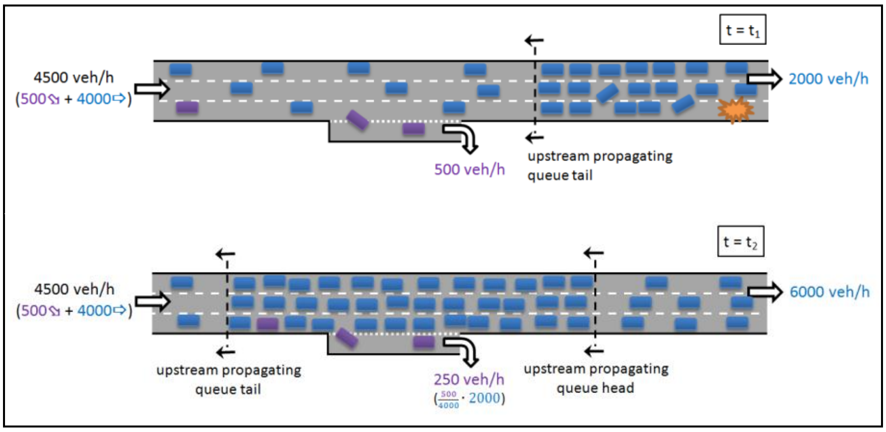
Figure 2.26: Blocking back of a queue to an upstream off-ramp



It is important to be aware of the fact that in order for the queue to block back to the off-ramp, it not necessarily needs to have a length greater than or equal to the distance between the bottleneck location and the off-ramp. This is related to the fact that queues might 'travel' upstream. This is explained using Figure 2.27.

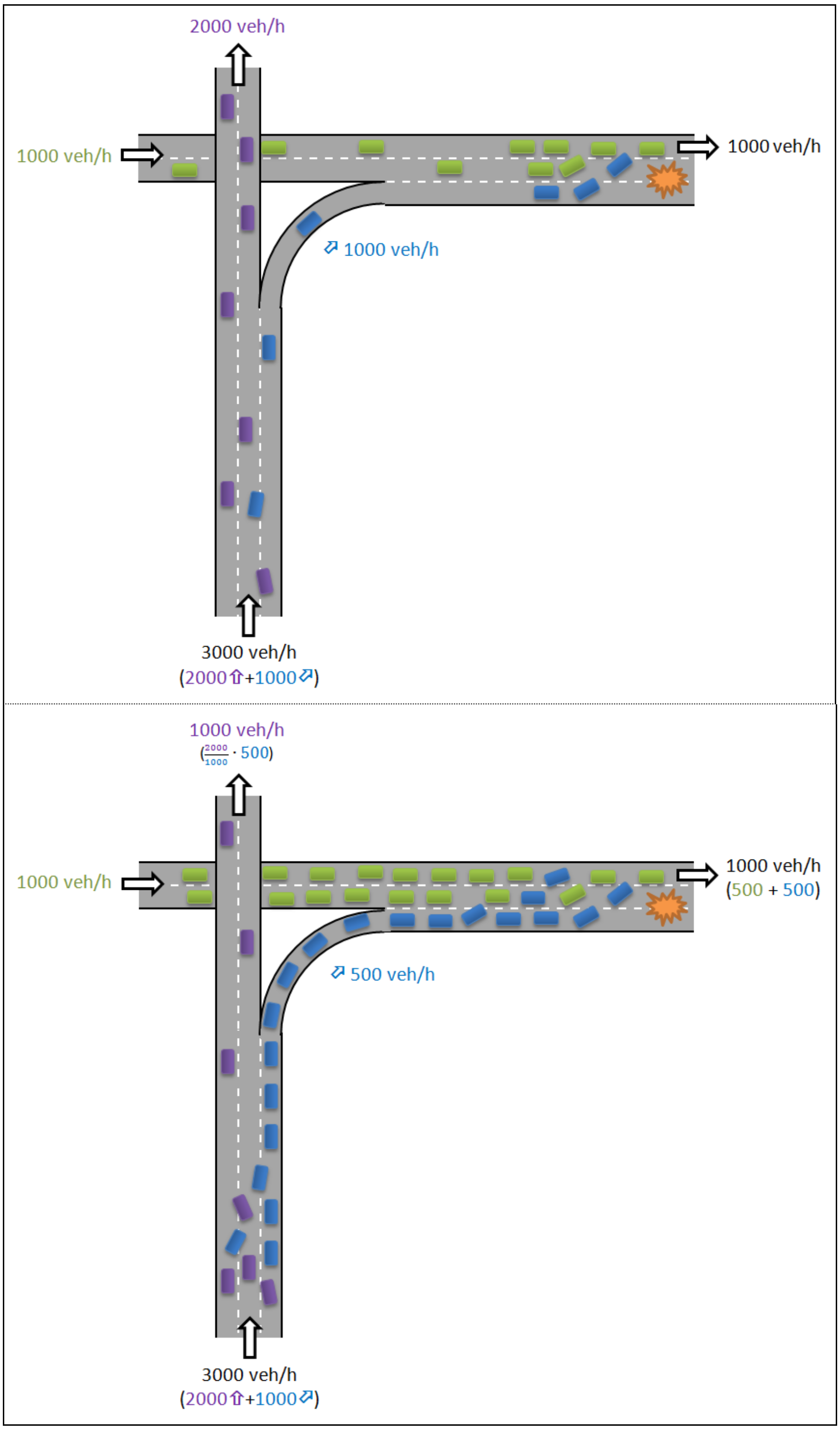
At a certain moment in time the available capacity at the (original) bottleneck location might be increased (for example due to the clearance of the incident that created the bottleneck), resulting in the queue starting to dissolve from its front side. This results in the head of the queue propagating upstream. However, in the meantime the queue often keeps on growing at the rear end, resulting in the tail of the queue propagating upstream as well. This makes the queue 'traveling' upstream. Due to the capacity drop (i.e. the phenomenon that the queue discharge rate is lower than the free flow capacity, which forms the upper limit of the inflow to the queue) such an upstream traveling queue may survive for quite a long time (for instance until the end of the peak period). Within this period it may travel a substantial distance. Along its way it might block other traffic streams, as illustrated in Figure 2.27.

Figure 2.27: Blocking back of an upstream traveling queue to an upstream off-ramp



The phenomenon of blocking back does not necessarily refer to the situation that a mainstream queue blocks the traffic stream to an off-ramp. It also occurs in other configurations. This is illustrated in Figure 2.28. In the situation depicted in the lower half of this figure a queue on the connecting roadway blocks the through traffic on the main roadway heading northwards.

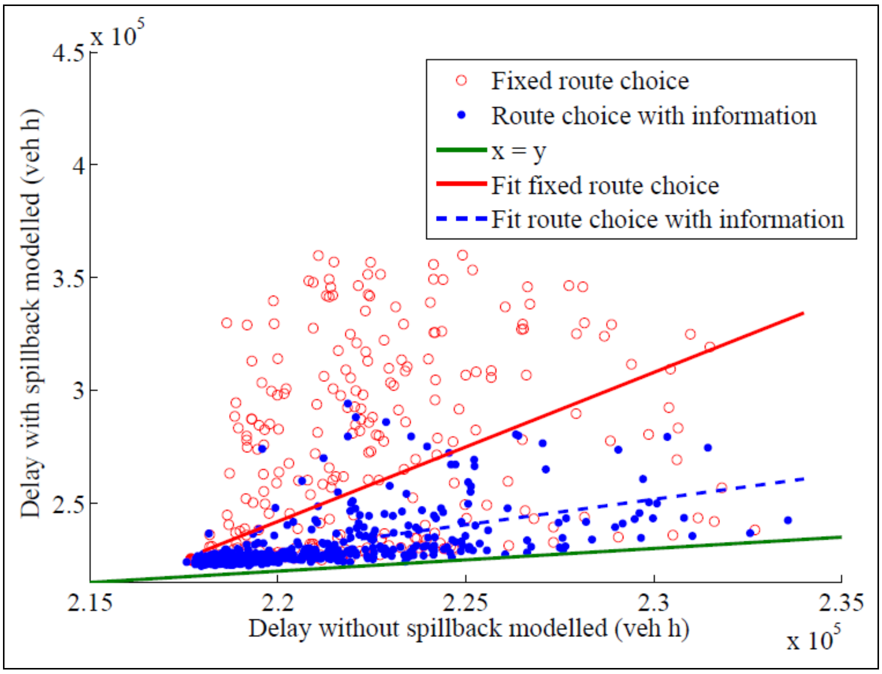
Figure 2.28: Blocking back of a queue to another road



In fact, the situation depicted in Figure 2.28 is much worse than the one shown in Figure 2.26. Due to the blocking back effect, the total outflow of the system is reduced by one third[31]. The growth rate of the total queue length (expressed in the total number of queuing vehicles, combining the queue on the west-east road and the queue on the south-north road) is increased by as much as 100%[32] (i.e. is doubled)!

Knoop (2009) shows that the delays due to blocking back may form a substantial part of the total amount of network delay caused by an incident. For a large number of incident scenarios, Knoop computed the total amount of network delay with a simulation model in which blocking back effects are accounted for, and compared the results with those obtained when these effects are not taken into account in the simulation model. This is illustrated in Figure 2.29. Clearly, in many cases the computed delay is much too low if blocking back effects are not taken into account, especially if the drivers are assumed not to deviate from their intended routes. The exact degree of underestimation is of course dependent on the incident scenario considered.

Figure 2.29: Comparison of total amount of network delay in case blocking back is modeled and total amount of network delay in case blocking back is not modeled, for a large set of incident scenarios
(Source: Knoop, 2009)



2.3.2 Temporal redistribution effect (filtering and releasing)

By restricting the traffic throughput, and thereby delaying the traffic, the occurrence of traffic congestion on a certain road section influences the traffic demand on the downstream road sections. This can be illustrated with Figure 2.30. The upper part of this figure shows the 'normal' (recurring) traffic situation at a certain road during one of both peak periods. The road has a bottleneck, resulting in a queue.

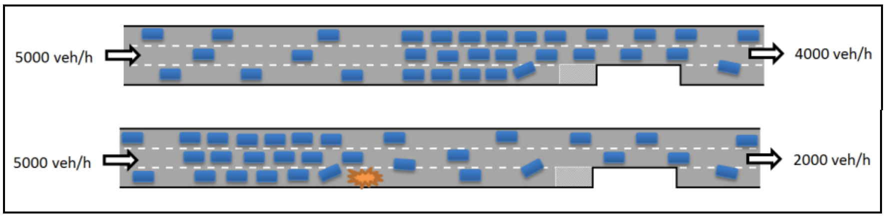
[31] ((2000+1000)-(1000+1000))/(2000+1000)*100%=(3000-2000)/3000*100%=33%

[32] ((3000+1000-1000-1000)-(3000+1000-2000-1000))/(3000+1000-2000-1000)*100%=(2000-1000)/1000*100%=100%

Now assume that on a certain day an incident occurs at a location upstream of the bottleneck. This situation is depicted in the lower part of the figure. Now the bottleneck location is 'under the lee' of the incident bottleneck location: at the incident location the throughput of traffic is restricted, reducing the traffic demand on the usual bottleneck location. As a consequence, at the latter location no traffic congestion occurs. All in all, the detriment brought about by the incident can be concluded to be smaller than it initially seems (i.e. if only looking at the incident location itself), since the incident results in an improvement of the traffic conditions further downstream.

Of course, after the road is cleared again (and the traffic from the queue is 'released'), the normal bottleneck location might still induce a certain amount of traffic congestion. However, if the incident lasts long enough (e.g. until well after the end of the peak period), in the meantime the queue in front of the incident location may have dissolved already. In this case, due to the incident, the traffic demand for the normal bottleneck location is redistributed over time in such a way that no congestion at all occurs at this location.

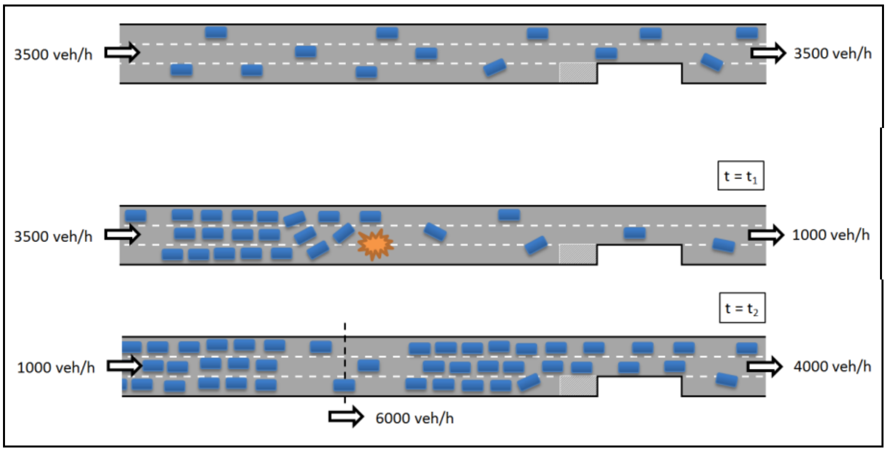
Figure 2.30: Positive influence of traffic congestion on downstream traffic conditions (filter effect)



The considered example might give the impression that traffic congestion always influences the downstream traffic demand in a favorable way. This however is not the case. This is illustrated with another example, depicted in Figure 2.31. Again the upper part of this figure shows the 'normal' (recurring) traffic situation at a certain road during one of both peak periods. One of the road sections has a lower capacity than the other parts of the road. This does not give rise to the occurrence of traffic congestion though, since this lower capacity is still sufficient to cope with the traffic demand.

Now assume that on a certain day an incident occurs at a location upstream of the section with lower capacity. This incident results in the creation of traffic congestion upstream of the incident location (middle part of Figure 2.31). Now consider the situation in which the road is cleared at a moment in time when the queue is still there. This results in a large amount of traffic being 'released', resulting in a traffic demand which is larger than the capacity of the narrower section. As a consequence, traffic congestion is created at this location (which is shown in the lower part of the figure). In this case, the total detriment brought about by the incident is thus larger (instead of smaller) than it initially seems (i.e. if only looking at the incident location itself): the incident results in a deterioration of the traffic conditions further downstream as well (by concentrating the traffic demand in a shorter time span).

Figure 2.31: Negative influence of traffic congestion on downstream traffic conditions (release effect)



The conclusion of the foregoing is that the occurrence of traffic congestion might affect the downstream traffic conditions in both a positive and a negative way. Please note that while in the examples the (upstream) traffic congestion was created by incidents, this is of course equally applicable to congestion with other causes.

2.3.3 Route choice effect

In situations in which the traffic conditions (i.e. levels of traffic congestion) are within their 'normal' range (or better), road users generally will stick to their standard routes. However, if the traffic conditions are significantly worse than what is considered 'normal' (as a consequence of for example an incident, an event or road works), road users might divert to other routes (possibly using the secondary road network), in order to get around the affected road segments. This will lead to a reduction of the traffic demand on the routes where the affected road segments are part of, which will positively influence the traffic conditions on these routes (meaning that these traffic conditions will deteriorate less than otherwise would have been the case). On road sections that are part of the alternative routes, however, traffic demand will be increased, affecting the traffic conditions in a *negative* way.

It is not possible to give some universally applicable values for the extent of this route choice effect. This is due to the fact that the actual magnitude of this effect strongly varies from case to case, since it depends on:

- the existence of route alternatives and the quality of these,
- the extent to which the road users are aware of these route alternatives (from their own knowledge or by being informed/advised on them),
- the degree to which the road users are informed on the traffic conditions on their routes (referring to both the current situation and the expected near-future evolution of this situation),
- the extent to which the road users are prepared to deviate from their standard routes (probably depending on the composition of the driver population), and
- the characteristics of the underlying causes of the 'unusual' level of traffic congestion (in case of an incident for example the duration of this incident).

Knoop (2009) investigated how many people change their route when faced with unexpected congestion due to an incident, by comparing actual route choice behavior for five incidents on a Dutch motorway (A13) with the corresponding behavior on similar days without incidents. He found that the severity of the capacity reduction and therefore the severity of the delays are important factors for the degree to which the route choice is influenced.

For incidents with minor impacts on the delays it seems that travelers do not deviate from their intended route, even in situations in which the alternative route would be slightly faster. For incidents with a considerable impact on the traffic conditions, however, Knoop found high percentages of users changing routes (even causing congestion on the alternative route). For one of the incidents this percentage was even higher than 50%. Knoop notes that these percentages in fact should be considered as lower bounds, since the 'decision point' considered actually was not the last opportunity to take an alternative route. Moreover, no correction was applied for the fact that part of the traffic had its destination in between the 'decision point' and the incident location (rather than downstream of the incident location). This part of the traffic might have had no other option than to stick to its intended route.

The reaction in route choice behavior is found to be delayed relative to the traffic conditions. This might be due to the delay in the information to which the road users react.

Kraaijeveld (2008) conducted a study similar to that of Knoop. In four out of five incident cases considered (on some Dutch motorways), no clear changes in route choice behavior were observed, probably due to a lack of good quality alternative routes and/or a limited impact of the incident on the traffic conditions. For one of the incidents (i.e. one with a much more severe impact, and relatively good alternative routes being available), however, a somewhat larger change in route choice behavior was observed. For this case it was concluded that 7% of the drivers switched to another route. This value is still far below the value of 50% reported by Knoop. This difference might be caused by for example differences in the quality of the available route alternatives and differences in the degree to which the road users were informed on the traffic conditions and the available alternatives.

From Figure 2.29 (section 2.3.1) it is clear that changes in route choice (that are made in order to get around road segments on which the traffic conditions are worse than what is considered 'normal', in this case due to the occurrence of an incident) might have significant effects on the delays incurred by the road users. In the traffic simulations in which the route choice was assumed fixed (i.e. not adaptable to the traffic conditions at hand) often a much larger total amount of network delay was found than in the corresponding traffic simulations in which part of the road users were assumed to adapt their route choices to the traffic conditions. At the level of individual routes the relative difference in delay can be even greater.

3. Criteria for the level of traffic congestion

3.1 Selecting appropriate criteria – introduction

In order to be able to evaluate the performance of the traffic system with regard to traffic congestion (taking into account its variable nature), first of all a clear view is needed on when the traffic system actually is considered to perform well and when it is not, because this is not something obvious. In fact, the consideration of the inherent variability adds an extra dimension to this. Therefore, one of the first parts of the research project has been devoted to the issue of finding one or more appropriate criterion(s) for this evaluation.

There does not exist a sharp 'failure boundary' with respect to the amount of traffic congestion (a threshold above which the motorway system can be considered to 'fail', and below which the system can be considered to 'function'). It could be suggested to consider the occurrence of traffic congestion as 'failure' (no matter how large this traffic congestion is). This does not really make any sense however, as the crux is in the *extent* of this traffic congestion. A little bit of traffic congestion occurring every day is not very harmful. Besides, even in quiet traffic conditions (without any traffic jams), travel time already gradually increases with rising traffic volume.

In the end, it is all about the costs that traffic congestion causes to society. Therefore, first of all consideration has been given to the question which features describing the traffic congestion phenomenon can be identified as being most decisive in bringing about costs to society (section 3.2). These typically are the features that need to be incorporated in the indicators.

Prior to the selection of indicators to be used in the remainder of the project, it has been checked which indicators are used in international literature, and which norms were used in the Dutch national traffic policy during the past few decades (section 3.3).

The final selection of indicators is discussed in section 3.4. Insofar as the known criterions (discussed in section 3.3) turned out to be inadequate (considering the findings from section 3.2), new or modified criterions were included in the selection.

In various empirical studies, strong relationships are found between the average travel time (or the average delay, corresponding to the difference between the average travel time and the free flow travel time) and other indicators based on the travel time distribution. These findings and their implications for this research project are discussed in section 3.5.

3.2 Societal costs of traffic congestion

3.2.1 Complexity of the topic

This section attempts to clarify how traffic congestion imposes costs on society. This is a complicated matter, because of the fact that costs in the restricted sense actually represent only part of the story. This can best be illustrated by a simple example. Consider a hypothetical situation in which cities A en B are not connected by the road network. Because of the fact that there is no connecting road, there is no traffic and consequently no traffic congestion costs either. Now a connecting road is constructed. This road attracts a lot of traffic, resulting in serious traffic congestion. Despite the large costs associated with this traffic congestion, the road users still gain from the road (otherwise they would not use it, obviously). This is because of the *benefits* that they derive from their trips, which apparently are larger than the costs experienced as a result of the traffic congestion.

If the road would be wider (resulting in a lower level of traffic congestion), even more trips would be made, since more travelers would be able to derive a positive net benefit (individual benefits derived from the trip, minus congestion costs experienced during the trip) from the trip in this case. Clearly, these additional net benefits because of extra trips are something to be taken into account in the evaluation of the effectiveness of measures aimed at improving the situation with respect to traffic congestion problems. It is quite arbitrary whether to consider these additional net benefits (that are missed if traffic congestion problems are not reduced) as a *potential gain* that *can be attained* by alleviating traffic congestion, or as a *cost* of traffic congestion that *can be removed* by alleviating traffic congestion. In this thesis the latter point of view is taken. That is, 'costs' is defined here in a broad sense: the reduction in net benefits due to traffic congestion.

3.2.2 Types of societal costs

Traffic congestion imposes costs on society in various ways. Below, a description will be given of the different types of costs. Next, consideration will be given to the issue of how these different costs are related to variables describing the traffic conditions. This has to provide important information regarding the indicators to be selected in order to be able to evaluate the traffic system's performance (with regard to traffic congestion).

The following types of societal costs are distinguished:

1) loss of time due to longer travel times,
2) costs due to uncertainty in travel times,
3) increase of both fuel consumption and wear and tear,
4) travelers diverting or staying away,
5) increase of the number of accidents,
6) larger load on the environment,
7) indirect costs related to settlement behavior and logistic chains,
8) discomfort.

Below, these cost items will be described one by one.

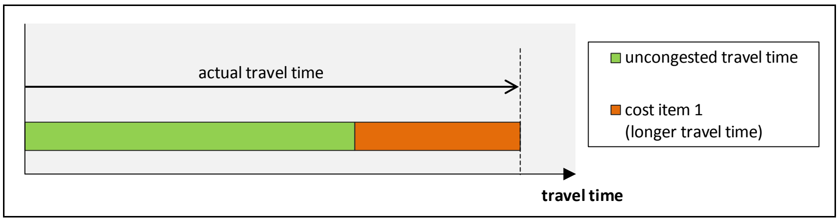
1) *Loss of time due to longer travel times*

Due to the traffic congestion, travel times are longer (Figure 3.1). In a situation without traffic congestion, this additional time could have been used for some other purpose. People could have made better use of their time, and capital goods (like trucks) could have been used more efficiently. As far as the transport of perishable products is concerned, longer travel times result in a reduction of the time available for selling these products. Please note that in this thesis, 'traffic congestion' is not limited to the occurrence of traffic jams. Also in free flow traffic, travel times gradually increase with rising traffic density/volume. The delay associated with this rise in travel time is understood as a cost of 'traffic congestion' as well.

When comparing the traffic system's performance in different situations, one has to be aware of the fact that a comparison based on delay only makes sense if this delay is expressed in relation to a (reasoned) fixed reference level of travel time/speed. This can be illustrated with the following example. Consider a rural road connecting two cities with a travel time of about 2 hours. There is no delay, because of the fact that the road is hardly used (owing to the long travel time). Now assume that this rural road is replaced by a motorway, reducing the free flow travel time to about 1 hour. Because of this shorter free flow travel time, a lot of people start traveling between the two cities, using the new motorway. This results in the occurrence of traffic congestion, increasing the travel time up to 1.5 hours.

Using the two different free flow travel times as reference levels (instead of a fixed, common reference level), a comparison of the two situations would result in the conclusion that traffic conditions are worsened: in the new situation there is a delay of half an hour, while in the previous situation there was no delay at all. Obviously, this conclusion is wrong. After all, in despite of the delay, the travel time (free flow travel time + delay) in the new situation is still shorter than in the original situation. Using a fixed reference level for delay can prevent such false conclusions. Comparing situations based on travel time instead of delay is another option.

Figure 3.1: Costs due to longer travel time



2) *Costs due to uncertainty in travel times*

Typically, traffic congestion not only results in *longer* travel times, but in *uncertainty* in these travel times as well. This uncertainty is another source of societal costs. The costs involved can be divided in two subcategories: costs due to late arrivals (item 2a), and costs due to waste of time related to arriving early (item 2b). Due to the uncertainty in travel times, the arrival time of a trip (or the other way around: the departure time required to arrive at a desired moment in time) is uncertain. Travelers deal with this uncertainty by adding a certain buffer to the travel time that they expect for their trip.

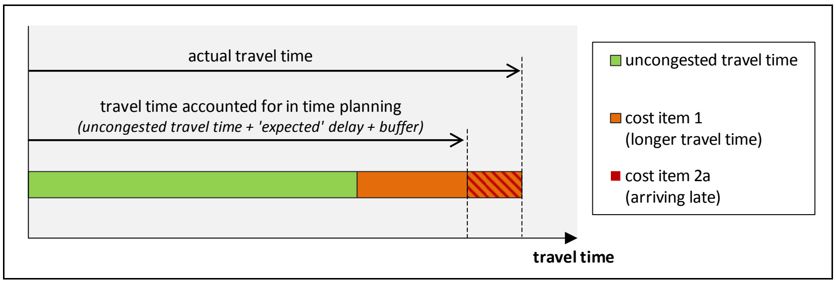
The size of these buffers mainly depends on:

- the variability in travel times,
- the predictability of this variation[33] (governed by regularity in the variation, knowledge, travel information, and travel time stability),
- the possibilities for diverting (governed by the availability of both alternative travel options and travel information),
- the travel purpose, and
- personal characteristics.

Note that such a buffer not necessarily is an *addition* to the 'expected' travel time. A negative 'buffer' might occur as well. This might be the case if arriving late is no problem at all, or if the traveler in question is rather risk prone.

2a) Costs due to late arrivals (Figure 3.2)
If the delay due to the traffic congestion was not (completely) anticipated upon (by including a sufficiently large buffer in one's time planning), one arrives late. Often, the costs of arriving *x* minutes late are higher than the costs of just having an *x* minutes longer travel time (cost item 1). This is because of the fact that late arrivals have consequences like appointments being missed or delayed, or supplies being run out of (resulting in disruption of production processes, and/or loss of sales). In order to limit the probability of running out of supplies, firms might opt for keeping larger stocks. However, because of the costs associated with keeping larger stocks, modern strategy in business is just the reverse: minimizing the need for stocks by 'just-in-time' deliveries.

Figure 3.2: Extra costs due to late arrival

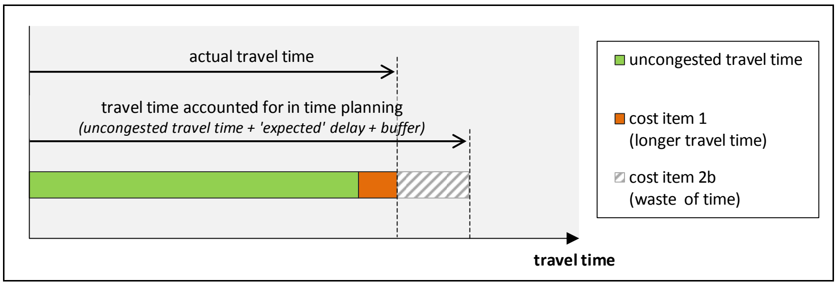


2b) Costs due to waste of time related to early arrivals (Figure 3.3)
If the actual travel time turns out to be shorter than the travel time planned for ('expected' travel time + buffer), a certain amount of time is left over. Often, this remaining time period is partially or

[33] Please note the difference between *variability* and *uncertainty*. The costs are related to the *uncertainty* in travel times. If the variation in travel times would be completely predictable, there would not be any uncertainty at all. In this situation, this cost item would not be there either.

entirely wasted (due to the fact that it cannot be efficiently used for another purpose anymore).

Figure 3.3: Extra costs related to arriving early



In total, the costs associated with the uncertainty in travel times are believed to be of the same order of magnitude as the costs associated with the increase in travel times alone (cost item 1):

- The day-to-day variation in the amount of delay is found to be of the same order of magnitude as the average amount of delay (Van Toorenburg, 2003).
- The appraisal of one minute standard deviation is believed to be approximately equal to the appraisal of one minute average travel time: for freight transport a ratio (‘reliability ratio’) of 1.24 has been derived, and for passenger cars usually a ratio of about 0.8 is found (Kouwenhoven et al., 2005).

3) *Increase of both fuel consumption and wear and tear*

In congested traffic operations, fuel consumption is often higher than in free flow traffic conditions. To a large extent, this is caused by the fact that in congested traffic conditions often large fluctuations in speed occur (‘stop-and-go-traffic’), resulting in a lot of acceleration and deceleration. Besides in higher fuel consumption, this also results in larger wear and tear of the vehicles involved.

4) *Travelers diverting or staying away*

Because of traffic congestion, part of the (potential) travelers opts for another trip (other departure time, other route, other transport mode, or other destination), or decides to stay at home. This way, the people in question save themselves the costs that they would be confronted with when ‘taking part’ in the traffic congestion. However, these costs are exchanged for other (lower) ‘costs’. As discussed in section 3.2.1, ‘costs’ should be taken broadly here: any reduction in net benefits[34] as compared to the situation of making the preferred trip without any traffic congestion should be considered as ‘costs’.

This can be illustrated with the following example. Consider a traveler selecting another destination than his preferred one, because of the fact that the routes to his preferred destination are

[34] Net benefits = benefits derived from the trip – costs incurred by making the trip.

heavily congested. By doing this, the traveler saves himself congestion costs. However, visiting the alternative destination is less valuable to him than visiting his preferred destination. In this way, the costs related to the traffic congestion are eliminated only for a limited part. A large part of the 'true' congestion costs simply is exchanged for other 'costs' (reduction in the benefits derived from the trip). These 'costs' should be attributed to the traffic congestion as well. (They could be considered as a kind of 'substitute' costs for the 'true' traffic congestion costs.)

5) *Increase of the number of accidents*

Traffic congestion probably results in more traffic accidents occurring (head-tail collisions). These accidents lead to physical injuries and material damage.

6) *Larger load on the environment*

By emitting harmful exhaust fumes, depleting fossil fuels, and producing noise and odors, traffic imposes a load on the environment. Traffic congestion causes this load to be larger.

7) *Indirect costs related to settlement behavior and logistic chains*

Traffic congestion problems have a negative influence on the settlement behavior of firms: firms might decide to settle in another area, where accessibility is better. Furthermore, traffic congestion problems might negatively affect logistic chains (from an economical point of view). For example, shippers may decide to start forwarding their freight to the European hinterland via Antwerp instead of Rotterdam. These processes have all sorts of (long-term) effects, like effects on employment and effects on tax revenues.

8) *Discomfort*

Though maybe not literary a societal cost, also discomfort may be mentioned in this list. Due to traffic congestion, driving comfort is lower. After all, in congested conditions there is less freedom to maneuver, desired driving speeds cannot be attained, the driving task may be more demanding (think of having to accelerate and decelerate all the time, due to shock waves in the traffic flow), and one might get stressed due to the uncertainty regarding whether one will arrive on time or not.

3.2.3 Cost items in relation to variables describing the traffic conditions

In the previous subsection, only a general description of the various cost items was provided. In this subsection, consideration will be given to the issue of how these different costs are related to variables describing the traffic conditions. As indicated before, this has to provide important information regarding the indicators to be selected in order to be able to evaluate the traffic system's performance (with regard to traffic congestion). Note that in this subsection, the various cost items are still considered separately. Obviously, it is desirable to combine them into a more limited number of indicators. The attempt to do this is described in a following section.

1) *Loss of time due to longer travel times*

This cost item probably is the one that can be expressed in terms of variables describing the traffic conditions most straightforwardly. It can simply be addressed by looking at the average travel time or delay of the road users. This average has to be calculated over a period of time that is sufficiently long to properly grasp the influence of all (systematic and random) variations over time. This means that at least a period of a year should be considered, but preferably longer.

As was explained before (in subsection 3.2.2), it is important to take a fixed reference level if travel time costs are expressed in terms of delay. Rijkswaterstaat uses a level of 100 km/h for this. Even if such a fixed reference level is used, the indicator 'delay' is still not the most appropriate one however. This is related to the fact that this indicator fails to properly represent the costs of congestion if drivers make detours to avoid congested locations. If delay is defined as the amount of additional travel time incurred due to the speed being below the reference level, this detour making behavior results in the average amount of delay being lower[35]. However, this reduction in delay time obviously is partially compensated for by an increase in the free flow travel time (related to an increase in the distance to be covered). While this is not reflected in the indicator 'delay', it *is* reflected in the indicator 'travel time'. Therefore it seems better to address the cost item by looking at the average travel time, instead of looking at the average delay.

The average travel time can be considered at different aggregation levels: at road section level, at route level, at origin-destination level (combining the travel times on different routes between the origin and destination), or even at network level (averaging the average travel times on all elements that are part of the network (road sections / routes / origin-destination relations), weighted by the relative numbers of trips). On the one hand it is convenient to employ the highest possible aggregation level. After all, this makes the analysis more manageable. On the other hand, the higher the aggregation level, the more (potentially relevant) information is lost. Indeed, at a higher aggregation level improvements and deteriorations at different locations might offset one another.

Since road users might start making detours if traffic conditions on a certain road section / route get too bad (i.e. traffic might shift between different (parallel) road sections / routes), in fact the levels of road sections and routes are too low to be selected as level of

[35] After all, in general travel speeds on detour routes are higher than those on the original routes, because otherwise it would not make sense to make the detour.

analysis. The network level however is too high, since too much information is lost on this level. Therefore it is preferred to consider the average travel times on the level of origin-destination relations. Of course, just like that road users might switch to another route in response to changed traffic conditions, they might change their destination as well. This might make one concluding that in fact the origin-destination level is not appropriate either. However, since this process of changing destinations is believed to be weaker (and at least much slower) than the process of switching routes, and in view of the fact that it is considered really undesirable to use a higher aggregation level (because of the loss of information associated with this), the origin-destination level is selected as the preferential level of analysis nonetheless.

As far as the time domain of the travel time average is concerned, it might both be decided to consider only one general average, or to make a distinction according to various periods of the day and/or week. In the latter case different periods are distinguished (like morning peak – evening peak – off-peak, or weekday – weekend day), for which separate averages are computed. While the former option is convenient to make the analyses more manageable, using the latter approach might provide valuable additional information.

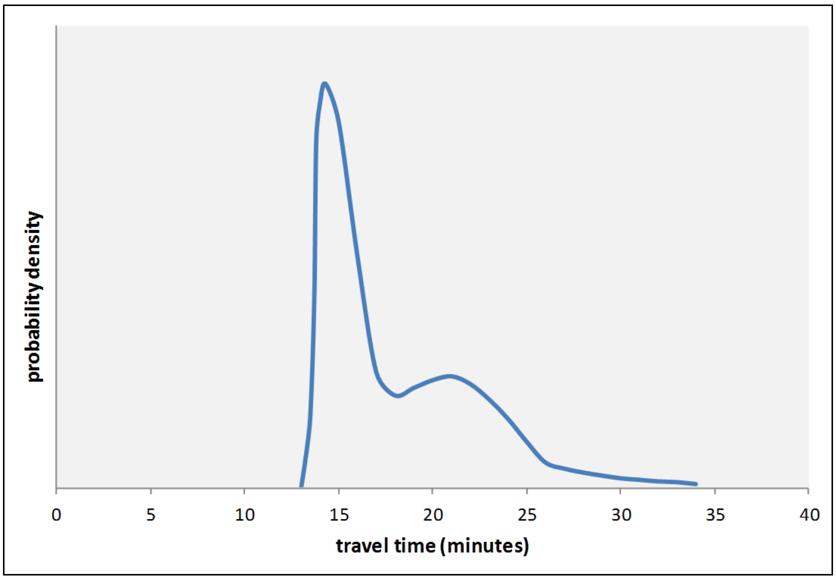
Finally, it should be noted that it is important to be aware of the fact that not all the delay / increase in travel times can be attributed to the occurrence of traffic congestion. Speed limit restrictions related to road works and incidents, and speed reductions in bad weather conditions can be sources of delay as well.

2) Costs due to uncertainty in travel times

Travel time distribution

For this cost item it is much more difficult to relate the costs to variables describing the traffic conditions. There is still no consensus on the issue of which quantitative measure(s) best reflect(s) travel time uncertainty (Bogers et al, 2008). Obviously, the costs due to this uncertainty are strongly dependent on the distribution of the travel times (Figure 3.4). Especially the width of this distribution (representing the variation in travel times) is very important in this context: the larger the variation in travel times, the larger the loss suffered due to arriving late and arriving early.

Figure 3.4: A travel time distribution



What travel time distribution?

When claiming that 'the distribution of the travel time' is important, it is necessary to clearly define what is actually meant by this distribution. The variation in travel time can be subdivided in the following two main components:

- 'within-day' variation: variation over de course of the day;
- 'day-to-day' variation: variation between days, for a given time of the day.

Consequently, one can think of different types of travel time distributions:

- the day-to-day distribution, for a given time of the day;
- the within-day distribution, for a given day; or
- the global (overall) probability distribution, for all moments in time (i.e. combinations of day and time of the day) together.

The issue of which of these distribution types to consider is closely related to the trip making behavior. This trip making behavior is the result of people's activity patterns. These activity patterns partly have a repetitive nature. This repetitive part of the activity patterns results in trips that are repeated at fixed moments in time. For example, many commuters go to their work every weekday at about the same point in time. For these trips, within-day variation in travel time is not important at all; the day-to-day variation is the only variation that is experienced by the road users involved. As far as this kind of trips is concerned, consideration thus has to be given to the *day-to-day distribution* of travel time. Commuting (i.e. home-work) trips are not the only example of this kind of trips. Freight transport trips often have fixed schedules as well. Part of the business trips is made on a regular basis too. For social-recreational trips the same applies.

The other part of the trips however does not take place repeatedly at fixed moments in time. These are solitary trips, and trips that do take place repetitively, but scattered over time instead of on a regular basis. For these trips not the day-to-day travel time distribution, but rather the *global (overall) distribution* should be considered. Because of the fact that for part of the trips the global (overall) distribution is important and for the other part the day-to-day distribution, in this project consideration should be given to (features of) *both* of these distributions.

Another important issue related to the definition of the travel time distribution is whether it is defined with respect to *time* (resulting in statements like: "in one on five *days*, travel time is larger than 10 minutes"), or with respect to *trips* (resulting in statements like: "in one on five *trips* (not necessarily of one and the same person), travel time is larger than 10 minutes")[36]. Opting for the one or the other definition might result in different distributions, due to any possible dependencies between the amount of trips (per unit of time) and the travel time. (For instance: in busier traffic conditions travel times are more likely to be longer.)

It is hard to say which option is best in this context. In this project, it has been chosen to define the travel time distribution with respect to time (the easiest option from a computational point of view). For the global (overall) travel time distribution, this choice may seem less obvious than for the day-to-day travel time distribution. One might argue that a definition with respect to time is rather 'unfair' in this case, because of the fact that trips are very unevenly distributed over the day. Here it should be born in mind however that this global travel time distribution is considered in relation to only a *subset* of the trips (see above). In this subset the commuting trips are not included. These commuting trips are to a large extent responsible for the unevenness in the distribution of the trips over the day. The other (i.e. non-commuting) trips actually are relatively evenly distributed over time (if the night period is left out of consideration).

One more important issue related to the definition of the travel time distribution is whether this distribution should be considered at *road section* level (i.e. the distribution of *road section travel times*), or at the level of *routes* (i.e. the distribution of *route travel times*). These two approaches are likely to give different results, among other things due to the fact that travel time fluctuations on the different road sections in a route might compensate for each other (which would result in the *relative* travel time variation on route level being smaller than the relative variance on section level). In fact, choosing between the two approaches is quite straightforward. After all, the societal costs associated with travel time uncertainty are due to the uncertainty in travel times *of trips*.

[36] If the travel time distribution is to be derived from empirical or simulated data, in the latter case (i.e. defining the distribution with respect to trips) all travel times should be weighted according to the amount of trips / road users involved, while in the first case (i.e. defining the distribution with respect to time), all travel times (each corresponding to a certain measurement interval) would be weighted equally.

Therefore, instead of considering the distribution of road section travel times, rather the distribution of route travel times should be considered[37].

In fact, this approach is still not optimal though. This is because of the fact that in part of the cases, road users have the possibility to deviate from their 'standard' route, by which they can get around heavily congested locations. This results in the travel time distribution between the origin and destination being less unfavorable than the *route* travel time distribution suggests. Consequently, the costs associated with the uncertainty in travel time (i.e. costs associated with arriving early or late) are lower than may be deduced from the route travel time distribution.

This leads to the conclusion that in fact not this *route* travel time distribution should be considered, but the distribution of the travel time *on the origin-destination relation.* However, this distribution is rather difficult to obtain, because it is highly dependent on the information received by the road users (information regarding the traffic conditions and the available alternatives), as well as on the extent to which the road users actually use this information / these alternatives. Furthermore, the various road users will all be different from one another in this respect. Some drivers will be better informed than others, and some drivers will be more prepared to deviate from their 'standard' routes than others. Therefore, it is chosen to consider both extremes: road users that always stick to their 'standard' routes on the one hand, and fully informed road users that always select the best route alternative available on the other hand[38]. For the first group, simply the *route* travel time distribution can be considered. For the latter group, the travel time distribution should be constructed by selecting for each 'measurement' time interval the shortest travel time available (among the travel times of the various alternative routes).

Now that it is decided to consider (indicators derived from) the distribution of *origin-destination* travel times, it still has to be decided in what range the distances of the origin-destination relations to be considered should be: should we focus on long origin-destination distances, on short ones, or both? Considering long distance or short distance origin-destination relations is likely to give different results, for example due to the facts that:

---

[37] In fact, these should be routes 'from door to door'. However, in this project routes between nodes in the motorway network will be considered. This is because of the facts that:

- Nearly all road users have a different combination of origin and destination, resulting in a door-to-door analysis being virtually impossible.
- In this project, focus is on congestion problems at the motorway network (and not on congestion problems on the underlying network).

In the ensuing of this chapter, an 'origin-destination relation' should thus not be interpreted as a 'true' door-to-door relation, but rather as a relation connecting two network nodes.

[38] This is considered a more correct approach than considering a kind of 'averaged' distribution.

- On longer distance origin-destination relations, travel time fluctuations on various parts of the routes are more likely to compensate for each other than on shorter distance origin-destination relations (which might result in longer distance origin-destination relations having a relatively smaller travel time variation).
- On longer distance origin-destination relations there may be more route alternatives than on shorter distance origin-destination relations.

Generally, the main function of the motorway network is considered to be the facilitation of long distance trips (with an adequate level of service). Therefore, one might be inclined to say that only longer distance origin-destination relations should be considered. This would be a rather naïve point of view however. Currently, the motorway network is being used by a lot of short distance traffic as well. This is mainly caused by the sharp focus on the main road network in the Dutch traffic policy of the past. In this traffic policy, facilitating long distance traffic was considered the main function of the main road network, but where possible, short distance traffic had to be handled via this network as well, in view of the associated benefits regarding traffic safety and quality of life. This has resulted in the underlying network remaining underdeveloped, and the main road network being used by a lot of short distance traffic.

Regardless of the view as to whether this situation actually is desirable or not, it is undeniable that the current situation *is* like this, and that consequently traffic congestion imposes costs on society not only by affecting long distance traffic, but also by affecting short distance traffic. Therefore, it is preferable to look at shorter distance origin-destination relations as well.

Finally, it should be decided for which time(s) of the day the day-to-day travel time distribution should be considered. Of course, considering only one day-to-day distribution (i.e. for only one specific time of the day) is not sufficient, since the day-to-day distribution will be very different for different periods of the day. Therefore, *a number* of day-to-day distributions should be considered, each for a different time of the day. In view of the differences within a peak period – the travel time distributions for the shoulders of a peak period are very different from the distribution for the peak of the peak period (see for example Van Lint et al, 2008) – considering one day-to-day distribution for each individual peak or off-peak period is still not sufficient. On the other hand, considering travel time distributions for each single minute of the day would be overdone. After all, travel times in two consecutive minutes will be almost equal.

For the peak periods, it was decided to consider three different travel time distributions (per peak): one in both shoulders of the peak, and one in the middle of the peak period. For the off-peak periods, one travel time distribution (per off-peak period: night, midday and evening) is considered sufficient, in view of the following considerations:

- The off-peak period travel time distribution is less variable over time than the peak period travel time distribution.[39]
- Although the off-peak distributions are certainly not irrelevant, they are considered less important than the peak period ones. After all, traffic congestion problems are typically largest during the peak periods.

Distribution skewness

Above, it has been stated that especially the width of the travel time distribution (representing the variation in travel times) is a very important factor for the costs associated with the uncertainty in travel times. It is very likely however, that these costs are not only related to the width of the travel time distribution (expressed in for example variance or standard deviation), but to the shape of the distribution as well. In this respect, an important characteristic is that travel time distributions often are skewed (left skew; long tail to the right). Van Lint et al (2008) state that the (economic) consequences of this skew are substantial: the consequences of extremely long delays may be much more severe than those of modest delays, since their result may be that appointments or trip-chain connections are completely missed instead of just delayed. Therefore, they argue that it should be preferred to use indicators in which besides the width of the travel time distribution, also its skew is incorporated.

How large the relative importance of the skew really is (i.e. its relative contribution to the total uncertainty-related costs), actually is not precisely known. To a large extent, this relative importance is determined by:

- the precise relationships between the amount of time arriving early or late and the costs associated with this (which of course will vary from trip to trip, depending on for instance travel purpose), and
- the way in which travelers respond to / anticipate on the characteristics of the travel time distribution (including the skewness), in scheduling their trips (including the selection of a certain 'buffer time').

Despite the uncertainty regarding the exact importance of the skewness, it seems reasonable to assume that its influence on the costs is quite significant indeed. For that reason, it seems wise to include this characteristic of the travel time distribution in the indicator(s).

[39] To avoid any confusion, please note that it is the temporal variability of the travel time *distribution* which is considered here, and thus not the temporal variability of the travel time *itself*. Contrary to the travel time distribution, the travel time itself might obviously strongly vary during the off-peak period, for example due to the occurrence of incidents or road works.

The role of the unpredictability of travel time variations

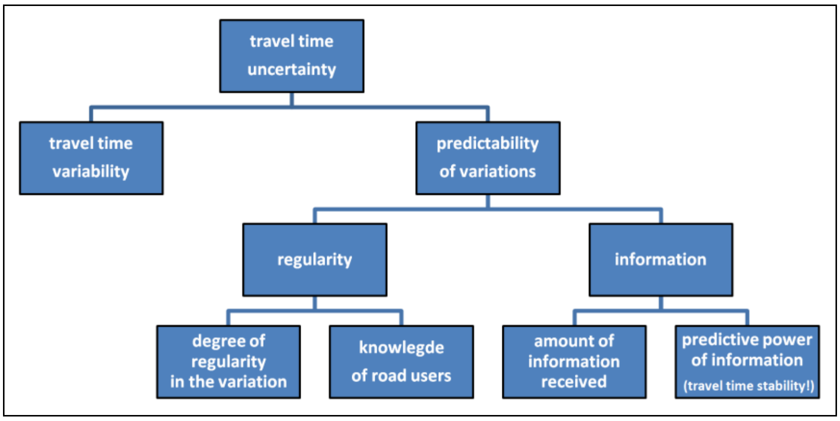
While the travel time distribution is an *important* indicator for the costs due to uncertainty, it is certainly not a *sufficient* one. The travel time distribution only describes the *variability* in travel times. This is not the same as the *uncertainty* in travel times. Given the variability in travel times, uncertainty is governed by the degree of *unpredictability* of this variability (see Figure 3.5). In the hypothetical situation in which the variation is completely predictable, of course there would be no uncertainty at all (and consequently no costs associated to this uncertainty either[40]). This makes the problem more difficult, since the variability in travel times is directly observable, but the degree to which this variation is predictable to the road users is not. In fact, this predictability varies greatly among the individual road users.

Basically, the predictability is governed by:

- the degree of regularity in the variation, and the extent to which any possible recurring patterns are known to the road users, and
- the amount of information that road users receive (traffic information, but also weather forecasts and announcements of road works), and the predictive power of this information.

First, the issue of regularity in the variation will be discussed. After that, the influence of information will be addressed.

Figure 3.5: Determinants of travel time uncertainty



[40] Late arrivals will not occur in this situation, because of the fact that the travel time variations can be planned for. Early arrivals (resulting in a (partial) waste of buffer times) will not occur either.

Regularity in the variation

Related to the existence of certain regular human activity patterns, some significant regular patterns can be observed in the travel time variation. The extent to which these recurring patterns are known to the road users is dependent on the amount of experience they have in the traffic system, and the relative contribution of the recurring variation to the total variation.

As argued before, the variation in travel times can be subdivided in two main components: 'within-day' variation and 'day-to-day' variation. On average, in the within-day variation clearly some regular pattern can be discerned: during the night travel time typically is lowest, and during morning and evening peak travel time is highest. Most travelers will have a certain (rough) awareness of this regular pattern. As such, this within-day variation is predictable to a certain extent, resulting in the uncertainty being lower than the global (i.e. overall) travel time distribution might suggest. It is virtually impossible to express this effect in numbers (or in modifications to the travel time distribution) though, if only because of the fact that each road user is different in this respect. Therefore, only the two extremes will be considered: road users that are completely unaware of the regular pattern in the within-day variation, and road users that have full knowledge of the within-day variation. It is assumed that the first group of road users is completely unaware not only of the regular pattern in the within-day variation, but of any possible regular patterns in the day-to-day variation (to be discussed below) as well.

As far as this first group is concerned, consideration should be given to the (unmodified) global (i.e. overall) travel time distribution. (Uncertainty = variation for these road users.) For the road users in the second group (having full knowledge of the within-day variation), only the day-to-day distribution is left over as source of uncertainty. As far as these road users are concerned, it thus is sufficient to consider the day-to-day distribution of travel times (like is the case for the road users making trips at a fixed time of the day).

Also in the day-to-day variation certain regular patterns can be identified. First of all, there is a certain regular variation between the various days of the week. In particular, there is a large difference between weekdays and weekend days. In weekend days, morning and evening peaks typically are absent in the travel times. However, events or other destinations attracting a lot of recreational traffic might result in peak-hour-like traffic conditions on some parts of the road network, during certain periods of the day. When mutually comparing the different *weekdays*, a certain regular pattern can be observed as well. Friday usually deviates most from the other weekdays. Travel times on this day typically have a less pronounced morning peak, and a longer lasting evening peak. Road users making the same trip every weekday probably will know this difference between Friday and the other weekdays.

Whether they know the mutual differences between the other weekdays however is doubtful. In this project, it is assumed that this is not the case. As a consequence, separate day-to-day distributions of travel time should be considered for:

- Monday-Thursday
- Friday
- Saturday
- Sunday

For the latter three categories then only the variation between weeks (for the given day of the week and a given time period of the day) is left over as a source of uncertainty.

However, also in the travel time variation between weeks (for a given day of the week) again some regular pattern can be identified. This regular variation over the months/seasons could be referred to as 'seasonal variation'. The most important component in this seasonal variation is the variation due to vacation periods. In these periods travel times are typically lower than they are normally, especially in the long vacation. The morning and evening peaks in travel time are less pronounced, or even (almost) absent. Most road users will know the difference in travel times between vacation periods and other periods (or otherwise they will learn this difference in a few days time, by experience). Therefore in fact separate (day-to-day) travel time distributions should be considered for 'normal' periods and vacation periods. However, in view of the fact that vacation periods account for only a limited part of the year, here only (indicators derived from) the travel time distribution for 'normal' periods (excluding vacation periods) will be considered.

At the start and the end of vacation periods just the opposite effect occurs: on these days traffic conditions typically are worse than normal, due to traffic volumes being larger (vacation rush). It is questionable whether much of the road users actually are able to take this into account in their time planning. It is very well possible that many road users do not realize in time that a vacation period is about to begin or end. (Of course this does not apply to the vacation traffic itself.) In this project, it is assumed that indeed the road users are not aware of it (in time). The travel times on the days concerned are therefore simply included in the travel time distributions relating to 'normal' days.

Yet another regular pattern in the travel time variation is related to holidays. On holidays, morning and evening peaks in the travel times are typically missing. In particular due to recreational traffic, there might still be traffic congestion on these days however. Because travel times on holidays are quite similar to those on Sundays, in this project holidays will be considered as being Sundays.

Finally, road users generally also have a certain idea of the impact (on travel time) of disturbing conditions like bad weather, accidents or road works. However, at the time at which such conditions manifest themselves to the road users, often it is already too late (i.e. the time planning cannot be modified anymore), *unless* some information on them is received in an earlier stage. The influence of this kind of information is the topic of the following subsection.

Influence of information

If travelers have access to / are provided with information on the conditions on the road network, they are better able to predict their travel times. This results in their uncertainty (and the associated costs) being lower than the travel time distribution suggests. The information referred to is not limited to traffic information. Other types of information, like weather forecasts and announcements of road works, may play a part as well.

Note that the influence of information is partly accounted for already (see above), by considering *origin-destination* travel time distributions (assuming fully informed road users, always selecting the shortest[41] route available) next to route travel time distributions. This way however only the reduction of travel time *variation* is considered, leaving the other component of uncertainty reduction (i.e. the uncertainty reduction for given travel time distribution) out of account.

The value of information is highly dependent on the point in time at which the information is obtained. The later the information is received, the lower its value (and consequently, the lower the reduction in the costs due to uncertainty). This is because of the fact that the possibilities to respond to the information decrease as a function of time.

Since planned (large scale) road works are usually announced well in advance, and moreover, usually last for a substantial period (enabling the regular road users to 'learn' the effects on the traffic conditions), it does not seem 'fair' to include the travel times in situations with this kind of road works in the (day-to-day) travel time distribution. It is better to leave these situations out of account, or (preferably) consider a separate travel time distribution for them. Of course, for emergency repairs this is completely different: travel times in situations with this kind of road works should be included in the 'normal' travel time distributions.

The reduction in uncertainty due to weather forecasts (or personal weather observation) could be taken into account by considering separate travel time distributions for different weather conditions. However, this would correspond to the (implicit) assumption that travelers have full knowledge on the differences in these distributions and that they are fully informed on the (future) weather conditions well in advance. Obviously, this is not realistic. Therefore, no separate travel time distributions will be considered. This corresponds to the implicit assumption that the road users have no weather related knowledge and information at all. Of

[41] 'shortest' in terms of travel time

course, this is not realistic either, but the previous assumption is estimated even more unrealistic.

It is very difficult to explicitly take into account the reduction in uncertainty due to information on the traffic conditions. This reduction is dependent on the amount of information that road users obtain, and on the quality of this information (and on the point in time at which the information is obtained, as mentioned before). The quality of the information refers to the extent to which the (future) travel times can be predicted correctly with this information. An important factor in this is the stability of the travel times. The more instable the travel times are, the lower the value of the information will be (i.e. the lower the reduction in the uncertainty in travel time and the associated costs). Here instability should be understood as the extent to which travel times can change rapidly and unpredictably. A high instability results in the travel times being ill predictable.

The instability of travel times can be expressed in various ways. Here it is proposed to look at the probability distribution of the difference between the *instantaneous route travel time*[42] (which typically is the travel time obtained from conventional traffic information) and the *actual route travel time*[43], for given time of the day. This will be done for several times of the day. Note that we are not only interested in the mean of the difference between the two travel times. In fact, this mean might be quite predictable: it may be related to a certain 'regular' variation over the day. Other characteristics of the probability distribution of the difference are important as well. If the distribution is very wide, with long tails, travel time instability is high.

3) *Increase of both fuel consumption and wear and tear*

Fuel consumption is dependent on the traffic performance (the total amount of vehicle-kilometers traveled) and the driving speeds. Obviously, total fuel consumption increases with the traffic performance. The relation with the driving speeds is more complicated. The fuel consumption is lowest for moderate speeds. For speeds larger than, say, 100 km/h, fuel consumption increases rapidly. For very low speeds, fuel consumption is larger as well. Also the extent of fluctuation in the driving speeds has an important impact on the fuel consumption. The greater these fluctuations are, the larger the fuel consumption will be.

By reducing the driving speeds from 120 km/h to for example 90 km/h (still free flow), the occurrence of busy traffic conditions might result in a reduction of the fuel consumption. However, if

[42] The instantaneous travel time is computed from the current traffic velocities on the network. It is based on the assumption that the traffic conditions do not change while traversing the route.

[43] The actual travel time for a given moment in time is the travel time that is truly experienced by road users departing at that moment in time.

traffic conditions get truly congested (speeds < 80 km/h), fuel consumption will increase. This is not (only) because of the lower (average) speed itself, but because of intensification of speed fluctuations as well (due to the emergence of stop-and-go traffic). Because of this intensification of speed fluctuations, wear and tear of the vehicles is increased as well. Furthermore, fuel consumption might be increased by traffic deviating to other (longer) routes, in order to get around heavily congested locations.

It may be concluded that this cost item can be taken into account by considering the total amount of vehicle-kilometers traveled (on network level or origin-destination level) and the total amount of lost vehicle hours relative to a reference level of, say, 80 km/h (on network, origin-destination, route, or road section level). It should be noted that this approach is not completely correct, since the fuel *savings* related to the speed reduction from 120 (or 100) to 80 km/h are not considered. However, in literature the additional fuel costs due to traffic congestion are estimated at only a few percent of the total congestion costs at the very most (see for instance KIM, 2009). Probably the costs related to the extra wear and tear are even less. It is thus not very important that the indicators to be used in this project very accurately represent this cost item.

4) Travelers diverting or staying away

This cost item can be addressed by considering the change in the number of road users. Of course a limitation of this approach is that the cost item can only be addressed in a *relative* way (i.e. comparing different scenarios, like the scenario *without* a certain measure and the scenario *with* this measure). It is not possible to give an indication of the *absolute value* of the magnitude of this cost item for a given scenario, using this approach. However, this actually is not really a problem, since it is not really necessary for the purpose of this research project.

It is not appropriate to consider the change in the number of road users at the level of individual road sections or routes. This way the results would be affected by detour making, while the costs associated to this are already addressed by considering the average travel times on the various origin-destination relations (cost item 1). Considering the change in the number of road users at the network level is not very appropriate either, since changes on different origin-destination relations might offset one another, resulting in these effects to remain unnoticed. Therefore, it seems best to consider the changes at the level of the origin-destination relations.

In order to properly grasp the influence of all (systematic and random) variations, the numbers of road users should be determined over a sufficiently long period of time (i.e. at least a year). By considering the numbers of road users separately for various days of the week and periods of the day, more detailed information is obtained. After all, since the levels of traffic congestion are variable over time, the extent to which travelers divert or stay away will strongly vary over time as well. More importantly, though, is that if all periods of the day are simply combined, the effect of travelers deviating from their preferred departure time (in order to avoid traffic congestion costs) is completely neglected, while there are definitely certain costs associated with this diverting behavior as well. This is another reason to consider the different periods of the day separately: this would visualize any possible shifts in departure times. While distinguishing between various days of the week and periods of the day thus certainly is preferable from a completeness point of view, it should be noted however that it would make the analyses more cumbersome.

5) *Increase of the number of accidents*

The relationship between traffic congestion and traffic accidents is not easy. On average, speeds are lower and mutual speed differences (between individual vehicles) are smaller in congested traffic, which will positively affect the frequency and severity of accidents. On the other hand, traffic density is larger, which will have a negative influence. More important, however, is that there are shock waves created in the traffic flow, in which the traffic has to slow down rather abruptly. This results in additional (head-tail) collisions.

This cost item can be taken into account by considering the total amount of lost-vehicle-hours relative to a reference level of, say, 80 km/h (on network, origin-destination, route, or road section level). The reasoning here is that a larger total amount of these lost-vehicle-hours corresponds to a larger number of shock waves, resulting in a larger number of accidents. This is a rather rough way of dealing with this cost item, but unraveling the exact relationships between traffic conditions and accidents is not the purpose of this study.

6) *Larger load on the environment*

The emissions of harmful exhaust fumes and the depletion of fossil fuels are respectively strongly and fully connected to the fuel consumption. Since these are the most important components of the environmental impact of traffic congestion, this cost item can thus be expressed in terms of variables describing the traffic conditions in rather the same way as cost item 3 (see above).

7) *Indirect costs related to settlement behavior and logistic chains*

This cost item is strongly related to cost items 1 and 2 (and 3): The longer and more uncertain the travel times are (and the larger the fuel consumption is), the larger the negative impact on settlement behavior and logistic chains will be. Therefore, this cost item can be related to the same traffic variables as cost items 1 and 2 (and 3) can be related to (see above).

8) *Discomfort*

The discomfort is strongly dependent on the (relative) occupancy of the road. Since the occupancy influences the travel speed, and therefore the travel time, the discomfort can be related to travel times. (The longer the travel times are, the larger the discomfort will be.) Furthermore, discomfort might be associated with uncertainty regarding the remaining part of the trip. Therefore, discomfort can be related to the uncertainty in travel times as well. This leads to the conclusion that the discomfort can be addressed in the same way as cost items 1 and 2.

The discomfort associated to the uncertainty regarding the remaining part of the trip is also influenced by the amount of information that travelers receive: generally travelers are less annoyed about delay if they are informed on the extent of this delay (Snelder et al, 2009). For the selection of an indicator / indicators for the level of traffic congestion this is not that relevant however.

Concluding observation

From the considerations above, it can be concluded that all societal costs in fact can be related to one or more of the following few characteristics of the traffic conditions:

- the average travel times (on origin-destination level),
- the uncertainty in the travel times (on origin-destination level),
- the number of road users (on origin-destination level),
- the total number of vehicle-kilometers traveled (on network or origin-destination level), and
- the total number of lost vehicle hours relative to a certain reference level of, say, 80 km/h (representing the 'boundary' between free flowing traffic states and congested traffic states) (on network, origin-destination, route, or road section level).

Therefore the criterions to look for should be criterions focusing on these aspects.

Of these five aspects, the first three are considered the most important ones. The fourth one is only included to properly take into account the additional fuel costs and environmental costs. As mentioned before, the additional fuel costs are only a few percent of the total congestion costs at the very most. Probably the additional environmental costs are of the same order of magnitude. The fifth aspect is included for the same purpose (i.e. properly representing the additional fuel costs and environmental costs). However, this one also has a function in representing the additional *accident* costs. Therefore, it is considered more important than the fourth one. The first three aspects are estimated even more important though.

3.3 Known criterions for the evaluation of traffic congestion

3.3.1 Outline

In this section the possibilities for defining criterion(s) for the level of traffic congestion are explored by considering which of such criterions already exist. First of all, a general overview is given of the various indicators that are used in practice or proposed in literature (subsection 3.3.2). Here, also their suitability in the context of this research project is discussed (taking into account the findings from the previous section). This is followed by an overview of the norms that have been put on traffic congestion in the Dutch national traffic policy of the past few decades (subsection 3.3.3). Here not only *the indicators* used in these norms are of interest, but the *limits set on these* as well. After all, from the value of an indicator alone one cannot infer whether the system actually performs well or poorly (in terms of traffic congestion)[44]. For this it is necessary to have an idea on the range of the indicator values for which the system can be considered to perform well.

The next section (3.4) describes the indicators finally selected. Insofar as the known criterions turned out to be inadequate, new or modified criterions were included in the selection.

3.3.2 Indicators used in practice or proposed in literature

In this section an overview is given of the various traffic congestion indicators that are used in practice, or proposed in literature. Their suitability for use in this research project is discussed as well. The indicators are considered in three groups. First, indicators (directly) relating to travel times, speeds or delays are discussed. After this, other indicators describing the quality of the traffic flow are considered. Although more indirectly, in a way these indicators relate to travel times, speeds or delays as well. Finally the indicators that do not really relate to the quality of the traffic flow, but rather to the amount of traffic, are discussed.

Indicators (directly) relating to travel times, speeds or delays

In practice, the total amount of lost vehicle hours (i.e. the sum of the delay of all vehicles, relative to a norm speed of for example 100 km/h) is the most common indicator for the severity of traffic congestion. Clearly this is not a sufficient indicator for the societal costs, since certain parts of these costs are completely disregarded, like those related to the uncertainty in travel times. For the average route velocity and the average travel time / delay (which are other indicators used in practice, representing the yearly average of the speed with which a

[44] Striving for a 'zero' congestion level typically is not optimal, since the costs that must be made to make the last bit of traffic congestion disappear are generally much larger than the associated benefits (i.e. the eliminated congestion costs).

certain route is traversed, and the yearly average of the travel time / delay on a certain route, respectively[45]), this is no different.

In the probabilistic design philosophy (see section 1.1), a widely used indicator for the evaluation of designs, or the mutual comparison of various decision options, is risk, often defined as the product of probability and consequence (or rather as the summation of the products of probability and consequence of the various possible outcomes). In the context of traffic congestion, the consequence typically would be the travel time, resulting in the risk value to represent the expected (or average) value of the travel time. Clearly then risk is not a sufficiently representative indicator for the social costs of traffic congestion either, again because of the fact that the additional costs due to the uncertainty in travel times are completely disregarded.

If the uncertainty-related costs *are* given consideration, often they are considered separately from the costs associated with the average/total travel times (expressed in e.g. the number of vehicle hours lost). For this, a certain unreliability indicator is used. In literature, various types of unreliability are discerned, like capacity reliability, terminal (or connectivity) reliability, encountered reliability, flow decrement reliability, and travel time reliability (see for example Nicholson et al, 2003). In practice the uncertainty-related costs are often addressed by considering the latter type of unreliability (i.e. the travel time reliability, or rather *un*reliability), since this type of reliability directly represents the impact on the road users, unlike some of the other types of reliability.

As mentioned before (section 3.2.3), there is no consensus on which indicator(s) to use for travel time unreliability. Most unreliability measures used/proposed in practice relate to the day-to-day variation on a particular route, for a particular time (period) of the day, limited to workdays (sometimes by day of the week). Many different indicators have been proposed. These can be divided in various classes. In (Van Lint et al, 2008) the following categorization is given:

- statistical range indicators,
- buffer time indicators,
- tardy trip indicators, and
- probabilistic indicators.

Statistical range indicators (like variance, standard deviation [46], coefficient of variation, or time windows defined by expected travel time plus/minus a certain number of times the standard deviation) only consider the width of the travel time distribution. The other indicator

[45] Both of these two indicators often are evaluated for a specific time period, usually being the peak or off-peak period of weekdays.

[46] In the Netherlands, it is suggested to use the standard deviation of travel times as a basis for the monetarization of travel time reliability effects in cost-benefit analyses of infrastructure projects. (See for instance Rand Europe, 2005.)

types focus on the tail(s) of the distribution. This way, to a certain extent they take into account ('combine') both the width and the skewness of the travel time distribution (since the tail is not only affected by the width of the distribution, but by the skewness (and other properties of the distribution) as well).

As the name implies, *buffer time indicators* consider a so-called 'buffer time'. This is the extra amount of time that a traveler should take into account in order to arrive on time, usually defined as the difference between the 90th or 95th percentile travel time and the average travel time. *Tardy trip indicators* focus on the extra delay (compared to the average travel time) incurred during the worst trips. An example of this type of indicators is the misery index, representing the relative distance between the average travel time of the 20% worst trips and the overall average travel time [47]. *Probabilistic indicators* finally consider the probability that the travel time between a certain origin and destination is within a certain predefined threshold or time window. Note that buffer time indicators and probabilistic indicators in fact are very closely connected. Buffer time indicators consider the buffer time for predefined probability of exceedance, while probabilistic indicators consider the probability of exceedance for predefined buffer time.

Van Lint et al. (2008) propose a new indicator for travel time unreliability (not covered by any of the above categories), explicitly combining width and skew of the travel time distribution. In this indicator, for width and skew not the 'classic' statistical measures are used, but new metrics based on percentiles of the travel time distribution. This is because of the fact that the well-known statistical measures are sensitive to outliers in the travel time data. Indicators based on percentiles of the distribution are more robust in this respect.

In their paper, Van Lint et al. (2008) illustrate that the various indicators differ significantly in judging the reliability of situations. In fact this is quite logical, since all these indicators depict only part of the large amount of information contained in the travel time distribution. This problem is inherent to using indicators. Another source of the discrepancies is the fact that part of the indicators are based on statistics that are sensitive to outliers in the travel time data (statistics like mean and variance), resulting in these indicators being sensitive to these outliers themselves as well.

Recently Tu (2008) defined a new indicator for travel time unreliability, combining travel time variability and travel time instability. Tu considers travel time unreliability as a kind of risk: the risk with respect to experiencing a traffic flow breakdown (or traffic congestion), faced by road users. This risk is considered as a function of the traffic inflow to the route. Tu uses another definition of risk than the one discussed in

[47] Other definitions of the misery index are possible as well (for example: instead of the 20% worst trips also the 15% or 10% worst trips might be considered, or instead of the overall average travel time the free flow travel time might be taken).

the beginning of this subsection[48]. The general concept (risk is the probability of outcome *i* times the consequence of outcome *i*, summed over all possible outcomes *i*) is the same, but the possible outcomes are defined in a different way.

Tu discerns two possible outcomes: the traffic state being congested and the traffic state being free flow (i.e. non-congested). The respective travel time variabilities of these two traffic states (i.e. the congested travel time variability and the free flow travel time variability, respectively) are considered as the consequences of the two possible outcomes. The breakdown probability (i.e. the probability of the traffic state transforming from free flow into congested) and its complement (i.e. the probability of the traffic state remaining free flow) are considered as the corresponding probabilities of occurrence.

From this it can be concluded that the indicator of Tu actually corresponds to the *expected value* of the travel time variability. Since the probabilities of occurrence are defined as the probabilities of breakdown and 'no breakdown', rather than being defined as the probabilities of congestion and 'no congestion', this expected value is not the *overall* expected value, but rather the expected value *assuming initially free flow conditions.*

In fact, this is no good indicator for representing the total uncertainty. For one thing, this is because of the fact that the indicator does not take into account the possibility that traffic conditions are congested already (due to a breakdown of the traffic flow at an earlier moment in time). Secondly, one of the components of the total uncertainty is neglected in the indicator, namely the uncertainty regarding whether the traffic conditions will be congested or not.

This can be illustrated with the following (theoretical) example. Consider a situation in which the traffic conditions are either free flow or congested, and variability in both cases is zero (i.e., there is only one possible free flow travel time, and likewise only one possible congested travel time). Of course the congested travel time is longer than the free flow travel time. The travel time unreliability according to the definition

---

[48] The indicator proposed by Tu is computed as:

$$TTUR(q_{in}^r) = (1 - p_r^{br}(q_{in}^r)) \times TTV^f(q_{in}^r) + p_r^{br}(q_{in}^r)) \times TTV^j(q_{in}^r) \, .$$

$TTUR(q_{in}^r)$ = Travel time unreliability for a given inflow level q_{in} on route r.

$p_r^{br}(q_{in}^r)$ = Probability of traffic breakdown for a given inflow level q_{in} on route r. (Note that the probability of *breakdown* is something different than the probability of *congested traffic flow*. In the first case the transition from free flow to congested flow is meant, while in the latter case the traffic flow might have been congested already.)

$TTV^f(q_{in}^r)$ = Travel time variability before breakdown (i.e. in free flow conditions) for a given inflow level q_{in} on route r (computed as the difference between the 10th and 90th percentile travel times).

$TTV^j(q_{in}^r)$ = Travel time variability after breakdown (i.e. in congested conditions) for a given inflow level q_{in} on route r (computed as the difference between the 10th and 90th percentile travel times).

of Tu (i.e., the expected value of the variability) would be zero. Still, there is of course uncertainty involved, namely the uncertainty associated with the question whether the traffic conditions will be free flow or congested, i.e. whether one will face the free flow travel time or the congested one.

Other indicators for the quality of the traffic flow

While all the indicators above are directly related to travel times, in practice also indicators not directly related to travel times are used for describing the quality of the traffic flow. The most well-known example of these indicators is the ratio of traffic flow (q) and capacity (c), indicating the relative loading of a certain road section. In this ratio for q and c certain 'representative' values are taken. For q this for example could be the maximum flow rate in the governing peak hour on an 'average' working day, or the flow that is exceeded a certain given number of hours a year (say, 30). For c it for example could be the average capacity in normal conditions. The indicator q/c is frequently used for assessing the quality of the traffic operations (also indicated as 'LOS': Level of Service) on (designed/planned) road sections. The higher the flow-capacity ratio, the lower the level of service, expressed in terms of comfort, freedom to maneuver (including for example overtaking possibilities), freedom to drive at the desired speed, travel time, and predictability of travel time. A general rule of thumb is that a value smaller than 0.8 corresponds to good quality traffic operations (i.e. hardly any traffic congestion), while a value of about 1 or above corresponds to bad traffic operations (i.e. a substantial amount of traffic congestion) (AVV, 2002; Rijkswaterstaat, 2009).

Another indicator that can be used for assessing the quality of the traffic flow is the traffic density (i.e. the amount of vehicles per unit length of roadway/lane). The larger the traffic density is, the lower the quality of the traffic operations will be. Based on traffic densities, in the 'Highway Capacity Manual' (HCM) (Transportation Research Board, 2000) six levels of service are distinguished, ranging from completely free flow conditions in which vehicles are virtually unaffected by each other (LOS A), to forced flow conditions in queues resulting from traffic demand exceeding available capacity (LOS F). This level of service classification has been widely used in planning, design and operational studies for several decades now.

In fact, the traffic density and the flow-capacity ratio are strongly connected. After all, traffic density and traffic flow can simply be converted into one another using standard relationships from traffic flow theory.

Yet another indicator used in practice is the 'probability of congestion'. For a given road section, this indicator gives the fraction of traffic experiencing traffic congestion.

Finally, an indicator which is commonly used in the Netherlands is the summation of the lengths of all traffic jams on the network, integrated over time (in Dutch referred to as the 'filezwaarte' or 'filedruk'). Here a traffic jam is supposed to exists on a road section if the speed on that road section is below 50 km/h. This is clearly not a good indicator, however, in view of the facts that:

- The length of a traffic jam is not a good measure for the amount of road users involved. For this the number of lanes and the traffic density should have been taken into account as well.
- The indicator does not differentiate between different speed levels (below the 50 km/h threshold), while these speed differences may have major consequences for the amount of delay incurred in a traffic jam.

While all these indicators (except for the latter) certainly give relevant information on the quality of the traffic flow, in fact the indicators discussed before (i.e. those related to travel times) are better indicators for the societal costs, since they describe more directly the magnitude of the effects that actually bring about these costs. Of course, there is a relation between the indicators discussed in this subsection and the indicators related to travel times. As far as the traffic density and flow-capacity ratio are concerned, on average it will be true that the (total/average) travel time and uncertainty in travel time are larger on road sections with higher flow-capacity ratios / traffic densities. However, this relation is only a very rough one. This is due to the fact that important factors like network structure and variations in both traffic demand and capacity are not reflected in the indicators flow-capacity ratio and traffic density. Consequently these indicators cannot be considered substitutes for indicators directly related to travel times[49].

The 'probability of congestion' indicator is much stronger related to the (total/average) travel time and the uncertainty in travel time than the flow-capacity ratio and traffic density. A larger probability of congestion generally will correspond to a larger total/average travel time and a larger uncertainty in travel time. This relation cannot be regarded as a hard and fast rule though. Therefore also this indicator cannot be considered a substitute for indicators directly related to travel times.

Other indicators

Beside the indicators that relate to the *quality* of the traffic flow, there are the indicators concerning the *amount* of traffic (which after all might be affected by traffic congestion as well). In the Netherlands, the number of vehicle-kilometers probably is the most common indicator for measuring the total amount of traffic. A problem related to this indicator is that it is not always clear how to explain an increase (or decrease) in the number of vehicle-kilometers. On the one hand, such

[49] In spite of the fact that the flow-capacity ratio and traffic density are not very appropriate indicators for assessing the actual size of the traffic congestion problems, in design processes they may be useful, since they might provide the traffic engineer with information on underlying causes of traffic congestion problems.

an increase could be due to more traffic being facilitated by the road network, which is generally considered a positive effect (in view of the economical/societal benefits that are derived from the additional trips)[50]. On the other hand, such an increase could also be due to more road users making (larger) detours in order to avoid congested locations, which of course is considered a negative effect. Therefore it seems better to measure the amount of traffic that is facilitated by the road network in numbers of trips (per origin-destination relation).

Brilon (2005) combined the amount of traffic being facilitated by a road section and the quality of the traffic flow, represented by the speed on this section, into one indicator (the 'traffic efficiency'), by taking the product of these two. This indicator indeed is a good measure for the efficiency with which the potential of the existing infrastructure is exploited. However, it is not that appropriate as indicator for the societal costs of traffic congestion, since detours (made in order to avoid congested locations) are not negatively valued.

3.3.3 Norms on traffic congestion in the Dutch traffic policy

This section zooms in on the norms that have been set on traffic congestion in the Dutch traffic policy of the past few decades. For the most part, this policy is laid down in the main policy documents on traffic and transport of the Dutch government. Over the past few decades, four of these documents have been released:

- the 'Traffic and Transport Structure Plan' (In Dutch: 'Structuurschema Verkeer en Vervoer', abbreviated as SVV), published in 1979;
- the 'Second Traffic and Transport Structure Plan' ('Tweede Structuurschema Verkeer en Vervoer', abbreviated as SVVII), released in 1990;
- the 'National Traffic and Transport Plan' ('Nationaal Verkeers- en Vervoersplan', abbreviated as NVVP), drawn up in 2000; and finally the most recent one:
- the 'Mobility Policy Document' ('Nota Mobiliteit', abbreviated as NoMo), dating from 2004.

SVV (1979)

In the SVV, policy objectives were based on the 'levels of service' as defined in the American Highway Capacity Manual (a classification in six levels of service A to F, based on the traffic density[51]). For weekdays, LOS C was selected as the policy objective for rural areas, and LOS D (limit E) for urban areas (Van der Hoorn, 2007). In fact, the Highway Capacity Manual (HCM) classification had already been used for a long time in the Netherlands when the SVV was issued. As discussed before, the traffic density and consequentially also the HCM classification however are not the most appropriate indicators for expressing the

[50] On the underlying road network such an increase however could be considered undesirable, in view of a possible deterioration of the traffic safety or quality of life.

[51] See the previous subsection.

level of traffic congestion. Indicators directly related to travel times are more appropriate for this.

SVVII (1990)

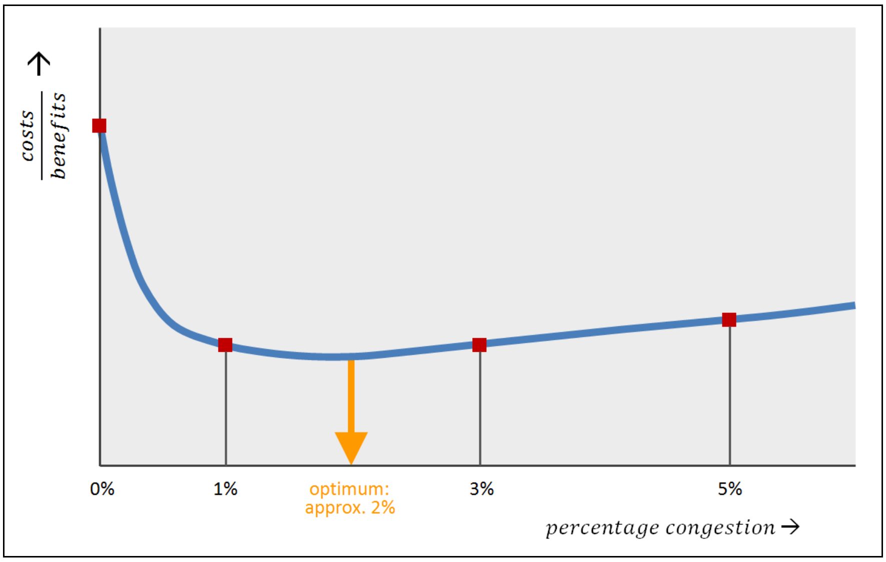
In the course of years, in the decision-making process for infrastructure investments it became increasingly important to map out the (economic) effects of alternative options quantitatively (Stembord, 1991). Because of the somewhat 'abstract' qualitative nature of the HCM levels of service (resulting in administrators having difficulties in interpreting these), a need was felt for a more intuitive indicator: an indicator with which it would be possible to *quantify* the level of traffic congestion. Therefore, in SVVII such a new indicator was introduced: the 'congestion probability' (discussed in the previous subsection). For a given road section, this congestion probability was defined as the fraction of all road users in a year (limited to weekdays) that is confronted with traffic congestion. In the computation of this congestion probability, within-day and day-to-day variations in the traffic demand were taken into account. To a certain extent, the variations in the capacity of the road section were taken into account as well.

Typically, a 'zero' congestion probability is not optimal: the infrastructure expansion needed to get rid of the last bit of congestion is much more costly than this last bit of congestion itself. Therefore a societal cost-benefit analysis was performed in order to find a norm for the congestion probability. In this cost-benefit analysis, improvements in travel time and traffic safety (i.e. the societal benefits) were weighted against the construction costs of new infrastructure, the maintenance costs of infrastructure, and the environmental effects of the construction of new infrastructure (i.e. the societal costs).

Overall (i.e. considering all individual road sections of the main road network together), a congestion probability of 2% was found to be the economic optimum (see Figure 3.6). In the end, only for the so-called 'hinterland connections' this 2% actually was selected as the norm. In view of budgetary restrictions, for the other roads of the main road network a less stringent norm of 5% was selected. (The traffic on the hinterland connections was considered to be of a larger economic value, resulting in travel time losses on these roads to be considered more costly than travel time losses on other roads.)

It should be noted that these norms were not to be interpreted as strict rules to be complied with for individual road sections. In fact, they were rather meant as a basis for a nationwide reservation of financial resources and land. It was stated that for actual projects a case-by-case assessment of the local optimum congestion probability had to be performed.

Figure 3.6: Societal optimization of the congestion probability
(Adapted from: Dienst Verkeerskunde, 1992)



In subsection 3.3.2 it was already noted that the congestion probability in fact is not the most appropriate indicator for expressing the level of traffic congestion. It does indicate the proportion of the road users that is confronted with traffic congestion, but does not give adequate information on the actual extent of this traffic congestion (in terms of attributes that are decisive for the societal costs). For this it is better to consider indicators (directly) related to travel times.

Concerning the societal optimization from which the 2% norm was derived, some critical remarks can be made as well. The societal costs and benefits of infrastructure expansion were accounted for in a rather simplistic way. In reality, assessing these costs and benefits is a rather complex problem. First of all, as far as the delays are concerned, only the direct travel time losses were considered. The additional costs due to uncertainly in travel times were left out of account. In view of the fact that that these costs contribute significantly to the total costs of traffic congestion (see section 3.2.2), this is an important limitation of the analysis performed. The indirect economic costs related to the effects on the location choice behavior of firms were left out of account as well. Furthermore, the feedback process from infrastructure supply to traffic demand (related to the notion of the so-called 'latent demand') seems to be neglected. This feedback process refers to the increase in traffic demand that can be expected if traffic conditions are improved (as a result of infrastructure expansion). This increase results in both additional costs (related to an increase in the amount of traffic congestion) and additional benefits (related to the larger amount of trips being facilitated), which should have been taken into account.

Another limitation of the analysis relates to the fact that certain causes of congestion were not taken into account explicitly (causes like incidents, road works, heavy traffic in weekends, etc.). These sources of congestion were accounted for by multiplying the calculated travel time losses by a factor of 2. In spite of the fact that this factor had an empirical basis, the accuracy of this approach is doubtful. Yet another limitation of the assessment was that all individual road sections were considered separately (i.e. as if they were isolated from each other). Interactions between bottlenecks were thus not taken into account in the assessment.

NVVP (2000)

In the NVVP the indicator 'congestion probability' was discarded again, because it was considered not closely enough related to the perception of the travelers (AVV, 2000). It was replaced by the so-called 'route speeds' (in Dutch: 'trajectsnelheden'), defined as the average speeds on motorway stretches of 30 km or longer, for the busiest hour of the day (averaged over all weekdays of a year). Based on feasibility, a lower limit of 60 km/h was selected as the policy objective for these route speeds.

Initially, in the computations of the route speeds for future situations, fluctuations in traffic conditions were not taken into account. Instead, simply speeds for a kind of 'representative situation' were calculated. Later on, Transpute (2003) developed a model with which route speeds could be calculated as true averages (i.e. taking into account (part of the) fluctuations in traffic conditions).

Note that using average speeds as indicator is in fact no different from using average travel times as indicator, since speed is just distance over time (where the distance has a fixed value). This can be considered to be a better indicator than indicators not directly related to travel times, but, as discussed before (see section 3.3.2), it obviously is not a *sufficient* indicator for the societal costs of traffic congestion, since it completely neglects, among others things, the costs that can be related to the uncertainty in travel times. Furthermore it should be mentioned that the theoretical basis for the norm speed of 60 km/h seems rather weak.

NoMo (2004)

When the NoMo (i.e. the most recent policy document) was put into force, the policy was changed once again. The policy objective from the NVVP (relating to the average route speeds) was modified, and a completely new policy objective was added to it. This new objective refers to the *reliability* of travel times, which is considered a very important issue in the NoMo. In fact it is thus recognized in the NoMo what was stated above, in relation to the policy objective in the NVVP, i.e. that the average speed (or travel time) is not a sufficient indicator for the level of traffic congestion. By adding the policy objective regarding the reliability of travel times, to a certain extent the uncertainty in travel times is addressed as well. It should be noted however that this reliability policy objective rather considers the *variability* in travel time, which is not exactly the same as the *uncertainty* in travel time (as discussed in section 3.2.3).

For the average travel times, the following targets have been defined (Ministerie van Verkeer & Waterstaat, 2004):

- On motorways, the average travel time during peak periods should not exceed 1.5 times the off-peak travel time.
- On urban ring roads (a subcategory of the motorways) and non-motorways that are part of the main road network, the average travel time during peak periods should not exceed 2 times the off-peak travel time.

In the above, the off-peak travel times should be calculated based on an assumed speed of 100 km/h (except for the non-motorways). The factors of 1.5 and 2 are referred to as 'travel time factors'.

The less stringent norm for urban ring roads and non-motorways is motivated by the facts that:

- For urban ring roads capacity expansion is often only possible at very high costs.
- On urban ring roads and non-motorways, the percentage of regional traffic (as opposed to national and international traffic) is relatively high. For this regional traffic (which travels shorter distances) a less stringent norm results in only a few minutes extra travel time.
- On non-motorways there are often elements like roundabouts, intersections and other elements negatively influencing the travel times.

