Network-wide Traffic State Estimation using the Macroscopic Fundamental Diagram

A data fusion approach

Marianthi Mermygka

TU Delft SWECO

Network-wide Traffic State Estimation using the Macroscopic Fundamental Diagram

A data fusion approach

by

Marianthi Mermygka

to obtain the degree of Master of Science

at the Delft University of Technology,

to be defended publicly on Monday October 24, 2016 at 11:00 am.

Student number:	4416996	
Project duration:	March 23, 2016 – October 24, 2016	
Thesis committee:	Prof. dr. ir. J.W.C. van Lint,	TU Delft, Chairman
	Dr. V. L. Knoop,	TU Delft, Daily Supervisor
	Ir. N. Henkens,	Sweco, Supervisor
	Prof. dr. ir. A. Verbraeck,	TU Delft, External Supervisor
	Ir. P. B.L. Wiggenraad,	TU Delft, Thesis Coordinator

This thesis is conducted in cooperation with Sweco.

"... I can live with doubt, and uncertainty, and not knowing. I think it's much more interesting to live not knowing than to have answers which might be wrong."

Richard P. Feynman

Preface

What I always liked about my studies in Transport and Planning is that I learnt how to solve problems that can actually affect people's lives and make a difference in our everyday environment. However, in the beginning of my thesis, I was confused on what problem to solve, and all I knew was that I wanted to work with traffic data. Little did I know, that I would end up diving so deep in the theories of traffic flow and the Macroscopic Fundamental Diagram, but my interest about it only kept growing. I feel very lucky that I finally did a thesis project that kept me excited throughout all the seven months of my work, finding motivation that my traffic state estimation may help one day to alleviate congestion in our cities.

Nevertheless, none of these would have been possible without Prof. Hans van Lint, who I would really like to thank for showing me the trust to do my Master thesis with him. All of his advice and insights have been of extreme importance. Moreover, I would like to thank Gerbrand Klijn and Bas van der Bijl from Sweco that supported the start of this thesis project. I would also really like to thank my supervisor from Sweco, Niels Henkens, for always helping me and being excited with my progress. Most important of all, I would like to thank my daily supervisor, Victor Knoop, who not only offered his advice and support, but he was actively thinking with me how to solve any difficulties that occurred. Our meetings were always enlightening and the main reason that the progress of this thesis was so smooth.

As my student time in Delft comes soon to an end, I would like to thank all of my friends and my master classmates that were along with me these past two years, and especially, my beloved housemates, my friend Ana, and my Yamas friends. Special thanks go to Karel Devriendt who helped me so much with his ideas on the uncertainties estimation. Probably the next person after me who knows the contents of this thesis is Lotfi Massarweh, who I would like to thank from the bottom of my heart, not only for his amazing help with my thesis, but mainly for being next to me, always supportive and positive at the most difficult moments during the last months of the thesis. Closing my acknowledgements, I would like to thank my parents and my brother for always being my side, and especially my mother that she has learnt me to be strong, follow my dreams and she has been a true role model for me.

Marianthi Mermygka
Delft, October 2016

Summary

As the percentage of urban population rises, the demand for urban mobility constantly increases leading to serious traffic congestion issues. In the past, new infrastructure was constructed to serve this demand, but this is neither cost-effective nor sustainable any more. Therefore, clever solutions that optimize the existing traffic system need to be explored. For example, solutions can be found in the domain of smart traffic control strategies that can contribute to improve the operation of the road network. Nevertheless, in order to accomplish a successful traffic control strategy, it is a crucial prerequisite to describe the traffic state accurately. Especially when the traffic state is described at a network level, it can be more precise, because the interaction between intersections is taken into consideration.

In this context, the goal of this research project is to solve the problem of knowing the traffic situation network-wide. Many approaches have been proposed by literature to describe the traffic state for a network, such as kinematic wave theory, cell transmission models or macroscopic traffic simulation models. However, most of them have limitations or require a lot of computation time. For this reason, researchers have been examining the existence of a simple and fast way that can describe sufficiently the dynamics of a road network. As a result, the concept of the Macroscopic Fundamental Diagram (MFD) was developed, which is a single function relating the average flow to the average density of a network, capturing the prevailing network situation. The results of the studies exploring the MFD so far, indicated that it can be used successfully for traffic monitoring, perimeter control or support and evaluation of traffic control strategies. Such results are encouraging to develop the idea of the MFD further so that it can become a broadly used traffic state estimation tool.

Thus, the objective of this project is to estimate the traffic state of urban networks using the MFD. Although the required variables for the MFD are only the network flow and the network density, it is not always easy to estimate them, due to data limitations. Hence, systematic investigation is required to explore the necessary traffic data to obtain the MFD. If the MFD of the network is known, all that is needed to have an accurate traffic state estimation is to know where we are on the MFD at any desired moment. Nevertheless, traffic data are always erroneous, so the certainty of the derived traffic state using the MFD needs to be examined. The use of the MFD for traffic state estimation can also be tested in special cases, such as in the case of an accident occurring in the network.

Within this framework, the following main research question was formed to accomplish the objective of this project:

Main Research Question: *How to derive a network-wide traffic state estimation using the Macroscopic Fundamental Diagram?*

The vehicle trajectories are the ideal way to obtain the MFD accurately, but knowledge of 100% of the vehicle trajectories is an unrealistic situation. For this reason, a data fusion approach was explored in this project, taking advantage of the valuable information that also a low fraction of vehicle trajectories can offer and combining it with the - almost always available - detector data. The main idea is that if the proportion of the known vehicle trajectories in relation to the total number of vehicles in the network is calculated, then the information from the subset of vehicle trajectories can be adjusted to represent all the vehicles. The proportion of the known vehicle trajectories can be found dividing the total number of vehicles that cross the detectors in every time interval with

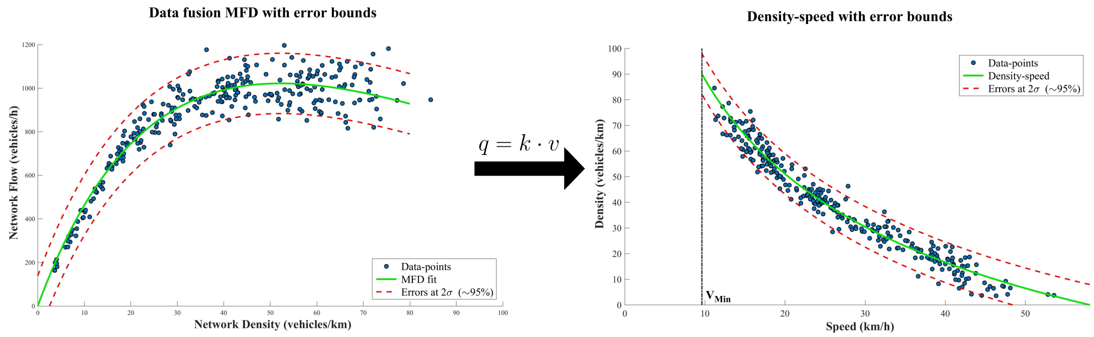


Figure 1: Traffic State Estimation Process with the MFD

the number of vehicle trajectories that traverse the links with the respective detectors. Then, the average network density and flow can be calculated to obtain the MFD.

The obtained MFD can then be combined with the fundamental relationship of the flow to calculate the relationship between density and speed for the network. In the density-speed relationship, each speed value corresponds to a unique density value. Taking this into advantage, speed measurements can be used to get the average network density and thus, the point on the MFD that the traffic network is performing. This process can be seen in Figure 1.

In order to validate the proposed traffic state estimation process with the MFD, it was applied to the simulated network of the area of Leidschendam-Voorburg in Paramics, which is a microscopic traffic simulation software. The MFD was obtained fusing subsets of vehicle trajectories with detector data for different runs of the simulation model with various demand levels. The obtained MFD was compared with the ground-truth MFD resulting from 100% vehicle trajectories. The results showed that the obtained MFD was very similar to the ground-truth MFD and even shared many common points. Afterwards, speed measurements using different fractions of vehicle trajectories were taken to derive the network density and the point on the MFD that the network is performing. As expected, the higher the fraction of the known vehicle trajectories was, the lower the error was and the derived traffic state was closer to the simulation situation, with the errors ranging from 7% to 26% when the fraction of the vehicle trajectories is 30% and 1%, respectively.

The effect that the errors of both the obtained MFD and the speed measurements have on the derived traffic state was investigated thoroughly with an extensive uncertainty analysis. The probability density function of the error of the obtained MFD was combined with the probability density function of the error of the speed measurements to produce the joint probability that a density value will occur in the network. For any given speed measurement, the density value is derived with an error bound at 95% level of certainty. The results showed that the error bounds are larger when the network is in the congested state compared to the free flow state. For instance, if the given speed is 45 km/h, the derived network density is 11 ± 9 vehicles/km, which corresponds to 323 ± 263 vehicles in the network at 95% confidence level. Whereas if the given speed is 15 km/h, the derived network density is 65 ± 17 vehicles/km, which in vehicle accumulation is 1941 ± 502 vehicles at 95% confidence level. This indicates that the densities derived from the proposed process are more accurate when the network is in the uncongested state, because there are more variations in the speed throughout the network when it is congested. Nevertheless, in order to express whether the estimated traffic state is accurate enough, the purpose that it will be used for is necessary.

Furthermore, the proposed method to derive the network-wide traffic state was tested in the case that an incident occurred in the network. Incidents of the nature of blocking one lane for one hour were simulated in seven network spots with different road types. The results showed that when

an incident occurred in roads with low demand, the proposed method was still able to predict the traffic state. However, in some cases that the incident was placed in spots with higher demand, the derived traffic state was out of bound. In these cases, it was observed that the average network speed drop after five minutes was remarkably high and over 20%. This indicates that the speed drop can potentially be used as a sign that an incident occurs in the network and the MFD cannot describe the traffic state accurately any more.

Concluding, this project managed to estimate the traffic state of a simulated network using the MFD and speed data. The results showed that the derived traffic state was within acceptable error bounds to describe accurately the traffic situation in the network. This means that the traffic state derived from the proposed process can be used as a solid and reliable base for traffic state prediction and traffic control strategies aiming at the optimization of the traffic system. Further research could focus on exploring how efficiently the proposed process can be used for these purposes. Furthermore, the application of the same traffic state estimation process in other networks could offer insights in the potential generalization of the method and the establishment of the MFD as a simple but powerful traffic state estimation tool.

Contents

List of Figures

List of Tables

1

Introduction

1.1. Motivation

As more and more people choose to reside in cities, the demand for urban mobility constantly incsreases and will continue to rapidly increase in the future. Serving this demand is very challenging, especially when it comes to road traffic, where high demand leads to serious congestion issues. According to Rijkswaterstaat (2007), 44 million hours are lost every year in the Netherlands, due to congestion that corresponds to a monetary loss of 700 million euros (as cited in Knoop (2009)).