In fact these norms are very similar to the 'route speed norm' defined in the NVVP. However, for the urban ring roads this norm is relaxed a bit, while for the other motorways the norm is tightened a little. After all, an average travel time of 2 times the off-peak travel time corresponds to an average speed of 100/2 = 50 km/h (assuming an off-peak speed of 100 km/h), and an average travel time of 1.5 times the off-peak travel time corresponds to an average speed of 100/1.5 = 67 km/h (where the NVVP norm was 60 km/h). It should be noted however that the NVVP norm referred to the busiest hour of the day, whereas the new norms apply to peak periods as a whole (being longer than just one hour). With this in mind, the 67 km/h norm in the NoMo actually might even be *less* stringent than the 60 km/h norm in the NVVP.

Next to these norms on *route* level, the NoMo also defines a norm on the level of the *main road network as a whole*. By 2020, the total amount of vehicle hours lost in traffic jams should be back at the level of 1992 (Ministerie van Verkeer & Waterstaat, 2006). In this context, traffic jams are to be understood as traffic with a speed below 50 km/h (Savelberg, 2008).

For the reliability of travel times, in the NoMo a probabilistic indicator (see section 3.3.2) is used. The policy objective states that 95% of all trips during the peak periods should be in time (Ministerie van Verkeer

& Waterstaat, 2004)[52]. For longer distance trips (above 50 km) 'in time' is defined as arriving 20% earlier or later than the expected travel time at the most. (In other words: the realized travel time should be within the interval bounded by respectively 80% and 120% of the expected travel time.) For shorter distance trips, 'in time' is defined as arriving within 10 minutes of the expected travel time (i.e. arriving 10 minutes early or late at the most). In these definitions, the 'expected travel time' is defined as the median of the travel time distribution.

Van Lint et al. (2008) criticize the arbitrary nature of the parameterization of these norms (i.e. the arbitrariness of the 95% level and the 20% and 10 minutes thresholds). If these parameters are chosen differently, the results might give a very different picture of the travel time reliability. The choice for using a *probabilistic* indicator (rather than one of the other types of reliability indicators discussed in section 3.3.2) to a certain degree is arbitrary too. In combination with the fact that these different types of indicators differ significantly in judging the reliability of situations, this makes the norms rather arbitrary as well.

Consequently, the value of these norms as policy criteria in fact is fairly limited. This can only be improved by making the norms more objective (both regarding the selection of an indicator and the parameterization of the norm). For this, it will be necessary to improve the knowledge on how the costs of uncertainty in travel times actually are related to the characteristics of the travel time distribution. Ideally, the parameterization of the norms should be based on a societal cost-benefit analysis. However, from the discussion in relation to the SVVII it might be clear that such a cost-benefit analysis is very complex indeed. Furthermore, for different parts of the motorway network (the results of) this cost-benefit evaluation might well be rather different.

3.4 Selected criterions

From the previous considerations it is clear that traffic congestion causes costs to society in several ways. The various cost items are related to the characteristics of the traffic congestion in different ways. One might be inclined to say that, ideally, the level of traffic congestion should be expressed in one single indicator, being the total amount of societal costs. However, for this it would be necessary to monetarize all individual cost items, which is not easily accomplished (since it is not entirely clear how the costs exactly relate to the characteristics of the traffic congestion). Actually, it (i.e. considering the level of traffic congestion in terms of total costs) would not be very insightful either. For these two reasons, it is more appropriate to express the level of traffic congestion in variables describing the traffic conditions. Since the different cost items are related to the traffic conditions in different

[52] In this norm 'all trips' refers to the complete set of all trips on the entire main road network. The norm is thus not applicable to individual routes, roads or road sections (Rijkswaterstaat, 2009)

ways, one indicator is not sufficient then. Instead, a *set of indicators* needs to be considered. Based on the considerations in the previous sections, the following indicators were selected:

I. The 90th percentile travel time (on origin-destination level)

This indicator was selected because of the fact that it (to a certain degree) implicitly comprises three features that are very important for the level of the societal costs of traffic congestion, namely the average travel time, the variance of the travel time and the skewness of the travel time distribution. After all, for each of these three features it applies that the larger its value is (holding the other two constant), the higher the 90th percentile travel time will be. The 90th percentile travel time therefore is a reasonable indicator for (part of) the societal costs due to traffic congestion.

The 90th percentile travel time is made dimensionless by dividing it by the free flow travel time. This has two main advantages:

- it puts the level of traffic congestion into perspective, and
- it makes the values on the different origin-destination relations better comparable.

The latter allows to plot the values for the various origin-destination relations in one diagram (think of a bar chart). This helps to facilitate the analyses.

In view of the arbitrariness of the choice for the *90th* percentile travel time, it seems advisable to consider some other percentiles as well. (For example the 80th and 95th percentile travel times).

II. The average travel time (on origin-destination level)

The 90th percentile travel time (and the other two percentile travel times considered) represent only a very limited part of the information contained in the travel time distribution (and consequently give only limited insight into the amount of additional travel time and the uncertainty in travel times). Therefore it is necessary to also consider some other indicators for these two aspects[53]. One of the indicators selected for this is the average travel time (on origin-destination level). Again this indicator is made dimensionless by dividing it by the free flow travel time.

III. The median travel time (on origin-destination level)

In section 3.3.2 it was discussed that the classic statistics, like average and variance, are sensitive to outliers in the (measured or simulated) travel time data, which is a drawback of these indicators. This is not the case for indicators based on percentile values of the travel time distribution (provided that not the most

[53] A better approach might be to consider the travel time distribution as a whole (by visual inspection). This way no information is lost. However, comparison between (many) different situations is more difficult then. The advantage of using indicators is that these allow for quick comparisons. Therefore it was decided to hold on to the strategy of using indicators. It is however advisable to give some (limited) consideration to the distribution as a whole as well, in order to avoid missing important information.

extreme percentiles are considered of course). Therefore besides the average travel time also the median travel time (or equivalently, the 50th percentile travel time) has been included in the set of indicators. Of course this could make one wonder why then the average travel time still is included in the set of indicators. This can be explained by the fact that the median travel time in fact is not an ideal indicator either. After all, the upper (or lower) half of the travel time distribution can significantly improve or deteriorate, without being 'detected' by the median (where the average would be affected). In view of both the average and the median having their individual drawbacks, it was decided to consider both of them in this study (i.e. one as 'safeguard' for the other). Also the median is made dimensionless by dividing it by the free flow travel time.

IV. The width of the travel time distribution (on origin-destination level)
In section 3.2.3 the variation in the travel times was identified as an important factor (though not the only one!) for the travel time uncertainty. Therefore the width of the travel time distribution was selected as an indicator too. In order to avoid sensitivity for outliers in (measured or simulated) data, this width is not defined in terms of the classic statistics (like standard deviation, variance or coefficient of variation), but rather as the difference between the 90th and 10th percentile travel time values. Also this width indicator is made dimensionless by dividing it by the free flow travel time.

V. The skewness of the travel time distribution (on origin-destination level)
As indicated in section 3.2.3, also the skewness of the travel time distribution might be an important factor for the costs associated with the uncertainty in travel times. For that reason, an indicator representing this skewness was selected as well. This indicator is defined as the quotient of the difference between the 90th and 50th percentile travel time values and the difference between the 50th and 10th percentile travel time values (i.e. $(TT_{90}—TT_{50})/\ (TT_{50}—TT_{10})$). The choice for a definition in terms of percentile values (rather than using the classic skewness statistic) again was motivated by the wish to avoid sensitivity to outliers in (measured or simulated) data.

VI. The distribution of the difference between the instantaneous and actual route travel time (on route level)
In section 3.2.3 it was explained that the instability of the travel times is an important factor for the travel time uncertainty. In case of a higher travel time instability the predictive power of traffic/travel information (which pertains usually to the instantaneous situation only) will be lower, resulting in the travel time uncertainty to be larger (for the same level of travel time *variability*). In the same section the distribution of the difference between the instantaneous and actual route travel time was identified as an appropriate basis for evaluating this travel time instability.

In fact, one would rather prefer to use a real indicator for this (i.e. to express the instability in a single number). However, it is difficult to define such an indicator without risking losing important information contained in the distribution. (As noted in section 3.2.3, we are not *only* interested in the mean of this distribution, since this mean actually might be related to a rather regular, predictable variation over the day). That is why it was decided to consider the distribution as a whole. Because of the fact that this results in the analyses to be more demanding (after all, comparing whole distributions is much more laborious than comparing indicator values), the travel time instability is evaluated only for a limited number of times of the day.

VII. The average number of road users (on origin-destination level)
This indicator was selected as a measure for the costs due to travelers diverting (to other transport modes, destinations, or departure times) or staying away because of the traffic congestion (or the other way around: as a measure for the benefits derived from additional trips induced by an improvement of the traffic conditions). It should be evaluated separately for the peak periods, the 'shoulders' of the peak periods, and the off-peak periods, in order to be able to show shifts in departure time period.

VIII. The total number of vehicle-kilometers traveled (on network level)
In order to allow for the additional fuel consumption and the environmental effects of the traffic congestion as well, the total number of vehicle-kilometers traveled has been included in the set of indicators, in combination with the total number of lost vehicle hours (relative to a certain reference level; see below).

IX. The total number of lost vehicle hours relative to a reference level of 80 km/h [54] (on network level)
In combination with the indicator above, this indicator was included as a measure for the additional fuel consumption and the environmental effects of the traffic congestion. It was also included however to allow for the additional accident costs due to traffic congestion (based on the line of reasoning that a larger amount of congestion hours corresponds to a larger amount of shock waves, which in turn corresponds to a larger number of (congestion-related) accidents; see section 3.2.3).

[54] Note that normally, in Dutch traffic policy a reference level of 100 km/h is used for the number of lost vehicle hours. In such cases, this indicator is used as a measure for *delays*, however, whereas here it is used as a measure for *additional fuel consumption*, *accident costs* and *environmental effects*. (Delays are considered by means of other indicators.) In this case, a reference level of 100 km/h is not appropriate. Rather, a reference level is required which can be considered to represent the 'boundary' between free flowing traffic states and congested traffic states. For this, a level of 80 km/h is taken.

All indicators on origin-destination level (I–V and VII) or route level (VI) should be considered for several origin-destination relations / routes (well distributed over the network, and comprising both shorter and longer distances), and for various times of the day.

Without specifying to which travel time distributions the indicators I to V (inclusive) relate, they of course do not have a real meaning yet. Based on the considerations described in section 3.2.3, it was decided that the following distributions should be considered:

i. The **global** (overall) travel time distribution (i.e. the travel time distribution for all moments in time combined, comprising both the day-to-day variation and the within-day variation), assuming a **fixed route choice** (i.e. assuming that the road users in all situations hold on to their standard/intended routes).

ii. The **global** (overall) travel time distribution (i.e. the travel time distribution for all moments in time combined, comprising both the day-to-day variation and the within-day variation), assuming an **optimal route choice** (i.e. assuming that the road users at all times are able (and willing) to select the route that will yield them the shortest travel time)[55].

iii. The **day-to-day** travel time distributions (i.e. the travel time distributions for fixed times of the day, compromising only the day-to-day variation), assuming a **fixed route choice** (i.e. assuming that the road users in all situations hold on to their standard routes).

iv. The **day-to-day** travel time distributions (i.e. the travel time distributions for fixed times of the day, compromising only the day-to-day variation), assuming an **optimal route choice** (i.e. assuming that the road users at all times are able (and willing) to select the route that will yield them the shortest travel time)[55].

All of these distributions are defined with respect to time (as opposed to being defined with respect to trips), a choice that was based on the considerations in section 3.2.3. Distribution types iii and iv are defined separately not only for the different times of the day, but for the different categories of the day of the week as well (Monday-Thursday, Friday, Saturday, and Sunday). Here public holidays should be considered as Sundays. Vacation periods and periods with large-scale road works should be excluded from the day-to-day distributions[56].

[55] This travel time distribution is constructed by selecting for each 'measurement' time interval the shortest travel time available (among the travel times of the various alternative routes on the origin-destination relation). Note that the indicators in this case are made dimensionless by dividing the indicator value by the *smallest free flow travel time* (considering all routes between the origin and destination).

[56] In section 3.2.3 it was argued that it seems not 'fair' to include these periods in the day-to-day travel time distributions, since this would result in the uncertainty in the travel times to be overestimated. Obviously, it would even be better (though more time-consuming) to consider separate day-to-day travel time distributions for these periods (rather than just leaving them out of account).

In section 3.3.1 it was indicated that it is desirable to have some reference values for the indicators considered (indicating when the traffic system can be considered to perform well), since striving for a 'zero' congestion level typically is not optimal. For this in the previous section the norms put on traffic congestion in the Dutch traffic policy were considered. It turned out that the theoretical basis for the parameterization of these norms actually seems to be rather weak. It was discussed that this parameterization in fact should be based on a comprehensive societal cost-benefit analysis. However, it is not within the scope of this study to find the societal/economical optimum. Therefore, the desired reference values cannot be given here.

3.5 Empirical relations between the average travel time and other indicators based on travel time statistics

In various empirical studies, strong relationships are found between the average travel time (or the average delay, corresponding to the difference between the average travel time and the free flow travel time) and other indicators based on the travel time distribution. This section discusses these findings (subsection 3.5.1) and their implications for this research project (subsection 3.5.2).

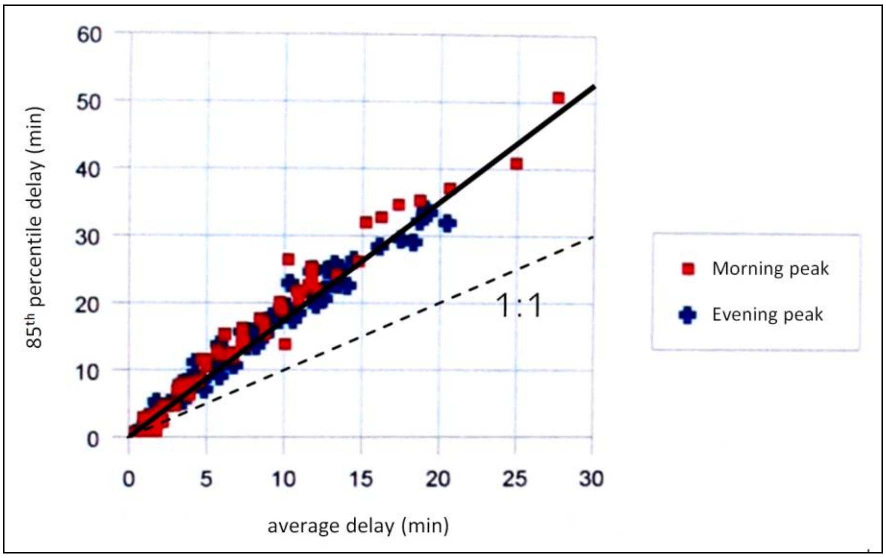
3.5.1 Empirical evidence suggesting the existence of relationships between the average travel time and other travel time statistics

Van Toorenburg (2003), Rand Europe (2004) and Margiotta (2009) all found empirical evidence suggesting the existence of certain relationships between the average travel time (or delay) and other indicators based on travel time statistics. Below, the findings of these different researchers are briefly discussed (one-by-one).

Van Toorenburg (2003) plotted the 85^{th} percentile delay against the average delay (both in the busiest hour) for 36 routes of about 30 kilometers throughout the Netherlands, and found a rather strong relation between these two (see Figure 3.7). It should be noted that the existence of a positive interdependency in itself is of course quite logical. After all, a larger 85^{th} percentile value indicates that large delays do occur more frequently, which will be reflected in the average value as well.

From the figure it is apparent that the differences in average delay between the different routes are considerably smaller than the differences in 85^{th} percentile delay. Each additional minute 85^{th} percentile delay roughly corresponds to 0.6 minutes additional average delay. (Or formulated the other way around: each additional minute average delay roughly corresponds to 1.75 minutes additional 85^{th} percentile delay.) From this it can be concluded that a larger average delay and a larger (day-to-day) variability in the delay go hand in hand. From the figure, this relation seems to be fairly strong.

Figure 3.7: Relation between 85th percentile delay and average delay (in the busiest hour) for 36 routes of about 30 kilometers throughout the Netherlands
(Adapted from Van Toorenburg, 2003)



The existence of this relationship between average delay and variability in delay is not surprising. Larger traffic loads generally do not only result in longer average travel times, but also in an increased sensitivity to disturbances. This latter causes the travel time variability to be larger as well. In addition, the existence of a speed limit might play a role. Due to this speed limit, the travel time distribution is bounded to the left. Because of this, an increase in variability generally corresponds to the travel time distribution extending further to the *right*, which on its turn corresponds to a larger average value.

Based on the apparent strength of the relationship, and given the fact that the data points represent a wide variety of circumstances (rural roads versus urban roads, roads with much freight traffic versus roads with little freight traffic, etc.), Van Toorenburg concludes that it will not be easy to attain a state that deviates from this relation (by taking measures).

In a research project of Rand Europe (2004) comparable relations were found. This can be illustrated with Figure 3.8 and Figure 3.9, showing empirical data for 212 routes on various parts of the Dutch main road network[57], for the morning and evening peaks (7-9am and 4-6pm, respectively).

Figure 3.8 shows plots of two speed percentiles against the 'average speed' (computed as travel distance over average travel time), based on data for 154 weekdays in 2002[58]. The 10th percentile speed and the 'average speed' (corresponding to the 90th percentile travel time and the average travel time, respectively) are clearly related to each other. Lower 'average speeds' coincide with lower 10th percentile speeds. From the fact that the data points are gathered within a fairly narrow

[57] These are routes of variable lengths. The distribution of these lengths corresponds well to the distribution of travel distances observed in reality on the main road network.

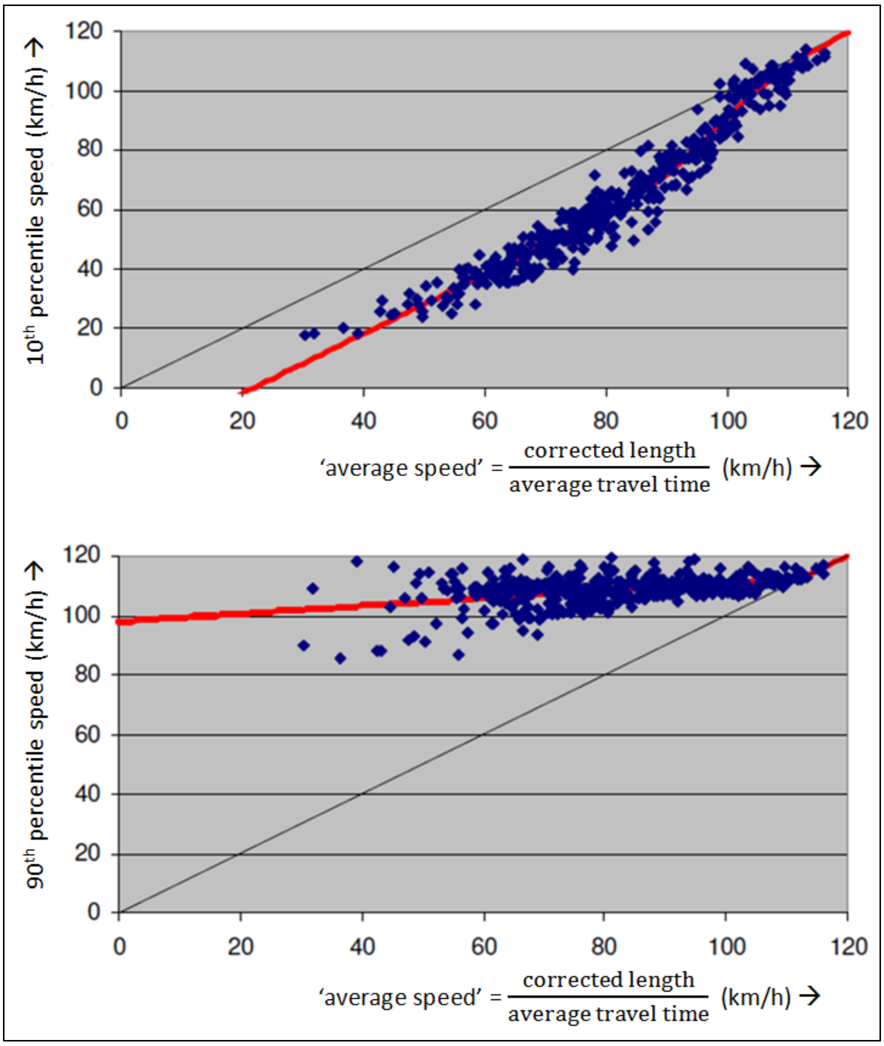
[58] Both the speed percentiles and the average travel time refer to the *day-to-day* distribution of the travel times.

band[59], this relation can be concluded to be fairly strong, just like the relation between the average and 85th percentile delay found by Van Toorenburg.

In contrast to the 10th percentile speed, the 90th percentile speed (corresponding to the 10th percentile travel time) apparently is not that strongly related to the 'average speed'. In all cases the 90th percentile speed had a fairly large value (within the 'free flow domain'), irrespective of whether it concerned a situation with a high 'average speed' or a situation with a low 'average speed'.

Combining the information provided by the two diagrams, it can be concluded that the width of the speed distribution strongly increases with decreasing 'average speed'. Put differently, a larger average travel time and a larger travel time variability are clearly linked.

Figure 3.8: Relationships between the 'average speed' and the 10th (above) and 90th (below) percentile speed respectively[60]
(Adapted from Rand Europe, 2004)

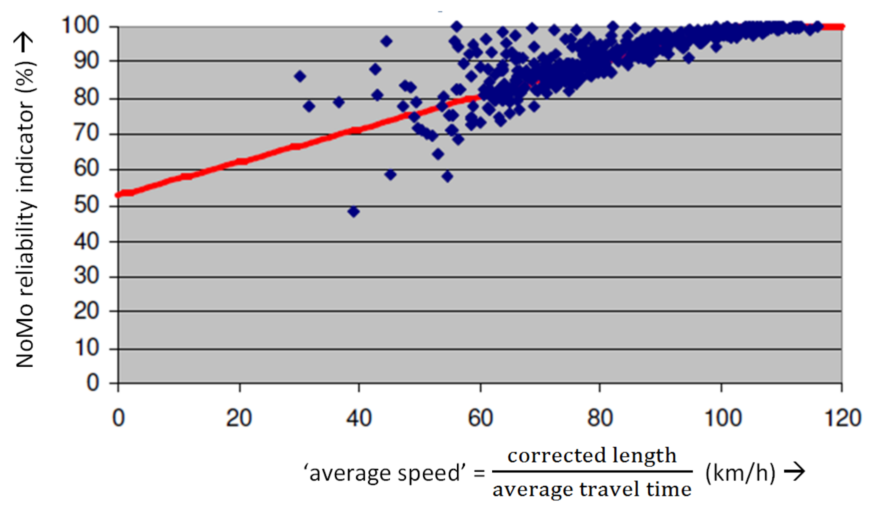


[59] It should be noted however that for low speeds, a narrow bandwidth in terms of speed corresponds to a much larger bandwidth in terms of travel times.

[60] all speeds corrected for differences in speed limit

Figure 3.9 shows another diagram from the same research project. In this diagram, the 'average speed' is plotted against the travel time reliability indicator used in the NoMo (i.e. the percentage of trips arriving within certain bounds of the median travel time). From the diagram it can be concluded that there is some interdependency between these two variables. On average, a lower 'average speed' (i.e. a larger average travel time) corresponds to a lower percentage of trips being realized 'in time' (indicating that extreme travel times occur more frequently). However, this relationship is not extremely strong: a substantial part of the variation in the NoMo indicator remains unexplained (in spite of the fact that a correction had been applied for the variation in route length, which is another important explaining variable).

Figure 3.9: Relation between the 'average speed'[61] and the NoMo reliability indicator (the percentage of trips 'in time', i.e. with a travel time within +/- 10 minutes from the median) [62]
(Adapted from Rand Europe, 2004)



In a research project conducted in the United States (see Margiotta, 2009), similar results were found as well. In this research project, a large empirical data set for 51 (relatively short[63]) urban freeway sections in the United States was analyzed. For all these sections, various indicators for the day-to-day travel time reliability in the morning or evening peak were calculated, as well as the average travel time. For part of the sections, the indicators were calculated both for a period *before* implementing a certain measure (like ramp meters or an aggressive incident clearance program) and for a period *after* this implementation.

Based on a statistical analysis on the calculated indicator values, Margiotta concluded that all considered indictors could be predicted as a function of the average travel time index (the average travel time divided by the free flow travel time). Figure 3.10 shows one of the relationships that were found. In this figure, the 95th percentile travel time index (i.e. the 95th percentile travel time divided by the free flow travel time) is plotted against the average travel time index, for all

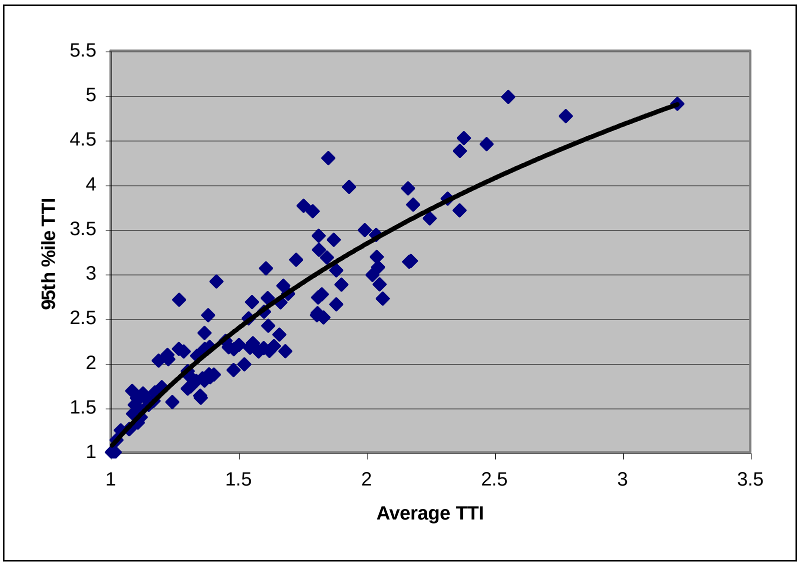
[61] corrected for differences in speed limit

[62] corrected to a standard route length of 20 km

[63] on average about 7 km long

freeway sections and all years included in the data set. The relation closely resembles the relation between the 85th percentile delay and the average delay found by Van Toorenburg (Figure 3.7).

Figure 3.10: Relation between 95th percentile travel time index and average travel time index in the morning/evening peak, for a large set of freeway sections in the United States
(Source: Margiotta, 2009)



3.5.2 Implications for this research project

The empirical findings discussed above might give the impression that it is in fact relatively easy to take into account the inherent variability in traffic conditions when evaluating traffic congestion (for example in the context of comparing the scenarios *with* and *without* a certain measure, in order to assess the benefits of this measure). Compared with the traditional approach, in which this inherent variability is not considered in a systematic way, the only extra step that seems necessary is to calculate some additional travel time statistics (like percentile values, variance and skew) from the 'average' travel time calculated with a traditional model, using the empirical relations discussed above (i.e. the empirical relations between the average travel time and all the other travel time statistics).

In reality however it is not so simple. This is due to the following reasons:

1) For most of the empirical relations between the average and the other statistics, the good quality of the fit is mainly achieved in the domain of low congestion (corresponding to low average travel times / delays, or, equivalently, high average speeds), where most of the data points are found. For the part of the data points which represent situations with heavier congestion (i.e. higher average travel times or lower average speeds), the quality of the fit is much lower. For these situations the unexplained variation is large. Due to this unexplained variation, it is uncertain to which extent the effects of measures will follow the fitted relationships. This means that the empirical relations are not accurate enough to predict all other effects of

a measure from its effect on the average only. If the effect on the average is small compared to the unexplained variation, it cannot even be inferred with certainty whether the other statistics will be affected *positively* or *negatively*.

2) Even if the empirical relations would have fitted perfectly to the data from which they have been derived (i.e. without any unexplained variation), it could not be excluded that the effects of certain measures would 'break' these relations.

3) The 'average' travel time predicted by a traditional calculation model is typically *not a real average travel time* (including the effects of all sources of variability), but rather a kind of 'representative' travel time (computed by performing a 'point calculation' for representative traffic demand and supply conditions, not taking into account variability in traffic conditions). Of course these two travel times are not equal to one another. The real average most likely will be higher than the 'representative' travel time, since in the computation of the latter disturbances (like accidents) are not taken into account. This means that all the empirical relations discussed above in fact cannot be used, as these have been derived for *real averages*. These empirical relations can thus only be used if real averages are calculated first (rather than 'representative' travel times). If these real averages are computed there is no point anymore in using the empirical relations however, since it takes only a small extra effort to also directly calculate the other travel time statistics if those real averages are computed.

   One might wonder whether it would not be possible to calculate the real average travel time directly from the 'representative' travel time (after which the other travel time statistics could be computed using the empirical relations discussed above), using an empirical relation between these representative and real average travel time. The existence of such a relation is to be doubted however. Based on his study Margiotta claims that the average travel time index is 15 to 20% larger than the 'recurring only' travel time index (a kind of 'representative' travel time index). Information on the quality of this relation (i.e. the extent to which practical situations deviate from this relationship) is unfortunately not provided, however.

   In the research project of Rand Europe (2004), the distribution of the 'representative' travel times (calculated with a model) over the various road sections considered in that project[64] was compared with the distribution of the average travel times over the same road sections. These two distributions turned out to be rather different, indicating that 'representative' travel times and average travel times indeed may differ significantly. Rand

[64] Please note that in contrast to (practically) all other distributions considered in this project, this is <u>not</u> a distribution over various moments in *time*, but over various sections in *space*.

Europe has not attempted, however, to find a proper relation between the two.

Even if there *would* be some rough relation between 'representative' and average travel times, it would not be preserved in situations in which certain types of measures are taken. Consider for example incident management measures (measures aimed at clearing the road as soon as possible after the occurrence of an incident). These measures will improve the average travel time (and other indicators), while having not any effect on the 'representative' travel time. One could also think of measures that improve the 'representative' situation but increase the traffic systems' sensitivity to disturbances, resulting in the average travel time to be increased.

4) From research results of Van Lint et al (2008) it can be concluded that the relationships between different travel time statistics may be different for different periods of the day (i.e. the off-peak part of the day, the different shoulders of the peak periods, and the central parts of these peak periods). Although not considered in their investigation, this might also be the case for the relationships between the *average* and the other statistics. This is not reflected in the empirical relations discussed above.

5) In earlier sections it turned out that in order to satisfactorily describe the extent of traffic congestion, besides indicators relating to travel time distributions some other types of indicators should be considered as well. Also for these indicators we are not interested in a 'representative' value (calculated in the traditional way), but rather in a value in which all sources of variability are reflected, which might be significantly different. One of the indicators selected in section 3.4) cannot even be obtained from a 'representative' calculation: in order to evaluate the travel time instability we want to consider the *probability distribution* of the difference between the instantaneous and actual route travel time (indicator VI), while a 'representative' calculation only provides us with a single ('representative') value for this. Obviously, the empirical relations discussed above cannot offer solutions here.

From the above it can be concluded that the empirical relations found in literature certainly provide some relevant information (especially that, on average, a larger average delay and a larger travel time variability go hand in hand), but that they do not remove the need to consider the various sources of variability in a more explicit way when calculating the extent of traffic congestion.

4. Quantification methodology

4.1 Introduction

Now that we discussed the basic (probabilistic) mechanisms behind the traffic congestion phenomenon, and explained which indicators for traffic congestion should be considered, we can explain the reasoning behind the selection of the method that was used to address the main research objective.

First of all we bring the main research objective back to mind:
'To reveal what kind of additional or revised insights can be obtained from evaluations of the traffic system's performance (in the context of considering taking strategic measures to alleviate congestion) when the inherent variable nature of daily traffic congestion on the motorway network is explicitly taken into account. (As opposed to the insights obtained by evaluations according to the more 'traditional' approach, in which only a kind of 'representative' situation is evaluated.) '

Two different *types* of such additional (or revised) insights are distinguished. Firstly, we might obtain new insights into the *relative importance of the various primary sources of traffic congestion* (like nominal demand levels exceeding nominal capacities, intrinsic variations in human travel or driving behavior, adverse weather conditions, incidents, etc.). These insights (referred to as 'insights of type A') typically cannot be obtained with the 'traditional' evaluation approach, since most of these primary sources are not even considered then. Yet these insights can be quite valuable, by indicating at which of these primary sources measures should be directed in order to alleviate traffic congestion most effectively.

The second type of insights (referred to as 'insights of type B') relates to the *effectiveness of specific measures*. If we assess the effectiveness of measures according to the 'new' approach (i.e. based on the indicators defined in chapter 3, and taking into account all sources of variability discussed in chapter 2), we might very well obtain results that extend, refine, or revise the insights obtained from a 'traditional' assessment (focusing on a kind of 'representative' situation).

These two types of additional insights are considered in two separate parts. In the first part, it is demonstrated that new insights might be obtained into the relative importance of the various primary sources of traffic congestion (Figure 4.1). Here a given traffic system is considered 'as it is', meaning that we do not consider the effects of measures yet. In the second part, an example of a modification to this traffic system is considered (i.e. a measure aimed at alleviating traffic congestion). For this measure it is analyzed whether – and if so, in what way – an assessment of its effectiveness according to the 'new' approach yields different conclusions than an assessment according to the 'traditional' approach. This is shown schematically in Figure 4.2.

For the analyses that need to be performed for these two parts, different types of methods could be used. From chapter 2 and 3 it might be clear that methods based on qualitative reasoning are not adequate, because of the complexities involved (such as the non-linearity in the interaction between demand and supply, the network effects of traffic congestion and the interdependencies between different sources of variability). In order to properly grasp these complexities, quantitative analyses will be needed. The selection of a quantification methodology for the first part of the analyses (relating to the insights of type A) is discussed in section 4.2. Next, section 4.3 considers the selection of a methodology for the second part (relating to the insights of type B). After that, these methods are further elaborated.

Figure 4.1: Schematic representation of the way in which the additional insights of type 'A' (i.e. into the relative importance of the various primary sources of traffic congestion) are demonstrated

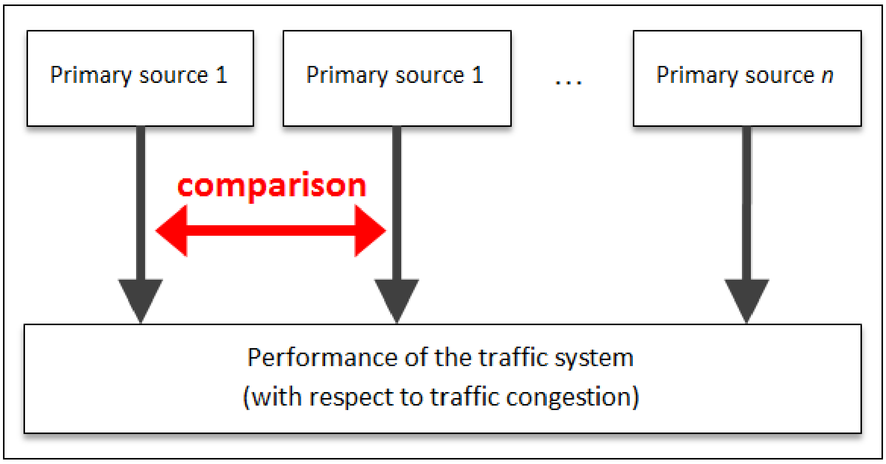
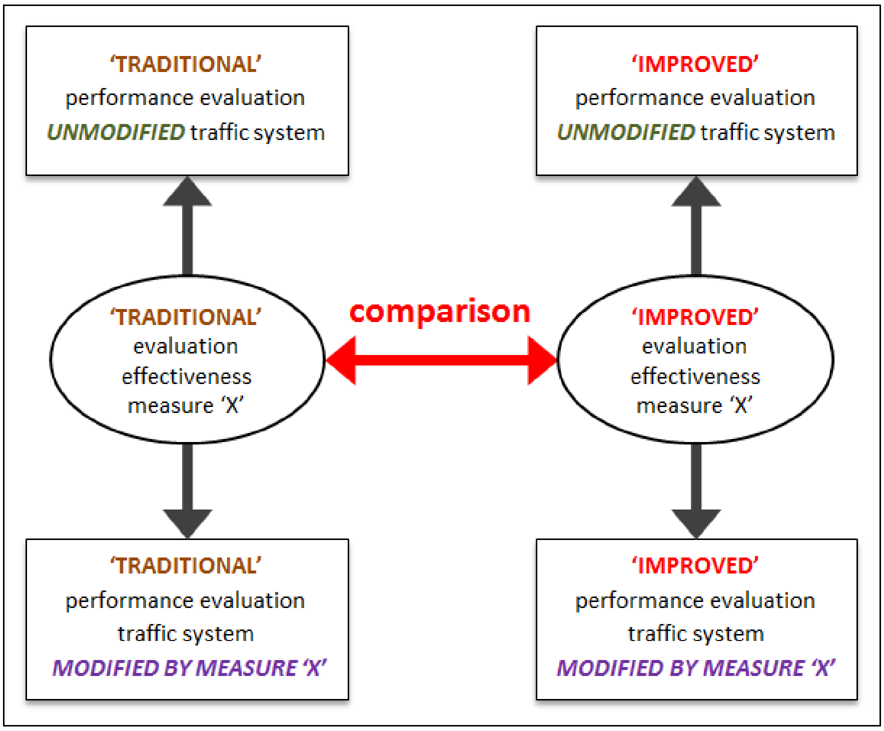


Figure 4.2: Schematic representation of the way in which the additional insights of type 'B' (i.e. into the effectiveness of a specific measure) are demonstrated



4.2 Methodology for the illustration of insights of type A

This section discusses the selection of a quantification methodology for the illustration of the acquisition of new insights into the relative importance of the various primary sources of traffic congestion (i.e. their relative contribution to the performance indicators selected in chapter 3). Two main categories of methods can be identified for this: *analysis of empirical data* and *model-based analysis*.

Examples of possible strategies that are based on analysis of empirical data could be:

- Grouping the empirical data for similar circumstances (with respect to the various primary sources of congestion) in clusters, after which the average performances per cluster (expressed in terms of the indicators selected in chapter 3) can be mutually compared in order to find the relative impacts of the various primary congestion sources.
- Fitting statistical relations to the data, relating the traffic system's performance (in terms of the indicators defined in chapter 3) to the different primary sources of traffic congestion.

Examples of possible strategies that are based on model-based analysis could be:

- Deriving an analytical model, describing the relationships between, on the one hand, variables that describe the various primary sources of traffic congestion, and, on the other hand, the performance indicators defined in chapter 3.
- Repeatedly performing a Monte Carlo series of traffic simulation model runs, each time omitting one of the primary sources of traffic congestion, in order to assess its influence on the indicators defined in chapter 3.

The main advantage of using empirical data is that these data represent exactly what reality is like (aside from measurement errors and inaccuracies, of course), while models do not. Models typically are a simplification of reality, meaning that they are usually incomplete and even erroneous to a certain extent. In spite of this strong argument for using empirical data analysis, yet the model-based methodology was selected, because of the following important considerations:

- The data that would be required for the empirical analyses will probably be partially unavailable. This relates to data describing the status of some of the primary congestion sources. If a model-based method is used, such data unavailability is less problematic.
- Empirical data may be distorted by trends in the demand levels and changes in the supply characteristics (such as the realization of additional infrastructure, or the implementation of traffic management measures).
- It is difficult to properly isolate the individual influences of the various primary congestion sources from empirical datasets, particularly because some of these sources are interdependent (see chapter 2). When a model-based method is used, the

different primary sources can easily be individually 'switched off'.

Note that by opting for a model-based methodology, the following drawbacks of using a model are accepted (besides the fact that models are a simplification of reality, as was already mentioned above):

- Developing a model in which the various sources of variability are appropriately accounted for (if no suitable existing model is found) will be a time-consuming task.
- Models require inputs and parameter values, which might be difficult to obtain.

It should be noted, however, that not too much importance should be attached to the latter drawback. This is because of the fact that it is not intended here to come up with firm quantitative inferences with respect to a specific existing situation. Instead, it is only aimed for to *illustrate* the gain of any possible new insights. The only requirement is then that the situation considered *could have been* a real-life situation. Consequently, all inputs and model parameters can be given any value within the range in which they occur in reality.

After the choice for a model-based method, it still had to be decided what *type* of model was to be used. At the most rudimentary level, we can make a distinction between analytical and numerical techniques. From chapter 2 it is evident however that the problem is too complex to be addressed in an analytical way, implying that a numerical method had to be used.

If opting for a numerical methodology, using a *traffic simulation model* is the most obvious choice (since this kind of models has a good capability in handling the complexity involved). The different sources of variability (discussed in chapter 2) could be accounted for in such a model by randomizing the various input variables and model parameters. For this purpose the Monte Carlo method can be used.

Two main categories of traffic simulation models can be distinguished: microscopic models and macroscopic models. In microscopic models all vehicles are modeled separately, while in macroscopic models traffic flows are modeled at an aggregate level. Accordingly, the output of microscopic models is much more detailed. The price paid for this, however, is a much longer calculation time per simulation run. In view of the facts that a large number of simulation runs have to be performed in order to obtain results in which the different variabilities are adequately reflected, and that a whole network is to be simulated, this long calculation time in fact makes microscopic simulation models unsuitable for the task. In addition to this, we are actually not really interested in the detailed results of a microscopic simulation. Aggregated results at the level of traffic flows are sufficient to calculate the performance indicators selected in chapter 3.

Another important reason for using a macroscopic model is that in this type of models all macroscopic supply characteristics of road sections (such as capacities) are *inputs* to the traffic simulations. This makes it possible to vary these characteristics, which is essential for the research tasks at hand. In microscopic models, by contrast, these macroscopic supply characteristics are not inputs, but *outputs* of the simulations. If such a model would be used, the supply effects of the different sources of variability would have to be accounted for by varying the parameters of the microscopic traffic behavior. This would be much more difficult, however. In addition, one should note that in microscopic simulation models, demands and capacities are randomly varying by nature[65]. For the research task at hand this is undesirable. After all, we want to be able to 'deactivate' the random fluctuations in the demands and capacities, in order to reveal their relative impacts on the congestion indicators. In macroscopic traffic simulation models this problem does not exist, since these are deterministic in nature.

Macroscopic traffic simulation models come in all shapes and sizes. In section 4.4 it is discussed what requirements a model must meet for the task at hand. Subsequently, in section 4.5 existing models specifically designed for addressing the variability in traffic conditions are considered. Here it is discussed to what extent these models meet the requirements set in section 4.4.

4.3 Methodology for the illustration of insights of type B

This section discusses the selection of a methodology for the illustration of the acquisition of new insights into the effectiveness of specific measures (aimed at alleviating traffic congestion). For this illustration, one particular measure will be considered as an example. For this measure it is to be considered then whether (and if so, in what way) an assessment of its effectiveness according to the 'new' approach (i.e. based on the indicators defined in chapter 3, and taking into account all sources of variability discussed in chapter 2) would yield different conclusions than an assessment according to the 'traditional' approach (focusing on a kind of 'representative' situation).

In order to achieve this, the following procedure is followed:

- Step 1: Assessment of the effectiveness of the measure according to the *traditional* approach.
- Step 2: Assessment of the effectiveness of the measure according to the *new* approach.
- Step 3: Comparison of the results of the two assessments.

For steps 1 and 2 (i.e. the assessments of the effectiveness of the measure considered) we can choose between the same two types of methods as for the illustration of the new insights of type A (section 4.2), namely *analysis of empirical data* or *model-based analysis*.

[65] Note that this stochasticity embodies only a very limited part of all variabilities identified in chapter 2.

In the first case, the effectiveness of the considered measure is assessed by comparing empirical data measured *before* the implementation of this measure with empirical data measured *after* the implementation. Also in the second case the effectiveness of the measure is assessed by comparing the traffic system's performance for the situations *with* and *without* this measure. However, in this case these performances are not derived from empirical data. Instead, they are calculated using a model.

For step 1 (i.e. the assessment according to the *traditional* approach, meaning that the effectiveness is assessed in terms of the improvement or deterioration of the *'representative'* traffic conditions), the choice between these two types of methods is actually rather obvious. This is because of the fact that the 'representative' situation, as it is understood here, is in fact an artificial concept. It is the situation in which all influence factors are at their representative level (i.e. their median, average, or whatever else is considered representative). In other words, it is a situation without any influences of variability, except for the regular demand variation with the time of the day. This situation typically only occurs in a model environment, and not in reality (or only with an almost infinite small probability). In reality, demand and supply are, after all, always fluctuating, due to:

- the variations in the driver and vehicle populations
- the variations in the luminance conditions
- the intrinsic randomness in people's personal travel choices
- the intrinsic randomness in human driving behavior

Since the representative situation does not occur in reality, it obviously cannot be found in empirical data either. This makes the empirical evaluation method unsuitable for step 1. This means that only model-based methods are appropriate for this step.

Of course, this argument does not apply to step 2 (i.e. the assessment according to the *new* approach). Therefore, for this step we still have the choice between empirical data analysis and model-based analysis. Below, an overview is given of the pros and cons of these two types of methods.

Pros and cons of using empirical data analysis:

+ Empirical data represent the 'factual truth' (apart from measurement errors and inaccuracies).
- The assessment results can be distorted by trends, random fluctuations or other inhomogeneities in the various influencing factors (causing differences between the data measured *before* and the data measured *after* the implementation of the measure considered, which are not due to this measure itself).
- The assessment will be time-consuming (due to the extensive data processing that will be required).
- The difference between the results of the two effectiveness assessments might be distorted by the use of two different methods (i.e. model-based analysis for step 1, and empirical data-analysis for step 2).

Pros and cons of using model-based analysis:

+ The assessment results will not be distorted by trends, random fluctuations [66] or other inhomogeneities in the various influencing factors.
+ The difference between the results of the two effectiveness assessments cannot be distorted by the use of different methods.
- Models are a (partially incomplete/erroneous) simplification of reality, with the result that the outcomes of these models to a greater or lesser extent will deviate from reality as well.
- Developing a model in which the various sources of variability are appropriately accounted for (if no suitable existing model is found) will be a time-consuming task.
- Models require inputs and parameter values, which might be difficult to obtain.

It should be noted, however, that not too much importance should be attached to the last-mentioned drawback, for the same reason as given in section 4.2.

All in all, it was considered the best option to use the same type of methodology for step 2 as for step 1, i.e. based on model computations. As far as the *type* of model is concerned, a *macroscopic traffic simulation model* was selected, for exactly the same reasons as given in section 4.2. Since there is a strong overlap between the requirements for the models needed to illustrate the new insights of types A and B, it was decided to select/develop *one* model satisfying both sets of requirements. In section 4.4 the combined set of requirements is given. Subsequently, in section 4.5 existing models specifically designed for addressing the variability in traffic conditions are considered. Here it is discussed to what extent these models meet the requirements specified in section 4.4. Since none of the models was found to be sufficiently adequate for the tasks at hand, a new model was developed. This model is discussed in chapter 5.

4.4 Requirements for the quantification model

In order to be able to satisfactorily perform the required analyses, several requirements have to be met by the quantification model. These are listed below. The requirements do not only relate to the way in which the traffic is modeled. Requirements relating to the model input and output are included as well.

The model is required to:

- be able to reproduce the dynamic processes of queue formation and dissipation, based on the evolution of traffic demand and capacity over time (see section 2.1);

[66] By using the same random seeds for the 'before' and 'after' scenarios, distortions by randomness can be prevented.

- include the capacity drop phenomenon (section 2.1);
- include the (equilibrium) relation between the density of traffic and the traffic speed (or, equivalently, between density and volume or speed and volume), as represented by the fundamental diagram (section 2.1);
- include the blocking back effect of traffic congestion (section 2.3.1);
- include the 'filtering' and 'releasing' effects of traffic congestion (also referred to as the 'temporal redistribution effect', see section 2.3.2);
- include the effect of (non-recurrent) traffic congestion on route choice (section 2.3.3), including the possibility that traffic diverts from the motorway network to the secondary network;
- capture all sources of variations in the traffic demand (both within-day and day-to-day) discussed in section 2.2.4 (except for the ones that were identified as occurring too infrequently to be considered a source of *daily* variability in traffic congestion);
- capture all sources of variations in the traffic supply (both within-day and day-to-day) discussed in section 2.2.5 (again except for the ones that cannot really be considered to contribute to the *daily* variability);
- include the interdependencies between the various sources of variations (section 2.2.6);
- include the possibility to calculate the 'representative' situation (section 4.3), by 'switching off' all sources of variability, except for the systematic within-day variability in traffic demands[67];
- include the possibility to *separately* 'switch off' each individual source of variability (section 4.2);
- offer sufficient possibilities to enter the (possibly dynamic) effects of traffic measures, by modifying the model input or the characteristics/parameters of the model itself (section 4.3);
- be able to perform a large number of simulation runs within a limited calculation time;
- provide the performance indicators selected in section 3.4 as output, or alternatively provide all calculated traffic conditions as output (i.e. the speeds and traffic volumes for all road sections in the network, for all simulated time intervals), from which the indicators can be computed in a post-processing step;
- use a time step that is small enough to:
    - prevent (within-day) variations in the traffic conditions from being smoothed out too much, and to
    - model the propagation of the traffic over the network in a sufficiently accurate (and stable) way.[68]

[67] In practice it is quite common to perform a *dynamic* traffic simulation (taking into account the systematic within-day variation in traffic demands) for the ex-ante evaluation of a given measure. Therefore, in this thesis the systematic within-day variation in traffic demands is considered part of the 'representative situation'.

[68] In practice, the latter of these two requirements will be decisive.

4.5 Existing quantification models

Various existing quantification models (specifically designed for addressing the variability in traffic conditions) have been assessed on their suitability for use in this project. That is, it has been considered to which extent the requirements specified in the previous section are met by these models.

Here the focus has been on models used by the Dutch road authority:
- **SMARA** (Simulation Model for Analyzing the Reliability of Accessibility)
- **LMS-BT** (National Model System - Reliability Tool)
- **Waiting time model for main roads** (Traffic Quality / ESIM)
- **Traffic Quality – Network version**

However, also some models reported in (recent) international literature have been considered:
- **Queuing Model for determination variability recurrent congestion**
- **KAPASIM**
- **Travel time variability model of Mehran and Nakamura**

It should be noted here that models focusing on only one type of variability (like incidents) are not considered.

In most models that are used in practice or proposed in literature, only one individual road section or bottleneck is considered. From the foregoing sections and chapters it is clear, however, that for the tasks at hand a *whole network* is to be considered, requiring a network model. In spite of this fact, the aforementioned models (relating to an individual segment) are included in this section as well. Not only for completeness, but also in view of the fact that it in principle can be considered to extend such a model to a network version, if the model – apart from being limited to an individual segment – is deemed very appropriate for the tasks at hand.

Detailed descriptions of the models and their appropriateness in terms of fulfilling the requirements of section 4.4 are given in Appendix 1. Here only an overview of the positive and negative aspects of the different models is presented. This overview is shown by Table 4.1.

From this table it can be concluded that all of the models turn out to have quite a number of major drawbacks. None of the models satisfactorily accounts for the route choice effect of traffic congestion. The other two network effects (i.e. blocking back and temporal redistribution) are adequately modeled by only one of the models. There is no single model that covers all relevant sources of variability. Furthermore, often important interdependencies are omitted. Another problem is that most models cannot provide the desired indicators. Only one of the models provides rough output data from which these indicators could be calculated in a post-processing step.

Overall, the model 'Traffic Quality – Network version' (using the dynamic macroscopic traffic simulation model 'Flowsimulator' as computational core) comes out best. Drawbacks of this model however are that route choice effects cannot be properly accounted for, and that fluctuations in the traffic demands cannot be addressed in an ideal manner. This latter is due to the fact that the traffic demands are specified on the *link* level (rather than on the *origin-destination* or *route* level), and propagated using aggregate split fractions.

Consequently, none of the models was considered adequate for the tasks at hand. Therefore, in this project a new model was developed, specifically designed for these tasks. This newly developed quantification model is discussed in the next chapter.

Table 4.1: Overview of the models' scores regarding the requirements from section 4.5

Model \ Requirement	Dynamic processes of queue formation and dissipation	Capacity drop	Speed-flow/density relationship	Blocking back	Temporal redistribution (filtering + releasing)	Route choice effect of traffic congestion	Diversion to secondary road network	Sources of variations demand	Sources of variations supply	Interdependencies sources of variations	Calculation representative situation	Individually 'switch off' sources	Express effects of measures	Limited calculation time	Performance indicators / rough output	Time step small enough
SMARA	-	-	o	-	-	o	o	o	o	-	+	+	o	+	-	-
LMS-BT	o	o	o	o	o	o	o	-	-	-	+	-	-	+	-	-
WTM – ESIM	+	-	+	-	-	-	-	-	-	-	+	o	-	+	-	-
WTM – Traffic Quality	o	-	-	-	-	-	-	o	-	-	-	-	-	+	-	-
Traffic Quality – Network version	+	+	+	+	+	-	-	-	o	o	+	+	o/+	+	+	+
KAPASIM	+	+	o	-	-	-	-	o	+	-	+	o	o	+	-	+
Queuing Model TU Delft	+	-	-	-	-	-	-	o	-	-	+	+	o	+	-	o
Model of Mehran & Nakamura	+	+	+	-	-	-	-	+	o	+	+	+	o	+	-	+

5. The developed quantification model

5.1 Introduction – general concept

In this chapter the developed quantification model is described. The model is based on the same general concept as most of the models discussed in the last section of the previous chapter. That is, a large number of traffic simulations are performed for varying model inputs, reflecting the various types of variability in traffic demand and supply characteristics (generated using the Monte Carlo technique). Subsequently, the desired performance indicators can be computed from the combined set of all simulation results. While this general concept might be the same as that of other models, its further elaboration is different, in order to overcome drawbacks of the models discussed in the section 4.5.

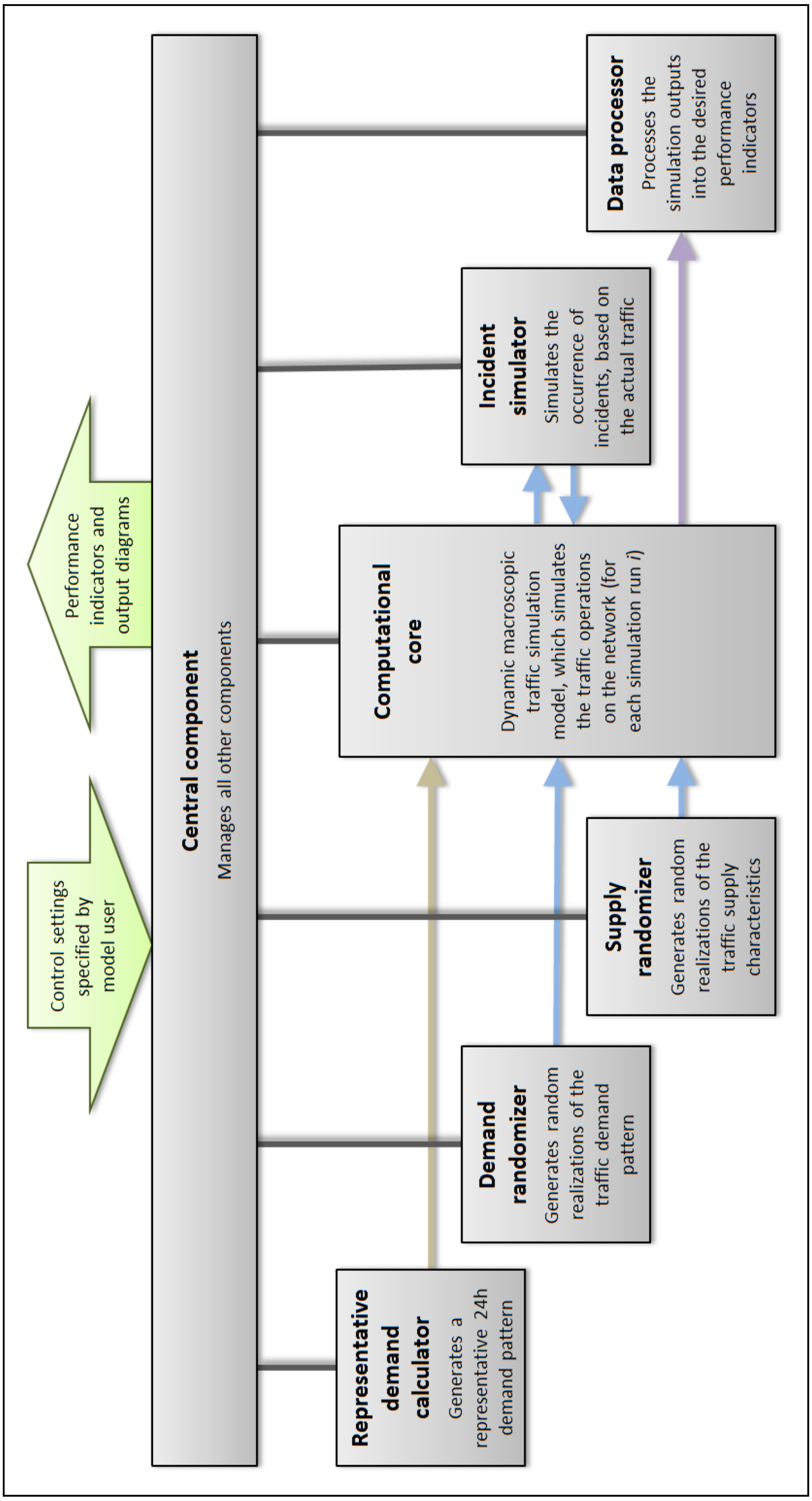
Figure 5.1 schematically shows the main structure of the model. The model consists of seven components. The computational core of the model is a dynamic macroscopic traffic simulation model. For each simulation that is performed, this component simulates the traffic operations on the network, given the various traffic demand and supply characteristics. The other six components are built around it. These other components are:

- A '*central component*':
  This component manages all other components, and provides the options for model users to interact with the model. All kinds of control settings can be manipulated by the model user, which subsequently are translated in corresponding model actions.

- A '*representative demand calculator*':
  This component calculates a representative 24-hours demand pattern for all origin-destination relations in the network, using a static origin-destination matrix for the morning peak period as its only external input.

- A '*demand randomizer*':
  This module generates random realizations of the traffic demand pattern, based on a stochastic simulation of the various influencing factors involved. The representative 24-hours demand pattern calculated by the representative demand calculator is used as the base situation in the randomization procedure.

Figure 5.1: Main structure of the developed model



- A '*supply randomizer*':
  Comparable to the demand randomizer generating random realizations of the traffic demand pattern, the supply randomizer generates random realizations of the traffic supply characteristics, again based on the simulation of the various influencing factors involved. Here the deterministic supply values specified in the network model are assumed to represent the 'representative' base situation.

- An '*incident simulator*':
  Since the occurrence of incidents is strongly dependent on the actual traffic conditions, these incidents cannot be randomly generated in advance (i.e. before actually simulating the traffic operations on the network). Since the supply randomizer is programmed to compute the supply characteristics in advance for the whole 24-hours period (like the demand randomizer is programmed to generate the 24-hours traffic demand pattern in advance), this means that the generation of traffic incidents (which mainly affect the traffic operations by their effects on the supply characteristics) cannot be included in the supply randomizer, but are to be modeled separately. This is the task of the incident simulator.

- A '*data processor*':
  This component collects the outputs of all traffic simulation model runs, and subsequently processes these into the desired performance indicators.

While the computational core of the model (i.e. the dynamic traffic simulation model) is implemented in the JAVA programming language, the other parts of the model are all implemented in the mathematical programming language MATLAB. In the course of the model development some routines of these other parts were moved to the JAVA code of the dynamic traffic simulation model, however, in order to (significantly) improve the computation speed.

The horizontal arrows in Figure 5.1 show the most important data flows between the different model components. Please note that the real model involves more data flows, which have been omitted from the figure for the sake of readability. One of these other data flows for example concerns the representative traffic demand pattern over the day, which is transferred from the representative demand calculator to the demand randomizer, which uses it as starting point for the randomization.

In the following sections, the most important aspects of the model will be discussed in more detail. These successively are:
- the traffic flow modeling (section 5.2),
- the general approach to the modeling of the variations in traffic demand and supply (section 5.3), and
- the modeling of the various individual variation factors (section 5.4).

In terms of the different model components described above, the first aspect is dealt with in the computational core of the model, while the other two aspects are managed in the demand and supply randomizer, the incident simulator and the representative demand calculator.

The next chapter (Ch. 6) discusses some modeling issues requiring further consideration, which have come to light during the development of the model. Since these generally are issues which require substantial further research, it was not possible to actually solve them within this master thesis research project. However, chapter 6 does discuss some possible strategies for this. Also a possible solution strategy for reducing the required number of simulations is proposed here. At this moment, at least a few thousand simulations are necessary to obtain results with a sufficient level of statistical accuracy. Given the simulation time of about 1.5 to 2 minutes per simulation, one simulation series would take multiple days, if not multiple weeks.

The validity of the developed model is discussed in chapter 7.

5.2 Traffic flow modeling

As was already mentioned in the previous section, the simulation of the traffic propagation over the network (i.e. the computation of the resulting traffic conditions for given traffic demand and supply characteristics, generated by other model components) is taken care of by the computational core of the model. For this use is made of an existing dynamic macroscopic traffic simulation model. Subsection 5.2.1 discusses the choice of this model. Subsequently subsection 5.2.2 describes which adaptations were made to the model. Finally in subsection 5.2.3 it is discussed how this model was used to model the traffic flow over the network, considering aspects like the (spatial and temporal) discretization, the fundamental diagram and the (absence of) route choice modeling.

5.2.1 Model selection

A large number of dynamic macroscopic traffic simulation models have been considered for the modeling of the traffic operations. These models are listed below:

- Flowsimulator,
- Indy (link transmission version),
- METANET,
- MaDAM,
- MARPLE,
- Fastlane,
- The EVAQ Hybrid Route Choice Model,
- DSMART, and
- JDSMART.

These models have been mutually compared on the following aspects:

- Ability to reproduce the processes of growth and dissipation of traffic jams (including blocking back effects) in a realistic way, at least consistent with first order traffic flow theory.
- The modeling of the capacity drop phenomenon.
- The way of modeling the split fractions at the network nodes. (For a consistent modeling of the traffic flows over the network *fixed* split fractions are not appropriate, because in reality they may be affected by the traffic conditions in upstream parts of the network.)
- The ability to simulate the effect that traffic might deviate from its standard routes in case of (seriously) non-recurrent traffic conditions.
- The possibility to perform an (once-only) equilibrium assignment (in order to obtain a basic demand pattern, which subsequently would be varied according to the various sources of variation).
- The possibility to introduce the variations in the traffic demands in the model in a consistent way.
- The possibility to introduce the variations in the traffic supply characteristics in the model in a consistent way.
- The possibilities for modeling time dependent supply effects of measures.
- The possibilities for modeling time dependent demand effects of measures.
- The calculation time.
- The model complexity (taking into account the fact that while a higher complexity might allow for a potentially more realistic traffic modeling, it usually also results in a larger amount of potential sources of error, and might adversely affect the interpretability of the results).
- The flexibility of the model (i.e. the possibility to adapt the model to the tasks at hand, which for example might be needed if one or more of the other criteria are not met).
- The generation of the desired output indicators (or rough output data, from which the desired output indicators can be computed).
- The interfaces with the other components of the developed model: in view of the large required number of simulations it should be (made) possible to make MATLAB invoke the traffic simulation model in an automated way.
- The ease of use of the simulation model
- The availability of the model.
- The availability of a readily usable test network, including a nominal demand pattern.

The assessment revealed that there is no single model that meets all the requirements. Therefore, inevitably some concessions had to be made.

There is only one model that can deal with the route choice effect of traffic congestion (i.e. the effect that traffic might deviate from its standard routes in case of (seriously) non-recurrent traffic conditions) in a sound way, namely the EVAQ Hybrid Route Choice Model (Pel et al, 2009). This is the only model that combines the process of *pre-trip* route choice (in which travelers base their route choices on their past experiences, which results in the phenomenon that the traffic conditions have the tendency to approximate an equilibrium[69]) with the process of *en-route* route choice (in which travelers base their route choices on the current traffic conditions). All other models account for only one of these two route choice processes (mostly pre-trip route choice), or do not model route choice at all. In the Hybrid Route Choice Model travelers in principle follow their 'equilibrium routes' (based on their past experiences), but will deviate from these if the traffic conditions differ significantly from the traffic conditions in the 'equilibrium' situation.

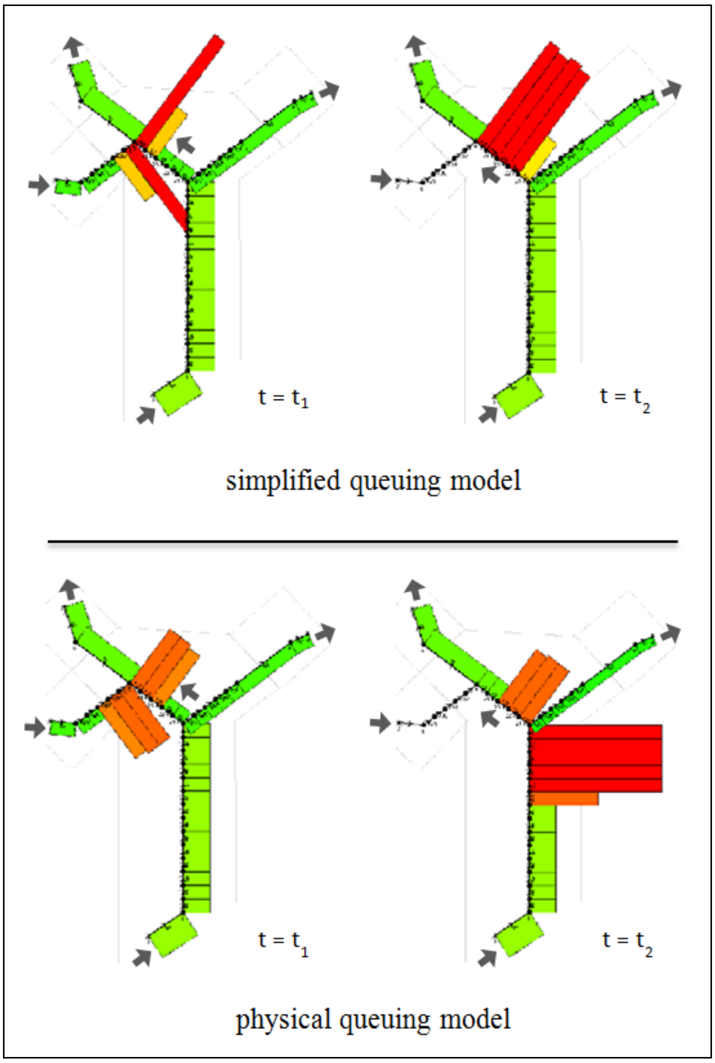
This clearly is a strong argument for the use of the EVAQ Hybrid Route Choice Model as computational core of the developed model. However, two clear drawbacks of this simulation model were identified, which make it less appropriate for the tasks at hand:

- Instead of using a physical queuing model (such as in cell or link transmission models), the model uses a simplified queuing model in which each link is artificially split up in a free flowing part and a queuing part. The head of this queuing part is fixed at the end of the link, and a constant queue density is assumed, which are clearly unrealistic simplifications. Consequently, the modeling of blocking back effects is not very accurate. This can be illustrated with Figure 5.2, which shows both the evolution of a queue as computed with the simplified queuing model, and the evolution of the same queue as computed with a physical queuing model.
- The model requires more computation time than other models, due to the fact that the individual traffic flows following a certain route need to be tracked (which is not done in en-route route choice models and part of the pre-trip route choice models), *and* route choice sets need to be generated during the traffic simulation (which is not done in pre-trip route choice models). Since the computation time of the selected model is critical for the success of the proposed approach, this was an important aspect to be taken into account.

Because of these two drawbacks, the possibility to use the EVAQ Hybrid Route Choice Model was rejected. Since this is the only model that can realistically deal with the route choice effects of traffic congestion (i.e. combining pre-trip and en-route route choice behavior), this means that this desired model property was sacrificed in favor of other desired model properties.

[69] That is, a situation in which none of the travelers can improve his or her (perceived) travel costs by unilaterally switching routes.

Figure 5.2: Evolution of one and the same queue according to the simplified queuing model and according to the link transmission model (a physical queuing model) (Adapted from: Snelder and Schrijver, 2008)



Based on the mutual comparison of all pros and cons of the different traffic simulation models, finally the choice was made to use the model JDSMART (Zuurbier, 2010). This is a first order cell transmission model implemented in JAVA, using the Godunov numerical solution scheme.

The choice for this particular model was mainly motivated by the following facts:

- Due to its nature (i.e. being a cell transmission model), the model can very adequately reproduce the processes of queue build-up and dissipation (including blocking back effects). It has to be noted, however, that the numerical diffusion associated with this type of models is a point of attention.
- The model uses destination-specific split fractions at the network nodes. Consequently, the aggregated split fractions (i.e. aggregated over the different destinations) are not fixed (as in many other models), but are dependent on the traffic conditions in upstream parts of the network, which is more

realistic[70]. Furthermore, this destination-specific nature of the split fractions allows one to introduce certain (origin-destination specific) demand variations in the model in a more consistent way[71]. Finally, the fact that the model uses destination-specific splits fractions leaves the door open to add some functionality regarding the modeling of the route choice effect of traffic congestion in any possible continuation studies.

- The model was readily available at Delft University of Technology, including a suitable test network with a nominal demand belonging to it (albeit being a static demand matrix for the morning peak only).
- The model could easily be operated from MATLAB, using some readily available MATLAB-scripts to invoke the model and load the test network.
- Due to its implementation in JAVA and operation from MATLAB, the simulation model also is reasonably flexible, in the sense that it can be adapted to the specific tasks at hand (mainly referring to the inclusion of facilities that enable the model to handle the provided input and deliver the desired output; see also the next subsection).

It should be noted that the first two aspects of course come at the cost of a larger computational complexity, which will result in the computation time of the model to be longer than that of some other types of dynamic macroscopic traffic simulation models. First of all a cell transmission model is quite computationally intensive by its nature, since every link (except for very short links) is divided in multiple cells, for each of which every time step the traffic conditions have to be recalculated. Secondly, the destination-specific modeling of the traffic propagation over the network consumes additional computation time, since it requires that the destinations of the vehicles are tracked during this propagation.

As far as this second point is concerned, it should be mentioned however that JDSMART performs this ‘multi-class’ aspect of the traffic flow modeling in a more efficient way than other multi-class models. Instead of propagating every single user class (heading for one specific destination) individually over the network, it just propagations only the aggregated traffic flow over the network, while separately propagating changes in the traffic composition (with respect to the destinations the traffic is heading for) by means of floating particles that carry information on these composition changes (Zuurbier, 2010).

[70] After all, the composition (by destination) of the traffic arriving at a network node can be affected by variations in the amount of traffic congestion in upstream parts of the network. Without having an effect on the *destination-specific* split fractions, this may introduce variations in the *aggregated* split fractions.

[71] In a model using aggregated split fractions this is not possible (without recalculating these fractions), since an addition to the demand level of a certain origin would be divided over all destinations by these aggregated split fractions.

5.2.2 Adaptations to the model

In order to make JDSMART suitable for its role as computational core of the model, some modifications were implemented, namely:

1) Transition from a *link*-based fundamental diagram to a *cell*-based fundamental diagram (as far as its parameters are concerned).
2) Addition of some functionality to update fundamental diagram parameters (representing the traffic supply characteristics of the cells) during the traffic simulation via an efficient MATLAB routine.
3) Implementation of the possibility to model the capacity drop in a certain way (which will be explained at a later stage).
4) Addition of a routine that keeps track of the amount of vehicle-kilometers traveled per cell (which forms the input required by the incident simulator component).
5) Addition of a routine that logs the amount of vehicle-hours lost due to the occurrence of traffic congestion (which is one of the desired output indicators), with respect to a certain reference speed.
6) Addition of a routine that keeps track of the traffic dynamics on a user-specified cell/link, in order to be able to visualize these dynamics after finishing the simulation.

Most of these modifications speak for themselves, and consequently do not need further explanation. Only modifications 1 and 3 might need some clarification. For modification 3 this explanation is given in the following subsection, discussing the traffic modeling. Therefore, here only modification 1 is explained in more detail.

Initially, in JDSMART all fundamental diagram parameters (representing the various traffic supply characteristics) were defined at the *link*-level. This means that all these traffic supply characteristics are assumed to be homogeneous over the length of the link (i.e. constant over all cells in which this link is partitioned). While this more or less will be true for the time-averaged values of these supply characteristics, it will not generally be true at individual moments in time.

This follows from the work of Brilon et al (2005), who show that free flow capacity distributions cannot only be derived for clearly distinguishable bottlenecks (which in the model typically are represented by the network nodes, where links with different supply characteristics are connected with each other), but also for uniform road sections, without a distinct bottleneck (corresponding to individual segments of the (longer) links in the model).

Furthermore it should be considered that insofar incidents are concerned, the traffic supply conditions are affected only locally. In accordance with this, the parameters of the fundamental diagram should be adapted to these incidents only over a limited length, and not over the whole length of (long) links. Also note that incidents can occur everywhere along the length of the link[72]. It might be relevant to properly account for this aspect in the model, by modeling incidents not at the level of links as a whole, but at the level of individual cells. The location along a link of an incident after all could be important for the amount of spill-back of the incident-related congestion to upstream links.

For these reasons, it was considered desirable to be able to vary the traffic supply characteristics on the level of individual cells, rather than on the level of links as a whole. Therefore the model was adapted to facilitate this, by changing the link-based definition of the fundamental diagram parameters into a cell-based definition.

5.2.3 Traffic modeling

This subsection discusses the way in which JDSMART was employed to model the traffic propagation over the network. Aspects that will be considered are the (spatial and temporal) discretization, the fundamental diagram, the capacity drop, route choice, the warming-up time and the cooling-down time.

Discretization

For the simulation of the traffic propagation a suitable temporal and spatial discretization should be chosen (i.e. time step and cell size, respectively). That is, the time step and cell size should be such small that the level of accuracy of the traffic flow modeling is sufficient for the final results to make sense. On the other hand, one should be well aware of the fact that if one reduces both the time step and the cell size with a factor x, this will increase the computation time by a factor of approximately x^2.

In this field of tension a time step of 5 seconds was chosen. For the cell size the corresponding minimum value was used, calculated by multiplying this time step by the (link dependent) speed limit. This relation between time step and minimal cell size follows from the Courant Friedrichs Lewy condition, which entails that within one time step 'information' in the traffic flow might not cover more than one cell length, in order for the numerical solution scheme to converge to the original continuous equations. From a small test with DSMART (the MATLAB-version of JDSMART) it was concluded that a (significantly) larger time step results in clearly different traffic flow patterns. For a smaller time step it seemed that the traffic flow patterns do not change importantly anymore.

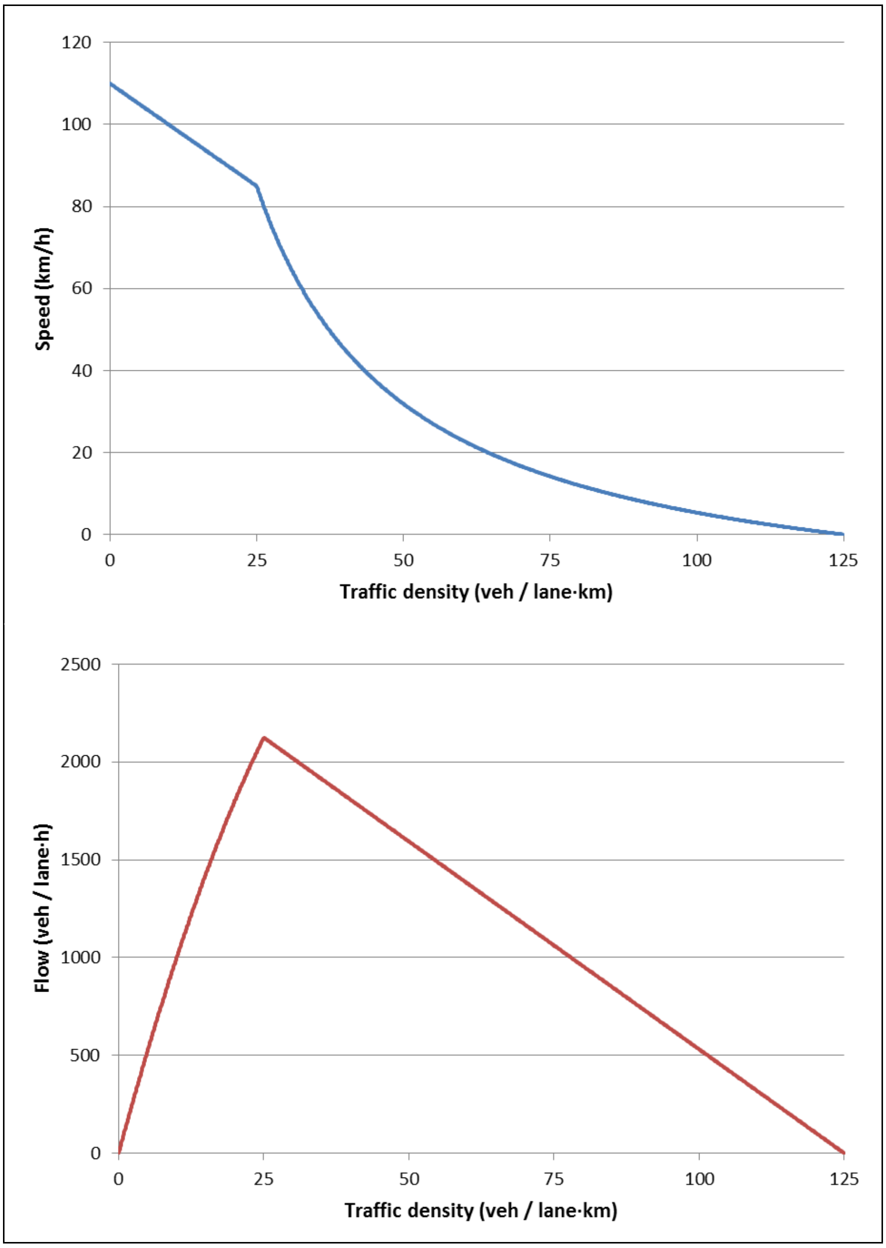
It has to be noted that for a time step of 5 seconds the numerical diffusion seems not negligible yet. This means that a smaller time step in fact would be preferable. A time step smaller than 5 seconds was computationally not feasible however.

[72] Of course in reality accidents occur relatively frequently at locations with geometrical discontinuities, and thus near link ends, but vehicle break downs (which together form the largest subset of incidents) of course will be distributed more evenly over space.

Fundamental diagram and capacity drop

For the modeling of the traffic flow on the motorways the well-known fundamental diagram of Smulders was selected (one of the default fundamental diagrams available in JDSMART). In this fundamental diagram (depicted in Figure 5.3) the speed decreases linearly with the density from the free speed at a zero density to the critical speed at the critical density (for which the traffic flow attains its maximum value). In the congested branch of the fundamental diagram the speed decreases hyperbolically with the density from the critical speed at the critical density to a zero speed at the jam density, corresponding to a linear decrease of the traffic flow.

Figure 5.3: Fundamental diagram of Smulders (upper part: speed versus density; lower part: flow versus density)



This fundamental diagram however does not include a capacity drop (i.e. a drop in the capacity which is observed after the onset of congestion), a phenomenon which was discussed in chapter 2. Since it was considered to be of potential importance for the final results, it was decided to add the capacity drop phenomenon to the traffic flow modeling. This potential importance can be explained in the following way. The capacity drop results in the phenomenon that a traffic jam once it is created has the tendency to enhance itself, meaning that this traffic jam will remain existent for a longer period of time (and reach a longer length) than without this drop in capacity. Since taking into account all sources of variation is hypothesized to result in more traffic congestion occurring in the simulation outputs (considered at an overall level), the capacity drop might play a more important role when all sources of variation are taken into account. This way, the capacity drop could affect the differences between the results obtained from a 'traditional' evaluation (looking at a representative situation only) and those obtained from an evaluation in which all sources of variability are taken into account, and consequentially affect the additional insights that the latter approach could provide us with.

Modeling the capacity drop in a first order traffic flow model is not a straightforward thing however. If the capacity drop is simply added to the fundamental diagram as shown in Figure 2.1 and Figure 2.2, the physical model would enable shock waves traveling upstream with an unrealistically high or even infinite speed, causing the capacity to drop everywhere on its way. In the discretized model this upstream speed of course would be maximized at one cell length per time step (corresponding to the free speed). This however is still much too fast, since the real maximum speed with which traffic congestion spills back is about 20 km/h.

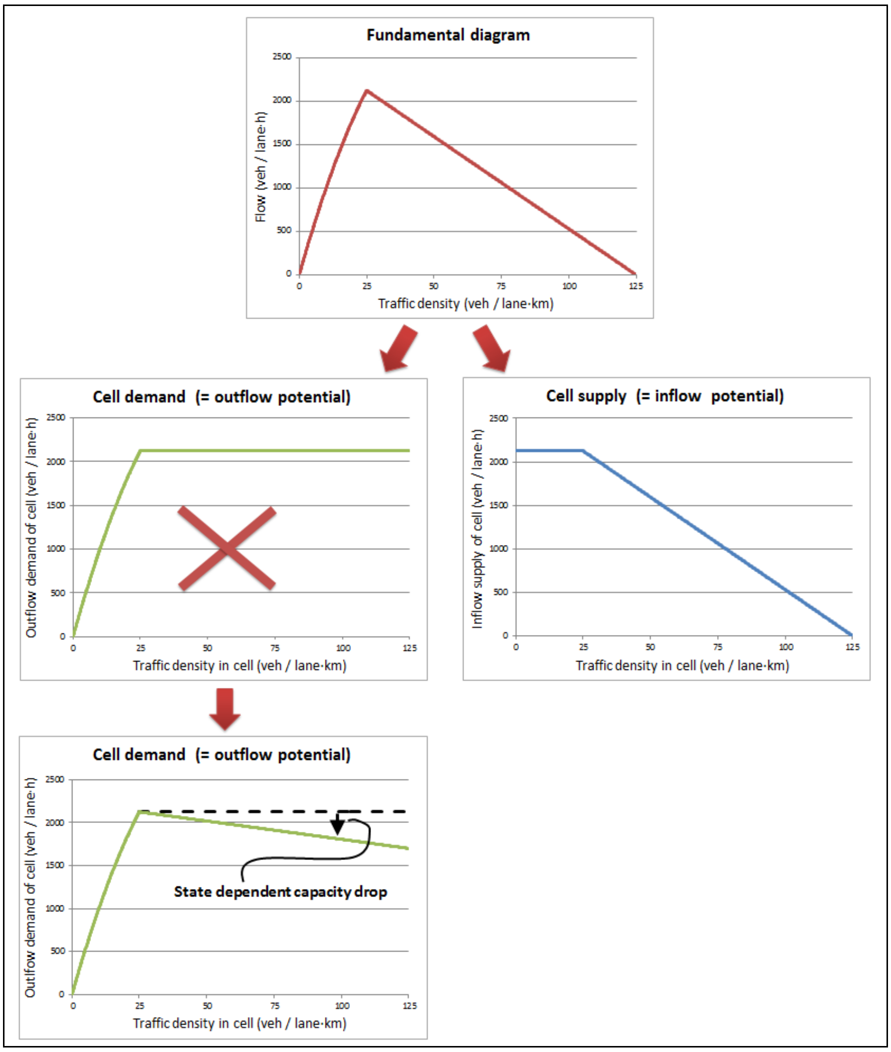
Two possible explanations can be suggested for the fact that this phenomenon (i.e. very high shock wave speeds induced by the capacity drop) is not observed in reality:

- After the occurrence of a traffic breakdown the speed on a completely saturated road in reality is not instantaneously lowered, as predicted by the fundamental diagram with a capacity drop. Instead, drivers will gradually decelerate, resulting in decreasing headways (and thus in an increasing traffic density). Accordingly, the traffic congestion will spill back at a much lower speed. This second order phenomenon is not taken into account in a first order traffic flow model.

- In reality the capacity drop might be dependent on the traffic state within the queue. It seems intuitively right to assume that the worse the traffic conditions in the traffic jam are (corresponding to a higher density and lower speed), the larger the capacity drop will be. Imagine for example the situation in which vehicles that flow out of the queue have to accelerate from a standstill. In this case probably much larger headway gaps between the outflowing vehicles will occur than in a situation in which the vehicles flow out from a queue in which the traffic is still reasonably flowing, with a speed of for example 60 km/h. If the capacity drop indeed is significant only in situations with outflow from queues in which the density is far above the critical density, it will not result in the emergence of shock waves with very high speeds. Again this (possible) effect is not taken into account in a first order traffic flow model.

A known solution (see Appendix A1.5) to prevent the emergence of unrealistically fast shock waves in a traffic model with capacity drop is to define this drop as a function of the state of the congested traffic, in line with the second explanation above. This approach was also selected in this case. For the maximal capacity drop (corresponding to outflow from standstill) a value of 20% was assumed. Looking at empirical values that are found for the capacity drop, this seems a reasonable though slightly conservative value. Figure 5.4 shows the corresponding change in the cell demand function of the Godunov discretization scheme. Please note that due to this modification the model cannot longer be considered a discretization of the continuous LWR (Lighthill-Witham-Richards) model anymore.

Figure 5.4: Modification in the cell demand function of the Godunov discretization scheme, implemented to account for the capacity drop



Route choice

As mentioned before, the desired model property of realistically dealing with the route choice effects of traffic congestion (which would only be possible by using the EVAQ Hybrid Route Choice Model) was sacrificed in favor of other desired model properties.

If pre-trip route choice and en-route route choice cannot be combined (for which the EVAQ Hybrid Route Choice Model would be needed), two possibilities for modeling the route choice process logically remain, namely assuming pre-trip route choice, or assuming en-route route choice. Although the first option is not realistic, the second option clearly is even more unrealistic. Modeling route choice in that way would correspond to the assumption that road users base their route choice decisions entirely on the instantaneous traffic conditions (or those of a short time interval ago, depending on the exact modeling choices), without considering their past travel experiences. In view of the fact that research shows that a serious disruption is needed to make a substantial part of the road users decide to deviate from their 'standard' routes (see section 2.3.3), this assumption clearly is not realistic. Probably, for many situations it is still more realistic then to assume that drivers do not deviate from their 'standard' routes at all.