To alleviate traffic congestion, new roads are constructed, extra lanes are added or traffic control is improved. Nevertheless, the construction of new roads and lane additions are not always the most effective choices, because they lead to higher vehicle usage, meaning that congestion will appear again sooner or later. Hence, the solutions should focus on ways that optimize the use of the existing network infrastructure. Such solutions can be the use of clever and cost-efficient traffic control strategies that manage traffic in a smart way and can thus, improve the operation of the road network and reduce congestion substantially.

So far, traffic control is mainly performed at a local level at the urban intersections. However, research has shown that applying traffic control at a network level or at parts of a network, e.g. zones can be more effective compared to the local level (Taale, 2008). Macroscopic control of cities can lead to a more efficient operating urban network. This is due to the fact that traffic control at one intersection can influence the traffic state at another intersection because of spillback effects.

Looking at the general scheme of the traffic control system at Figure 1.1, it can be seen that traffic control strategies can be applied to the traffic system through different kinds of control actuators. Their effect can be examined from the outputs produced by the traffic sensors. The sensors can vary and can provide different types of data, like loop detector data, floating car data, Bluetooth, GPS, camera data. These data need to be translated into useful information to determine control actions necessary for the actuators.

Subsequently, the key to having a successful control strategy is to translate the traffic sensors data correctly. The crucial process needed for this is to have a reliable and accurate estimation of the traffic state that fully describes the traffic situation. The estimated traffic state can be used to predict the traffic state in the next time step. The appropriate control action can then be chosen and optimized based on the expected traffic state. Thus, the critical step to have an efficient traffic system is the accurate traffic state estimation that acts as the solid base for the traffic state prediction and the control optimization.

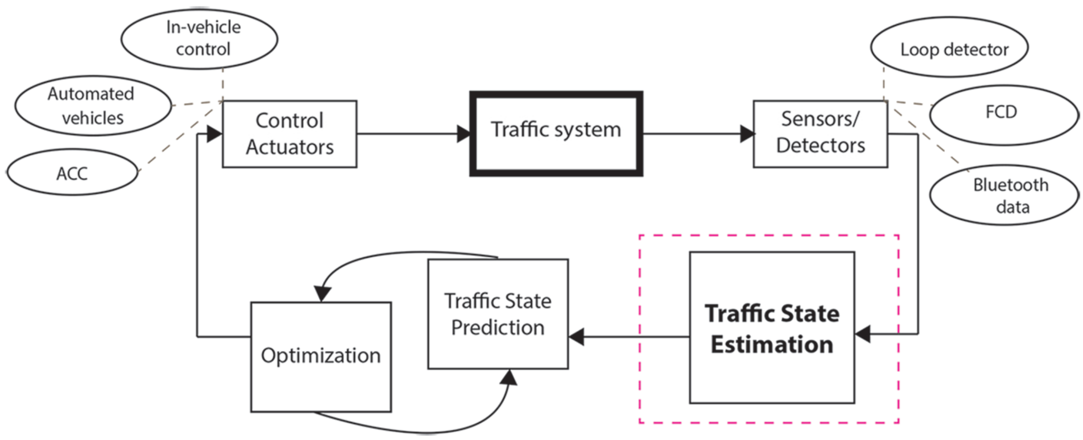


Figure 1.1: General traffic control system (inspired by Van Lint (2014))

Nevertheless, the macroscopic traffic control requires the collection and process of a lot of data to derive the required information to estimate the prevailing traffic state in the entire network. So far, some approaches that have been proposed by literature suggest the use of kinematic wave theory, cell transmission models or macroscopic traffic simulation models to describe the traffic state for a network. However, most of these approaches have limited applicability or have very high demands in computational resources and time.

In this perspective, the motivation for this research project originates from the need to find alternative ways to provide the information for macroscopic traffic control and estimate the traffic state of a network. The research interest within this subject has turned towards the examination of simple and fast ways to describe adequately the traffic situation in a road network. As a result, the development of the so-called Macroscopic Fundamental Diagram (MFD) was accomplished (Daganzo (2007), Geroliminis and Daganzo (2008)). The MFD relates the network flow to the network density in a single function and its knowledge can fully capture the prevailing network situation of the urban network at all times.

1.2. Research Objective and Questions

In this context, the goal of this master thesis is to solve the problem of knowing the traffic situation network-wide and to offer insights into the use of the Macroscopic Fundamental Diagram (MFD) for this purpose. Thus, the objective of this project is: *to estimate the traffic state of urban networks using the MFD*. In order to reach this objective, the main research question is formed as:

Main Research Question: *How to derive a network-wide traffic state estimation using the Macroscopic Fundamental Diagram?*

In order to answer the main research question, four sub-questions related to it are given. The subquestions aim to provide a simplified and step-wise approach on this complex problem. Starting with the construction of the MFD, it needs to be noted here that, although the only required variables are the network flow and the network density, they are not always easy to estimate. This is mainly due to data limitations, but also because of inconsistencies between the different data sources. Therefore, systematic investigation of the necessary requirements of traffic data to obtain the MFD is going to be performed in this research project. Within this framework, the first subquestion is formed as:

Sub-question 1: *What are the types and the amount of traffic data necessary to obtain the Macroscopic Fundamental Diagram?*

If the MFD of a network is known, all that is needed to have an accurate traffic state estimation is to know on which point of the MFD the traffic network is performing at any desired moment. Consequently, the second main concern of this research are the data requirements to obtain the traffic state on the MFD. Thus, the second sub-question is as follows:

Sub-question 2: *What are the types and the amount of traffic data necessary to know the traffic state that the network is performing at on the Macroscopic Fundamental Diagram?*

Nevertheless, traffic data are always erroneous, causing uncertainty on the confidence of the estimations. Hence, the certainty of the derived traffic state using the MFD needs to be thoroughly examined, in order to provide a complete depiction of the acquired result. So, the third sub-question is:

Sub-question 3: *What is the effect of the uncertainty that the obtained Macroscopic Fundamental Diagram encompasses on the derived traffic state?*

An additional point of interest is to examine the accuracy of the traffic state estimation with the MFD, not only in regular traffic conditions, but also in special cases. Such cases could be any road incident such an accident, a public event or harsh weather conditions. In this perspective, the fourth and last sub-question that is formed is:

Sub-question 4: *How is it possible to know if the Macroscopic Fundamental Diagram is not valid in the case of an incident?*

The four sub-questions that are formed contribute to fully answer the main research question and hence, fulfil the objective of this research project. The next section will discuss how to approach the answer of each one of the questions.

1.3. Research Approach

In order to set a solid base for the analysis that is performed throughout the project, a description of the MFD theory is necessary in the beginning. Furthermore, studies that explored the estimation of the MFD and the traffic state in urban context are analysed thoroughly in Chapter 2. This offers a good overview of proposed methodologies to be followed and how these could be improved or developed. The goal is to derive useful insights in the potential applications of the MFD, the data used by other studies to obtain it and the factors that have an effect on its construction.

After investigating the findings of recent literature on the MFD, the necessary traffic data to obtain the MFD are examined. The use of a data fusion technique is incorporated to acquire the MFD taking advantage of the combination of the available traffic data from loop detectors and floating car data. This leads to the proposal of a data fusion approach to estimate the traffic state using the MFD in Chapter 3. The proposed process includes two steps that provide the traffic state estimation of an urban network. In the first step, data points of network flow and density are obtained and the function connecting them, thus the function of the MFD, is found. In the second step, speed data are used to indicate the traffic density on the obtained MFD. A thorough description of the steps of the traffic state estimation and remarks on how the process can be applied are made within this chapter. In this way, the two first research sub-questions are answered.

In order to validate the 2-step traffic state estimation process, a test application is performed in Chapter 4 on the microsimulation network of Leidschendam-Voorburg in Paramics, which is a microscopic traffic simulation software. The results from the application of the traffic state estimation

of the simulated network are presented in Chapter 5. The goal of this chapter is to showcase the advantages of the proposed process. Moreover, a detailed description is given on how the necessary variables are calculated and how a formula fitting the MFD points can be found.

As mentioned, traffic data have errors and noise which needs to be taken into consideration when using them to estimate the traffic state. For this reason, the results from the test application of Chapter 5 are used next, to analyse the uncertainty that the estimation encompasses. In order to examine the uncertainty of the proposed traffic state estimation, a detailed error analysis is performed in Chapter 6. The third research sub-question is answered within the content of this chapter. The parts that constitute the uncertainty of the final estimation are discussed and the way of combining them is presented.

Last, the traffic state estimation is tested in the extreme case of an incident occurring in the simulated network. The purpose is to assess the validity of the traffic state estimation in irregular conditions. Road incidents are simulated in Chapter 7 to test how the estimation process works in such cases. The answer to the fourth and last research sub-question is given within this chapter.

After answering all four sub-questions, Chapter 8 presents a critical discussion of the steps that were performed during the thesis. Concluding, after the results have been finalized and the research questions have been answered, the conclusions of the research project are drawn in Chapter 9. Moreover, recommendations are made in this chapter both for practical work and for further scientific research.

1.4. Thesis outline

A short description of the outline of the research project is given collectively in this section. The approach that was made to meet the research objective of this thesis forms the following chapters with the respective content:

1. **Introduction**
   Research motivation leading to the research objective and questions.
2. **Theoretical Background and Literature Review**
   Description of the MFD theory and the studies about it.
3. **Traffic State Estimation Process**
   Proposal of a 2-step traffic state estimation process.
4. **Application of the Traffic State Estimation**
   Validation of the 2-step traffic state estimation by applying it to a microsimulation network.
5. **Results of the Traffic State Estimation**
   Resulting traffic state estimation of the microsimulation network.
6. **Uncertainty of the Traffic State Estimation**
   Detailed error analysis of the estimated traffic state.
7. **Traffic State Estimation in the case of an Incident**
   Simulation of road incidents to indicate the invalidity of the traffic state.
8. **Discussion**
   Critical assessment of important aspects of the thesis.
9. **Conclusions and recommendations**
   Final results and concluding remarks followed by future recommendations.

2 Theoretical Background and Literature Review

As mentioned, the main objective of this thesis project is to estimate the traffic state of urban networks using the MFD. Thus, the research focus of this work is on investigating the necessary data to obtain the MFD and to use it as a traffic state estimation tool. In this chapter, the basic theory of the MFD and previous studies that explored it are discussed. The goal is to offer an overview of the methodologies that have been followed in the studies and analyse their results. More specifically, the topics that we aim to address with the literature review are:

1. What is the range of the possible applications of the MFD supporting the allegation that it can be a very simple but very useful tool?
2. What are the data that previous studies used for the MFD and is it possible to combine different data types to obtain it?
3. What are the factors that determine the construction of a well-defined MFD?

In Section 2.1, the main theory of the MFD is presented. Next, in Section 2.2, the various ways that the MFD has been used are presented aiming to answer the first question. Section 2.3 investigates the data that other studies have used to produce the MFD, as mentioned in the second question. Section 2.4 looks into the factors that affect the shape of the MFD in order to answer the third question. Last but not least, Section 2.5 presents an overview of the answers to the questions under investigation and discusses how the literature study is useful for the current project.

2.1. Basic idea of the MFD

The main idea of the Network or Macroscopic Fundamental Diagram (MFD) is that it can describe traffic at an aggregated level. In a similar way that the Fundamental Diagram (FD) relates the flow and the density at a link or a road section, the MFD extends this relationship at an urban area or a network. More specifically, the MFD relates the number of the vehicles in the network, which is called accumulation, to the outflow from the network, which is expressed by the production (vehicles/hour). In other words, the MFD is a function between the average network density (vehicles/km) and the average network flow (vehicles/hour).

As it can be seen in the MFD in Figure 2.1, three different traffic states can be distinguished, like in the FD. When the network density is low, then the flow is also low and the network is in the free flow

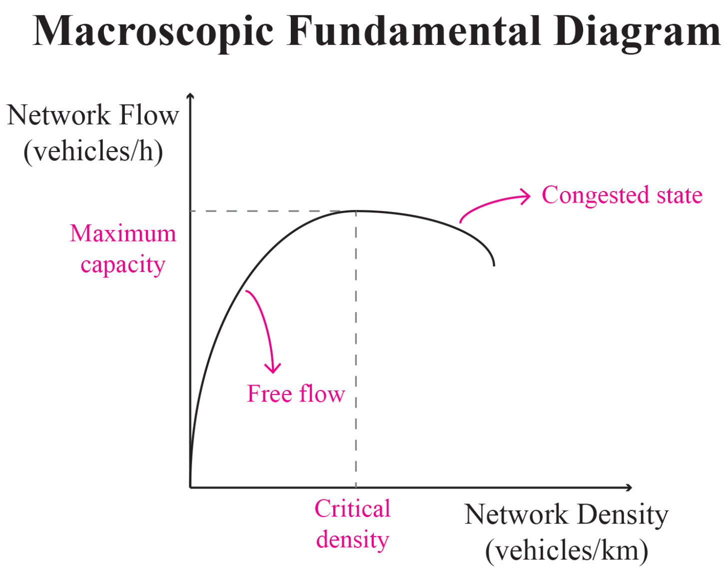


Figure 2.1: The Macroscopic Fundamental Diagram

state. As the number of the vehicles that are using the network increases, the network flow reaches the maximum capacity of the network. When this happens and the network density exceeds the critical density, the flow starts to decrease. This happens because there are more vehicles in the network that it can handle. At this point, the network enters the congested state and the drivers experience delays. As seen, the MFD is not drawn up to the jam density ($k_{jam} = k_{max}$) as in the case of the FD. That is because there will always be some links in the network that the vehicles can still move at least at a low speed.

The efforts to obtain a network-wide traffic relationship had started already from the '60s (Smeed (1966), Godfrey (1969), Zahavi (1972)) but it was not until 2007 that Daganzo (2007) and later in 2008 that Geroliminis and Daganzo (2008) found a well-defined relationship between flow and density and showed that the MFD can indeed exist in urban networks. Since then, different applications of the MFD have been proposed showing the extent of its utility. Furthermore, research has been conducted on how to derive the MFD, what data are needed to acquire a graph without scatter and what are the factors that influence its shape. All these points will be further analysed in the following sections.

2.2. Applications

In this section, the first question about the range of the possible applications of the MFD that was set in the content of the literature review is answered. The MFD can have various applications aiming to improve mobility in urban networks. It can act as a very useful tool for traffic managers to monitor their system and assess if it is operating at the desired level. The MFD can also be used to provide the necessary information for efficient traffic control. In Keyvan-Ekbatani et al. (2012), it is stated that although the MFD is still under research, the knowledge that we have on it so far, allows us to say that it can be used as a very good and reliable base for control strategies. Research is ongoing towards exploring the vast field of the useful MFD applications.

In Geroliminis and Daganzo (2008), the outflow rate provided by the MFD is proposed to be used as a way to evaluate the city's accessibility and determine how it can be improved with certain measures. If a new control strategy is implemented or an infrastructure change occurs, the MFD will most

probably change and the maximum production will increase or decrease showing if the measure was successful or not. Of course the evaluation of the measurements could also take place only in the area that the measure was applied but the effect could be falsely overestimated due to the narrow exploration of the results. For example, the implementation of a new traffic control algorithm at one intersection could improve the flow at the intersection, but maybe that is only because drivers choose another intersection now. This false assessment could have been avoided by examining the results in the complete network.

Daganzo (2007) proposed a solution to decrease congestion by developing an adaptive perimeter control mechanism that uses the MFD to monitor and control the total vehicle amount that enters a neighborhood. The idea was tested by Geroliminis and Daganzo (2007) at two simulated sites in Lincoln Avenue in Los Angeles and in Downtown San Fransisco. They tried to maximize the outflow by maintaining the total number of vehicles inside the network at the optimal level and restricting more vehicles to enter. The results showed that perimeter control can indeed work as expected, since the outflow increased by 34%. In this way, they also proved that the outflow and the travel production are linearly related.

Keyvan-Ekbatani et al. (2012) also analysed the application of a perimeter control strategy. More specifically, they proposed the use of the MFD to apply a "feedback-based gating". A test application of the gating strategy with simulation offers encouraging results with less delays and higher speeds in the simulated network. Geroliminis et al. (2013) proposed a perimeter control as well, but by applying a model predictive control solution. In order to test their method, they used various examples of two-region urban networks with different levels of congestion and different amounts of noise and errors in the data. The results are very interesting, showing that the model predictive algorithm performs much better than the "greedy" feedback control. These encouraging results of the perimeter control can be further used for the development of efficient traffic control strategies in any urban network.

Another example of application of the MFD is a congestion pricing strategy that was proposed by Geroliminis and Levinson (2009). Their research utilizes the MFD to sketch a network-based congestion pricing scheme. It was tested in the same site that was used by Geroliminis and Daganzo (2008) at Yokohama, Japan for the morning peak hour. Their results showed that the toll-case works very well compared to the no-toll case. With the application of the optimal toll price, delays were eliminated and the duration of the peak hour decreased. Simoni et al. (2015) also proposed a methodology to derive time-dependent toll prices using the Network Fundamental Diagram. They tested their methodology in a simulated case study of the city of Zurich. They suggest that this approach is more realistic than the analytic methods that are usually applied to decide tolling schemes. Furthermore, the proposed approach needs only a few information and offers a very good representation of the traffic dynamics. Thus, it is believed that it could be implemented in real occasions very soon.

Instead of applying a limit to the inflow in the network as it was proposed by the preceding gating and pricing strategies, the MFD could also be used to implement a routing strategy to spread the vehicles over the network. Such a routing strategy is proposed by Knoop et al. (2012). They suggested to route traffic using the MFD in a way that oversaturation is avoided and indeed their results showed an improvement in traffic flow. Another paper that proposed a routing strategy using the MFD is by Yildirimoglu et al. (2015). They proposed a "route guidance advisory control system" and their results showed that their method can produce a system optimum state for the network. Although homogeneity conditions play an important role in the routing strategies and further research is required on this aspect in the future, the results are quite promising towards improving network efficiency.

All in all, it can be seen that various studies have already investigated the advantages of applying the MFD to support new traffic control strategies or to assess existing strategies. These findings strengthen the allegation that although the MFD is simple and parsimonious, it can be very useful and powerful. We could actually say that its simplicity is one of its strongest points and this motivates us to research more about it.

2.3. Data and Methodology

This section aims to answer the second question about the necessary data to produce a well-defined MFD. The basic requirement is to have sufficient traffic data to estimate the traffic variables of network density and network flow. Most of the research performed so far has used the following terms to refer to these two variables:
Accumulation A: the number of vehicles in the network (unit: vehicles or vehicles/km)
Production P: internal flow in the network (unit: vehicles/hour)

The MFDs that have been produced so far have been based either on traffic data from the real world or on simulation data. Instead of using traffic data, efforts have been made by Daganzo and Geroliminis (2008) and later by Geroliminis and Boyacı (2012) to derive an analytical method to derive the MFD so that the need for large amounts of data is avoided. Their goal is to produce a general model that can be used to acquire the MFD for any network. Nevertheless, they acknowledged that networks are affected by multiple and complicated variables, so they tried to derive an estimation method with as few parameters as possible. Their proposed method produces an upper bound of the average flow in the network if it complies with some regularity conditions. Leclercq and Geroliminis (2013) further extended and improved this methodology by relaxing some of the regularity conditions. However, the proposed methodology still has the disadvantage that it requires homogeneous congestion levels in the network.

Finding a methodology to acquire the MFD that does not require large amounts of data is a challenging task. So far, most of the researchers that explored the MFD used simulation data, due to the difficulty to collect all the necessary data to enhance the complex traffic dynamics that are present in urban networks. A proposed solution could be the implementation of data fusion techniques. By aggregating and fusing different types of data, improved information can be obtained for the traffic state that can be used to accurately estimate the MFD of the urban network.

Next, the most noteworthy papers that used either real or simulation data to obtain the MFD and the basics of their methodology will be described. Moreover, papers that have pursued to follow a data fusion direction will also be described to show the advantages of this approach and the reason that it is recommended for this project.

2.3.1. Using real traffic data

Geroliminis and Daganzo (2008) were the first to derive a smooth network fundamental diagram from real data. For the purposes of their study, they used loop detector data from downtown Yokohama in Japan. They realized that the flow-occupancy relationship produced by the data from individual detectors contained a lot of scatter, as seen in Figure 2.2. However, when aggregating the detector data, a smooth relationship resulted as it can be seen in Figure 2.3.