Based on the above, it was decided to assume pre-trip route choice. (In fact there was not really an alternative either, since as opposed to its MATLAB-version DSMART, JDSMART does not possess an en-route route choice component). The next issue then of course is how this pre-trip route choice pattern should be calculated. Basically, there are two options for this: calculating a dynamic equilibrium route choice (for which in JDSMART an equilibrium assignment component is available), or simply performing a shortest path calculation on the free flow travel times. In the latter case the route choice is assumed constant over the day. Please note that if an equilibrium assignment would be taken as point of departure, such an equilibrium calculation of course would be calculated only once, namely for the representative situation. The resulting route choice pattern would then be applied in every other simulation run (with randomized traffic demand and supply characteristics). After all, it would not make any sense at all to recalculate this equilibrium for every individual randomized simulation, since this equilibrium is a long-term notion. It should be noted here that long-term road works will not be included in the simulations, for reasons explained in section 5.4.13. Clearly, for this type of road works the equilibrium would have to be recalculated.

Although the route choice pattern resulting from a dynamic traffic equilibrium calculation could theoretically result in a significant more realistic network loading, it yet was decided to assume the route choice pattern resulting from a shortest path calculation on the free flow travel times. The associated reduction in realism of the network loading was not considered of vital importance for the research tasks at hand.

There are two reasons for not calculating a dynamic traffic equilibrium. First of all, an important difficulty with respect to this calculation would have been that the JDSMART equilibrium assignment component did not provide the facilities to deal with a traffic demand pattern in which the various origin-destination demands vary independently from one another (which actually is the case in the generated dynamic origin-destination matrix). Only for demand patterns in which the various origin-destination demands vary according to one and the same temporal pattern, a dynamic equilibrium could be calculated in a straightforward way.

The second reason is that in JDSMART, dynamic changes in the route choices (derived from the equilibrium calculation) are not only applied on newly departing traffic, but also on all vehicles that are already present in the network. This is related to the fact that in JDSMART vehicles are not tracked according to their routes, but only according to their destination. As a result of this latter, changes in route choice have to be applied via the (destination-specific) split fractions in the network nodes.

This principle of applying dynamic changes in route choice on all traffic that is already present in the network is considered very undesirable within the context of this project. To illustrate this, consider two randomized simulation runs, one in which hardly any traffic congestion occurs, and another in which traffic conditions get heavily congested. Assume that in both cases at a certain moment in time the route choices are updated, on the basis of the (same) pre-calculated dynamic equilibrium. Although in both cases the change in route choice is updated at the same moment in time, in the second simulation most vehicles will have made much less progress on their journey than in the first simulation run, due to the traffic congestion. This means that their route choices are altered at a different stage of their trip, meaning that the routes that are traveled in the second simulation run in the end are different from those in the first simulation run. This means that the model would automatically generate a certain variation in route choice, which does not have any theoretical basis at all. This of course would be highly undesirable (i.e. much more undesirable than the reduction in realism of the network loading that is associated with simply assuming the free flow route choice pattern).

Warming-up period

At the start of the simulation the traffic network is still empty. It will then take some time before all network parts are loaded in a way which is more or less in 'equilibrium' with the imposed traffic demand and supply characteristics. As a result, the first part of the traffic simulation will yield erroneous results (in terms of the calculated traffic conditions). Therefore a warming-up period is added to the simulated period. That is, before the actual start of the period that is to be simulated, the traffic simulation model is already run for a certain pre-specified amount of time, with traffic demand and supply values equal to those for the first time interval of the actual simulation period.

Since the default simulated period in the model is the whole 24-hours period of a day, a relatively short warming-up period is sufficient (because of the low traffic volumes at midnight). Considering that the amount of time required for reaching the other end of the network is about 15 to 20 minutes (in free flow conditions), a default warming-up period of half an hour was chosen.

Cooling-down period

For the computation of travel times JDSMART uses floating particles, which are propagated along with the flow. At the start of each 5-minute time interval the model releases new instances of such particles at the start of the different routes for which the travel time distributions are to be evaluated. Of course it should be taken into account then that at the end of the simulated period not yet all of these particles reached the end of their routes. Consequently, the travel times for the last intervals of the simulated period are not known yet either. Therefore the traffic simulation run is extended until the last travel time particle finishes. During the whole of this 'cooling-down period', the demand and supply values of the last 5-minute interval of the day are used.

5.3 General approach to the modeling of the variations in traffic demand and supply

This section describes the general methodology that is used in the modeling of the variations in the traffic demand and supply. The specific aspects in the modeling of the different individual sources of variability are discussed in the next section. The current section starts with a discussion on the most basic principles of the modeling of the variations, which are applicable to both the demand and supply characteristics. After that, it is explained how variations in the traffic demand are modeled. The last subsection explains how this is done for the traffic supply characteristics.

5.3.1 General concept

Performance of a large number of randomized simulation runs

As was already explained before, the model deals with the variability in traffic demand and supply by performing a large number of randomized traffic simulation runs. For each of these simulation runs, the demand and supply randomization components generate tables with the traffic demand and supply values (per 5-minute interval) that are to be used in that specific run. In these tables the various sources of variability discussed in chapter 2 are reflected. Here the within-day variability is reflected *within* a table for a specific run, while the day-to-day variability is reflected in the differences *between* the tables for different runs.

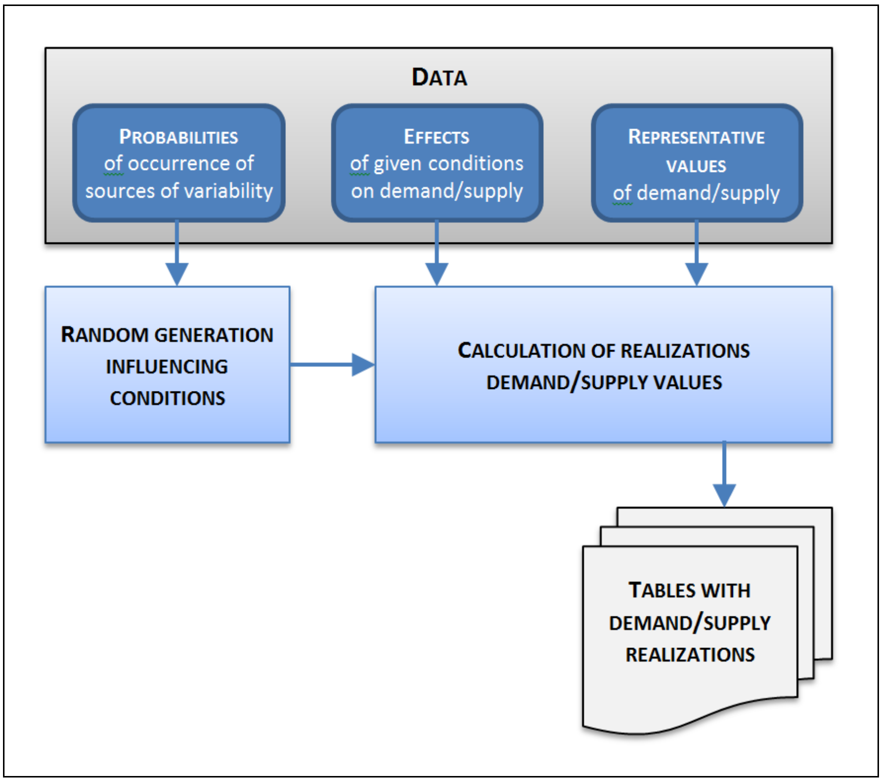
Random generation of demand and supply tables

The demand and supply tables are generated using the Monte Carlo technique and the information on effects and probabilities/patterns of occurrence provided in chapter 2. Important interdependencies between the different sources of variability are taken into account by using conditional probability specifications. An important limitation of the information from chapter 2 is that it is not always specific enough for use in the demand and supply randomizers. Since the required information has not been found in literature, in fact empirical research into a number of aspects would be needed. Because of the fact that this however was not feasible within the scope of this project, assumptions had to be made on these aspects.

It is important to note here that the (supply) effects of any possible *incidents* are not yet reflected in the tables generated by the supply randomization component. This is because of the fact that the occurrence of incidents is strongly dependent on the prevailing traffic conditions. Therefore, random realizations of incidents are not generated in advance of the actual traffic flow simulation (like all other demand and supply influencing factors), but concurrently with this simulation. For every generated incident, the supply tables are adjusted accordingly in real-time.

Figure 5.5 schematically explains the working of the demand and supply randomization components of the model. They operate in two steps. First, random realizations of the different influencing factors (also indicated as 'sources of variability') are generated. For this, data on the probabilities/frequencies of occurrence of the different possible conditions are used. These data are specified in the form of discrete or continuous probability distributions. In order to account for the fact that some influence factors are dependent on other ones, part of these probability distributions is specified conditionally. Note that this means that some of the sources of variability have to be dealt with before others in the randomization procedure, since the randomly generated statuses of the formers are needed as input for the randomization of the latters.

Figure 5.5: Schematic explanation of the working of the demand and supply randomization components



When the random realizations of the influencing factors have been generated, these are translated in effects on the traffic demands and/or traffic supply characteristics, using tables in which these effects are specified in terms of correction factors with respect to the representative values. By applying the correction factors on the representative values of the demand and supply characteristics, the stochastic realizations of these demand and supply characteristics are found. These realizations are stored in tables, which are passed on to the computational core of the model. The latter subsequently will simulate the traffic conditions that would arise from these demand and supply characteristics.

With respect to the correction factors mentioned above, it should be noted that it is assumed in the model that multiple correction factors (expressing the effects of different sources of variability) can be applied simultaneously on the representative demand and capacity values without any further corrections for any possible interaction effects. This corresponds to the assumption that the effects of different sources of variability are independent of one another. In reality this is not the case. Some effects might strengthen or weaken each other, or partly overlap with one another. These interaction effects however are often too difficult to estimate to be taken into account in the model.

Replication of the stochastic realizations

An important desirable property of any stochastic model is that its stochastic realizations are reproducible. This entails that if one performs the same simulation run again, exactly the same stochastics realizations are generated, and consequently exactly the same output is obtained. This model property is especially important in comparative evaluations. After all, in such situations one typically does not want the difference between the results of two model runs to be distorted by the use of different pseudo-random number values.

Obviously, the requirement that the stochastic realizations can be replicated applies equally well to the model at hand. Therefore, the model was programmed to reset the random number generator to its initial state each time a new model run is started. This ensures that all model runs are based on the same pseudo-random number stream.

This however is not sufficient to obtain equal stochastic realizations in different model runs. For this it is not only required that the same pseudo-random number stream is generated, but also that these pseudo-random numbers are employed at exactly the same locations within the model.

This additional requirement can be illustrated as follows. Imagine that we want to asses the relative contribution of weather influences to a certain congestion indicator. This can be done by comparing the results of two different model runs: one in which weather influences are switched *on*, and one in which these are switched *off*. In the latter case no stochastic realizations of the weather conditions would be taken anymore, which would decrease the amount of pseudo-random numbers used per simulation. The consequence of this latter would be that in the second model run every pseudo-random number is used at a *different* location than in the first run. As a result, completely different stochastic realizations would be generated, in spite of the fact that the random number stream itself is exactly the same. Obviously, this is highly undesirable.

It is stressed that the occurrence of this problem is not limited to situations in which a certain influence factors is switched on or off, as the above example might suggest. The amount of pseudo-random numbers used per simulation may also vary under influence of the fact that some 'supplementary' stochatisic realizations are only needed if certain other realizations turn out in a certain way. An example of such a 'supplementary' stochastic realization is the random generation of an incident duration, which obviously is only needed when an incident is generated.

This problem can only be prevented by ensuring that all elements of the model employ a *constant* amount of pseudo-random numbers, no matter what the model settings and outcomes of other stochastic realizations are. This has been achieved by implementing the following modeling principles:

- Switching off a source of variation does ***not*** affect the generation of stochastic realizations for this source. Instead, it just prevents these stochastic realizations to have effects on demand and supply, by blocking their correction factors.
- In all cases where pseudo-random numbers might be needed for the generation of 'supplementary' stochastic realizations, these are just ***always*** generated, regardless of whether they actually will be used or not.

For the modeling of incidents the latter for example means that a pseudo-random number for the stochastic realization of the incident duration is generated for each and every *potential* incident, irrespective of whether this incident indeed is generated or not. Since there is a potential incident for every possible combination of network cell and time interval, this means that actually a very large amount of pseudo-random numbers have to generated, most of which finally are not even used. From this it may be concluded that this solution comes at the expense of a sharp deterioration in the computational efficiency of the model. Relatively speaking, however, the increase in computation time is only very limited.

5.3.2 Variability in the traffic demands

Demand variables

The output of the demand randomizer consists of a dynamic origin-destination matrix. This three-dimensional table contains the origin-destination traffic demands per five-minute interval. The destination-specific split fractions at the network nodes (corresponding to the route choices) are not varied, but are assumed fixed. This is not entirely correct, but too little information is available to vary these split fractions in a sound way. It is difficult to assess to which extent this simplification is reasonable. On the one hand, the strong habitual component in route choice behavior suggests that the variability in these split fractions will not be very large. On the other hand, the variation in the population of road users may obviously bring a certain degree of variability with it. However, in the test network that was considered within this project (see chapter 8) all motorway trips have only a very limited number of route alternatives. Many of these trips even have only one realistic route possibility. This means that the variability in route choices can safely be neglected, without significantly harming the final model outcomes.

Note that in spite of the fact that the *destination specific* split fractions are assumed fixed (per time interval), the *aggregated* split fractions are *not fixed* in the traffic simulation, since the composition of the traffic (by destination) may vary under the influence of variations in the levels of origin-destination traffic demands, as well as under the influence of traffic congestion in upstream parts of the network.

Method of implementing the sources of variation

As apparent from chapter 2, most of the sources of variation in traffic demands have clearly different impacts on the traffic demands during peak periods and those during off-peak periods. This means that multiplication of the whole 24-hours demand pattern with one uniform scaling factor would not be an appropriated method to deal with these demand variations. Instead a method is required in which a distinction can be made between these different parts of the day.

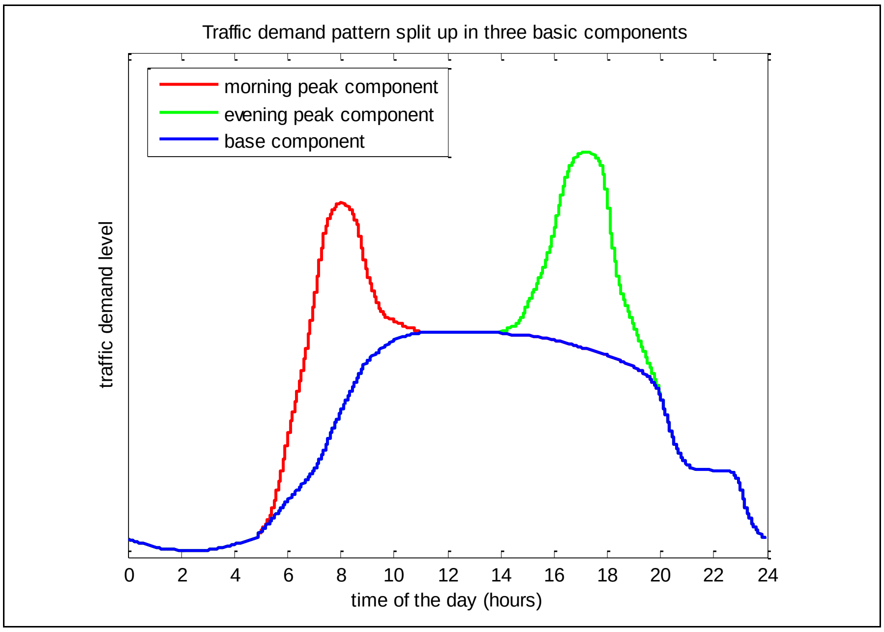
However, simply applying different correction factors for different parts of the day (without taking care of gradual transitions) would not be an appropriated method either. Such an approach would lead to abrupt jumps in the traffic volumes on the network, resulting in unrealistic traffic conditions (such as the emergence of congestion at predetermined times of the day). Therefore, a method is desired that computes demand pattern realizations in which the peak and off-peak gradually run over into one another.

To meet the above requirements, a method was devised in which the 24-hours demand pattern of an origin-destination relation is considered to be composed of three basic components[73]:

- a *base component*, which describes the traffic demand pattern as if there were no peak periods;
- a *morning peak component*, which – as its name implies – adds a morning peak to the traffic demand pattern; and
- an *evening peak component*, which – as its name implies – adds an evening peak to the demand pattern.

This principle is illustrated in Figure 5.6.

Figure 5.6: Traffic demand pattern for an origin-destination relation considered to be composed of three basic components



[73] Inspired by the so-called 'T-values Model' (also known as the 'Tones Methodology'), described in (AVV, 1997) and (Transpute, 2003),

For each origin-destination relation a realization of the traffic demand pattern is now calculated in the following three steps:

- First, stochastic realizations are generated for the following three aggregated demand values::
    - the total morning peak traffic demand (7-9h)
    - the total evening peak traffic demand (16-18h)
    - the total off-peak traffic demand

  This stochastic generation proceeds according to the procedure that has been described in the previous section.

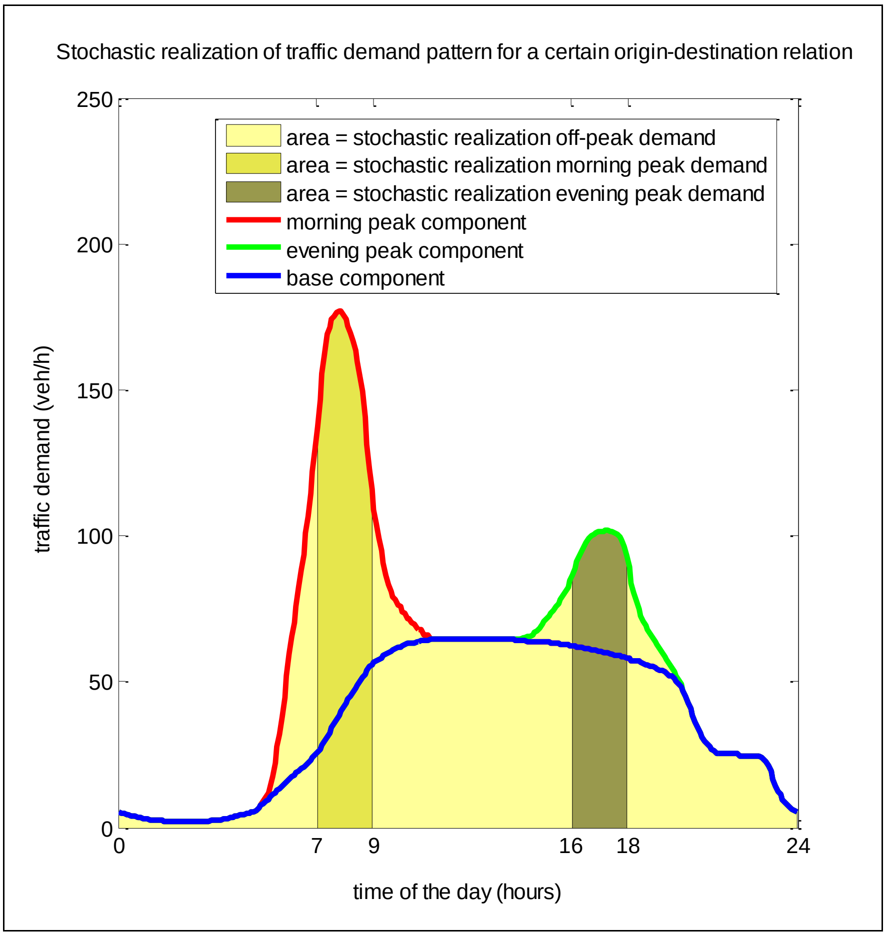
- Next, the combination of the three basic demand pattern components (depicted in Figure 5.6) is fitted to these three aggregated demand values, using linear algebra. An example of this is shown in Figure 5.7. This fitting procedure results in a weight for each of these three basic components.
  It should be noted here that a negative weight is not accepted. If such a situation occurs, this is dealt with by performing a different fitting procedure, in which the component concerned is set at zero. The obvious consequence of this is that it becomes impossible to satisfy all three demand constraints. This means that one of these constraints has to be abandoned. For this the constraint on the off-peak demand is chosen, in view of the fact that most of the traffic congestion occurs during the peak periods. By carefully tuning the representative demand patterns (in such a way that all weights have a reasonable buffer for corrections) it was more or less guaranteed however that any possible deviations from the desired off-peak corrections will be only marginal.

- Finally the realization of the demand pattern is obtained by multiplying the three basic components with their calculated weights, and superimposing them.

Obviously not all influence factors can be modeled in this way. This is the case for events (which typically are associated with an arrival and a departure peak, which obviously do not coincide with the global peaks in traffic demand) and weekend days and (part of the) public holidays (which do not have peak periods). The way of modeling of these effects will be discussed in section 5.4, which considers the modeling of the different sources of variability on an individual basis.

It should be noted that the pattern calculated according to the above procedure actually is not the final realization of the demand pattern yet. After the calculation of this pattern still a final step is to be carried out. In this step a certain ‘random noise’ is added to it, in order to account for the intrinsic random variability in the travel decisions of individual travelers. This will be discussed in more detail in section 5.4 as well.

Figure 5.7: Fitting of the combination of the three basic demand components to the three (stochastically generated) aggregated demand values



5.3.3 Variability in the traffic supply characteristics

Supply variables

The output of the supply randomizer consists of tables with the stochastically generated values of a number of supply characteristics. These supply characteristics are the parameters that define the fundamental diagram discussed in section 5.2.3. Separate values are given for each cell of the road network and for every 5-minute interval of the day.

All parameters of the fundamental diagram are considered in the randomization procedure. These are:

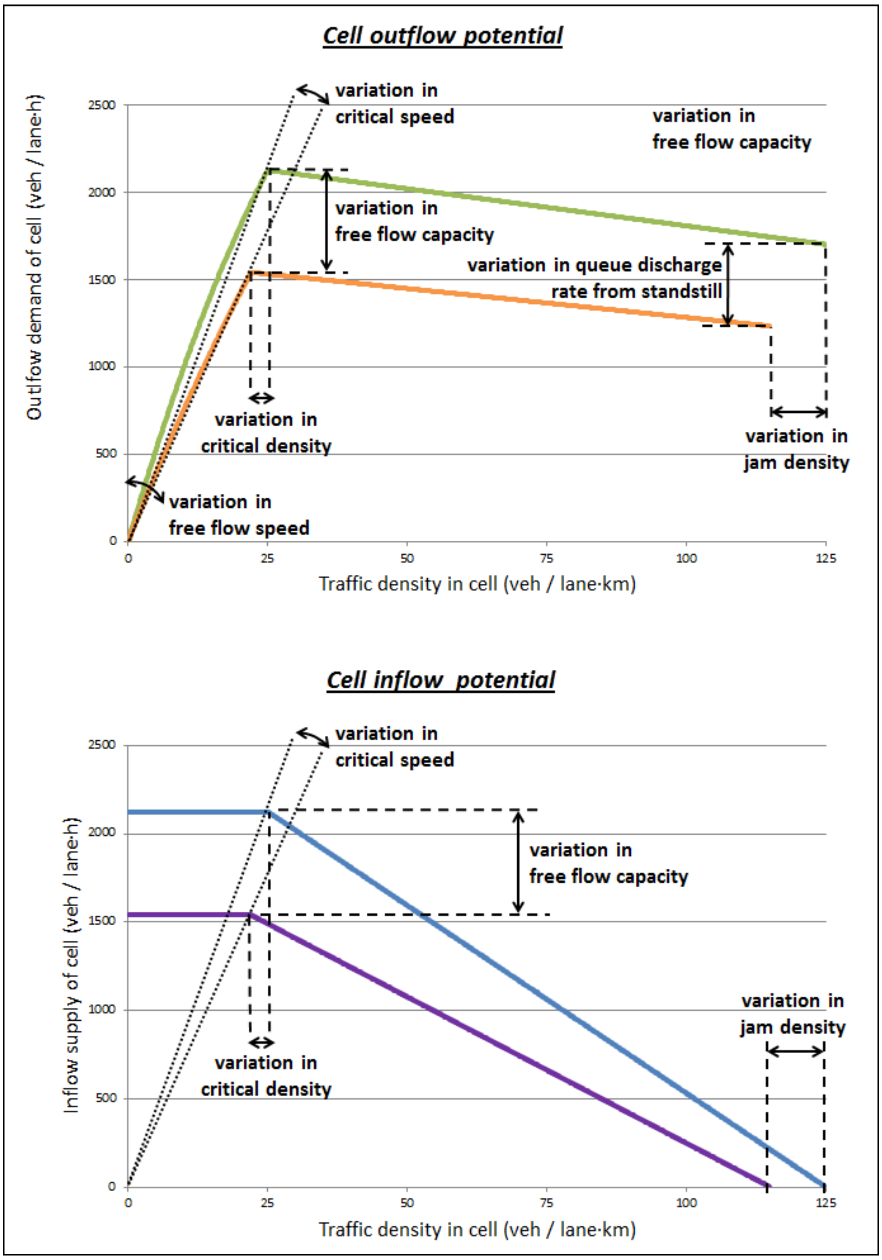
- the *free speed*: the speed for an (almost) zero traffic density;
- the *number of available lanes;*
- the *critical density* (per lane): the traffic density for which the maximum flow (i.e. the free flow capacity) is achieved;
- the *critical speed:* the speed connected to the critical density;
- the *free flow capacity* (per lane): the maximum traffic flow that can be achieved under free flow conditions (equal to the product of critical density and critical speed);
- the *queue discharge rate from standstill* (per lane): the outflow from a queue in which the traffic has come to a standstill; and

- the *jam density* (per lane): the density of traffic that has come to a standstill.

An example of the variation of these parameters is given in Figure 5.8, which shows two stochastic realizations of the inflow and outflow curves of a cell. Note that the values expressed per lane can easily be converted into values for the roadway as a whole, by multiplication with the number of available lanes.

It should be noted that not all of the above mentioned parameters are varied for each and every source of variability. Some sources of variability have significant effects on part of the parameters only, while leaving the others (more or less) unaffected. The jam density is an example of a parameter which is relatively constant. This parameter is only varied with the variations in vehicle population.

Figure 5.8: Example of the variation introduced in the traffic supply characteristics. The upper diagram shows two stochastic realizations of the curve describing the outflow potential of a cell. The lower diagram shows the corresponding realizations of the curve describing the inflow potential of this cell.



Method of implementing the sources of variation

As far as the *free flow capacity, queue discharge rate from standstill* and *jam density* are concerned (all of them expressed per lane), the different sources of variation are accounted for exactly in the manner described in subsection 5.3.1 (which can be considered the 'standard' manner). That is:

- First, stochastic realizations of the different influencing conditions (corresponding to the different sources of variability) are generated.
- Next, these stochastically generated influencing conditions are translated into correction factors on the different supply parameters, representing their effects on these parameters.
- Finally, these correction factors are applied on the representative values of the parameters in question, in order to obtain the stochastic realizations of the latters.

For the free flow capacities and queue discharge rates there is however one additional step to be taken. In this step a certain amount of 'random noise' is added, in order to account for the intrinsic random variability in human driving behavior (similar to the 'random noise' that is added to the traffic demands, in order to account for the intrinsic random variability in the travel decisions of individual travelers). This will be discussed in more detail in a subsection of paragraph 5.4, which is specifically devoted to the modeling of this intrinsic random variability in driving behavior.

For the *number of available* lanes the above procedure is different in the sense that no correction factors are used. The obvious reason for this is that the number of available lanes is not a continuous variable, but a discrete one. Instead of using a correction factor, for each stochastically generated influence factor it is determined how many lanes it blocks (if any). This number of lanes is then subtracted from the total number of available lanes. For the number of remaining lanes a minimum of one is assumed. This means that complete blockages are not considered. The reason for this is that in such situations an important feedback from traffic conditions to route choice would occur. Due to the fact that this feedback is not taken into account, the model would produce very unrealistic output for complete blockages. Also note that in reality in part of such cases the hard shoulder is opened to traffic, which results in a situation which actually is more or less equivalent to the modeled situation with one remaining lane.

Also for the *free speed* the applied procedure is slightly different from the standard one given above. This is for the following two reasons:

- The relative effects on the free speed are typically dependent on the 'undisturbed' free speed value (which in turn depends on the speed limit of the road), which means that these effects cannot be expressed in universally applicable correction factors. For example, the relative free speed reduction associated with road works will be much larger on 120 km/h roads than on 80 km/h roads. Note that in this thesis, it is assumed that if a temporary speed limit reduction is put in place, the reduced limit is strictly respected. In reality this is obviously not necessarily the case, since speed limits can be exceeded by road users. This effect is neglected here, however. It is not relevant for the research task at hand.

- A specific characteristic of the free speed is that the effects of different simultaneous influence factors are often clearly not (fully) additive or multiplicative. Consider for example a situation in which a vehicle breakdown occurs at a location where road works are going on. Imagine that this vehicle breakdown would have caused a free speed reduction from 100 to 90 km/h. However, due to the road works this free speed is already at a lower level of 70 km/h. In this situation it is rather unlikely that the vehicle breakdown would lead to any additional free speed reduction.

Because of the fact that the way in which the speed effects are accounted for is different for the different sources of variability involved, no general description of this can be given here. Individual descriptions per source of variability are given in section 5.4.

The correction factors for the *critical density* and *critical speed* are computed in a special way. By definition, the product of the critical density and the critical speed is equal to the free flow capacity. For this reason, the correction factors for the critical density and critical speed are calculated from the correction factor of the free flow capacity. More specifically, the correction factor of the free flow capacity is divided over the critical density and critical speed. Here the ratio of the corrections on critical density and critical speed is chosen differently for the different sources of variation involved. After all, some sources of variation can be assumed to have a larger effect on the critical density, while others will primarily affect the critical speed. In the model this is expressed in a special 'division parameter' z_i for every source of variability *i*.

Equations 6.1 and 6.2 indicate how the correction factors for the critical density and critical speed (indicated as crf_{dens_i} and crf_{speed_i}) can be calculated from the correction factor for the free flow capacity (crf_{cap_i}), using this division parameter z_i. Note that the product of both correction factors gives back the correction factor for the free flow capacity crf_{cap_i}, as required by the definitions of the variables involved.

Equation 5.1

$$crf_{dens_i} = crf_{cap_i}^{z_i}$$

Equation 5.2

$$crf_{speed_i} = crf_{cap_i}^{1-z_i}$$

The critical speed can obviously not be larger than the free flow speed. This therefore is not allowed in the model. If nessecary, this is compensated for by a larger correction factor on the critical density.

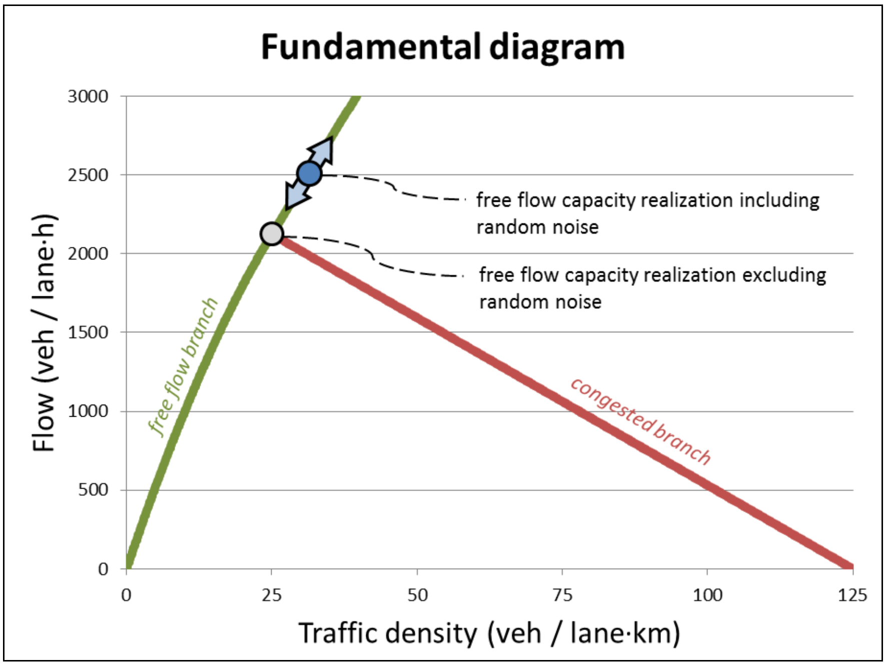
Undesirable side effect

An important note to be made here is that the implementation of random noise in the free flow capacity value actually has a negative side effect, via the associated variations in the critical speed and critical density. These variations affect the course of the whole flow-density relation. This way, they introduce variations in the traffic flows in the cells, which are probably not very realistic. Although the final impact on the traffic conditions seems limited, it is not clear to which extent this leads to undesired model behavior.

This effect can be prevented by implementing the random capacity noise only in the upper bounds on the inflow and outflow of cells. This means that this random noise is not accounted for in the basic flow-density relation. The obvious drawback of this approach is, however, that inconsistencies are introduced in the fundamental diagram.

Therefore another modeling approach was developed, which has no such inconsistencies. According to this modeling approach, the random noise in the free flow capacity corresponds to a variation *along* the free flow branch of the fundamental diagram, and not to a variation of the fundamental diagram itself. This is illustrated in Figure 5.9. In Appendix 2 a more detailed description is given of this approach.

Figure 5.9: New modeling approach, according to which the random noise in the free flow capacity corresponds to a variation along the free flow branch of the fundamental diagram, instead of to a variation of the fundamental diagram itself.



In a small test this new modeling approach seemed to produce reasonable model behavior. It was considered too risky however, to apply this modeling approach in the model evaluations presented in chapter 8. For this, the behavior of this modeling approach would first have to be studied in much more detail, in order to find out whether it really results in a realistic traffic flow modeling.

5.4 Modeling of the different sources of variation

In this section the modeling of the individual sources of variation will be considered in more detail. To each of these sources a separate subsection is devoted. These subsections are divided in two parts: one considering the modeling of the occurrence of the source of variability, and the other considering the modeling of its effects.

5.4.1 Time of the day

Classification:

- Modeled as a **CONTINUOUS** influence factor,
- with a **DETERMINISTIC** occurrence,
- and a **NETWORK-WIDE** effect
- on the traffic **DEMANDS**.

Special position within the set of all sources of variability:

Is the only source of variability which is also taken into account in studies according to the more traditional approach (focusing on the representative situation).

Modeled interdependencies:

Table 5.1: Modeled dependencies between the time of the day and other sources of variability

Modeled dependencies on other sources of variability	Considered source of variability	Modeled influences on other sources of variability
None	Time of the day	Events Weather conditions Darkness Driver population Vehicle population Road works

Procedure for simulating the time of the day:

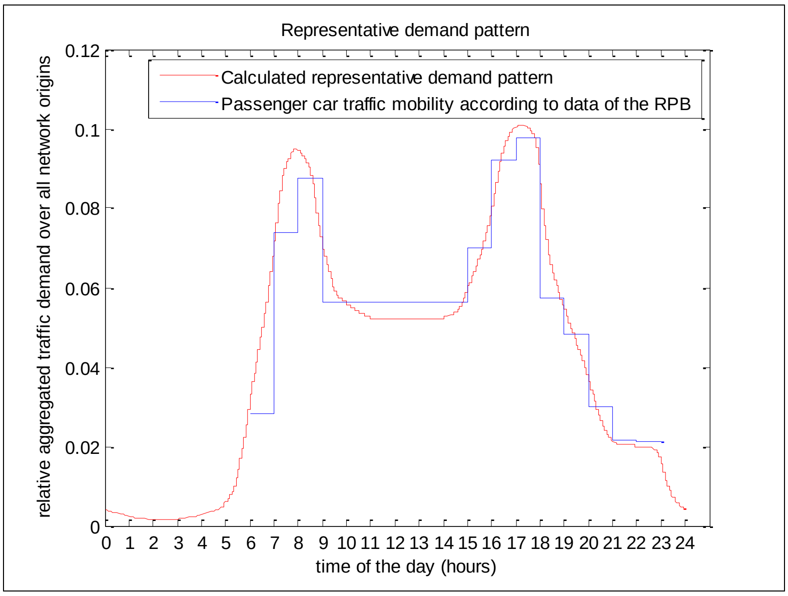
The total simulated time period of 24 hours is divided in 288 time intervals of 5 minutes. In the simulations these different time intervals are considered one after the other.

Modeling of the effects:

- The relation between the time of the day and the traffic demand is contained in three basic demand components, (depicted in Figure 5.6), which have to be fitted to certain origin-destination specific demand levels, as explained in section 5.3.2.

- These basic components have been calibrated in such a way that if they are fitted to the *representative* demand levels, an over the network aggregated demand pattern is obtained which approximately corresponds to the general passenger car mobility pattern provided in Figure 2.5. This calibration result is shown in the figure below. Note that at some points the calibrated pattern *intentionally* departs from the more general car mobility pattern. This is related to the following observations:
    - In motorway traffic the onset of the morning peak typically occurs earlier than reflected in the general car mobility pattern.
    - In motorway traffic the difference between morning and evening peak is assumed to be smaller than in the overall car mobility pattern. Therefore, a value of 1.10 was used for the ratio between the evening and morning peak demands, rather than a value of 1.18.
    - On motorways the traffic demands typically reach their peak *in the middle* of the peak periods (or even before), rather than in the second hour of these.

  Finally it should be mentioned that the ratio between the total peak demand and the total off-peak demand has been assumed at 0.55, corresponding to an average peak hour fraction (relative to the 24h-demand) of 8.9%.



5.4.2 Day of the week

Classification:

- Modeled as a **DISCRETE** influence factor,
- with a **DETERMINISTIC** occurrence,
- and a **NETWORK-WIDE** effect
- on the traffic **DEMANDS**.

Modeled interdependencies:

Table 5.2: Modeled dependencies between the day of the week and other sources of variability

Modeled dependencies on other sources of variability	Considered source of variability	Modeled influences on other sources of variability
None	Day of the week	Vacations Special days Events Driver population Vehicle population Road works

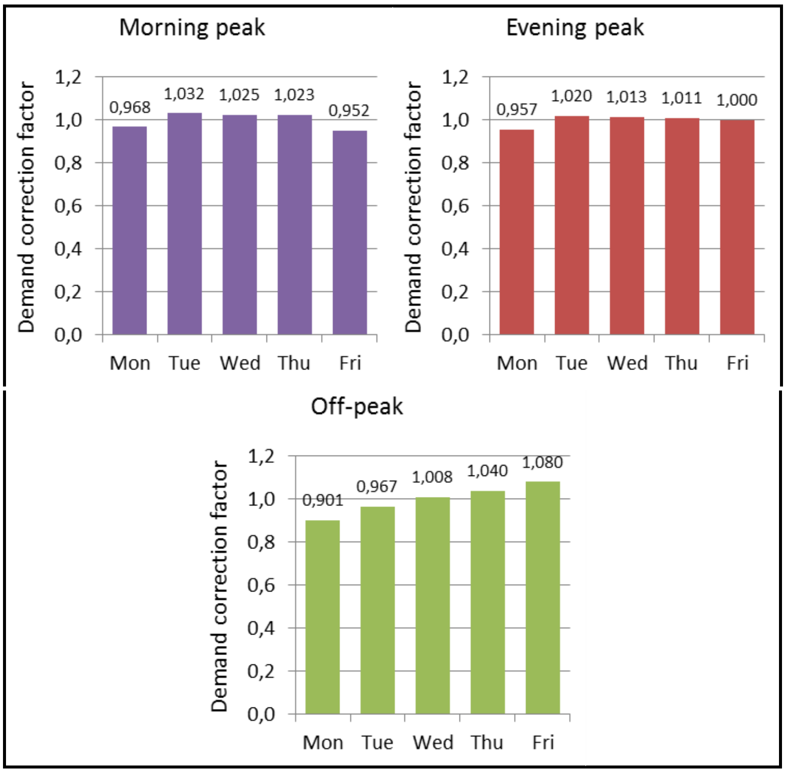
Procedure for selecting the day of the week:

In the series of simulation runs, the different days of the week are simulated one after the other. That is, every first simulation is considered to be a Monday, every second one a Tuesday, and so on. After the seventh simulation (a Sunday) the cycle restarts, meaning that the eighth simulation is considered to be a Monday again. This approach ensures that each day of the week is simulated equally often, which improves the statistical accuracy/reliability of the final results (as compared with a situation in which the day of the week is randomly drawn).[74]

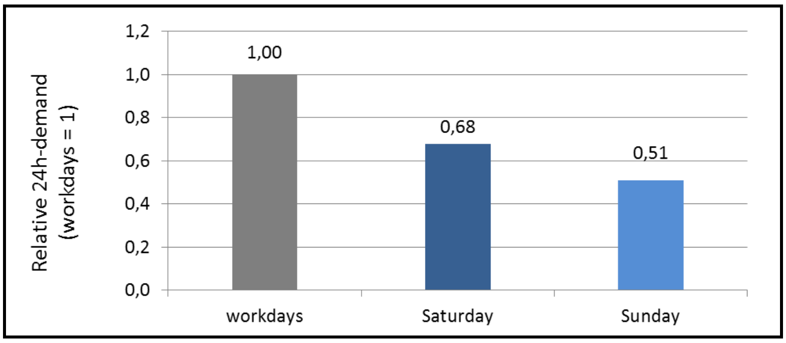
[74] Note that if multiple sources of variation were to be dealt with in this deterministic way, there would be a risk of introducing artificial dependencies between these, which could cause a bias in the final results. Since the day of the week is the only source of variation which is treated in such a way, this will not pose any problems, however.

Modeling of the effects:

After the selection of a day of the week, the peak and off-peak traffic demand levels are adjusted with the day-dependent correction factors shown below. These correction factors are based on the values indicated Table 2.5. A shortcoming of these data is however that one *combined* correction factor is given for the peak periods of the day, while especially for Fridays distinctly different correction factors should be used for the morning peak demand on the one hand and the evening peak demand on the other. Therefore, assumptions have been made with respect to these differences.

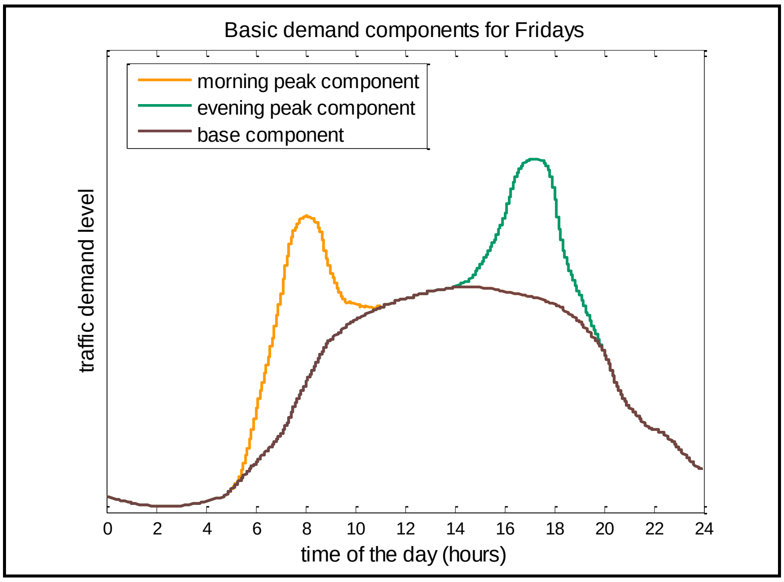


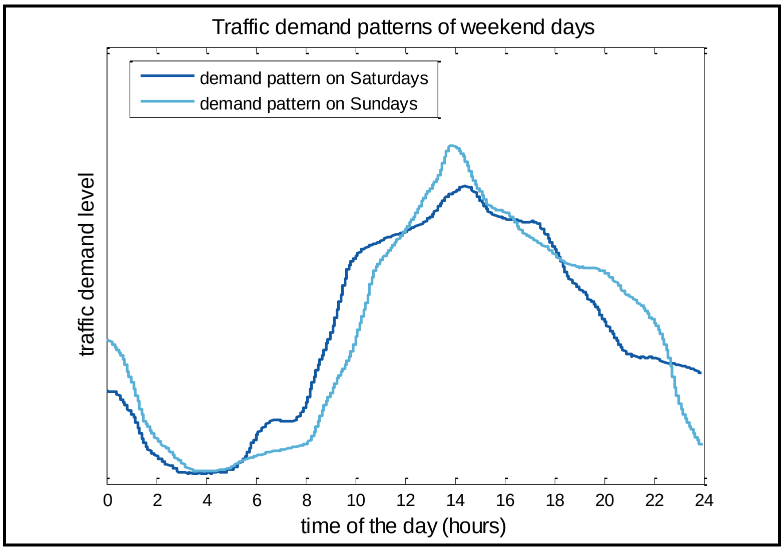
Literature does not provide information on the relative traffic demands on weekend days (as compared with the traffic demands on workdays). Based on some very limited empirical data from a motorway in the urban agglomeration of Western Holland, the following values have been assumed:



Obviously, not only the traffic demand *levels* have to be adapted to the day of the week, but their *patterns* as well. This is implemented in the following way:

- for Monday to Thursday the same basic demand components are used as for the representative situation (see Figure 5.6);
- for Friday other basic demand components are used (depicted in the figure below), to account for the fact that Fridays have a different off-peak pattern (see chapter 2);
- for Saturday and Sunday other demand patterns are used (based on some empirical data), which do not have morning and evening peaks and are not composed of different basic components (see the second figure below).





5.4.3 Month of the year

Classification:

- Modeled as a **DISCRETE** influence factor,
- with a **STOCHASTICALLY GENERATED** occurrence,
- and a **NETWORK-WIDE** effect
- on the traffic **DEMANDS**.

Modeled interdependencies:

Table 5.3: Modeled dependencies between the month of the year and other sources of variability

Modeled dependencies on other sources of variability	Considered source of variability	Modeled influences on other sources of variability
None	Month of the year	Vacations Special days Events Weather conditions Darkness Road works

Procedure for randomly selecting the month of the year:

The month of the year is drawn from a discrete probability distribution function, which reflects the differences in the lengths of the months.

Modeling of the effects:

After the random selection of a month, the peak and off-peak traffic demand levels are adjusted with the month-dependent correction factors shown in Figure 2.7 (peak) and Figure 2.8 (off-peak). It is assumed that for the weekend day traffic demand the same correction factors can be used as for the off-peaks on workdays.

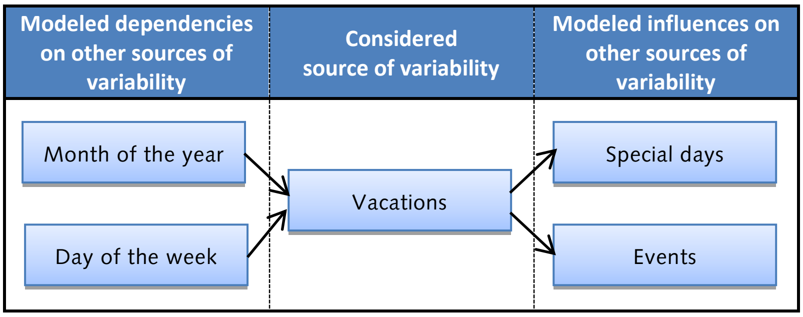
5.4.4 Vacations

Classification:

- Modeled as a **DISCRETE** influence factor,
- with a **STOCHASTICALLY GENERATED** occurrence,
- and a **NETWORK-WIDE** effect
- on the traffic **DEMANDS**.

Modeled interdependencies:

Table 5.4: Modeled dependencies between vacations and other sources of variability

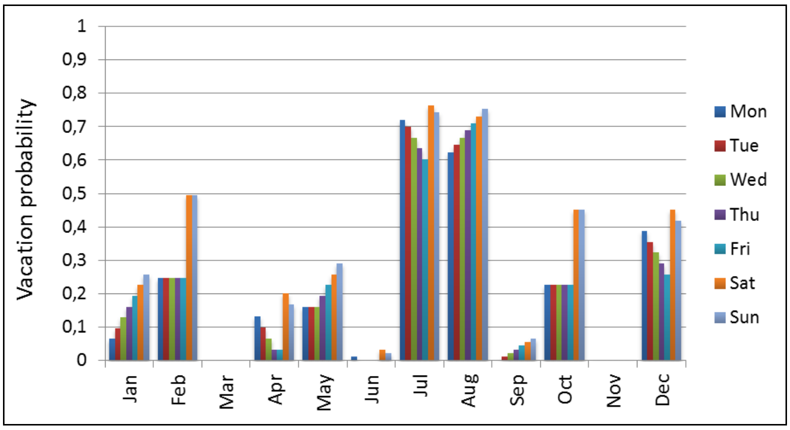


Procedure for randomly generating vacation

Whether or not the simulated day falls within a vacation period is randomly determined using a 'vacation probability table'. This table contains a separate vacation probability value for every single combination of month of the year and day of the week. In this table account has been made for:

- the cyclical shift in the vacation data (which for example results in the fact that January has one or two 'vacation Sundays' every year, but a 'vacation Monday' only twice in seven years); and
- the staggering of the summer vacation period.

A graphical representation of this table is shown below.



Modeling of the effects:

Depending on the outcome of the above, the peak and off-peak traffic demands are reduced or increased by the vacation/non-vacation correction factors given in Figure 2.7 (peak) and Figure 2.8 (off-peak). It is assumed that the weekend day traffic demands are not affected. Any possible differences *between* different vacation periods are ignored in the model.

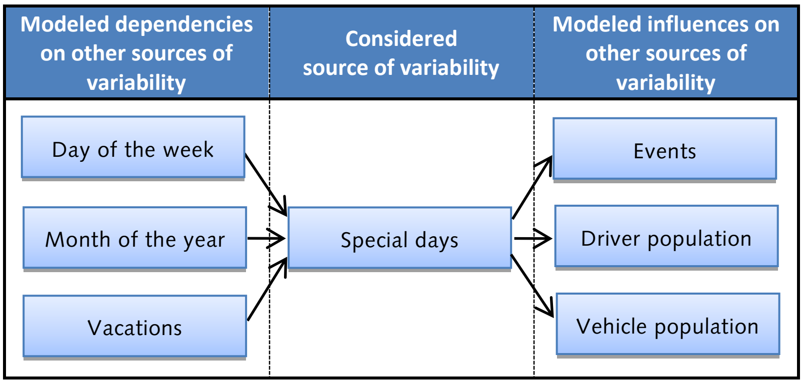
5.4.5 Special days (public holidays, long weekend days, vacation peaks)

Classification:

- Modeled as a **DISCRETE** influence factor,
- with a **STOCHASTICALLY GENERATED** occurrence,
- and a **NETWORK-WIDE** effect
- on the traffic **DEMANDS**.

Modeled interdependencies:

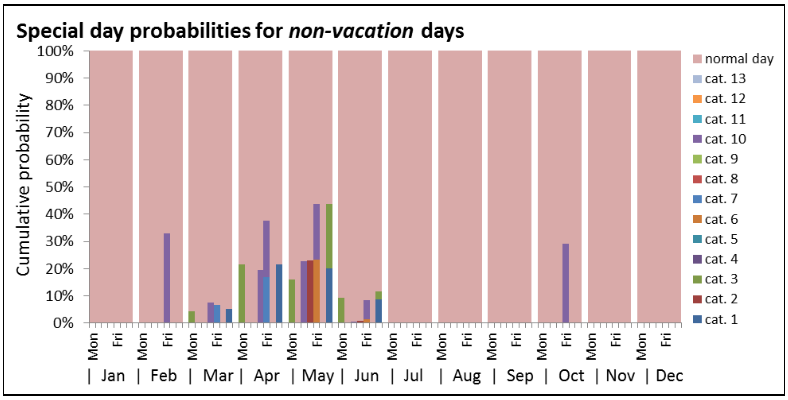
Table 5.5: Modeled dependencies between special days and other sources of variability

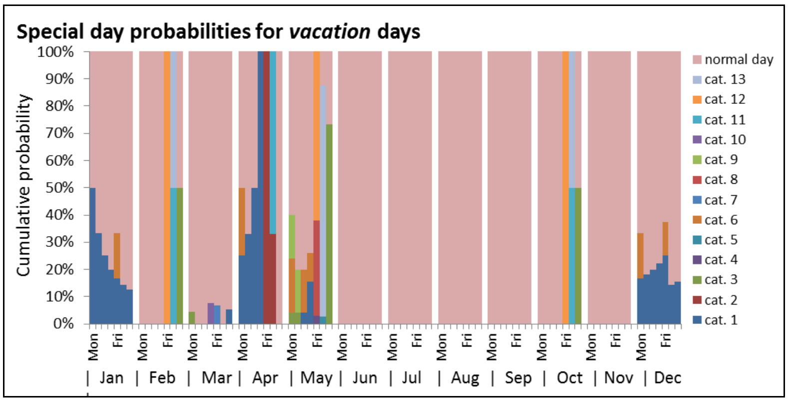


Procedure for simulating the occurrence of a special day:

Whether or not the simulated day is a special day of a certain type is randomly determined using a four-dimensional 'probability cube'. This probability cube defines the discrete cumulative probability distribution function of the type of special day (including the category '*no* special day'), conditional on the day of the week, the month of the year and the 'vacation situation'.

This is illustrated in the two figures below. In these figures, the first two dimensions of the probability cube (representing the month of the year and the day of the week) are on the horizontal axis. The third dimension (representing the 'vacation situation') is reflected in the distinction between the upper and lower figure. Finally, the fourth dimension (representing the different categories of special days) is indicated by using different colors.





As shown in the figures, 13 different categories of special days are distinguished. These are:

- Cat.1: Public holidays without a vacation peak in traffic
- Cat.2: Public holidays with a vacation *departure* peak in traffic
- Cat.3: Public holidays (Mon/Tue/Wed/Thu/Sun) or normal Sundays with a vacation *return* peak in traffic
- Cat.4: Public holidays (Fri) with a vacation *return* peak in traffic
- Cat.5: Public holidays (Sat) with a vacation *return* peak in traffic
- Cat.6: 'Semi-official' holidays / long weekend days without a vacation peak in traffic
- Cat.7: 'Semi-official' holidays (Fri) with a vacation *departure* peak
- Cat.8: 'Semi-official' holidays (Fri) / long weekend Fridays with a vacation *return* peak in traffic
- Cat.9: 'Semi-official' holidays (Mon/Tue) with a vacation *return* peak in traffic
- Cat.10: Normal workdays with a vacation *departure* peak in traffic
- Cat.11: Normal Saturdays with a vacation *departure* peak in traffic
- Cat.12: Normal Fridays with a vacation *return* peak in traffic
- Cat.13: Normal Saturdays with a vacation *return* peak in traffic

It should be noted here that the relative importance[75] of these different categories of special days, considered on an individual basis, will obviously be limited, because of their limited frequencies of occurrence. The distinction between these different categories is however a prerequisite for obtaining a reasonable aggregate influence. For this we cannot simply consider one 'average' category of special days, because of the

[75] in terms of their influence on the travel time statistics, or other congestion indicators.

fact that the effects of the different special days are very dissimilar.

Modeling of the effects:

In the model, the different categories of special days are dealt with in the following ways:

- Cat.1: Modeled as a Sunday (which means that both the demand *level* and the demand *pattern* are assumed equal to those of Sunday).
- Cat.2: Modeled as a Sunday, with a correction factor of 1.15 on the normal Sunday demand level (to account for departing vacation traffic).
- Cat.3: Modeled as a Sunday, with a correction factor of 1.20 on the normal Sunday demand level (to account for returning vacation traffic).
- Cat.4: Modeled as a Sunday, with a correction factor of 1.075 on the normal Sunday demand level (to account for returning vacation traffic).
- Cat.5: Modeled as a Sunday, with a correction factor of 1.10 on the normal Sunday demand level (to account for returning vacation traffic).
- Cat.6: Modeled as the average of the workday concerned and Saturday.
- Cat.7: Modeled as the average of Friday and Saturday, where the off-peak and evening peak demands of Friday are adjusted with correction factors of 1.10 and 1.15 respectively (to account for departing vacation traffic).
- Cat.8: Modeled as the average of Friday and Saturday, where the demand level of Friday is adjusted with a correction factor of 1.05 and the demand level of Saturday with a correction factor of 1.075 (to account for returning vacation traffic).
- Cat.9: Modeled as the average of the workday concerned (Monday/Tuesday) and Saturday, where the demand level of Saturday is adjusted with a correction factor of 1.20 (to account for returning vacation traffic).
- Cat.10: Modeled as having the demand *pattern* of a Friday, while keeping its original demand *level* (i.e. not necessarily that of a Friday), although the off-peak and evening peak demand levels are adjusted with correction factors of 1.10 and 1.15 respectively (to account for departing vacation traffic).
- Cat.11: Modeled with a correction factor of 1.175 on the demand level (to account for departing vacation traffic).
- Cat.12: Modeled with a correction factor of 1.05 on the demand level (to account for returning vacation traffic).
- Cat.13: Modeled with a correction factor of 1.10 on the demand level (to account for returning vacation traffic).

5.4.6 Weather conditions

Classification:

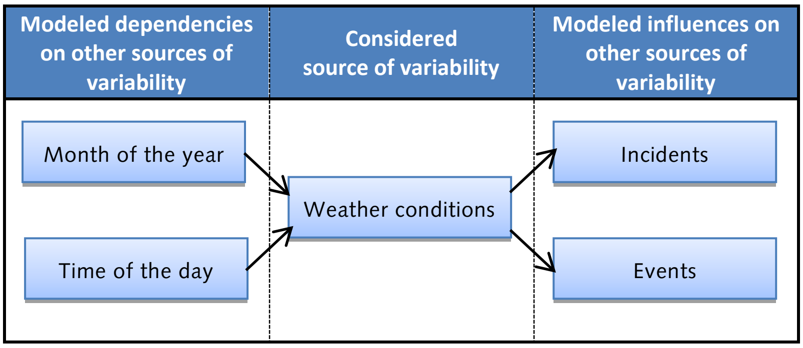
- Modeled as a **DISCRETE** influence factor,
- with a **STOCHASTIC** way of occurrence,
- and **NETWORK-WIDE** effects
- on both the traffic **DEMAND** and **SUPPLY** characteristics.

Stochastically generated variables:

- **PRESENCE OF (SIGNIFICANT) PRECIPITATION** per time-interval of the day
  (yes / no)
- **TYPE OF PRECIPITATION** (one for all precipitation of that day)
  (moderate snow / heavy snow / black ice / moderate rain / heavy rain / fog)
- **SUMMERY WEATHER CONDITIONS** on the simulated day
  (yes / no)

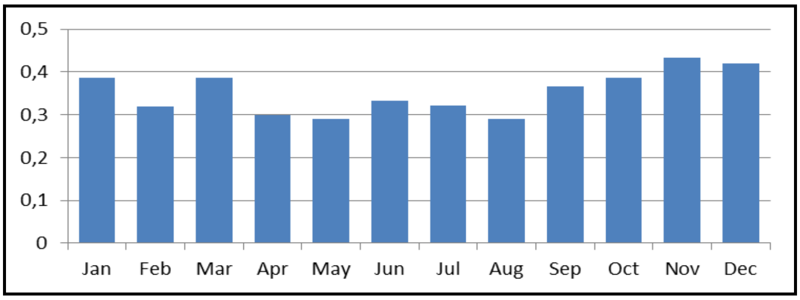
Modeled interdependencies:

Table 5.6: Modeled dependencies between weather conditions and other sources of variability

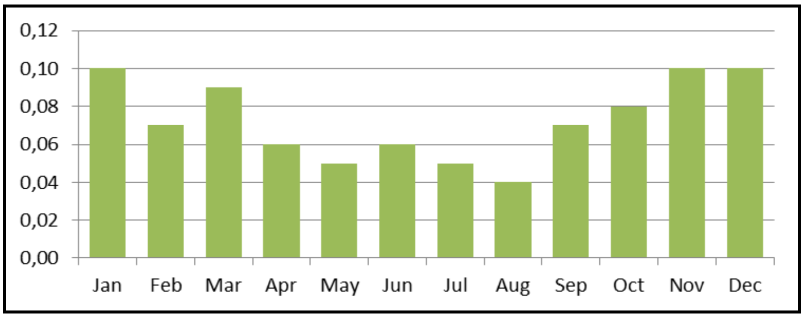


Probability input data:

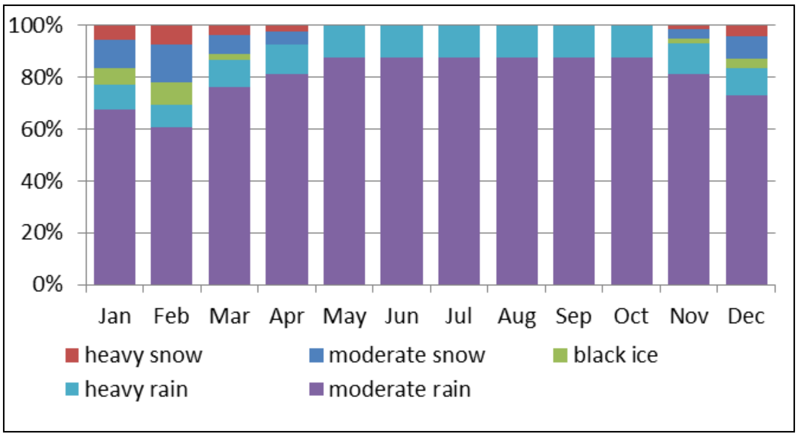
- Monthly frequencies of days with significant precipitation (>1mm):



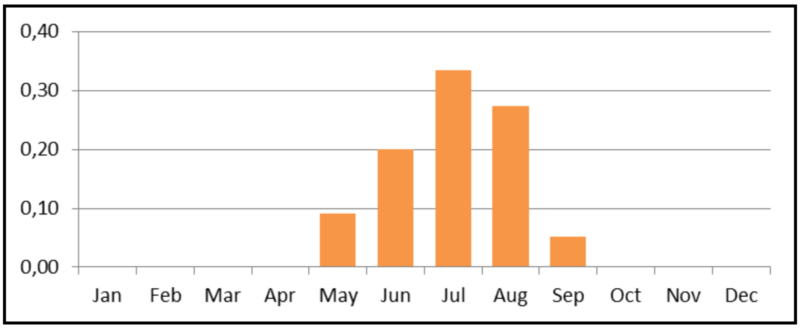
- Monthly percentages of time with precipitation:



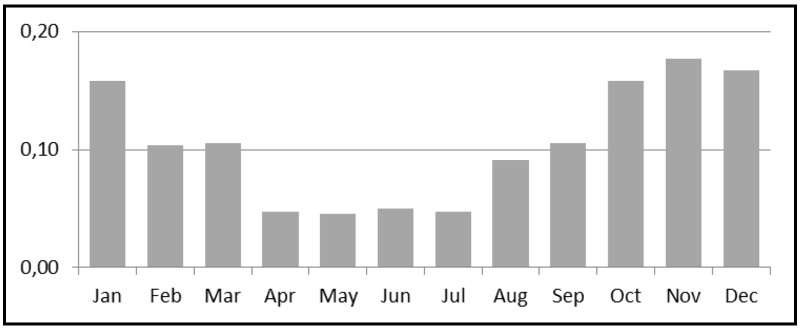
- Assumed correction factor for converting the percentages of time with precipitation to percentages of time with *significant* precipitation, for days with more than 1 mm: 0.05/0.0725.
- Monthly distributions of precipitation type:



- Monthly frequencies of summery days, relative to the total amount of day *without* significant precipitation:

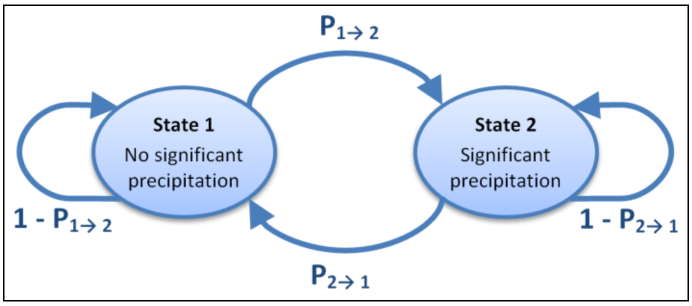


- Montly frequencies of days with fog, relative to the total amount of day *without* significant precipitation:



Procedure for simulating the occurrence of specific weather conditions:

- STEP 1. PRESENCE OF SIGNIFICANT PRECIPITATION PER 5-MINUTE INTERVAL:
    - Modeled as a Markov Chain[76] with the following two possible states:
        - Significant precipitation
        - No significant precipitation



    - Month-dependent transition probabilities ($P_{1\to 2}$ and $P_{2\to 1}$), calculated in such a way that:
        - the resulting monthly frequencies of days with significant precipitation are equal to those given above, and
        - the resulting monthly percentages of time with significant precipitation are equal to those given above (multiplied with the correction factor that was also given above).

- STEP 2. PRECIPITATION TYPE:
    - Random selection from the month-dependent discrete probability distribution given above.
    - The selected precipitation type is assumed to apply to *all* precipitation on the simulated day.

- STEP 3. SUMMERY WEATHER CONDITIONS:

  If significant precipitation was generated for one or more of the time intervals of the simulated day, this day is assumed not to be a summery one. If no significant precipitation was generated for the simulated day, the monthly probabilities given above are used to randomly determine whether it is a summery day or not.

- STEP 4. FOG
    - If significant precipitation was generated for one or more time intervals of the simulated day, it is assumed that no fog occurs on this day. If no significant precipitation was generated for the simulated day, the monthly probabilities given above are used to randomly determine whether or not fog occurs on this day.
    - If fog is simulated to take place, it is assumed to occur in one unbroken period, starting sometime during the night

[76] In order to account for the temporal coherence in precipitation: precipitation in time interval *i* is more likely to occur if precipitation occurred in time interval *i-1*, than if no precipitation occurred in time interval *i-1* (and the other way around).

and ending within a certain (month-dependent) period of time after sunrise.

- The start time of fog is randomly drawn from the following probability density function:

$$f(t_s) = \frac{1}{4 * d_{night}} + \frac{3 \cdot {t_s}^3}{{d_{night}}^4}$$

where t_s ($0 \leq t_s \leq d_{night}$) is the time at which the fog starts, relative to the time of sunset, and d_{night} is the (month-dependent) length of the night (i.e. the time between sunset and sunrise).

- The end time of fog is randomly drawn from the following probability density function:

$$g(t_e) = \frac{4}{{t_{e_{max}}}^4} \cdot (t_{e_{max}} - t_e)^3$$

where t_e ($0 \leq t_e \leq t_{e\_max}$) is the time at which the fog ends, relative to the time of sunrise, and t_{e_max} is its (month-dependent) maximum value.

Modeling of the effects:

- **EFFECTS ON THE TRAFFIC DEMANDS:**
  - In case of rain for one hour or more between 8 and 19h, the off-peak workday traffic demand is reduced by 2%, and the 24h weekend demand by 5%. The peak demands are not reduced in case of rain.
  - Demand reductions for other adverse weather conditions are indicated in the table below.

Table 5.7: Modeled demand effects of adverse weather conditions (other than rain)

Type of weather	Time window	Duration	Demand effect
Moderate snow	5:30 – 8:30	≥ 1 interval	Morning peak -5% Evening peak -2.5%
	15:00 – 19:00	≥ 1 interval	Morning peak -2% Evening peak -2.5%
	8:00 – 19:00	≥ 1 h.	Off-peak -7% Weekend -18%
Heavy snow	5:00 – 8:30	≥ 1 interval	Morning peak -20% Evening peak -15%
	15:00 – 19:00	≥ 1 interval	Morning peak -5% Evening peak -8%
	8:00 – 19:00	≥ 1 h.	Off-peak -30% Weekend -50%
Black ice	6:00 – 8:30	≥ 1 interval	Morning peak -4% Evening peak -1%
	15:00 – 19:00	≥ 1 interval	Morning peak -0.5% Evening peak -1%
	8:00 – 19:00	≥ 1 h.	Off-peak -5% Weekend -15%

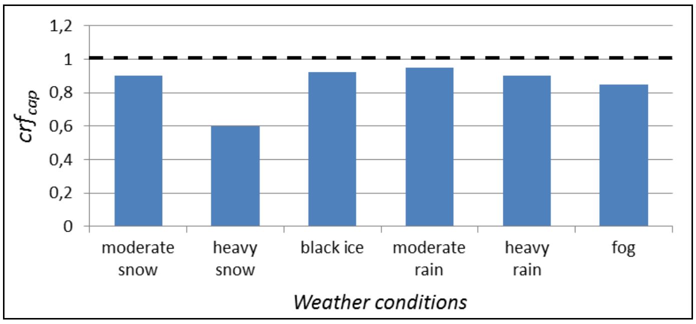
  - In case of a summery day, the off-peak workday traffic demand is increased by 1.5% and the 24 h weekend demand by 5%. The peak demands are not adjusted for summery days.

- **Effects on the traffic supply characteristics:**
  - The *free speeds* in adverse weather are calculated as: $v_{Free} = v_{Free\_basic} + f_{weather} \cdot (v_{Free\_nom} - v_{Free\_basic})$, where v_{Free} is the free speed in adverse weather, v_{Free_nom} is the nominal free speed value (i.e. in favorable weather), v_{Free_basic} is a certain weather-dependent reference level, and $f_{weather}$ is a weather-dependent factor expressing which part of the difference between v_{Free_nom} and v_{Free_basic} remains under the adverse weather conditions. Values for v_{Free_basic} and $f_{weather}$ are indicated in the table below.

Table 5.8: Values of the parameters v_{Free_basic} and $f_{weather}$, which are used for the modeling of the effects of adverse weather conditions on the free speeds.

Type of weather	v_{Free_basic} (km/h)	$f_{weather}$
Moderate snow	70	0.75
Heavy snow	50	0.15
Black ice	70	0.80
Moderate rain	80	0.88
Heavy rain	60	0.60
Fog	60	0.70

  - In case of heavy snow, the *number of available lanes* is assumed to be reduced by half.
  - The *capacities* are adjusted with the following correction factors:



    In the absence of data regarding any possible differences in the effects on the *free flow capacity* on the one hand, and the *queue discharge rate* on the other, no distinction is made between these.
  - In line with (Hranac et al, 2006), it is assumed that the largest part of the weather effect on the free flow capacity corresponds to an effect on the *critical speed* (rather than on the *critical density*). Therefore, a division parameter $z_{weather}$ (see subsection 5.3.3) of 0.8 is used.

5.4.7 Low sun

The possible influences of low sun (see chapter 2) are not taken into account in the model. Too little is known about its effects. Furthermore, there will be a strong dependency in these effects on the orientation of the road, and the season-dependent positioning of the sun. These aspects cannot easily be incorporated into a model.

5.4.8 Darkness

Classification:

- Modeled as a **DISCRETE** influence factor,
- with a **DETERMINISTIC** way of occurrence,
- and **NETWORK-WIDE** effects
- on the traffic **SUPPLY** characteristics.

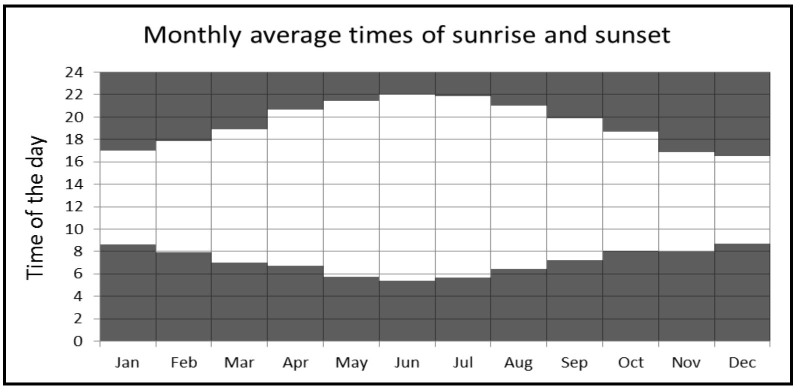
Modeled interdependencies:

Table 5.9: Modeled dependencies between darkness and other sources of variability

Modeled dependencies on other sources of variability	Considered source of variability	Modeled influences on other sources of variability
Time of the day Month of the year	Darkness	None

Procedure for simulating the occurrence of darkness:

For all time intervals before the monthly average time of sunrise and after the monthly average time of sunset, conditions are assumed to be dark.



Modeling of the effects:

- Any possible effects on the *free speeds* are neglected.
- For dark conditions the *capacities* are adjusted with a correction factor of 0.985 (corresponding to a 1.5% reduction). In the absence of data regarding any possible difference in the effect on the *free flow capacity* on the one hand, and the *queue discharge rate* on the other, no distinction is made between these.
- It is assumed that the effect on the free flow capacity is distributed over the *critical speed* and the *critical density* in the same way as the effects of adverse weather conditions. Therefore, a division parameter $z_{darkness} = z_{weather} = 0.8$ is used.

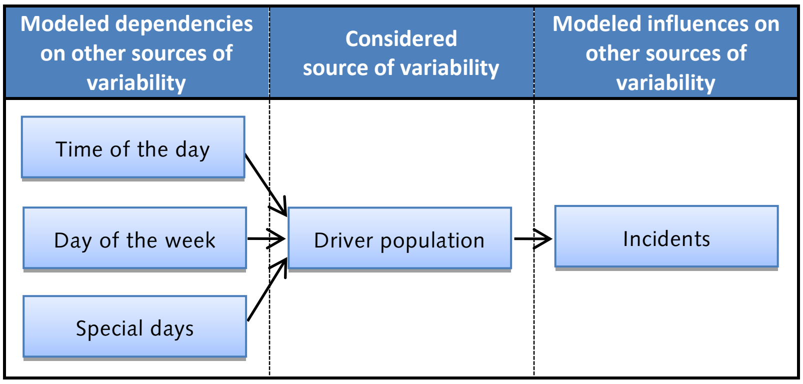
5.4.9 Driver population

Classification:

- Modeled as a **DISCRETE** influence factor,
- with a **DETERMINISTIC** way of occurrence,
- and a **NETWORK-WIDE** effect
- on the traffic **SUPPLY CHARACTERISTICS**.

Modeled interdependencies:

Table 5.10: Modeled dependencies between the composition of the driver population and other sources of variability



Procedure for simulating the variation in the driver population:

- Four typical driver populations are distinguished:
  - *Peak driver population* (largely commuter traffic)
  - *Off-peak driver population* (a mix of mainly: social traffic, leisure traffic, commercial traffic and commuter traffic)
  - *Saturday driver population* (largely social and leisure traffic)
  - *Sunday driver population* (almost exclusively social and leisure traffic)

  Here the Saturday driver population is assumed to be more experienced (i.e. better acquainted with the local traffic situations) than the Sunday driver population. This assumption is based on the supposition that people's activity patterns are more regular on Saturdays than on Sundays, owing to some typical recurrent Saturday activities as shopping and sporting. On Sundays probably a larger part of the trips is of an occasional nature.
- Depending on the simulated day of the week and the time of the day, one of these different driver populations is selected, as indicated in the table below:

Table 5.11: Selection of a driver population as a function of the day of the week and the time of the day

Day of the week	Time window	Type of driver population
weekday	6:30 – 9:30	Peak drivers population
	15:30 – 18:30	Off-peak drivers population
Saturday	whole day	Saturday drivers population
Sunday	whole day	Sunday drivers population

- If a *special* day is simulated, the type of driver population might have to be corrected, depending on the type of special day. These corrections are as follows:

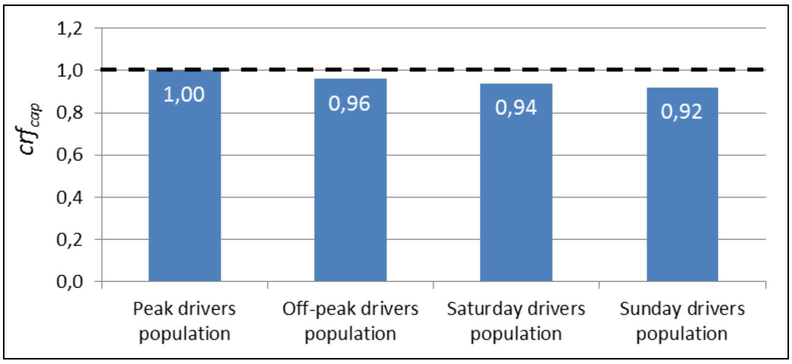
Table 5.12: Corrections of the selected driver population for certain types of special days

Type of special day	Time window	Type of driver population
1 – 5	whole day	Sunday drivers population
6 – 9	whole day	Off-peak drivers population
10	15:30 – 18:30	Off-peak drivers population
12	whole day	Off-peak drivers population

- Driver population differences between vacation and non-vacation periods are assumed to be negligible.

Modeling of the effects:

- Any possible effects on the *free speeds* are neglected.
- The *capacities* are adjusted with the following correction factors:



The peak drivers population is associated with the highest capacity values. Such drivers typically are very experienced, probably resulting in a more efficient driving behavior. In the absence of information regarding any possible difference in the effect on the *free flow capacity* on the one hand, and the *queue discharge rate* on the other, no distinction is made between these.

Note that a correction factor with value *one* is used for the peak drivers population. This actually means that this population is considered the 'representative one'. The reasons for this are that capacities are mainly measured during peak periods, and that models (or network files) are usually calibrated on a peak period. As a result, the default capacities of any model or network file are likely to represent the values for the peak drivers population.

- In the absence of any information on this, it is assumed that the effect on the free flow capacity is equally distributed over the *critical speed* and the *critical density*. Therefore, a division parameter $z_{drivers}$ = 0.5 is used. Schedule

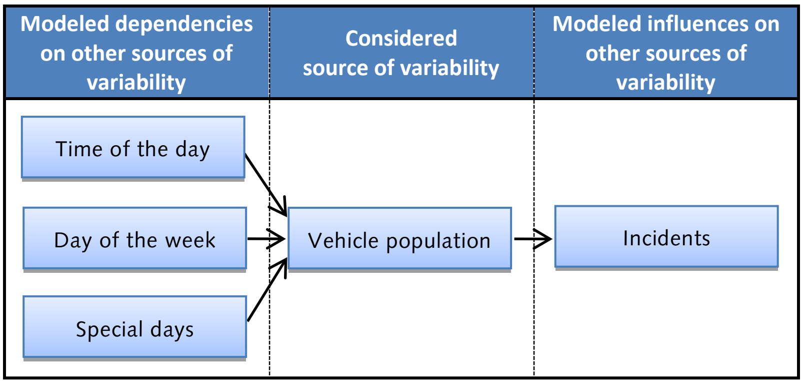
5.4.10 Vehicle population

Classification:

- Modeled as a **CONTINUOUS** influence factor,
- with a **DETERMINISTIC** way of occurrence,
- and a **NETWORK-WIDE** effect
- on the traffic **SUPPLY CHARACTERISTICS**.

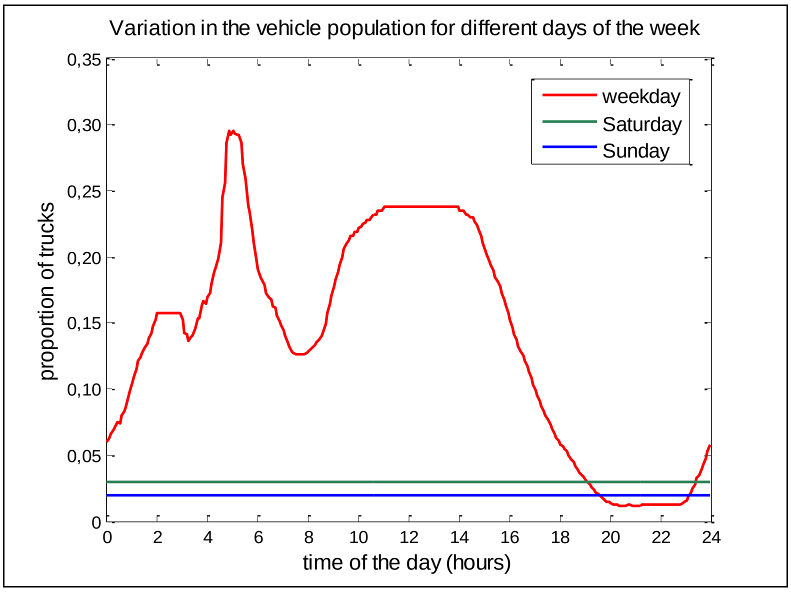
Table 5.13: Modeled dependencies between the composition of the vehicle population and other sources of variability

Modeled interdependencies:



Procedure for simulating the variation in the vehicle population:

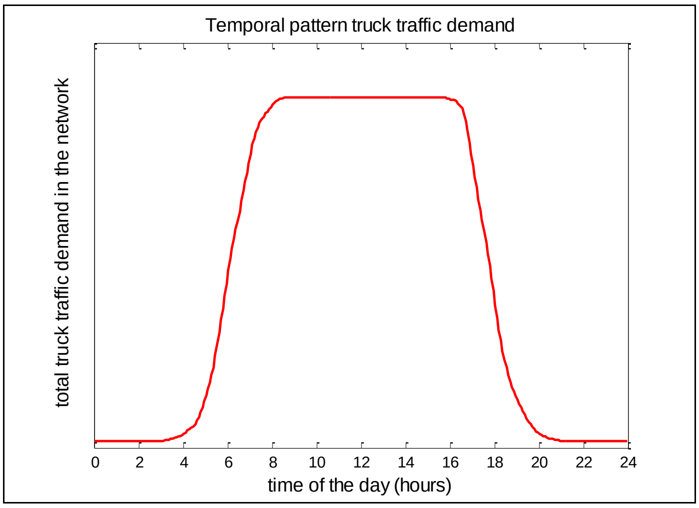
In fact, the only appropriate way to model the influence of variations in the vehicle population would be to use a dynamic traffic simulator which explicitly considers passenger car and freight traffic as two individual, interacting traffic flows. JDSMART (i.e. the traffic simulator used here) considers only one type of vehicles, however[77]. Therefore, the influence of the variability in the vehicle population had to be modeled in another (less optimal) way. In this alternative approach the vehicle composition is assumed to be uniform across the entire network, and to vary over time according to some predetermined pattern. Depending on the simulated day of the week, one out of three different patterns is chosen. These different patterns are shown in the figure below.



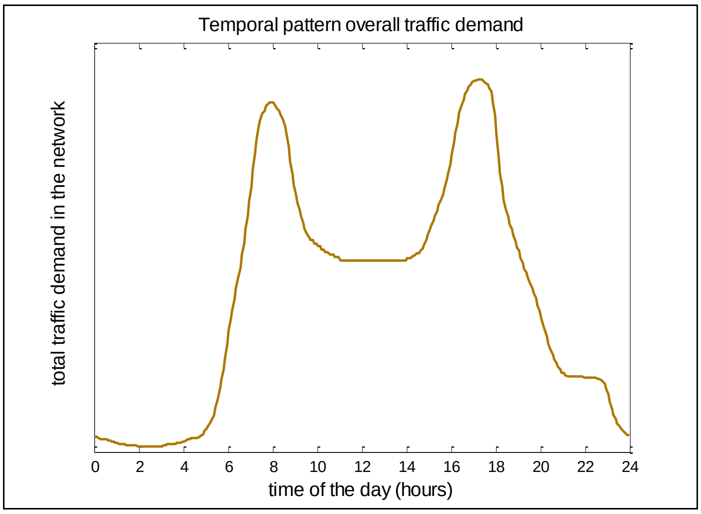
[77] Note that extending JDSMART to a multi-vehicle-class model in fact would not be that difficult, since it is already a multi-user-class model. In the current version these different user classes are however not related to different *vehicle types*, but to the different *destinations* of the vehicles involved. Consequently, the only adjustment required would be to extend this concept of considering multiple user classes to the vehicle type. This was however not possible within the scope of this project.

As shown in the figure, for Saturdays and Sundays constant truck shares of 2 and 3 percent are assumed. For weekdays a (much higher) time-varying truck percentage is used. The pattern of this variation has been obtained in the following three steps:

- For the total *truck* traffic demand in the network, the following temporal pattern was assumed:



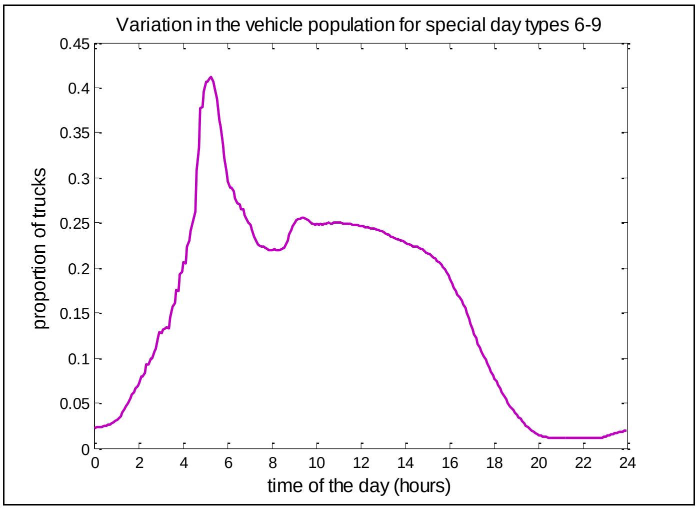
- Based on some empirical data, a 24-hours truck fraction of 14.5% was assumed.
- By combining these two with the (representative) temporal pattern of the *overall* traffic demand in the network (depicted in the figure below), the temporal pattern of the truck fraction is found.



Note that due to this modeling approach, the short-term local variations in traffic composition are not taken into account. Since these variations are actually observed to be considerable, this is not a very desirable situation. It is estimated however that it will not be too detrimental to the overall model results, since the extent to which the variance of free flow capacities is attributable to variations in truck percentage is empirically found to be relatively small (Geistefeldt, 2009).

For certain types of special days the temporal pattern of the vehicle population is adapted:

- For categories 1-5 the pattern is assumed to be equal to that of Sundays.
- For categories 6-9 the *truck* traffic demand is assumed to be unaffected, while the *total* traffic demand is reduced (see section 5.4.5). This results in the following modified pattern:



The effects of vacations and any possible variations with the months of the year are not considered.