Because the detectors did not cover the entire network, they tried to produce the same relationship using GPS taxi-data that had a full network coverage. From these data, they created the relationship between space-mean speed and density. Indeed, a smooth graph was again produced indicating that a MFD can exist for the entire urban network and that it is a characteristic of the network and

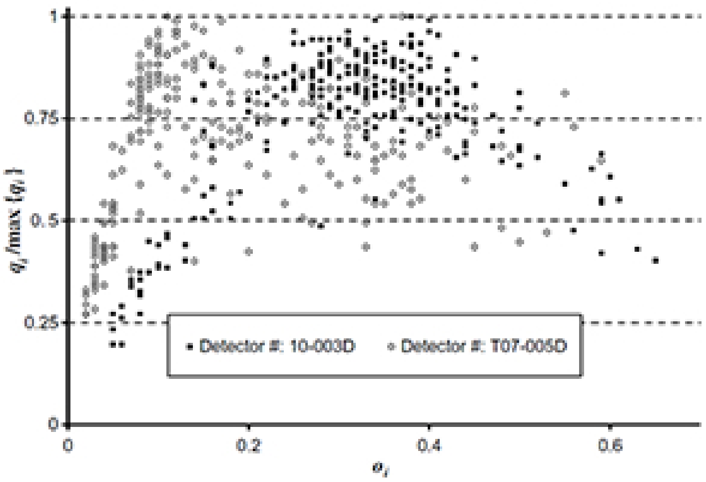


Figure 2.2: Graph of flow vs. occupancy for two individual detectors of Yokohama (Geroliminis and Daganzo, 2008)

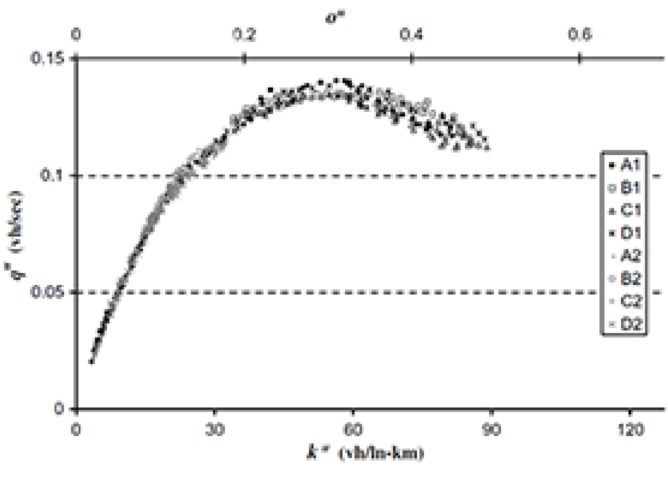


Figure 2.3: The Macroscopic Fundamental Diagram after aggregating the detector data of Yokohama (Geroliminis and Daganzo, 2008)

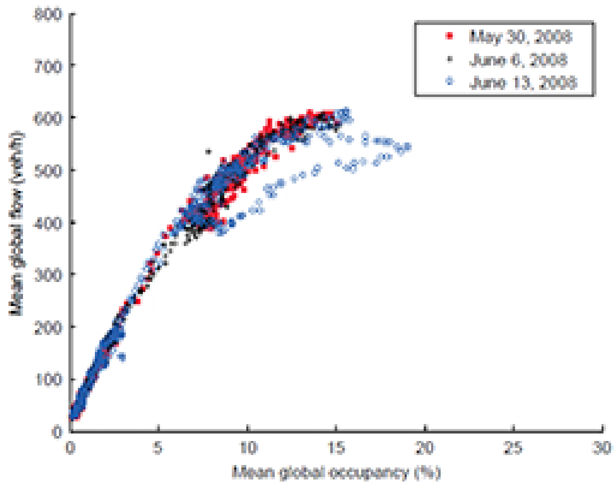


Figure 2.4: The Macroscopic Fundamental Diagram from Buisson and Ladier (2009)

not of the detectors. The formulas that they used for traffic flow q, occupancy o and density k derived from the generalized definitions of Edie (1963). They used either the weighted averages:

$$q^w = \frac{\Sigma_i q_i l_i}{\Sigma_i l_i} \qquad \text{and} \qquad o^w = k^w s = \frac{\Sigma_i o_i l_i}{\Sigma_i l_i}$$

or the unweighted averages:

$$q^u = \frac{\Sigma_i q_i}{\Sigma_i n} \qquad \text{and} \qquad o^u = k^u s = \frac{\Sigma_i o_i}{\Sigma_i n}$$

In these formulas, i is a road lane segment, l_i is its length and n is the number of lanes on the link. These formulas have been mainly used also in subsequent articles with small alterations.

Buisson and Ladier (2009) used the unweighted mean formulas with loop detector data from both the highway and the surface road network in order to create a MFD in Toulouse, France (Figure 2.4). They observed the evolution of flow and occupancy throughout the chosen days of the analysis with the goal to see what happens if heterogeneity applies in the network. More specifically, they examined the homogeneity issue in concern to the different road types, the distance of the loop detectors from the stop line and the congestion levels throughout the network and investigated how each one of these parameters influence the scatter on the MFD. Their results on the factors influencing the shape of the MFD are more thoroughly analysed in Section 2.4.

Another study that used real data to estimate the MFD is by Cassidy et al. (2011). In this study they used detailed vehicle trajectories data from freeway stretches. Their results showed that the MFD can be produced only if the network is either in the congested or in the uncongested state and not in a mixed state. They also discovered that the MFDs for the freeways can be produced also by loop detector data as long as there is at least one detector per link and the data are filtered so as to meet the single regime condition.

2.3.2. Using traffic simulation data

Geroliminis and Daganzo (2007) performed various microsimulations in downtown San Fransisco and were the first to produce a diagram relating the production to the accumulation of vehicles with the aim of applying macroscopic feedback control strategies. They used Edie's definitions (Edie, 1963) to estimate mean flow and mean density.

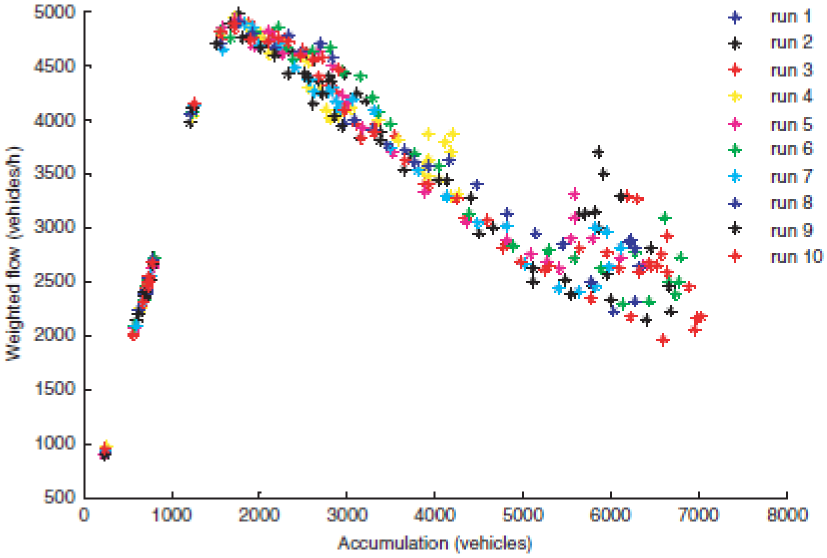


Figure 2.5: The Macroscopic Fundamental Diagram with simulation data of Amsterdam from Ji et al. (2010)

Ji et al. (2010) also used microsimulation data for the city of Amsterdam because the collected traffic data were not adequate to describe the traffic flow. Their resulting MFD can be seen in Figure 2.5. The goal of their research was to examine the influencing factors of the shape of the MFD. Their results on this interesting matter are further referenced in Section 2.4. To calculate the flow, they used the formula of the weighted average flow q^w as in Geroliminis and Daganzo (2008) and for the accumulation n_i they used the following:

$$n_i = \Sigma_i k_i l_i$$

In Courbon and Leclercq (2011) they also used a microsimulation model. The goal of their research is to compare the results of three different ways to produce the MFD. For this reason, they want to avoid the bias resulting from the use of empirical data and have complete control over the urban environment and the traffic phenomena occurring at it. The three different approaches that they are investigating are the following:

1. Using loop detector data and the weighted averages as in Geroliminis and Daganzo (2008).
2. The analytical method with cuts proposed by Daganzo and Geroliminis (2008).
3. Using data from vehicle trajectories.

They used a very simple network with road sections of similar length and the same traffic light cycle. Their results showed that the trajectory method can produce the MFD very accurately in all network shapes and thus, they suggest that this method can be used as a base to evaluate and compare other methods.

Keyvan-Ekbatani et al. (2012) used simulation data of the city of Chania, Greece to produce the MFD and test their proposed gating measures to improve mobility. They state that a MFD can be either "ideal" when it includes the precise traffic data of all the network links, so it can only be derived from simulation environments, or "operational" when it includes the available traffic data from a subset

of the network links. An "operational" MFD can be "complete" if the available traffic data cover the entire set of network links. For their test, they produced a "complete operational" MFD which in combination with a moderate amount of real-time measurements can describe the traffic situation and determine the appropriate gating strategy. For their case, they used all of the measurements of the links to detect the point on the MFD that their network was operating. However, they suggest that this is not needed and that the information required for gating can be extracted by using less real-time measurements.

Another interesting study with simulation data is by Ortigosa et al. (2014). They used a microsimulation model of the inner city of Zurich. They only used loop detector data, because they support that floating car data are not still broadly available and they aspire that their methodology could be implemented in any urban areas. The purpose of their study was to investigate the required number of links to obtain the MFD. In order to achieve that, they created the complete MFD from all the links and incomplete MFDs with only some of the links. Then, they compared the incomplete MFDs to the complete MFD based on the density ratios and evaluated their accuracy. In order to estimate the average flow and density the weighted averages proposed by Geroliminis and Daganzo (2008) were used again in this study. Their results showed that a network coverage of minimum 25% provides sufficient accuracy with small error at the density ratios.

2.3.3. Fusing data

Gayah and Dixit (2013) propose an indirect traffic state estimation combining data from mobile probes with the MFD. Their goal is to derive an accurate real-time estimation of the network density which can be used to support network-wide traffic control strategies. In order to test their methodology, they used a simulated network of the city of Orlando. Their methodology takes advantage of the fact that each point on the MFD is related to a unique speed value. Thus, knowing the average speed from the probe vehicles they can then see on the MFD to what density that speed corresponds. However promising their methodology seems to be, it had the issue of being able to provide the traffic density only in congested conditions and it cannot be used if the MFD is not known.

Nagle and Gayah (2014) suggested a methodology that overcomes these disadvantages. Their method proposed the use of the generalized definitions of Edie (Edie, 1963) to estimate the network-wide variables from probe vehicle data. However, in order to apply these formulas, the data of all the vehicle trajectories are necessary, but usually only a small number of vehicles serves as probes. They suggested that they can overcome this difficulty as long as the ratio of the probe vehicles is known. In order to acquire the ratio, they proposed dividing the number of vehicles that were tracked by GPS in the analysis area for a specific time period to the number of vehicles that crossed the detectors in the same area and period. Microsimulation tests of their methodology showed that a 20% probe penetration rate can provide accurate estimations for any traffic state, making this a robust methodology to acquire the MFD. Nevertheless, this study has some limitations since it assumes that the probe vehicles are uniformly distributed across the network which is not always realistic.

Leclercq et al. (2014) also combined data from probe vehicles with loop detector data but in a different way. From the probe vehicles, they derived the average network speed and from the loop detectors, the average network flow. Their results showed that a 20% probe penetration rate can significantly improve the estimation of the network speed for the MFD. However, this study has the limitation that the dynamics of the network are not captured completely, since the loop detectors are not placed everywhere in the network.

In Du et al. (2015), they tried to overcome these limitations by proposing a MFD estimation method without the conditions that the probe penetration rate is homogeneous and the detectors are placed

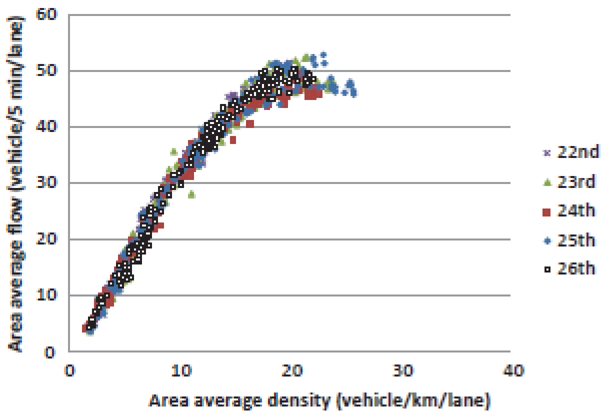


Figure 2.6: The Macroscopic Fundamental Diagram with real data of Brisbane from Tsubota et al. (2014)

in all links. Their approach suggested to estimate the appropriate average probe penetration rates from the weighted harmonic means with the weights being the travel times and the distances of the individual probe vehicles. Their methodology was tested with microsimulation and the results showed that it is a very effective approach. Du et al. (2015) suggest that since mobile probe data are becoming more and more available, this methodology can be applied also with real data.

All of the studies described before attempted a data fusion approach using simulation data. Tsubota et al. (2014) is one of the few studies that fused real data for the city of Brisbane, Australia. Their approach was to estimate the network density combining loop detector data, traffic signal timings and probe vehicle data. The method that they used is the so called "Cumulative Plots and Probe Integration for Travel Time Estimation (CUPRITE)". This method has been created to estimate the travel time on links but it can also estimate the link density. Their resulting MFD can be seen in Figure 2.6. As it can be seen, the observations are mainly situated in the free flow branch. In order to explore better the characteristics of their system, they also divided the network in four parts and produced one MFD for each part. In this case, the MFDs had obvious differences in their shape. This indicates that network partitioning can help researchers to understand their network and determine the appropriate size of the zones for the application of traffic control measures.

2.4. Influencing factors

After analysing the data requirements to acquire the MFD, the third question about the parameters that affect the shape of the MFD will be answered within this section. The factors that have impact on the shape of the MFD have been the topic of interest in a few studies. Their findings suggest that the main parameters that influence the MFD are: the homogeneity levels of congestion, the road types of the network, the location of the loop detectors and the characteristics of the traffic lights cycle. Each of these factors and its effect on the estimation of the MFD will be further described in this section.

2.4.1. Homogeneity of congestion

One of the most important factors influencing the shape of the MFD and that has been studied more is the homogeneity of congestion over the network. Daganzo and Geroliminis (2008) mention that "universal MFDs should not be expected" in cases that congestion is not distributed in a consistent

way over the urban network. This can occur when different road types are present in the network or when there are more than one highly-congested centers. A thorough research on this allegation and the impact of homogeneity at the shape of the MFD was further performed by Buisson and Ladier (2009). They found that indeed, homogeneity at the congestion level of the network is necessary to have a MFD. Heterogeneous increase of congestion in space led to hysteresis effects on the MFD and so they concluded that "the congestion spreading must be homogeneous". Another paper that supported the same allegation was from Ji et al. (2010) using simulation data. They found that when congestion starts and resolves inhomogeneously a reduction in flow occurs. They suggested a solution to this problem by using ramp metering to create homogeneous traffic states in the network.

According to the previous references, in order to have a MFD, all links in the network need to have similar density, so similar traffic state. Nevertheless, this is a quite strict precondition. For this reason, first, Mazloumian et al. (2010) and then, Geroliminis and Sun (2011) tried to explore further the homogeneity assumption. Mazloumian et al. (2010), using simulation data, found that the spatial variation of the density plays a highly important role in determining the network performance. Geroliminis and Sun (2011) further supported this finding, using the real data from Yokohama, Japan and from the Twin Cities in Minnesota, USA. Their results showed that it is not absolutely necessary to have homogeneous traffic states in the entire network in order to have a well-defined MFD. As long as the spatial distribution of congestion was similar for two different time intervals that had the same average network occupancy, a MFD without scatter formed. This means that they managed to relax the strict need of homogeneous congestion in all the network to the condition that spatial variability of density should be similar for times when average network occupancy is about the same. Following this finding, Ji and Geroliminis (2012) investigated a method to divide the network in homogeneous zones with low variation in density so that perimeter control strategies could be applied in them.

Knoop et al. (2015) further supported the finding that the variation of the density is highly important and suggested a more reliable way to estimate the MFD by taking it into consideration. The way to achieve this is by including besides the average density, the standard deviation of the network density in the estimations. They called the resulting graph the Generalized Macroscopic Fundamental Diagram. Their approach is very interesting because by relating the production to both the network accumulation and the standard deviation of density, the hysteresis effect was also included and so, traffic flow was described better.

2.4.2. Road types

Geroliminis and Daganzo (2008) separated the signalized and non-signalized roads of Yokohama in order to derive a MFD without scatter and they did not consider at all the highways. They did this because they consider that homogeneous road types need to be used for an accurate MFD. The same conclusion was drawn from Buisson and Ladier (2009) and they recommended that the city should be divided into zones taking into consideration geographic characteristics and the road types. Within zones with the same road type, they recommended that is also important to separate signalized and non-signalized roads to avoid scatter in the MFD.

In Geroliminis and Sun (2011), they found that the MFD does not hold for freeway networks due to their different characteristics such as the absence of traffic signals, hysteresis effect and different speed levels. This finding is based on a real data experiment that they conducted in Twin Cities, Minnesota. In contrast, Cassidy et al. (2011) suggested that the MFD can be also produced in freeways. This is true for any freeway network as long as all the links are either in the congested or the uncongested state.

2.4.3. Loop detectors location

Buisson and Ladier (2009) separated the detector data in three categories according to the distance of the detector from the traffic light. They found that this distance has a strong impact on the slope of the MFD. Therefore, they recommended the use of detectors that have similar distance from the traffic signal. However, in Courbon and Leclercq (2011) they discovered that when the loop detectors have the exact same distance from the traffic signal, the results were not representative. Instead, they suggested that the detectors should be "spread upstream, downstream of the traffic signal and in the middle of the section" so that they cover all the possible traffic states.

2.4.4. Traffic light cycle

Laval and Castrillón (2015) investigated the effects of signal timing on the MFD. Especially, they focused on the length of the cycle, the ratio of green to red and the coordination. In order to perform their study, they used both simulation and the empirical data from Yokohama, Japan. Their results suggested that there are two parameters that play the most important role on the shape of the MFD: the mean ratio of the block length to green and the mean ratio of red to green light. De Jong et al. (2013) also investigated the effect of traffic lights on the MFD and found that when the signals settings change, then the shape of the MFD also becomes different. For this reason, they suggested to not use the MFD to determine the signal cycle for traffic control strategies.

2.5. Concluding remarks

Although the idea of the MFD is relatively new, the literature review showed that quite some studies have been performed about it. This indicates the important contribution that the MFD could have for traffic managers to improve urban mobility. The goal of this chapter was to answer the three questions about: the applications of the MFD, the data to obtain it and the factors that affect its shape. The answer to these questions provides useful information and can act as a solid foundation for the realization of this research project.

Concluding from the studies about its applications, it was found that the MFD can potentially monitor the traffic system and evaluate the application of traffic control strategies. More specifically, it can be used to support perimeter control mechanisms, congestion pricing schemes or routing strategies. The results of the studies that applied the MFD for such purposes are encouraging to develop it further so that it can become a broadly used traffic estimation tool.

Regarding the required data to obtain the MFD, it was shown that real data have limitations in producing the desirable result. Mostly, the problem is that there is not complete control on the congestion levels. Whereas using simulation data, different demand scenarios can be tested easily and fast compared to reality. Furthermore, the real data do not always have sufficient coverage of the complete network, so additional data are needed. Another problem is that loop detectors data have many errors and malfunctions that limit their usefulness.

The majority of the studies that produced MFDs used microsimulation data to estimate the variables of the average network flow and density to create them. It was concluded that the average network flow can be estimated accurately from the loop detectors. However, the average network density is not recommended to be estimated from the loop detectors data. Leclercq et al. (2014) clearly showed that the use of only loop detectors data should not be applied because the loop detectors cannot capture properly the spatial dynamics over the links. Trajectories produced from probe vehicles data can offer a much more accurate estimation of the network density or the network speed. Ideally, 100% knowledge of the trajectories would provide an exact MFD estimation. Nevertheless,

acknowledging the fact that it is not possible to have all the required trajectories data, this project suggests that the focus should be on fusing the existent data sources to the utmost.

The factors that were found to have significant impact on the shape of the MFD mainly regard the homogeneity of congestion and network characteristics, such as the road types. Additionally, an impact factor that needs to be emphasized is the location of the detectors in the network and how substantially it can affect the resulting variables for the MFD. These parameters will be taken into consideration when estimating the MFD in this project.

The literature review showed that the available literature cannot sufficiently address the research questions set for this thesis project. None of the studies done so far have managed to investigate systematically the types and the amount of traffic data necessary to estimate the MFD. Although some studies have used different types of data to confirm their findings, none of these have tested if data fusion can provide us with an efficient and robust methodology to estimate the MFD and the traffic state reliably. Thus, the goal of this project is to investigate these open matters and obtain the network-wide traffic state with the MFD by fusing different data types.

3

Traffic State Estimation Process

As mentioned in Section 1.2, the goal of this research project is to solve the problem of knowing the traffic situation network-wide. The literature review in Chapter 2 showed that the Macroscopic Fundamental Diagram (MFD), relating the flow to the density, is a simple and fast way to describe the traffic situation that can occur in a network. The studies on the MFD have so far been showing encouraging results to establish the MFD as a broadly used traffic state estimation tool. In this context, the objective of this project is to estimate the traffic state of urban networks using the MFD. Thus, the main research question focuses on determining how a network-wide traffic state estimation can be derived using the MFD. Within this chapter, the two first sub-questions that were formed to answer the main research question, will be answered:

1. What are the types and the amount of traffic data necessary to obtain the Macroscopic Fundamental Diagram?
2. What are the types and the amount of traffic data necessary to know the traffic state that the network is performing at on the Macroscopic Fundamental Diagram?

In order to accomplish the research objective, a traffic state estimation process is proposed. In this chapter, first, the main concept of the proposed network-wide traffic state estimation process is presented in Section 3.1. Next, a thorough description of the steps to perform the traffic state estimation is introduced in Section 3.2. Some concluding remarks and the next steps regarding the application of the traffic state estimation are presented in Section 3.3.

3.1. Main Concept

The basic idea of the proposed traffic state estimation derives from the fact that if the MFD of a network is known, all that is needed to have an accurate traffic state estimation is to know where we are on the MFD at any desired moment. As it was also suggested by Gayah and Dixit (2013), as long as the MFD is given, information about the network speed from any data source, can be used to estimate the network density and determine the point on the MFD that the network is performing. Hence, the main concept of the traffic state estimation process is based on this observation and includes two basic steps: first, to obtain the MFD and then, to use speed data to indicate the traffic state of the network on the MFD. This concept can also be seen schematically in Figure 3.1.