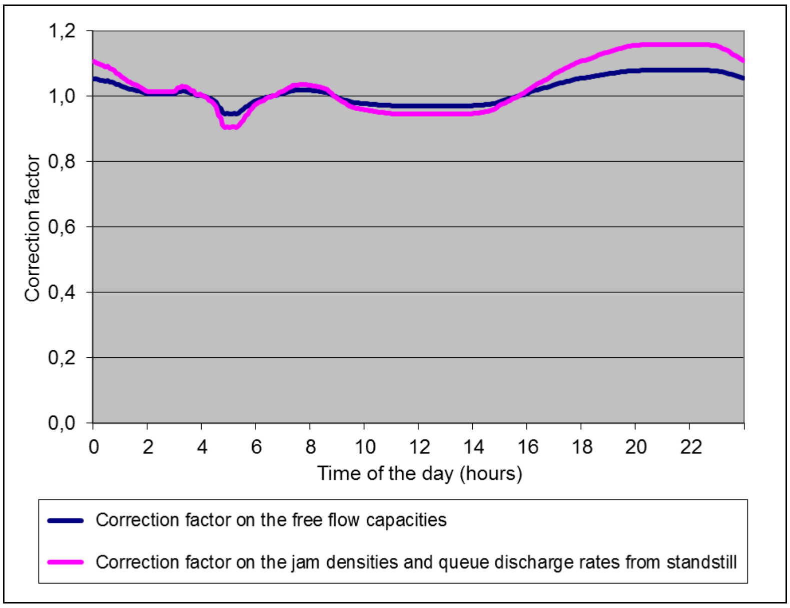
Modeling of the effects:

- Any possible effects on the *free speeds* are neglected.
- The effects on the capacities and jam densities are expressed in terms of a 'passenger car equivalency factor' (PCE). This factor indicates the number of passenger cars that one truck could be considered to be equivalent with (not only considering its physical dimensions, but also its operating capabilities). For the *free flow capacities* the commonly used PCE-value of 1.5 is applied. In congested conditions the impact of heavy vehicles turns out to be larger than in free flow conditions (see section 2.2.5). Therefore, for the *queue discharge rates (from standstill)* a higher PCE-value is used. For this a value of 2.0 is assumed. For the *jam densities* a PCE-value of 2.0 is used as well.
- Using these PCE-values, the correction factors on the capacities and jam densities can now be calculated as:

$$crf = \frac{truck\_frac_{rep} \cdot PCE + (1 - truck\_frac_{rep}) \cdot 1}{truck\_frac(t) \cdot PCE + (1 - truck\_frac(t)) \cdot 1}$$

In this equation $truck\_frac(t)$ represents the truck fraction simulated for time interval t (according to one of the truck fraction patterns given above). $truck\_frac_{rep}$ represents the *representative* truck fraction, corresponding to the default capacity values in the model or network file. Since these default capacity values typically have been measured and/or calibrated for peak periods, this representative truck fraction typically is associated with peak periods as well. Because of the fact that the test network used in this project was originally calibrated on a certain period around the morning peak, here the representative truck fraction is taken equal to the average truck fraction during that period, corresponding to a value of 17.2%.

An example of the resulting correction factors (for normal weekdays) is shown in the figure below:



- It seems reasonable to assume that the vehicles-effect on the free flow capacity is in its entirety associated with an effect on the *critical density* (corresponding to the *critical speed* being unaffected). Therefore, a division parameter $z_{vehicles}$ = 0 is used.

5.4.11 Events

Classification:

- Modeled as a **DISCRETE** influence factor,
- with a **STOCHASTIC** way of occurrence,
- and **ORIGIN-DESTINATION SPECIFIC** effects
- on traffic **DEMANDS**.

Modeled interdependencies:

Table 5.14: Modeled dependencies between events and other sources of variability

Modeled dependencies on other sources of variability	Considered source of variability	Modeled influences on other sources of variability
Time of the day Day of the week Month of the year Vacations Special days Weather conditions	Events	None

Input data:

The influence of events is not easily incorporated in a model. After all, there is no generally applicable frequency of occurrence and no generally applicable demand effect either. In real life widely varying frequencies and effects are found. This means that inevitably there will be a certain degree of arbitrariness in the modeling of events. This does not mean, however, that we might just as well use some assumed (i.e. fictitious) frequencies and effects. In this case there would be a very real risk that these frequencies and effects (or their combinations) would be out of their 'realistic range', resulting in unrealistic simulation results.

Because of this, we are more or less forced to use rather specific real-life event data, relating to the particular network considered in the model calculations, or at least to a comparable network. In view of the fact that in this project the motorway network around the Dutch city of Rotterdam was used as a test network (see chapter 8), here event data of the Rotterdam area have been considered. This area includes four of the 45 most traffic generating event locations in the Netherlands (Meeuwissen et al, 2004). These are:

- Ahoy (a multifunctional events accommodation),
- De Kuip (a football stadium),
- Blijdorp (a large zoo), and
- Rotterdam's city center.

For these four locations event frequency data have been obtained from a number of event calendars[78]. Data regarding their traffic generation have been derived from (Meeuwissen et al, 2004).

Procedure for simulating the occurrence of events:

- Events are randomly selected using the frequency data mentioned above. These frequency data have been translated into conditional event probabilities. The conditioning is with respect to:
  - the month of the year,
  - the day of the week,
  - the 'vacation situation', and
  - the special day category (including the category 'no special day')

Fifteen different categories of events are considered. These categories are different in:

- location (Blijdorp, De Kuip, Ahoy or Rotterdam Center),
- time of the day (evening or daytime),
- pattern of traffic generation (peaked or more evenly spread out over time), and/or
- type of event (corresponding to a certain volume of generated traffic).

An overview of all different categories of events is given in the table below:

Table 5.15: Overview of the different categories of events

Event category	Location	Time of day	Type of traffic generation	Type of event
1	Blijdorp	daytime	spread	busy day
2 a	De Kuip	daytm./even.	peaked	league match
b	De Kuip	evening	peaked	int. match / concert
3 a	Ahoy	evening	peaked	show / sporting event /...
b	Ahoy	daytime	peaked	show / sporting event /...
c	Ahoy	daytime	spread	exhibition / fair
4 a	City center	daytime	spread	very large weekend event
b	City center	evening	peaked	very large evening event
c	City center	daytime	spread	very large weekday event
d	City center	daytime	spread	large daytime event
e	City center	evening	peaked	large evening event
f	City center	daytime	peaked	congress
g	City center	evening	peaked	other evening events
h	City center	afternoon	peaked	other weekend events
i	City center	morning	peaked	other weekend events

The model takes into account that the events of categories 2a and 2b are mutually exclusive (because of sharing the same accommodation). The same applies to the events of categories 3b and 3c. All other combinations are possible, because of the events being separated in time and/or accommodation.

[78] www.ahoy.nl; www.dekuip.nl; www.blijdorp.nl; www.rotterdam.info; www.dedoelen.nl

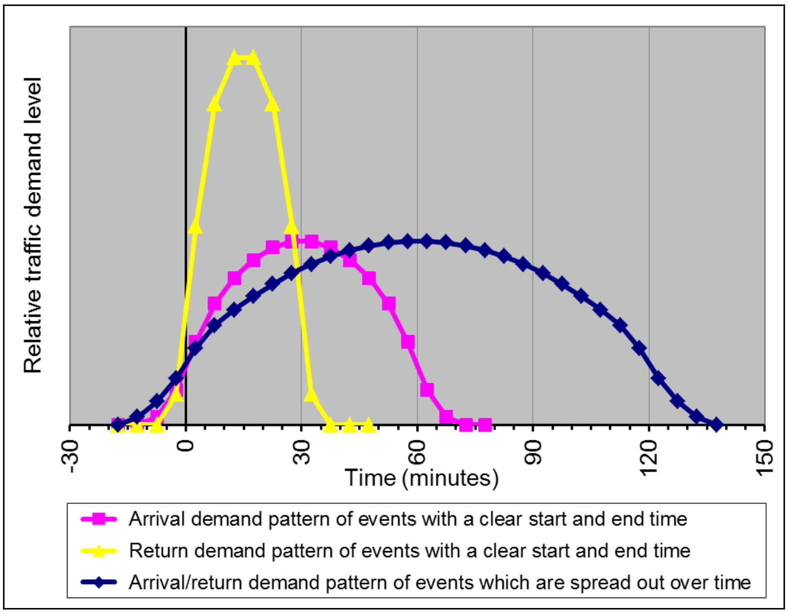
It should be noted that the relative importance[79] of the *individual* categories will obviously be limited, because of their limited frequencies of occurrence. The distinction between these different categories is however essential for obtaining a reasonable *aggregate* influence. For this we cannot simply consider one 'average' category of events, because of the fact that the effects of the different events are very dissimilar.

- After the random selection of events, it is checked whether the weather conditions are not too unfavorable for these events. In case of heavy snowfall, outdoor events are modeled to be cancelled.

Modeling of the effects:

- Every event category has the following attributes:
  - the volume of traffic generated (in veh/h),
  - the start time of the 'arrival flow',
  - the start time of the 'return flow', and
  - a set of factors expressing the relative shares of the event traffic that come from the external motorway origin/destination zones.

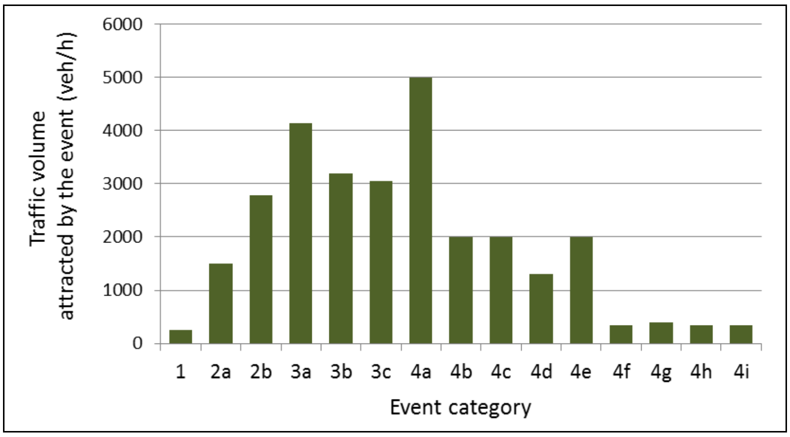
  Furthermore, each event category is associated with one or two of the following demand *patterns* (one for the 'arrival flow' and possibly another one for the 'return flow'):

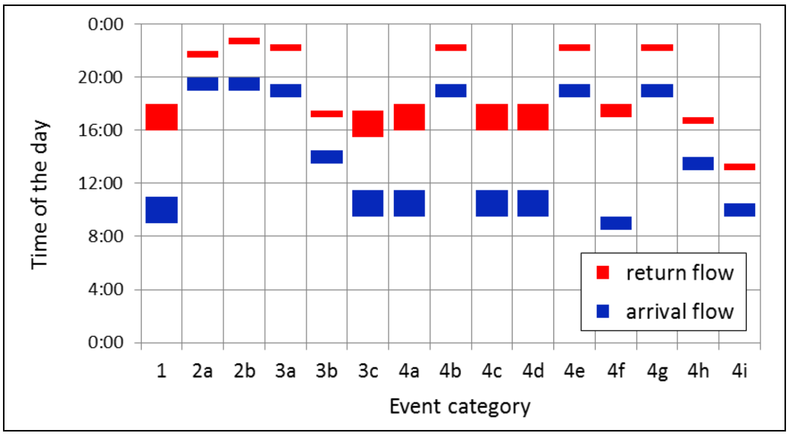


  Note that the *return* flow of events with well-defined start and end times is assumed to be more peaked than the corresponding *arrival* flow. Approximately 90% of the returning traffic leaves within 30 minutes time. For the arriving traffic, this 90% is distributed over a period which is twice as long (i.e. 60 minutes). For the events with less distinct start and end times (like exhibitions and fairs) it is assumed that 90% of the traffic occurs within two hours time.

[79] in terms of their influence on the travel time statistics, or other congestion indicators

The figures below show the traffic volumes associated with the different types of events, as well as their ‘scheduling’.

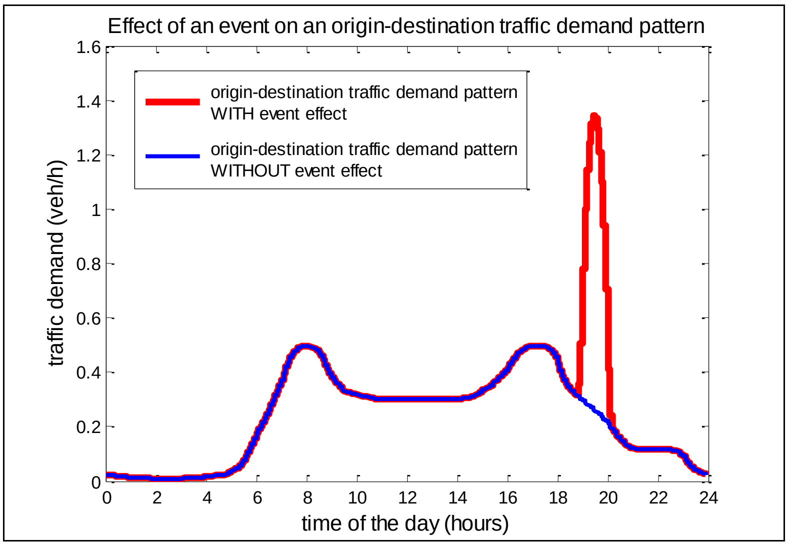




In case of adverse weather conditions, the amounts of event traffic are reduced. Here the reductions for *outdoor* events obviously are larger than those for *indoor* events. In case of rainy weather, the traffic demands for indoor events are even not adjusted at all, or corrected in the opposite direction (i.e. *in*creased, instead of *de*creased). For days with summery weather conditions, the traffic demands generated by outdoor events are assumed to be larger.

An important aspect in the modeling of large-scale events is that the associated traffic flows cannot simply be distributed over the origins/destinations in proportion to the normal origin-destination traffic demands. This would lead to a relatively large part of the event traffic having its origin in nearby zones. For large-scale events this is clearly not realistic. For such events a relatively large part of the attracted traffic comes from other regions, or even from other parts of the country. This is accounted for in the model by explicitly specifying the shares of the event traffic coming from the external motorway zones. Only the remaining traffic is then distributed over the (remaining) zones in proportion to the normal demands.

- By combining all elements mentioned above, for every event its demand effect per origin/destination can be evaluated. These demand effects can then be superimposed on the origin-destination demand patterns. An example of an origin-destination demand pattern after the superposition of an event effect is shown below.



5.4.12 Incidents

Classification:

- Modeled as a **DISCRETE** influence factor,
- with a **STOCHASTIC** way of occurrence,
- and **LOCAL** effects
- on the traffic **SUPPLY** characteristics.

Stochastically generated variables:

- **OCCURRENCE OF ACCIDENTS** per cell of the network and time-interval of the day
- **OCCURRENCE OF VEHICLE BREAKDOWNS** per cell of the network and time-interval of the day
- **DURATION OF INCIDENTS**
- **NUMBER OF BLOCKED LANES**

Modeled interdependencies:

Table 5.16: Modeled dependencies between incidents and other sources of variability

Modeled dependencies on other sources of variability	Considered source of variability	Modeled influences on other sources of variability
Weather conditions Driver population Vehicle population Local amount of traffic All sources of variability	Events	None

Probability input data:

- In the model only the two main types of incidents are considered: vehicle breakdowns and accidents.
- Based on literature, the following rates of occurrence are assumed for these two categories of incidents:
    - 1.5E-6 vehicle breakdowns per vehicle-kilometer
    - 0.5E-6 accidents per vehicle-kilometer[80]
- These average rates of occurrence are corrected for the:
    - *driver population* (accident rate only):

Type of driver population	Correction factor on accident rate
Peak drivers population	0.93
Off-peak drivers population	1.02
Saturday drivers population	1.10
Sunday drivers population	1.10

    - *vehicle population* (both incident rates):

Type of vehicle	Correction factor accident rate	Correction factor breakdown rate
Passenger car	0.98	0.98
Truck	1.17	1.17

    - *weather conditions* (accident rate only):

Weather conditions	Correction factor on accident rate
No adverse weather	0.95
Snow	0.95[81]
Black ice	2.5
Fog	1.9

No corrections are made for:

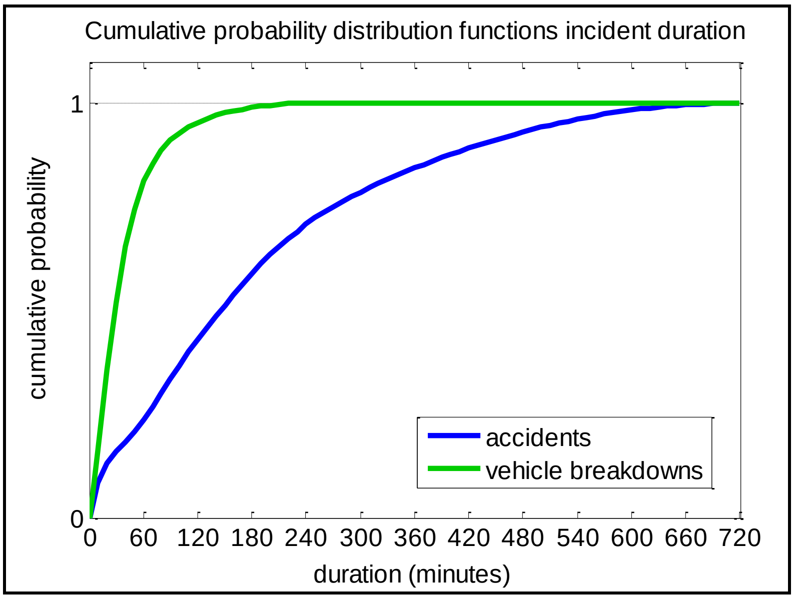
- *road works* (too much uncertainty with regard to their influence)
- *darkness* (visibility is worse in darkness, but on the other hand vehicles are better distinguishable due to their lights)
- *low sun* (too complicated to be accounted for, due to the dependencies on the orientation of the road, the season-dependent positioning of the sun, and the cloud coverage)
- *traffic conditions* [82] (too little known on the quantitative relations between traffic conditions and accident rate (and vehicle breakdown rate) for Dutch circumstances)

[80] comprising both accidents involving injuries and accidents involving property damage only

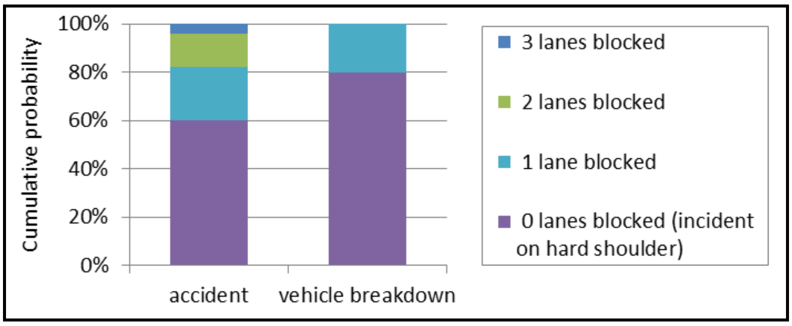
[81] In practice, both increases and decreases of accident rates are observed for snowy weather. On the one hand, visibility and car controllability are reduced (which *increases* the accident rate), but on the other hand road users drive more carefully (which *decreases* the accident rate).

[82] Note that the dependency on traffic conditions is partly accounted for already by expressing the rates of occurrence in terms of the number of occurrences *per vehicle-kilometer*.

- Based on the empirical distribution depicted in Figure 2.20 and assumed differences between accidents on the one hand and vehicle breakdowns on the other, the following cumulative probability distribution functions have been assumed for the incident duration:



- Based on the data provided in Figure 2.21 and some further assumptions, the following cumulative probability distributions have been assumed for the incident severity (expressed in the number of lanes that is blocked):



The interdependency between incident duration and incident severity (expressed in the number of lanes that it blocks) is not accounted for in the model.

Procedure for simulating the occurrence of incidents:

- In order to account for the fact that incidents are more likely to occur when more traffic is present, the occurrence of incidents is simulated in a traffic-dependent way. This means that incidents are not generated *prior to* the simulation of the traffic operations (as all other sources of variability), but simultaneously with this simulation.

- After each simulated 5-minute interval it is randomly determined for each of the cells whether or not an accident or a breakdown event occurred in that particular time-interval. For this the model keeps track of the amounts of vehicle-kilometers that are traveled within the different cells. Assuming independency in the accident occurrence between the individual vehicle-kilometers, the probability of occurrence of an accident is then calculated as:

$$P_{acc}(i,t) = 1 - (1 - p_{acc})^{d(i,t)}$$

In this formula, $P_{acc}(i,t)$ is the probability of the occurrence of an accident in cell *i* within time-interval *t*; *d(i,t)* is the amount of vehicle-kilometers traveled in cell *i* in time-interval *t*; and p_{acc} is the accident rate per vehicle kilometer, for which a value was given above.

Similarly, the probablility of occurrence of a vehicle breakdown event is calculated as:

$$P_{brk}(i,t) = 1 - (1 - p_{brk})^{d(i,t)}$$

By confronting these probabilities with random numbers between zero and one, incidents can randomly be generated then.

- For every newly generated incident the procedure is now as follows:
    - The duration of the incident is randomly drawn from the cumulative probability distribution function of the duration given above. The end time of the incident can then be calculated as the summation of its start time (which equals the current time in the simulation) and duration.
    - The severity of the incident (i.e. the number of lanes that it blocks) is randomly drawn from the cumulative distribution function of the incident severity given above.
    - All data of the incident are stored in a 'dynamic incident table'. This table contains all incidents that are currently 'active'. After each time interval it is checked whether the table contains incidents that have 'expired'. As soon as an incident has expired, it is removed.
    - In order to account for the effect that an incident diverts the attention of the drivers on the roadway *in the opposite direction* (which significantly reduces the capacity over there: see section 2.2.5) for that roadway an (artificial) incident is generated as well. Obviously, the duration of this artificial incident is equal to that of the factual incident where it belongs to.

Modeling of the effects:

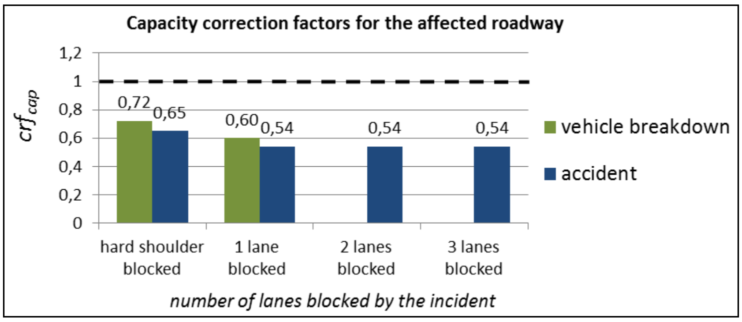
Prior to each simulated 5-minute interval, it is checked whether the 'dynamic incident table' contains any incidents. If that is indeed the case, for all incident cells the supply characteristics are adjusted in the following ways:

- For incidents on the hard shoulder, the speed limit (and thus the *free speed*) is assumed to be reduced to 90 km/h.

For incidents blocking one or more lanes, a reduction to 70 km/h is assumed. For the 'artificial' incidents (representing the effects of incidents on the roadway in the opposite direction) no speed reduction is applied.

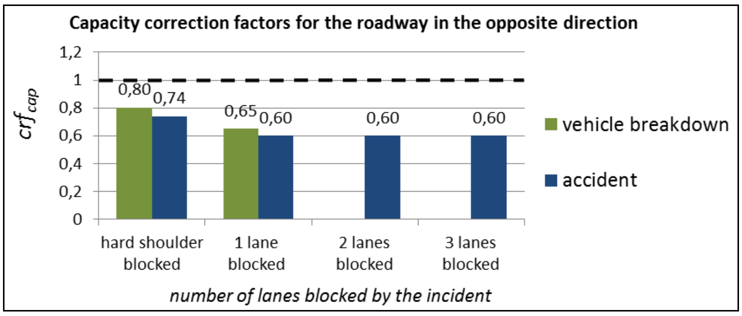
- The *number of available lanes* is reduced by the stochastically generated number of lanes that are blocked by the incident. Obviously no reductions are associated with the 'artificial' incidents.

- The *queue discharge rates (from standstill)* of the remaining lanes are adjusted with the following correction factors:



These assumed correction factors are based on the data of Knoop (2009) discussed in section 2.2.5. In the absence of data regarding the effects on *free flow capacities*, for these capacities the same correction factors are assumed to apply.

- For the 'artificial' incidents (representing the effects of incidents on the roadway in the opposite direction) the capacities are adjusted with the following correction factors (reflecting the effects of diverted attention):



These factors are roughly based on the factor of 0.69 found by Knoop (not differentating between the different types of incidents).

- In the absence of any information on this, it is assumed that the effect on the free flow capacity is equally distributed over the *critical speed* and the *critical density*. Therefore, a division parameter $z_{incident}$ = 0.5 is used.

5.4.13 Road works

Classification:

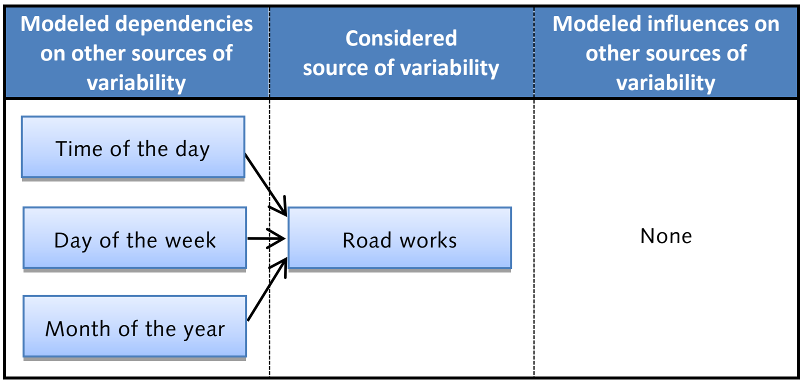
- Modeled as a DISCRETE influence factor,
- with a STOCHASTIC way of occurrence,
- and LOCAL effects
- on the traffic SUPPLY characteristics.

Stochastically generated variables:

- OCCURRENCE OF (SHORT-TERM) ROAD WORKS per link of the motorway network
- START TIME OF ROAD WORKS
- DURATION OF ROAD WORKS
- NUMBER OF CLOSED LANES

Modeled interdependencies:

Table 5.17: Modeled dependencies between road works and other sources of variability



General characteristics of the simulated road works:

- In the model only SHORT-TERM road works are simulated. Here 'short-term' is understood as lasting less than 24 hours. The exclusion of longer lasting road works can be explained as follows:
  - The longer the duration of road works, the larger their demand effects can be expected to be.
  - These demand effects cannot (easily) be included in the model, since they are likely to be rather context-specific (while the model would need to deal with them in an automated way).
  - Instead of making some crude, not well-founded assumptions on these demand effects (with an unclear impact on the realism of the overall model outcomes), it was considered a better approach to make the clear choice to exclude the longer lasting road works. For the remaining (i.e. short-term) road works it is assumed acceptable to leave any possible demand effects out of account.

- It is assumed that these short-term road works can be modeled as occurring randomly over the network. Obviously, for longer lasting road works (including large scale reconstructions) this would be an unrealistic assumption. For the short-term road works (including repair works and regular maintenance activities) it is considered acceptable, however.
- It seems reasonable to assume that work zones with a short *duration* generally are short in *physical length* as well. Because of this, it is considered acceptable to assume independency between the different links of the network. If road works are simulated to occur on a given link, these road works are supposed to extend from the start to the end of that link.
- The model uses an equal road works probability for all links of the network. This corresponds to the assumption that the frequency of road works is independent of both the link length and the number of lanes on the link, which is clearly a simplification of reality. The only exception to this is that one-lane roadways are assumed to have a 50% lower road works frequency.

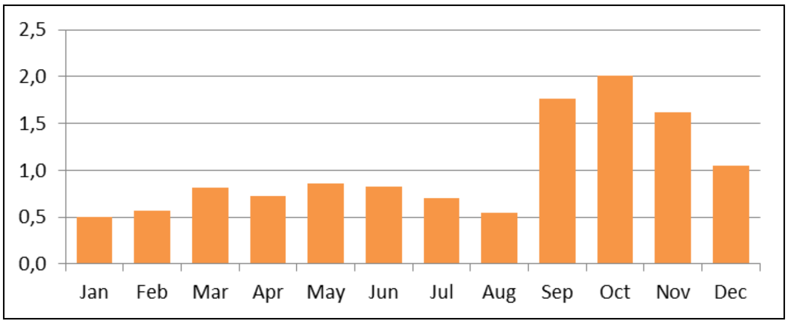
Probability input data:

- The probabilities of occurrence of road works are derived from the 'Meldwerk' report mentioned in chapter 2. This report provides some statistical data of road works on the main road network of the Netherlands, collected over a one year period.
- Using that:
    - in 2002, a total number of 25143 road works was registered (on motorways);
    - 95% of the road works on the main road network was reported to last for less than 24 hours;
    - in 2002, the total length of the motorway network amounted to 2300 kilometers;
    - in the model, the average motorway link length amounts to 1.2 kilometers,

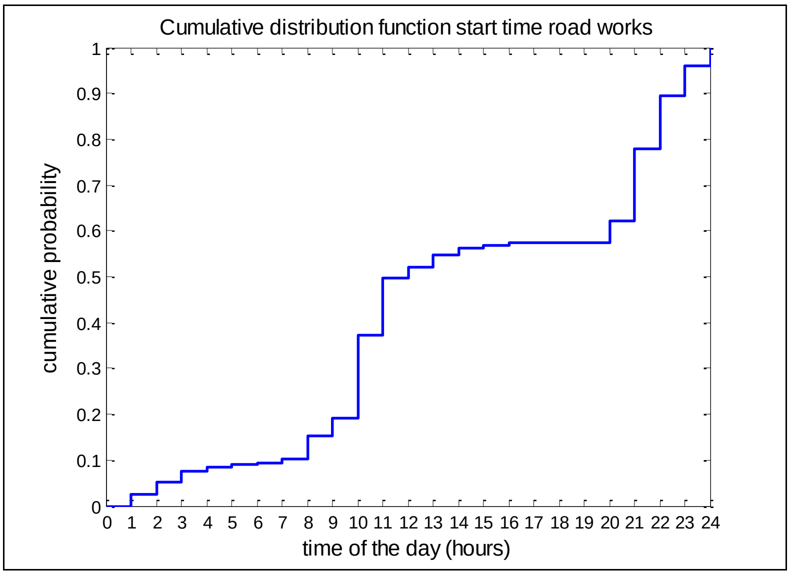
  and assuming that on weekend days 50% more road works are carried out than on weekdays, the following probabilities of occurrence of road works have been obtained[83]:
    - weekdays: 2.9%
    - weekend days: 4.4%

  These probabilities are adjusted with the following correction factors for the month of year, in order to account for the month-dependency in the execution of road works:

[83] representing the probability that at some time on the simulated day road works are started on the considered link



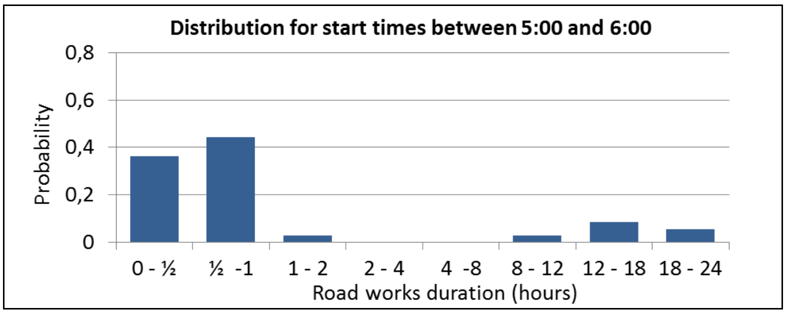
- For the random generation of the *start time* of the road works, the following empirical cumulative probability distribution is used[84]:

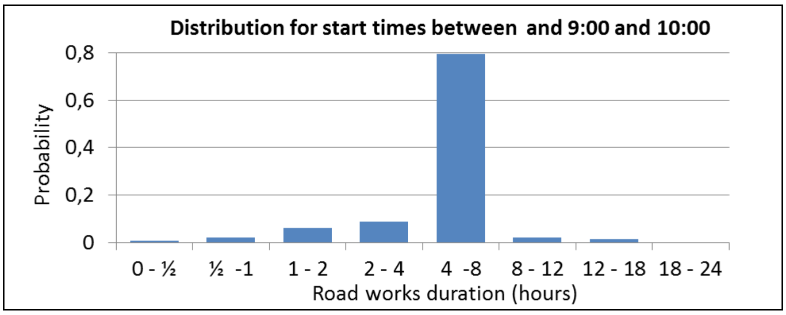


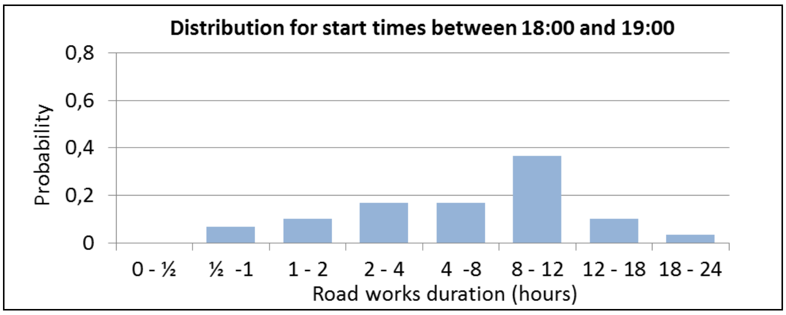
This distribution is valid for road works on weekdays. In the absence of the required empirical data, for weekend days a uniform distribution is assumed.

- For the random generation of the *duration* of the road works, an empirical probability distribution is used which is conditional on the start time of the road works. This way, it is fully reflected in the model that fewer road works are being carried out during the peak periods. The conditionality on the start time is illustrated in the figures below, which show the probability distribution of the duration for three different times of the day.

[84] In this diagram, the changes in cumulative probability have been depicted at the *upper bounds* of the start time classes.

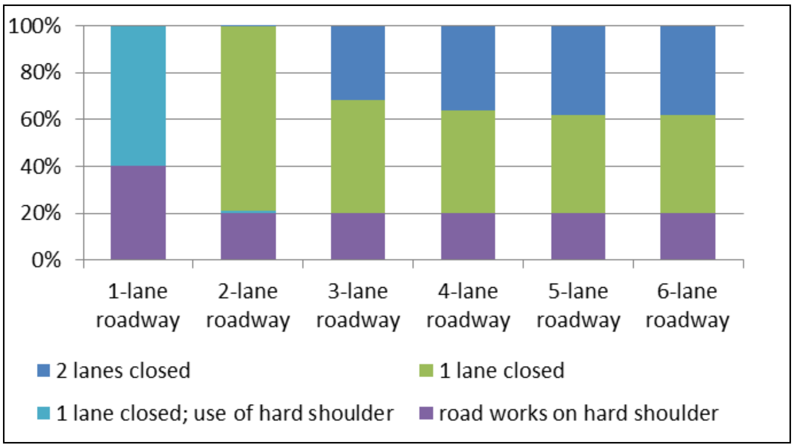






For the duration of road works on weekend days another probability distribution is used, which is not conditional on the start time. In the absence of the required empirical data, this distribution is taken equal to the average (i.e. unconditional) distribution of weekdays.

- Based on the 'Meldwerk' data and some further assumptions, the following cumulative probability distributions have been assumed for the 'severity' of the road works (i.e. the number of lanes that are closed):



The interdependency between severity on the one hand and start time and duration on the other, are not taken into account. This is obviously an important limitation of the model.

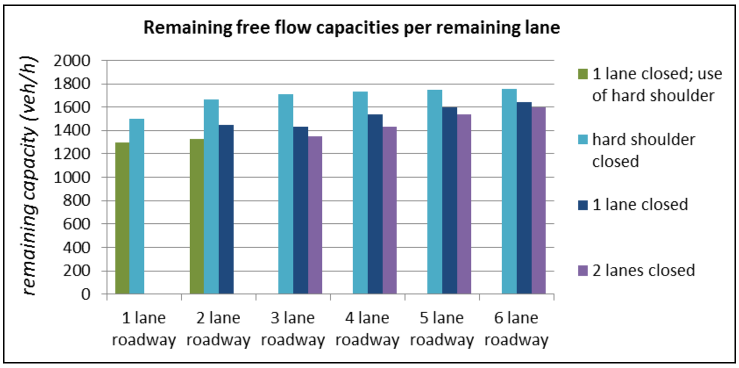
Procedure for simulating the occurrence of road works:

- For every link it is randomly determined whether or not road works are started on the simulated day, using the probabilities of occurrence mentioned above.
- Next, for each of the simulated road works a start time class is drawn from the start time distribution given above. The exact start time is determined by linearly interpolating with a random number.
- After that, for all of the simulated road works a duration class is drawn from the (conditional) duration probability distribution. Again the exact value is determined by performing a uniform draw between the class boundaries.
- Finally, for each of the road works it is randomly determined how many lanes are closed, using the probability distributions given above.
- It is important to be aware of the fact that there might also be road works on the simulated day that have *started already on the day before*. Therefore, the above procedure is repeated for a hypothetical previous day. It is checked then whether there are any road works that extend into the current day. If so, this is accounted for by the model.

Modeling of the effects:

For each of the simulated road works, the following supply effects are implemented:

- The speed limit (and thus the *free speed*) is assumed to be reduced to 70 km/h.
- The *number of available lanes* is reduced by the stochastically generated number of lanes that are closed for the road works.
- Based on values from the 'Handboek Capaciteitswaarden Infrastructuur Autosnelwegen' (AVV, 2002) and some further assumptions, the *free flow capacities* of the remaining lanes are adjusted to the following values:



In the absence of information regarding any possible differences in the effects on the free flow capacity on the one hand, and the queue discharge rate (from standstill) on the other, the calculation of adjusted values for the latter is based on the assumption that the relative capacity drop is unaffected. Obviously, this corresponds to the assumption that the relative effect on the queue discharge rate is equal to the relative effect on the free flow capacity.

- Because of the fact that the free speed is reduced to 70 km/h, the *critical speed* has to be reduced to 70 km/h as well. This already accounts for part of the effect on the free flow capacity (which equals the product of critical speed and critical density). It is assumed that the remaining part of this effect can be fully assigned to the *critical density* (meaning that the critical speed is not lowered any further).

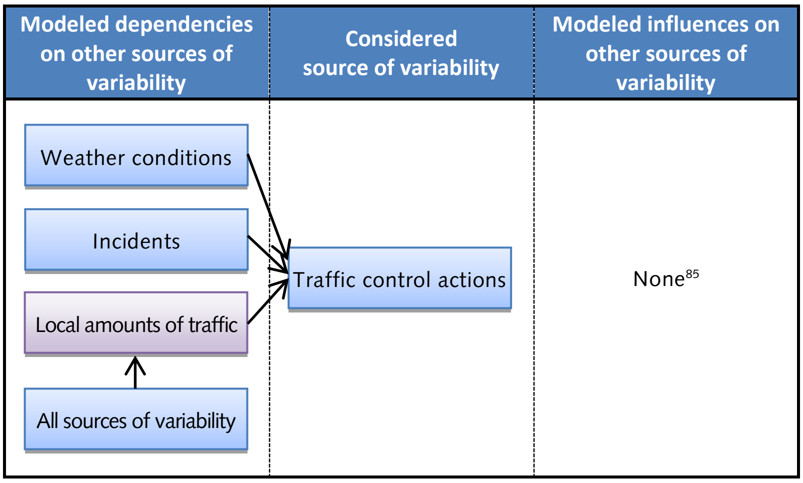
5.4.14 Traffic control actions (rush-hour lanes)

Classification:

- Modeled as a **DISCRETE** influence factor,
- with a **TRAFFIC RESPONSIVE DYNAMIC** way of occurrence,
- and **LOCAL** effects
- on the traffic **SUPPLY** characteristics.

Modeled interdependencies:

Table 5.18: Modeled dependencies between traffic control actions (the opening and closure of rush-hour lanes) and other sources of variability



General characteristics of the simulated traffic control actions:

By default, no traffic control actions are simulated in the model. However, in order to be able to demonstrate that additional/revised insights may be obtained when evaluating a proposed traffic measure according to an approach in which the various sources of variability are explicitly accounted for, a possibility was included to simulate one specific example of dynamic traffic management measures, namely rush-hour lanes. These rush-hour lanes can be introduced on any link of the motorway network. They are supposed to involve a dynamic use of the hard shoulder of the road.

[85] Note that a potentially important dependency is omitted here. Due to the temporary absence of a hard shoulder (during the period in which it is opened to traffic), incidents will more often cause lane blockages.

Procedure for simulating the dynamic behavior of rush-hour lanes:

- After every simulated 5-minute interval it is checked for every individual rush-hour lane section whether it should be opened (if it is currently closed) or closed (if it is currently opened). For the opening of a section, a threshold of a (5-minute averaged) flow on this section of 1500 veh/h $\cdot$ n_{lanes} is used, where n_{lane} is the (nominal) number of lanes. For the closure of a section a lower threshold is used, in order to avoid (high frequency) alternating behavior: 1200 veh/h $\cdot$ n_{lanes}. This latter threshold is however not sufficient, since a low flow might be caused by congestion as well. Therefore, another requirement was added, which states that the (5-minute averaged) speed should be above 55 km/h.
- If a rush-hour lane is going to be opened, it is checked whether the concerning link is incident free. If not, the opening of the rush-hour lane is cancelled. In such cases the hard shoulder will usually be blocked by the incident or used by emergency services, meaning that it cannot be opened to traffic.
- Under certain weather conditions, the rush-hour lane cannot be opened either:
    - In case of limited visibility (due to fog or snow), the road traffic controllers cannot observe whether the hard shoulder is obstacle free (using the cameras installed for this purpose).
    - In case of snow or black ice, the hard shoulder might be too slippery to be opened to traffic.

  This is modeled by the specification of some weather-related criteria.

Modeling of the effects:

The effects of the opening of a rush-hour lane are modeled in the following way:

- The speed limit (and thus the *free speed*) is assumed to be reduced to 80 km/h (for safety reasons, related to the smaller width of the rush-hour lane).
- The *number of available lanes* is increased by one.
- The *capacities* per lane, the *critical density* per lane, and the *jam density* per lane are not adapted. Note that this is a simplification of reality. For the analyses presented in this report, this simplification was acceptable, however.
- The *critical speed* is not adjusted, unless the reduction of the free speed necessitates this[86]. Note that in this latter case the *critical density* is to be adjusted as well, in order to keep the product of critical speed and critical density equal to the free flow capacity, which by definition must be the case.

[86] A critical speed which is larger than the (reduced) free speed is unrealistic, and therefore not accepted.

5.4.15 Intrinsic random variability in human driving behavior

Classification:

- Modeled as a **CONTINUOUS** influence factor,
- with a **STOCHASTIC** way of occurrence,
- and **LOCAL** effects
- on the traffic **SUPPLY** characteristics.

Modeled interdependencies:

Table 5.19: Modeled dependencies between the intrinsic randomness in human driving behavior and other sources of variability

Modeled dependencies on other sources of variability	Considered source of variability	Modeled influences on other sources of variability
None	Intrinsic variability driving behavior	None

General characteristics of the simulated intrinsic random variability:

- The intrinsic random variations in the *free speeds* are neglected. In view of the fact that these local variations will largely counterbalance each other at the level of routes, they are considered less important.
- The intrinsic random variations in the jam densities are neglected as well. In practice, these variations are found to be relatively limited. Furthermore, it should be noted that as far as their effect on queue lengths is considered, the local variations can be expected to partially counterbalance each other.
- The intrinsic random variations in the *free flow capacities* are modeled by assuming them to be Weibull distributed, in line with the research results discussed in section 2.2.5. Roughly based on the findings of Brilon et al (2005) and Geistefeldt (2009), the coefficient of variation of these distributions is assumed equal to 7%. The mean of the distributions is taken equal to the capacity value as obtained by correcting the nominal/representative capacity value for all other sources of variability.
- No information was found on the characteristics of the intrinsic random variation in the *queue discharge rates from standstill*. It is assumed that this variation can be modeled with a Weibull distribution as well, though with a lower variability. A coefficient of variation of 55% of that of the free flow capacities is assumed.
- The value of 7% given above applies to roadways with three lanes. Likely, for roadways with other numbers of lanes different coefficients of variation are found. This is accounted for by using the following equation to calculate the coefficients of variation for roadways with other numbers of lanes:

$$CoV(n_{lanes}) = \frac{\sqrt{3}}{\sqrt{n_{lanes}}} \cdot CoV(3)$$

In this formula, $CoV(n_{lanes})$ is the coefficient of variation for roadways with n_{lanes} lanes. Accordingly, $CoV(n_{lanes})$ is the coefficient of variation for roadways with 3 lanes (0.07).

Note that this formula is based on the assumption that the stochastic roadway capacity is equal to the summation of n_{lanes} independent identically distributed lane capacities. Although not *entirely* correct, independency between the different lane capacities seems a reasonable assumption [87] (at least more reasonable than the assumption of full *dependency*). The direct consequence of this assumption is that the *relative* variability of the roadway capacity is larger for roadways with a lower number of lanes, which seems reasonable. Empirical research will be required, however, to find out whether this is indeed the case.

- When the number of lanes is simulated to change (due to some other source of variability), the coefficients of variation of the capacity distributions are adjusted accordingly, using the formula given above.
- It is important to be aware of the fact that if the average of the capacity distribution changes (due to one of the other sources of variability, affecting the capacity per lane), this will cause the absolute width of this distribution (reflecting the variation in the capacity values) to change as well, if the coefficient of variation is kept constant. (After all, the standard deviation of a distribution is by definition equal to the product of the mean and the coefficient of variation). Since this might not be realistic, the model corrects for this by multiplying the coefficient of variation with the ratio of the representative capacity value and the newly computed average capacity value.
- The developed model also provides the option to include any possible influences that other sources of variability may have on the degree of intrinsic randomness of the capacities. In the quantitative analyses presented in this report this option has not been used, however, because of a lack of information regarding such potential influences.

Procedure for simulating the effect of the intrinsic random variability

For every cell in the network, independent random realizations of the Weibull distributed capacities are generated for every 5-minute interval. For this, the scale and shape parameters of the Weibull distributions are required. These parameters are uniquely related to the mean and coefficient of variation of the distribution. The parameters cannot be calculated from the

[87] Note here that the fact that vehicles can switch from one lane to another is already accounted for by confronting the *total* traffic demand with the *total* roadway capacity (rather than confronting demand and capacity lane by lane). Also note that the common sources of variability (causing interdependencies between the capacities of the different lanes) are largely accounted for already by explicitly adapting their average values to the variations in these influence factors.

mean and coefficient of variation in analytical way, however. Therefore, a numerical approximation is required.

Related to the fact that this numerical approximation is to be repeated a very large number of times (because of the fact that the mean and coefficient of variation are variable over time and space), this was computationally too demanding, however. Therefore, another approach was devised. This approach uses a large precompiled conversion table, which contains the shape parameters for a large number of coefficients of variation. This way, the required shape parameter can be found in virtually no time and yet with a high accuracy.

Once the shape parameter is known, the scale parameter can straightforwardly be calculated from the combination of this shape parameter and the mean capacity value, using an analytical relationship.

5.4.16 Intrinsic random variability in human travel behavior

Classification:

- Modeled as a **CONTINUOUS** influence factor,
- with a **STOCHASTIC** way of occurrence,
- and **ORIGIN-DESTINATION SPECIFIC** effects
- on the traffic **DEMANDS**.

Modeled interdependencies:

Table 5.20: Modeled dependencies between the intrinsic randomness in human travel behavior and other sources of variability

Modeled dependencies on other sources of variability	Considered source of variability	Modeled influences on other sources of variability
None	Intrinsic random variability demands	None

Procedure for simulating the effect of the intrinsic random variability

It is assumed that all other sources of variability in the traffic demands together fully explain the part of the demand variation for which a relation/dependency exists between the individual travelers (i.e. the part of the variation in their travel behavior that is attributable to some *common* external influences). What remains then, is the *independent* part of the variation in the travel behavior of individual travelers. That is, the part of the variation that can only be explained by personal influence factors.

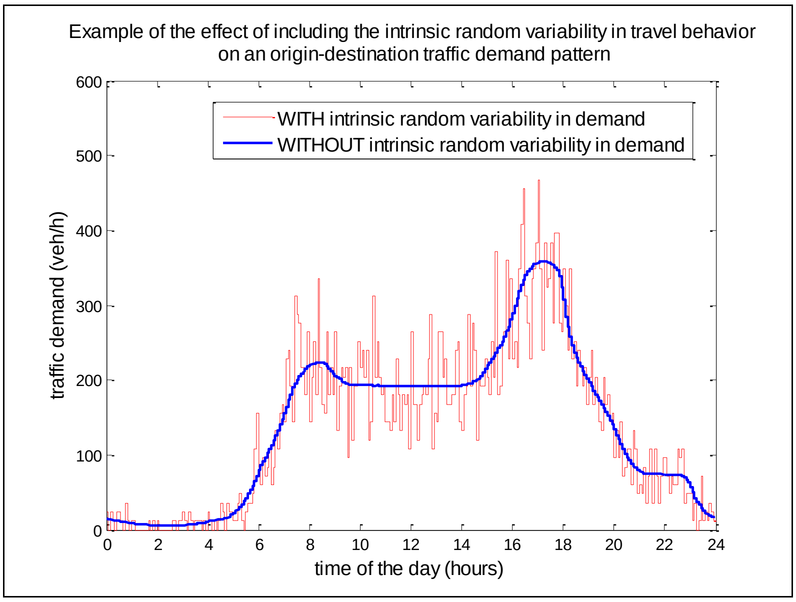
In this case, the number of travelers on a given origin-destination relation in a given 5-minute interval can be considered as the summation of n Bern(p_i)-distributed variables, where n represents the (imaginary) total number of *potential* travelers, and p_i represents the probability that potential traveler i decides to make the trip. This probability has a different value for each (potential) traveler i, and is conditional on the various common external influences mentioned above.

This summation of n Bern(p_i)-distributed variables can be approximated by one single Binom(n,$<p_i>$)-distributed variable, where $<p_i>$ represents the average probability over all potential travelers. Note that the variance of this variable is larger than the variance of the original summation. This does not really matter, however. In fact, one could even say that this will partially compensate for the fact that inevitably some of the 'common external influence variation' will have been 'overlooked' in the model.

The main problem is now the choice of appropriate values for n and $<p_i>$. This problem can be avoided by approximating the Binom(n,$<p_i>$)-distributed variable by a Poiss(λ)-distributed variable, where $\lambda = n \cdot <p_i>$. Now it is no longer necessary to choose values for n and $<p_i>$. A value for their product is sufficient. This product corresponds to the expected value of the Poisson distribution. This expected value can be considered to be given by the demand value following from all other sources of variability (computed for the given origin-destination relation and 5-minute interval). This makes the Poisson distribution conditional on all external influence factors, as it obviously should be.

Note that by approximating the Binom(n,$<p_i>$)-distributed variable by a Poiss(λ)-distributed variable, it is implicitly assumed that $n \rightarrow \infty$ and $<p_i> \rightarrow 0$. Again, this results in an overestimation of the variance of the traffic demands. As argued before, this does not seem too problematic, however.

The figure below shows an example of the effect of the addition of the intrinsic random demand variability on an origin-destination demand pattern over the day.



5.5 Issues requiring further consideration

In the previous sections, the developed quantification model has been described. During the development of this model, some modeling issues have come to light which require further consideration. These issues are dealt with in the next chapter.

6. Remaining modeling issues

6.1 Introduction

As noted in the final section of the previous chapter, in the development of the model some modeling issues have come to light which require further consideration. These issues will be discussed in this chapter. Since they generally require substantial further research, it was not possible to actually solve them within this master thesis research project. However, besides explaining the different problems, this chapter also tries to suggest some possible strategies to overcome these problems. This includes a possible solution strategy for reducing the required number of simulations. At this moment, at least a few thousand simulations are necessary to obtain results with a sufficient level of statistical accuracy. Given the simulation time of about 1.5 to 2 minutes per simulation, one simulation series would take multiple days, if not multiple weeks.

6.2 Spatial scale of the random capacity variations

As explained before, the road capacities are randomly varied per cell of the network (modeling these as Weibull distributed variables), in order to account for the intrinsic randomness in human driving behavior (both between and 'within' drivers). This means that it is implicitly assumed that the spatial scale of these random capacity variations matches the cell size of the numerical solution scheme. Obviously this assumption is purely a pragmatic one, without any theoretical foundation.

Obviously, the stochastic capacities of two very closely spaced cross-sections of a road will be highly correlated with each other. This correlation arises from the fact that during a 5-minute interval, almost the same traffic – being an 'arrangement' of individual driver-vehicle entities – passes by at these two cross-sections. For an increasing spacing between the two cross-sections, this mutual correlation will decrease. This is due to the facts that, to an increasing extent:

- the passing traffic will partly consist of *other driver-vehicle entities* at the two cross-sections
- the driver-vehicle entities will be *differently 'arranged' within the traffic flow* (due to lane changes and speed differences)
- *variations in the behavior of an individual* may take place between the two cross-sections (e.g. due to fluctuations in attention level)

To the best knowledge of the author, the spatial dependency in the stochastic capacities has never been studied. The only thing found in international literature is a remark by Brilon et al (2005), stating that is seems reasonable to assume independency of the capacity values if the lengths of the sections on which they are applied are chosen sufficiently large.

In spite of the little attention in international research, it is not unlikely that this aspect could have a significant effect on the final outcomes of the developed model. After all, the larger the number of independent capacity realizations on a link of a given length (corresponding to a shorter section length for which one uniform capacity value is assumed), the lower the minimum of these realizations is expected to be. This minimum of the capacity values on the link is obviously decisive for the amount of traffic that can traverse the link without inducing congestion (Figure 6.1). As a result, a smaller capacity length scale will on average result in more traffic congestion occurring.

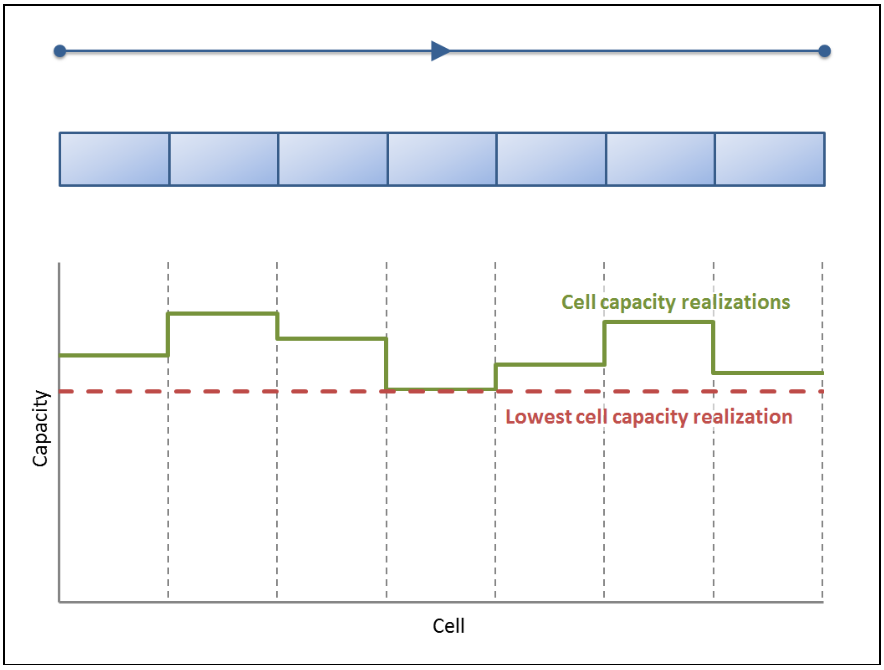


Figure 6.1: Modeling of a link (top) as a series of cells (middle) with independent capacity realizations (assuming a uniform capacity within a cell), of which the minimum value is decisive for the occurrence of traffic congestion on the link (bottom)

This latter effect is indeed observed in the model results. In order to illustrate this, an example is given in Figure 6.2. In this figure, the traffic speed on a link is shown as a function of the time of the day (horizontal axis) and the location along this link (vertical axis). The upper part of the figure shows the speeds as computed when independent capacity realizations are drawn per cell of the numerical discretization scheme (the default modeling approach in the developed model). The lower part on the other hand shows the speeds that are obtained when only one capacity value is drawn for the link as a whole. This latter modeling approach corresponds to the (unrealistic) situation in which the capacities on the link would be fully interdependent.

The figure shows that all traffic congestion that is calculated with the first modeling approach is absent in the results of the second modeling approach, except for the traffic congestion induced by two incidents that were simulated around 8 and 17 o'clock. In order to avoid giving a wrong impression, it has to be noted here that the difference between the outputs of the two modeling approaches is not on all links as big as for the link for which this figure was obtained.

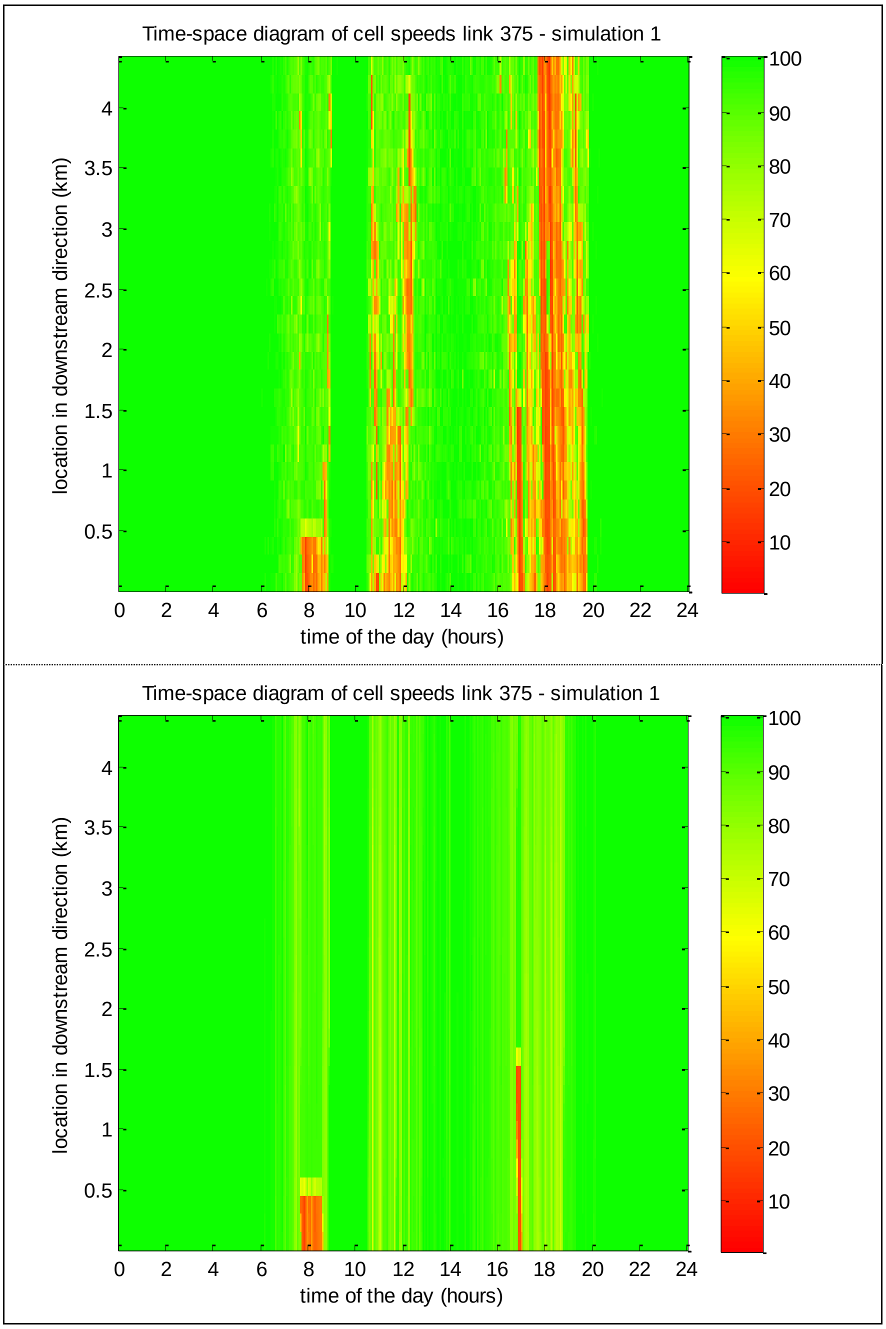


Figure 6.2: Time-space diagrams of the speed at a link as computed with a cell-based capacity variation approach (upper part) and as computed with a link-based capacity variation approach (lower part)

Let us now take a closer look at the nature of the traffic congestion computed with the first modeling approach (ignoring the incident-induced congestion). Part of this congestion can be recognized as spill-back from a downstream bottleneck (the thick red band near 18:15 hours), while the other part of the congestion consists of shock waves that are generated somewhere on the link itself. The probable explanation for the fact that the spill-back is not observed in the output of the second simulation is in the fact that in the first simulation the 'net' bottleneck capacity was likely to be lower. After all, in this simulation *n* independent capacity realizations were associated with the bottleneck-forming link (where *n* is the number of cells of this link), of which the minimum is decisive for the bottleneck capacity (as explained with Figure 6.1). In the second simulation only one capacity realization was taken for the link as a whole.

The congestion waves having their origin on the link itself are created when the propagating traffic flow pattern meets a cell with a capacity realization which is too low for it. In the figure it can be observed that many of these congestion waves have already dissolved again before reaching the upstream end of the link. Also note that most of the congestion waves are reasonably short. As a consequence, their final effects on the travel times are probably limited.

It should nevertheless be noted here that in real empirical data measured at the Dutch motorways we do not see that many congestion waves emerging at arbitrary places along a link as suggested by the model output depicted in the upper part of Figure 6.2. This does not have to say that the second modeling approach (i.e. using a *link*-based capacity variation procedure, corresponding to the assumption of a full dependency between the capacities at the different locations along a link) is more realistic, though. Such a full dependency seems rather unlikely. After all, the only variation that is considered here is the *variation due to the intrinsic randomness in human driving behavior*, which is expected to have a spatial correlation that clearly decreases with increasing distance, as explained above. All other sources of spatial dependencies in the capacity values, which might have much larger dependency length scales (such as the effects of bad weather, which often are network-wide) are not relevant here, since these have been explicitly accounted for already in the parameters of the capacity distributions.

Instead of abandoning the concept of a cell-based capacity variation, maybe exactly the opposite strategy should be followed in order to improve the validity of the modeling approach. That is, maybe the cell size of the capacity randomization should be taken *smaller* than the cell size currently used in the model. After all, if a very small cell size would be chosen, the higher traffic demands would likely be ‘filtered out’ already within short distances after bottleneck locations (typically found at the network nodes, i.e. at the points where different links are connected to each other). As a result, less traffic congestion would originate at arbitrary locations along the link, resulting in a better resemblance of the model output to the empirical patterns found in reality. However, the overall level of traffic congestion would be higher in this case, since a very small cell size would correspond to very large numbers of independent capacity realizations, of which the minimums would be decisive for the quality of the traffic flow. Considering that the current model already produces too much congestion in many of the simulation runs (see section 7.5.1), on this aspect the resemblance to reality would deteriorate.

Apart from the possible role of the capacity cell size, some other possible causes for the discrepancy between the model output and real-life data (i.e. the fact that we do not really observe that many congestion waves originating at arbitrary locations along a link as predicted by the model) have been identified as well:

- In reality, the (free flow) capacity of a bottleneck itself (i.e. the part of the road in the direct vicinity of the discontinuity in geometry) may be significantly lower than the capacity of the rest of the bottleneck-forming link, due to the specific traffic maneuvers that take place near the bottleneck. This difference can explain why high traffic demands usually break down already at the bottleneck location itself (i.e. directly at the beginning of a link in the model), and thus less often somewhere at 'arbitrary' locations along the link. From practice it is known that this is the case for weaving sections. For on-ramp bottlenecks, however, exactly the opposite is true: over a short length near the on-ramp significantly *higher* capacities are achieved than elsewhere on the link.
- In reality the traffic demand on links downstream of active bottlenecks may be significantly reduced by the capacity drop, resulting in the traffic demands remaining well below the capacities along these links. In the model, the capacity drop at fixed bottlenecks (represented by network nodes) is not correctly modeled (which is discussed in section 6.4), resulting in unrealistically high traffic demands downstream of such bottlenecks, which might generate congestion waves.
- From real-life data measured on Dutch motorways it appears that most traffic breakdowns at road sections without a distinct bottleneck can still be explained by some other, less pronounced spatial discontinuity, such as a gradual curve in the road axis, or a certain traffic monitoring device above the road, which locally reduces the capacity of the motorway. As a result, most of the real-life traffic breakdowns occurring 'somewhere' along a link are concentrated at a limited number of locations with such a distinguishing feature. This is not taken into account in the model.

From the discussion above it might be clear that there is no clear solution for the problem at hand. More research into the spatial (or rather: *spatiotemporal*) dependencies in capacities seems indispensable, in combination with research into the (possible) phenomena listed above. In any case, capacity distributions for road sections without distinct bottleneck – as derived by Brilon et al. (2005) – have little value without a corresponding length scale for which they have been derived.

After an investigation of the spatial (or spatiotemporal) dependencies, the newly obtained knowledge can be incorporated in the model by using it for making a well-founded choice for the lengths over which (mutually independent) uniform capacity values are applied, or by using it to add a certain correlation between the capacities of adjoining cells. Ideally, one would stop using capacity variables that are related to *fixed* cells of the network, and rather use capacity variables that are *propagated* along with the traffic flow, while varying in an autocorrelated way. It would be very difficult, however, to derive appropriate distribution functions and autocorrelation coefficients for this.

6.3 Mismatch between the demands and the capacity realizations

As described in section 5.3.1, traffic demand and capacity values are both generated on a 5-minute basis. As far as the demand values are concerned, this means that *at the network boundaries* traffic demands are constant within these 5-minute intervals. *Within the network* this however is not the case. This is due to the fact that the computational core of the model (i.e. the dynamic traffic simulation model) uses a much shorter time step (of 5 seconds), which is required for a sufficiently accurate traffic flow modeling. As a result, the traffic demands *within the network* show an unintended variation with a much shorter time interval, as illustrated by Figure 6.3. The vertical grid lines in this figure indicate the boundaries of the 5-minute intervals.

The consequence of the above is that in the traffic simulation 5-minute free flow capacity values are confronted with more frequently varying traffic demand values. This is obviously not correct, since these 5-minute capacity values are valid for 5-minute demand values only. For shorter time intervals different capacity distribution functions would apply. After all, the shorter the time interval that is considered, the less likely it is that a given traffic demand level induces a traffic breakdown during this interval.

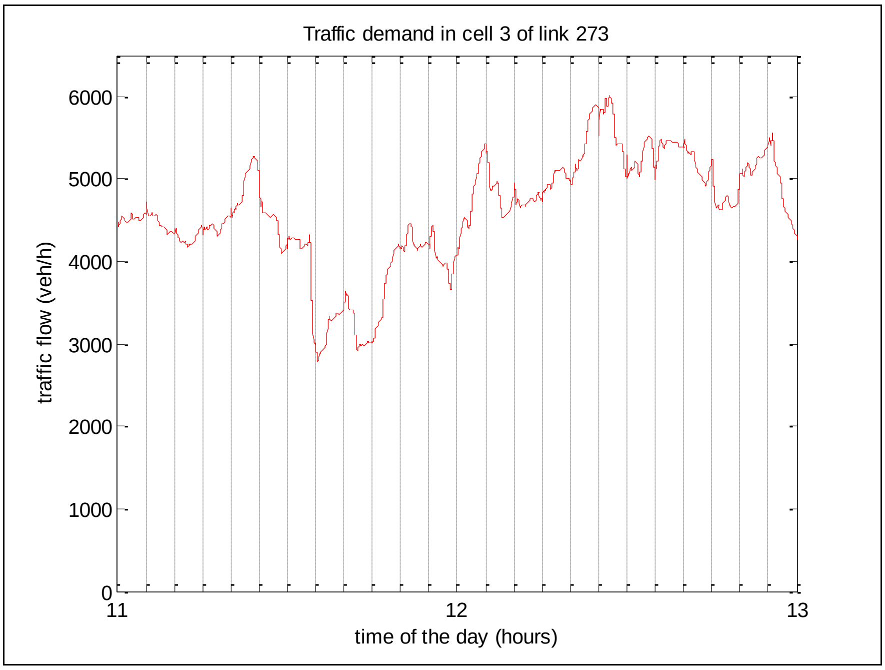


Figure 6.3: Graph of the traffic demand in an arbitrary cell of the network (for two arbitrary hours of the day), illustrating that this traffic demand is not constant within the 5-minute intervals (of which the boundaries are indicated by the vertical grid lines)

As a result of this mismatch between demand and capacity, capacity will be exceeded too frequently in the model. The final impact on the traffic conditions is unclear however. In any case, it is obvious that the net effect will be negative. After all, the additional congestion that is created when the traffic demand temporarily exceeds the 5-minute capacity value, *while its 5-minute aggregated value would not have exceeded it*, will not be compensated for at other moments in time. This is due to the fact that 'negative congestion' (i.e. a positively valued counterpart of congestion) does not exist, which makes traffic a non-linear system. It is estimated however that the amount of additional congestion remains limited, because of the fact that the extra capacity exceedances have limited durations and probably are usually also limited in size.

Note that there is no obvious solution for this problem. Basically, there are two types of solutions that could be suggested, which both turn out to be inappropriate when examined more closely:

- varying the capacity on a *5-second* instead of *5-minute* basis (using 5-second capacity distributions);
- using a cumulative approach for the confrontation of traffic demand with capacity.

The first approach is problematic because of the facts that:

- consecutive 5-second capacity realizations will probably not be independent of one another anymore, due to microscopic traffic flow phenomena with times scales in this order of magnitude;
- confronting the traffic demand values with 5-second capacity values actually would be erroneous too, since these traffic demand values are only *to a limited extent* variable per 5-second interval. Only if the demand randomization at the network boundaries would be based on 5-second intervals too, this would be a valid approach[88]. In such small time intervals traffic demands however are strongly determined by microscopic traffic processes, like platooning.

The second approach is inappropriate because it would imply that initially too much traffic could be let through, which should be compensated for by using a capacity value of zero from the moment the aggregated demand has reached the capacity. Clearly, this may yield unrealistic traffic conditions. In order to avoid this, one actually should redistribute the aggregated traffic demand over the 5-minute interval. This however would demand from the model that it could predict the future or, alternatively, go back in time.

6.4 Incomplete modeling of the capacity drop

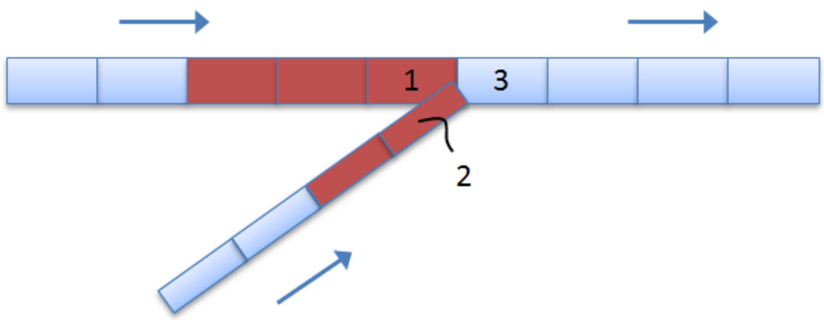
In section 5.2.3 it was described how the capacity drop is taken into account in the model. While this way of modeling indeed results in the capacity drop being taken into account for situations in which the number of lanes directly downstream of the head of the queue is equal to (or larger than) the number of lanes directly upstream of the head of

[88] Although it could very well be that these 5-second demand variations would be smoothed out too much by numerical diffusion in the traffic flow simulation.

the queue (a situation which typically occurs when a queue dissolves from its head, meaning that this head of the queue travels upstream), it does not result in the capacity drop being taken into account for situations in which the number of lanes directly downstream of the head of the queue is smaller than the number of lanes directly upstream of the head of the queue.

After all, in this latter case the inflow capacity of the downstream cell might well be lower than the outflow capacity of the (upstream) congested cell, even if in the computation of the latter a capacity drop is accounted for (see Figure 6.4). This makes the inflow capacity of the downstream cell decisive for the outflow of the queue, and not the outflow capacity of the congested cell. This means that in such situations the capacity drop should have been accounted for in the inflow capacity of the downstream cell, while currently the capacity drop is only applied on the outflow capacity of congested cells themselves.

Figure 6.4: Example of a situation in which the capacity drop remains without effect, due to an incompleteness in the modeling approach. For the situation shown in this figure this will have to be solved by modification of the node model.



Current calculation procedure:

Total outflow capacity congested cells:

cell 1: 2 lanes * 2000 = 4000 veh/h
cell 2: 1 lane * 2000 = 2000 veh/h
--------------- +
6000 veh/h
capacity drop: 1000 veh/h
--------------- –
5000 veh/h

Total inflow capacity downstream cell:

cell 3: 2 lanes * 2000 = 4000 veh/h

4000 veh/h

<u>Flux = min (5000, 4000) = 4000 veh/h</u>

Correct calculation procedure:

Total outflow capacity congested cells:

cell 1: 2 lanes * 2000 = 4000 veh/h
cell 2: 1 lane * 2000 = 2000 veh/h
--------------- +
6000 veh/h
capacity drop = 1000 veh/h
--------------- –
5000 veh/h

Total inflow capacity downstream cell:

cell 3: 2 lanes * 2000 = 4000 veh/h
capacity drop = 1000 veh/h
--------------- –
3000 veh/h

<u>Flux = min (5000, 3000) = 3000 veh/h</u>

Since this situation is the typical situation found at bottlenecks, where most of the traffic congestion emerges, this incomplete modeling of the capacity drop might well result in capacity drop having less effect in the model than in reality. This actually seems to be confirmed by some model test runs in which the simulated traffic conditions with and without the capacity drop were compared. In these tests the influence of the capacity drop seemed to be relatively modest.

Solving this problem is not so easy, since bottlenecks typically coincide with network nodes. This means that the downstream cell on which the capacity drop is to be applied on its inflow capacity is typically located on another link than the first congested cell of the queue. As a result, the problem cannot be solved by a simple adaptation in the flux model applied at the cell boundaries within a link. Instead, the flux model for the network nodes will have to be adapted, which will be more complicated.

6.5 Replication of the random number sequence

As explained in section 5.3.1, special care has been taken to make sure that the same pseudo-random numbers would be used in different model runs, in order to improve the comparability of the outputs of these runs. This means that all different parts of the model have been programmed in such a way that they always generate the same number of pseudo-random numbers, irrespective of certain model settings. In a certain test run it was discovered however that some subcomponent of the demand randomizer does not always obeys to this principle, messing up the random number sequence.

Closer examination revealed that the problem is in the Poisson generator of MATLAB (which is used at the end of the demand randomization procedure, for obtaining the final realizations of the demand values, given their expected values). The algorithm underneath this generator turned out to use a randomly varying number of (uniform) pseudorandom numbers per evaluation.

Since there was no straightforward solution for this (except for a computationally too demanding one), the problem has not really been solved. Instead, its final impact on the random number stream has been suppressed. This was done by programming the model to manipulate the random number generator in such a way after the completion of the Poisson procedure, that it arrives in a state which is a certain predefined amount beyond[89] the state that it had at the start of this procedure.

This approach ensures that the effects of the problem remain limited to the Poisson variations themselves only. Considering that the relative importance of these Poisson variations is found to be minor (see section 8.6), these remaining effects are unlikely to have significant consequences.

[89] in terms of the number of random values that is has generated

6.6 Required number of simulations

This section deals with the problem of the extremely large computation time that is associated with performing analyses with the model. Below, this problem first is explained in more detail. After that, different solution approaches are discussed. Finally one of these strategies is developed in more detail.

6.6.1 Problem explanation

The required number of simulation runs is dependent on:

- the accuracy with which the performance indicators are to be computed,
- the desired confidence level (i.e. the level of confidence with which the desired accuracy is to be achieved),
- the variation in the output of the traffic system, and
- the performance indicators to be evaluated.

Note that the required number of simulation runs is not directly dependent on the number of degrees of freedom included in the model, or their variation. A larger number of degrees of freedom or a larger variation in these does only result in an increase of the required number of simulation runs to the extent that this contributes to a larger variation in the output of the traffic system.

Of the various statistics that are to be calculated from the simulation results (see chapter 3), probably the 90th percentile travel time (which is not only one of the indicators itself, but is also used in the computation of two other indicators [90]) is decisive for the required number of simulations[91]. This is due to the fact that this statistic is 'located' in the long (i.e. low-density) tail of the travel time distribution, which results in the required number of simulations for a given accuracy and confidence level to be relatively high.

It is well known that the variance of a quantile estimator (or equivalently, *percentile* estimator) based on order statistics is given by the following equation (Chen and Kelton, 2001):

Equation 6.1

$$\mathrm{Var}\left(\hat{x}_p\right) = \frac{p(1-p)}{(n+2)f^2(x_p)} + O(1/n^2) \quad \text{for } \hat{x}_p \text{ given by } \hat{x}_p = Y_{[np]}$$

In this equation,

- $\hat{x}_p$ is the estimator for x_p, which is the p quantile (or, equivalently, the $100p^{th}$ percentile) of a probability distribution,
- n is the number of simulations (i.e. independent realizations from one and the same probability distribution $F(x)$,

[90] the width and the skewness of the travel time distribution

[91] It should be noted that it actually is not impossible that not the 90th percentile travel time, but rather the travel time instability 'indicator' is decisive for the required number of simulations. However, since this 'indicator' is a whole probability distribution (instead of a single-valued quantity; see section 3.4), it is difficult to make an estimate of the number of simulations that it requires.

- $f(x)$ is the probability density function of the random variable X (in this case the travel time), and
- Y_j $(j = 1 \dots n)$ is the j^{th} order statistic of the $X_i' s$ $(i = 1 \dots n)$, being the n independent realizations of X.

In order to get an indication of the required number of simulation runs (n), this equation was applied on a hypothetical travel time distribution. Imagine that we would require that $\mathrm{Var}(\hat{x}_p) \leq 1$ (corresponding to a maximal standard deviation of the estimator of 1 minute[92]), and that $f(x_{0.9})$ = 0.005 min^{-1} (which follows from the hypothetical travel time distribution). Then it follows directly from Equation 6.1 (neglecting the second order term) that for the calculation of the 90th percentile travel time ($p = 0.9$) about 3600 simulation runs are required.

It should be noted that of course not only the variance of the estimator should be sufficiently small, but its bias as well. This poses another requirement on the number of simulations. However, given the fact that the bias is likely to be relatively small compared to the variance[93], it is estimated that the requirement on the variance will be decisive.

As expected, for the calculation of the 10th and 50th percentile travel times (which in chapter 3 were identified as (elements in) important performance indicators as well) a much lower number of simulation runs would be required (for the same requirement $\mathrm{Var}(\hat{x}_p) \leq 1$). For the 50th percentile $n \geq 623$ is found (assuming $f(x_{0.5})$ = 0.020 min^{-1}), and for the 10th percentile $n \geq 81$ (assuming $f(x_{0.1})$ = 0.033 min^{-1}).

Since one simulation run takes about 1.5 to 2 minutes, the in the above estimated required number of 3600 simulation runs would correspond to a total simulation time of multiple days. In chapter 3 it was argued, however, that in an ideal evaluation the 90th percentile travel times should be evaluated separately for different subsets of the simulation runs. For example, the travel times simulated for Fridays should be considered separately from those simulated for other weekdays, since regular road users will be aware of the systematic differences between these different days of the week, which means that these differences do not contribute to their travel time uncertainty. If the 90th percentiles indeed are evaluated separately for different subsets of the simulation runs, even more simulation runs would be needed, resulting in a total simulation time of multiple weeks. Since it ultimately is aimed for to compare the traffic system's performance for different scenarios (for example *with* and *without* certain traffic measures), this multiple weeks lasting procedure would have to be run multiple times (at least twice), resulting in the total amount of time required for a complete analysis being even several times as large.

In view of this extremely large computation time, it is desirable to look for ways to reduce the required number of simulations.

[92] This actually can be seen as a combined accuracy and confidence level requirement.

[93] For large n, $\mathrm{Bias}(\hat{x}_p) \approx -\frac{f'(x_p)}{2f(x_p)} \mathrm{Var}(\hat{x}_p)$.

6.6.2 Solution strategies

Obviously, one possibility would be to remove the 90th percentile travel time from the set of indicators to be evaluated. In view of the important role of this statistic (both as separate indicator and as element of other indicators), this is clearly no option, however.

Another solution that might be suggested could be to abandon the Monte Carlo based approach, and to resort to a *scenario-based* simulation approach (meaning that only a number of predefined demand/supply scenarios would be simulated). After all, at first sight it might seem somewhat inefficient to 'simply' generate random realizations from the full spectrum of possible outcomes (i.e. possible traffic conditions). However, it can easily be seen that a scenario-based approached is definitely not appropriate for the task at hand, since:

- Using a scenario-based approach would require to make assumptions on the relative importance of the various influencing factors, while it is precisely one of the objectives of this project to find out whether explicitly considering the inherent variability in the traffic system can provide us with additional insights into these relative contributions.
- It would be very difficult (if not impossible) to derive statistical indicators (characterizing probability distributions) from the results of a set of predefined scenarios.

In view of the above, we are more or less stuck to using the approach using Monte Carlo simulation. However, a possible strategy to reduce the required number of simulation runs could of course be to look for more advanced variants of the Monte Carlo technique, which are more efficient.

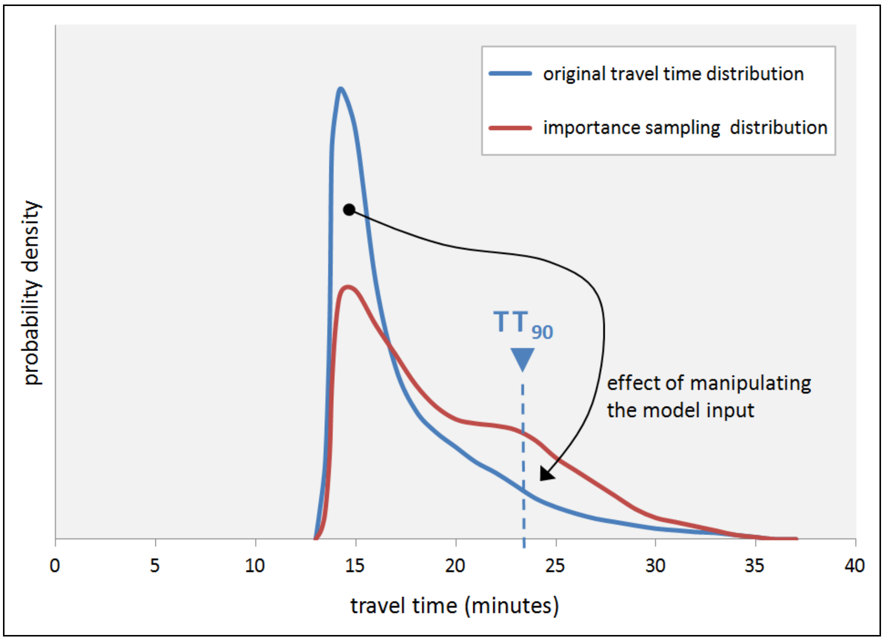
These more advanced ways of sampling can be found in the techniques that are known as 'Latin Hypercube Sampling' and 'Importance Sampling'. With Latin Hypercube Sampling, the total probability space of the random variables in question is divided in a number of equally sized intervals (equally sized in terms of probability). This number of intervals should be equal to the total number of simulations to be performed. In the sampling procedure a random realization of a variable now is taken in two stages. First, randomly one of the intervals is selected. After that, a random realization is taken from this specific interval. The key of Latin Hypercube Sampling is now that the interval is selected from the subset of intervals *that have not been selected in one of the simulations before*. Consequently, after finishing the last simulation, all intervals have been selected once. This way, the randomly generated variable values are very likely more evenly spread over the total probability space as compared with those generated with the basic Monte Carlo technique. As a result, the generated outcomes can be assumed to better represent the total space of all possible outcomes. This then results in the number of simulations required to achieve a certain statistical accuracy being lower.

It might be clear however that Latin Hypercube Sampling is yet not the most efficient sampling technique that one conceptually could think of. After all, the estimation of the 90th percentile travel time will still require far more simulations than the estimation of for example the median of the travel time, because of its location in the tail of the distribution. It would therefore be more efficient if the sampling procedure would give extra emphasis to this tail of the travel time distribution. This basically is the idea of the Importance Sampling technique. This technique is explained in more detail below. Here it is also put forward how this technique could be used in the model at hand. Note that an integration of this technique with the Latin Hypercube Sampling technique described above of course would be even more efficient.

6.6.3 Elaboration of the importance sampling technique

Assuming that the 90th percentile travel time statistic by far requires the largest number of simulation runs (see the discussion in section 6.6.1), and is therefore decisive for the number of runs to be performed, this required number of simulation runs can be reduced by making 'unfavorable' simulation outcomes (involving travel times which are found in the right tail of the travel time distributions) more likely, by manipulating the inputs of the dynamic traffic simulation model (i.e. the traffic demand and supply characteristics generated by the randomization components). This is illustrated in Figure 6.5. Note that this manipulation should not be exaggerated, since other indicators (like the median of the travel time distribution) might become decisive for the required number of simulations then, meaning that the problem would be moved.

Figure 6.5: Importance sampling: increasing the likelihood of certain outcomes of the simulation runs by manipulating the input of the dynamic traffic simulation model (i.e. the randomly generated traffic demand and supply characteristics)



From Equation 6.1 it appears that this technique in principal should allow for a drastic reduction of the required number of simulation runs, since this required number is proportional to the *square* of the probability density corresponding to the 90th percentile travel time

$f(x_{0.9})$. It is estimated that in this context the maximum achievable effect will be a reduction from a few thousand to only one thousand simulation runs.

Of course, after the simulation process a correction should be applied on the outcomes, in order to correct for the manipulation of the input. Currently, in the calculations that are performed by the data processing component of the model all outcomes of the different simulations runs o_j $(j = 1 \dots n)$ are of course weighted equally. If importance sampling is applied, however, each individual simulation outcome o_j should be weighted by the ratio of its *original* probability density $f(o_j)$ and its *importance sampling* (or *manipulated*) probability density $f_{IS}(o_j)$, in order to correct for its too high or too low likelihood of occurrence in the simulation process:

Equation 6.2

$$w_j = \frac{f(o_j)}{f_{IS}(o_j)}$$

In this equation, w_j is de weight (or correction factor) of simulation outcome o_j. The simulation outcome for example could be a travel time or the number of lost vehicle hours.