In order to apply the proposed concept, traffic data are required in order to obtain the MFD and the network speed. In the remaining of the section, firstly, the necessary traffic data will be discussed.

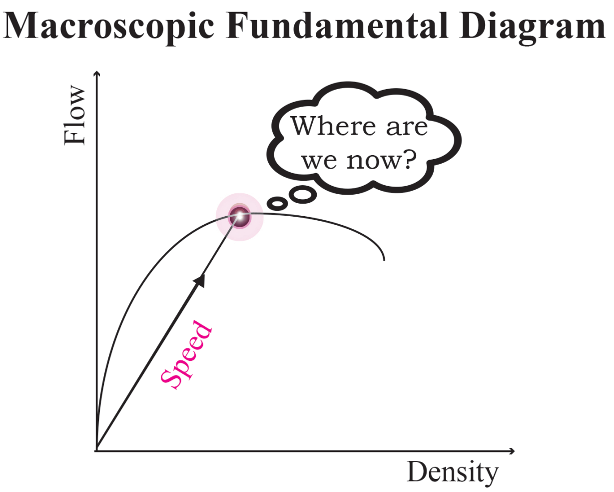


Figure 3.1: Conceptual Scheme of the Traffic State Estimation with the MFD

However, due to data limitations, the necessary traffic data are not always easily available. For this reason, a data fusion approach is investigated within this project and is presented next.

3.1.1. Need for Data

The construction of the MFD requires accurate and reliable data that cover the entire network area, so that they can represent the traffic situation in the network. The variables needed for this purpose are the network flow and the network density calculated for the selected time interval, usually defined at 5 minutes. Moreover, the variable of the average network speed during the time interval is needed in order to indicate the prevailing traffic state on the MFD.

In real traffic networks, the two more usually available traffic data types are:

1. Loop detector data collected from vehicle detection loops
2. Vehicle trajectories or else called floating car data (FCD) collected from GPS devices, bluetooth, etc.

Loop Detector Formulas

In the case that loop detector data are used to obtain the MFD, the following formulas of weighted flow q (vehicles/hour) and weighted density k (vehicles/km), as presented in Geroliminis and Daganzo (2008), can be used:

$$q = \frac{\Sigma_i q_i l_i}{\Sigma_i l_i} \tag{3.1}$$

$$k = \frac{\Sigma_i k_i l_i}{\Sigma_i l_i} \tag{3.2}$$

where

i: road lane segment

l_i: length of the road lane segment i

q_i: flow of the road lane segment i

k_i: density of the road lane segment i

Regarding the estimation of the network speed v (km/h) from loop detector data, both the arithmetic and the harmonic mean speed can be calculated. This is because, as mentioned in Daamen et al. (2014), the arithmetic mean speed can result to up to 25% overestimation of the speed in congested conditions, due to the larger impact of the faster vehicles. Consequently, the formulas used for the speed v in the case of the loop detector data are the following:

$$v_{arithmetic} = \frac{\Sigma_i v_{ij}}{\Sigma_i n_i} \tag{3.3}$$

$$v_{harmonic} = \frac{\Sigma_i n_i}{\Sigma_i \frac{1}{v_{ij}}} \tag{3.4}$$

where

v_{ij} : speed of vehicle j crossing road lane segment i

n_i : total number of vehicles crossing road lane segment i

Vehicle Trajectories Formulas

Regarding the case that vehicle trajectories are available to obtain the MFD, the formulas from Edie's definitions (Edie, 1963) can be used to extract the valuable information that the vehicle trajectories can offer. According to Edie's formulas, the average vehicle speed v (km/h), the average density k (vehicles/km), the average flow q (vehicles/hour) and the vehicle accumulation A (vehicles) in the network for the time period of the analysis, can be calculated as following:

$$v = \frac{\sum_i^n d_i}{\sum_i^n t_i} \tag{3.5}$$

$$k = \frac{\sum_i^n t_i}{LT} \tag{3.6}$$

$$q = \frac{\sum_i^n d_i}{LT} \tag{3.7}$$

$$A = kL \tag{3.8}$$

where

d_i : travel distance of vehicle i

t_i : travel time of vehicle i

n : total number of vehicles that use the network during the time interval

LT : time-space area (total network length $L \cdot$ time interval T)

As it also concluded in Leclercq et al. (2014), where they compared different methods to estimate the MFD, knowledge of 100% of the vehicle trajectories is the ideal way to estimate the MFD. The results

of Leclercq et al. (2014) showed that the trajectory method can produce the MFD very accurately in all network shapes and thus, they suggest that this method can be used as a base to evaluate and compare other methods. This means that the trajectories data can undoubtedly be used to produce the ground-truth MFD. Nevertheless, full information of 100% of the vehicle trajectories is an idealistic situation and can only occur in simulated traffic networks.

If only detector data are available to obtain the MFD, the network flow can be calculated sufficiently and accurately. However, the same does not apply for the vehicle speed and the network density. As the literature review has already shown, such as in Buisson and Ladier (2009) or in Courbon and Leclercq (2011), the location of the detectors can influence significantly the network speed estimation. When the detectors are near the stop line, most of the captured vehicle speeds are low and consequently, the situation further upstream of the traffic light is not taken into account. This issue creates doubts on the credibility of the detector data and hence, on the accuracy of the MFD obtained by the detector data.

3.1.2. Data fusion approach

In the case that the MFD estimation is based only on the known vehicle trajectories, the derived information would not be explanatory enough, because, as mentioned, the whole set of vehicle trajectories is not possible to be known in real networks. Consequently, some network areas would remain uncovered if none of the known vehicle trajectories traversed them.

Regarding the estimation exclusively from detector data, there is high uncertainty on the network density and the speed calculated from detectors. This is due to the impact that the detector's distance from the stop line has. When the detectors are very close to the traffic light, they capture only low vehicle speeds and they are not representative of the traffic situation more upstream. Nevertheless, loop detector data are available in many intersections in almost every traffic network, and the issue of the location of the detectors creates a disturbing situation on whether the detector data can be trusted or not.

On account of the insufficient availability of vehicle trajectories and the ambiguity of the estimations from loop detectors, a data fusion approach is explored in this project. The proposed approach takes advantage of the valuable information that also a low fraction of vehicle trajectories can offer and combines it with the - almost always available - loop detector data. The combination of these two data types can potentially provide a more accurate description than if only one of these data sources is used.

The way to fuse the two data types is to scale the variables of the flow and the density estimated from the subset of known vehicle trajectories to the whole set of vehicles. This can be achieved by dividing the flow and the density calculated from the trajectories' subset with the proportion a_{traj} of the known vehicle trajectories in relation to the total detector flows. In this way, the network variables that are needed to derive the MFD can be calculated. An extensive explanation of the data fusion process to obtain the MFD follows in the next section.

3.2. Steps of the process

As mentioned in the previous section, the low fraction of vehicle trajectories is not sufficient to describe the entire network dynamics and there is high uncertainty regarding the network density estimated from detector data. For these reasons, a data fusion process is recommended in this project to obtain the MFD. Succeeding the data fusion process to obtain the MFD, speed data can then be used to indicate the traffic state that the network is performing at on the obtained MFD. Hence, the

proposed traffic state estimation process can be described in the following two steps:

Step 1: Data fusion of vehicle trajectories and detector data to obtain the MFD
Step 2: Use of the obtained MFD and speed data to derive the network density

In more detail, each one of the steps is described in the next subsections.

3.2.1. Step 1: Data fusion to obtain the MFD

More analytically, starting with the 1st step, the main idea on how to fuse the data is based on the fact that the vehicle trajectories offer representative information, but only a subset of them is available. However, if the size of the subset is known, then the variables calculated from the subset can be scaled to the full set of vehicles. In other words, if the proportion a_{traj} of the known vehicle trajectories in relation to the total number of vehicles is calculated, the information from the subset of vehicle trajectories can be divided with the subset to represent the entire network traffic state.

The proportion of the known vehicle trajectories can be calculated by relating the total number of vehicles that cross the detectors in every time interval with the number of vehicle trajectories that traverse the links with the respective detectors. The average flow and the average density calculated by the subset of the known vehicle trajectories using Edie's definitions can then be divided with the proportion to represent the total set of vehicles. In an algorithm format, the proposed data fusion approach of the 1st step is:

1. From the detector data, calculate the total number of vehicles crossing the detectors in every time interval N_{det}.
2. From the subset of known vehicle trajectories, calculate the density k_{traj} (Equation 3.6), the flow q_{traj} (Equation 3.7) and the number of vehicle trajectories that traverse the links with detectors N_{traj}.
3. For each time interval, calculate the proportion of the known vehicle trajectories in relation to the total number of vehicles as:

$$a_{traj} = \frac{N_{traj}}{N_{det}} \tag{3.9}$$

4. For each time interval, calculate the average network density k_{net} and the average network flow q_{net}, as

$$k_{net} = \frac{k_{traj}}{a_{traj}} \tag{3.10}$$

$$q_{net} = \frac{q_{traj}}{a_{traj}} \tag{3.11}$$

5. Construct the MFD of the network finding the mathematical function that best fits the data points of k_{net} and q_{net}.

A schematic flow chart of the data fusion process is illustrated in the first part of Figure 3.2.

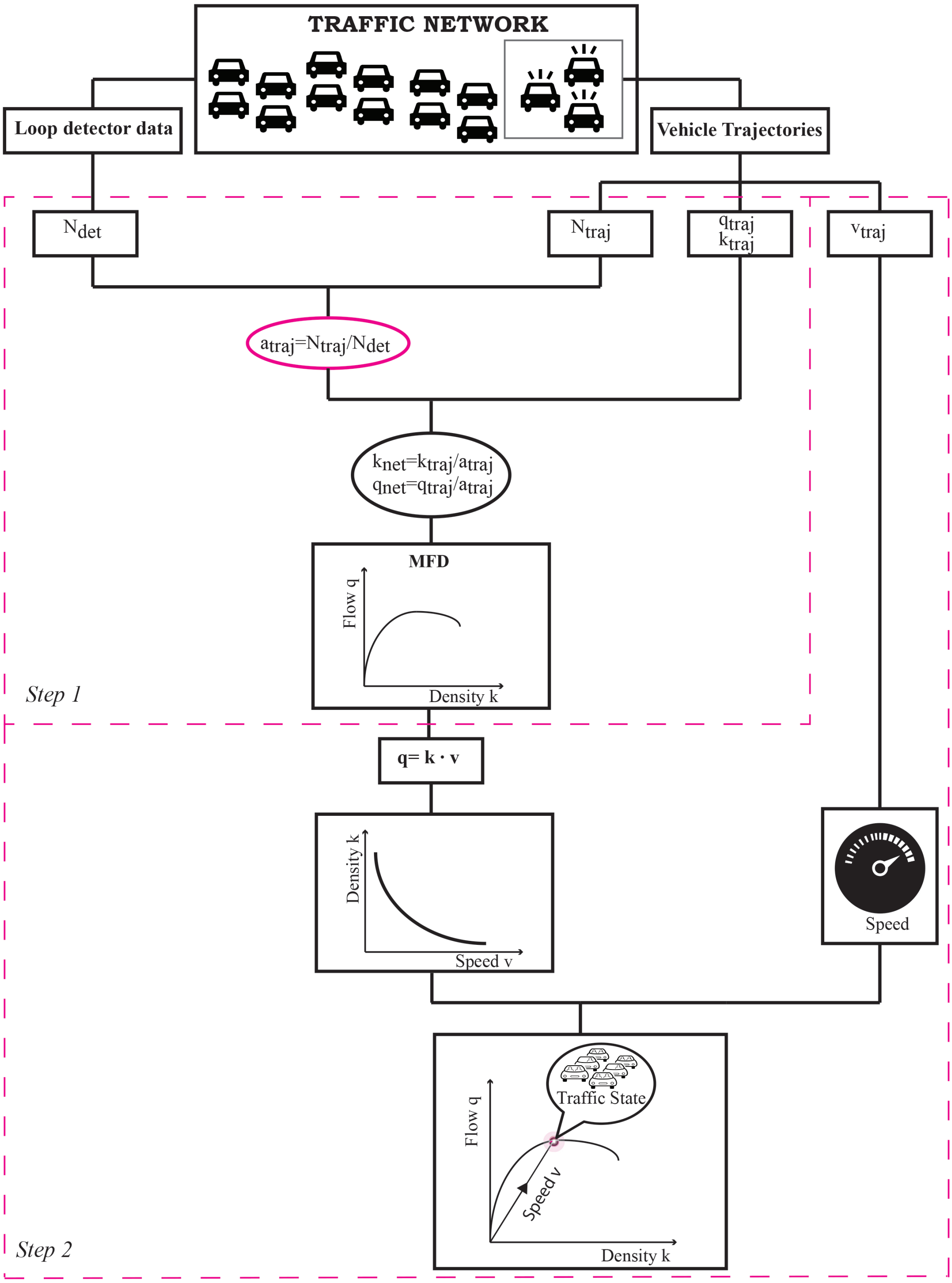


Figure 3.2: Schematic representation of the two steps of the traffic state estimation process

3.2.2. Step 2: Traffic state using the MFD and speed data

The MFD obtained from the 1st step can be used in the 2nd step to calculate the relationship between speed and density for the network. This can be done by applying the fundamental relationship of the flow $q = kv$. In the relationship between speed and density for the network, each speed value corresponds to a unique density value. Taking advantage of this, speed data from any data source can be used to derive the average network density and thus, the point on the MFD that the traffic network is. In an algorithm format, the proposed density estimation approach of the 2nd step is:

1. Given the obtained MFD formula, calculate the relationship between speed and density for the network from $q = kv$.
2. Given the relationship between speed and density for the network and speed measurements from any available data source, calculate the average network density.
3. Use the derived average network density to indicate the traffic state on the MFD.

The second part of the schematic flow chart in Figure 3.2 depicts the 2nd step of the process.

3.3. Concluding remarks and next steps

Concluding, the main concept that was proposed to estimate the traffic state includes 2 steps: first, data fusion of the known vehicle trajectories and detector data to obtain the MFD and second, use of the obtained MFD and speed data to derive the traffic state. With regard to the two research subquestions that we wanted to answer within this chapter, the main concept gives the answer to both of them. The required data to obtain the MFD are the available fraction of vehicle trajectories and the detector flows. The required data to indicate the prevailing traffic state on the MFD are speed measurements. Consequently, the data fusion approach that is suggested contributes to a low amount of required data. Moreover, the required data types are normally available in every traffic network, which makes the proposed process very easy and simple to apply.

In order to validate the applicability and the accuracy of the traffic state estimation process, a test application is necessary to compare the values of the estimated variables with the real values. The resulting data fusion MFD should be compared to the ground-truth MFD from 100% vehicle trajectories and the derived average network density should be compared to the real density. For these validation purposes, the traffic state estimation is applied to a simulated network in the following chapter. The characteristics of the simulated network and the collection of the data needed for the application are presented in Chapter 4. Then, in Chapter 5, the results of the application are shown and discussed.

4

Application of the Traffic State Estimation

In Chapter 3, we proposed a process to estimate the traffic state network-wide in two steps. In the 1st step, the MFD of urban networks is obtained by fusing the available subset of vehicle trajectories and detector data. Then, in the 2nd step, the obtained MFD is used in combination with speed measurements to derive the network density, and thus the traffic state that the network is performing on the MFD. In order to test the accuracy of the 2-step proposed process, it is applied to a microsimulation traffic model created in Paramics, which is a microscopic traffic simulation software that can simulate the vehicles and the traffic lights of the traffic network.

In this chapter, the application of the process to estimate the traffic state of the simulated network is presented. First, in Section 4.1, the requirements that the simulation environment needs to fulfil, the characteristics of the simulated network and the reasons that this particular network was chosen, are presented. For the purposes of the test application, vehicle trajectories and loop detector data are necessary. The collection of these data from the simulation model is presented in Section 4.2. Afterwards, Section 4.3 describes the preparatory actions that are performed stepwise to obtain the MFD and estimate the traffic state with the collected data. This section also offers insights on the issues encompassed while preparing the data. Moreover, the recommended number of simulation runs to obtain the final reliable results is discussed in Section 4.4.

4.1. Network setup

Undoubtedly, the application environment of the traffic state estimation needs to be as realistic as possible, since theoretically, the same process could also be applied with real data. Thus, the main requirements that the simulation model needs to fulfil are to assign trips based on an accurate demand model, to be dynamic, so that the impact of spillback is taken into consideration and to include characteristics of real driver's behaviour, such as gap acceptance, preferred speed, lane-changing and car-following choices. Furthermore, traffic lights should be included, in order to have both controlled and priority intersections, such as in real networks. Additionally, the interactions between different transport modes needs to be taken into account by considering not only vehicles, but also public transport and cyclists. The microscopic model in Paramics fulfils all the above requirements and offers a high level traffic simulation. Thus, it is considered sufficient for the requirements of our experiment application.

The Paramics model that was used is a simulated network of part of the municipality of Leidschendam-Voorburg that is located in the province of South Holland. It should be noted, that this simulation

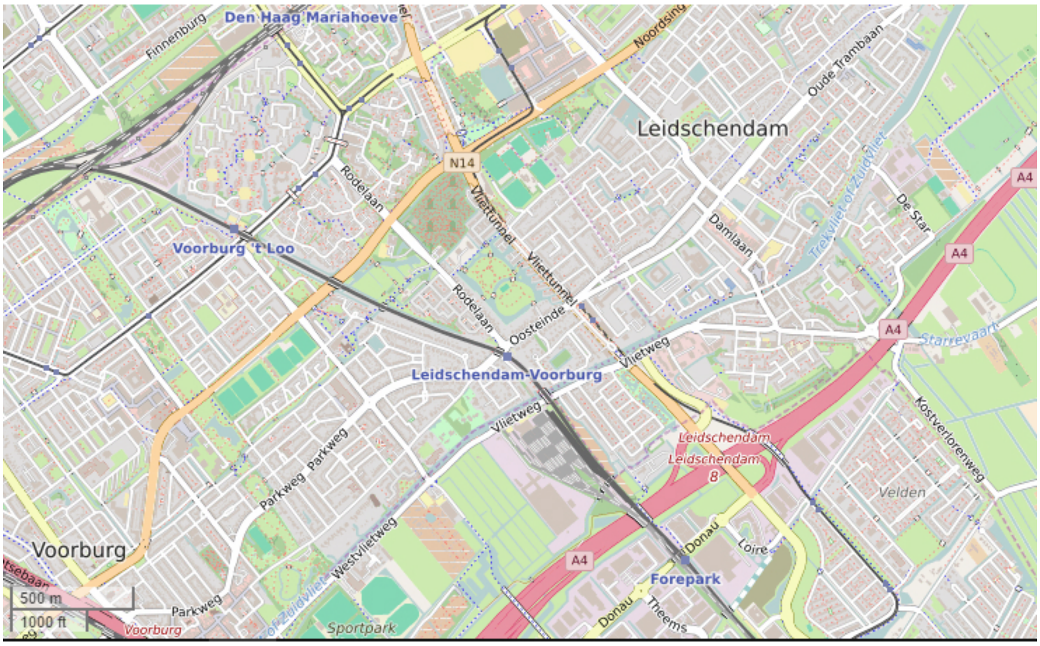


Figure 4.1: Map of the simulated area of Leidschendam-Voorburg (source: OpenStreetMap)

model was already available and was not created by the author. In the following of this section, firstly, the basic configuration characteristics and secondly, the traffic demand of the simulated network of Leidschendam-Voorburg will be described. Leading up from these two factors, the reasons are provided that this network is appropriate to test the traffic state estimation.

The traffic network of the town of Leidschendam-Voorburg includes different types of roads such as freeways, arterial roads and urban streets. It is connected to the A4 highway and the N14 crosses through the city via tunnels. A map of the area that is simulated is presented in Figure 4.1.

The simulated traffic network of Leidschendam-Voorburg in Paramics covers an area of about 7 km^2 and has a total road length of 29.9 km. It includes 67 zones connected with 491 nodes and 1068 links. The road types that are present in the network are urban roads with speed limits ranging from 30 km/h to 80 km/h and highways with a speed limit ranging from 80 km/h to 120 km/h. In total, the simulated network has 63 junctions, of which 17 are signalised and the rest are flow controlled. The vehicle types in the simulation are single vehicle units, e.g. cars or LGV, bus units, trams and HGV units and bicycles. The network is covered with 65 traffic loop detectors, of which 6 are used for the tram and/ or the bus. The configuration of the simulated network of Leidschendam-Voorburg can be seen in Figure 4.2. The location of the loop detectors across the network's intersections can be seen in green color.

The simulated time period of the analysis is the morning peak period from 06:00-10:00, when higher congestion is observed in the real network. The profile of the demand during the morning peak period is presented in Figure 4.3. The demand is presented in bars which have a time duration of 5 minutes. As it can be seen, traffic increases gradually with a peak from 07:30 to 09:00 and then, demand decreases until 10:00.