The problem of Equation 6.2 however is that $f(o_j)$ and $f_{IS}(o_j)$ are both unknown. This is commonly solved by calculating the weight factor from the (original and manipulated) joint probability density values of the *input*:

Equation 6.3

$$w_j = \frac{h(i_{1_j}, i_{2_j}, \dots, i_{m_j})}{h_{IS}(i_{1_j}, i_{2_j}, \dots, i_{m_j})}$$

Here $h(i_{1_j}, i_{2_j}, \dots, i_{m_j})$ is the *original* joint probability density of the input values i_{k_j} $(k = 1 \dots m)$, which are the realizations of the input variables I_k $(k = 1 \dots m)$ for simulation run j, describing the traffic demand and supply characteristics.
Similarly, $h_{IS}(i_{1_j}, i_{2_j}, \dots, i_{m_j})$ is the *importance sampling* (or *manipulated)* joint probability density of the input values i_{k_j} $(k = 1 \dots m)$.

In this case, this new equation does not really solve the problem, however, because of the difficulties in calculating $h(i_{1_j}, i_{2_j}, \dots, i_{m_j})$ and $h_{IS}(i_{1_j}, i_{2_j}, \dots, i_{m_j})$, which arise from the interdependencies between various of the input variables (i.e. demand and supply variables) I_k. Therefore, another approach is suggested. In this approach, not directly the probability distribution functions of the input variables themselves are manipulated, but rather the – originally uniform – probability distributions of the random numbers that are generated for the calculation of these input variables. The main advantage of this approach is that all these random numbers are mutually independent, which significantly simplifies the calculation of the weight factors:

Equation 6.4

$$w_j = \frac{g(r_{1_j}, r_{2_j}, \ldots, r_{z_j})}{g_{IS}(r_{1_j}, r_{2_j}, \ldots, r_{z_j})} = \frac{u(r_{1_j})u(r_{2_j}) \ldots u(r_{z_j})}{u_{IS}(r_{1_j})u_{IS}(r_{2_j}) \ldots u_{IS}(r_{z_j})}$$
$$= \frac{1 \cdot 1 \cdot \ldots \cdot 1}{u_{IS}(r_{1_j})u_{IS}(r_{2_j}) \ldots u_{IS}(r_{z_j})} = \frac{1}{u_{IS}(r_{1_j})u_{IS}(r_{2_j}) \ldots u_{IS}(r_{z_j})}$$

In this equation:

- w_j is de weight (or correction factor) of simulation outcome o_j.
- r_{t_j} $(t = 1 \ldots z)$ is the realization of the random number variable R_t for simulation run j.
- $g(r_{1_j}, r_{2_j}, \ldots, r_{z_j})$ is the *original* joint probability density of the random number realizations r_{t_j} $(t = 1 \ldots z)$.
- $g_{IS}(r_{1_j}, r_{2_j}, \ldots, r_{z_j})$ is the *importance sampling* (or *manipulated)* joint probability density of the random number realizations r_{t_j} $(t = 1 \ldots z)$.
- $u(r_{t_j})$ $(t = 1 \ldots z)$ is the *original* probability density of the random number realizations r_{t_j}. Since R_t is uniformly distributed between zero and one, obviously $u(r_{t_j}) = 1$.
- $u_{IS}(r_{t_j})$ $(t = 1 \ldots z)$ is the *importance sampling* (or *manipulated)* probability density of the random number realizations r_{t_j}. This probability density obviously is not necessarily equal to one.

Since only part of the random numbers will be manipulated (say only the R_t's for $t = 1 \ldots v$, resulting in $u_{IS}(r_{t_j}) = u(r_{t_j}) = 1$ for $t = (v + 1) \ldots z$), Equation 6.4 further simplifies to:

Equation 6.5

$$w_j = \frac{1}{u_{IS}(r_{1_j})u_{IS}(r_{2_j}) \ldots u_{IS}(r_{z_j})}$$
$$= \frac{1}{u_{IS}(r_{1_j})u_{IS}(r_{2_j}) \ldots u_{IS}(r_{v_j})u_{IS}(r_{v+1_j}) \ldots u_{IS}(r_{z_j})}$$
$$= \frac{1}{u_{IS}(r_{1_j})u_{IS}(r_{2_j}) \ldots u_{IS}(r_{v_j}) \cdot 1 \cdot \ldots \cdot 1}$$
$$= \frac{1}{u_{IS}(r_{1_j})u_{IS}(r_{2_j}) \ldots u_{IS}(r_{v_j})}$$

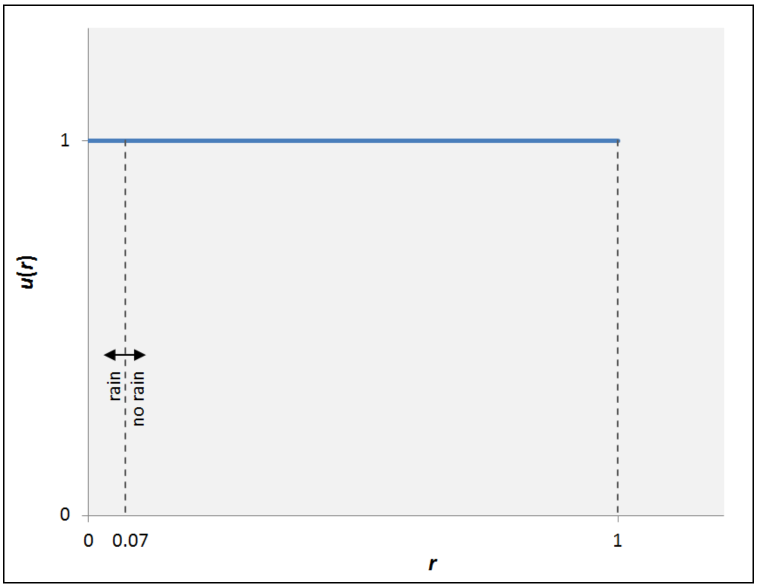
To illustrate the manipulation process, a simple example is considered.

Imagine that we would simulate the occurrence of rainy weather in the following (simplified) way (using standard Monte Carlo simulation, i.e. without applying importance sampling):

1) Draw a random number *r* (i.e. a realization of *R*, which is uniformly distributed between one and zero).
2) If $r \leq 0.07$ (the probability of occurrence of rainy weather), rainy weather will be simulated (by adapting the relevant traffic demand and supply characteristics).
3) If $r > 0.07$ no rainy weather will be simulated.

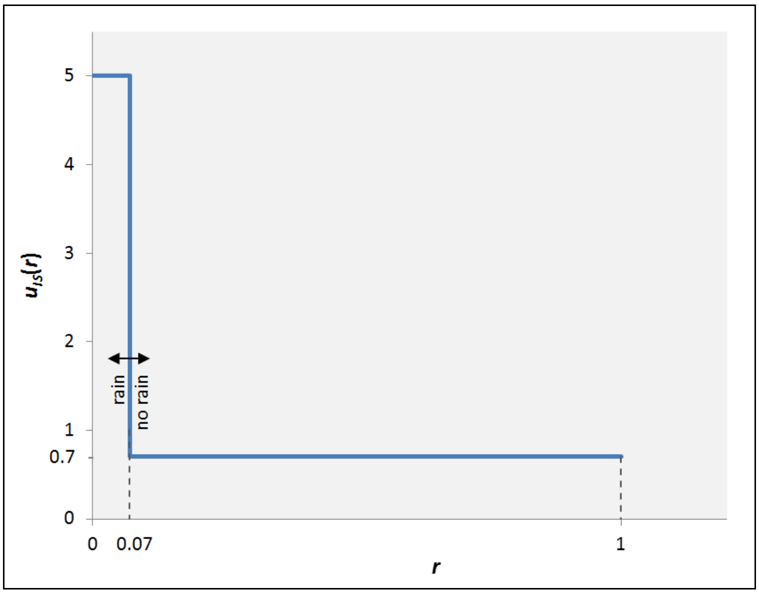
This is shown in Figure 6.6.

Figure 6.6: Simulation of the occurrence of rainy weather using a random number between zero and one, as applied within the context of a standard Monte Carlo simulation approach



Now we can artificially increase the probability of occurrence of rainy weather in the simulation process by manipulating the probability distribution of *R*, which is illustrated in Figure 6.7.

Figure 6.7: Simulation of the occurrence of rainy weather using a manipulated random number between zero and one, as applied within the context of the importance sampling approach



Imagine that we would draw the random number r = 0.04 from this manipulated probability distribution for a certain simulation run j (meaning that the occurrence of rainy weather will be assumed for this specific simulation run). If none of the other random number variables were manipulated, the outcome of this simulation run is then to be weighted (relative to the outcomes of other simulation runs) with a correction factor w_j given by:

$$w_j = \frac{u(r)}{u_{IS}(r)} = \frac{u(0.04)}{u_{IS}(0.04)} = \frac{1}{5} = 0.2$$

Similarly, if a random number r > 0.07 would have been drawn (corresponding to the situation 'no rain'), the outcome of the concerning simulation run would have to be weighted with a correction factor w_j given by:

$$w_j = \frac{u(r)}{u_{IS}(r)} = \frac{u(r > 0.07)}{u_{IS}(r > 0.07)} = \frac{1}{0.7} \approx 1.43$$

Of course, the main question is now: which of the various random number variables (each of which belongs to a certain source of variability in traffic demand and/or supply) should be manipulated, and how, in order for the *indirect* manipulation of the travel time distributions to be most effective (i.e. achieving the most significant reduction in the required number of simulation runs)? Unfortunately, this question cannot be answered. Therefore, this would have to be found out by a process of trial and error. Obviously, the *absolute optimum* will certainly not be found then. This however is not required for the method to have a significant effect.

7. Model validity

7.1 Introduction

A model is a simplified representation of a part of reality. In order to be able to make sound inferences with such a model, it has to be sufficiently valid for the task at hand. In this chapter, the validity of the developed model is discussed. This model should be sufficiently valid for showing what kind of additional insights may be obtained (into the relative importance of different primary sources of congestion, as well as into the effectiveness of proposed measures aimed at alleviating congestion) when the inherent variable nature of daily traffic congestion is explicitly taken into account.

The validity of a simulation model (for a specific task) is determined by its inputs, parameters and underlying theories and assumptions. Three different levels of validity of a simulation model can be distinguished (Van Lint, 2008):

- face validity (or content validity)
- construct validity
- predictive validity (or generality)

These impose increasingly stronger constraints on the model.

In sections 7.2 - 7.4, the developed model is assessed on these three levels of validity. This is based solely on theoretical considerations. Normally, one would assess the final validity of a model (i.e. its predictive validity) by means of a quantitative validation procedure. In section 7.5 it is argued, however, that the model at hand cannot be quantitatively validated in the usual way. Yet, some quantitative considerations are given in this section. These are considerations of a more general nature, relating to the computed congestion levels.

7.2 Face validity

A model is said to be **face valid** if its equations, parameters and characteristics are logically related to the characteristics of the system at hand and if it encompasses the minimally required detail to tackle the problem in question[94]. Apart from any possible remaining bugs in the model (for which a thorough check of its code is strongly desirable), it can be argued that the model is largely face valid for the task at hand (i.e. showing the gain of additional insights). This is ensured by the inclusion of all relevant sources of variability, and the use of a traffic flow modeling approach which is consistent with first order traffic flow theory (to which the capacity drop phenomenon has been added).

[94] definition taken from Van Lint, 2008

This does not imply that the model is *completely* face valid, however. Two important deficiencies can be identified:

- The absence of feedbacks from the actual traffic conditions to the traffic demands. In section 5.2.1 it was already discussed that the desired model property of realistically dealing with the route choice effects of traffic congestion (i.e. combining pre-trip and en-route route choice behavior) was sacrificed in favor of other desired model properties. The feedback effects on the other travel choices (i.e. trip making decisions, destination choices, mode choices and departure time choices) are not included either.
- The lack of a sound theoretical basis for the way of modeling the part of the capacity variations that is due to the intrinsic randomness in human driving behavior. This relates to one specific aspect of these variations, namely their spatial dependencies. This problem has been explained in chapter 6 (section 6.2). Note that some other imperfections discussed in that chapter will affect the face validity as well. The two deficiencies mentioned here are considered the most important ones, however.

These two deficiencies cannot easily be remedied. Including the route choice effects of traffic congestion is currently not possible in terms of required computation time. Incorporation of the effects on the other travel choices (trip making decisions, destination choices, mode choices and departure time choices) is unfeasible due to the lack of sufficient knowledge on these effects. This knowledge can only be obtained by conducting dedicated research. The elimination of the second deficiency would require substantial further research as well.

7.3 Construct validity

The second level of validity is **construct validity**. A model is construct valid if it is face valid and if its parameter values and inputs are mutually consistent and consistent with observations in reality[94]. Provided that the model is face valid, this construct validity can be obtained by calibrating the model to real-life observations (i.e. tuning the inputs and model parameters in such a way that the outputs of the model match the real-life observations to a satisfactory degree).

In principle, an extensive calibration procedure would not be necessary for the research tasks at hand. For these research tasks, construct validity is *relatively* easily accomplished. This is because of the fact that it is not intended to come up with firm quantitative inferences with respect to a specific existing situation. Instead, it is only aimed for to *illustrate* the gain of any possible additional/revised insights. The only requirement is then that the situation considered *could have been* a real-life situation. Consequently, all inputs and model parameters can be given any value within the range in which they occur in reality.

This turned out not to be as easy as it seems, however. For many of the parameters describing the effects of the different sources of variability values reported for other countries had to be used (because of a lack of information specifically relating to the situation in the Netherlands), or even *assumed* values (because of a general lack of knowledge). Consequently, it cannot be excluded that part of these parameter values are actually *outside* their 'realistic range'. Although to a more limited extent, the same applies to part of the parameters describing the frequencies/patterns of occurrence of the different sources of variability.

There are two basic strategies to overcome this (i.e. to enhance the construct validity of the model):

- Conducting additional research into the parameters in question.
- Calibration of the model (i.e. fitting to observations in reality).

Note that a calibration of the model could not be limited to part of the parameters only. In this case, *all* inputs and parameters would have to be adapted to the specific real-life situation considered for the calibration. Otherwise, inadequacies of the values which are not calibrated would hamper the calibration of the others. Also note that for the calibration to a specific real-life situation, in fact, the model should be extended with additional parameters. This is because of the fact that in real life, certain sources of (temporal) variability are clearly variable over space[95]. Consequently, we cannot suffice with one (uniformly applied) parameter for the influence factors in question, when calibrating the model to a real-life situation. For the research tasks at hand (which relate to the *temporal* variability), the consideration of these spatial variations is not very important, which was the reason for not including them in the model.

Obviously, the first strategy is very time-consuming, and likely to give a less optimal end result than the second one. Yet we might have to put up with it, however, because of the fact that the feasibility of the second strategy is very doubtful. This is due to the following:

1. The deficiencies with respect to the face validity (discussed above) will prohibit a proper calibration of the model parameters. There is a risk of unintentionally compensating these deficiencies by errors in the calibrated values of certain model parameters, which is obviously very undesirable (because of the associated deterioration of the model's predictive validity).
2. The mathematical feasibility of the calibration procedure is to be doubted. This is because of the large number of parameters involved, and the large number of simulations that would have to be performed for each iteration of the calibration process again.

[95] Clear examples of this are the spatial variations in the demand effects of vacations and in the frequency of occurrence of incidents.

3. A very long homogeneous series of empirical data would be required to obtain a sufficient statistical accuracy. Such a series is not available, however, due to:
   - (gradual) trends in especially the traffic demands (but in supply characteristics as well)
   - more abrupt changes in the traffic system, due to infrastructure changes (including dynamic traffic management measures) or changes in the spatial setting (such as the realization of a new residential area)
4. The problem is likely underdetermined. After all, there is a very large number of input variables and model parameters to be fitted, while there are only a few output indicators (which are, moreover, only partially independent of each other). Of course, the number of output indicators could be increased by considering a larger number of different statistical indicators, or by considering the route-based indicators for a larger number of different routes. The addition of more and more indicators will however be increasingly less effective, due to the dependencies between the different indicators.

   If the problem is underdetermined, the model will be over-fitted. The calibration result makes little sense then, resulting in a very limited predictive validity (which might even be lower than in the situation in which part of the parameter values are based on data from other countries or on common sense assumptions).

The last two problems might be (partially) overcome by calibrating the model at a lower level. That is, not at the level of the *final model outputs* (i.e. the considered indicators), but rather at the level of the *results of the individual simulations.* In this case we would use the model to compute the traffic conditions for *empirical scenarios* (reflecting the various sources of variability), and tune these computed traffic conditions to their empirical counterparts, by adjusting the model parameters that represent the effects of the sources of variability. Note that the model would have to be adapted for this, because of the fact that the influencing circumstances are input here, while normally being part of the output as well. Also note that the model parameters relating to the frequencies/patterns of occurrence of the different sources of variability are not considered in this calibration approach. Therefore, these would have to be considered separately in this case (using empirical data on these frequencies/patterns of occurrence).

By calibrating at the level of the results of the individual simulations, a lot more independent observations are obtained (one set of observations per individual simulation run), while the number of parameters to be fitted remains the same. This would solve problem 4. Problem 3 would be partially solved as well: although still a very long series of empirical observations would be required, this series would not have to be completely homogeneous anymore. Some changes in the traffic system could be compensated for, by adapting the model inputs (such as the network definition) for the simulation runs concerned.

The first two problems are not remedied by this alternative calibration approach, however. This means that even if this alternative approach would be used, a proper calibration seems unfeasible.

7.4 Predictive validity

The third level of model validity is **predictive validity**. A model has predictive validity if the model is both face and construct valid and if the predictions made with the model are consistent with the observed evolution of the system[94]. Obviously, the predictive validity of a model will not be better than its construct validity (which in turn is limited by any possible deficiencies in the face validity). As a result, the critical remarks regarding the face and construct validity of the model will apply equally well to its predictive validity. The relevance of the distinction between construct validity and predictive validity lies in the fact that if the model has been calibrated to a certain dataset (in order to achieve construct validity), it might still perform poorly in predicting the system behavior under slightly different circumstances (for instance due to over-fitting, as discussed above). The developed model has not been calibrated to empirical data, however. Therefore, the distinction between construct and predictive validity is not really relevant here.

7.5 Quantitative considerations on the model validity

In the above, the validity of the model has been discussed on the basis of theoretical considerations only. In fact, it would be desirable to assess this validity from a more practical perspective as well, by considering to what extent the indicator values computed for some (real-life) test scenario are consistent with or different from their empirical counterparts. From the theoretical considerations it is clear that that there will be differences, but not to what extent. However, it does not seem very worthwhile to perform such a quantitative validation procedure if the model has not been calibrated for the real-life situation in question[96] (which is considered unfeasible, as explained above).

Yet, some quantitative considerations are given in this section. These are considerations of a more general nature, relating to the computed congestion levels. Based on some test runs (using the motorway network around the city of Rotterdam), two important observations have been made regarding these congestion levels:

- In many of the traffic simulations, the computed traffic congestion seems unrealistically heavy for the circumstances at hand (i.e. the stochastic realizations of weather, incidents, road works, day of the week, month of the year, etc.). Certain types

[96] Note that it is not meant here that the model should be calibrated on the same *dataset* as the one that would be used for the validation. Obviously, for the calibration and validation of a model two separate datasets have to be used.

of circumstances even frequently result in the occurrence of a gridlock in the traffic network.

- The traffic congestion generated during the peak hours dissolves (much) too late in the traffic simulations.

These two observations and their (possible) explanations will be discussed in the two subsections below.

7.5.1 Unrealistically heavy traffic congestion

Introduction

In many of the traffic simulations, the computed traffic congestion seems unrealistically heavy for the circumstances at hand (i.e. the stochastic realizations of weather, incidents, road works, day of the week, month of the year, etc.). This does not seem to be a general problem of a too high base level of traffic demand or too unfavorable basis values for the traffic supply characteristics, however. This is concluded from the fact that the amount of traffic congestion for the representative situation appears to be reasonably realistic. This for example can be observed from Figure 8.4 and Figure 8.5, presented in chapter 8.

Possible causes

Different possible causes for the too heavy traffic congestion have been identified. First of all, there are the two deficiencies with respect to the face validity of the model, which were discussed in section 7.2:

- The absence of feedbacks from the actual traffic conditions to the traffic demands.
- The lack of a sound theoretical basis for the spatial dimension of the capacity randomization.

It is evident that the absence of feedbacks to the traffic demands will play a role in the problem. In reality, travelers will react to 'abnormalities' in the traffic conditions, by adapting their travel choices, as has been described in chapter 2. The larger these 'abnormalities' are, the larger the fraction of travelers to which this applies. The most important effect probably consists of changes in route choices, which are made in order to circumvent congested network parts. Obviously, these demand effects will have a limiting influence on the amount of traffic congestion. As discussed in section 7.2, these effects are not incorporated in the model, however (and cannot easily be added either). This will result in the model calculating too much traffic congestion, in particular for situations with a more serious local disruption, like an incident or road works.

It is uncertain to what extent the second deficiency plays a part in the problem. In section 6.2 it was discussed that the *minimum* of the capacity realizations for all cells of a link is decisive for the amount of traffic that can traverse that link without inducing congestion. The smaller the cell length of the capacity variation is chosen (corresponding to a larger number of independent capacity realizations per link), the lower the expected value of this minimum will be. Due to this effect, the cell length of the capacity variation is an important factor in the resulting traffic congestion levels. This is indeed observed if capacity realizations are uniformly applied to links *as a whole* (rather than generating multiple capacity realizations for different cells of a link). In this case, the amount of 'excess congestion' is substantially lower.

This observation does however not automatically imply that this *link*-based capacity variation is *more realistic* than the currently implemented *cell*-based variation (or, in other words, that this cell-based capacity variation is a cause of the too heavy traffic congestion). If the capacities would be varied at the level of *links*, this would correspond to assuming a full dependency between the capacities at the different locations along a link. Such a full dependency is rather unrealistic, as explained in section 6.2. It could of course be, however, that the current modeling approach (which assumes complete dependence *within* cells, and complete independence *between* cells) does not adequately reflect the *partial* spatial dependency involved. Maybe larger cell sizes should be used for the randomization, or a modeling approach in which the dependency is gradually reduced with increasing distance, according to some correlation function. More clarity on this can however only be obtained by new research into the properties of the spatial dependencies.

Other possible causes for the too heavy traffic congestion have been sought in:

- the inclusion of the capacity drop in the traffic flow model (section 5.2.3)
- the way of introducing the stochastic capacity realizations into the traffic flow modeling (section 5.3.3)
- the mismatch between the demands and the capacity realizations (section 6.3)
- the absence of physical traffic flow separations in the network definition of the test network
- the random generation of road works (section 5.4.13)
- the model parameter settings (section 7.3)

By comparing the traffic conditions in simulations *with* and *without* the capacity drop, it was found that this capacity drop does not contribute importantly to the problem of the too heavy traffic congestion. As a matter of fact, the effects of the capacity drop even turned out to be smaller than expected. This might be related to the problem discussed in section 6.4.

The role of the way in which the stochastic capacity realizations are introduced into the traffic flow modeling appeared not really important either. This was studied by employing alternative methods for this, which have been discussed in section 5.3.3. These alternative methods turned out to decrease the amount of traffic congestion somewhat, but not to a very important degree.

Section 6.3 discussed the problem that the 5-minute capacity realizations are confronted with traffic demands that inevitably develop variations with the time step of the traffic flow simulation (which is much smaller). Due to the non-linearity of the traffic system (meaning that a temporary capacity shortage is not cancelled out by an equal capacity surplus in another time interval), this will result in extra congestion, although it is not clear to which extent. It is estimated that the effect is relatively limited, however.

Another possible cause of the too heavy congestion is found in the specification of the network used in the test runs (being the motorway network around the city of Rotterdam). In this network definition all motorways are modeled as a single roadway per direction. In reality, however, many motorways in this network consist of multiple physically separated roadways per direction. Due to the absence of this physical separation in the model, traffic congestion spreads out over the network more easily than in reality. This might be a significant contributory factor to the discrepancy between the (high) levels of traffic congestion observed in the model and those observed in reality.

A more specific partial cause of the problem is found in the stochastic simulation of (short-term) road works. These road works are regularly observed to have an unrealistically severe impact on the traffic conditions in the network. To an important extent, this can be attributed to the fact that the model only to a limited degree takes into account that road works in reality are very carefully planned, taking into account the expected traffic conditions during the road works.

The model *does* account for the fact that there are less road works going on during the busier times of the day, by using:

- an empirical probability distribution for the *starting time* of the road works, and
- an empirical probability distribution for the *duration* of the road works, which is *conditional on the generated starting time*.

For the severity of the road works (referring to the number of lanes that are closed) an empirical probability distribution is used as well. This distribution is *not* conditional on the time of the day, however. This is an important limitation, since it results in the simulation of severe road works on busy times of the day, which is clearly unrealistic. Another important limitation is that it is not accounted for that some roads have less spare capacity than others, which influences the likelihood of the execution of road works during certain parts of the day.

Finally, any possible inadequacies in the model parameter settings may play a role in the problem as well. In section 7.3 it was explained that it cannot be excluded that some of the parameter values are outside their 'realistic range'. It was also explained here that this problem can only be remedied by conducting (time-consuming) dedicated research into the parameters in question, since a proper calibration of the model is considered unfeasible.

Gridlocks
For certain types of unfavorable demand/supply realizations, the too heavy congestion even often ends up in the occurrence of a gridlock in the traffic network. In such cases a traffic jam on the (ring-shaped) test network grows so long that finally its tail reaches its own head, after which the traffic comes to a complete stand-still. Once such a ring-shaped traffic jam has arisen, it will sustain itself, since car drivers basically are 'waiting for themselves'. In the vast majority of cases, the gridlock it is triggered by one or more incidents and/or road works. Here probably the absence of a feedback loop to the route choices of the road users and the difficulties in proper simulating road works play a dominant role.

In reality, true gridlocks do not occur, or at least will resolve within a limited period of time. This is due to the fact that road users will change their routes if traffic conditions on their originally intended route get too adverse, which is not accounted for in the model.

It might be clear that traffic simulations in which a gridlock occurs cannot give any useful output anymore. The travel times for example go to infinity in such a situation. Therefore, the model is programmed to remove the output data of such simulations from the output database on which the final statistical evaluations are performed. While this may currently be the only solution, it actually is a very undesirable one, since it is expected to create a huge bias in the statistical output indicators. After all, the gridlock phenomenon does not occur randomly over the different simulation runs, but typically in one specific subset of these, namely the subset of simulations with relatively unfavorable realizations of certain demand/supply characteristics.

Because of this, it is considered absolutely necessary to prevent these gridlocks from occurring. This can only be achieved by integrating a traffic-responsive route choice component in the model. For this, the hybrid route choice modeling approach described in (Pel et al, 2009) is advised. This is however technically not yet feasible within an acceptable computation time. Therefore, it is advised to study possibilities for reducing the required calculation time.

7.5.2 Traffic congestion dissolving too late

Besides the problem of the traffic congestion being too heavy in many of the simulation runs, another problem that was observed is that the traffic congestion generated during the peak hours dissolves (much) too late. Obviously, to some extent the problem discussed above will play a role in this. After all, the heavier the congestion is, the more time it will take for the traffic system to recover from this congestion. Another possible cause for this discrepancy with reality might be found in the phenomenon that on motorways often an early peak in the traffic demand is observed, caused by travelers that depart early, in order to pass through certain bottlenecks before the daily traffic jams emerge. This phenomenon is not taken into account in the model.

Although this issue is something to be aware of when interpreting the model results, it is considered not very problematic for the research tasks at hand.

7.6 Conclusion

Although the model is considered largely valid for the research tasks at hand (in which it is not aimed for to come up with firm quantitative inferences regarding specific existing situations), there are some deficiencies which cause too much congestion to be generated in many of the simulations. These are deficiencies in the mechanisms included in the model, but possibly also some inadequacies in the model parameter values. It has not been possible to solve this problem within this project, since most of these deficiencies require substantial further investigation.

Of course, this does not mean that these deficiencies can simply be ignored in the analyses with the model. They will be respected in the following ways:

- Incidents and road works will be left out of the analyses, in view of the apparent inability of the model to deal with these in a valid way.
- All model results will have to be interpreted with caution, taking into account that too much congestion is generated in many of the simulations (even if incidents and road works are excluded).

8. Model results

8.1 Introduction

In this chapter the results obtained with the developed model are presented. It should be noted however that only 350 simulation runs were performed per evaluation, in order to be able to generate some model output within an acceptable period of time. Obviously, this number of simulations is way too low to obtain results with any reasonable statistical accuracy / reliability. Therefore, all model results presented below should be interpreted with care. It is believed, though, that the presented mutual comparisons of different scenarios can still provide some valuable qualitative information, as a result of the modeling efforts to ensure comparison under ceteris paribus conditions (i.e. the efforts to make sure that for the simulations of different scenarios exactly the same pseudo-random numbers are used at the same places within the model[97], irrespective of the user control settings with respect to the 'activation' of the different sources of variation).

In the next section first of all a description is given of the network for which the model evaluations have been performed. Section 8.3 deals with the indicators that have been considered. After that, section 8.4 discusses the model outputs for the representative situation (i.e., a 'nominal' weekday, without any variable influences except for a deterministic traffic demand variation with the time of the day). This typically is the output obtained by a traditional calculation of the traffic conditions in a network. Next, section 8.5 treats the results obtained by a calculation in which all different sources of variability are taken into account, and compares these with the output for the representative situation. Section 8.6 then shows that the relative importance of these different sources can be studied by deactivating them in the model. This provides insights which cannot be obtained with a traditional evaluation. Section 8.7 finally considers the effects of a rush-hour lane, as computed with the new model, in which various sources of variability are taken into account, and compares these with the effects that would have been found with a calculation according to the traditional approach (considering a representative situation only).

The chapter ends with a section on the practical applicability of the developed model. In this section it is discussed whether/how this model could be used for practical application within the context of real-life evaluations of measures proposed to alleviate traffic congestion.

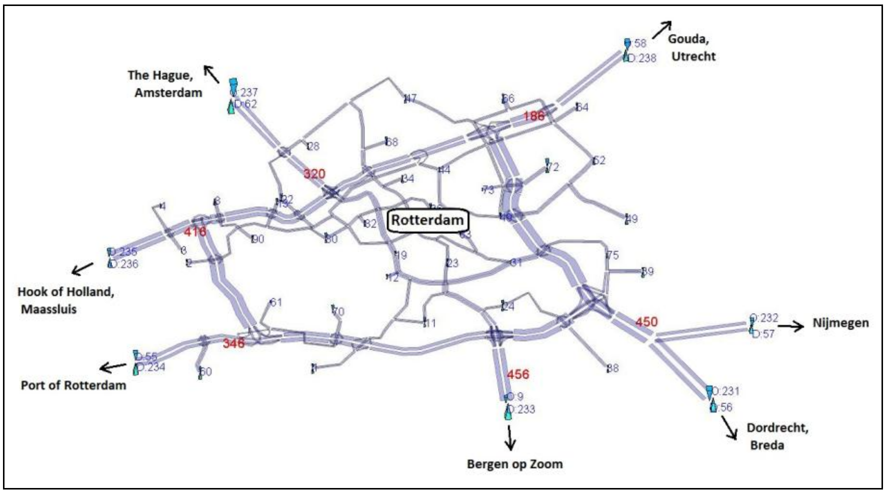
[97] except for the modeling of the intrinsic randomness in the travel demands, as discussed in section 6.5

8.2 Test network

For the quantitative evaluations the motorway network around the Dutch city of Rotterdam (see Figure 8.1) was selected as test network. This choice was mainly made because of the facts that this network:

- was readily available in JDSMART (including an accompanying origin-destination demand matrix, albeit a static one, which covers the morning peak only),
- is a reasonably sized real-life network, and
- has a layout which is quite representative for road networks with a lot of congestion problems, as found in many large urban areas (i.e., a layout with many on- and off-ramps and interchanges, potentially giving rise to blocking back effects).

Figure 8.1: The test network selected for the quantitative evaluations



It should be noted that it would have been better to use a larger sized network, in order to reduce the relative influence of its boundaries. Inevitably, these network boundaries will affect the results of the quantitative evaluations, because of the associated neglect of all network effects that 'come from outside'. Probably, this will artificially lower the variability in the (calculated) traffic conditions. The use of a larger sized network was prohibited by the requirement of a limited calculation time, however.

Note that there is no need for a thorough calibration of the model for the situation around Rotterdam. A crude calibration (i.e. making the computed network traffic conditions roughly resemble those observed in reality) was already performed in the research project from which the network and origin-destination matrix have been taken (Zuurbier et al, 2006). Even though the route choice modeling was different here, this is considered sufficient. After all, for addressing the research questions at hand it is not necessary to 'match' an actually existing situation. In fact, considering a complete fictitious situation would do as well, as long as it 'could have been' a real-life situation. The advantage of using a real-life network however is that this latter (i.e. that it 'could have been' a real-life situation) is made much more likely then, even without an extensive calibration. Obviously, this will only be true if the (representative) demand and supply settings make sense. It is assumed that this is ensured here by the crude calibration mentioned above.

As shown in Figure 8.1, besides the motorway network the test network also contains the most important roads of the secondary network. Since the focus is on the motorway network, and the route choice effect of traffic congestion (i.e. the effect that road users might deviate to other roads in case of significant disruptions on their 'standard' routes) is not considered (because of the inability to properly model this; see section 5.2), this is actually of little use, while contributing significantly to the required computation time. However, it should be appreciated that it would be a fairly laborious task to remove these secondary roads from the network, because of the fact that the (destination-specific) traffic demands from the origins on the secondary network then would have to be translated into corresponding (destination-specific) traffic demands at the on- and off-ramps of the motorway network. The difficulty here is in the fact that one origin (destination) does not have a one-on-one correspondence with a certain on-ramp (off-ramp).

8.3 Indicators considered

In chapter 3 it was extensively discussed which performance indicators should be considered in an evaluation in which the probabilistic nature of traffic congestion is explicitly taken into account. This finally resulted in a set of selected indicators, given in section 3.4.

In the quantitative evaluations presented in this chapter, not all of the indicators from this set are considered, however. This is directly related to the fact that the developed model does not include a feedback loop from the traffic conditions to the traffic demands. It obviously does not make sense then to consider indicators VII (the number of travelers on origin-destination relations) and VIII (the total number of vehicle-kilometers traveled on the network), which were selected precisely because of the existence of this feedback loop.

The various travel time statistics are considered in a simplified way. In chapter 3 it was argued that in fact two different types of travel time statistics should be considered, corresponding to two extremes regarding route choice:

- assuming a *fixed route choice* (i.e. assuming that the road users in all situations hold on to their standard/intended routes), and
- assuming an *optimal route choice* (i.e. assuming that the road users at all times are able (and willing) to select the route that will yield them the shortest travel time).

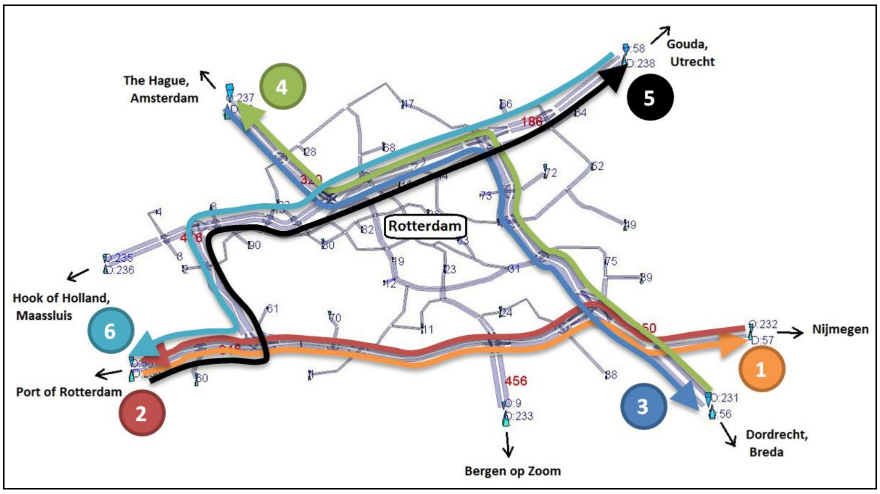
However, in the developed model the travel time statistics are only computed for *given* routes, corresponding to the assumption of a *fixed* route choice.

Six different routes are considered here (three in two directions):

1. A15 Eastbound
2. A15 Westbound
3. A13 → A20 → A16
4. A16 → A20 → A13
5. A15 → A4 → A20
6. A20 → A4 → A15

These routes are depicted in the figure below. Together, they cover a large part of the motorway network.

Figure 8.2: The six considered routes



Finally a remark should be made with respect to public holidays and vacation periods. In section 3.4 it was indicated that the travel times on public holidays in fact should be included in the travel time distributions for Sunday, rather than in the travel time distributions for the actual day. Further it was indicated that vacation periods actually should be excluded from the day-to-day travel time distributions. For reasons of simplification, here these two matters have been disregarded, however.

8.4 Representative situation

In the tables and figures below, the model outputs obtained for the *representative* situation are shown. These outputs represent the traffic system's performance according to the traditional evaluation approach, in which only this representative situation is considered. All sources of variability are ignored here, except for the systematic demand variation with the time of the day.

Please note that the travel time statistics for specific times of the day are not defined for the representative situation, due to the fact that they refer to a variability which is not modeled in this case. Similarly, the travel time instabilities cannot be given as distributions in this case, but only as deterministic values.

INDICATORS I–V (TRAVEL TIME STATISTICS):

Statistics of the overall travel time distribution:

Table 8.1: Representative values of the statistics of the overall travel time distributions

Route / Indicator	1	2	3	4	5	6
I. TT_90	1.15	1.19	1.10	1.10	1.11	1.11
II. TT_mean	1.11	1.08	1.05	1.05	1.05	1.05
III. TT_median	1.07	1.08	1.05	1.05	1.05	1.05
IV. TT_width	0.15	0.19	0.09	0.10	0.10	0.11
V. TT_skew	1.10	1.54	0.84	1.02	1.14	1.17

Note: all values of indicators I-IV are made dimensionless by division by the free flow travel time. Indicator V is already dimensionless by itself.

Statistics of the travel time distributions for specific times of the day:

TT_{90}, TT_{mean}, TT_{median}, TT_{width} and TT_{skew} are not defined for specific times of the day. Instead, only the representative travel time values can be given. These are shown in the figure below:

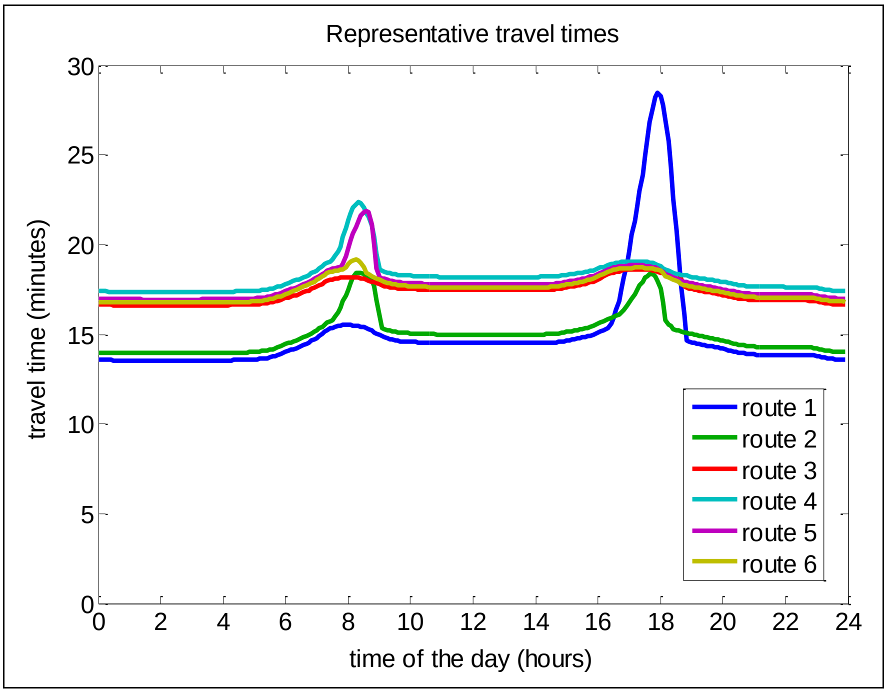


Figure 8.3: The representative travel times as a function of the time of the day (representing the travel times as obtained by a traditional performance evaluation)

In all travel time curves clear morning and evening peaks can be observed. For some of the routes these peaks in travel time are more pronounced than for others, which is logically connected to the presence or absence of bottlenecks on these routes. Further, it can be observed that some of the bottlenecks are active in both peak periods, while others are active in the morning or evening peak period only. The travel times clearly are lowest during the night.

INDICATOR VI (TRAVEL TIME INSTABILITY):

Table 8.2: Representative values of indicator VI (reflecting the travel time instability)

Route / Time of the day	1	2	3	4	5	6
03:00	0.00	0.00	0.00	0.00	0.00	0.00
07:00	0.01	0.01	0.01	0.01	0.01	0.01
08:00	0.00	0.01	0.00	0.03	0.04	0.02
09:00	-0.01	-0.04	0.00	-0.04	-0.08	0.00
13:00	0.00	0.00	0.00	0.00	0.00	0.00
16:00	0.01	0.01	0.00	0.00	0.01	0.01
17:00	0.08	0.03	0.00	0.00	0.00	0.00
18:00	0.00	-0.04	0.00	-0.01	-0.01	0.00
21:00	0.00	0.00	0.00	0.00	0.00	0.00

INDICATOR IX (NUMBER OF LOST VEHICLE HOURS):

LVH = 2644 hours (over a 24-hours period)

In order to give an impression of the spatial characteristics of the traffic congestion in the representative situation, the two figures below show the traffic states at 8:30 and 17:30, respectively. Green colors indicate *free flowing* traffic, while red colors indicate *congested* traffic.

Figure 8.4: The representative traffic situation at 8:30 (with green colors indicating free flowing traffic, and red colors congested traffic)

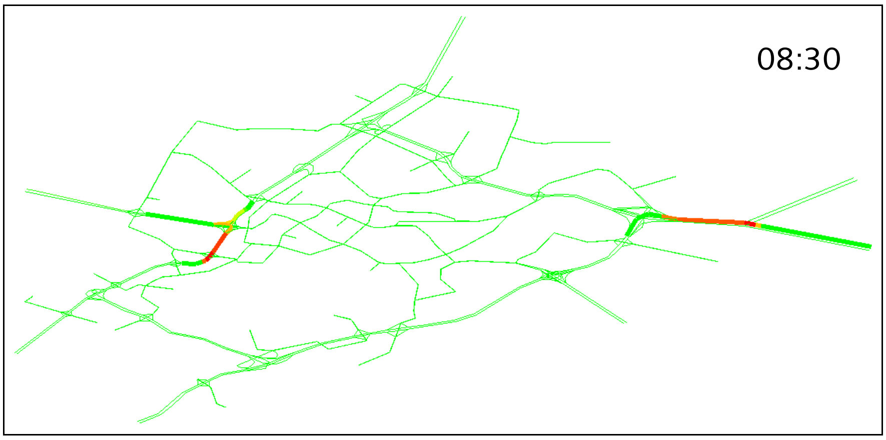
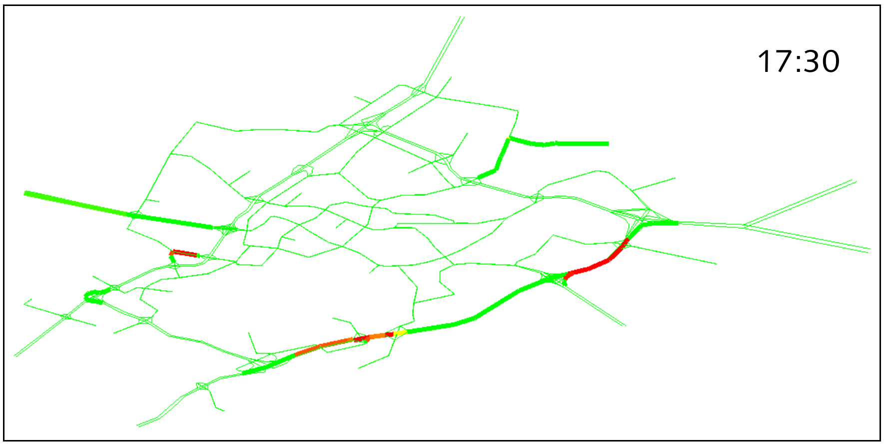


Figure 8.5: The representative traffic situation at 17:30 (with green colors indicating free flowing traffic, and red colors congested traffic)



8.5 Results of the model run with full variability

8.5.1 Introduction

In this section, the results of the model run with full variability (i.e. including the various sources of variability) are provided. In the remainder of this chapter, these results are used as a kind of ‘benchmark’ for analyzing the relative contributions of the different primary sources of congestion, and the effectiveness of a given traffic measure.

It should however be noted that two sources of variability have been omitted in this model run, namely road works and incidents. In section 7.6 it was concluded that these are better left out of the analyses, in view of the apparent current inability of the model to deal with these in a valid way. The inclusion of road works and/or incidents results too frequently in excessive traffic congestion being created (often giving rise to the occurrence of a gridlock). It is impossible then to make any sound inferences.

It should also be noted that the results of 3 of the 350 simulations had to be excluded from the model output, because of the fact that a gridlock occurred (in spite of the exclusion of incidents and road works). This problem has been discussed in section 7.5.1.

Section 8.5.2 presents the obtained results, and compares these with the 'representative' calculation results. Section 8.5.3 then draws some conclusions with respect to the observed differences, and tries to explain these. Finally, section 8.5.4 provides an analysis of the sensitivity of the results to an important model uncertainty.

8.5.2 Results

INDICATORS I–V (TRAVEL TIME STATISTICS):

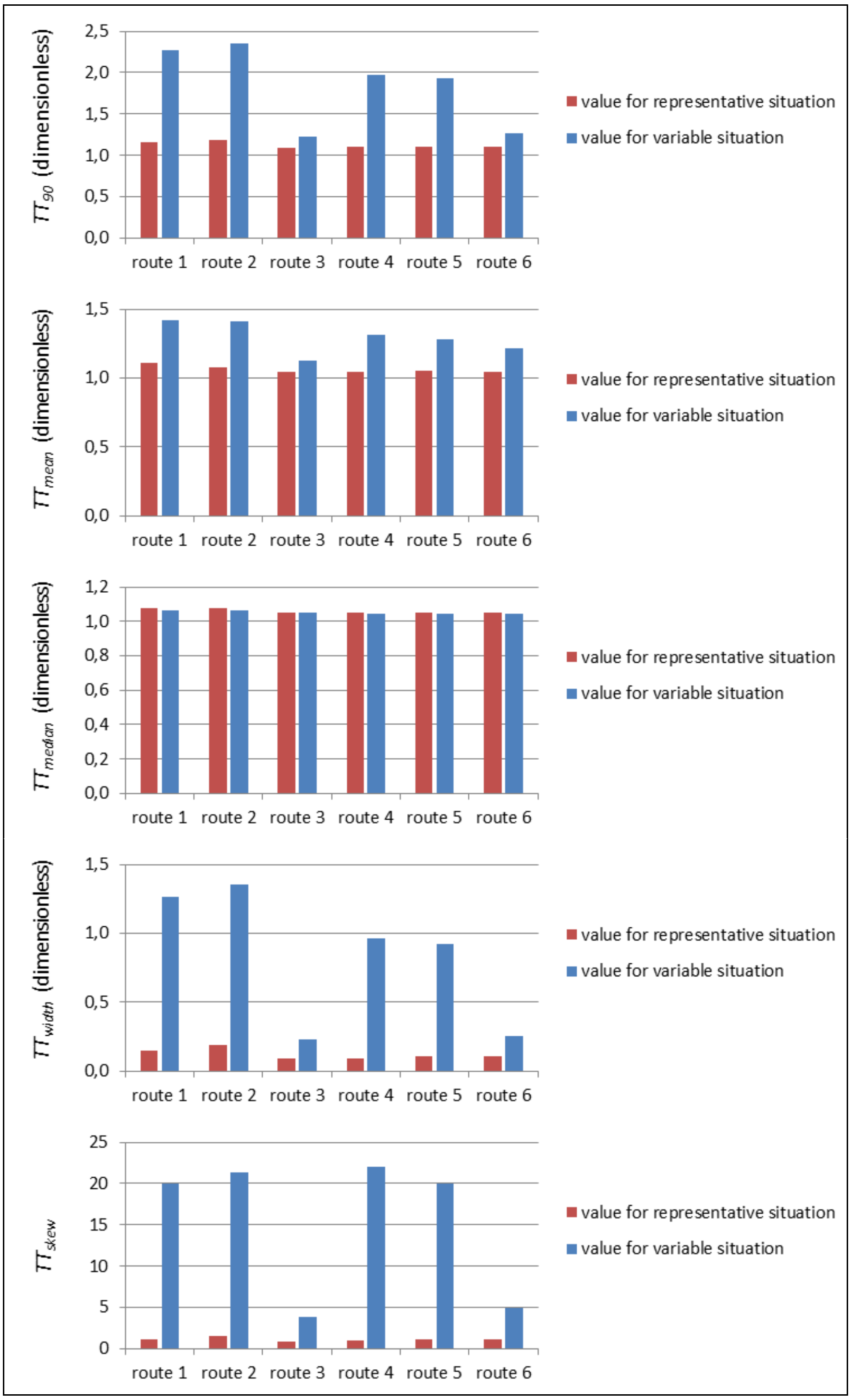
Statistics of the overall travel time distribution:

In the figure below, the computed statistics of the overall travel time distributions are compared with their 'representative' counterparts (provided in the previous section). Note that for the computation of these statistics only *one* travel time value per simulated day was used (per route), in order to avoid introducing a bias due to dependencies between the travel time values. The times of the day for which these travel times were taken were randomly selected per simulated day.

Clearly, the 90th percentile and mean travel times are much larger if the different sources of variability are accounted for. For the width and especially the skew of the travel time distributions this is even more so. It should be noted here, however, that the differences will be overestimated in the figure, due to the problem discussed in section 7.5.1 (i.e. the fact that the calculated traffic congestion is too heavy in many of the simulations). Especially the skew of the distributions will be sensitive to this problem.

The median travel times are not really affected by the inclusion of the various sources of variability. Here it should be noted, however, that this is probably partly due to the fact that weekend days are also included now, while the 'representative' situation in fact reflects a weekday. If the weekend day travel times would have been excluded from the computation of the statistics, probably the median travel times would have been (somewhat) higher than their 'representative' counterparts as well.

Figure 8.6: The computed statistics of the overall travel time distributions (90th-percentile, mean, median, width and skew), compared with their 'representative' counterparts

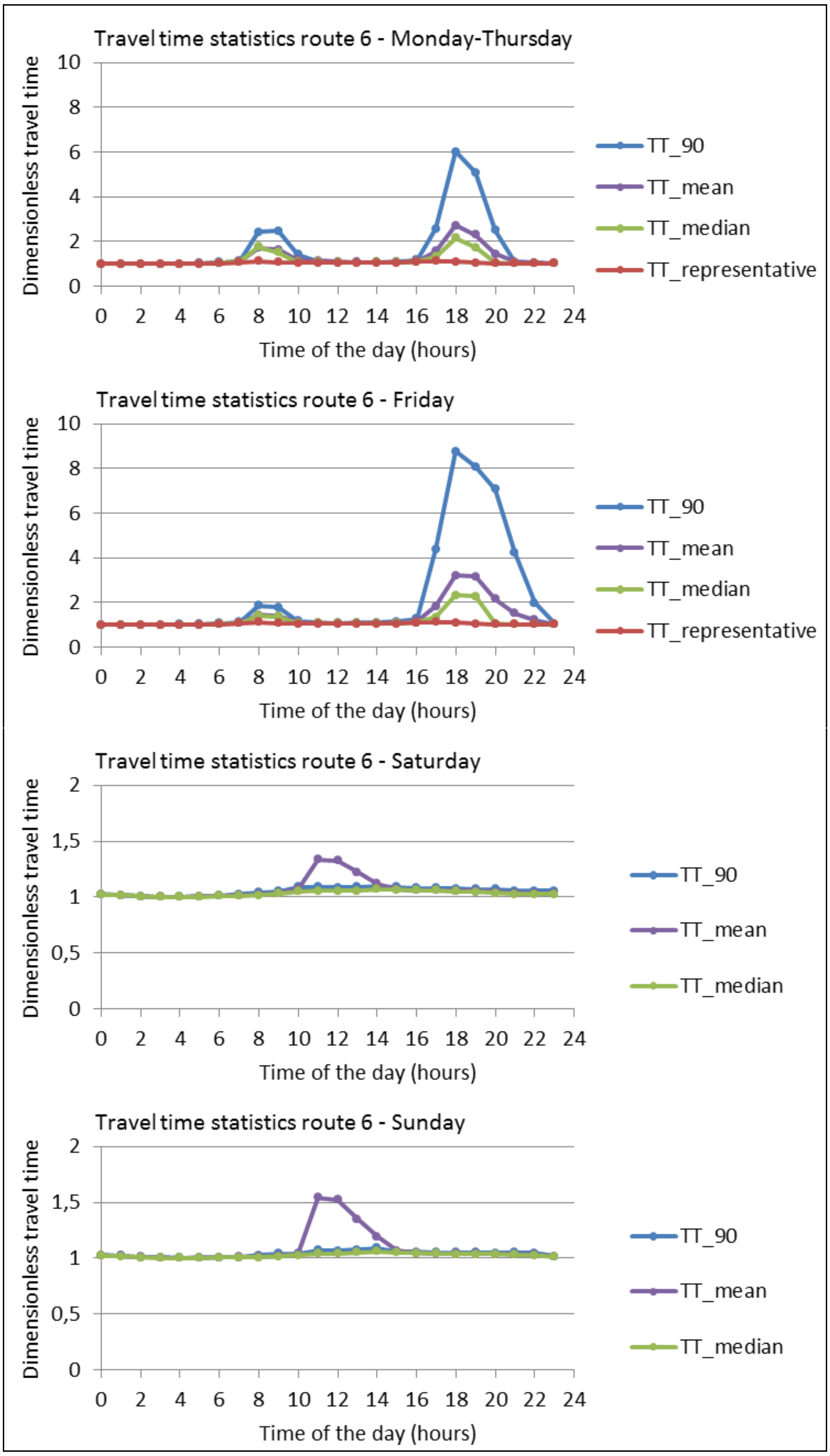


Note: all values of the first four indicators have been made dimensionless by division by the free flow travel time. The fifth indicator (i.e. the skew) is already dimensionless by itself.

Statistics of the travel time distributions for specific times of the day:

By way of example, below the travel time statistics I-III for one of the different routes are shown, as a function of the time of the day (separately for the different categories of days). Please note that for Saturday and Sunday a different scaling for the vertical axis has been used.

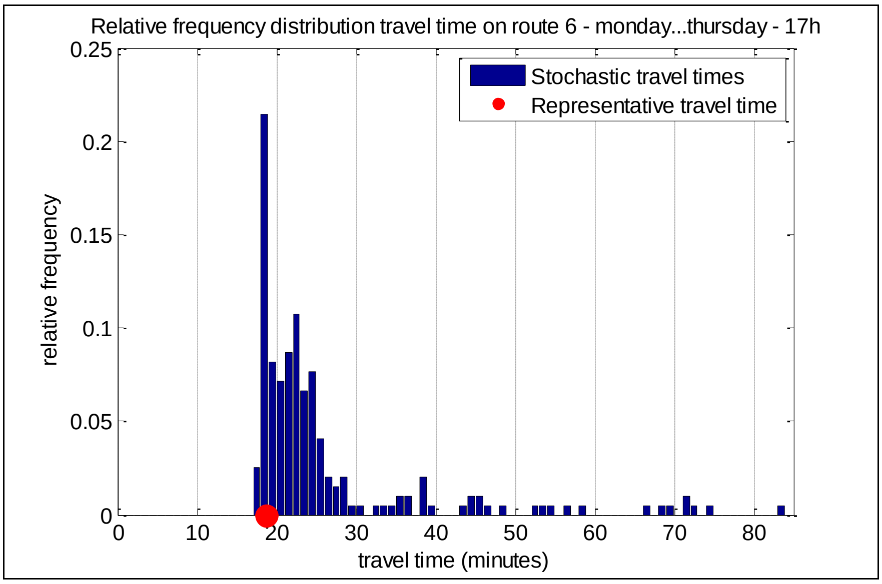
Figure 8.7: The travel time statistics (90^{th}-percentile, mean and median) computed for route 6, as a function of the time of day and the category of days. For the purpose of comparison, the representative travel time is included as well (for weekdays only).



The figure below provides an example of a travel time distribution computed for a particular time of the day. In Appendix 3 some travel time distributions for other times of the day are given.

Note that some values are very high indeed (especially if one considers that road works and incidents have not been included in the simulations). This is related to the simulation problem discussed in section 7.5.1.

Figure 8.8: Distribution of the travel time on route 6 at 17:00, computed for Monday – Thursday



In the figures it can clearly been seen that the representative calculation does not provide a good impression of the situation on this route. It predicts hardly any congestion, while the travel time statistics indicate that frequently a significant delay might be experienced.

In the evening peak, the mean travel time is clearly observed to be larger than the median travel time. This is obviously related to the skewness of the travel time distribution. The 90th percentile travel time reaches very high values during the evening peak. Such high values seem unrealistic. They can be related to the problem mentioned above. One of the causes is that it is not accounted for in the model that if the traffic conditions get really bad, travelers will reconsider their travel choices (like route and departure time), resulting in lower traffic demands and consequentially lower travel times.

On Fridays the morning peak clearly results in less congestion than on the other weekdays, while the evening peak causes more congestion problems. For Saturdays and Sundays little congestion is calculated. Remarkable is that for the middle of these days the mean travel time is found to be larger than the 90th percentile travel time. This must be due to some (relatively) large peaks, which occur in less than 10% of the cases.

Below, for the same route the travel time statistics IV and V (i.e. the distribution width and skewness) are shown, as a function of the time of the day. Note that the distribution widths for Saturday and Sunday have been omitted. The reason for this is that these widths are very small throughout the day (with maxima of only a few percent of the free flow travel time).

Figure 8.9: The width of the travel time distribution computed for route 6, as a function of the time of day (separately for Mon-Thu and Fri)

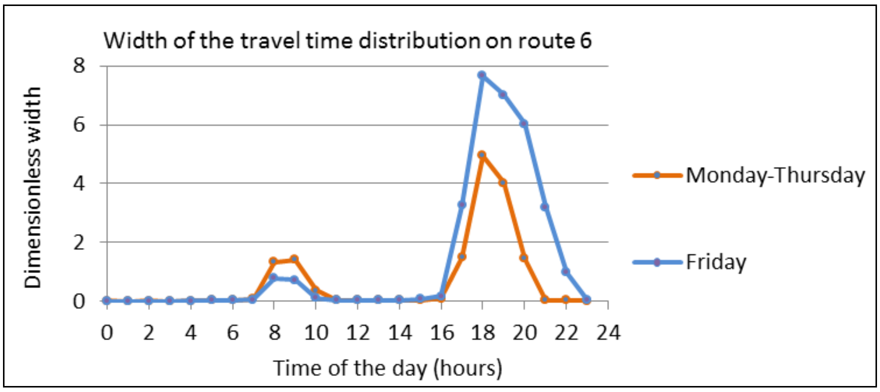
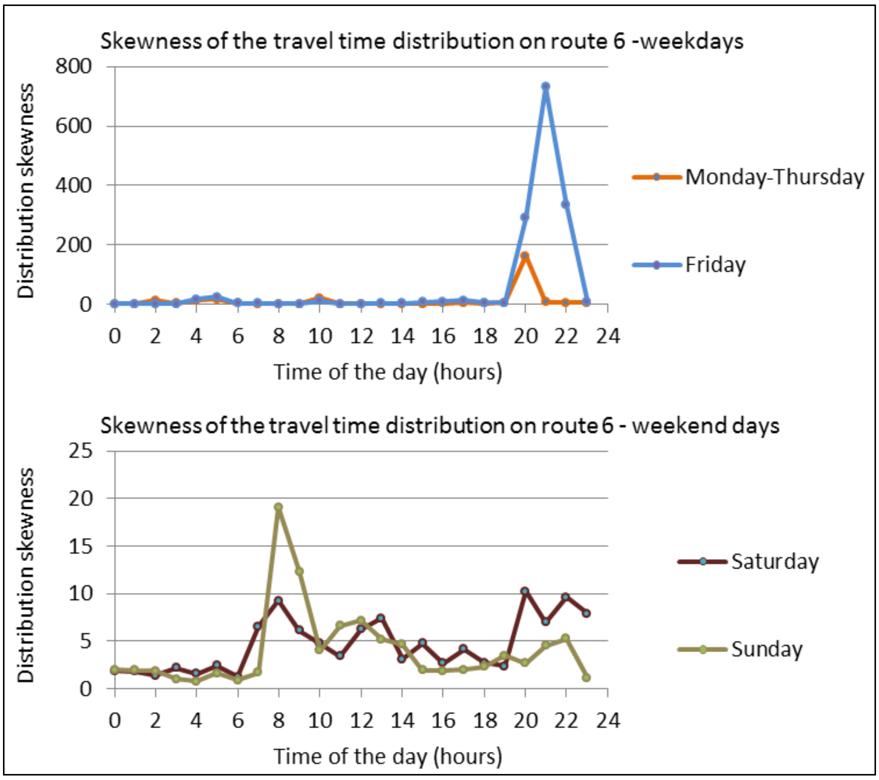


Figure 8.10: The skewness of the travel time distribution computed for route 6, as a function of the time of day (separately for the different categories of days)



For weekdays, it is found that the distribution skewness is largest during the periods *before* and *after* the peak periods. After the evening peak extremely high values are reached. This will undoubtedly be caused by the fact that on some days the evening peak congestion is not yet resolved then, while on most days practically all congestion has disappeared already. This results in a large difference between the 90th and 50th percentile travel times, while the difference between the 10th and 50th percentile travel times is almost zero. Since the skewness is computed as $(TT_{90}—TT_{50})/\ (TT_{50}—TT_{10})$ (see chapter 3), this will obviously result in extremely large skewness values.

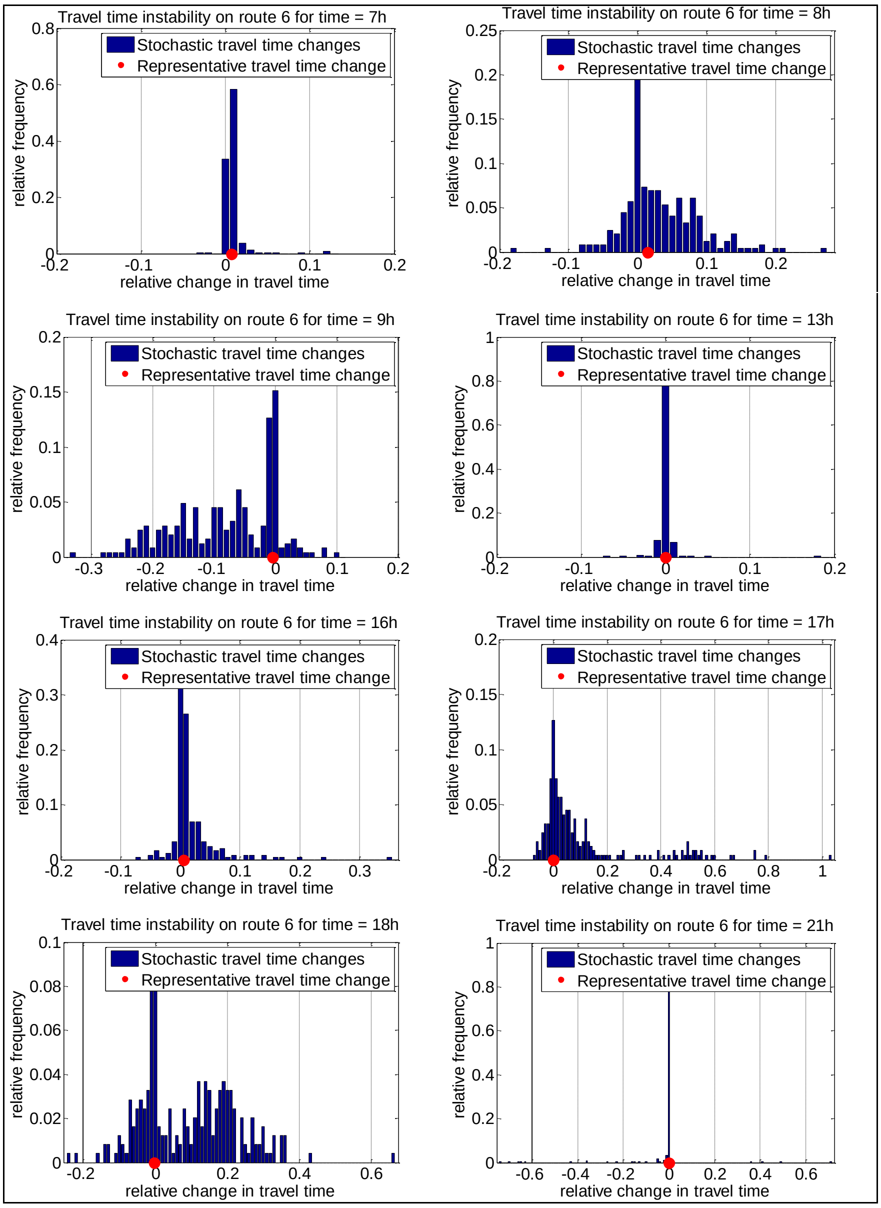
For the other five routes qualitatively similar pictures are obtained for the different travel time statistics as for the route considered above, except for the facts that:

- On part of the routes the daytime off-peak period is associated with higher (dimensionless) 90^{th}-percentile and mean travel times than the route considered above.
- On some of the routes the mean travel time is sometimes found to be lower than the median.
- On part of the routes the morning peak is associated with more congestion than the evening peak (instead of the other way around).
- On some routes also in the daytime off-peak period extremely high skewness values are reached.

INDICATOR VI (TRAVEL TIME INSTABILITY):

The figure below shows the computed diagrams of the travel time instability for different times of the day (relating to weekdays), again taking route 6 as example.

Figure 8.11: The travel time instability on route 6, for different times of the day (for weekdays only)



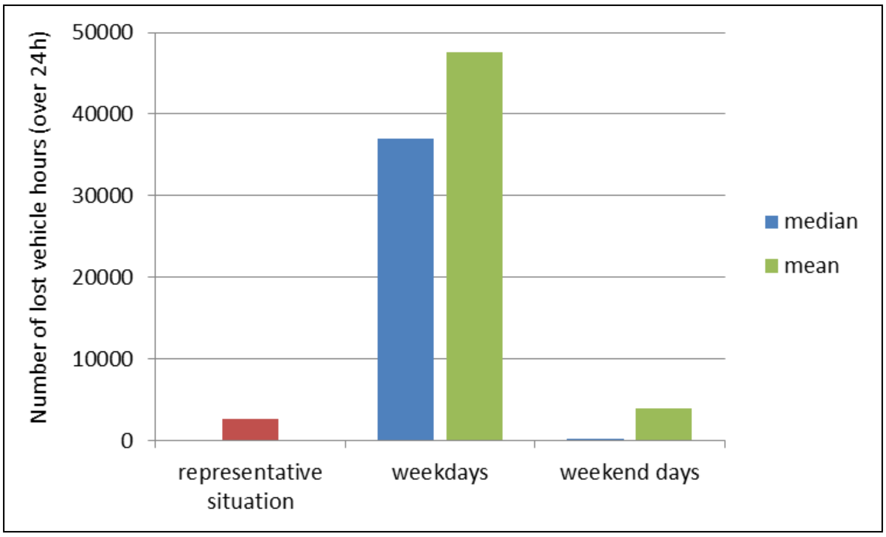
From the figure it can be seen that the travel time is most unstable during the peak periods and (shortly) afterwards. The results clearly indicate that the traffic conditions may change much faster during one's trip than reflected in the results of the representative calculation. It can also be seen that the changes in travel time are highly uncertain.

The result of this instability in the travel times is that road users cannot accurately predict their future travel times from the traffic conditions at the moment of departure. This results in costs to society, related to late arrivals and unused buffer time, as discussed in chapter 3.

INDICATOR IX (NUMBER OF LOST VEHICLE HOURS):

The figure below shows the amounts of lost vehicle hours (incurred within the boundaries of the network), separately for weekdays and weekend days. The representative calculation clearly results in an underestimation of the amount of traffic congestion. It the figure the difference is exaggerated, however, due to the problem discussed in section 7.5.1.

Figure 8.12: The computed values of the mean and median numbers of lost vehicle hours (incurred within the boundaries of the network), compared with the number of lost vehicle hours in the representative situation



The figure shows that the mean number of lost vehicle hours is larger than the median value. This indicates that its probability distribution is skewed. This is logically related to the fact that traffic conditions can be very bad (virtually without any upper bound), but not better than free flow.

8.5.3 Discussion

From the provided results it is clear that the 'representative' calculation does not give a good impression of the performance of the traffic system. This is not only due to the obvious fact that the (day-to-day) uncertainty aspect of this performance is disregarded (due to the neglect of the day-to-day variability in the traffic conditions). Also, the 'representative' calculation turns out to result in an underestimation of the congestion indicators. That is, the traffic congestion calculated for the 'representative' situation (i.e. the situation in which all demand and supply characteristics are at their 'representative' level, which for example could be the mean or median value) is not so 'representative' itself.

Two explanations can be given for this. The first explanation can be found in the fact that some of the sources of variability affect the traffic conditions *by nature* purely in a negative way. Examples of these are adverse weather conditions and incidents. By neglecting such sources of variability, the traffic congestion is obviously underestimated by the representative calculation.

Other sources of variability do not by nature affect the traffic conditions in a purely negative way. Consider for example the capacity variation due to the intrinsic randomness in human driving behavior. Sometimes the capacity is below its mean (or median) value, resulting in heavier traffic congestion, but at other times it is above it, resulting in *less* congestion. Something similar applies in the spatial dimension: at some locations the capacity will be below its mean or median value (resulting in heavier congestion), whereas at other locations (in the network / along a route) it may be above it (resulting in less congestion)[98].

Often, these positive and negative realizations do not cancel out, however. This is due to the non-linearity in the traffic system. This non-linearity is in the facts that:

- Traffic conditions can be very severe (virtually without any upper limit), but conversely not better than free flow.
- A capacity exceedance that is twice as large (or twice as long in duration) causes more than twice as much delay.

Due to this non-linearity in the behavior of the traffic system, it is likely that the detriments of the 'negative occurrences' (i.e. lower capacity realizations or higher demand realizations) are larger than the benefits of the 'positive occurrences' (i.e. higher capacity realizations or lower demand realizations). This way, sources of variability may have a *net negative* impact on the traffic conditions, even if these do not act by nature in a purely negative way. This is the second explanation for the fact that the neglect of the variability results in an underestimation of the values of the congestion indicators.

[98] Obviously, this only applies to the extent that the capacities (and demands) at different locations are independent from each other.

From the presented results it can be observed that the mean indicator values generally differ more from the representative values than the median indicator values. This is due both to the fact that the traffic system behaves in a non-linear way (as explained above) and to the fact that some sources of variability occur only on an occasional basis (and consequently hardly affect the median traffic conditions).

In practice, the underestimation by the 'representative' calculation would typically be counteracted to some extent by calibrating the traffic simulation model to an (empirical) congestion level that is considered representative. In such a calibration procedure, the neglect of the variability is partly compensated for by adjusting the values of the inputs and/or model parameters (usually mainly the bottleneck capacities). This is clearly not an optimal way of 'accounting for the variability', though. First of all, the calibrated model can be expected to have a limited predictive validity. After all, under changed conditions (reflecting some future scenario to be evaluated with the model) the parameter adjustments needed to compensate for the variability might be different from those found for the calibration case.

Secondly, it might be impossible to compensate for the variability (by means of parameter adjustments) in a consistent way. This depends on what kind of level is considered representative (e.g. the *median* or the *average*). If the traffic model is calibrated in such a way that it properly reproduces the median congestion levels at the individual bottleneck locations, it cannot simultaneously reproduce the medians of route or network based congestion indicators in a proper way. This is due to the fact that the sum of the medians of the congestion levels will not be equal to the median of the sum of these levels. If *average* levels are considered representative (instead of *medians*), this is different, however. After all, the sum of the averages of variables is always equal to the average of the sum of these variables.

Even apart from these two issues, however, such a calibration will not satisfactorily resolve the shortcomings of the 'representative' evaluation approach that have been identified above, because:

- In terms of the average performance, the traffic congestion will still be underestimated. This is due to the fact that the empirical congestion level to which the model is calibrated is typically not the overall average level, but rather a level which is considered representative for *regular* circumstances (i.e. excluding situations with disturbances such as incidents and possibly also bad weather events).
- The (day-to-day) uncertainty aspect of the performance remains ignored in the evaluation.

8.5.4 Sensitivity to the spatial dependencies in the capacity randomness

In section 6.2 and chapter 7, the lack of a sound theoretical basis for the modeling of the spatial dependencies in the part of the capacity variability that is due to the intrinsic randomness in driving behavior was identified as a deficiency in the model. Because of the fact that the model outcomes appear reasonably sensitive to the spatial dimension of these dependencies, a small sensitivity analysis has been performed on this aspect.

The dimension of the spatial dependencies is reflected in the size of the cells that are used for the capacity randomization. *Within* these cells, the local capacities are assumed to be fully dependent. This is reflected in the fact that per cell one single capacity value is drawn (from the probability distribution function of the capacity), which is assumed to apply to the whole of the cell. *Between* the different cells, complete independency is assumed. This is reflected in the fact that the capacity values for the different cells are independently drawn from the probability distribution.

In the model, the cell size for the capacity randomization is chosen equal to the cell size of the numerical solution scheme of the dynamic traffic simulator. As discussed in section 6.2, this choice was a purely pragmatic one, related to the fact that no knowledge is available on the spatial length scales of the dependencies. In section 7.5.1 it was discussed that in many of the traffic simulations unrealistically heavy traffic congestion is generated. It was also explained here that, hypothetically, a (partial) cause for this *could* be that the cell size of the capacity randomization is currently too small (corresponding to too little spatial dependency in the capacity randomness). Therefore, in the sensitivity analysis a situation with a *larger* cell size was considered[99]. More specifically, the 'extreme' case of the cell size being equal to the *link length* was studied, which means that the capacities were varied at the level of links as a whole. This corresponds to the implicit assumption that the capacities are fully dependent over the whole length of the link. Although this is clearly not realistic, it should be noted that on the other hand the capacities of the different links are still completely independent from each other, whereas in reality there might be some (limited) dependencies between the capacities of adjoining links as well.

[99] Of course, this only relates to the cell size of the *capacity randomization*. The cell size of the numerical solution scheme of the dynamic traffic simulator is not altered.

The sensitivity analysis has been performed by recalculating all output diagrams presented in this section (i.e. Figure 8.6 – Figure 8.12). The results of this calculation with the modified model are provided in Appendix A4.1. From these results it is clear that in the traffic simulations generally much less traffic congestion is generated then, which is in line with the discussion in section 6.2 [100]. Since the (unmodified) model generates *too much* congestion in many of the simulation runs (as discussed in section 7.5.1), this decrease in the amount of congestion results in the traffic conditions looking more realistic. This observation does not automatically imply, however, that this modified way of modeling the random capacity variation is also more realistic *itself*. Theoretically, it could also be that this change in the model does not improve its realism, but just 'luckily' compensates for other deficiencies. More clarity on this can only be obtained by conducting dedicated research into the spatial dependencies in the random capacity variation.

Due to the fact that much less congestion is generated by the modified model, the quantitative differences between the variable case and the representative situation are *much smaller* then. Qualitatively, the main conclusion drawn in section 8.5.3 remains largely valid, however. It is still concluded that the 'representative' calculation does not give a good impression of the performance of the traffic system, which is not only due to the obvious fact that the (day-to-day) uncertainty aspect of this performance is disregarded, but also due to the underestimation of some congestion indicators.

This underestimation mainly applies to:

- *The amount of lost vehicle hours (Figure A4.7):*
  Both the mean and median (for weekdays) are significantly larger than the representative amount of lost vehicle hours.
- *The mean values of the day-to-day travel time distributions (Figure A4.2):*
  During the peak periods, the average travel time is well above the representative travel time[101].

[100] In this section, it was explained that the minimum of the capacity realizations for all cells of a link is decisive for the amount of traffic that can traverse that link without inducing congestion. The shorter the cell length of the capacity variation is chosen (corresponding to a larger number of independent capacity realizations per link), the lower the expected value of this minimum will be, resulting in more congestion. Since in this sensitivity analysis the cell length is chosen equal to the *link length* (corresponding to a situation with only *one* capacity realization per link), the expected value of this minimum is at its highest possible value here (i.e. simply equal to the expected value of the cell capacity distribution function itself). It is obvious that this may result in much less congestion being calculated.

[101] Note that in Figure A4.2 only one route is considered. It cannot be excluded that for other routes different findings would be obtained.

Important differences in the results are that:

- According to the outcomes of the modified model (Figure A4.2), the representative travel time is a very good representation of the median travel time during the evening peak and Friday morning peak, whereas according to the outcomes presented in section 8.5.2 (Figure 8.7) it significantly underestimates this median during these peak periods[101].
- Although it is still found that the statistics of the overall travel time distributions (i.e. the distributions relating to the combination of all days of week and times of day) are underestimated by the representative calculation, the differences are very small according to the outcomes of the modified model (Figure A4.1). This seems to contradict the *substantial* underestimations (in the amount of lost vehicle hours and the representative travel times) that were mentioned above. However, this small size of the differences can be logically explained from the facts that:
    - The different routes are free of congestion during by far the largest part of the time, both in the 'representative' situation and in the variable case. As a result, the median, mean and 90th-percentile travel times are all reasonably close to the free flow travel times, in both the representative situation and the variable case. As a result, all differences are relatively small. It should be noted, however, that – as far the mean and 90th-percentile values are concerned – the relative differences in terms of *delay* (rather than *travel time*) are substantial.
    - The overall travel time distribution obtained by the representative calculation refers to a ('nominal') *weekday*, while its counterpart for the variable case covers *weekend days as well*. This means that the comparison is in fact 'unfair', since weekend days are typically less congested.

8.6 Relative importance of different sources of variation

8.6.1 Introduction

This section illustrates that explicitly accounting for the different sources of variability can provide us with new insights into the relative importance of these different influence factors. By way of example, six of these factors will be mutually compared on their relative contributions to the congestion indicators. These are:

- the intrinsic randomness in human driving behavior,
- the intrinsic randomness in people's travel decisions,
- weather conditions,
- luminance conditions (daylight or darkness),
- events, and
- the demand variation with the month of the year.

The influences of these different contributory factors are analyzed in the following way:

- For each of these factors a new model run is performed, in which the concerning factor is 'switched off' (i.e. is omitted from the model).
- The results of these model runs are compared with those of the model run with full variability (i.e. including the various sources of variability – see section 8.5), which is used as a kind of 'benchmark'. That is, it is computed to what extent the various congestion indicators are changed by 'switching off' the different sources of variability.
- These relative changes in the indicators values are compared between the different sources of variability, in order to establish the relative influences of the latters.

The results of these analyses are presented in subsection 8.6.2. Next, subsection 8.6.3 tries to draw some general conclusions from these results, and discusses the practical implications of these conclusions. Finally, subsection 8.6.4 discusses the results of a small sensitivity analysis of the results, comparable to the sensitivity analysis discussed in section 8.5.4.

Please note that the relative importance of the different contributory factors is studied by *'switching off' the factor in question*, rather than by isolating it (i.e. 'switching off' all other contributory factors). The main reason for this is in the dependencies between the different sources of variability. Due to these dependencies, most sources of variability affect the traffic conditions not only by their 'own' (i.e. *direct*) effects on demand and supply, but also by their effects on other sources of variability (affecting demand and supply in an *indirect* way). If the (influences of) the different sources of variability would be considered in an isolated way, these latter effects would be ignored. This means that the comparison between the different sources of variability would be incomplete then.

Another important reason for using this approach of 'switching off' the factor in question (rather than isolating it) is that the occurring traffic congestion is the end result of the *combination* of all individual contributory factors. This combined impact is much larger than the sum of all individual impacts, determined by considering the individual factors in an isolated way (i.e. as if all other contributory factors did not exist). If the individual contributory factors are considered in an isolated way, some of these will even be found to have no influence at all, while playing a significant role in the combined impact of the different contributory factors. This means that the final impact of a contributory factor (i.e. its impact on the quality of the traffic conditions) also depends on the other contributory factors, even if there are no dependencies in occurrence and (demand/supply) effects.

For the same two reasons, the mutual comparison of the influences of the six contributory factors mentioned above cannot be based on model evaluations in which only these six factors are included. All other sources of variability have to be included as well, even though their relative contributions will not be studied. Again, incidents and road works are omitted, however, for the reason explained in the previous section.

8.6.2 Results

INDICATORS I–V (TRAVEL TIME STATISTICS):

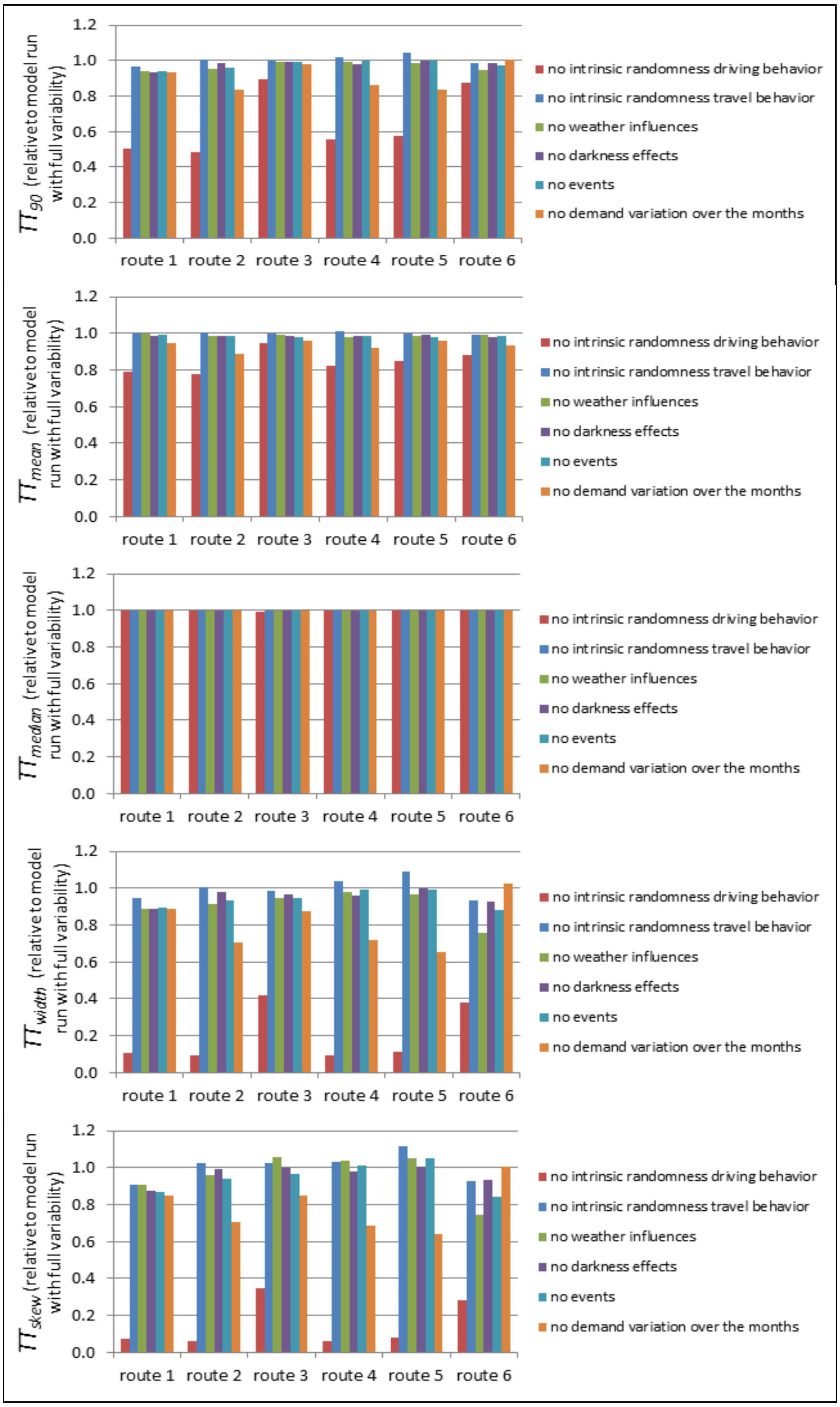
Statistics of the overall travel time distribution:

In the figure below, for each of the statistics of the overall travel time distribution the ratios are given of its 'modified values' (i.e. leaving out one of the sources of variability) and its 'benchmark value' (i.e. accounting for all sources of variability).

The intrinsic randomness in human driving behavior is clearly the most important contributor to the congestion problems, followed by the traffic demand variation over the months of the year. The other considered sources of variation turn out to have a much smaller influence.

From the figure it can be seen that the effects on the width and skewness of the distributions are the largest. The considered sources of variability do not have significant effects on the medians of the travel time distributions. This is logically related to the fact that here the *overall* travel time distributions are considered (i.e. of all days of the week and times of the day combined). Because of the fact that the largest part of the days is congestion free, these medians will always relate to congestion free situations as well. As a consequence they will not differ much from the free flow travel times, whether a certain source of variation is accounted for or not.

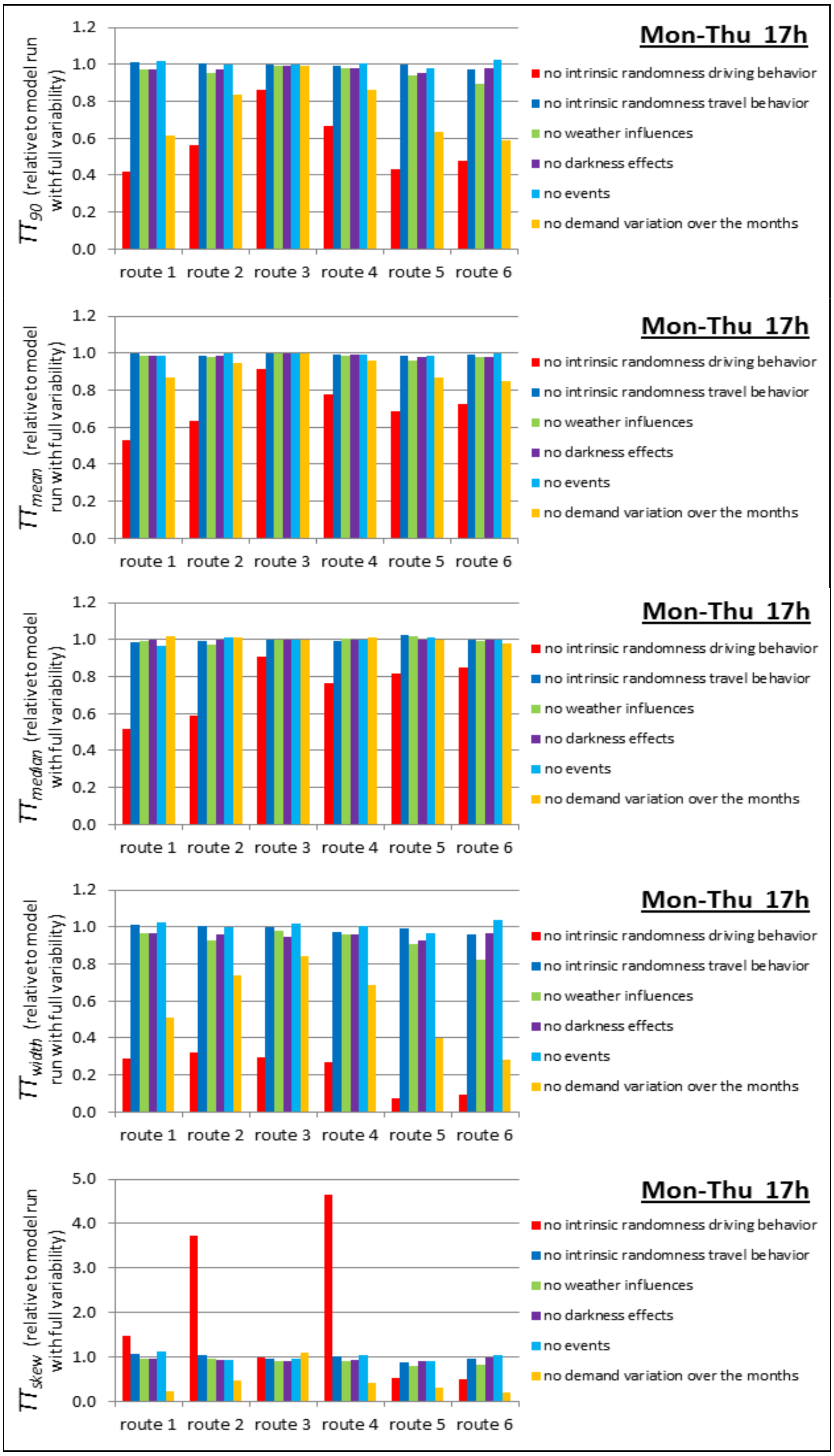
Figure 8.13: The statistics of the overall travel time distributions (90th-percentile, mean, median, width and skew), computed for situations in which one of the different sources of variability is omitted from the model. (All values expressed as a ratio to the value obtained from the model run with <u>full</u> variability.)



Statistics of the travel time distributions for specific times of the day:

By way of example, below the results for one of the different categories of days (that is, Monday-Thursday) and one specific time of the day (17:00 hours) are given. Again, for each of the travel time statistics the ratios are given of its 'modified values' (i.e. leaving out one of the sources of variability) and its 'benchmark value' (i.e. accounting for all sources of variability).

Figure 8.14: The travel time statistics for Mon-Thu – 17:00 (90th-percentile, mean, median, width and skew), computed for situations in which one of the different sources of variability is omitted from the model. (All values expressed as a ratio to the value obtained from the model run with full variability.)

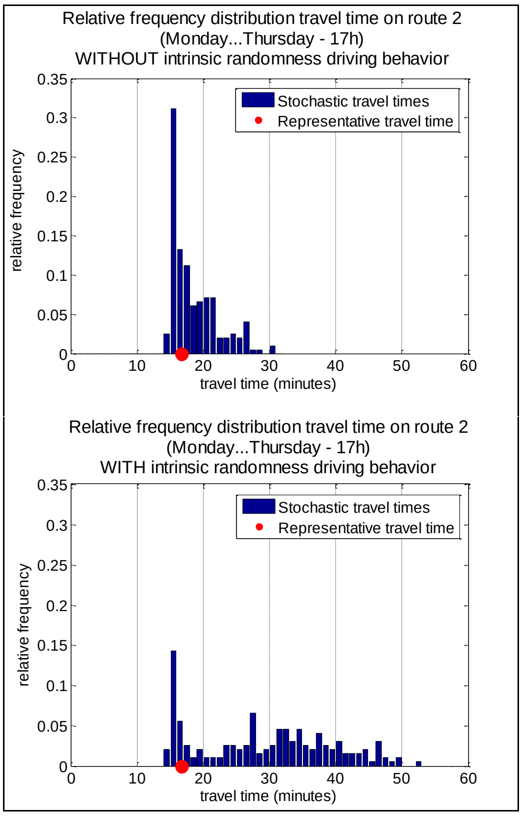


From this figure more or less the same conclusion can be drawn with respect to the relative importance of the different factors as the one that was drawn in relation to the *overall* travel time distributions. That is, the intrinsic randomness in driving behavior turns out to be the most important contributor to the congestion indicators, followed by the demand variation over the months of the year. Apparently, the other considered sources of variation do not contribute very substantially to the congestion indicators.

There are, however, also some differences to be observed:

- The intrinsic randomness in driving behavior and especially the demand variation over the months generally have a larger relative impact on the travel time distributions for this specific time of the day, than on the *overall* travel time distributions, which seems logical.
- The intrinsic randomness in driving behavior *does* affect the medians of the travel time distributions for this specific time of the day, while it does *not* affect the medians of the *overall* travel time distributions. This is logically related to the fact that the 'median traffic situation' for this specific time of the day probably is a congested one, while the *overall* 'median traffic situation' is congestion free.
- For some routes the omission of the modeling of the intrinsic randomness in driving behavior is observed to result in a much higher skewness value, which means that the existence of this randomness actually *lowers* the skewness of the travel time distributions on these routes. This can also be seen from the difference between the two diagrams in the figure below.

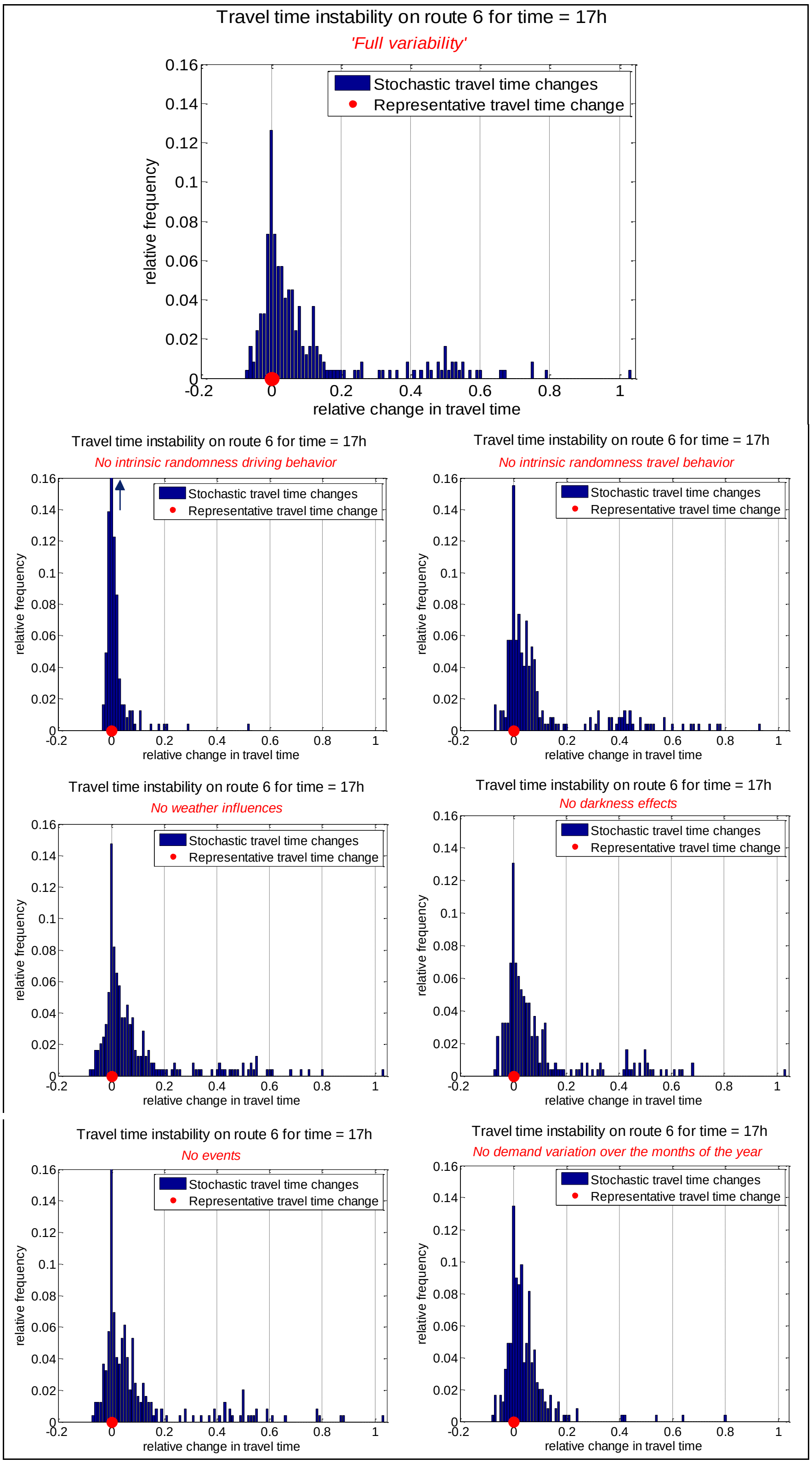
Figure 8.15: Computed distribution of the travel time on route 2 (for Mon-Thu 17:00), which is much less skewed if the capacity variation due to the intrinsic randomness in driving behavior is accounted for in the model (lower diagram) than if this variation is not accounted for (upper diagram)



INDICATOR VI (TRAVEL TIME INSTABILITY):

The figure below shows an example of the relative influences on the travel time instability.

Figure 8.16: Example of the effects on the travel time instability if given sources of variability are omitted from the model



Also for this indicator the intrinsic randomness in driving behavior proves the most important contributory factor. The demand variation over the months of the year turns out to have an important role in especially the right tail of the travel time instability graph. Apparently, the relatively high demands in some of the months may result in very sharply rising travel times at this time of the day.

INDICATOR IX (NUMBER OF LOST VEHICLE HOURS):

The figure below shows the relative effects of a number of different sources of variability on the number of lost vehicle hours (over a 24-hours period), as computed for weekdays. Again, the intrinsic randomness in human driving behavior is clearly by far the most important factor. The demand variation with the month of the year has an important influence on the mean number of lost vehicle hours as well.

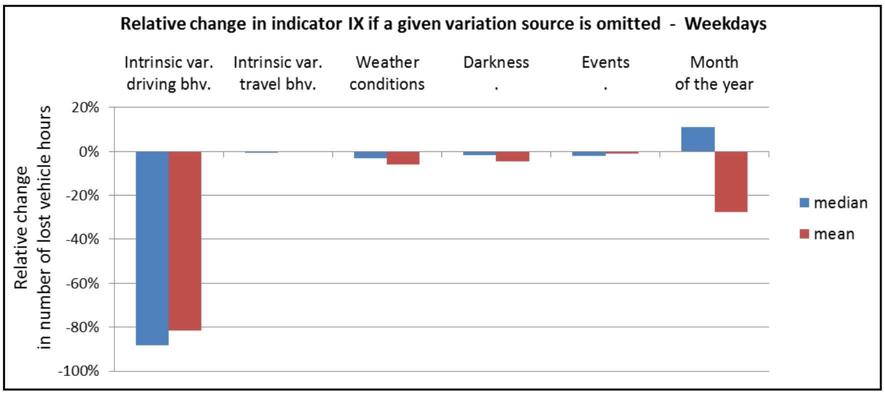


Figure 8.17: The relative changes in the mean and median numbers of lost vehicle hours if a given source of variability is omitted from the model (as compared with the numbers obtained from the model run with full variability), for weekdays

It might look surprising that the median number of lost vehicle hours is found to be larger (instead of lower) if the demand variation over the months is omitted from the model. This can probably be explained by the fact that the average of the monthly traffic demands (which is used as the *representative* monthly demand in the model) is larger than the median of these monthly demand values.

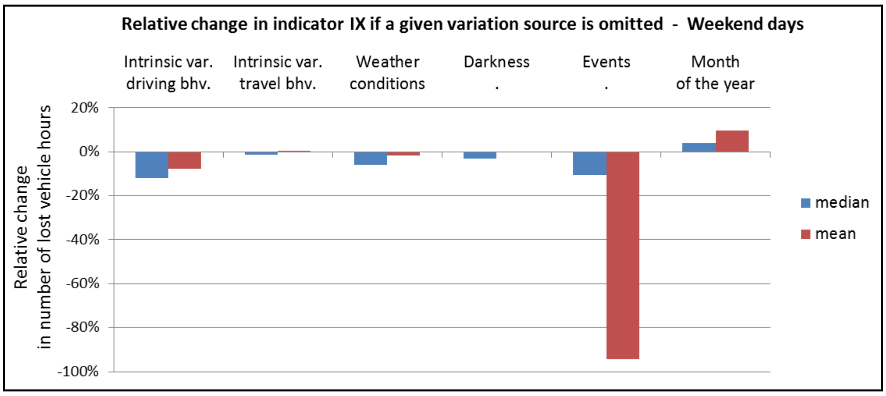
The impact of the intrinsic randomness in people's travel decisions turns out to be negligible compared to the impacts of the other sources of variability. At network level, the influence of events proves minor too. The relative contributions of the ambient conditions (i.e. weather and daylight/darkness) are small as well, though not negligible.

The relative contribution of weather conditions may seem smaller than expected. Here it should be considered, however, that the demand effects of adverse weather partially compensate for its negative supply effects. It should also be noted that the relative influence of weather conditions may actually be underestimated here, due to the following:

- As noted in section 8.1, the number of simulations performed (350) is way too low to obtain results with reasonable statistical accuracy/reliability (meaning that all presented model results have to be interpreted with care). By chance, this low number of simulations has resulted in the absence of heavy snowfall occurrences in the simulation results (while the expected value was about 3 occurrences). Since heavy snowfall is one of the most detrimental types of weather, its absence in the results might lead to a significant underestimation of the relative influence of weather conditions.
- As mentioned in section 8.5, three of the 350 simulations had to be excluded from the model output, because of the fact that a gridlock occurred. In two of these cases, adverse weather conditions were involved (moderate rain and heavy rain, respectively). By excluding these simulations, the relative influence of weather conditions is underestimated.

The figure below shows the same diagram for *weekend days*. On these days the relative influence of the intrinsic randomness in driving behavior appears to be much smaller. Here events seem dominating[102]. Obviously, events have a much larger effect on the mean than on the median traffic conditions, due to their limited frequency of occurrence.

Figure 8.18: The relative changes in the mean and median numbers of lost vehicle hours if a given source of variability is omitted from the model (as compared with the numbers obtained from the model run with full variability), for weekend days



[102] It should be noted, however, that some extremely high travel times were computed for certain event situations (in the order of magnitude of two hours). In reality these probably would have been lower, due to the fact that part of the travellers will reconsider their travel choices in such situations of heavy traffic congestion (which is not included in the model). As a result, in reality the relative importance of the influence factor 'events' might be somewhat less extreme than indicated in the figure.

8.6.3 Discussion

In the above, an example has been given of a comparison between the relative contributions of different sources of variability to the congestion indicators. Although the relative influences of only a part of all different sources of variability were considered, and only one specific spatiotemporal configuration (with respect to both the network and the traffic demand pattern), some general conclusions can already be drawn:

1. The capacity variations due to the intrinsic randomness in human driving behavior seem to play a central role in the (recurrent) peak period-related traffic congestion, considerably increasing both the:
   - mean and median levels of the amount of traffic congestion
   - variability of the amount of traffic congestion
   - instability of the traffic conditions (i.e. the extent to which one's realized travel time may deviate from the instantaneous travel time at the time of departure)

   Explanations for this large influence may be found in:
   - the considerable size of these variations (for the free flow capacity: coefficients of variation ranging from 5 to 12%, depending on the number of lanes)
   - the fact that the lowest cell capacity value is decisive for the capacity of the link (see section 6.2)
   - the fact that these intrinsic capacity variations are continuously present, unlike some of the other sources of variability

2. The seasonal demand variation over the months of the year seems to play a significant role in the peak period-related traffic congestion as well. Apparently the relatively high traffic demands in some of the months contribute importantly to the occurrence of heavy traffic congestion in these months, involving very high and unstable travel times.

3. The intrinsic randomness in travel behavior, the ambient conditions (weather and daylight/darkness) and events seem to play a much smaller role in the peak period-related congestion. The influences of events and the intrinsic randomness in travel behavior even seem negligible.

   The limited role of *events* will probably be mainly due to their limited frequencies of occurrence and more local nature. The limited contribution of *darkness* can be explained from the fact that its capacity effect is relatively small (1.5%).

   To some extent, the relative influence of *weather conditions* might have been underestimated in the presented analysis, as explained at the end of the previous subsection. Its limited influence will however also be largely due to the limit frequency of occurrence of special weather conditions. Due to this limited frequency of occurrence, the consequences of these weather conditions will hardly be reflected in indicators I, III, IV and V, and only to a limited extent in indicators II (the average travel time), VI (the travel time instability) and IX (the total number of lost vehicle hours).

It should be noted, however, that the top 10 most congested situations ever (on the Dutch main road network) can all be attributed to extremely adverse weather conditions. Such situations occur only very occasionally, but their disruptive effects on society are huge. Because of this latter, one could argue that these extreme situations are insufficiently reflected in the performance indicators. This is a direct consequence of the deliberate choice to use *robust* indicators, which are relatively insensitive to 'outliers' in the simulated data (see sections 3.3.2 and 3.4). This choice was made because of the facts that:

- For the same number of simulation runs, values of outlier-sensitive indicators have a considerably lower statistical accuracy/reliability. Therefore, the use of such indicators would make the results of the analyses vulnerable to randomness, especially if only a relatively small number of simulation runs is performed.
- The 'outliers' are likely to be inaccurate, since extreme circumstances will lead to a greater impact of model deficiencies.

Finally, the limited contribution of the intrinsic randomness in travel behavior can be attributed to the fact that its impact on the traffic volumes on the network is relatively limited. That is, the resulting fluctuations are small compared to the *average* traffic volumes, as illustrated by the left figure below. Also, these fluctuations are small compared to the random fluctuations in the *capacities*, which can be seen by comparing the two figures below.

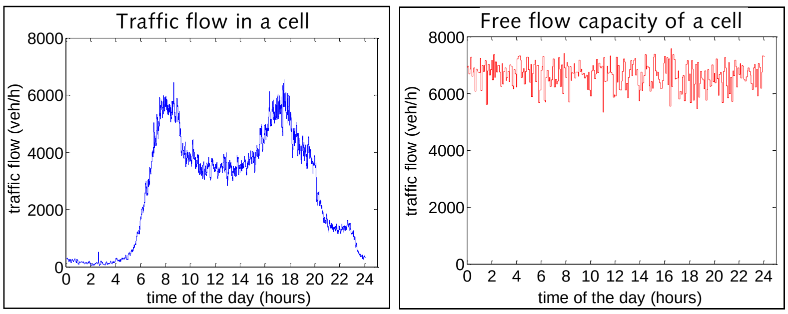


Figure 8.19: Random fluctuation of the traffic flow in a cell of the network, as a result of the intrinsic randomness in travel behavior (left), and random fluctuation of the free flow capacity of a cell of the network, as a result of the intrinsic randomness in driving behavior (right). All other sources of variability were excluded in this case, except for the (regular) demand variation with the time of the day.

4. Ignoring the influences of incidents and road works, events seem to be the most important source of weekend day traffic congestion. The relative influence of the capacity variations due to the intrinsic randomness in driving behavior is much smaller on weekend days.

It should be noted here that some caution should be taken regarding the first conclusion. As discussed in section 6.2, it is not entirely clear yet how the capacity effect of the intrinsic randomness in driving behavior can best be modeled. Up to a certain degree, the use of a different modeling approach could affect the findings with regard to the relative importance of this source of variability.

The above conclusions can be translated into the following implications for the selection of strategies to alleviate traffic congestion:

1. Peak period-related traffic congestion might relatively effectively be alleviated by taking measures which reduce the effects of the intrinsic randomness in human driving behavior. These measures might however not be so easy to find. Results of Brilon et al. (2005) indicate that the implementation of traffic adaptive variable speed limits might be a measure with the desired effect. Also the implementation of Advanced Driver Assistance Systems might have the desired effect, insofar as these exclude the human factor, by actively taking over control[103].
2. Theoretically it would be desirable to strive for a more even distribution of the traffic demand over the months of the year. This is clearly unfeasible, however, since the variation over the months is directly related to seasonal differences in human activity patterns and travel choices, which are virtually unchangeable.
3. Aside from potentially yielding large benefits in really *extreme* weather, taking measures directed at mitigating the traffic effects of adverse weather seems not a very effective strategy to alleviate traffic congestion.
4. Up to a certain extent, weekend day traffic congestion can be avoided by a careful planning of events, including the provision of route guidance, adequate parking facilities and appropriate alternative transport modes.

[103] Note that if such a system allows drivers to make their own settings (based on their own driving style), the reduction in the intrinsic capacity randomness will obviously be much smaller.

8.6.4 Sensitivity to the spatial dependencies in the capacity randomness

In section 8.5, the model results were subjected to a small sensitivity analysis. In this analysis, it was considered to what extent the results are sensitive to the modeling assumption on the spatial dependency in the part of the capacity variation that is due to the intrinsic randomness in driving behavior. This modeling assumption is reflected in the cell size chosen for the capacity randomization. In the sensitivity analysis, the 'extreme' case of the cell size being equal to the *link length* was considered, which means that the capacities were varied at the level of links as a whole. This corresponds to the implicit assumption that the capacities are fully dependent over the whole length of the link.

For the model results presented above, a similar sensitivity analysis has been performed. The results of this analysis are provided in Appendix A4.2. These results consist of recalculated versions of the output diagrams presented in section 8.6.2 (except for the travel time instability diagrams), as obtained by model runs in which the cell-based capacity randomization is replaced by a link-based randomization.

From the results it is clear that the conclusions are not really affected by the use of the different modeling approach. Of course, the relative influence of the intrinsic randomness in human driving behavior is considerably smaller if its capacity effect is assumed to be uniform (i.e. fully correlated) over the length of a link. Together with the seasonal demand variation over the months of the year, it is, however, still the most important factor in the peak period-related traffic congestion. Again, the other considered sources of variability are found to have a much smaller influence.

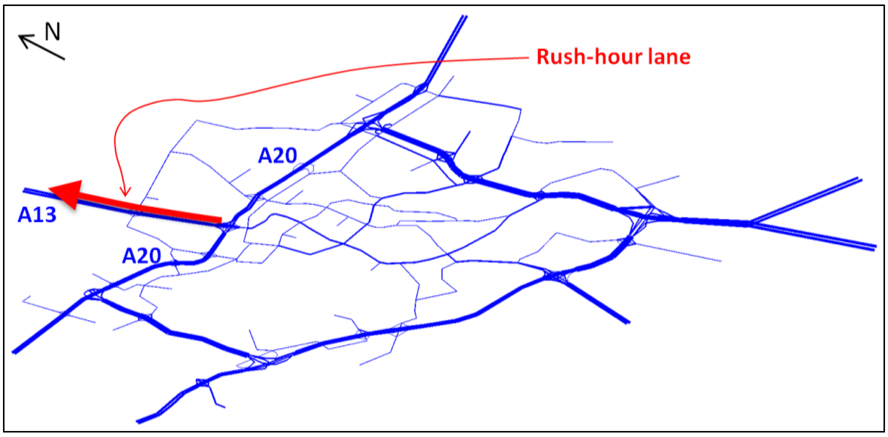
Note that in the mean, median and 90^{th}-percentile values of the *overall* travel time distributions, the impacts of the different sources of variability are hardly observable in case of this alternative modeling approach, as apparent from Figure A4.8. This is obviously due to the fact that even in case of full variability (i.e. without any of the sources of variability being omitted), these three indicators are already close to the free flow travel time, as can be seen in Figure A4.1 (and was explained in section 8.5.4). As a result, these indicators will be relatively insensitive to the omission of a source of variability, whichever of the different sources it is.

8.7 Effects of a rush-hour lane

8.7.1 Introduction

This section shows that explicitly accounting for the different sources of variability can provide us with new insights into the effectiveness of specific measures that are proposed to alleviate traffic congestion. For this, the example of a rush-hour lane is considered. This rush-hour lane is supposed to be proposed for the A13 motorway, on the roadway in the northbound direction. It is supposed to extend from the motorway junction with the A20 to a location just before the network boundary, as illustrated in the figure below. Note that this is just a hypothetical scenario, which does not need to be realistic.

Figure 8.20: Location of the considered rush-hour lane



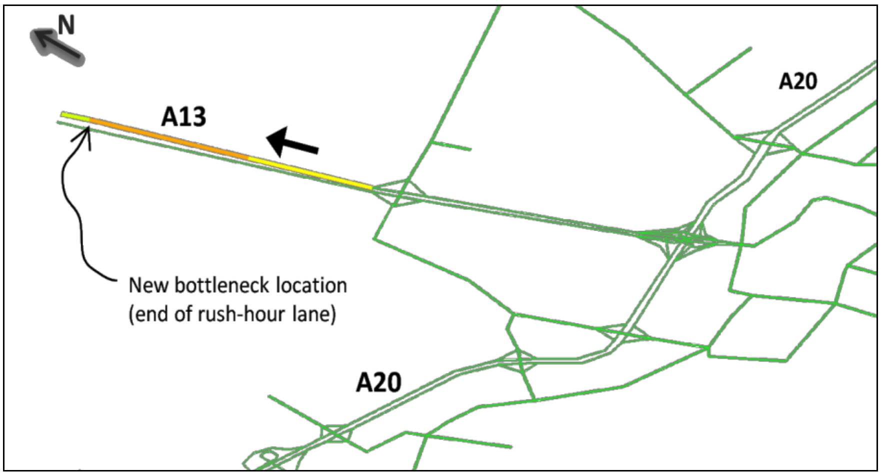
This rush-hour lane is supposed to reduce the following problem:

In the 'representative' morning peak, the capacity of the A13 is insufficient to cope with the traffic demands from the A20. As a result, traffic jams are created on both branches of this A20, as shown by the figure below. An important observation is now that these traffic jams do not only consist of traffic wanting to enter the A13 motorway. It also consists of through traffic (i.e. traffic that continues its way over the A20 motorway), which is blocked by the traffic heading for the A13. The result of this is that this through traffic experiences 'unnecessary delay': delay due to a capacity shortage at a bottleneck that it does not have to pass.

Figure 8.21: Congestion spillback from the A13-bottleneck to both branches of the A20 motorway

The rush-hour lane will reduce this problem by moving the bottleneck to a location further downstream, as illustrated in the figure below. As a result, the likelihood of spill-back to the A20 motorway will be significantly reduced.

Figure 8.22: New congestion location after the realization of the rush-hour lane



The (supply) effects of the rush-hour lane are modeled in the way described in section 5.4.14. It is important to be aware of the fact that in reality, the implementation of this measure might significantly change the traffic demand on the motorway network (regarding both its total volume, as well as its distribution over time and space), by influencing travelers' choices with respect to: whether or not making a trip, destination, mode, route, and departure time. When assessing the effectiveness of a measure *in the context of a concrete real-life project*, all these effects would need to be given due consideration. Here it is not intended however to make statements on the exact effectiveness of this particular rush-hour lane. Rather, the intention is to illustrate that there might be *differences* in the effectiveness according to a traditional evaluation (focusing on a 'representative' situation) and the effectiveness according to an evaluation in which the various sources of variability are accounted for. Because of this, it is considered acceptable to neglect these demand effects here, for reasons of simplification.

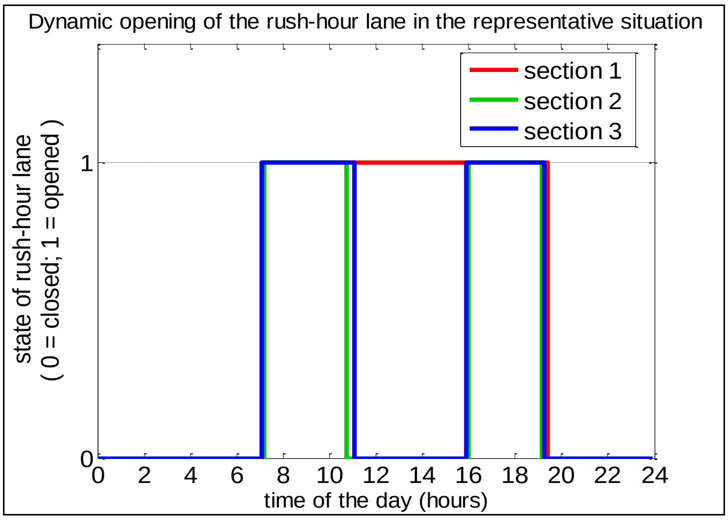
Below, first of all the evaluation according to the more traditional approach will be given (focusing on a 'representative' situation). After that, subsection 8.7.3 describes the outcomes of the evaluation in which the various sources of variability are explicitly accounted for, and discusses which additional/revised insights these provide (as compared with the outcomes of the traditional evaluation). Subsection 8.7.4 summarizes these new insights, and draws some more general conclusions. Finally, subsection 8.7.5 discusses the results of a small sensitivity analysis of the results, comparable to the sensitivity analyses discussed in sections 8.5.4 and 8.6.4.

8.7.2 Traditional evaluation

In the tables and figures below, the effects of the rush-hour lane as obtained for the *representative* situation are shown. Together, these effects reflect the effectiveness of the rush-hour lane according to the traditional evaluation approach, in which only this representative situation is considered. All sources of variability are ignored here, except for the systematic demand variation with the time of the day.

Before turning to its effects, the figure below first of all shows the (traffic responsive) dynamic behavior of the rush-hour lane in the representative situation. Note that the rush-hour lane consists of three consecutive sections, corresponding to different links of the network[104]. These different sections may be opened or closed independently from each another. From the figure it can be observed that the second and third sections are opened during the peak periods only, while the first section is opened during the whole of the daytime period. The dynamic lane considered here is thus not a rush-hour lane in its most restricted sense.

Figure 8.23: Dynamic opening of the rush-hour lane in the representative situation



INDICATORS I–V (TRAVEL TIME STATISTICS):

<u>*Statistics of the overall travel time distribution:*</u>

The table below shows the statistics of the (representative) overall travel time distributions for the scenario with the rush-hour lane. Values that deviate from those obtained for the base scenario without the rush-hour lane (see section 8.4) are explicitly marked. For the purpose of comparison, the original value (for the base scenario) is given as well in such cases (between brackets).

[104] Please note that the curves in the figure are overlapping to some extent.

Table 8.3: Representative values of the statistics of the overall travel time distributions for the scenario with the rush-hour lane

Indicator \ Route	1	2	3	4	5	6
I. TT_90	1,15	1,19	1,10	**1,11** (1,10)	**1,10** (1,11)	1,11
II. TT_mean	1,11	1,08	1,05	**1,06** (1,05)	**1,04** (1,05)	**1,04** (1.05)
III. TT_median	1,07	1,08	1,05	1,05	1,05	1,05
IV. TT_width	0,15	0,19	**0,10** (0.09)	**0,11** (0.10)	0,10	**0,10** (0.11)
V. TT_skew	1,10	1,54	**0,85** (0.84)	**1,32** (1,02)	**1,05** (1,14)	**1,12** (1,17)

Note: all values of indicators I-IV are made dimensionless by division by the free flow travel time. Indicator V is already dimensionless by itself.

As apparent from the table, the effects on the overall travel time distributions are relatively limited. This is in fact rather logical, since:

- The overall travel time distributions relate to *the whole of the day*, while the problem existed only during the morning peak.
- The travel time distributions relate to entire routes, while the problem existed only locally.

From the table it can be seen that the 'A20 routes' (route numbers 5 and 6) benefit from the presence of the rush-hour lane (as expected), while the directly affected route (i.e. the route over the road stretch with the rush-hour lane, corresponding to number 4) suffers from its presence. This latter can be explained by the facts that:

- The rush-hour lane does nothing more than moving the A13 bottleneck to a location further downstream, implying that the A13 road users will still face more or less the same amount of congestion.
- During the periods in which the rush-hour lane is open to traffic, the speed limit on the road stretch concerned is reduced to 80 km/h[105] (for safety reasons), which increases the travel time.

The rush-hour lane appears to have a (very small) negative influence on the overall travel time distribution of route 3 (the north-south route from the A13 to the A16) as well. This can probably be attributed to the fact that the traffic congestion on the A20 (which disappears after the implementation of the rush-hour lane) has a favorable effect on the traffic conditions on some road stretches that are part of this route, by delaying some of the traffic that is on its way to these road stretches.

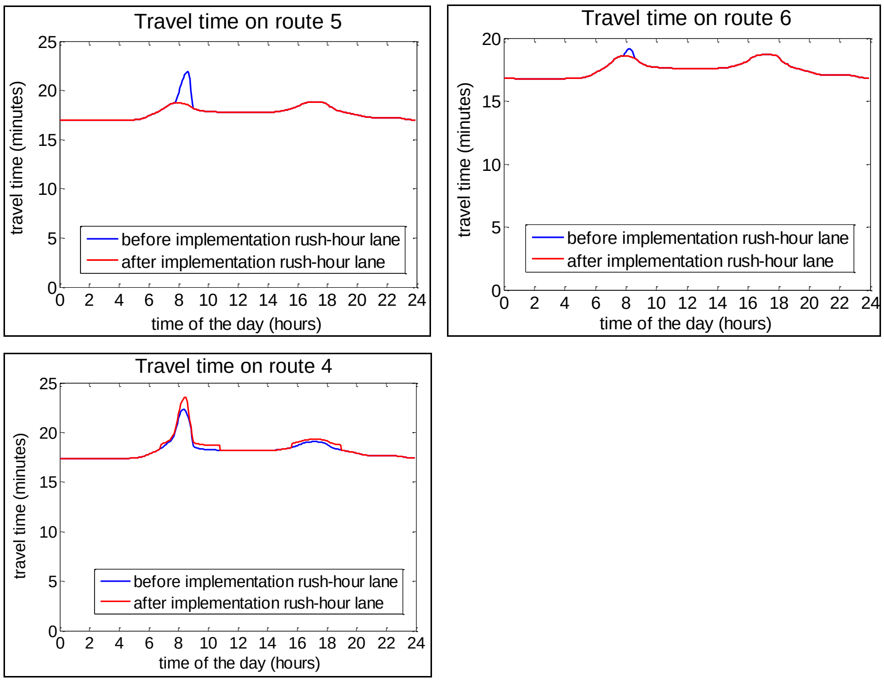
Statistics of the travel time distributions for specific times of the day:

Obviously, the travel time statistics for specific times of the day are not defined for the representative situation, due to the fact that they refer to a variability which is not modeled in this case. Instead, only the representative travel time patterns over the day can be given. These are shown in the figures below. The travel time improvement on route 5 can clearly be observed. The improvement on route 6 is smaller, however. On route 4 the travel time clearly deteriorates, which is due

[105] Note that this applies to only a part of this road stretch. On the other part the speed limit is 80 km/h already, because it is located within an 80 km/h zone.

to the speed limit reduction during the periods in which the rush-hour lane is opened (as discussed above). The travel time graphs of the other routes do not show significant changes and are therefore omitted here.

Figure 8.24: Effects of the rush-hour lane on the representative travel time patterns over the day



Indicator VI (travel time instability):

The table below shows the travel time instability values for the scenario with the rush-hour lane. Note that these travel time instabilities cannot be given as distributions in this representative evaluation, but only as deterministic values. Values that deviate from those obtained for the base scenario without the rush-hour lane (see section 8.4) are explicitly marked. For the purpose of comparison, the original value (for the base scenario) is given as well in such cases (between brackets).

Also in terms of travel time instability, the ‘A20 routes’ (route numbers 5 and 6) benefit from the presence of the rush-hour lane (as expected), while the directly affected route (i.e. the route over the road stretch with the rush-hour lane, corresponding to number 4) suffers from its presence.

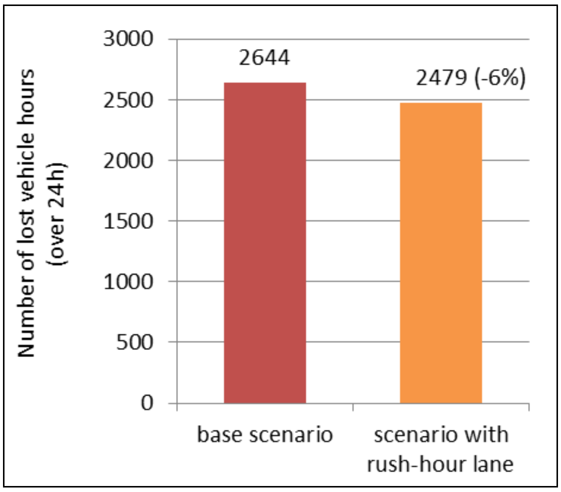
Table 8.4: Representative values of the travel time instability for the scenario with the rush-hour lane

Route / Time of the day	1	2	3	4	5	6
03:00	0,00	0,00	0,00	0,00	0,00	0,00
07:00	0,01	0,01	0,01	**0,03** (0.01)	0,01	0,01
08:00	0,00	0,01	0,00	**0,05** (0.03)	**0,00** (0.04)	**0,00** (0.02)
09:00	-0,01	-0,04	**-0,01** (0.00)	**-0,06** (-0.04)	**-0,01** (-0.08)	0,00
13:00	0,00	0,00	0,00	0,00	0,00	0,00
16:00	0,01	0,01	0,00	0,00	0,01	0,01
17:00	0,08	0,03	0,00	0,00	0,00	0,00
18:00	0,00	-0,04	0,00	**0,00** (-0.01)	-0,01	0,00
21:00	0,00	0,00	0,00	0,00	0,00	0,00

INDICATOR IX (NUMBER OF LOST VEHICLE HOURS):

The figure below shows the reduction in the total number of lost vehicle hours on the network (over the 24-hours period). This reduction is obviously limited, due to the fact that the rush-hour lane only reduces the congestion that is related to the A13 bottleneck. Other congestion problems in the network are not alleviated.

Figure 8.25: Effect of the rush-hour lane on the total number of lost vehicle hours in the representative situation



8.7.3 Evaluation of the effectiveness in the variable situation

This section discusses the effectiveness of the rush-hour lane as found by performing an evaluation in which the various sources of variability are explicitly accounted for. It should however be noted that two sources of variability have been omitted from this evaluation, namely road works and incidents. This is for the same reason as described in section 8.5.

INDICATORS I–V (TRAVEL TIME STATISTICS):

Statistics of the overall travel time distribution:

The table below shows the statistics of the overall travel time distributions for the scenario with the rush-hour lane.

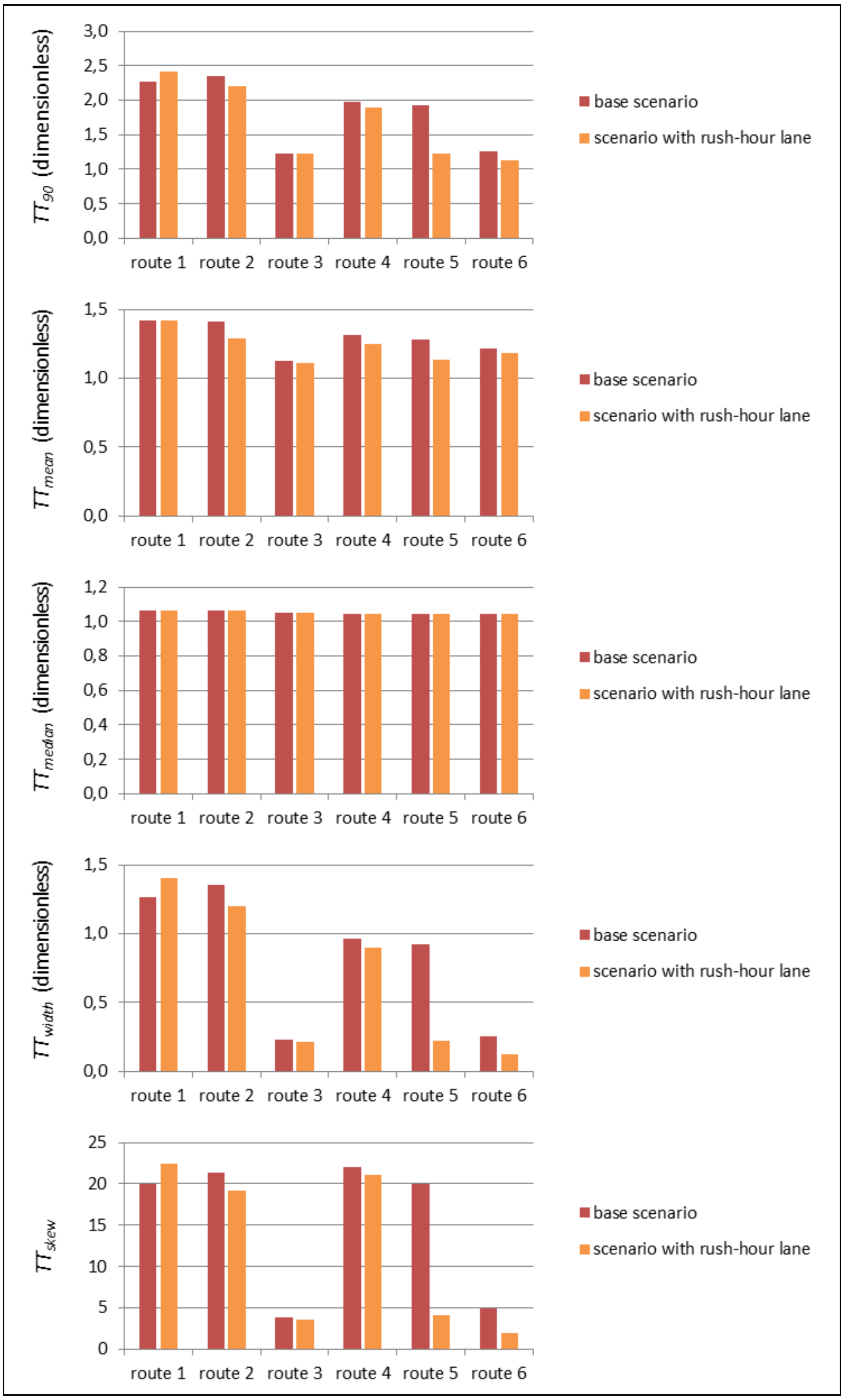
Table 8.5: Statistics of the overall travel time distributions for the scenario with the rush-hour lane

Route / Indicator	1	2	3	4	5	6
I. TT_90	2,41	2,21	1,22	1,90	1,23	1,13
II. TT_mean	1,42	1,29	1,11	1,25	1,14	1,18
III. TT_median	1,06	1,06	1,05	1,04	1,05	1,05
IV. TT_width	1,41	1,20	0,22	0,90	0,23	0,13
V. TT_skew	22,39	19,16	3,63	21,02	4,13	1,96

Note: all values of indicators I-IV are made dimensionless by division by the free flow travel time. Indicator V is already dimensionless by itself.

In the figure below, these indicator values are compared with the corresponding values for the base scenario (i.e. the situation without the rush-hour lane).

Figure 8.26: The statistics of the overall travel time distributions (90th-percentile, mean, median, width and skew) in the situation with the rush-hour lane, compared with those in the situation without this lane



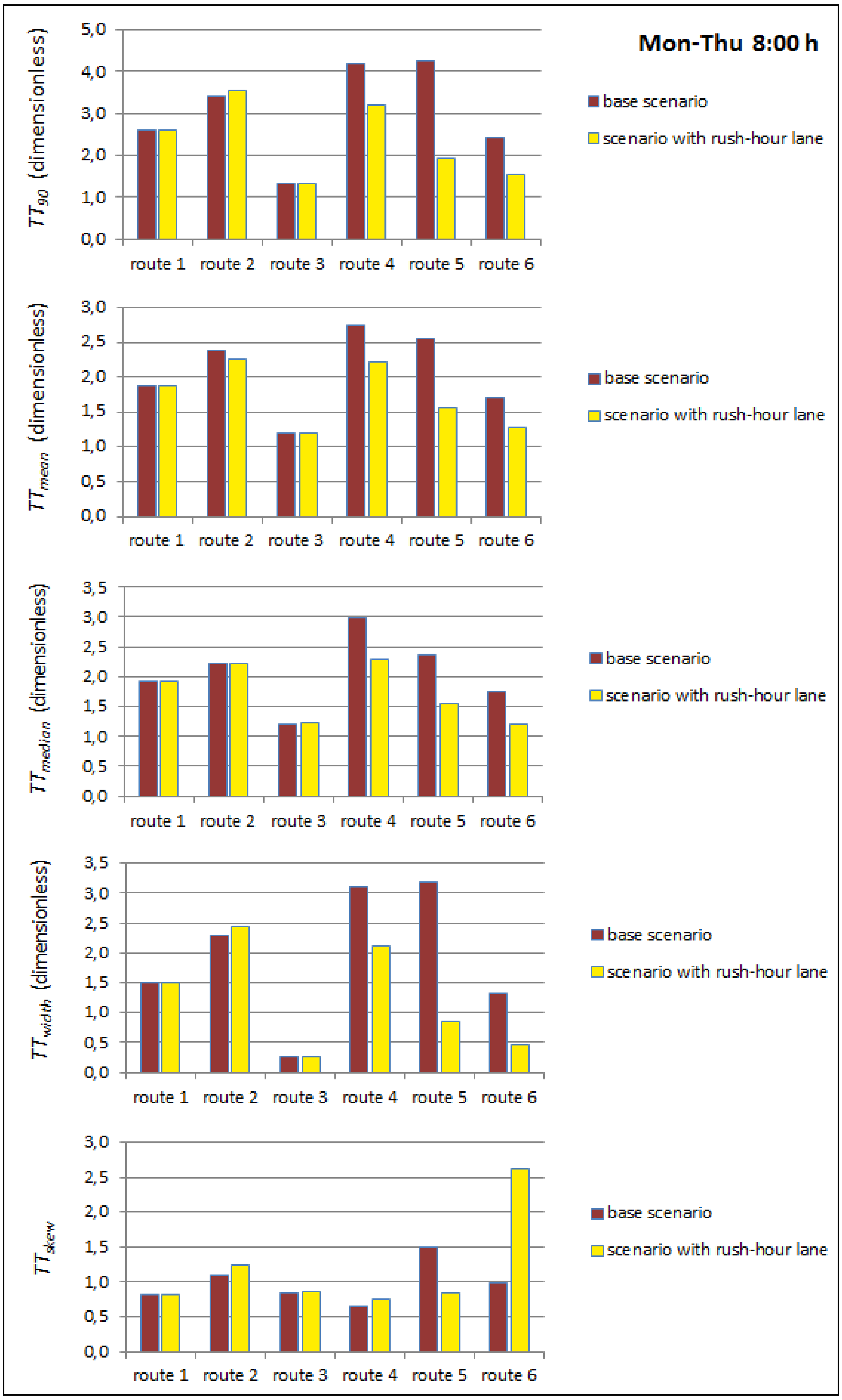
The effects according to the above figure are clearly different from those found in the evaluation according to the traditional approach (see section 8.7.2):

- The positive effects on the travel time distribution of route 5 (i.e. the eastbound A20-route) turn out to be much larger than indicated by the results of the traditional evaluation. The 90th percentile travel time even decreases by nearly 40%.
- The positive effects on the travel time distribution of route 6 (i.e. the westbound A20-route) are found to be larger as well.
- While the traditional evaluation predicted that the directly affected route (i.e. the A13-route over the roadway with the rush-hour lane, corresponding to number 4) would suffer from the presence of the rush-hour lane, here it is found that rather a (limited) improvement of the travel time statistics of this route is to be expected.
- While in the traditional evaluation no effect was found on the travel time distribution of route 2, here a significant improvement is found. This is due to the fact that in the base scenario the congestion generated by the A13-bottleneck often spills back all the way to the A15 motorway, which increases the travel time on route 2. By solving the bottleneck at the start of the A13, the rush-hour lane might prevent this spillback from occuring, thereby improving the travel time distribution of route 2. Since this spillback to the A15 does not occur in the 'representative' situation, this potential benefit of the rush-hour lane is not noticed in the evaluation according to the traditional approach.
- While in the traditional evaluation no effect was found on the travel time distribution of route 1, here its 90th pecentile value is found to increase. A possible explanation for this might be that the A13-related congestion sometimes has as a favorable effect on the traffic conditions on certain road stretches that are part of this route, by blocking/delaying traffic that is on its way to these road stretches. If the A13-related congestion is alleviated by the rush-hour lane, this favorable effect disappears.

Statistics of the travel time distributions for specific times of the day:

For certain times of the day very large improvements in the travel time statistics for route 4, 5 and 6 (i.e. the A13-route with the rush-hour lane, and both A20-routes) are found. This is illustrated by the figure below, which shows the computed changes in the travel time statistics for 8 am. Both the mean/median travel time and the travel time variability turn out to decrease on these routes. The calculated improvements in the mean/median travel time are much larger than the improvements in the representative travel time, as calculated in the evaluation according to the traditional approach. Remember that for route 4 even a (slight) deterioration was calculated for the representative situation.

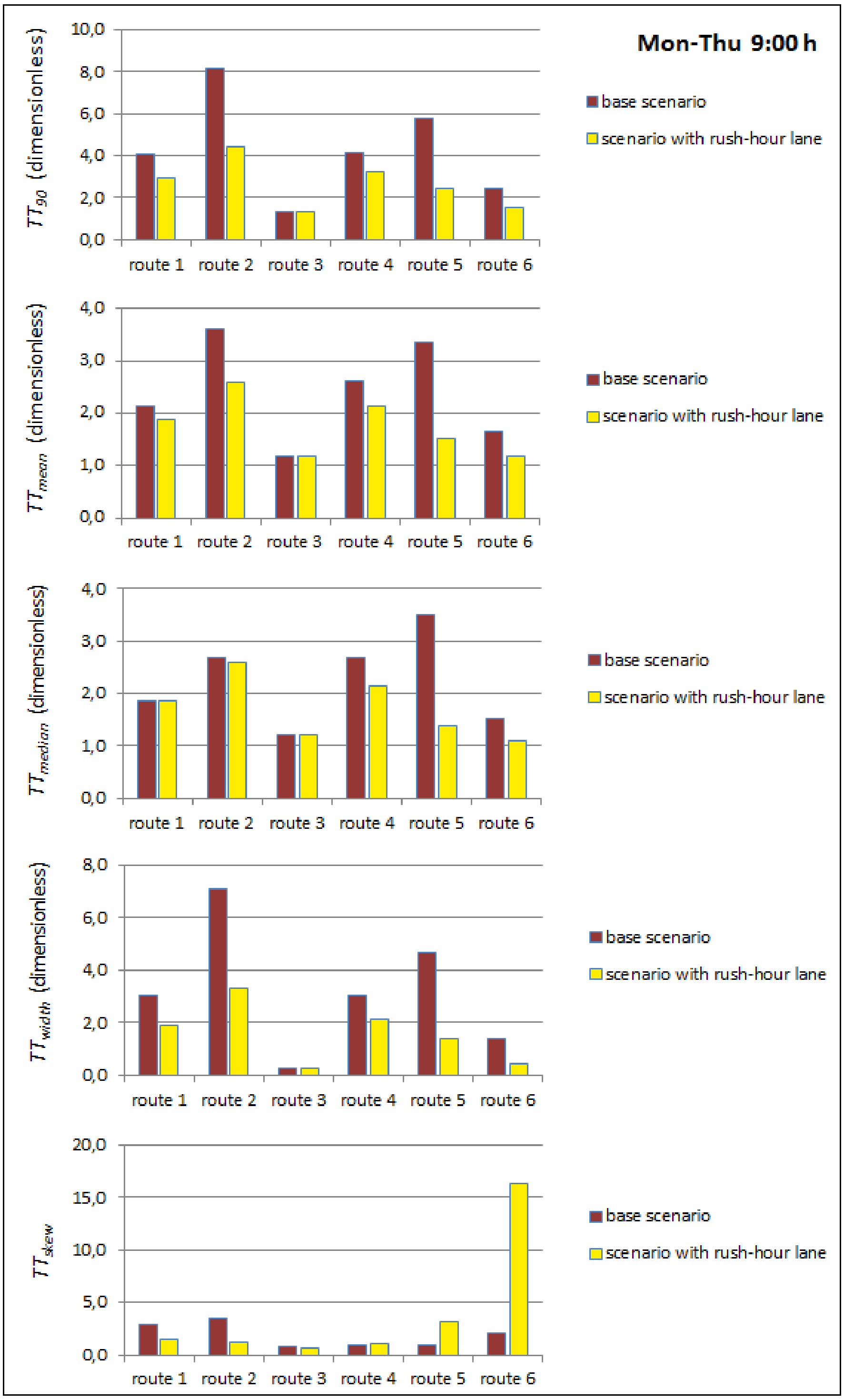
Figure 8.27: The travel time statistics for Mon-Thu – 08:00 (90th-percentile, mean, median, width and skew), in the situation with the rush-hour lane, compared with those in the situation without this lane



For routes 4, 5 and 6, similar improvements are found for 9am as for 8am, as apparent from the figure below. This is a difference with the results for the representative situation, in which the improvements are limited to a period *before* 9am only (see section 8.7.2).

Even more remarkable is that significant improvements are found for routes 1 and 2 (the A15-routes) as well, while the traditional evaluation suggested that the conditions on these routes would be unaffected by the rush-hour lane. Closer examination reveals that these results are not contradictory, however. To see this, note that hardly any change is found in the *median* travel times. This is in good agreement with the fact that no effect is observed in the representative situation. In the assessment according to the traditional approach it remains unnoticed however that the A13-related congestion in part of the occasions is much heavier than in other occasions. In such situations this congestion spills back all the way to the A15 motorway (where it apparently arrives between 8 and 9 am), causing significant delays over there. The rush-hour lane reduces this congestion spillback.

Figure 8.28: The travel time statistics for Mon-Thu – 09:00 (90th-percentile, mean, median, width and skew), in the situation with the rush-hour lane, compared with those in the situation without this lane



For the daytime off-peak period improvements in the 90th percentile and mean travel times are found as well (for routes 1, 2, 4 and 5). It is questionable to which extent this is realistic, however. At least part of the off-peak congestion concerned here is congestion originating from the morning peak, which dissolves much too late. This is related to the problem that the model generates too much traffic congestion in many of the simulation runs, as discussed in section 7.5.1. Obviously, this problem will affect the results for other times of the day as well. Therefore, not too much value should be attached to the exact quantitative values of the outcomes.

For the evening peak the effects of the rush-hour lane are found to be much smaller, which is in reasonable agreement with the results of the evaluation according to the traditional approach (see section 8.7.2). It is however not true that there are no effects on the evening peak conditions at all, as this traditional evaluation suggests. For route 5 (i.e. the eastbound A20 route) a significant improvement is found, not only in the 90th percentile and mean, but even in the median travel time.

For Fridays more or less the same effects are observed as for the other weekdays. The positive impact on route 1 and 2 at the end of the morning peak is smaller (or in case of route 1 even absent), however. This can be explained by the fact that the situation on Friday is better anyhow (i.e. not considering the rush-hour lane yet). Due to its lower morning peak traffic demand, less congestion spillback occurs from the A13 bottleneck. As a result, there is less room for improvement on this day.

For weekend days no significant effects are found. This is obviously related to the fact that the rush-hour lane is specifically aimed at solving a specific weekday congestion problem, which does not exist on weekend days.

INDICATOR VI (TRAVEL TIME INSTABILITY):

Due to the removal of the bottleneck at the start of the A13, the morning peak travel time instability on both A20 routes (route numbers 5 and 6) is significantly reduced (both in terms of the average travel time change and the uncertainty in this change). This is shown in the figures below. In absolute terms, the improvements of the averages are larger than the improvements of the corresponding instability values for the representative situation (see section 8.7.2).

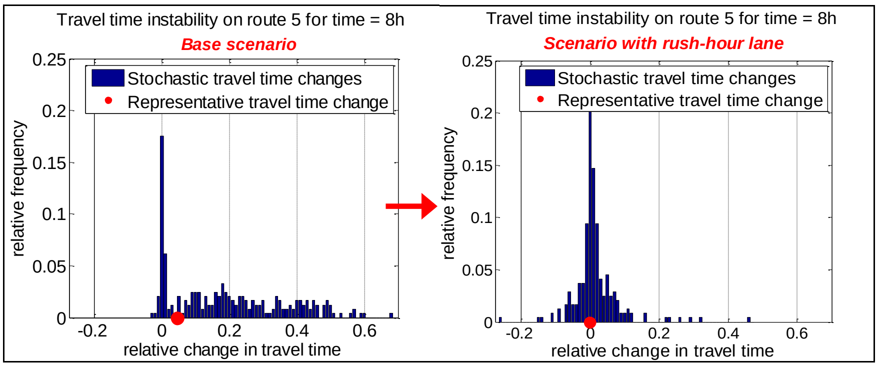


Figure 8.29: The effect on the travel time instability on route 5, for 08:00 (weekdays only)

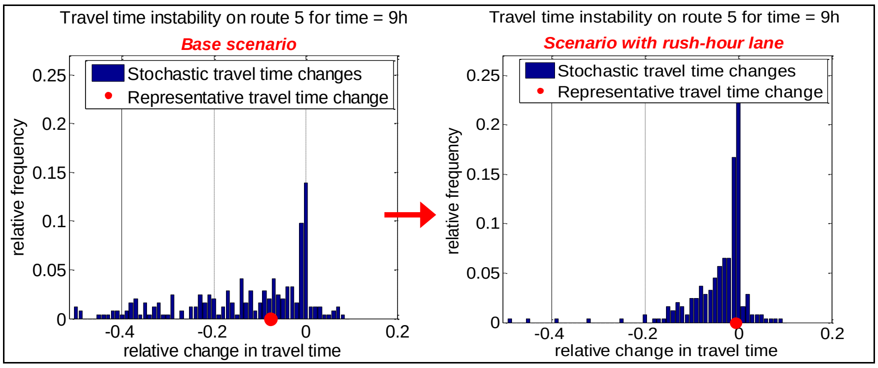


Figure 8.30: The effect on the travel time instability on route 5, for 09:00 (weekdays only)

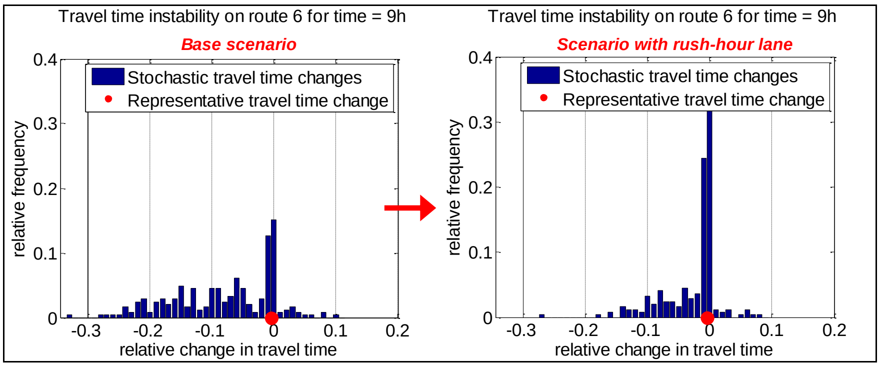


Figure 8.31: The effect on the travel time instability on route 6, for 09:00 (weekdays only)

For route 5 not only the *morning peak* travel time instability is reduced, but the *evening peak* travel time instability as well, as shown below. This improvement was missed by the evaluation according to the traditional approach (section 8.7.2).

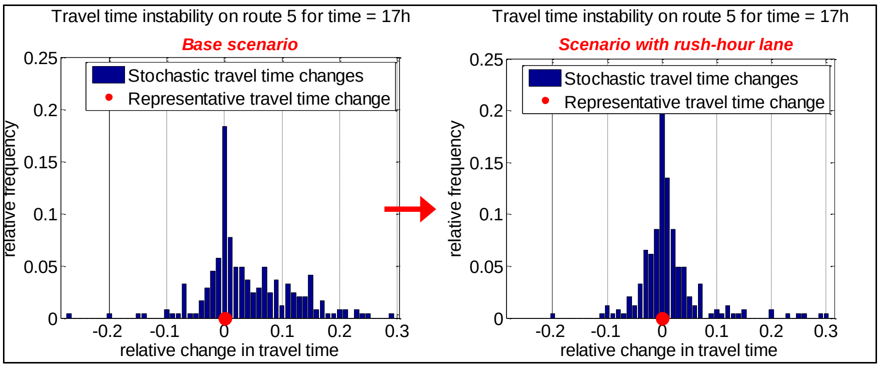


Figure 8.32: The effect on the travel time instability on route 5, for 17:00 (weekdays only)

On the route with the rush-hour lane itself (i.e. the northbound A13 route, corresponding to route number 4), the morning peak travel time instability is reduced as well, as apparent from the figures below. Since the evaluation according to the traditional approach showed a worsening instead of an improvement, this evaluation clearly gave a false impression of the effect on the travel time instability on this route.

Figure 8.33: The effect on the travel time instability on route 4, for 08:00 (weekdays only)

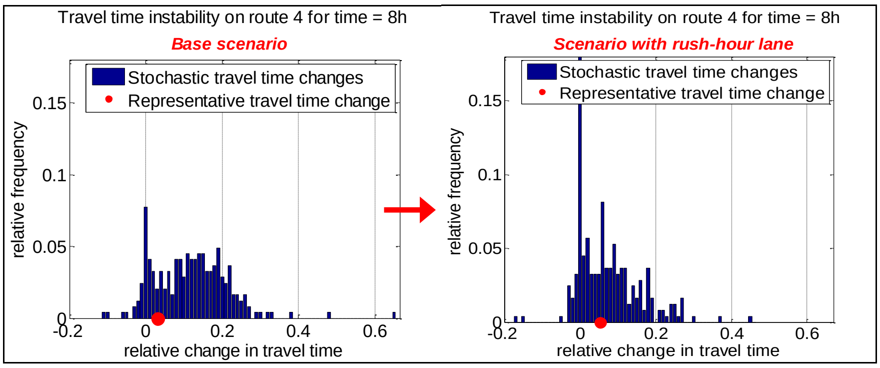
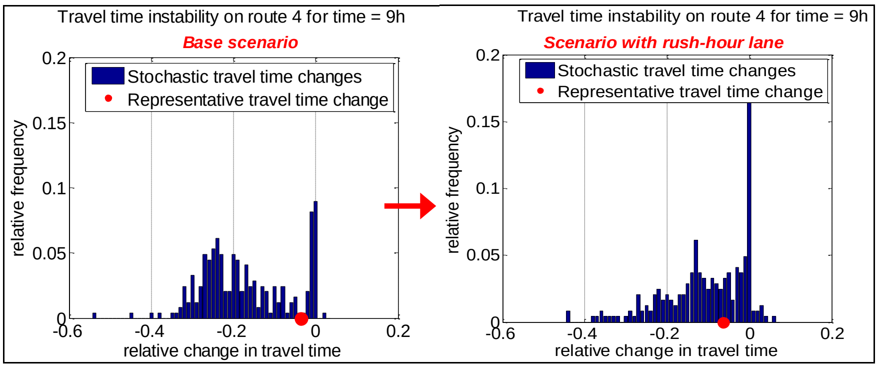
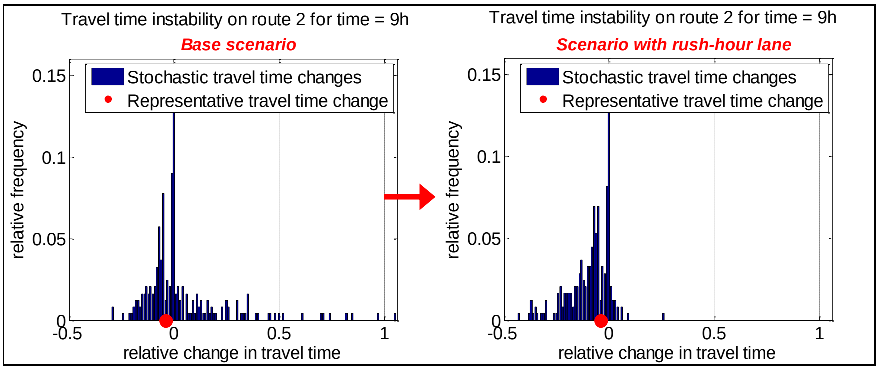


Figure 8.34: The effect on the travel time instability on route 4, for 09:00 (weekdays only)



Finally, a reduction in travel time instability is found on route 2, at the end of the morning peak (as shown in the figure below). On some days, spillback from the A13 bottleneck resulted in sudden sharp increases in travel time here. By removing this bottleneck, the rush-hour lane can prevent this phenomenon from occurring. In the evaluation according to the traditional approach, this potential benefit of the rush-hour lane remains unnoticed.

Figure 8.35: The effect on the travel time instability on route 2, for 09:00 (weekdays only)



INDICATOR IX (NUMBER OF LOST VEHICLE HOURS):

The figures below show the reductions in the total amount of lost vehicle hours on the network, separately for weekdays and weekend days. Clearly, on weekdays the reduction is much larger than found in the evaluation according to the traditional approach (focusing on the representative situation – see section 8.7.2), both in absolute and relative terms. This means that the 'representative' evaluation results in a significant underestimation of the benefits of the rush-hour lane.

For weekend days, the reduction is found to be much smaller (both in absolute and in relative terms). This is much less important, however, since the volume of traffic congestion on weekend days is relatively small. Probably, the small reduction on weekend days will be due to the fact that the A13-bottleneck is not often active on these days (as opposed to the situation on weekdays). In section 8.6 it was found that most of the congestion on weekend days is related to events. Obviously, the traffic generated by these events might cause congestion at other locations in the network than the weekday peak period traffic.

Figure 8.36: The mean and median numbers of lost vehicle hours on weekdays in the situation <u>with</u> the rush-hour lane, compared with those in the situation <u>without</u> this lane

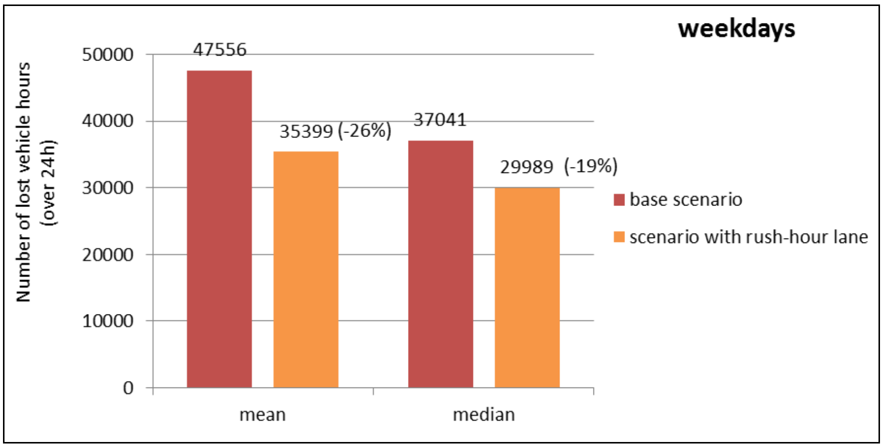
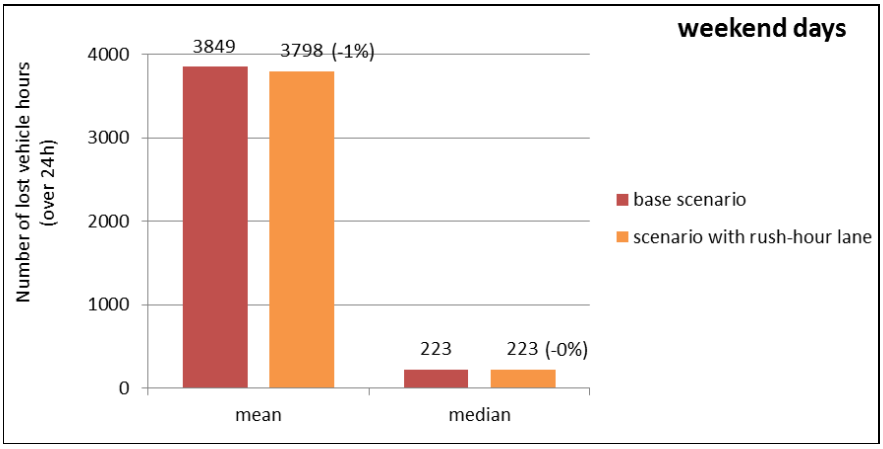


Figure 8.37: The mean and median numbers of lost vehicle hours on weekend days in the situation <u>with</u> the rush-hour lane, compared with those in the situation <u>without</u> this lane



8.7.4 Discussion

Clearly, the evaluation in which the various sources of variability were explicitly accounted for provided us with extra insights into the effectiveness of the considered rush-hour lane, which were not obtained from the evaluation according to the more traditional approach (focusing on the 'representative' situation). Furthermore, it turned out that some of the findings obtained for the 'representative' situation actually did not give a good impression of the typical effects of the rush-hour lane. That is, the effects calculated for the 'representative' situation prove not always that representative *themselves*.

An overview of the additional and revised insights obtained by explicitly accounting for the various sources of variability (as compared with those obtained by an evaluation according to the more traditional approach) is given below.

Additional/revised insights into the effects of the rush-hour lane:

- The positive effects on the travel times on the A20 routes are much larger than apparent from the results of the evaluation according to the traditional approach. (Even if one considers the *median* travel times.)

  This seems closely related to the fact that the 'representative' calculation underestimates the amount of traffic congestion (see section 8.5). Obviously, this results in the potential benefits of measures being underestimated as well.
- The travel time statistics and travel time stability on the directly affected route (i.e. the A13-route over the roadway with the rush-hour lane) will improve, while the evaluation according to the traditional approach predicted a (slight) deterioration of the travel times and travel time stability.
- The morning peak travel time instability on routes 4, 5 and 6 is significantly reduced. The reductions in the average instability values are larger than the reductions in the corresponding representative values. Also the day-to-day *variability* in the instability values is signifcantly reduced. This is not noticed in the evaluation according to the traditional approach, since this day-to-day variability is not considered in such an evaluation.
- The positive impact on the conditions on route 5 is not limited to the morning peak period (as indicated by the results of the evaluation according to the traditional approach), but is associated with the evening peak period as well (although to a more limited extent).
- The 90^{th} percentile and mean travel times on route 1 and 2 for the end of the morning peak will significantly improve (on Monday-Thursday), while the representative calculation did not show any effects at all. The same applies to the travel time stability on route 2.

In the representative calculation this potential benefit of the rush-hour lane was 'overlooked', due to the fact that the regularly occurring spillback to the A15 does *not* occur in the 'representative' situation.

- On routes 4, 5 and 6 the day-to-day variability in the morning peak period travel times is greatly reduced.

  This aspect is not noticed in the evaluation according to the traditional approach, since this day-to-day variability is not considered in such an evaluation.

- At network level the reduction in the total amount of lost vehicle hours is much larger than calculated for the representative situation. This does not only apply to the mean, but to the median value as well.

  This seems closely related again to the fact that the 'representative' calculation underestimates the amount of traffic congestion, as mentioned above.

- For weekend days significant effects are not to be expected.

These observations can be translated into the following *conclusions of a more general nature*:

1. Due to the fact that a 'representative' calculation (i.e. based on the representative values of the demand and supply characteristics) underestimates the traffic congestion in certain respects (see section 8.5.3), it underestimates the beneficial effects of proposed measures (aimed at alleviating this congestion) as well. In section 8.5.3 it was discussed that this deficiency will not be satisfactorily remedied by calibrating the representative calculation to the 'representative' congestion level.
2. In an evaluation according to the traditional approach, potential benefits of a traffic measure may remain unnoticed due to trend breaks in the behavior of the traffic system. This applies particularly to (the avoidance of) spillback of congestion to other network elements. If this spillback occurs only in part of the occasions (say less than 50%), it will not be included in the representative analyses. Consequently, the benefits achieved on these other network elements will not be reflected in the evaluation results.
3. In an evaluation according to the traditional approach, no information is obtained on the improvements in travel time uncertainty (due to the fact that the day-to-day variability in the traffic conditions is not considered), while this improvement might be an important component in the benefits of a traffic measure.

These conclusions imply that more systematic attention should be given to the inherent variability in traffic congestion, when evaluating the effectiveness of measures that are proposed to alleviate traffic congestion. Because of the complexity involved (especially in case of heavily loaded networks in highly urbanized areas), this would have to be done by using a model in which the different sources of variability are explicitly accounted for, such as the model developed in this project. In the final section of this chapter (section 8.8), it is discussed whether/how this model could be used for this purpose.

8.7.5 Sensitivity to the spatial dependencies in the capacity randomness

In sections 8.5 and 8.6, the model results were subjected to a small sensitivity analysis with respect to an important model uncertainty. In this analysis, it was considered to what extent the results are sensitive to the cell size chosen for the capacity randomization. Implicitly, choosing a certain cell size corresponds to assuming a certain degree of spatial dependency in the random capacity variation (i.e. the part of the capacity variation that is due to the intrinsic randomness in human driving behavior). In the sensitivity analysis, the 'extreme' case of the cell size being equal to the *link length* was considered, which means that the capacities were varied at the level of links as a whole. This corresponds to the implicit assumption that the capacities are fully dependent over the whole length of the link.

For the computed effects of the rush-hour lane, a similar sensitivity analysis has been performed. The results of this analysis are provided in Appendix A4.3. These results consist of recalculated versions of the output diagrams presented in subsection 8.7.3, as obtained by a model run in which the cell-based capacity randomization is replaced by a link-based randomization.

From the results it is clear that the conclusions are not really affected by the use of the different modeling approach. In absolute terms, the calculated improvements are generally smaller now, but in relative terms sometimes larger. The most important observation, however, is that the calculated improvements are still much larger and more comprehensive (in terms of both time and space) than those found in the evaluation according to the traditional approach.

Note that this is not the case for the improvements in the mean, median and 90th-percentile values of the *overall* travel time distributions (i.e. the distributions related to the combination of all times of day and days of week). In these statistics (Figure A4.12), the effects of the rush-hour lane are hardly noticeable, which is in good agreement with the findings obtained in the evaluation according to the traditional approach (Table 8.3). The cause of this (almost) negligible impact is obvious. Due to the fact that the traffic conditions are free of congestion during the main part of the time, the values of these three indicators are in the base scenario already close to the free flow travel times (meaning that they are not very significantly affected by congestion). Consequently, there is little room for improvement in these values.

8.8 Use of the developed model for real-life evaluations

8.8.1 Introduction

In this section it is discussed whether/how the developed model could be used for practical application within the context of real-life evaluations of measures proposed to alleviate traffic congestion. It should be stressed, however, that this model was developed solely for the research task considered in this project (i.e. *illustrating* the gain of additional insights), and thus *not* directly for practical application in the evaluation of concrete projects.

When considering the model's suitability for practical application in such evaluations, two aspects are of central importance:

- its validity for this task
- the required computation time

These two aspects are discussed in subsections 8.8.2 and 8.8.3, respectively.

First of all, however, let us note that section 8.6 showed that the model can be simplified for practical application, as far as the considered sources of variability are concerned. After all, some of the sources of variation appeared negligible compared to others, which implies that these can be ignored in practical studies. This for example seems the case for the intrinsic randomness in the demands and – as far as the (recurrent) peak period-related congestion is considered – the demand variations due to events. Note however that the example presented in section 8.6 clearly provides insufficient basis for definite conclusions on this, since the relative influences of only a part of all different sources of variability have been considered here, and only for one specific spatiotemporal configuration (with respect to both the network and the traffic demand pattern). Much additional analysis would be needed for this.

8.8.2 Validity for real-life evaluations

Recall from chapter 7 that different levels of model validity can be distinguished, the lowest of which is face validity. In the same chapter it was discussed that the developed model is largely face valid for the research tasks considered in this thesis, albeit with some deficiencies (which cannot easily be remedied). In case of a practical application for the evaluation of concrete projects, these deficiencies weigh more heavily (resulting in a lower face validity), however. This is because of the fact that in this case we are interested in the *quantitative values* of the effects, rather than merely in a *qualitative illustration* of the new insights obtained. Furthermore, an additional deficiency becomes relevant in this case. This concerns the fact that the route choices in the model are currently based on the shortest paths according to the *free flow* travel times. In particular for the peak periods, this is clearly unrealistic. For the research tasks considered in this project, this deficiency is not considered very important. In practical applications, it will be relevant however.

Of course, the deficiencies will lead to errors in both the results for the base scenario and the results for the scenario with the considered measure. As a result, these errors will partially cancel each other out. It is clear, however, that the results for the two scenarios will not be affected equally. This means that the deficiencies will lead to errors in the calculated effects of the measure.

In chapter 7 it was also discussed that for the research tasks at hand, construct validity is *relatively* easily achieved, because of the fact that all inputs and model parameters can be given any value within their 'realistic range'. In case of a practical application of the model, this is obviously not the case. In this situation, the model should be tailored to the specific situation at hand, which necessitates a proper calibration. In section 7.3 it was argued, however, that the feasibility of such a calibration is rather doubtful. The only way to 'tune' the model's parameter settings to a practical situation is then by conducting *location-specific* empirical research on the different influence factors to which the parameters relate. However, this will clearly be too time-consuming for a practical study. Moreover, a really good end result (in terms of construct/predictive validity[106]) is of course still not to be expected then.

In view of the above, it can be concluded that in practical applications, the model can only be used in a *qualitative* way, to find out whether certain effects (i.e. benefits or detriments) of a measure may be overlooked (or considerably underestimated) in the evaluation according to the traditional approach. It will not be possible to find the detailed quantitative values of these effects, because of a lack of sufficient model validity for this purpose.

8.8.3 Required computation time

For practical application of the model in evaluations of concrete projects, it is obviously strongly desirable that the required computation time is limited. This is currently not the case, however. In section 6.6.1 it was explained that the required computation time is in the order of multiple days or weeks, which is due to the large number of simulation runs required. Because of this, it is desirable to look for ways to reduce this computation time.

There are three ways to achieve such a reduction:

- by using faster computers (or multiple computers in parallel),
- by lowering the number of simulation runs performed, or
- by increasing the speed of the developed model (reducing the amount of computation time required per individual simulation run).

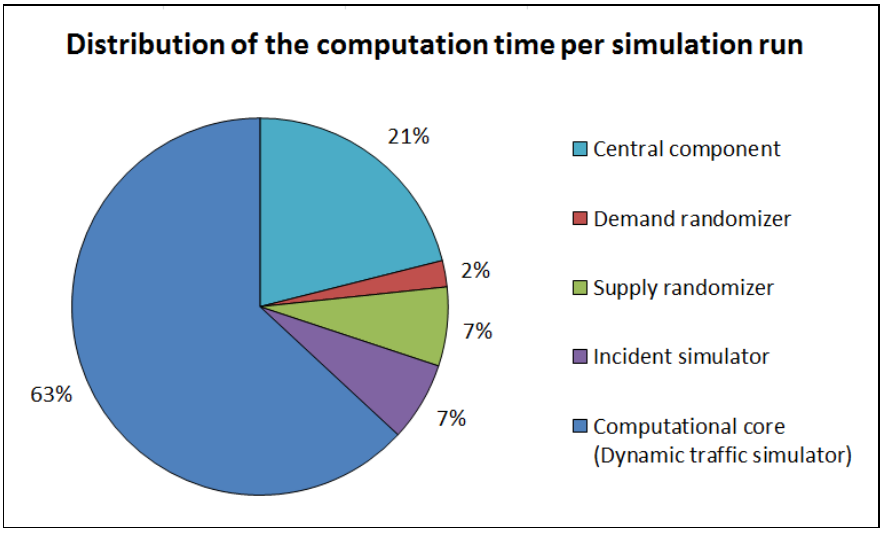
[106] Note that if a model is not calibrated to empirical data, the distinction between the construct and predictive validity of this model is not relevant.

The first strategy speaks for itself and consequently needs no further explanation. Possible ways to reduce the required number of simulation runs have been considered in section 6.6 (i.e. by implementing more efficient sampling techniques). It is important to note that the required number of simulation runs *cannot* be reduced by simplifying the model. For given performance indicators and a given requirement with respect to the statistical accuracy that is to be achieved, this number is directly governed by the statistical properties of the traffic conditions.

An important remark is to be made, however, with respect to the level of statistical accuracy that one requires. In view of the model's limited validity for practical applications (which limits its use to making qualitative observations), it obviously does not make sense to compute the considered performance indicators with a very high statistical accuracy. For increasing statistical accuracy, any further improvements of this statistical accuracy will gradually become more and more negligible compared to the inaccuracies that are due to deficiencies in the model itself, resulting in the overall accuracy of the model outputs hardly improving anymore. This means that a considerable amount of computation time can be saved by avoiding performing a senseless large number of simulation runs.

Finally, a reduction of the amount of computation time can be sought in the simulation speed of the model. Obviously, this involves a tradeoff between computation time and model accuracy. Currently, one simulation takes about 1.9 minutes (on a computer with a 2 GHz processor). The figure below shows the approximate distribution of this computation time over the different model components.

Figure 8.38: Distribution of the computation time of one simulation over the different model components



Clearly, the largest part of the computation time is consumed by the dynamic traffic simulator. Still, this simulator should be regarded as being rather fast, considering that it needs only slightly more than one minute to simulate the traffic operations

- on a reasonably sized network,
- over the period of a whole day,
- with physical queuing (based on a discretization of the kinematic wave equations),
- with a time step of 5 seconds and a (corresponding) cell length of 167 m or smaller (depending on the speed limit of the link),
- with a destination-specific modeling of the traffic propagation over the network, and
- with a simultaneous tracking of some output indicators (i.e. numbers of lost vehicle hours and realized travel times).

Increasing the speed of the computational core (the traffic simulator)

In order to increase the speed of the traffic simulator, one or more of these features would have to be given up. Obviously, we cannot give up considering a reasonably sized network. The network needs to be sufficiently large to include the network effects of traffic congestion (i.e. blocking back effects and the temporal redistribution effect[107]), and to limit the relative influence of the network boundaries. Of course, the tracking of certain output indicators during the simulation cannot be given up either.

If we are only interested in (the effects on) the peak period related traffic congestion, the computation time can be limited by simulating only the peak periods of the day. Obviously, important information with respect to off-peak period traffic congestion might be missed in this case. Furthermore, the calculation of the indicators relating to the *overall* travel time distributions is not possible then.

Theoretically, the required simulation time may significantly be reduced by using a dynamic traffic simulator without physical queuing. In this case, JDSMART would be replaced by a simulator with a non-physical queuing mechanism, like vertical queuing, horizontal queuing, or traffic propagation via travel time functions. Such simulators are typically much less computationally intensive than cell transmission models such as JDSMART. However, the modeling of blocking back effects is not accurate then, or even absent. In view of the very important role that these blocking back effects can have in the benefits of traffic measures (which may prevent these blocking back effects from occurring), the replacement of JDSMART by such a simulator with non-physical queue modeling is not advisable.

[107] See section 2.3 for a description of these network effects. Note that the route choice effect of traffic congestion is not taken into account in the developed model.

Given the model's limited validity for practical applications (which limits its use to making qualitative observations), it is acceptable to allow some inaccuracy (including some numerical diffusion) in the traffic flow modeling due to a relatively rough discretization. If the time step and cell lengths would be doubled (resulting in a new time step of 10 seconds and cell lengths of maximal 333 m), the required simulation time can be expected to decrease with a factor of approximately 4 (= 2^2). It is advisable, however, to perform some tests to find out whether this indeed is still acceptable.

The destination-specific modeling of the traffic propagation over the network was considered important in order to be able to:

- include the demand effects of events in the model (which are limited to specific origin-destination relations only),
- properly model the influence of the intrinsic randomness in the origin-destination traffic demands, and
- include the effect that the aggregated turn fractions at the network nodes may be affected by traffic congestion in upstream parts of the network.

In view of the fact that the results of section 8.6 suggest that the importance of the demand effects of events and the intrinsic randomness in travel behavior is only limited, it might be considered to abandon this destination-specific modeling. It is not clear, however, to what extent the congestion-related variation in the aggregated turn fractions is important. Before deciding whether or not to give up the destination-specific aspect of the traffic flow modeling, this would have to be studied first.

It should be noted, however, that in JDSMART the abandoning of the destination-specific traffic modeling will not yield such a large reduction in simulation time as in other simulators. This is due to the fact that JDSMART deals with this 'multi-class' aspect of the traffic in a more efficient way than other multi-class models, as discussed in section 5.2.1. However, if this multi-class aspect (which was one of the reasons for selecting JDSMART – see section 5.2.1) would be given up, it could also be considered whether it would be feasible to replace JDSMART by the (single-class) simulator 'Flowsim-Live'. This simulator is an 'online' version of the earlier mentioned model 'Flowsimulator' (Kijk in de Vegte & Van Toorenburg, 2009). This online version uses the computing power of the graphics card of the computer, which increases its speed by a factor of 30. This would obviously be very beneficial to the practical applicability of the model. Just like JDSMART, this simulator is a cell transmission model (with physical queuing), resulting in an accurate modeling of spillback effects.

It should be noted, however, that rather than abandoning the multi-class feature of JDSMART, it would actually be desirable to *expand* this concept even further, by distinguishing traffic flows not simply by destination, but rather by *route,* in order to be able to:

- properly model the (regular) variations in route choice with the time of the day, and
- enhance the model with a feedback mechanism from the (actual) traffic conditions to the route choices.

Obviously, this desire is incompatible with the need to reduce the computation time.

Increasing the speed of the other model components

Figure 8.38 showed that the largest part of the computation time is consumed by the dynamic traffic simulator. This certainly does not mean, however, that the computation time of the other model components is insignificant. Together, these account for more than one third of the total computation time. This means that if the total computation time is to be reduced really drastically, their computation time will have to be reduced as well.

From Figure 8.38 it can be seen that the 'central component' of the model (which manages all other components) is clearly the largest contributor to the computation time of these other model components. More than 60% of its computation time is spent on the updating of the demand and supply variables during the simulation process (based on the tables generated by the demand and supply randomizers).

The most effective way to reduce this updating time is by using a larger time interval for the randomization. In order for the variations in the traffic conditions to be properly reflected in the model output, it is important to choose this time interval sufficiently small compared to the time scales of the sources of variation. It is hypothesized that for this a time interval in the range of 5 to 15 minutes is required. In this project a safe value of 5 minutes has been chosen. It could be considered, however, to use a 10 or 15 minute interval, in order to reduce the computation time. Of course, this modification would have to be preceded by testing the output´s sensitivity to different time intervals. Note that the problem discussed in section 6.3 (i.e. the partial mismatch between the demands and the capacity realizations) will increase if the time interval of the randomization is increased.

The use of a larger randomization interval will not only reduce the computation time of the *central component* of the model. Obviously, the computation time of the demand and supply randomizers will be reduced as well, since less randomized values are to be generated in this case. The same applies to the computation time of the incident simulator.

Another way to reduce the updating time is by varying the supply parameters at the level of *links*, rather than at the level of the *individual cells of the links*. This strategy is however estimated to be less effective than the use of larger randomization time intervals. This is due to the fact that in the MATLAB code of the central component, the updating procedure is invoked at the *link level* already. The updating procedure itself (which considers the individual cells of the link in question) is implemented in the Java code of the traffic simulator, resulting in a relatively low computational cost.

Furthermore, this transition to a link-based variation approach will not yield a substantial reduction in the computation time of the supply randomizer, as the change of the randomization time interval does. This is due to the fact that the cell-based aspect accounts for only a few percent of the total computation time of the supply randomizer.

For the incident simulator, this is different. Almost 90% of its computation time is spent on the random generation of incidents per cell of the network. Therefore, a significant reduction in computation time could be achieved by generating these incidents at the level of links, rather than at the level of cells. As a second step, then, the generated incidents might be randomly assigned to one of the cells of the link in question. The implicit simplifying assumptions associated with this approach seem acceptable.

It should be noted that the computation time cannot be substantially reduced by omitting one or more of the sources of variability. By far, incidents and the intrinsic randomness in driving behavior are the sources of variability that are the most computationally expensive[108]. In view of the fact that the intrinsic randomness in driving behavior was observed to play a very important role in the performance of the traffic system (see section 8.6), this source of variability clearly cannot be omitted from the model. Although incidents have not been included in the model evaluations presented in the previous sections (because of a problem with respect to the computed congestion levels, as discussed in section 8.5), these are expected to play an important role as well. Therefore, these in fact should not be omitted from the model either. For these incidents there is however an opportunity to drastically reduce the associated computation time, as has been discussed above.

[108] As far as the intrinsic randomness in driving behavior is concerned, this is not so much due to the (cell-based) generation of capacity values from their (link-based) probability distribution functions, but rather due to the computationally expensive calculation of the parameters of these distributions. The variations in these parameters reflect the capacity effects of the other sources of variability, as discussed in chapter 6.

Summary of the possibilities to increase the speed of the model

In the above it was explained that in order to substantially increase the speed of the developed model (to make it suitable for practical application), both the speed of its computational core (i.e. the dynamic traffic simulator) and the speed of the other model components would have to be improved. Different possibilities have been identified for this. All of these possibilities come at the expense of the quality of the output of the model. In view of the importance of a limited computation time, we may simply have to put up with this, however. As long as the introduced errors/inaccuracies are limited in comparison with the deficiencies already present in the model (which limit its practical applicability to making qualitative observations), they might be considered acceptable.

The options identified for increasing the computation speed of the computational core are:

- the use of a coarser discretization (i.e. longer cells and a larger time step)
- the replacement of JDSMART by the simulator Flowsim-Live (which however does not model the destination-specific aspect of the traffic flow over the network)

For the other model components the following options have been identified:

- using a larger time interval for the demand and supply randomization (like a 10 or 15-minutes interval, instead of a 5-minutes interval)
- generating incidents at the *link* level, rather than at the *cell* level (after which the generated incidents might be randomly assigned to one of the cells of the link in question)

Before implementing one or more of these options, it should first be investigated whether these indeed are acceptable (by performing some comparative test runs). For the last mentioned option this seems not necessary, however.

9. Conclusions and recommendations

9.1 Introduction

In the research presented in this thesis it has been investigated what kind of additional or revised insights can be obtained from evaluations of the traffic system's performance (in the context of considering taking strategic measures to alleviate congestion) when the inherent variable nature of daily traffic congestion on the motorway network is explicitly taken into account. (As opposed to the insights obtained by evaluations according to the more 'traditional' approach, in which only a kind of 'representative' situation is evaluated.)

This final chapter presents the main conclusions that can be derived from this research (section 8.2). Furthermore, it discusses the practical implications of the research findings obtained (section 8.3), and it comes up with a number of recommendations for further research (section 8.4).

9.2 Conclusions

1. **A 'representative' calculation of the traffic conditions (that is, a calculation in which all traffic demand and supply variables are taken at their 'representative' levels, which for instance could be their means or medians), does not give a good impression of the performance of the traffic system.**

   This is due to facts that:

   - The (day-to-day) *travel time uncertainty* aspect of this performance is disregarded (since the day-to-day variability in the traffic conditions is completely neglected in such a calculation), while this was identified as an important component in the societal costs of traffic congestion.
   - The 'representative' calculation *underestimates* the traffic congestion in certain respects, meaning that the traffic congestion calculated for the 'representative' situation is not so 'representative' itself. This is related to the predominantly negative influence of the (neglected) variations. This predominantly negative influence arises from:
     - the purely negative nature of some of the sources of variability (such as incidents or bad weather events)
     - the non-linearity in the traffic system (i.e. the fact that the congestion level is a non-linear function of the difference between demand and supply, causing that the detriments of 'negative occurrences' are often larger than the benefits of 'positive occurrences')

This underestimation will not be satisfactorily remedied by calibrating the traffic model to the 'representative' congestion level.

2. **By explicitly considering the different sources of variability in evaluations of the performance of the traffic system, new insights can be obtained into the relative importance of these sources.**

   In the analyses that have been performed in this thesis (for the purpose of demonstration) it was found that:

   - The capacity variations due to the intrinsic randomness in human driving behavior play a central role in (peak period-related) traffic congestion on weekdays, considerably increasing both the mean and median levels of the amount of traffic congestion, the variability of the amount of traffic congestion, and the instability of the traffic conditions (i.e. the extent to which one's realized travel time might deviate from the instantaneous travel time at the moment of departure).
   - The seasonal demand variation over the months of the year plays an important role in weekday congestion as well
   - The intrinsic randomness in travel behavior, the ambient conditions (weather and daylight/darkness) and events play a much smaller role in weekday congestion. It is to be noted, however, that very occasionally, extreme weather conditions may disrupt traffic to a very severe extent, causing large costs to society.
   - Ignoring the influences of incidents and road works, events seem to be the most important source of weekend day traffic congestion.

3. **Explicitly accounting for the different sources of variability can provide us with new insights into the effectiveness of specific measures that are proposed to alleviate traffic congestion. More specifically, the 'traditional' way of evaluating the effectiveness of a measure (in which only a kind of 'representative' situation is considered) may result in a <u>SIGNIFICANT UNDERESTIMATION OF THE BENEFITS</u> of this measure.**

   This is due to facts that:

   - A 'representative' calculation underestimates the amount of traffic congestion (as discussed above), and thereby underestimates the potential benefits of proposed measures (aimed at alleviating this congestion) as well.
   - In an evaluation according to the traditional approach potential benefits of a considered measure may remain unnoticed due to nonlinearities and trend breaks in the behavior of the traffic system. This applies particularly to (the prevention of) spillback of congestion to other network elements. If this spillback occurs only in part of the occasions (say less than 50%), it will not be

included in the representative analyses. Consequently, the benefits achieved on these other network elements will not be reflected in the evaluation results.

- In an evaluation according to the traditional approach, no information is obtained on the improvements in travel time uncertainty (due to the fact that the day-to-day variability in the traffic conditions is not considered), while this improvement might be an important component in the benefits of a traffic measure.

The precise nature and extent of the additional/revised insights into the effectiveness of a measure will be highly context and measure specific. Of course, these new insights are not necessarily all positive in nature. Some more negative aspects of a measure could be brought to light as well.

9.3 Practical implications of the results

1. **The last conclusion mentioned above (no 3) implies that in practice more systematic attention should be given to the inherent variability in traffic, when evaluating the effectiveness of measures that are proposed to alleviate congestion. Because of the complexity involved (especially in case of heavily loaded networks in highly urbanized areas), this would have to be done by using a model in which the different sources of variability are explicitly accounted for, such as the model developed in this project.**

   It should be stressed, however, that this model was developed solely for the research task considered in this project, and thus not directly for practical application in the evaluation of concrete projects. In such practical evaluations, the model can only be used in a *qualitative* way, to find out whether certain effects (i.e. benefits or detriments) of a measure may be overlooked (or considerably underestimated) in the evaluation according to the traditional approach. The model is not sufficiently valid for making *quantitative inferences* on the effects of concrete real-life measures, due to the following problems:

   - In this case, some deficiencies in the model would become relevant, related to some modeling issues that require substantial further research (section 9.4.1 – sub 2-5). In the model applications presented in this thesis these deficiencies were less important, because of the fact that it was not intended to come up with firm quantitative conclusions for one specific real-life situation.
   - A proper calibration would be required, in order to tailor the model to the specific situation at hand. Such a calibration seems unfeasible, however. The only way to 'tune' the model's parameter settings to a practical situation is then by conducting location-specific empirical research on the different influencing factors to which the parameters relate. However, this will be too

time-consuming for a practical study. Moreover, a really good end result (in terms of model validity) is of course still not to be expected then.

Another issue relevant to the practical applicability of the model is its computation time. Currently, the computation time required for one model run is in the order of days or weeks, which is related to the large number of simulations that is to be performed. For practical applications, this computation time would have to be reduced. Such a reduction could be achieved in three different ways:

- Using faster computers (or multiple computers in parallel).
- Reducing the required number of simulation runs, by implementing a more efficient sampling technique (9.4.1 – sub 7). It is important to note that the required number of simulation runs *cannot* be reduced by simplifying the model. For given performance indicators and a given requirement with respect to the statistical accuracy that is to be achieved, this number is directly governed by the statistical properties of the traffic conditions.
- Increasing the speed of the developed model (i.e. reducing the amount of computation time required per individual simulation run), involving a tradeoff between computation time and model accuracy (9.4.1 – sub 8).

2. **When considering taking strategic measures to alleviate traffic congestion, priorities can be set by considering the relative importance of the various primary sources of congestion, using the results of model evaluations as performed in this project.**

   The results presented in this thesis indicate that:

   - Peak period-related traffic congestion might relatively effectively be alleviated by taking measures which reduce the effects of the intrinsic randomness in human driving behavior, although these might not be so easy to find.
   - Up to a certain extent, weekend day traffic congestion can be avoided by a careful planning of events, including the provision of route guidance, adequate parking facilities and appropriate alternative transport modes.

3. **In view of the important role of the intrinsic randomness in human driving behavior (section 9.2 – sub 2), it is hypothesized that the impact of future Advanced Driver Assistance Systems might actually be larger than currently anticipated upon. Insofar as these systems would <u>exclude</u> the human factor (by actively taking over control), they could reduce this intrinsic randomness in the capacities[109], and thereby significantly improve the traffic conditions.**

[109] Note that if such a system allows drivers to make their own settings (based on their own driving preferences), the reduction in the intrinsic capacity randomness will obviously be much smaller.

9.4 Recommendations for further research

9.4.1 Improvement of the developed model

1. **Much more empirical research is necessary with respect to the frequencies and effects of the different sources of variability, in order to give the model a more solid foundation.**

2. **As far as the random variations in capacities due to the intrinsic randomness in human driving behavior are concerned, this additional empirical research cannot be restricted to 'finding some empirical numbers'. For this source of variation additional research is necessary which is more fundamental in nature. This research should be directed mainly at the <u>spatial dependency</u> in this variation (which arises from the fact that it is partly the same traffic which traverses the subsequent sections of a road).**

   To the best knowledge of the author, this spatial dependency has never been studied. The modeling assumptions on this dependency appear to have an important effect on the generated congestion levels, however.

3. **More research is necessary into the issue of how to deal with the problem of confronting an n-minute capacity distribution with a traffic demand which varies with the time interval of the simulation of the traffic propagation (which has to be set at a few seconds or even less in order to obtain a sufficiently accurate traffic flow modeling).**

4. **It is advised to implement a feedback in the model from the actual traffic conditions to the route choices (reflecting the influence of traffic information provision).**

   For this, it is recommended to use the hybrid route choice model described in (Pel et al, 2009). A solution is then to be found for the problem that this is currently not feasible in terms of computation time.

5. **It is advised to incorporate feedbacks to the other demand governing travel choices (i.e. trip making decisions, destination choices, mode choices and departure time choices) as well.**

   Currently, this is unfeasible, however, due to a lack of knowledge on these effects. This knowledge can only be obtained by conducting dedicated research.

6. **It is recommended to give further consideration to the validity of the model.**

   In this project the evaluation of this validity has been largely limited to an analysis based on theoretical considerations (apart from some quantitative considerations on the generated congestion levels). No systematic validation on *empirical data* has been performed, because of the unfeasibility of a proper calibration of the model to a specific real-life situation[110]. However, an alternative approach could be to perform this validation at a more general level. In this case, it would be analyzed whether the *orders of magnitude* of the calculated indicator values (made dimensionless, for example by division by the free flow travel time) are roughly in the range observed in practice, and whether the *shapes* of the computed distributions are consistent with those found in reality (for example in terms of the ratios between different statistics of these distributions). It is also strongly advised to subject the model to a thorough check on possible remaining programming bugs.

7. **It is advisable to study the option to implement a more advanced (i.e. more efficient) sampling method, in order to reduce the required number of simulation runs.**

   Possible methods could be the Latin Hypercube Sampling approach, or the Importance Sampling technique. The latter is expected to yield the largest reduction.

8. **In order to increase the speed of the developed model (i.e. reduce the amount of computation time required per individual simulation run), it could be considered to implement the modifications listed below. Since these modifications come at the expense of the quality of the output of the model, it should first be studied whether they are acceptable. This involves a tradeoff between output accuracy and computation time.**

   - Using a coarser discretization in the dynamic traffic flow simulation (i.e. longer cells and a larger time step).
   - Replacing JDSMART (i.e. the dynamic traffic simulator which is used as the computational core of the model) by the simulator Flowsim-Live (which however does not model the destination-specific aspect of the traffic flow over the network).
   - Using a larger time interval for the demand and supply randomization (like a 10 or 15-minutes interval, instead of a 5-minutes interval).
   - Generating incidents at the link level, rather than at the cell level (after which the generated incidents might be randomly assigned to one of the cells of the link in question).

[110] Note that such a calibration to a specific situation was not required for the research tasks at hand, since it was not intended to come up with results for one particular actually existing situation.

9.4.2 Further exploration of (possible) new insights

1. **By way of illustration, in this project six sources of variability have been compared on their relative contribution to traffic congestion. By extending this investigation to the other sources of variability, and repeating it for different spatiotemporal configurations (with respect to both the infrastructure network and the associated demand pattern), a more complete and more general picture can be obtained of the relative importance of the different sources of variability.**

   This information may be valuable in two different ways:

   - It might yield important insights into how traffic congestion can be remedied most effectively. For example, if the relative contribution of incidents proves to be large, incident management measures might have a relative high effectiveness
   - Insofar as certain sources of variability are found to be negligible compared to others (as a general rule), these can be omitted in future model evaluations (both in research studies and in practical applications).

   Of course, a proper sensitivity assessment with respect to important model parameters should be part of this analysis as well.

2. **By studying different types of measures with the developed model, it could be investigated to what extent the explicit consideration of the variability in traffic provides us with revised insights into the relative effectiveness of these measures.**

   These measures may include *traffic management measures* (such as ramp metering, dynamic use of the hard shoulder lane, provision of route information/advices, and incident management measures), but also *demand management measures* (i.e. measures aimed at reducing or redistributing the traffic demand, such as road pricing) and *infrastructural measures* (such as the addition of an extra lane, the creation of traffic buffers, and the physical separation of long distance and local traffic). Obviously, the effectiveness of measures may be highly context-dependent. This means that the measures should be evaluated in multiple different spatiotemporal configurations (with respect to both the infrastructure network and the associated demand pattern), in order to arrive at conclusions with some general validity. Absolute generality is of course not attainable, however.

3. **In section 9.3 – sub 3, it was hypothesized that the impact of future Advanced Driver Assistance Systems might actually be larger than currently anticipated upon. Although the effects on the variability of the capacity would have to be investigated by means of microscopic analysis, the final impacts of these capacity effects could be studied with the model developed in this project.**

   It should be noted that it is not inconceivable that such systems also might increase the vulnerability of the traffic system to a certain extent, by introducing the possibility of 'common mode failure'. If all vehicles equipped with a certain actively controlling system would react in one uniform way to a certain (external or internal) condition, which may be a desirable reaction from the point of view of the individual vehicle, but not from the point of view of the traffic flow as a whole, this could potentially be very detrimental to the traffic conditions. Insofar as this kind of reactions on certain circumstances could be translated into effects on the capacity distribution functions, this issue could be studied as well with the developed model.

4. **It might be hypothesized that certain traffic management measures that help to increase the throughput of existing infrastructure might strongly deteriorate the performance of the traffic system under certain disturbing conditions, which could be studied with the developed model.**

   This for example might be the case for rush-hour lanes. The availability of such a rush-hour lane will attract additional traffic. If then a situation occurs in which this rush-hour lane cannot be opened to traffic (for example due to adverse weather conditions or an incident), this can be expected to have very detrimental consequences. Furthermore, if the hard shoulder of the road is sacrificed for the rush-hour lane, incidents are more likely to have larger disrupting effects on the traffic conditions.

   The model developed in this project could be used to investigate to what extent this type of effects is important in the overall performance of a traffic management measure, and under which conditions (such as the utilization rate of the infrastructure supply). This is an interesting topic for further research.

Literature

Adviesdienst Verkeer en Vervoer (AVV) (1997), *Kwaliteit Verkeersafwikkeling 1994*, Rotterdam

Adviesdienst Verkeer en Vervoer (AVV) (2000), *Mobiliteit verkend; Feiten achter het NVVP*, Rotterdam

Adviesdienst Verkeer en Vervoer (AVV) (2002), *Capaciteitswaarden Infrastructuur Autosnelwegen, Handboek, versie 2*

Adviesdienst Verkeer en Vervoer (AVV) (2002b), *Meldwerk Jaarbericht 2001, Overzicht van in Meldwerk geregistreerde wegafzettingen*

Adviesdienst Verkeer en Vervoer (AVV) (2002c), *Evaluatie groot onderhoud aan de A10-West, Integraal eindrapport*

Adviesdienst Verkeer en Vervoer (AVV) (2003), *Meldwerk Jaarbericht 2002, Overzicht van de in Meldwerk ingevoerde wegafzettingen*, Rotterdam

Adviesdienst Verkeer en Vervoer (AVV) (2004), *Fileverkenning, De ontwikkeling van de vertragingen op het Nederlandse autosnelwegennet*, Rotterdam.

Adviesdienst Verkeer en Vervoer (AVV) (2004b), *Ontwikkelingen Verkeer en Vervoer 1990 – 2020, Probleemverkenning voor de Nota Mobiliteit.*

Adviesdienst Verkeer en Vervoer (AVV) (2006), *Evaluatie van de Mobiliteitsbeïnvloedende maatregelen tijdens het groot onderhoud A4/A10 Zuid, Hoofdrapportage*

Adviesdienst Verkeer en Vervoer (AVV) (2006b), *Effecten Verkeersmanagement, Cijfers van meer dan 100 (praktijk)evaluaties uit Nederland* (PowerPoint-presentatie)

Adviesgroep voor verkeer en vervoer (AGV) (1997), *Aanbod-capaciteitsmodellen, Verkenning van mogelijkheden tot integratie*, Nieuwegein

Agarwal, M., T.H. Maze & R. Souleyrette (2005), *Impact of Weather on Urban Freeway Traffic Flow Characteristics and Facility Capacity*, Aurora, Iowa State University, Ames

Al-Kaisy, A.F., F.L. Hall & E.S. Reisman (2002), *Developing passenger car equivalents for heavy vehicles on freeways during queue discharge flow*, Transportation Research Part A, volume 36, pp. 725-742

Bexelius, S., and H.P.A.M. Kengen (1993), *Dagen met topdrukte in het verkeer, Kenmerken en voorspelbaarheid*, Verkeerskundige werkdagen 1993 (publikatie 73) deel II, CROW, Ede, pp. 717-729 (ISBN: 90 6628 153 7)

Bogers, E.A.I., J.W.C. van Lint & H.J. van Zuylen (2008), *Reliability of travel time: Effective measures from a behavioral point of view*, Transportation Research Record, volume 2082, pp. 27-34

Botma, H. (1999), *Queueing Model To Determine Variability of Recurrent Congestion at Freeway Link over a Year*, Transportation Research Record 1678, Transportation Research Board of the National Academies, Washington, pp. 83-89

Brilon, W., J. Geistefeldt, & M. Regler (2005), *Reliability of Freeway Traffic Flow: A Stochastic Concept of Capacity*, Proceedings of the 16th International Symposium on Transportation and Traffic Theory, Maryland, pp. 125 – 144 (ISBN: 0 08 044680 9)

Brilon, W., J. Geistefeldt & H. Zurlinden (2007), *Implementing the Concept of Reliability for Highway Capacity Analysis*, Transportation Research Record 2027, Transportation Research Board of the National Academies, Washington, pp. 1-8

Brilon, W., and M. Ponzlet (1996), *Variability of Speed-Flow Relationships on German Autobahns*, Transportation Research Record, Volume 1555, pp. 91-98

Buishand, T.A. & C.A. Velds (1980), *Neerslag en verdamping*, KNMI, De Bilt

Bureau Goudappel Coffeng (BGC) (1986), *Wetmatigheden in verkeersintensiteiten*, Deventer

Changnon, S.A. (1996), *Effects of summer precipitation on urban transportation*, Climatic Change, Vol. 32, no. 4, pp. 481-494

Chen, E.J. & W.D. Kelton (2001), *Quantile and histogram estimation*, Proceedings of the 2001 Winter Simulation Conference, B.A. Peters, J.S. Smith, D.J. Medeiros & M.W. Rohrer (eds.)

Chung, E., O. Ohtani, H. Warita, M. Kuwahara & H. Morita (2005), *Effect of rain on travel demand and traffic accidents*, Proceedings of the 8th International IEEE Conference on Intelligent Transportation Systems, Vienna, pp. 1080-1083

Chung, E., O. Ohtani, H. Warita, M. Kuwahara & H. Morita (2006), *Does weather affect highway capacity?*, Proceedings of the 5th International Symposium on Highway Capacity and Quality of Service, Vol. 1, Yokohama, pp. 139-146

Cools, M., E. Moons & G. Wets (2008), *Assessing the Impact of Weather on Traffic Intensity*, TRB 87th Annual Meeting Compendium of Papers DVD

Dienst Verkeerskunde (1992), *Kwaliteit Verkeersafwikkeling*, Rotterdam

Dienst Verkeer en Scheepvaart (DVS) (2010), *Kwartaalmonitor bereikbaarheidsontwikkeling Hoofdwegennet, 4e kwartaal 2009, 1 oktober – 31 december 2009.*

Geistefeldt, J. (2009), *Estimation of Passenger Car Equivalents Based on Capacity Variability*, Transportation Research Record, Volume 2130, Transportation Research Board of the National Academies, Washington, pp. 1-6

Grol, R. van, J. Baak & M. Kouwenhoven (2004), *Gebruikershandleiding LMS-BT, LMS/NRM Betrouwbaarheids-module*, RAND Europe, Leiden

Harms, L. (2003), *Mobiel in de tijd, Op weg naar een auto-afhankelijke maatschappij, 1975-200*, Sociaal en Cultureel Planbureau, Den Haag (ISBN: 90-377-0125-6)

Hilbers, H., J. Ritsema van Eck & D. Snellen (2004), *Behalve de dagelijkse files, Over betrouwbaarheid van reistijd*, RPB, Den Haag / NAi Uitgevers, Rotterdam (ISBN: 90 5662 375 3)

Hofman, F. (2010), *Nieuwe versies LMS/NRM*, Presentation at the PLATOS Colloquium 2010, Utrecht.

Hogema, J.H. (1996), *Effects of rain on daily traffic volume and on driving behaviour*¸TNO Human Factors Research Institute, Soesterberg

Hoogendoorn, S.P. (2007), *Traffic Flow Theory and Simulation, CT4821*, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft

Hoogendoorn S.P., V.L. Knoop & H.J. van Zuylen (2008), *Robust Control of Traffic Networks under Uncertain Conditions*, Journal of Advanced Transportation, Vol. 42, No. 3, pp. 357-377.

Hoorn, T. van der (2007), *Files normeren? Rijksbeleid ten aanzien van congestie en files van 1979 tot heden*, Tijdschrift Vervoerswetenschap, volume 43, issue 2, pp. 14-21

Hranac, R., E. Sterzin, D. Krechmer, H. Rakha & M. Farzaneh (2006), *Empirical Studies on Traffic Flow in Inclement Weather*, Federal Highway Administration, Washington

Huiskamp, A. (2010), *IJzel en ijsregen*, http://www.vwkweb.nl/cms/index.php?id=354&option=com_content&task=view (accessed on 29-04-2010), Vereniging voor Weerkunde en Klimatologie

Immers, B., A. Bleukx, J. Stada, C. Tampère & I. Yperman (2005), *Robustness And Resilience Of Road Network Structures*, Katholieke Universiteit Leuven, Heverlee

Jong, G. de, et al. (2005), *Uncertainty in traffic forecasts: literature review and new results for the Netherlands*, RAND Europe, Leiden.

Keay, K. and I. Simmonds (2005), *The association of rainfall and other weather variables with road traffic volume in Melbourne, Australia*, Accident Analysis and Prevention 37, pp. 109-124

Kennisinstituut voor Mobiliteitsbeleid (KIM) (2008), *Doelen en daden, Herijking van de Nota Mobiliteit*

Kennisinstituut voor Mobiliteitsbeleid (KIM) (2009), *Mobiliteitsbalans 2009*, Den Haag (ISBN: 978 90 8902 058 1)

Khattak, A.J, and A. de Palma (1997), *The impact of adverse weather conditions on the propensity to change travel decisions: a survey of Brussels commuters*, Transportation Research Part A: Policy and Practice, Vol. 31, Issue 3, pp. 181-203

Kijk in de Vegte, N. & J. van Toorenburg (2009), *Flowsim-Live: scenario's doorrekenen on-the-fly met een on-line verkeersmodel*, 8th Dynamic Traffic Management Symposium, Rotterdam, May 14, 2009

KNMI (2002), *Langjarige gemiddelden en extremen, tijdvak 1971-2000*, http://www.knmi.nl/klimatologie/normalen1971-2000/ index.html, May 2002.

KNMI (2008), *Nader verklaard, Mistdagen*, http://www.knmi.nl/cms/ content/25320/mistdagen, December 22, 2008

KNMI (2009a), *Mist en nevel zijn de afgelopen 30 jaar sterk afgenomen, Goed nieuws voor het verkeer en voor zonaanbidders*, http://www.knmi.nl/cms/content/69838 /mist_en_nevel_ zijn_ de_afgelopen_30_jaar_sterk_afgenomen, October 28, 2009

KNMI (2009b), *Risicosignalering Winterse neerslag*

KNMI (2009c), *Nader verklaard, Sneeuwweetjes*, http://www.knmi.nl/cms/content/23103 /sneeuwweetjes, December 17, 2009

KNMI (2009d), *Neerslagstatistiek*, http://www.knmi.nl/klimatologie/achtergrondinformatie/ neerslagstatistiek.pdf, KNMI, De Bilt

Knoop, V.L. (2009), *Road Incidents and Network Dynamics, Effects on driving behaviour and traffic congestion (TRAIL Thesis Series no. T2009/13)*, TRAIL Research School

Kouwenhoven, M., G. de Jong & P. Rietveld (2005), *Reliability Ratio's voor het Goederenvervoer*, Rand Europe / VU Amsterdam

Kouwenhoven, M., Siemonsma, H. & R. van Grol (2006), *Voertuigverliesuren door Incidenten, Hoeveel kan incident-management bijdragen aan de vermindering van files?*, Rand Europe, Leiden

Kraaijeveld, R. (2008), *Onderzoek naar de identificatie van kwetsbare wegvakken, Onderzoek naar de bepaling van kwetsbaarheid van wegvakken op basis van empirische gegevens*, ITS Edulab, Delft

Lint, J.W.C. van (ed) (2008), *Traffic Theory for Intelligent Transport Systems for Road Transport*

Lint, J.W.C. van, H.J. van Zuylen & H. Tu (2008), *Travel time unreliability on freeways: Why measures based on variance tell only half the story*, Transportation Research Part A 42, pp. 258-277

Margiotta, R. (2009), *Developing and Predicting Travel Time Reliability*, PowerPoint presentation presented to ITS Georgia, October 5, 2009

Maze, T.H., M. Agarwal & G. Burchett (2005), *Whether Weather Matters to Traffic Demand, Traffic Safety, and Traffic Flow, Final General Report*, Aurora, Iowa State University, Ames

Meeuwissen, A.M.H., M. Snelder & J.M. Schrijver (2004), *Statistische analyse variabiliteit reistijden voor SMARA*, TNO Inro, Delft, (ISBN: 90 5986 095 0)

Mehran, B. & H. Nakamura (2009), *Considering travel time reliability and safety for evaluation of congestion relief schemes on expressway segments*, IATSS Research Vol.33, No. 1, pp. 55-70.

Ministerie van Verkeer en Waterstaat (2004), *Nota Mobiliteit, Naar een betrouwbare en voorspelbare bereikbaarheid.*

Ministerie van Verkeer & Waterstaat (2006), *Nota Mobiliteit, Deel IV: Na parlementaire behandeling vastgestelde PKB*

Ministerie van Verkeer en Waterstaat (2008), *Beleidskader Benutten, Eén van de pijlers voor een betere bereikbaarheid.*

Nicholson, A.J., J.D. Schmöcker, M.G.H. Bell & Y Iida (2003), *Assessing Transport Reliability: Malevolence and User Knowledge*, in: The network reliability of transport – Proceedings of the 1st international symposium on transportation network reliability (INSTR) (ISBN: 0 080 44109 2), M.G.H. Bell & Y. Iida (eds.), Elsevier Science Ltd., 2003, pp. 1-22

Pel, A.J, M.C.J. Bliemer & S.P. Hoogendoorn (2009), *Hybrid Route Choice Modeling in Dynamic Traffic Assignment*, Transportation Research Record 2091, Transportation Research Board of the National Academies, Washington, pp. 100-107

PB Farradyne (2000), *Traffic incident management Handbook*

Rand Europe (2004), *Ontwikkeling ex ante Instrument Betrouwbaarheid, Eindrapport*, Leiden

Rand Europe (2005), *The value of reliability in transport, Provisional values for the Netherlands based on expert opinion*, Leiden

Rijkswaterstaat (2009), *Trajectnota/MER Rijksweg 13/16 Rotterdam, Deelnota Verkeer*, Rotterdam

RPB (2004), *Behalve de dagelijkse files, Over betrouwbaarheid van reistijd*, NAI Uitgevers, Rotterdam.

Savelberg, F. (2008), *Doelen en daden, Herijking van de Nota Mobiliteit*, Kennisinstituut voor Mobiliteitsbeleid (KIM), Den Haag (ISBN: 978-90-8902-038-3)

Scholtens, B. (2001), *Les van de snelweg*, De Volkskrant, august 25, 2001

Smith, B.L., K.G. Byrne, R.B. Copperman, S.M. Hennessy & N.J. Goodall (2003), *An investigation into the impact of rainfall on freeway traffic flow*, University of Virginia, Charlottesville

Snelder, M. & J. Schrijver (2008), *De toekomst voorspeld in de toekomst: files realistisch gemodelleerd*, Colloquium Vervoersplanologisch Speurwerk, Santpoort

Snelder, M., J. Schrijver & B. Immers (2009), *Nietsdoen is geen optie – methodes voor het bepalen van de robuustheid van wegennetwerken*, Colloquium Vervoersplanologisch Speurwerk, Antwerpen

Stembord, H.L. (1991), *Quality of service on the main road network in the Netherlands*, in: Highway capacity and level of service, proceedings of the international symposium on highway capacity (ISBN: 90 54 10 011 7), U. Brannolte (ed.), 1991, pp. 357-365

SWOV (2009), *SWOV Fact Sheet, The influence of weather on road safety,* Leidschendam

Toorenburg, J.A.C. van (1988), *Wachttijd-karakteristieken voor hoofdwegen*, Transpute, Gouda

Toorenburg, J.A.C. van (1991), *Performance of motorways and trunk routes at high traffic volumes*, Highway capacity and level of service, proceedings of the international symposium on highway capacity (ISBN: 90 54 10 011 7), U. Brannolte (ed.), pp. 413-418

Toorenburg, J.A.C. van (1997), *VQ-NET – Uitbreiding model 'Kwaliteit Verkeersafwikkeling' tot netwerk-effecten*, Transpute

Toorenburg, J.A.C. van (2002), *Notitie: Rijtijden en trajectsnelheden op A7/A8 berekend met het model Flowsimulator / Verkeerskwaliteit*, Transpute BV, Gouda

Toorenburg, J.A.C. van (2003), *Memo: 'Ontwikkeling onbetrouwbaarheid hoofdwegennet 2000-2020'*, Transpute BV

Transportation Research Board (TRB) (2000), *Highway Capacity Manual (HCM)*¸ National Research Council, Washington

Transpute (1989), *Wachttijd-karakteristieken voor hoofdwegen, Grafieken*, Gouda

Transpute (1989b), *KWAST/SVV-norm – Uitwerking op strengen*, Gouda

Transpute (2003), *Flowsimulator, Beschrijving van het model*, Gouda

Tu, H. (2008), *Monitoring travel time reliability on freeways (TRAIL Thesis Series no. T2008/7)*, TRAIL Research School

Vries, L. de, and J. Praagman (1995), *Variatiestructuur Verkeersintensiteit Hoofdwegennet*, CQM BV, Eindhoven

Weijermars, W.A.M, and E.C. van Berkum (2005), *Road traffic patterns, Analysis of daily patterns on a Dutch Highway using cluster analysis*, University of Twente, Enschede

Weijermars, W.A.M (2007), *Analysis of urban traffic patterns using clustering (TRAIL Thesis Series no. T2007/3)*, TRAIL Research School

Zuurbier, F.S., H.J. van Zuylen, S.P. Hoogendoorn & Y. Chen (2006), *A generic approach to generating optimal controlled prescriptive route guidance in realistic traffic networks*, Proceedings of the 85th Annual Meeting of the Transportation Research Board, Washington

Zuurbier, F.S. (2010), *Intelligent Route Guidance (TRAIL Thesis Series no. T2010/8)*, TRAIL Research School

Appendix 1 - Existing models dealing with variability

A1.1 SMARA
(Simulation Model for Analyzing the Reliability of Accessibility)

SMARA is a model for analyzing the variability in travel times, developed by TNO Inro (a Dutch institute for applied scientific research in the field of traffic and transport) for the Dutch Spatial Planning Agency. It calculates travel time distributions between origin and destination zones that together cover the whole of the Netherlands, based on variations in the demand and supply variables. The model is developed for analyzing the situation on working days only. The situation on weekend days (or public holidays) cannot be considered with the model. The model is primarily meant as a tool enabling to make statements at the strategic level about the reliability of the origin-destination travel times, for example relating to the (expected) development of this reliability in the coming years, the mutual differences between different regions, and the effects of infrastructural measures or new residential areas (Meeuwissen et al, 2004). SMARA does not only consider the motorway network. Up to a certain degree the secondary and local networks are included as well.

SMARA is based on a static traffic assignment model. This means that the evolution of the traffic conditions over time is not considered. The model considers two different time periods of the day: a 'typical peak hour' [111] and a 'typical off-peak hour' [112]. The fact that the model combines all different peak hours (or all different off-peak hours, excluding the period 0-6h) into one number or probability distribution actually is a negative feature of the model. This way results are obtained that are less connected to the way in which traffic congestion causes costs to society (as described in chapter 3).

The variability in traffic demands and supply characteristics is addressed by means of the Monte Carlo simulation technique. The model performs a large number of simulation runs (one for each working day of the year). For each of these runs the nominal values of all origin-destination traffic demands and road section capacities are multiplied by a number of 'correction factors'. By these correction factors the influences of the different sources of variability are expressed. Each of them is drawn from a (discrete) probability distribution (a so-called 'probability table'), expressing the effects and the corresponding probabilities/frequencies of the various sources of variability. These

[111] A typical peak hour is in terms of level and composition equal to the average of the morning peak (7-9h) and the evening peak (16-18h), and in terms of direction equal to the morning peak (Meeuwissen et al, 2004)

[112] A typical off-peak hour is defined as the average of the periods 6-7h, 9-16h, and 18-24h.

effects and probabilities/frequencies have been derived from literature or analysis of empirical data. The nominal traffic demand levels for all origin-destination relations are obtained from another model (viz. a 'standard' traffic model).

For two of the sources of traffic supply variability not only the capacities of the road sections are 'corrected', but the speed levels as well:

- In the case of road works an upper limit is set on the speed on the affected road sections. This upper limit represents the (reduced) speed limit in the work zone.
- For the various combinations of weather and luminance conditions speed correction factors are applied.

Not all sources of fluctuations identified in chapter 2 are taken into account in the model. As far as the fluctuations in traffic demand are concerned, the following sources are included:

- time of day (up to a certain degree)
- day of week[113] (working days only)
- period of year[113]
- events
- other variations in human travel behavior[113] (up to a certain degree)

This means that the following sources of variations in traffic demand are not taken into account:

- time of day (remaining part)
- weekend days versus weekdays
- public holidays
- variations in weather conditions
- road works
- traffic information dissemination
- other variations in human travel behavior (remaining part)

For the fluctuations in traffic supply, the following sources are taken into account in the model:

- variations in weather conditions
- variations in luminance
- incidents
- road works

Not included are:

- traffic control actions
- variations in vehicle population
- variations in driver population
- variations in human behavior

Possible interdependencies between the various sources of variations (like the interdependency between capacity reduction due to adverse weather conditions and the increased probability on the occurrence of

[113] 'Day of week', 'period of year' and 'other variations in human travel behavior' are taken together as one source of variation.

accidents due to the same weather conditions) are not taken into account in the model.

The different sources of variations can easily be 'switched on or off' in the model (in order to find out their individual contributions to the performance indicators considered). It is just a matter of setting their probabilities/frequencies in the input files to zero, or setting the associated correction factors to 100 percent (corresponding to a zero correction). By 'switching off' all different sources of variability simultaneously, it would be possible to calculate the representative situation. However, there is no need to perform this as a separate (manual) step, since SMARA automatically includes this representative situation in its output.

Due to the fact that SMARA is based on a *static* traffic assignment model, the travel times cannot be calculated from the process of queue formation and dissolution. Instead, the travel times are derived from a predefined relationship relating the average speed on a road section to the ratio of demand and capacity on this road section. It should be noted that this relation is not fully comparable to the fundamental diagram discussed in section 2.1. After all, the relation used in static traffic assignment models allows for demands larger than the capacity, which in reality physically is not possible, and therefore is not possible according to the fundamental diagram either.

Drawbacks of such a static traffic assignment approach are:
- The traffic congestion occurs at the wrong location (i.e. *in* the bottleneck, instead of upstream of it).
- Blocking back effects are not modeled (since there are not really queues in static traffic assignment models).
- The temporal redistribution effect of traffic congestion is not modeled (since the vehicles are at all different road sections that are part of their route *at the same time*[114]).
- Overflow from the previous or to the next time interval is not taken into account (since all trips are assumed to be completed within the single time interval considered, even though travel times might be longer than this time interval).
- En-route changes in route choice (in response to changed traffic conditions) cannot be modeled.
- The effect of the capacity drop cannot be modeled explicitly.

Clearly, this way of calculating the traffic conditions is not very accurate. In SMARA the static assignment approach was chosen in view of its relatively limited calculation time (Hilbers et al, 2004). This calculation time was important because of the fact that it was aimed for to perform simulations for the whole of the Netherlands.

In SMARA the origin-destination traffic demands can be assigned to the different route options in two different ways:
- by using a fixed-path assignment, or
- by using a fully adaptive assignment.

[114] After all, only one time interval is considered.

In case of a fixed-path assignment for all simulation runs one and the same set of route choice fractions is used. These are route choice fractions calculated for the 'nominal' situation (i.e. the situation without any of the correction factors for the influences of the sources of variability). Using the same route choice fractions for all simulation runs corresponds to the assumption that road users do not change their route choice in response to 'abnormalities' in the traffic conditions. In case of minor differences in traffic conditions this might be a reasonable assumption. From section 2.3.3 it may be clear though that in case of severe disruptions this is not a realistic assumption anymore.

In case of the fully adaptive assignment for each individual simulation run the route choice fractions are calculated anew. This corresponds to the assumption that road users are fully aware of the (current and future) traffic conditions on the different route alternatives, and (based on this knowledge) optimally adapt their route choice in response to 'abnormalities' in the traffic conditions. Obviously this is not a realistic assumption either, since the largest part of the road users generally will hold to their 'standard' routes. In reality, the situation will be somewhere in between the situations modeled by a fixed-path assignment and a fully adaptive assignment. In SMARA such an 'intermediate' assignment option is not available however (Hilbers et al, 2005)[115].

As output SMARA provides both the results on the level of individual road sections and the results on the level of origin-destination relations. In the latter case, the travel times for the different route alternatives are averaged using the route choice fractions as weight. As the output is aggregated/processed to some extent already (rather than being the 'rough' simulation results), a large part of the indicators selected in chapter 3 cannot be calculated from it. Note that indicator VI (representing the travel time instability) obviously could not be calculated from the results of a *static* traffic assignment model anyhow.

In chapter 3 it was explained that it is desirable to exclude the vacation periods and periods with planned (large scale) road works from the day-to-day distributions of travel times, and to consider separate day-to-day distributions for the different days of the week. In SMARA this is not possible however, since 'day of the week', 'period of the year' (including vacations) and 'other variations in human travel behavior' are considered *together* as *one* source of demand fluctuations (rather than explicitly specifying each of these sources individually), and no distinction is made in planned (large scale) road works and emergency repair.

As far as the incorporation of the effects of the measures to be considered (aimed at alleviating traffic congestion) are concerned, SMARA offers the following possibilities:

[115] It cannot be excluded however that this has been changed in any possible newer versions of SMARA (since 2005).

- modifications to the network structure;
- changes of the nominal demand levels (separately for the peak and off-peak period) for given origin-destination relations;
- changes of the nominal capacities of given road sections;
- adjustment of the relationship relating the speed to the demand-capacity ratio for given road sections;
- adjustment of the correction factors expressing the effects of the various sources of variability;
- adjustment of the probabilities/frequencies of the various sources of variability.

In spite of these possibilities, part of the conceivable measures cannot satisfactorily be incorporated in the model however. This primarily relates to measures whose effect is highly dynamic in nature. Since SMARA is not dynamic but static in nature, such effects cannot properly be accounted for. Measures directed at influencing sources of variation that are not included in SMARA obviously cannot be incorporated in the model either.

A1.2 LMS-BT (National Model System - Reliability Tool)[116]

LMS-BT (developed by Rand Europe, by order of the Dutch national road authority) is a tool that can be used for estimating travel time reliability from the output of the LMS (National Model System[117]) or NRM (New Regional Model[118]). These are strategic traffic models, with which traffic prognoses are calculated in the Netherlands. The output of these models consists of the traffic volumes and speeds on all main links of the national road network (LMS) or regional network (NRM) on an 'average' working day, for three parts of the day: the morning peak, the evening peak and the rest of the day. LMS-BT calculates five reliability indicators from these output variables (again separately for the morning peak, evening peak and rest of the day), using predefined relationships.

These five reliability indicators are:
- the NoMo reliability indicator (i.e. the percentage of trips 'in time' – see section 3.3.3);
- an adapted version of the NoMo reliability indicator (i.e. the percentage of trips 'not too late');
- the travel time gain (based on the route flows and the differences between the 90th percentile route speed and the median route speed);
- the travel time loss (based on the route flows and the differences between the median route speed and the 10th percentile route speed);
- the costs of (un)reliability (based on the former two indicators and a user-specified monetary valuation of (un)reliability).

While the basic calculations in LMS-BT are performed at the level of individual routes, the output is aggregated at area level (i.e. at the level

[116] In Dutch: Landelijk Modelsysteem - Betrouwbaarheidstool

[117] In Dutch: Landelijk Modelsysteem

[118] In Dutch: Nieuw Regionaal Model

of the whole LMS/NRM area or one or more user-specified parts of this). In LMS-BT only the main road network is considered.

The predefined relationships that are used to calculate the reliability indicators from the LMS/NRM output have been derived from an empirical analysis. In this empirical analysis, the average route speed (averaged over the working days of a year) was found to be the most important explaining variable for the various reliability indicators considered (Rand Europe, 2004). Therefore, empirical relations between the average route speed on the one hand and the various reliability indicators on the other hand have been used as the basis for LMS-BT. For the NoMo indicator and its variant, the route length turned out to be an important variable as well. For longer routes, lower reliabilities were found. This has been included in LMS-BT by applying a correction for the length of the route considered.

Of course, the speeds calculated by the LMS or NRM (calculated for 'representative' traffic demand and supply conditions) cannot directly be used in LMS-BT as if it they were *average* speeds. Average speeds do include the effects of disturbances, while the 'representative' LMS/NRM speeds do not. Therefore, the LMS/NRM speeds can be expected to be higher than the corresponding average speeds found in real-life. However, there might be other systematic differences involved as well.

Obviously, for these differences a correction should be applied. The most obvious approach for this would be to consider a scatter plot of the LMS/NRM speeds and the corresponding[119] average speeds found in real-life, and look for a possible relationship. However, Rand Europe (2004) dealt with this problem in a rather different way. They converted both the LMS speeds and the empirical average speeds into travel times per kilometer, and subsequently plotted the distributions of these two. Then, they applied a transformation to the LMS data in order to make its distribution correspond better to the distribution of the empirical data. After this transformation there are still important differences between the two distributions though. Therefore, Rand Europe concludes that their tool may not be that accurate in calculating the *absolute* values of the reliability indicators. For *relative* studies (i.e. comparing different scenarios) on the other hand, the tool is judged to be suitable.

In LMS-BT, the frequency of some congestion causing disturbances (i.e. rainy weather, road works, accidents, and large accidents) can be varied (by specifying an index indicating the relative frequency with respect to the situation in the base year 2000[120]), in order to see their influence on the reliability indicators. These influences are calculated using empirical relations between the occurrence of the disturbances and average route speed. The impact of these disturbances can be

[119] i.e. for the same road sections / routes

[120] It should be noted that this index should be kept within the range of 0.5 to 2.0 (see Grol et al, 2004), in order to avoid too extreme extrapolations.

varied as well, by specifying indices indicating their average effect on speed (relative to the situation in the base year).

In terms of model type, LMS-BT in fact can be seen as a combination of a knowledge-based model and a black box model. The knowledge-based model (LMS or NRM) is used to calculate the representative traffic conditions in the network. Next, the black box model (consisting of the empirical relations between, on the one hand, the representative traffic conditions and the effects/frequencies of the disturbance factors, and, on the other hand, the performance indicators to be evaluated) is used to compute the values of the performance indicators. The main advantage of this method is its limited calculation time. Its quality leaves much to be desired, however.

First the quality of the knowledge-based model part (viz. the LMS or NRM) is discussed. An important limitation of the models LMS and NRM is that they, like SMARA, are not dynamic[121]. Unlike SMARA, the LMS and NRM however *do* take into account both the limited inflow to links and blocking phenomena (in the travel time calculations). Furthermore, in the latest update of the LMS and NRM some improvements have been made to model the location and length of traffic jams in a more realistic way (Hofman, 2010). However, due to the fact that the models are not really dynamic (and use a rather course spatial discretization), the modeling of the processes of queue formation/dissolution and blocking back cannot be better than a rough approximation. Obviously the phenomenon of the capacity drop is not taken into account in these kinds of models.

Now the quality of the black box model part is considered. In section 3.5.2 it was already discussed that the use of empirical relations such as included in LMS-BT in fact is not desirable, since:

- The empirical relations seem not accurate enough for evaluating measures with effects of only a few percent.
- It is not certain whether the empirical relations hold true for *all* conceivable types of measures (and in *every* conceivable situation).
- The existence of a reliable relation between the representative and average speed (or travel time) is to be doubted.
- In any case such a relation between the representative and average speed would not be preserved in situations in which certain types of measures are taken.

In order to give sound results, in fact the empirical (i.e. statistical) relations should implicitly ‘embody’ all underlying mechanisms in the traffic flow, like the processes of queue formation and dissolution, the capacity drop phenomenon, the relation between flow and speed, the blocking back phenomenon, the temporal redistribution effect of the occurrence of traffic congestion, and the effects of traffic congestion on route choice and departure time choice. Clearly this is not sufficiently the case though, if only because of the fact that many of these

[121] The LMS and NRM have always been static models. However, in their newest versions a *semi-dynamic* assignment method is used.

mechanisms are rather location-dependent [122], while the empirical relations do not take into account any location-dependent information, except for the local representative speed value (calculated with the LMS or NRM). This obviously is not sufficient however.

Once could also question the accurateness of the estimation of the empirical relations for the influence of the different disturbance factors. After all, in section 4.2 it was explained that:

- The data required for such an empirical analysis is typically incomplete.
- It is difficult to properly isolate the individual influences of the different disturbance factors from empirical datasets, particularly because there are mutual interdependencies with other variations.

A drawback of LMS-BT is that the possibilities for varying the individual influences (i.e. frequencies and effects) of the various sources of variability are very limited:

- Only for rainy weather, road works and accidents such possibilities are offered. For all other sources of variability identified in chapter 2 there are no such possibilities at all.
- The influences of these sources (i.e. rainy weather, road works and accidents) can be varied only within a limited range. They cannot be completely 'switched off', since the empirical relations may not be extrapolated that far.

As a consequence, this model in fact is not appropriate for studying the relative importance of the different sources of variability.

Another drawback of LMS-BT is that the possibilities for incorporating the effects of measures (aimed at alleviating traffic congestion) are more limited than in SMARA. This is due to the facts that:

- Most of the sources of variability cannot be changed in LMS-BT (see above), with the result that the effects of the measures directed at influencing these sources cannot be incorporated in the model.
- The effects of the sources of variability that *can* be changed in LMS-BT (i.e. rainy weather, road works and accidents) are defined in a rather implicit way (viz. in terms of their influence on the average speed), which makes it difficult to properly incorporate the effects of measures directed at influencing these.

Just like in SMARA, measures whose effect is highly dynamic in nature cannot satisfactorily be incorporated in LMS-BT.

The output of LMS-BT clearly does not meet the requirements set in section 4.4. Only five performance indicators are calculated, which do not match the performance indicators selected in section 3.4. Computing the desired performance indicators in a post-processing step is in this case not possible either, since there are no raw simulation data available for this (due to the fact that the model is largely based on empirical relations instead of traffic simulations).

[122] i.e. dependent on local factors such as the network structure

A1.3 Waiting time model for main roads[123] (WTM) – Traffic Quality / ESIM

The 'Waiting time model for main roads' is a model which has been developed in the Netherlands in order to be able to assess the performance of main roads (in terms of the occurrence of traffic congestion) in a more accurate way, to enable a more objective way of dealing with traffic congestion in traffic policy, roadway dimensioning, and timing of infrastructural projects (Toorenburg, 1988 & 1990). Actually, a first version of the model was used for the study into the 'optimal' congestion probability, conducted as part of the preparation of the national policy document SVVII (see section 3.3.3). However, after this study the calculation model has been further improved several times.

With the model, several indicators characterizing the traffic congestion can be calculated, as a function of the traffic demand and the road capacity. Instead of considering only one specific situation (like a 'design' peak hour), these indicators provide an *overall* characterization of the traffic conditions, considering all working days of the year. Both average congestion characteristics and the reliability of the traffic conditions can be evaluated with the model. The calculated indicators include the average daily sum of travel time losses, the average delay of vehicles involved in queues, the average daily duration of queuing, the percentage of vehicles experiencing traffic congestion, and the percentages experiencing more than 5, 10, 15 or 30 minutes of delay (Transpute, 1989).

In the model an individual 'isolated' bottleneck or road section is considered. All relations with other parts of the traffic network are neglected. This means that network effects (i.e. blocking back, the temporal redistribution effect, and route choice effects – see section 2.3) are not taken into account, which obviously is an important limitation of the model. The model output for a given bottleneck or road section therefore can be considered as the 'intrinsic weakness' of this bottleneck or road section (rather than a characterization of the congestion observed in reality).

The 'Waiting time model' has been implemented in different tools (viz. 'Traffic Quality' and 'ESIM'; see Toorenburg, 1997). In these tools, the 'Waiting time model' is used in different ways/forms. In 'Traffic Quality' it is used only in an *indirect* way. The output then is derived from a precompiled set of simulation results, using only a few input variables characterizing the situation at hand: the ratio of the annual average working day traffic demand and the (nominal) capacity of the road section under consideration, the percentage freight traffic, and the so-called 't-values' (two parameters describing the daily traffic demand

[123] in Dutch: 'Wachttijdmodel hoofdwegen'

profile[124]). This way quickly a first approximation of the traffic quality on a given road section can be obtained. It is a convenient approach for studies for which no very specific data are available, or for studies in which many different road sections are to be analyzed within a short period of time. Actually, the tool 'Traffic Quality' was used for performing a yearly assessment of the quality of the traffic operations on all sections of the Dutch main road network.

Clearly, such an approach is not appropriate for the task at hand however, since we want to make very specific changes to the traffic system, which result in situations for which no results can be found in such a precompiled set of simulation results. (Changes like 'switching' different sources of variability in traffic demand and supply 'off', and implementing measures whose effects cannot satisfactorily be expressed in terms of the input variables mentioned above)

In ESIM – meant for assessing the quality of a road design or an existing road section in the current or in a prognosis-situation – the 'Waiting time model' is used in a *direct* way: with this model dynamic traffic simulations are performed for the specific situation at hand[125]. This way, more accurate and more situation-specific results are obtained. In this case, more specific input data are used, such as the location-specific (average) working day traffic demand profile. The computations are based on a simple vertical queuing model. This means that the physical dimension of the queue is not considered[126]. All excess traffic demand (i.e. the demand exceeding the bottleneck capacity) is supposed to be 'stored' at the bottleneck location. For the delays incurred due to the capacity exceedance, this does not make any difference at all, as long as no network effects do occur. As mentioned above, network effects are neglected however in this model.

For each time step, the model compares the traffic demand at the bottleneck location and the capacity, and – if applicable – updates the number of vehicles in the queue with the difference between these two. The capacity drop (i.e. the difference between the free flow capacity and the queue discharge capacity; see section 2.1) is neglected in the model. From the computed course over time of the number of vehicles in the queue, and the inflow and outflow of the queue, rather

[124] These two parameters indicate how the daily traffic demand profile at the road section under consideration can be approximated by a combination of three predefined basic components of such daily traffic demand profiles. They could be termed 'shape parameters' of the demand pattern, since they do not give any information on the absolute size of the traffic demand. Obviously, the shape of the daily demand pattern has a strong influence on the level of traffic congestion on the road in question.

[125] However, ESIM also includes the option to fall back on a precompiled set of simulation results, for purpose of reference or for situations in which no detailed input data are available or only a rough approximation of the output is desired.

[126] It should be noted that a separate 'traffic jam module' is added to ESIM, which allows the length of the queue and the traffic density and speed within the queue to be calculated in a kind of post-processing step.

straightforwardly a large number of performance indicators can be computed, such as the average delay.

Obviously, also in free flowing traffic delays are incurred, due to the difference between the actual speed and the free speed (i.e. the average speed for a (almost) zero traffic volume; see section 2.1). In the first versions of the 'Waiting time model', these delays were neglected (i.e. only delays due to capacity exceedance were considered). In ESIM however these delays *are* taken into account, by using a separate model component for computing the free flow traffic conditions, added to the 'Waiting time model'. For this, the model user needs to specify the speed limit, the relationship between the actual speed and the volume-capacity ratio, and the length of the road section under consideration. Due to this additional model component, ESIM can also calculate indicators like the average speed or travel time on the road section (also separately for free flow traffic conditions and forced traffic conditions), and the statistical distribution of the speed or travel time.

In the simulations a time step of 1 hour is used (AGV, 1997). Obviously, this would be way too large for a sufficiently accurate (and stable) modeling of the propagation of the traffic flow over a network. In this model, this does not pose a problem though, since only an individual bottleneck location is considered, which removes the necessity to calculate propagation of the traffic. However, of course the time interval still should be short enough for a sufficiently accurate computation of the processes of queue formation and dissolution, and the delays resulting from these processes. In fact, a time interval of 1 hour seems not short enough for this.

The model repeats the simulations for a user-specified number of runs (with one run corresponding to one working day, and using different random seeds for different runs/days). Clearly, this number should be large enough to make sure that the variabilities are statistically sufficiently reflected in the model output. Nonetheless, due to its simplicity (and the fairly large time interval), the calculation time of the model is rather limited.

The model does not (or not explicitly) take into account many of the sources of variability identified in chapter 2. The following sources of variability in the capacity are taken into account:

- varying weather conditions (rainy weather only)
- intrinsic variability (a combination of the effects of variations in driver and vehicle population, and (other) variations in driving behavior)

This means that the other sources of capacity variations identified in chapter 2 (i.e. road works, incidents, variations in luminance, and traffic control actions) are not taken into account.

Based on empirical research, the effects of the varying weather conditions (i.e. rain) are simulated by reducing the available capacity by

12% [127], with a frequency of occurrence of 0.07 (AGV, 1997). Dependencies between the probabilities of occurrence of rain in subsequent time intervals likely are not taken into account[128]. The effect of the 'intrinsic' variability is modeled by drawing the capacity values from a normal distribution with a coefficient of variation of 7%[129]. It seems that variations in the relationship between volume-capacity ratio and speed are not taken into account.

Besides the (within-day) variability described by the average daily traffic demand profile, the model does also take into account other variations in the traffic demand, by adding a certain random element to the demand pattern. However, no information was found regarding the way in which this is actually done. In (AGV, 1997) it is stated that ESIM applies a traffic engineering reasoned random process on the user-specified demand pattern (using the Monte Carlo technique). It seems that no distinction is made in the various sources of the variability. Regularities in the variations (related to the influences of the day of the week and the season) are not taken into account. Interdependencies with the variations in capacity are not considered either.

Due to the fact that the different sources of variability in the traffic demand are not explicitly considered, they cannot separately be switched off (in order to assess their relative contributions to the congestion indicators). In combination with the fact that a number of sources of variability in the traffic supply conditions are not taken into account, this in fact makes the model rather inappropriate for assessing the relative contributions of the various sources of variability to the congestion indicators.

The 'representative' situation can easily be computed with the simulation model. ESIM seems to offer a special possibility to perform a simulation without randomness for this (see AGV, 1997). Alternatively, one could also 'manually' switch off all sources of variability, by changing the relevant model parameters (like putting the probability or effect of rainy weather on zero).

The possibilities that are offered by the model to evaluate measures that can be considered for alleviating traffic congestion are rather limited, due to the facts that:

- many sources of variations are not – or not *explicitly* – included in the model, resulting in the inability to evaluate measures having effects on these;

[127] In the Netherlands it nowadays is common to use a road surface of ZOAB (i.e. very porous asphalt concrete) on motorways. ESIM can take this into account by using a smaller capacity reduction for rainy weather.

[128] These dependencies at least were not taken into account in an earlier version of the 'Waiting time model' (see Toorenburg, 1988). It cannot be claimed with certainty that they are not taken into account in the version used in ESIM either.

[129] Assuming that in the version of the 'Waiting time model' implemented in ESIM this percentage has not been changed as compared with the earlier version described in (Toorenburg, 1988).

- the time interval of the model (1 hour) is relatively large, which makes it doubtful whether measures with effects that are highly dynamic in nature can be modeled in a sufficiently accurate way;
- in the model only an isolated bottleneck is considered, with the consequence that measures with effects with a clear 'network component' certainly cannot be incorporated in a satisfactory way.

The standard output variables do not match the desired output indicators specified in section 3.4. However, using the detailed raw simulation results of the model, one can evaluate self-defined indicators as well (i.e. indicators that correspond more closely to the desired ones). An insurmountable difference however would remain then in the fact that the desired indicators are defined at origin-destination or network level, while the model can only evaluate indicators at the level of an individual road section / bottleneck.

A1.4 Queuing Model for determination variability recurrent congestion (TU Delft)

The queuing model of the TU Delft is an analogy to the previously discussed 'Waiting time model', developed at the Delft University of Technology because of the fact that in generally accessible literature few details were available on the latter model, while its approach was considered interesting (Botma, 1999).

Just like the 'Waiting time model', the model consists of a vertical queuing model. By comparing the demand and capacity pattern for a bottleneck location, the processes of queue formation and dissolution are computed. A difference with the 'Waiting time model' is that different sources of systematic variations in the traffic demands are explicitly considered. These are the systematic variations with the time of the day, with the day of the week, and with the month of the year. The latter two are accounted for by multiplying the daily demand pattern by a day-of-week factor and a month factor[130]. Finally, a random fluctuation is added to the demand pattern. In order to avoid the demand pattern to fluctuate to widely, some autocorrelation was added to this random fluctuation. (This also seems reasonable, since it is plausible that if the demand during a certain time interval is relatively high, it is likely also relatively high during the next interval.)

Another difference between both models is that the queuing model of the TU Delft uses a time step of 15 minutes, whereas the 'Waiting time model' uses a time step of one hour. As a consequence, the processes of queue formation and dissolution are computed with greater accuracy in the queuing model of the TU Delft.

[130] Note that this is a rather simplified approach, since in reality not only the *height* of the daily demand pattern may be different for the different days of the week and the different months of the year, but its *shape* as well.

Apart from the differences described above, the two models are more or less the same, and consequently have the same advantages and drawbacks. For more information on these, the reader is referred to the section on the 'Waiting time model' (A1.3).

A1.5 Traffic Quality – Network version

In section A1.3, the 'Waiting time model' and the two tools in which this model has been implemented (i.e. the models 'Traffic Quality' and 'ESIM') were discussed. It was described that the 'Waiting time model' only considers one individual bottleneck, in a way as if this bottleneck was isolated from the rest of the network. As was explained, this results in the following two drawbacks of the 'Waiting time model':

- network effects are neglected, and
- calculation of travel times / speeds on the level of *routes over the network* is not possible.

Obviously, these drawbacks are equally applicable to ´Traffic Quality´ and ESIM.

To overcome these drawbacks (insofar the model 'Traffic Quality' was concerned), a *network version* of 'Traffic Quality' was developed. It was recognized that a network traffic simulation technique in fact is the only approach with which the mutual interactions between the bottlenecks could be accounted for in a consistent way. The use of a pre-compiled set of simulation results – as in the original model 'Traffic Quality' – was abandoned in the network version.

In 'Traffic Quality – Network version' use is made of the dynamic macroscopic traffic simulation model 'Flowsimulator', developed by the Dutch consultancy firm Transpute. This model can simulate the propagation of the traffic flow over a whole network. This way, mutual interactions between bottlenecks (related to the phenomena of blocking back, filtering and releasing – see section 2.3) are taken into account. Unlike many network traffic simulation models, Flowsimulator does not contain an origin-destination matrix and traffic assignment module for the distribution of the traffic over the various possible routes. Instead, it uses user-specified split fractions at the nodes of the network for this. These split fractions indicate how the outflow of a link is distributed over the different following links.

In order to obtain a network version of the model 'Traffic Quality', Flowsimulator has been adapted. With the purpose of making optimal use of the traffic demand data available within the original version of 'Traffic Quality' (i.e. the demand *per individual section* of the main road network), it was made possible to specify the traffic demands on the level of every single link of the network (Toorenburg, 2002). This is a very unusual characteristic. In conventional traffic flow models, the traffic demand is specified only on the *entry* links of the network (after which it is distributed over the different possible routes by means of a traffic assignment routine, or by using split fractions at the network nodes).

As a consequence of this way of specifying the traffic demand, the model is over-specified. In the simulations, this over-specification is dealt with by relaxing the principle of conservation of vehicles at the network nodes. This means that in the traffic simulations the total inflow and outflow of a node are not required to be equal (i.e., some inconsistency is allowed). In the simulation of the traffic propagation over the network, these inconsistencies are dealt with by multiplying the amounts of traffic 'passed on' at the nodes by a correction factor. Note that in this adapted version of Flowsimulator, the split fractions at the nodes do not have to be specified by the model user anymore, since they follow directly from the user-specified traffic demands at the links of the network.

As long as the discrepancies are not too large, the relaxation of the principle of conservation of vehicles is not very detrimental to the simulation results. The main advantage of the over-specification is that one can easily obtain a simulation model that is properly calibrated (to the measured traffic demands [131]) *over the whole network*. In a conventional traffic flow model, it is a difficult and time-consuming process to establish the traffic demand pattern (usually in the form of an origin-destination matrix) in such a way that the resulting traffic demands on the individual links do reasonably match the values observed in reality.

It should be noted, however, that for the research questions under consideration in this project, this characteristic of the model 'Traffic Quality – Network version' is not an advantage. After all, for addressing these research questions it is not necessary to 'match' an actually existing situation. (In fact, considering a complete fictitious situation would do as well, as long as it 'could have been' a real-life situation.) Consequently, there is no issue of calibration to observed traffic volumes anyhow.

In fact, for the task at hand, an approach in which the traffic demands are defined on *origin-destination or route level* is preferable to an approach in which the traffic demands are specified on *link level* (and propagated using split fractions), as used in 'Traffic Quality – Network version' / Flowsimulator. This is because of the following reasons:

1) Due to the definition of traffic demands on link-level, the model cannot properly deal with the route choice effects of traffic congestion. In its basic form these effects are not taken into account at all. A possible way to change this is by making split fractions dependent on the traffic conditions on certain links, but in general this certainly cannot offer a really

[131] It should be noted that the link traffic *demand* patterns to be specified by the model user in fact are not directly the traffic *volume* patterns as measured on the links concerned. After all, these latter are affected by the occurrence of traffic congestion, while the *demand* patterns required refer to the *unaffected* traffic quantities. This can be accounted for by fitting a combination of standard demand pattern components to the measured *volumes*, and using this fitted combination as link demand pattern (see Toorenburg, 2002).

satisfactory solution. For (very) simple networks it might be a solution however.

2) In reality, the temporal redistribution effect of traffic congestion (i.e. the filtering and subsequent release of traffic) does not only dynamically affect the traffic volumes on the links, but the split fractions in the network nodes as well (even apart from the route choice effects discussed above). This is not accounted for in models in which the traffic demands are inputted on link level. In models in which the traffic demands are specified on origin-destination or route level this problem does not exist, since no predefined nodal split fractions are used in these types of models.
3) Due to the definition of traffic demands on link-level, demand effects of traffic measures cannot easily be accounted for in a consistent way. As stated in (Transpute, 2002), due to this property the model in fact can only be applied on the 'current' situation (i.e. the situation for which the user-specified link traffic demands have been observed) or small variations thereof. It cannot be used for situations in which the traffic demands are significantly affected, unless some other model is used to compute the 'new' link travel demands.

The core of the model (i.e. the part that models the traffic propagation over the network) consists of a cell transmission model. This means that all roads are divided in a number of cells. For each time interval, the flux on all cell boundaries is determined on the basis of the conditions in both the downstream and the upstream cell. Each time interval, the number of vehicles in each cell is updated, using the fluxes on the cell boundaries (i.e. the in- and outflow of the cell under consideration). From the number of vehicles in each cell and the cell length, the traffic density within the cell can be calculated. (Note that within each cell and time-interval, the traffic conditions are assumed homogeneous and stationary.) From this density, the traffic speed within the cell is determined, using the fundamental diagram (see section 2.1).

Although the model (in its basic form described in Transpute, 2003) does not generate two-capacity regimes (in order to 'avoid unexpected undesirable behavior of the model'), it does take into account the capacity drop to some extent. That is, the maximum outflow of a congested cell is assumed to decrease linearly with the density, from the capacity (at the critical density) to the queue discharge rate from standstill (at the jam density). Here the queue discharge rate from standstill is a fixed percentage of the capacity.

In the model automatically a spatial discretization (i.e. cell length) of 200 m is used. This seems short enough for a computation of the spatial dynamics of the queues which is sufficiently accurate for the task at hand.

Since for the determination of the cell fluxes (i.e. the flows from one cell to another) only the traffic states in the two immediately adjoining cells are considered in the computational scheme, it is required that the 'information' in the traffic flow might not travel a larger distance within one time interval than one cell length (i.e. 200 m). In the upstream direction information travels with a speed of maximal about 20 km/h (the maximum speed with which a shock wave travels upstream). In the downstream direction information travels with a speed of maximal 120 km/h (the highest speed limit on Dutch motorways, thus neglecting over-speeding). Since the latter is higher, this one is decisive for the maximum time interval. By dividing the cell length (200 m) by this speed (120 km/h), one can compute that the maximum time step is 6 seconds then. Therefore, Flowsimulator uses a time interval of this duration (Transpute, 2003).

In spite of the fact that the traffic conditions on a whole network are calculated, and that a rather small time step is used, the calculation time of the model is not very large. (In about one minute, the traffic conditions on the main road network of all motorways in the Netherlands can be calculated for a one-day period.)

In the model a possibility is created to include variations in the circumstances governing the traffic conditions (other than the average within-day variations in the traffic demand, which of course are already included in the model by specifying the average working day traffic demand patterns). This possibility is created by the inclusion of a 'capacity table' in the model. This table consists of 'capacity factors' for every single road section, per quarter of an hour of the day. In the simulation process, the nominal link capacities are multiplied by these factors, by which these capacities might be temporarily raised or lowered. This way, within-day variations can be modeled. By generating different tables (reflecting different days), and repeating the simulation process for these different tables, day-to-day variations can be taken into account as well. Note that the different tables should be generated partially in a systematic way, and partially in a random way. After all, some of the sources of variation behave in a systematic way, while others are more random in nature (see chapter 2).

Note that the model contains only a *capacity* table. A counterpart for the traffic *demands* is not included. This means that the variations in the traffic demands have to be incorporated artificially by applying variations in the capacities 'in the opposite direction' (i.e. compensating for the neglected variations in the demands). It should be noted however that in the model fluctuations in the traffic demands in fact cannot be addressed in an ideal manner anyhow, since all model inputs (not only capacities, but demands as well) are specified on link level, while the variations in demands arise at the route (or origin-destination) level.

In (Transpute, 2003), a possible scheme for generating the capacity tables (in order to account for the variations in demand and capacity) is described, which was applied in a project in which the model was used. In this scheme, many of the sources of variability (and their mutual interactions) are not included however (like incidents and road works), or only in a strongly simplified way. However, the model user can easily generate such capacity tables himself (for example using a spreadsheet program with a random number generator), in which all different sources of variability and their mutual interdependencies are included in a proper way. It is very easy to 'switch' the various sources of variability 'off' in the model (in order to find out their individual contributions to the performance indicators, or in order to perform a 'representative' calculation of the traffic conditions). A simple change of the capacity tables (i.e. the removal of some or all of the influences factors) is sufficient.

Of course, the variations in traffic supply conditions are not limited to variations in the capacities. Other supply characteristics, such as the free speeds and the jam densities might be affected by the sources of variability as well[132]. This however is not taken into account by the model, which obviously is a limitation of it.

The model offers a wide range of possibilities to incorporate the effects of measures (aimed at alleviating traffic congestion). These possibilities include: adaptation of the network structure, adjustment of link or node characteristics (like the number of lanes, or spill back ratios), adaptation of the fundamental diagrams, and modification of the capacity table. Since the capacity table allows one to specify different values for different time periods (per quarter of an hour), some dynamic measures can be incorporated as well. Probably, some other inputs to the model cannot be specified in a time dependent way however, so that not all types of dynamic measures can be evaluated. Also note that part of the dynamic measures is traffic responsive in nature (meaning that these measures react on the actual traffic conditions). It seems that the model does not contain readily available possibilities to take this into account. Another limitation of the model is that effects on the traffic demands cannot be accounted for in an ideal way, as was discussed already in an earlier part of this section.

The output of Flowsimulator consists of the rough simulation data, i.e. the traffic conditions (speed and traffic volume) in all cells of the network, for every n^{th} time step. Using this output, all desired indicators (defined in section 3.4) can be computed in a post-processing step.

A1.6 KAPASIM

KAPASIM is a model developed at the Ruhr University in Bochum (Germany). It was developed to enable the assessment of traffic flow quality over a whole year (instead of during one single 'representative' peak hour), in the context of the design of road facilities. It is based on a comparison of annual patterns of traffic demand and freeway capacity. These patterns are simulated using the Monte Carlo technique. As output, the sum of delays and the total duration of congested flow conditions over a whole year are provided. Several other parameters relating to the reliability of the freeway operation are

[132] As far as the jam density is concerned, the most obvious cause of variability is the variability in the number of available lanes.

estimated as well, like the risk of being significantly delayed by congestion, or the number of traffic break downs (Brilon et al, 2007).

In the model, traffic conditions are computed with a simple vertical queuing model. At the end of each time interval the number of vehicles in the queue is updated with the difference between the traffic demand and capacity generated for this time interval. The delays due to traffic congestion (i.e. incurred in queues) are computed by multiplying the average of the numbers of queued vehicles at the beginning and end of the time-interval by the interval duration (Brilon et al., 2007). Obviously also in free flowing traffic delays are incurred. After all, on average the speed decreases with increasing traffic volume (within the free flow domain) – see also the fundamental diagram discussed in section 2.1. Brilon et al (2007) note that this can be addressed by using a combined traffic flow model based on standardized speed-flow curves, which are varied in accordance with the random capacity variation. It is not clear however whether this is actually implemented in KAPASIM.

In the model an individual freeway section is considered. This means that network effects (i.e. blocking back, the temporal redistribution effect, and route choice effects – see section 2.3) are not taken into account. The capacity drop (i.e. the difference between the free flow capacity and the queue discharge capacity – see section 2.1) is accounted for in the model.

In the model a time interval of 5 minutes is used. This interval seems short enough for a sufficiently accurate computation of the formation and dissolution of queues and the resulting delays (i.e. without the (within-day) variations in the traffic demand and supply and the resulting traffic conditions being smoothed out too much). It should be noted that a time interval of 5 minutes typically would not be short enough for the propagation of the traffic over the network to be modeled in a sufficiently accurate (and stable) way. Since in this model only an individual freeway section is considered, for this model a time interval of 5 minutes is sufficiently short however.

Not all sources of fluctuations identified in section 2.2 are taken into account in the model. As far as the fluctuations in traffic demand are concerned, the following sources are included:

- time of day
- day of week (explicitly or implicitly)
- period of year (explicitly or implicitly)
- 'other' variations in human travel and driving behavior[133] (only to a limited extent)

The time of day is taken into account by multiplying the daily traffic volumes by typical demand patterns for different weekdays. These typical demand patterns describe the shares of hourly demand values in the total daily traffic volume. The daily traffic volumes (reflecting the

[133] Here 'other' is defined with respect to *all* sources of demand variations identified in section 2.2.4 (i.e. not with respect to the time of day, day of week, and period of year only).

day-of-week and period-of-year fluctuations) are derived from loop detector data for an existing freeway, or – if no traffic data are available – by using typical demand patterns over a week and a year. For the German freeway network such patterns are available. In the first case the influence of the day of the week and the period of the year are thus taken into account in an implicit way, while in the second case they are considered explicitly. It should be noted that it in fact is not very desirable to use measured traffic volume data, since these data are likely to be affected by the occurrence of traffic congestion (at the road section under consideration, or elsewhere in the network). Consequently, these data do not correspond to the factual traffic demands. Furthermore they are valid only for a specific year, which may not be representative for an arbitrary year.

It seems that other sources of variation in traffic volumes (i.e. events, public holidays, variations in weather conditions, road works, traffic information dissemination, and other variations in human travel behavior) are not really taken into account in the model. Of course one could argue that if the daily traffic volumes are derived from loop detector data, these other sources of variation are implicitly reflected in these daily traffic volumes. However, this clearly is not a proper way to take these variations into account, since:

- The variations in demand are partially connected to the (separately simulated) variations in supply[134], which is neglected this way.
- The effects of some of these sources of variation (especially the effects of weather conditions, events and traffic information dissemination) are (often) concentrated on a limited part of the day, rather than increasing or decreasing the traffic demand over the whole day to the same extent. This is not taken into account in the model.

Apart from the systematic within-day demand patterns, the only within-day demand variation taken into account is the short-term 'white noise' in the traffic demand (which can be attributed to variations in human travel and driving behavior). This 'white noise' is included in the model by applying a normal-distributed factor with an expected value of 1, and a variance of 0.1 (Brilon et al, 2007).

For the fluctuations in traffic supply, the following sources are taken into account in the model:

- incidents
- variations in weather conditions (rainfall events only)
- variations in vehicle population (implicitly)
- variations in driver population (implicitly)
- variations in human behavior (implicitly)

[134] Examples of these interconnections are:

- The effects of weather conditions on the traffic demands coincide with the effects of weather conditions on the traffic supply.
- The effects of road works on the traffic demands coincide with the effects of road works on the traffic supply.
- The (effects of) traffic information dissemination is/are dependent on the traffic conditions, which on their turn are dependent on (among other things) the traffic supply conditions.

The capacity of the road section under consideration is drawn from a capacity distribution function. Based on empirical research, a Weibull distribution function with shape parameter $\alpha = 13$ is used for this. For the scale parameter β – which is derived from empirical research as well – no general value can be given. This value is dependent on the characteristics of the road section under consideration, such as number of lanes and gradient. In an implicit way, the effects of variations in vehicle population, driver population and human behavior are included in this capacity distribution function.

The influence of incidents is taken into account by randomly generating incidents, based on typical accident and car breakdown rates, and applying a capacity reduction estimated by using the reduction percentages from the Highway Capacity Manual (see section 2.2.5). This capacity reduction is applied by varying the scale parameter of the capacity distribution. As far as the influence of variations in weather conditions is concerned, only rainfall events are taken into account. It is argued that extreme weather conditions such as heavy snowfall and ice are rare in most parts of the world, and consequentially should not be addressed in highway dimensioning. The influence of variations in luminance is not considered either, since Brilon et al. (2005) found that darkness does not shift the capacity distributions. The influence of rainfall is taken into account by randomly generating rainfall events, based on monthly probabilities of rainfall. Again, the effect on capacity is taken into account by adapting the scale parameter of the capacity distribution function. It is not clear to which extent the effect on the speeds (in particular the free flow speeds) is accounted for.

It is not clear from (Brilon et al, 2007) whether the influence of road works on the traffic supply conditions is included in KAPASIM. It would not be very difficult to add this to the model however. Road works can be generated in a similar way as incidents and rainfall events. The effect on capacity can be accounted for in a way similar as well (i.e. by adapting the capacity distribution). The effect of the speed limit reductions in work zones would have to be incorporated by using adapted speed-flow curves for simulated roadwork situations.

The influence of traffic control actions can be included in the model by using adapted capacity distributions (i.e. with different scale and/or shape parameters). In (Brilon et al, 2005) an example is given of the effect of dynamic speed limit control on the capacity distribution (see also section 2.2.5). For part of the traffic control measures it is necessary to adapt the speed-flow curve (used for the situations in which the traffic is freely flowing) as well.

The model does not take into account all mutual interdependencies between the various sources of fluctuations. The influence of weather conditions on the rate of occurrence of accidents (resulting in an interdependency between the capacity effect of incidents and the capacity effect of weather conditions) for example is not taken into account. The dependency of incidents on the traffic volume probably *is* taken into account, since the random generation of incidents likely is based on the incident rates *per vehicle-kilometer*. Also the relation between the influence of the time of the day and the day of the week is taken into account, by applying separate daily patterns for the different days of the week.

In order to be able to assess the individual contributions of the various primary sources of traffic congestion, one of the requirements set in section 4.4 was that the model should include the possibility to *separately* 'switch off' each individual source of variability. In this model, all included sources of variability in the traffic supply can be 'switched off' by putting their frequencies/probabilities to zero or putting the shape parameter of the capacity distribution (α) to an 'infinitely large' (or very large) value (resulting in an (almost) zero variance). The variations due to fluctuations in vehicle population, driver population and driving behavior cannot be separated from one another, due to the fact that their effects are combined into one capacity distribution.

The systematic sources of fluctuations in traffic demands (i.e. time of day, day of week, and period of year) can be 'switched off' insofar they are included in the model by specifying their (individual) typical variation patterns (namely by replacing these patterns by one constant average level). Insofar as they are included in the form of measured daily traffic demands (obtained from loop detectors), they cannot easily be 'switched off' on an individual basis. The 'white noise' in the traffic demand can be 'switched off' by simply putting its variance to zero.

For the computation of the 'representative' situation, it should be possible to 'switch off' all sources of variability together, except for the systematic within-day variability in traffic demand (see section 4.4). This is easily possible by putting the frequencies/probabilities of all disturbances to zero and putting the variances of all systematic influence factors to zero (and giving them a 'representative' value).

The model offers a lot of possibilities for the incorporation of the effects of the measures to be considered (aimed at alleviating traffic congestion):

- modification of the (systematic) demand patterns,
- adjustment of the variance of the 'white noise' in the traffic demand,
- adjustment of the frequencies of the disturbing influences,
- modification of the parameters of the capacity distributions.

Due to fact that the model is dynamic, with a reasonably small time step, dynamic measures can be included as well. However, all measures with effects with a significant 'network-component' (like traffic buffers and route information provision) cannot satisfactorily be included, due to the fact that the model considers one individual freeway section only. This is a major drawback of the model. Obviously, measures with significant effects on the sources of variability that are not (explicitly) included in the model, cannot satisfactorily be included either.

The calculation time of the model is rather limited, due to the fact that only one freeway section is considered, using a simple (i.e. vertical) queuing model. Besides the fact that network effects are neglected with this approach (as was already mentioned earlier), another major drawback of this approach however is that the desired performance indicators (as specified in section 3.4) cannot be computed with the model. After all, these are performance indicators at origin-destination or network level, while the model can only provide indicators on the level of individual freeway sections. Actually, another difficulty is in the fact that the model does not directly provide travel times (or speeds) as output. Instead, more aggregated indicators are provided, like the sum of all delays.

A1.7 Travel time variability model of Mehran & Nakamura

In 2009, Mehran and Nakamura presented a methodology to estimate travel time reliability (and the impacts of congestion relief schemes thereon) based on modeling travel time variations as a function of demand and capacity, taking into account weather conditions and traffic accidents. In fact, at the same time the model is also meant for ex-ante assessing the traffic safety effects of congestion relief schemes (expressed in the number of accidents).

In the model, an expressway segment is considered. The patterns of demand and capacity over a year are simulated using the Monte Carlo technique. For each time interval, it is determined whether or not a queue starts to build up, by comparing demand and capacity. From the moment that a queue is created, the number of queuing vehicles is estimated for each time interval by means of shockwave analysis (i.e. considering the physical length of the queue). Next, travel times are estimated using speed-flow relationships (not only for congested time intervals, but for uncongested intervals as well). For congested intervals, for this a distinction is made between the operating speeds upstream of the queue, within the queue, and downstream of the queue.

It should be noted that in the model the location of the head of the queue is assumed fixed over time. In reality, this is not necessarily the case however. Furthermore, it is assumed that all queues have their head at one and the same location (to be determined by means of a calibration procedure), also in situations in which the road segment under consideration has no well-defined bottleneck. Clearly, this is not realistic either.

The approach based on shockwave analysis makes the model different from the other models for single road sections / bottlenecks that have been discussed in this appendix. In these models use is made of a vertical queuing model (in which the traffic conditions are directly derived from the demand and capacity patterns). In fact, the approach based on shockwave analysis is a rather roundabout method. After all, the delays can be equally well computed using a (much simpler) vertical queuing model, since it is not relevant how these delays are incurred (i.e. *upstream* of the queue or *within* the queue; with a *still reasonably high speed* within a *long* queue or with a *very low speed* within a *short* queue). Since network effects are not considered in the model (i.e. only an individual road segment is considered), enabling the consideration of blocking back effects cannot be the reason for considering the spatial properties of the queue either. The only possible reason for considering the physical queue characteristics in this case is that the model developers wanted to incorporate the dependency of the accident rate on the traffic conditions in the model. One might wonder however if this dependency could not have been accounted for in an easier way.

In the model, the capacity drop is accounted for. Under congested traffic conditions the capacity is taken to be 10% lower (on average) than under free flow traffic conditions.

Just like in KAPASIM, in the model of Mehran and Nakamura a time step of 5 minutes is used. As was discussed already for KAPASIM, such a time interval seems short enough for a model in which only one individual freeway section is considered. It results in a sufficiently accurate computation of the formation and dissolution of queues and the resulting delays. The calculation time of the model can be expected to be rather limited, in spite of the fact that the model is based on shockwave analysis (i.e. considering the physical dimension of the queue) rather than on a simple vertical queuing principle. This limited calculation time again can be attributed to the fact that only one freeway section is considered.

In the model, not all sources of variations are taken into account. As far as the variations in the traffic supply conditions are concerned, variations in the relationship between speed and flow-capacity ratio seem not to be taken into account at all. For the capacity, the following sources of variability are taken into account by the model:

- varying weather conditions (rainfall only)
- accidents
- variations in vehicle population (implicitly/explicitly)
- variations in driver population (implicitly)
- variations in driving behavior (implicitly)

The influences of luminance, road works, traffic control actions, and incidents other than accidents (like vehicle breakdowns or cargo spills) are neglected.

Unlike earlier discussed models taking into account the influence of weather conditions, the model of Mehran and Nakamura does not randomly generate bad weather events. Instead, historical meteorological data are used. Accidents are generated randomly. In contrast to most other models, the model does take into account the dependency of the occurrence of accidents on the traffic conditions (see chapter 2). For this relationships are used that link the accident rate to the traffic density. The influence of weather conditions on the accident rate is not considered, which is a limitation of the model.

Traffic accidents are modeled only *after* performing an initial calculation of the traffic conditions for all time intervals (without any traffic accidents). After the generation of accidents, the traffic conditions are computed anew. It is not clear why accidents are not immediately taken into account. This would save a considerable amount of calculation effort. Furthermore, secondary accidents are neglected now.

The three sources of capacity variation that were mentioned last in the list above are modeled in a combined way, by modeling the capacity as a random variable with a Weibull distribution. The effect of variability in vehicle population however to a certain extent is taken into account explicitly as well, by accounting for the variation in the fraction of heavy vehicles (using a capacity-conversion factor, as explained in chapter 2).

As far as the traffic demands are concerned, the following sources of variability are taken into account:

- hour of the day
- day of the week (including holidays (considered as Sundays), consecutive holidays and special days)
- month of the year
- varying weather conditions (amount of daily rainfall only)
- (other) variations in human travel behavior

As far as the mutual differences between the different days of the week are concerned, not only the differences in the total daily traffic volume are taken into account, but the differences in the pattern over the day as well. Note that of all models considered, this is the only model taking into account the demand effects of varying weather conditions. Also note that the interdependency between the capacity effects of weather conditions and the demand effects of weather conditions automatically is taken into account, since their occurrence is modeled on the basis of one and the same weather data set.

The last mentioned source of variability (i.e. 'other variations in human travel behavior') is taken into account by applying two randomly generated (normal distributed) correction factors: one on the level of the 5-minute intervals (to account for the short-term variations) and one on the level of the day (to account for the fact that if the traffic demand in a certain time interval is larger than average, it is more likely to be larger than average in the next time interval as well).

The demand effects of events, road works, and traffic information dissemination are not explicitly taken into account in the model. However, to a certain extent they may implicitly be included in the randomly generated correction factors mentioned above, since these are derived from empirical data.

A calculation of the 'representative' traffic conditions can easily be performed with the model, by adapting a number of model parameters (like putting the capacity adjustment factors associated with rainy weather to one, giving the scale parameter of the Weibull capacity distribution a very large value, putting the monthly demand correction factors to their average value, and setting the standard deviations of the random demand factors on zero). For the pattern of the traffic demand over the day a certain 'representative' profile should be chosen.

Related to the above, the relative influences of the various individual sources of variability can be assessed with the model rather easily as well. By adapting one (or a limited number) of the different model parameters, individual sources of variability can be 'switched off'. However, the influences of variations in the driver population and variations in the driving behavior (and partially also the influence of variations in vehicle population) obviously cannot be separated from one another, because of the fact that these sources of variability are combined in the model.

The model offers many possibilities for the incorporation of the effects of the measures to be considered (aimed at alleviating traffic congestion), like the possibilities to:
- adjust the various included sources of variability (in terms of effect, probability of occurrence, or pattern),
- adapt the general demand or capacity level,
- adjust the speed-volume relationship, and
- change the magnitude of the capacity drop.

Due to fact that the model calculates the traffic conditions dynamically (with a reasonably small time step), dynamic measures can be considered as well. However, as was already noted for all other models in which only one individual road section is considered, all measures with effects with a significant 'network-component' cannot satisfactorily be included, which is a major drawback of such models. Further, measures with significant effects on the sources of variability that are not included in the model (like road works, events and incidents other than accidents) obviously cannot satisfactorily be incorporated either.

As main output, the model provides the buffer time index [135] (as a function of the time of the day, separately for different categories of days), estimated from the distributions of the simulated travel times. However, the 'intermediate simulation results' (i.e. the traffic demand, the queue discharge flow, the number of queuing vehicles, the speeds, and the travel time, all of them for every single 5-minute interval) can be used to compute other indicators as well. As with the other models in which only one individual road section is considered, an important drawback however remains in the fact that the model can only evaluate indicators on the level of such an individual road section, while the *desired* indicators are defined at origin-destination or network level.

[135] See chapter 3 for the definition of the buffer time index.

Appendix 2 - Random capacity variation along the FD

This appendix describes an alternative approach for implementing the random noise of the free flow capacity in the model. According to this approach, this random noise is interpreted as a variation *along* the free flow branch of the fundamental diagram, instead of as a variation of the fundamental diagram itself. This alternative modeling approach is shown in Figure A2.1.

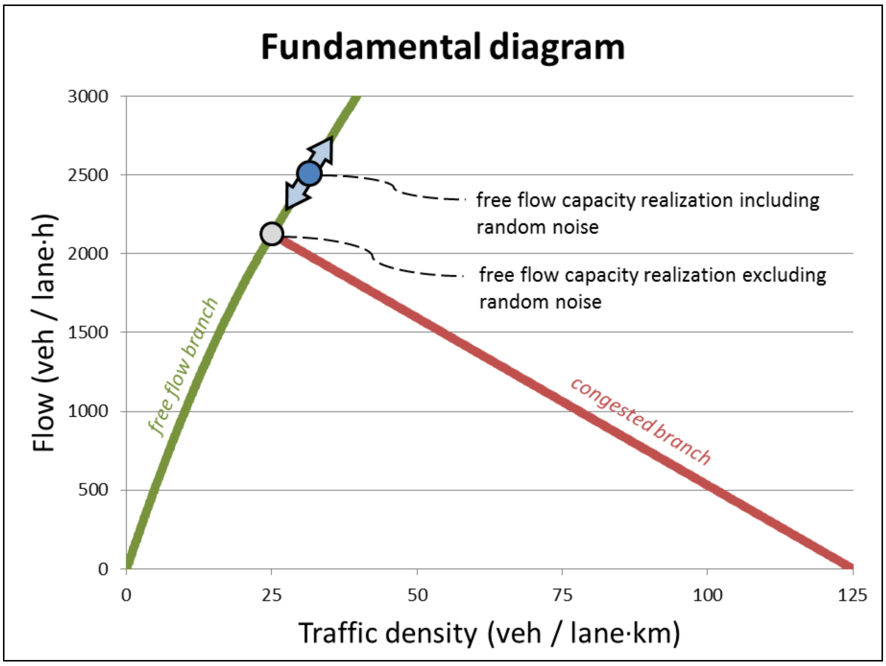


Figure A2.1: Alternative modeling approach, according to which the random noise in the free flow capacity corresponds to a variation along the free flow branch of the fundamental diagram, instead of to a variation of the fundamental diagram itself.

As a consequence of this approach, for part of the traffic density domain the traffic flow is not uniquely defined anymore. This means that the model should explicitly track now whether the traffic state in a cell is free flow or congested. Furthermore, some rules are required for the transitions between the two states. Figures A.2-A.4 show how the transition from free flow to congestion is proposed to be modeled.

Figure A2.2 shows the situation at t=1, when the traffic is still free flowing. Such a situation will persist as long as the traffic flow in the cell is lower than the existent free flow capacity realization. Now imagine that this free flow capacity goes down at a certain moment in time t=t2, to a value below the 'desired' flow (i.e. the flow that would be in 'equilibrium' with the traffic density in the cell, according to the free flow branch of the fundamental diagram). This situation is depicted in Figure A2.3. In this situation, the flow is limited at the free flow capacity. In addition to this, a state transition from free flow to congestion is triggered. This means that in the next time interval, the traffic state will be at the congested branch of the fundamental diagram, as shown in Figure A2.4.

Figure A2.2: Situation when the traffic is still free flowing

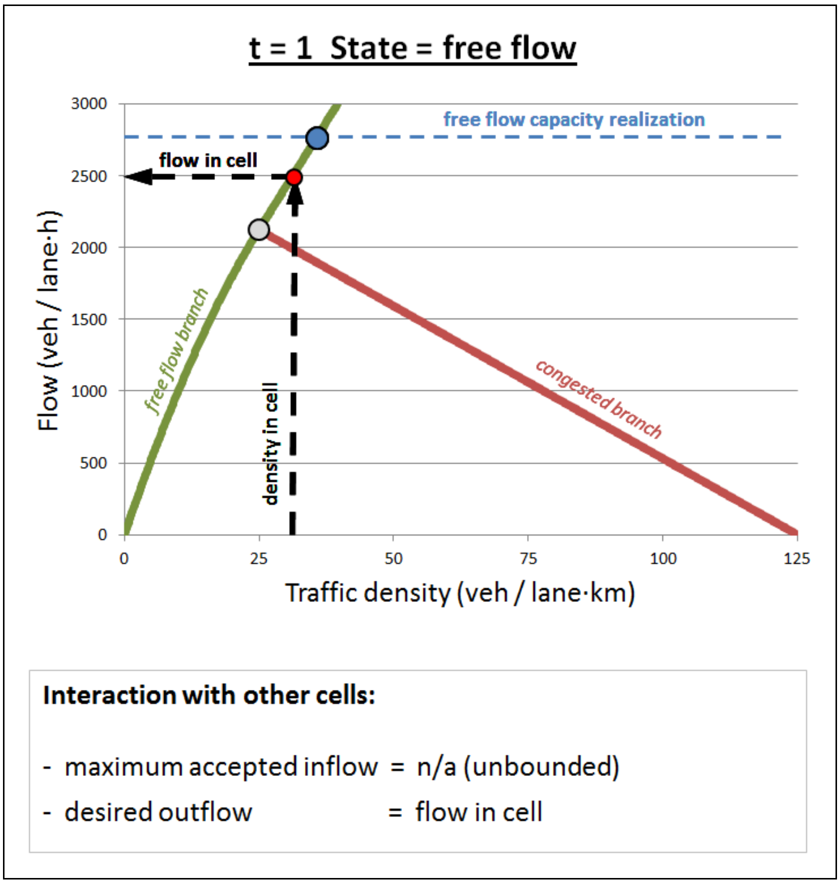


Figure A2.3: Situation in which the 'desired' traffic flow exceeds the available capacity, resulting in a transition to a congested traffic state

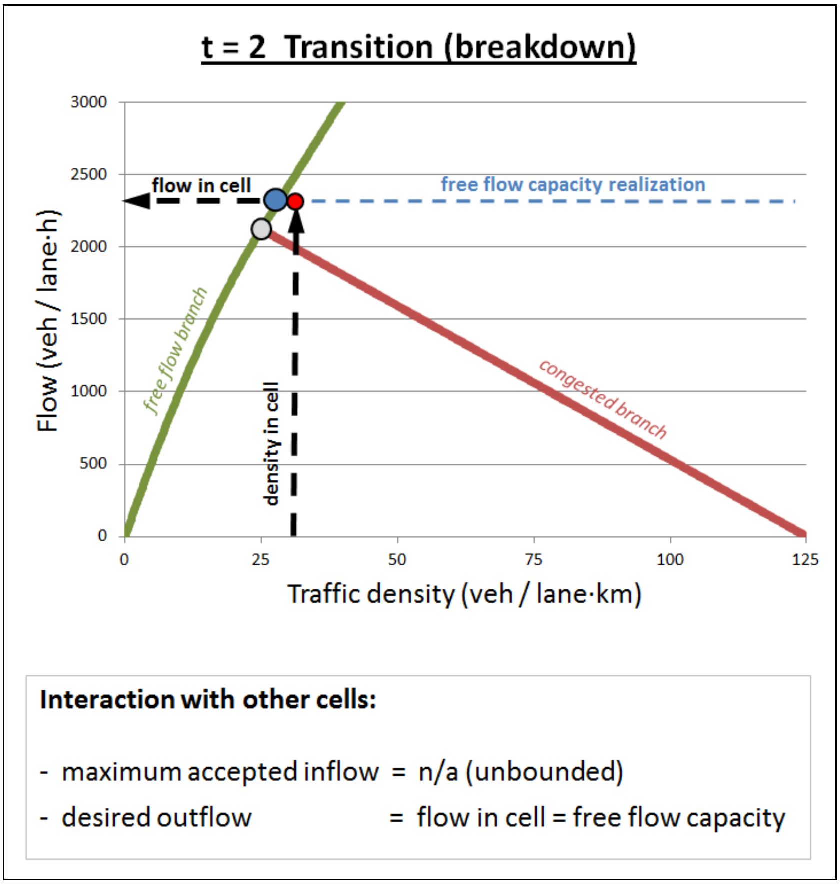
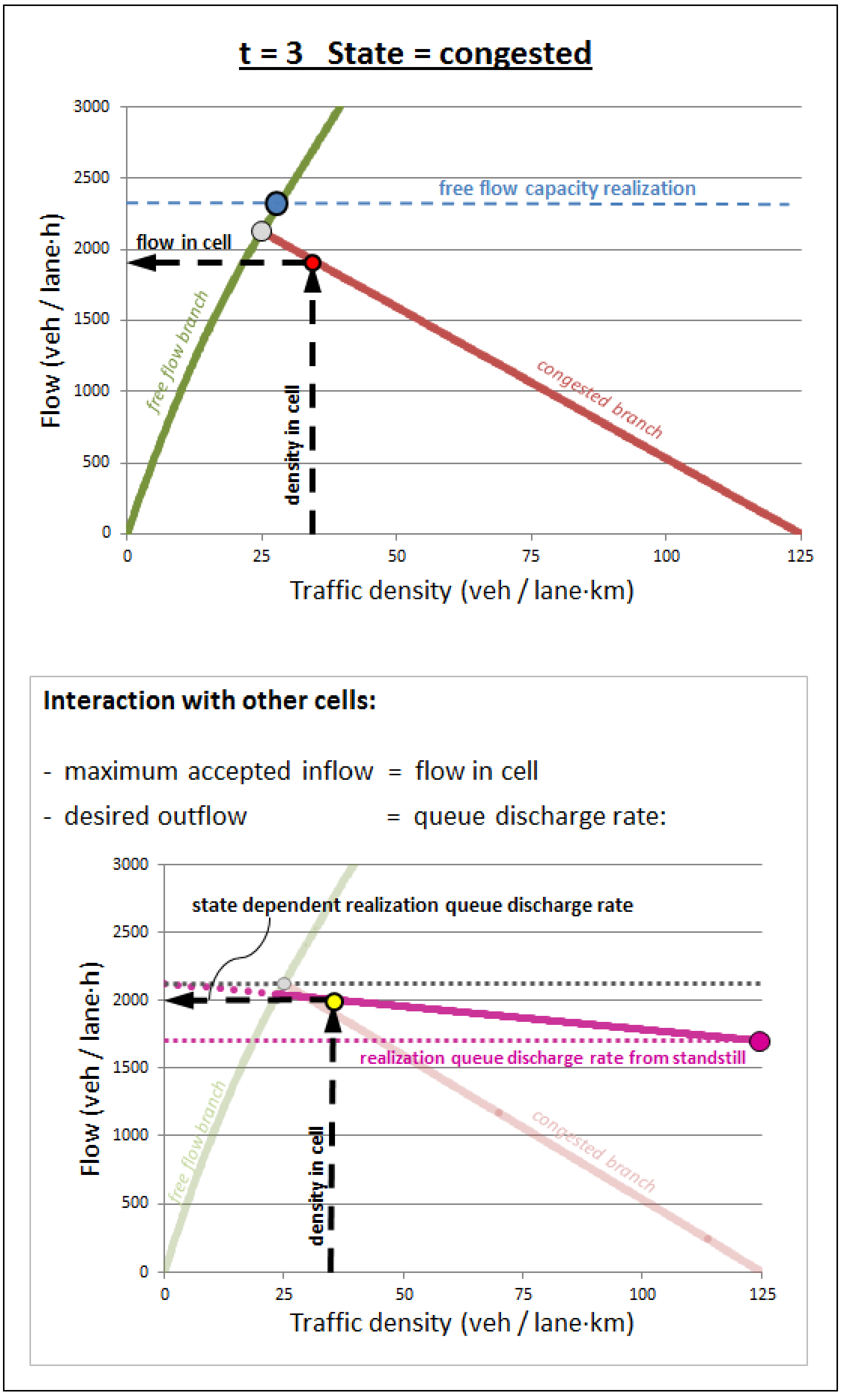


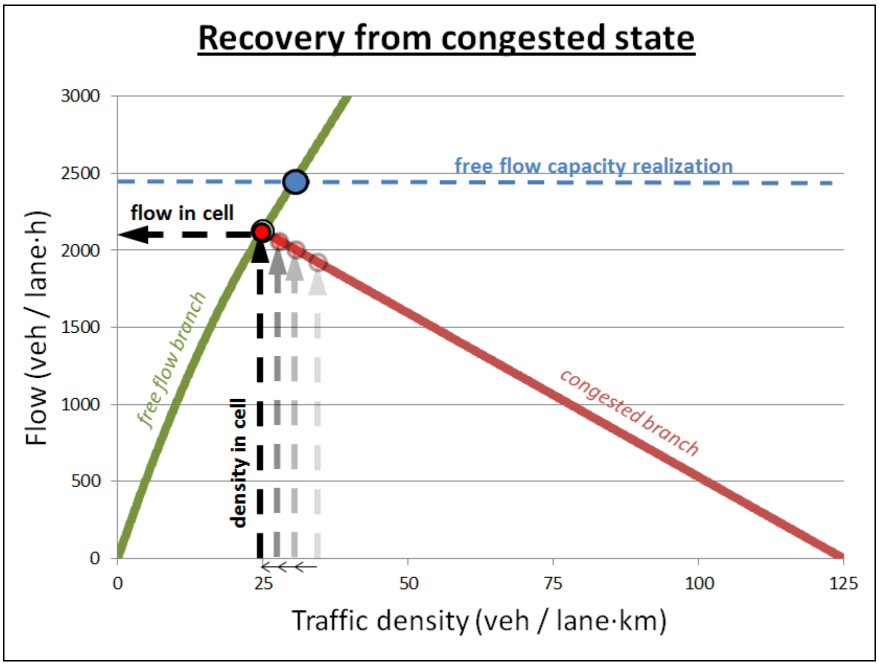
Figure A2.4: Situation after the transition to a congested traffic state



In the above figures, the rules for the interaction with the upstream and downstream cell have been indicated as well. Note that no longer a free flow inflow limitation is applied, as in the Godunov scheme. Limiting this inflow to the free flow capacity is in this case not really necessary anymore, since this free flow capacity has its effect now already via the state transition rule. (The only difference is that traffic congestion now will commonly be initiated in the first cell *downstream* of an oversaturated bottleneck location, rather than in its *upstream* cell. As long as the cell size is chosen sufficiently small, this does not matter however. In reality traffic actually most often breaks down at some distance beyond the bottleneck location as well.)

Figure A2.5 finally shows how the congestion gradually resolves again. As soon as the traffic density has decreased to a value below the lowest density in the congested branch of the fundamental diagram, the traffic state is restored to free flow.

Figure A2.5: Recovery to free flow conditions



Appendix 3 - Some computed travel time distributions

In section 8.5 an example was provided of a travel time distribution computed for a particular time of the day. This appendix gives some travel time distributions for other times of the day, obtained for the same route.

Figure A3.1: Travel time distribution on route 6, computed for Mon-Thu – 03:00

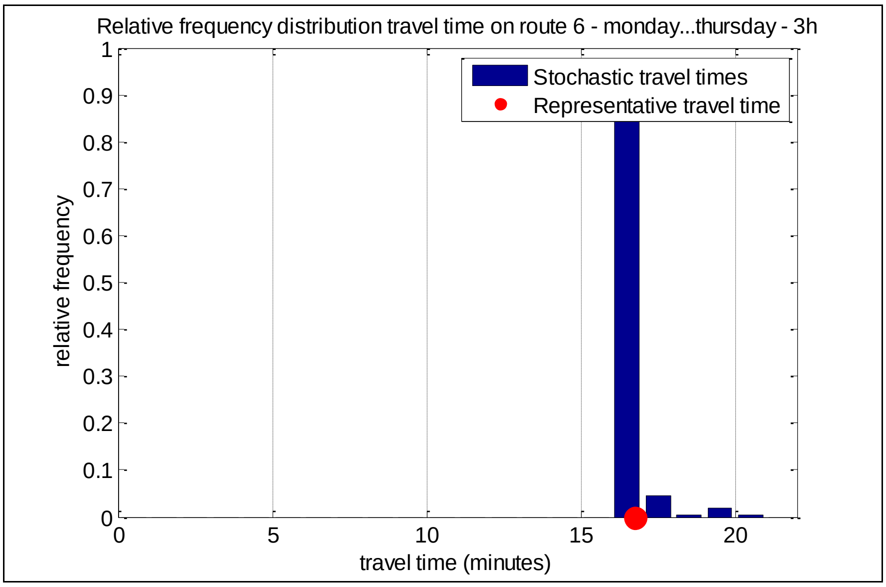


Figure A3.2: Travel time distribution on route 6, computed for Mon-Thu – 07:00

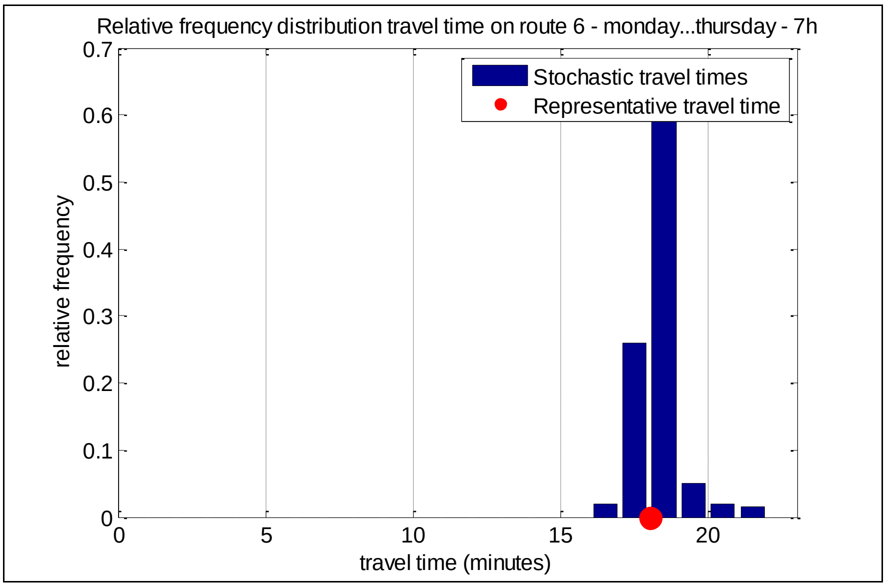


Figure A3.3: Travel time distribution on route 6, computed for Mon-Thu – 08:00

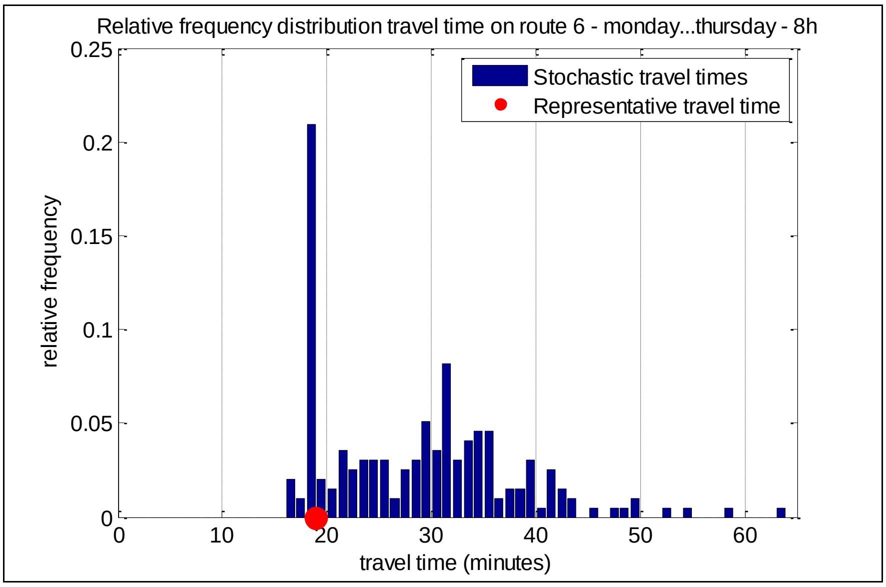


Figure A3.4: Travel time distribution on route 6, computed for Mon-Thu – 09:00

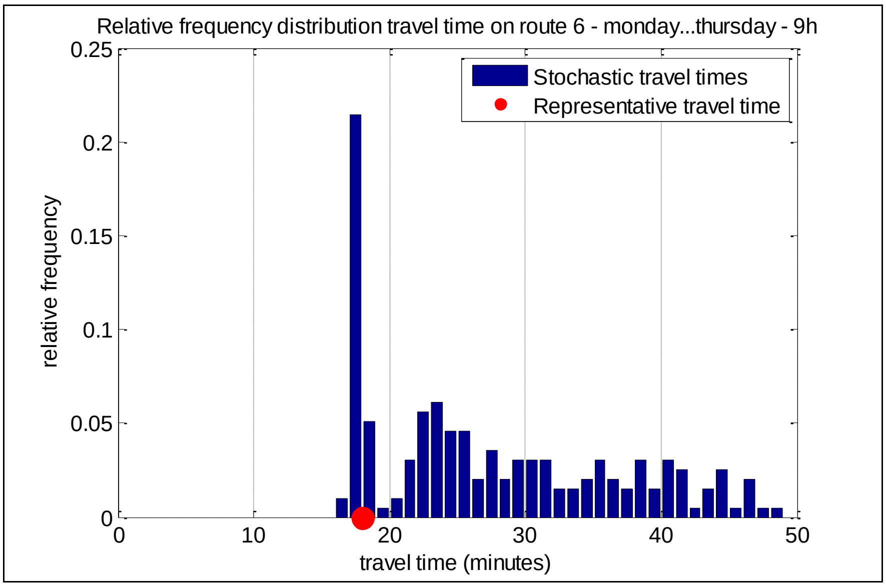


Figure A3.5: Travel time distribution on route 6, computed for Mon-Thu – 13:00

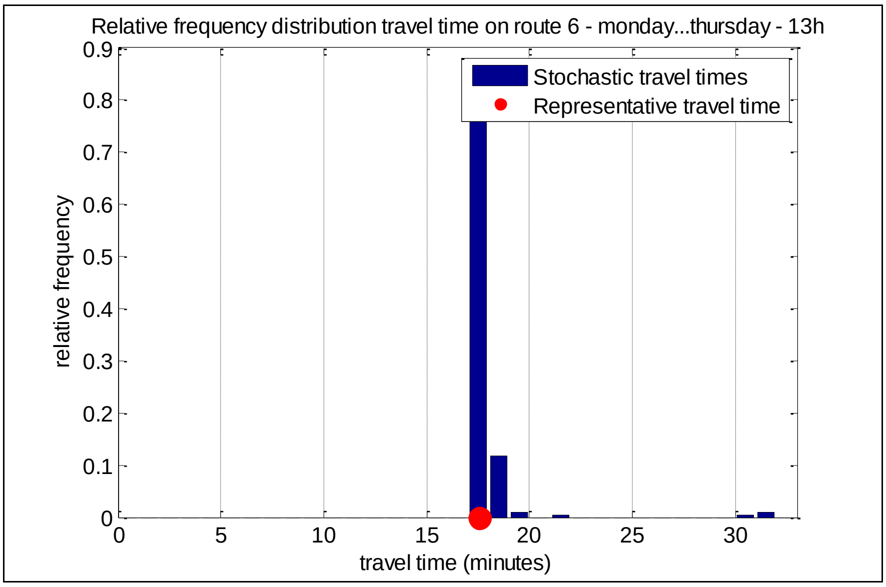


Figure A3.6: Travel time distribution on route 6, computed for Mon-Thu – 16:00

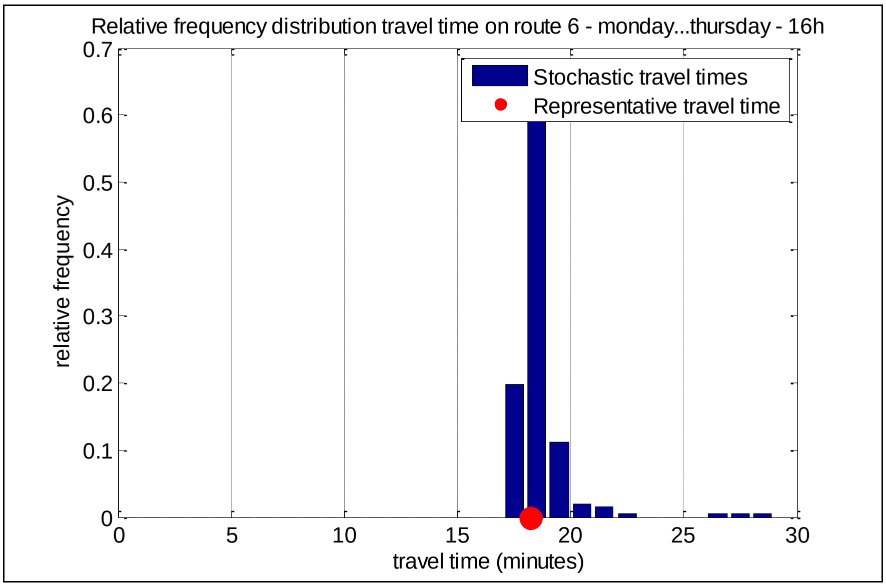


Figure A3.7: Travel time distribution on route 6, computed for Mon-Thu – 17:00

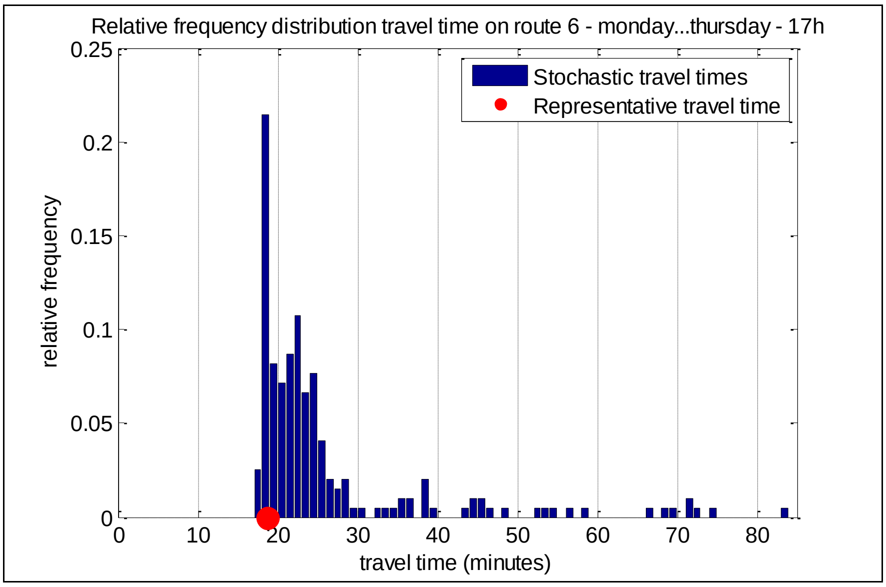


Figure A3.8: Travel time distribution on route 6, computed for Mon-Thu – 18:00

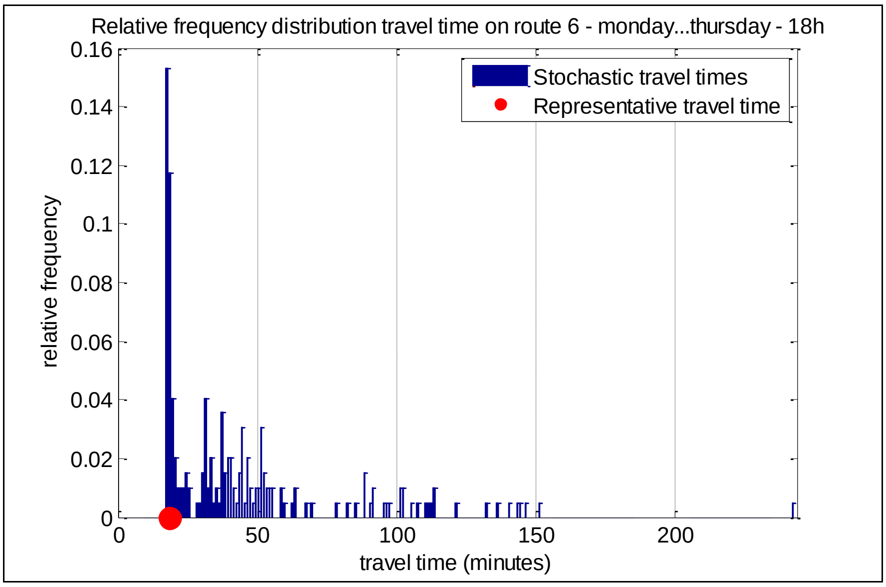
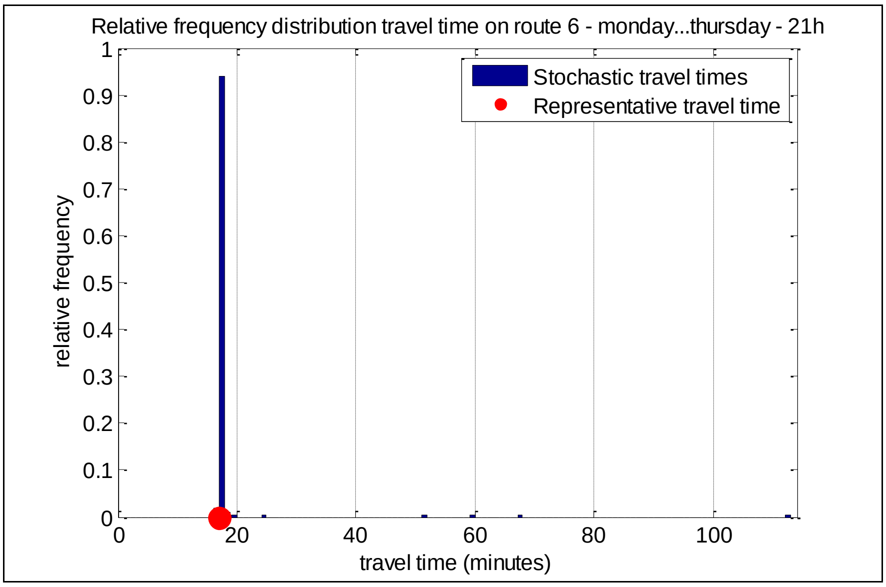


Figure A3.9: Travel time distribution on route 6, computed for Mon-Thu – 21:00



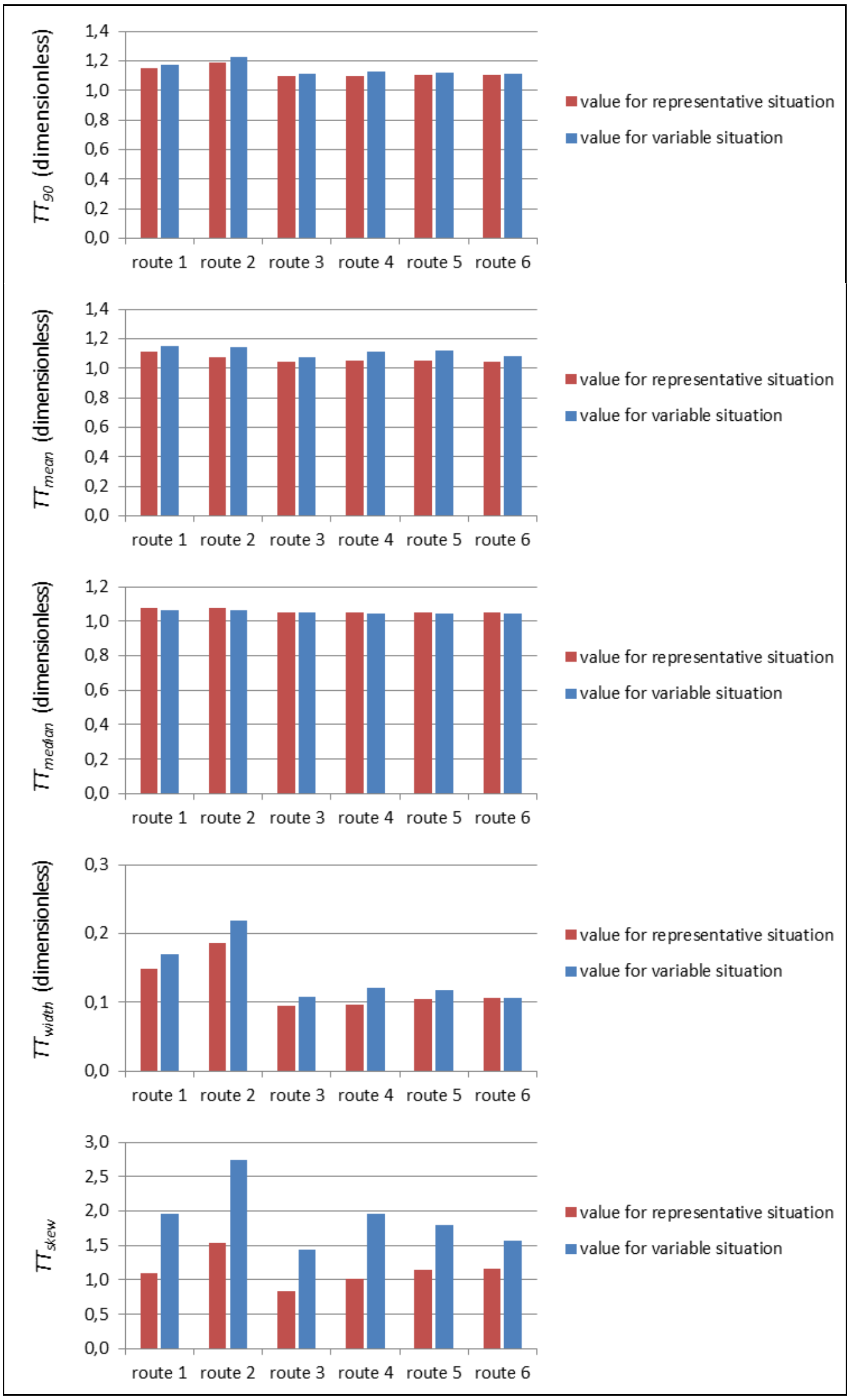
Appendix 4 - Sensitivity analysis capacity randomness

This appendix provides the results of the sensitivity analyses that have been performed on the spatial dependencies in the part of the capacity variability that is due to the intrinsic randomness in human driving behavior. The figures show the model output obtained for a situation in which the capacities are varied at the level of links, rather than at the level of individual cells of these links. This corresponds to the assumption that the local capacities are fully dependent over the whole length of the link.

Section A4.1 provides the outputs corresponding to the model results presented in section 8.5. Similarly, sections A4.2 en A4.3 provide the outputs that correspond to the model results presented in sections 8.6 and 8.7, respectively. For the interpretation of the presented results of the sensitivity analysis, the reader is referred to the respective sections in chapter 8.

A4.1 Results of the model run with full variability

Figure A4.1: The computed statistics of the overall travel time distributions (90th-percentile, mean, median, width and skew), compared with their 'representative' counterparts



Note: all values of the first four indicators have been made dimensionless by division by the free flow travel time. The fifth indicator (i.e. the skew) is already dimensionless by itself.

Figure A4.2: The travel time statistics (90th-percentile, mean and median) computed for route 6, as a function of the time of day and the category of days. For the purpose of comparison, the representative travel time is included as well (for weekdays only).

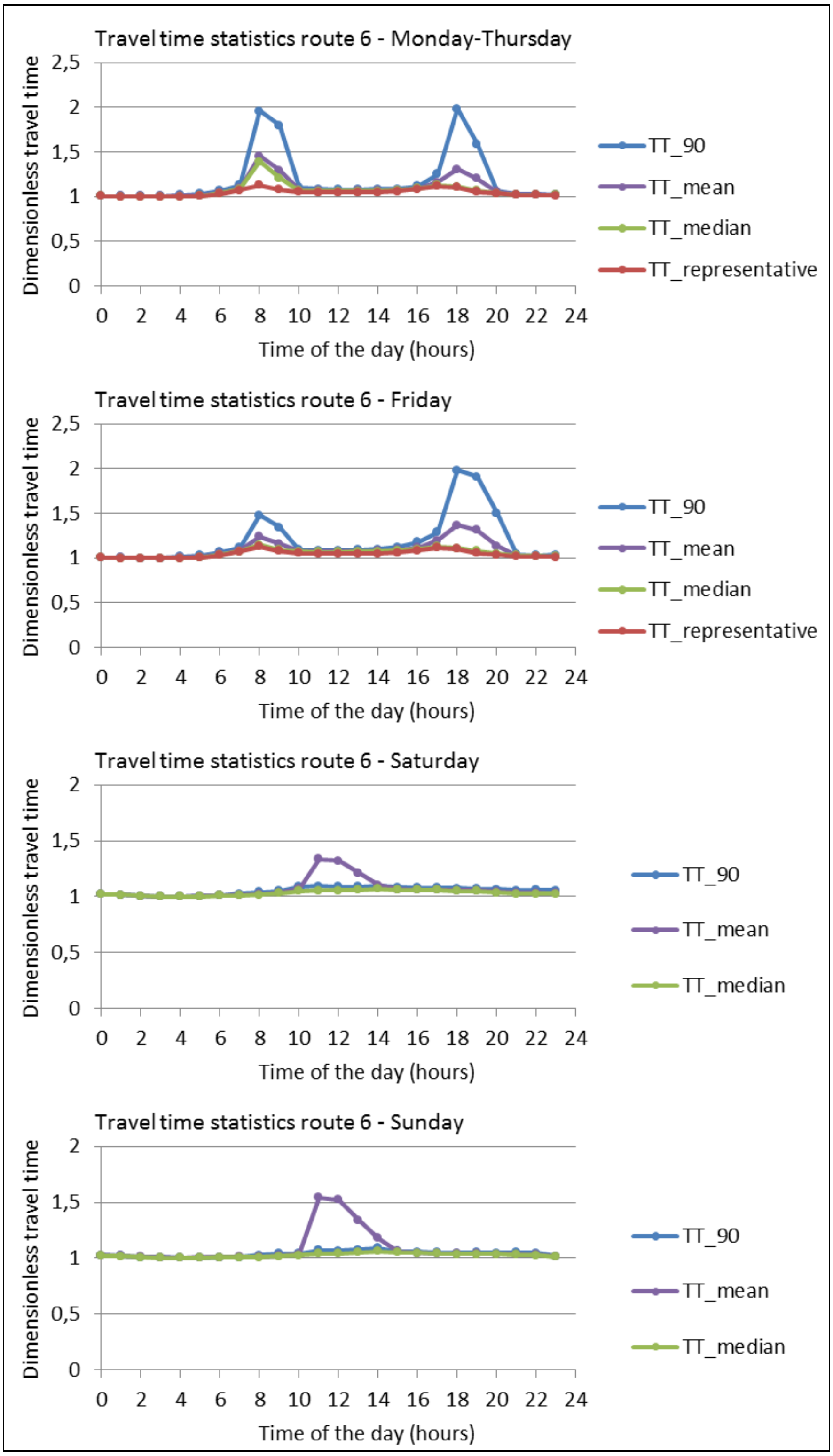


Figure A4.3: Distribution of the travel time on route 6 at 17:00, computed for Monday – Thursday

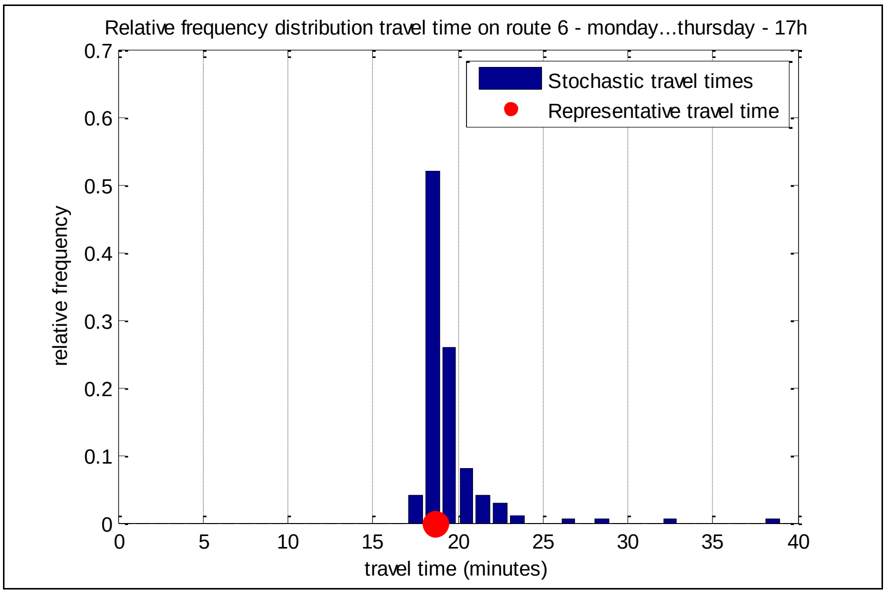


Figure A4.4: The width of the travel time distribution computed for route 6, as a function of the time of day (separately for Mon-Thu and Fri)

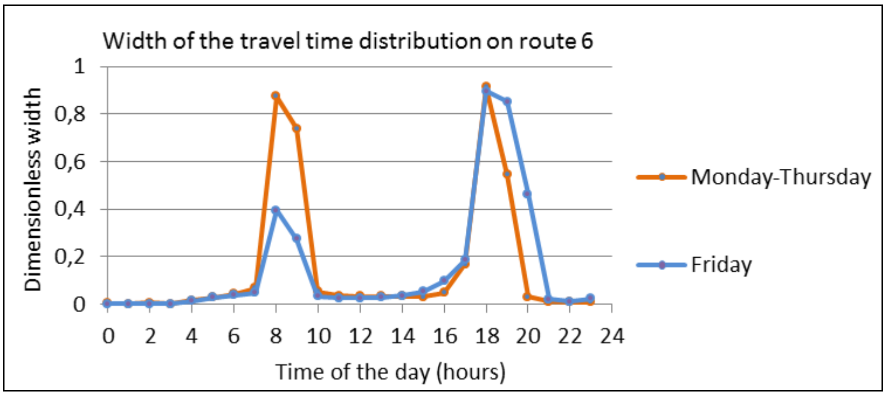


Figure A4.5: The skewness of the travel time distribution computed for route 6, as a function of the time of day (separately for the different categories of days)

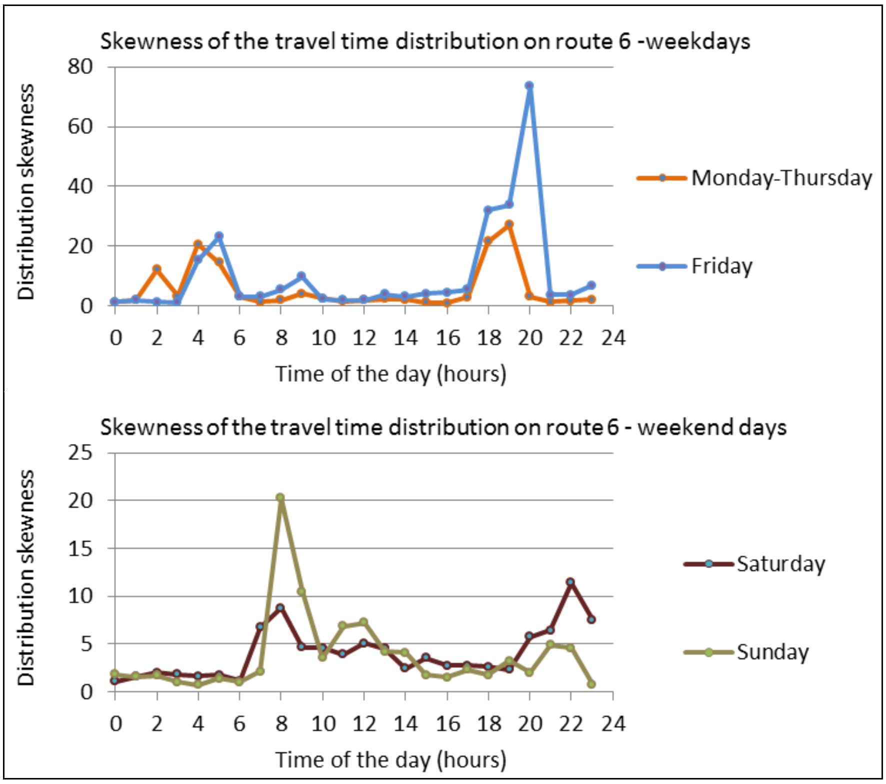
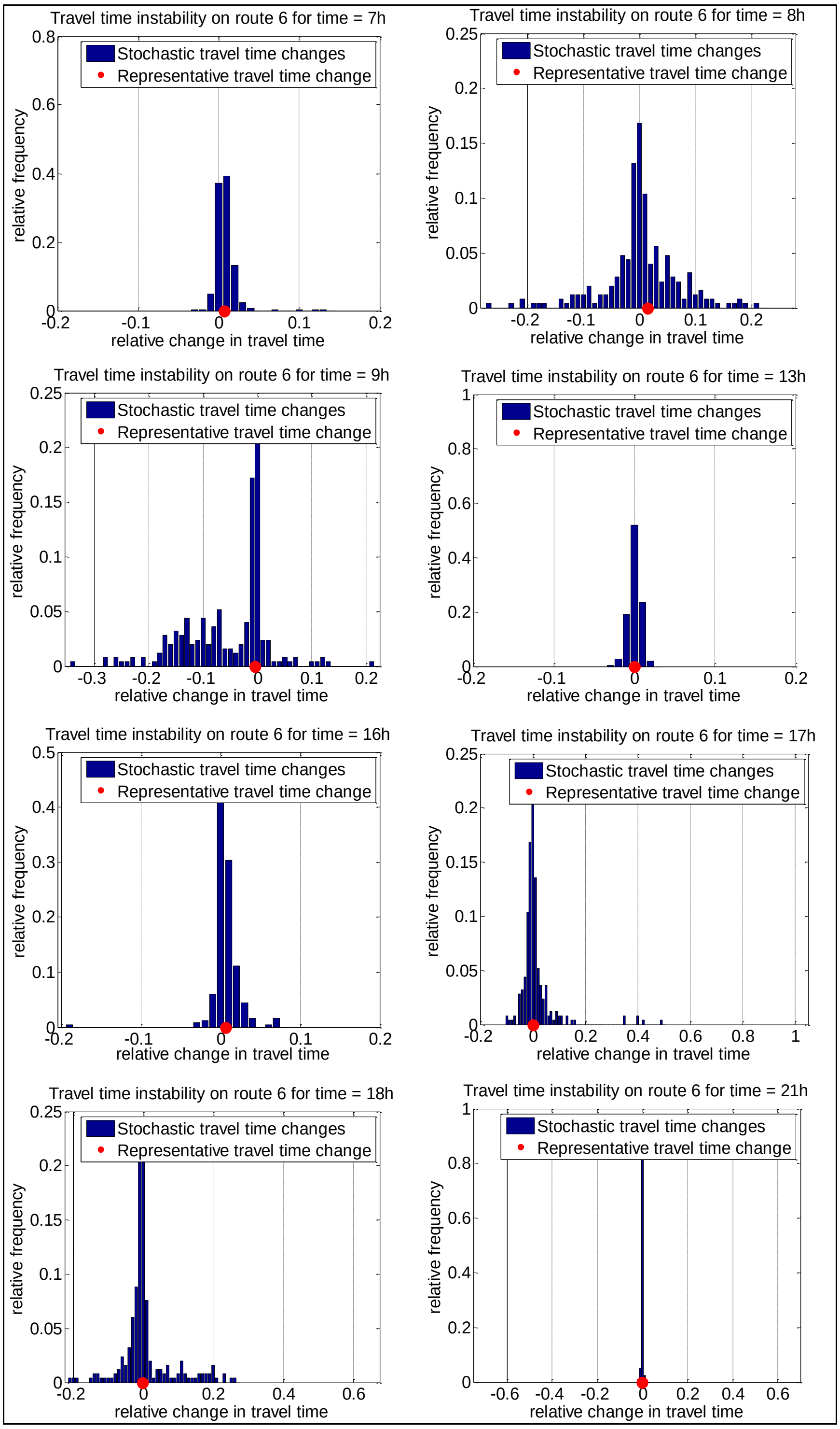
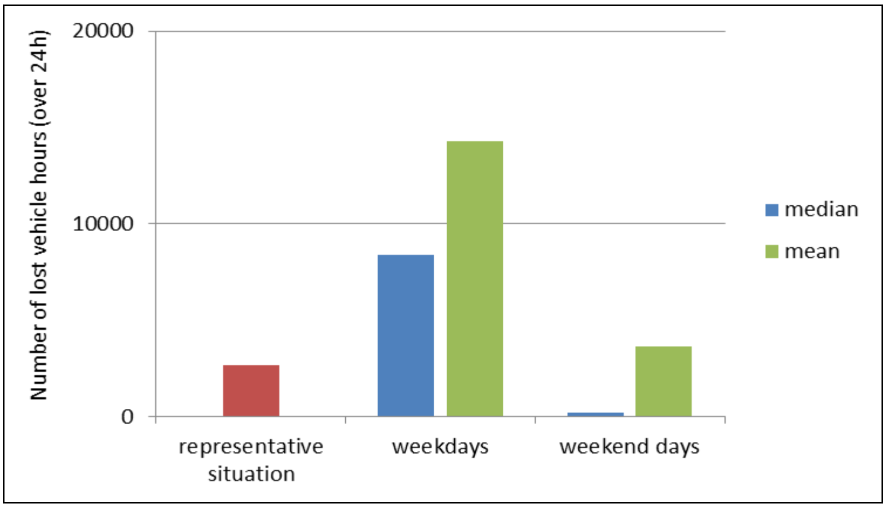


Figure A4.6: The travel time instability on route 6, for different times of the day (for weekdays only)



No important diffence for first half of the day (including midday). During and after evening peak travel time much more stable now.

Figure A4.7: The computed values of the mean and median numbers of lost vehicle hours (incurred within the boundaries of the network), compared with the number of lost vehicle hours in the representative situation



A4.2 Relative importance of different sources of variation

Figure A4.8: The statistics of the overall travel time distributions (90th-percentile, mean, median, width and skew), computed for situations in which one of the different sources of variability is omitted from the model. (All values expressed as a ratio to the value obtained from the model run with full variability.)

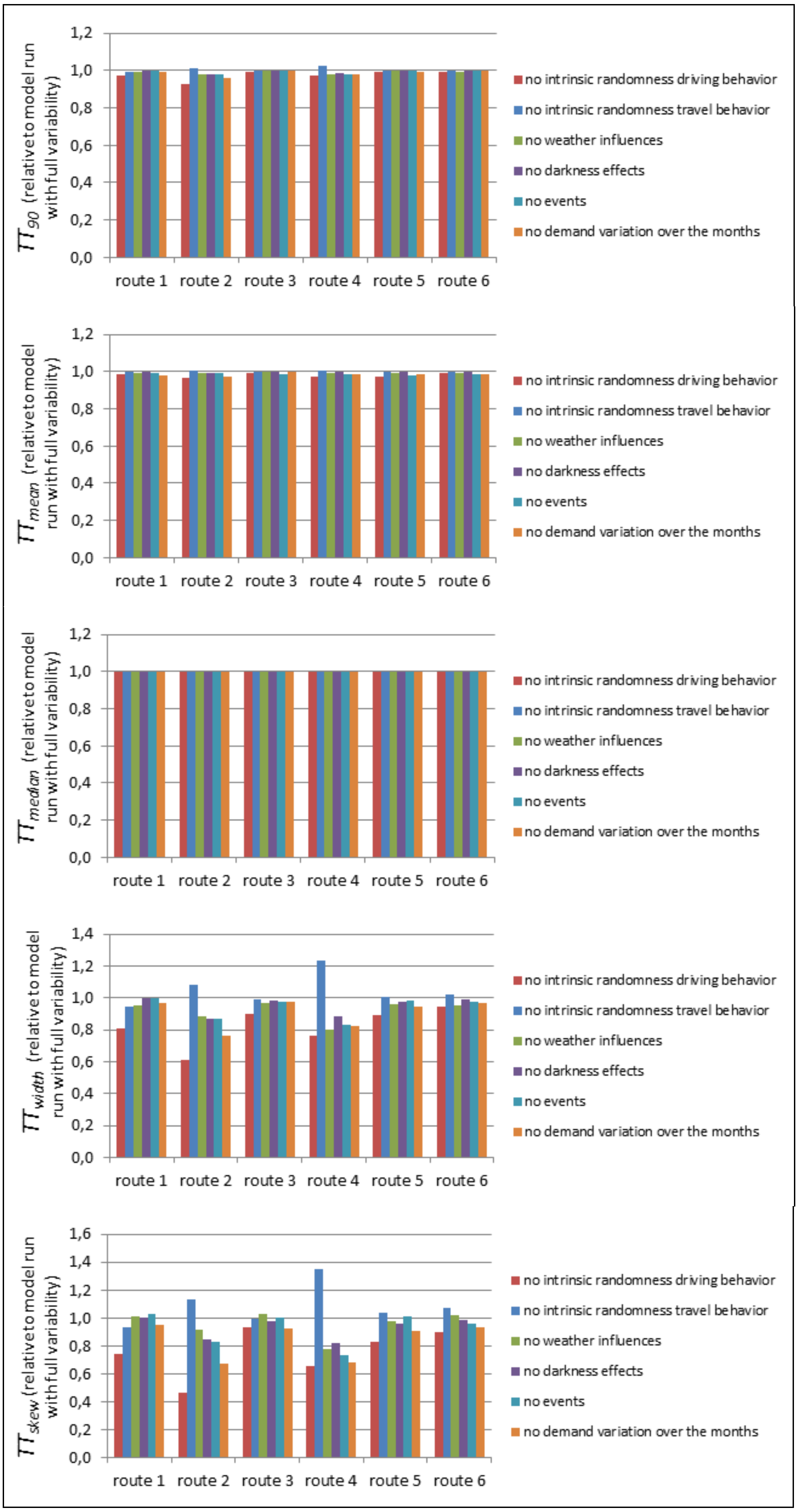


Figure A4.9: The travel time statistics for Mon-Thu – 17:00 (90th-percentile, mean, median, width and skew), computed for situations in which one of the different sources of variability is omitted from the model. (All values expressed as a ratio to the value obtained from the model run with full variability.)

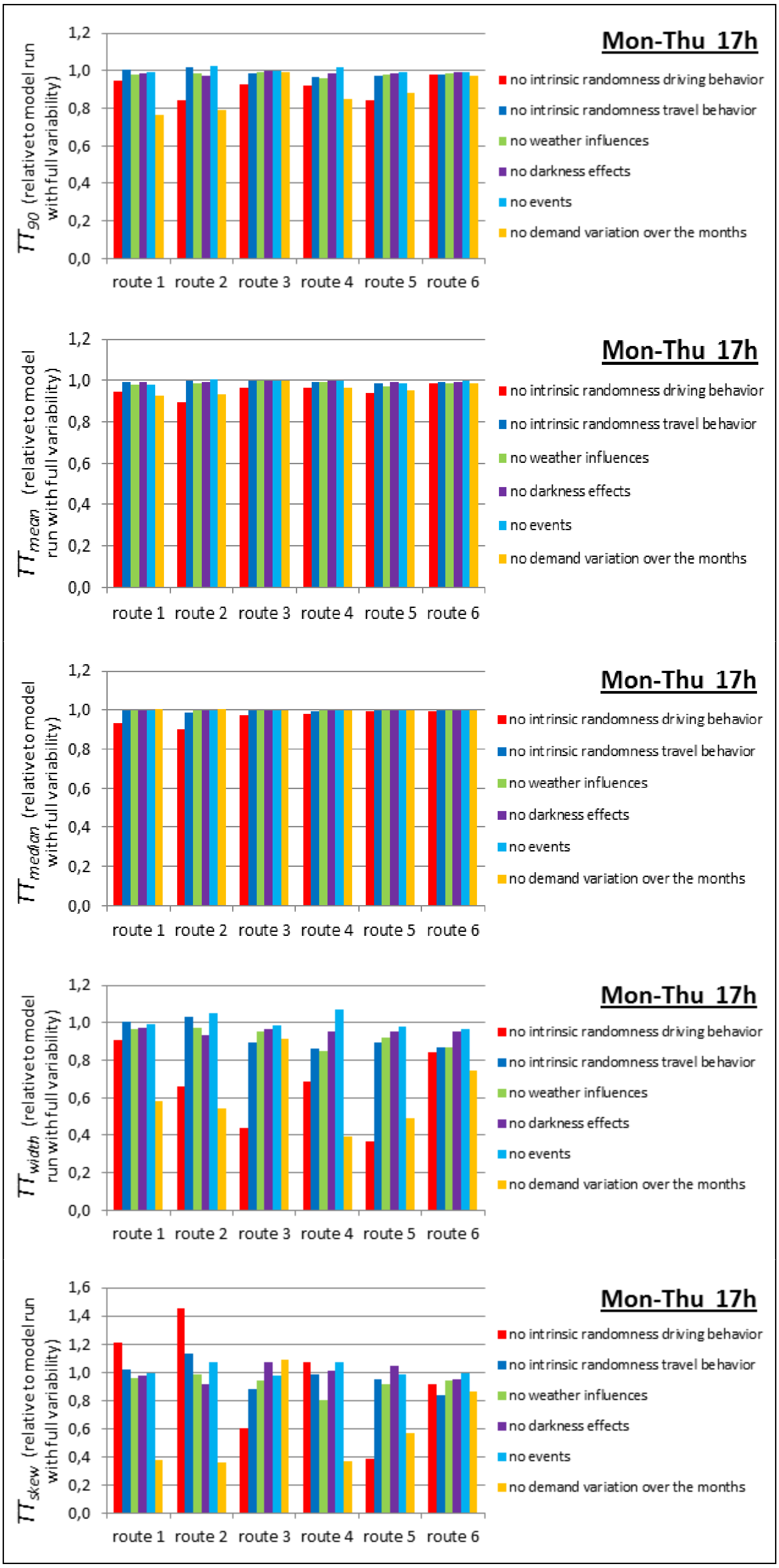


Figure A4.10: The relative changes in the mean and median numbers of lost vehicle hours if a given source of variability is omitted from the model (as compared with the numbers obtained from the model run with full variability), for weekdays

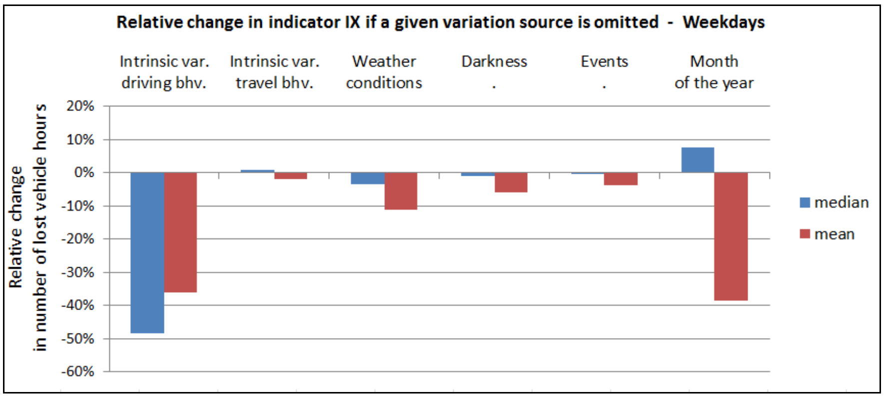
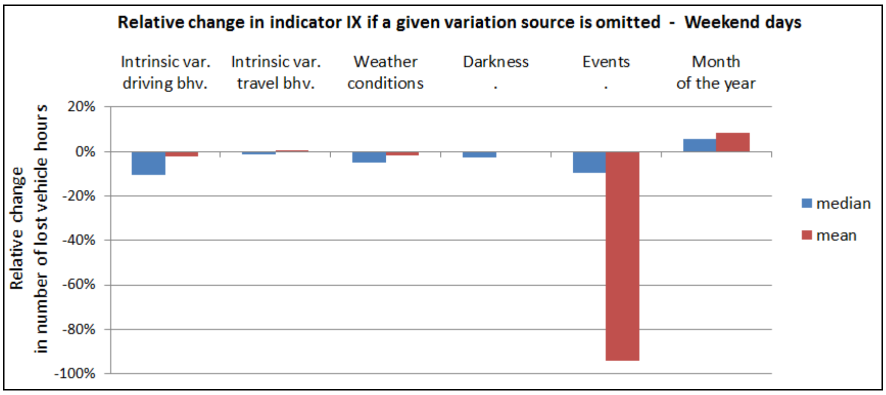


Figure A4.11: The relative changes in the mean and median numbers of lost vehicle hours if a given source of variability is omitted from the model (as compared with the numbers obtained from the model run with full variability), for weekend days



A4.3 Effects of a rush-hour lane

Figure A4.12: The statistics of the overall travel time distributions (90th-percentile, mean, median, width and skew) in the situation with the rush-hour lane, compared with those in the situation without this lane

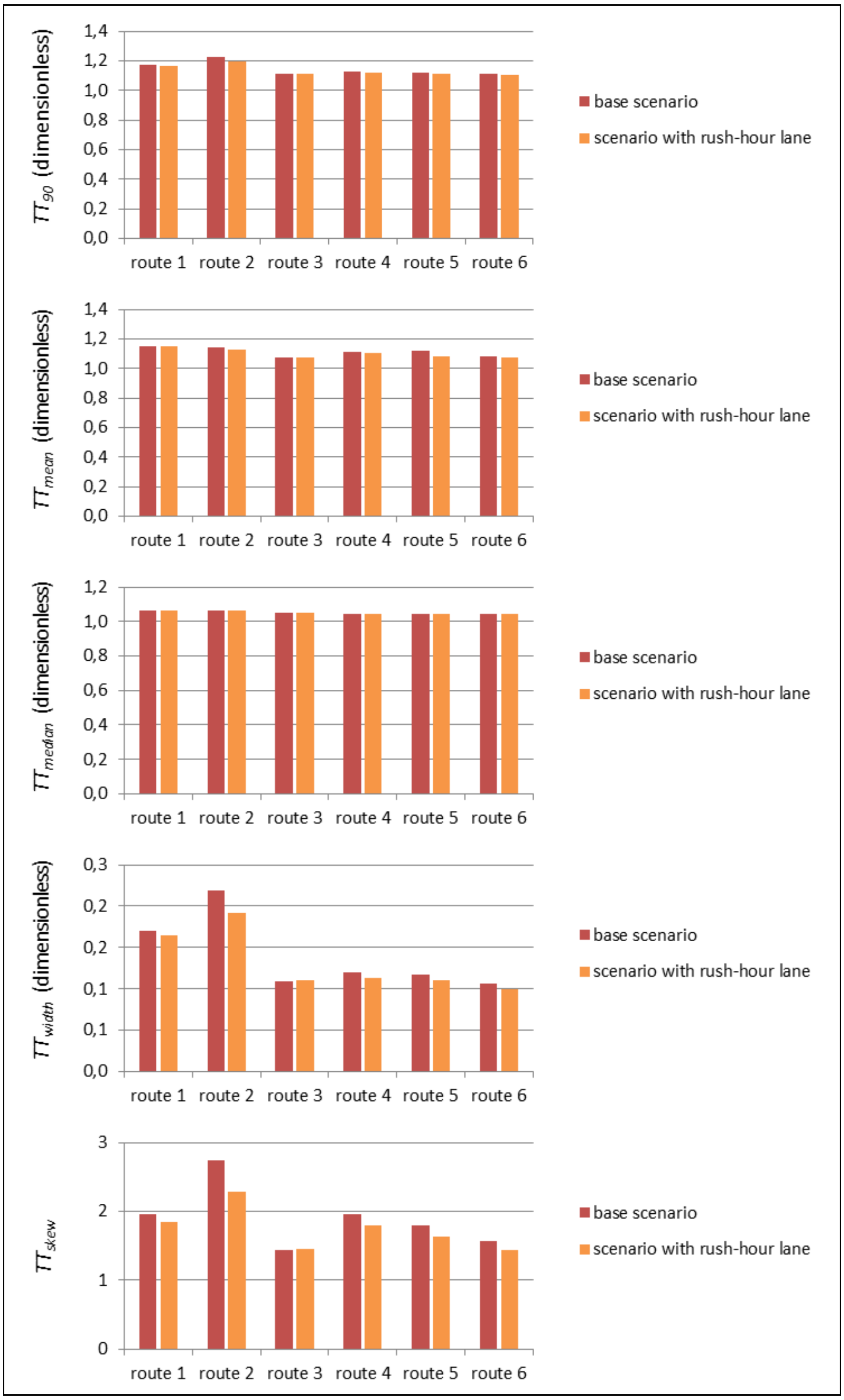


Figure A4.13: The travel time statistics for Mon-Thu – 08:00 (90th-percentile, mean, median, width and skew), in the situation with the rush-hour lane, compared with those in the situation without this lane

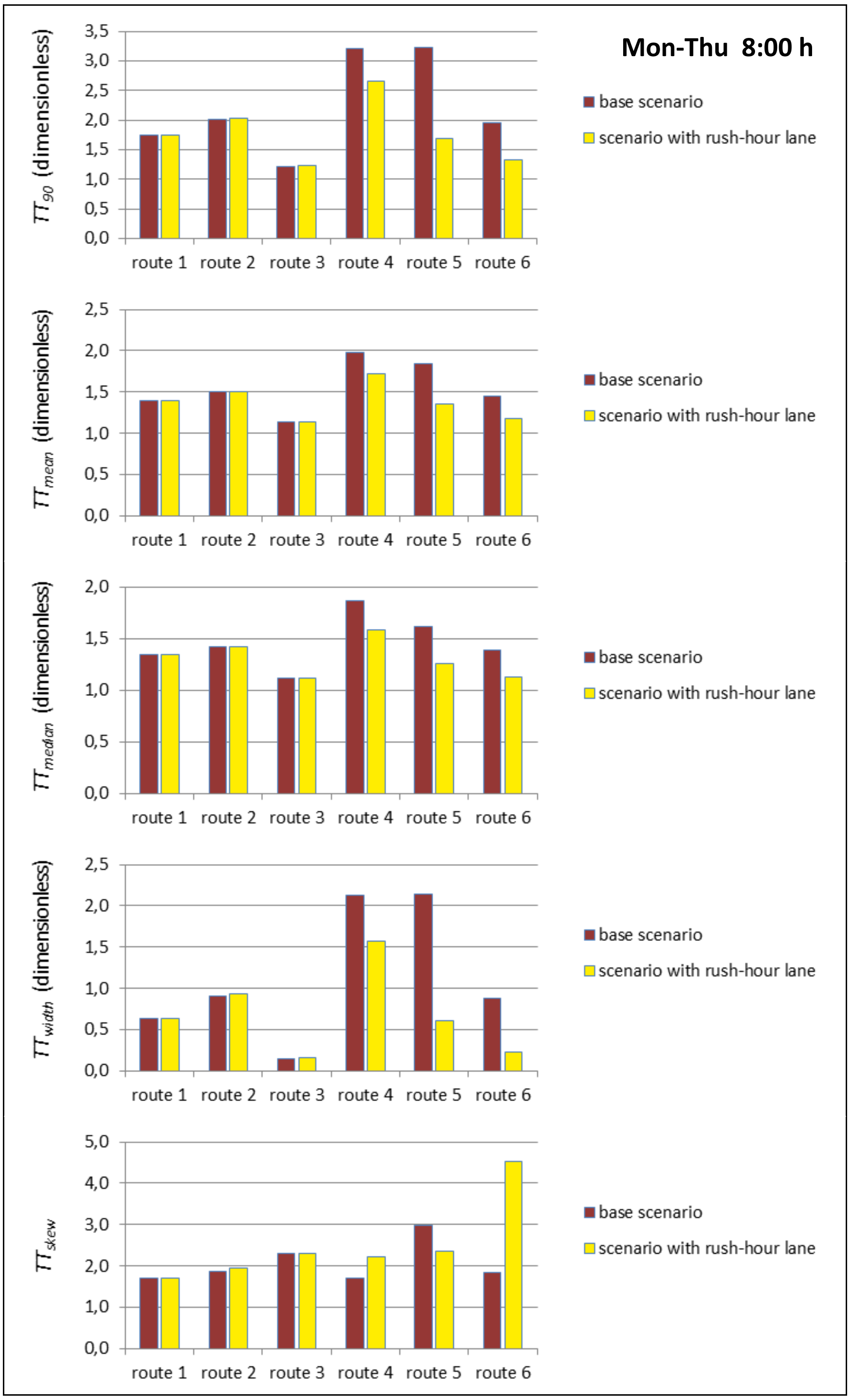


Figure A4.14: The travel time statistics for Mon-Thu – 09:00 (90th-percentile, mean, median, width and skew), in the situation <u>with</u> the rush-hour lane, compared with those in the situation <u>without</u> this lane

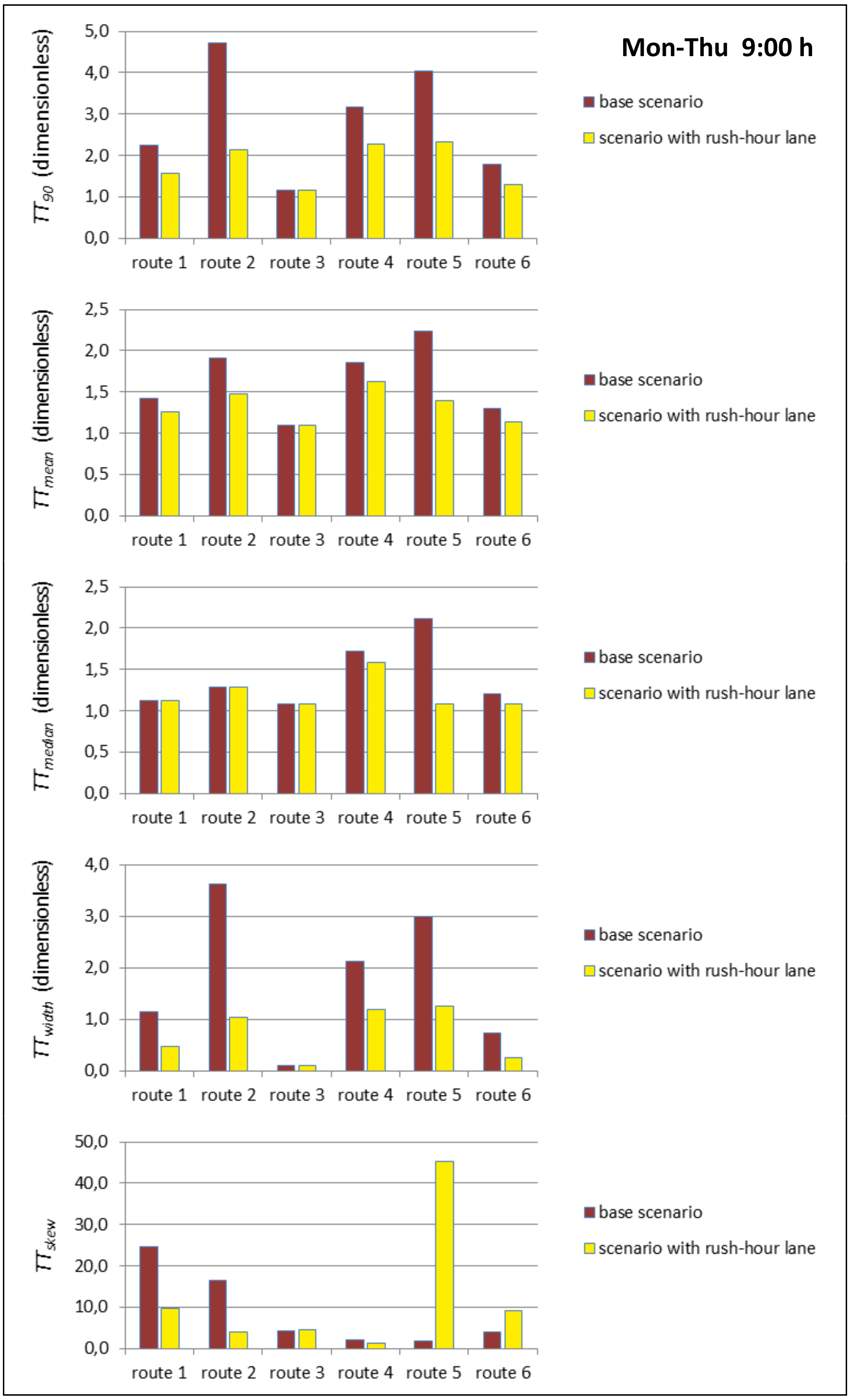


Figure A4.15: The effect on the travel time instability on route 5, for 08:00 (weekdays only)

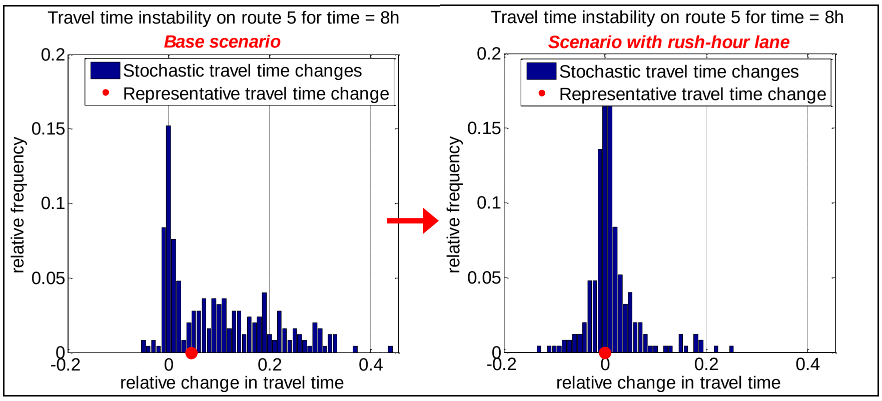


Figure A4.16: The effect on the travel time instability on route 5, for 09:00 (weekdays only)

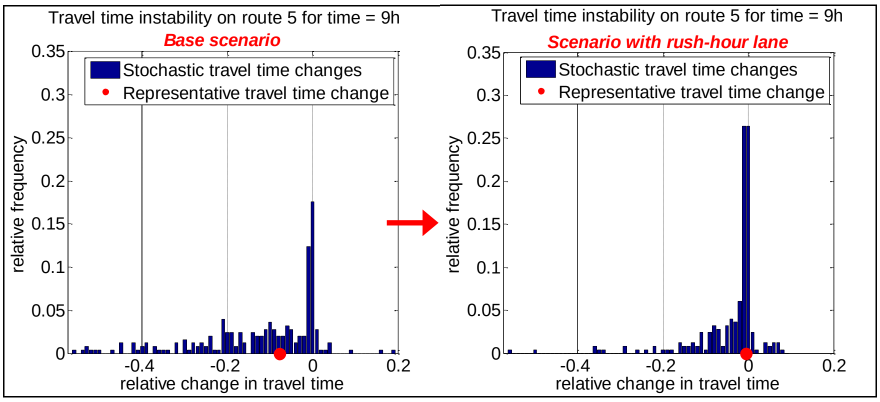


Figure A4.17: The effect on the travel time instability on route 6, for 09:00 (weekdays only)

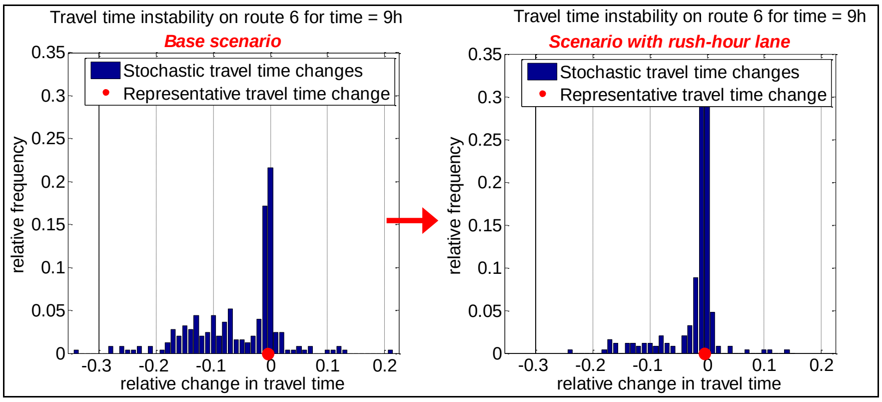
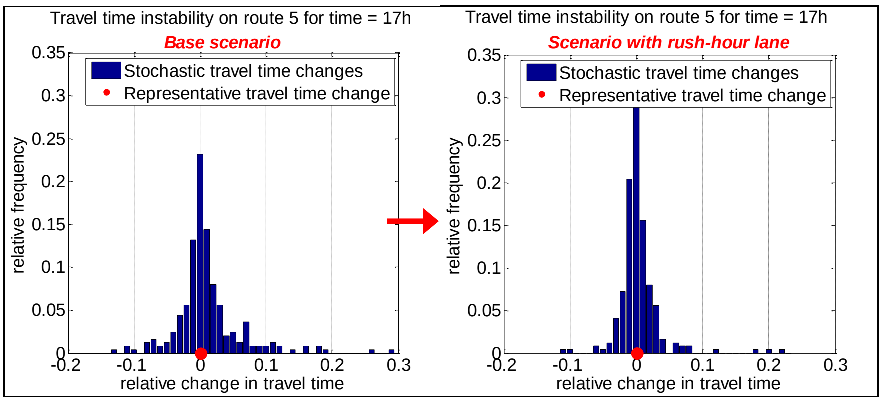


Figure A4.18: The effect on the travel time instability on route 5, for 17:00 (weekdays only)



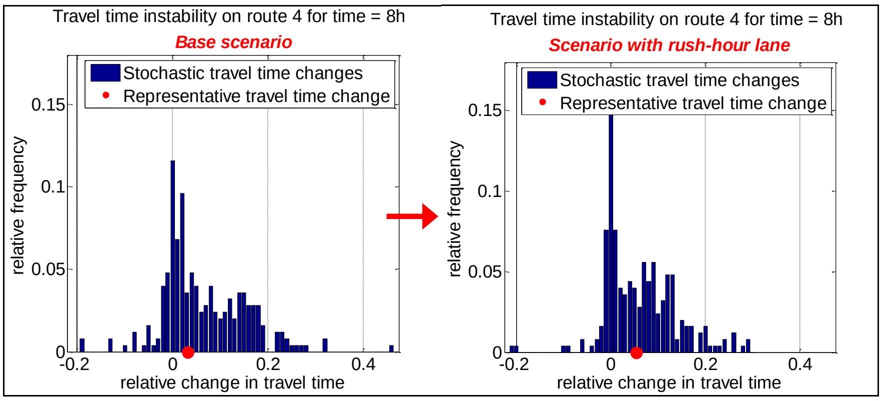


Figure A4.19: The effect on the travel time instability on route 4, for 08:00 (weekdays only)

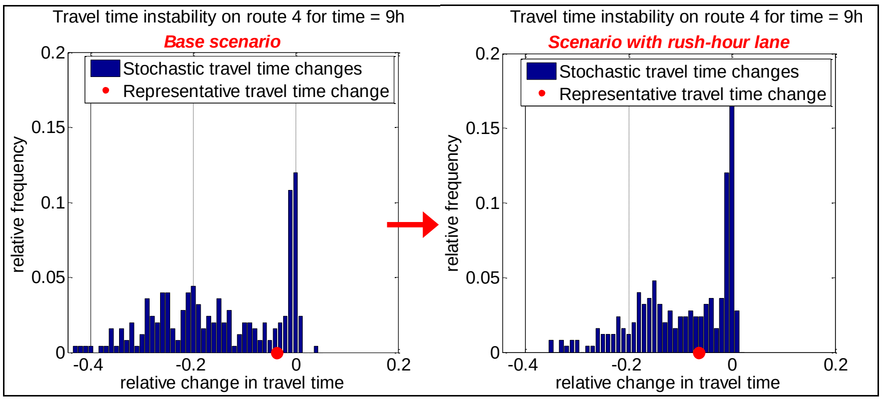


Figure A4.20: The effect on the travel time instability on route 4, for 09:00 (weekdays only)

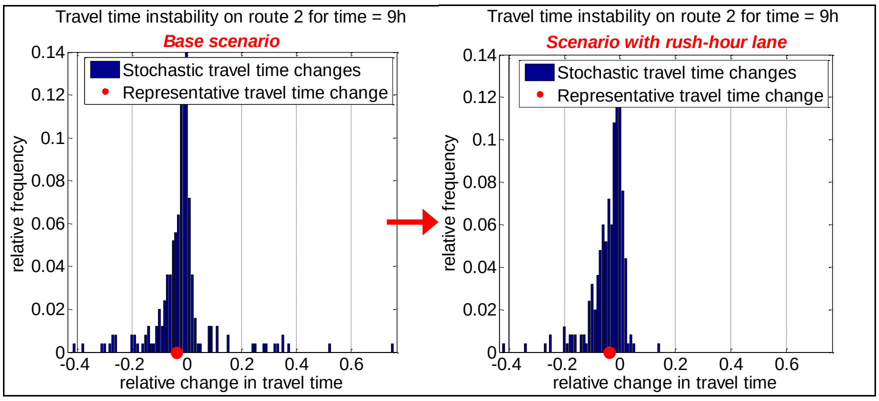


Figure A4.21: The effect on the travel time instability on route 2, for 09:00 (weekdays only)

Figure A4.22: The mean and median numbers of lost vehicle hours on weekdays in the situation with the rush-hour lane, compared with those in the situation without this lane

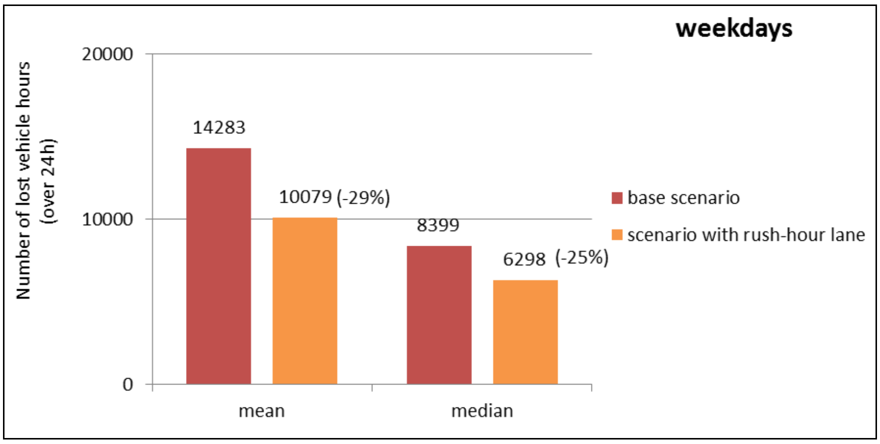


Figure A4.23: The mean and median numbers of lost vehicle hours on weekend days in the situation with the rush-hour lane, compared with those in the situation without this lane