The characteristics and the demand levels of the simulated network of Leidschendam-Voorburg constitute it as an appropriate environment to test the accuracy of the traffic state estimation. More precisely, this network was chosen for the following reasons:

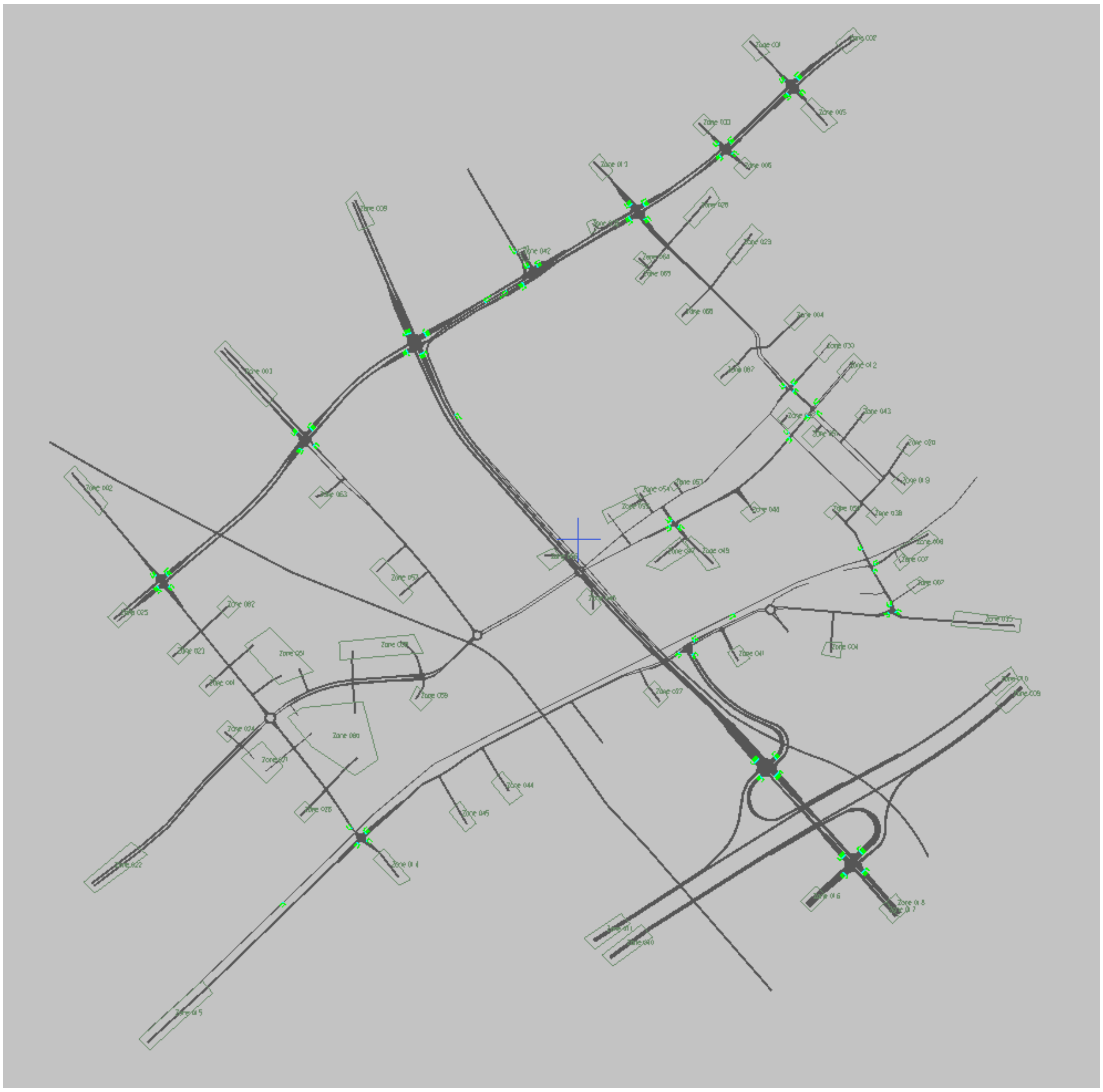


Figure 4.2: The simulated network of Leidschendam-Voorburg with the location of the loop detectors indicated with green lines

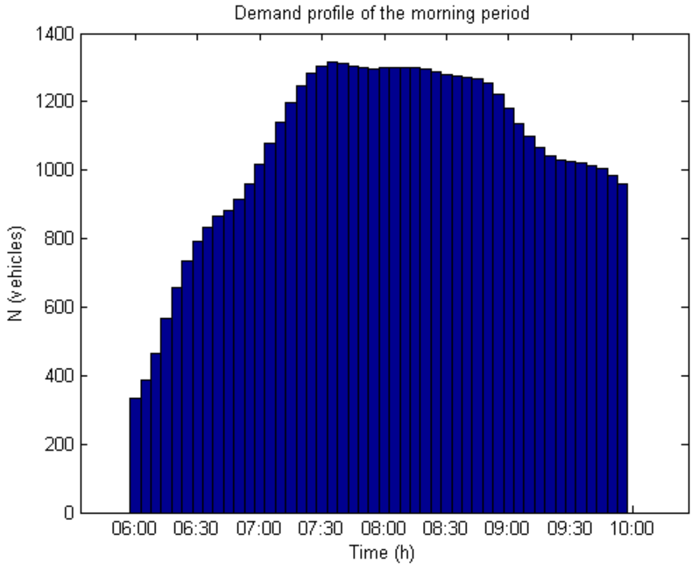


Figure 4.3: Traffic demand during the morning peak period of the simulation

1. its size is medium, so it is easy to use and observe
2. it has a reasonable amount of junctions (63), which allows to take into consideration the interactions between intersections
3. it includes the various road types that are also present in real traffic networks
4. it has significant congestion problems
5. it was already created and available to use

For all the above reasons, the Leidschendam-Voorburg network is considered an appropriate case study to test the accuracy of the proposed traffic estimation process of Chapter 3.

4.2. Data collection

The necessary data to apply the traffic state estimation process are vehicle trajectories and loop detector data. These two types of traffic data are also what can be usually collected from real-life networks. Furthermore, for our case, a file containing the statistics of the simulation model is produced, in order to check the accuracy of the variables estimated from the other two data sources.

In this section, the way that the three aforementioned data types are generated from Paramics is described. Moreover, a detailed description of the format of each data type is provided.

Different files that Paramics produces in the end of the simulation provide the necessary data. The respective files and the context of each data type are as following:

1. Vehicle Trajectories
   In order to extract vehicle trajectories data from Paramics, the Full Trip Analysis is selected for detailed link analysis and the *vehicleroutes* file is produced. This file contains for every vehicle, the entrance time in each link of its journey, so the route of every vehicle in the simulation. Thus, it can be considered as floating car data (FCD).

Table 4.1: Format of the traffic data files of the simulation (source: Paramics manual)

Data Type	Value	Description
Vehicle Trajectories	Vehicle Tag	Unique tag number for each vehicle
	Link	Link this vehicle used on its trip
	Time Entered (s)	Time vehicle entered this link
Detector data	Detector Name	Unique identifier of the loop detector
	Time (s)	Time that the vehicle crosses the loop
	Speed (km/h)	Vehicle speed as the vehicle crosses the loop
Simulation Statistics	Time	Time measurement was taken
	Current NV	Current number of vehicles in the system
	Mean Speed	Speed of vehicles in the system this last minute

2. Loop Detector data
   Loop detector data for the modeled network of Leidschendam-Voorburg are collected from 59 detectors through the file *alldetectors*. The loop detectors file is updated each time an event occurs. This means that when a vehicle crosses a loop detector, a line is added to the file with the detector ID, the time that the vehicle crossed and the vehicle's speed. The location of the detectors in the network can be seen in green color in Figure 4.2.

3. Simulation Statistics
   Simulation information is saved in the file *stats-general*, where a new row is added every minute with detailed information of the traffic situation in the simulation model. The data of this file provide the exact value of the traffic variables as collected from the simulation environment. They can be used to check the accuracy of the variables calculated from the two previous data types.

More analytically, each traffic data type and an exact description of the values that it contains, are presented in Table 4.1.

4.3. Data preparation and issues

After the data are collected from Paramics, as it was described in the previous section, they are imported in Matlab. This section describes the alterations that are performed in the data to prepare them for the application of the traffic state estimation. The necessary alterations are presented following the algorithm of the 2-step traffic state estimation process, as it was described in Section 3.2.

4.3.1. Step 1: Data fusion to obtain the MFD

Starting with the algorithm of the 1st step of the process, each of the algorithm points will be given, as described in Section 3.2.1, followed by the actions to apply it.

1. From the detector data, calculate the total number of vehicles crossing the detectors in every time interval N_{det}.

   To calculate N_{det}, the detector data need to be aggregated in the time interval of the analysis, which is 5 minutes. A 5-minute time interval is considered an appropriate amount of time to capture the changes in the variables due to the traffic light cycle and the demand variation.

The vehicle crossings are grouped in 5-minute time intervals matching the 48 5-minute time intervals of the time period of the analysis from 06:00 to 10:00.

2. From the subset of known vehicle trajectories, calculate the density k_{traj} (Equation 3.6), the flow q_{traj} (Equation 3.7) and the number of vehicle trajectories that traverse the links with detectors N_{traj}.

   For this point, a subset of vehicle trajectories needs to be selected from which the density and the flow are calculated. A random sampling function was used to generate different subsets that operated as the known vehicle trajectories.

   In order to calculate k_{traj} and q_{traj} from the subset using Edie's formulas (Equation 3.6 and 3.7), the travel times and travel distances of the vehicles are aggregated in the 48 5-minute time intervals of the analysis. As it can be seen in Figure 4.4, each vehicle traveled a specific distance in a specific time, within each one of the 5-min time intervals (represented with the red dashed area). The following formulas are used to calculate accurately the time and the distance that each vehicle traveled within each time interval:

$$tt^{i}_{(j,j+1)} = min(t_{j+1}, t_{k+1}) - max(t_j, t_k) \tag{4.1}$$

$$td^{i}_{(j,j+1)} = \frac{L_{k+1} - L_k}{t_{k+1} - t_k} \cdot tt^{i}_{(j,j+1)} \tag{4.2}$$

   where

   $tt^{i}_{(j,j+1)}$: travel time of vehicle i during time interval from t_j to t_{j+1}

   $td^{i}_{(j,j+1)}$: travel distance of vehicle i during time interval from t_j to t_{j+1}

   t_j : start of the time interval

   t_{j+1} : end of the time interval

   $L_{k+1} - L_k$: distance between links k and $k+1$ that vehicle i used in its trip

   t_k : time that vehicle i enters link L_k

   t_{k+1} : time that vehicle i enters link L_{k+1}

3. For each time interval, calculate the proportion of the known vehicle trajectories in relation to the total number of vehicles a_{traj}.

   For each one of the 48 5-minute time intervals of the time period of the analysis, a_{traj} is calculated, dividing the number of vehicle trajectories with the detector flows. As mentioned, using a random sampling function, different subsets of vehicle trajectories were selected to operate as the known vehicle trajectories. The reason for that is to simulate as realistic as possible the penetration rates of floating car data that may happen in real networks. The proportions that were sampled vary within the simulation time around the following rates: 1%, 3%, 5%, 10%, 20%, 30%.

4. For each time interval, calculate the average network density k_{net} and the average network flow q_{net}.

   Dividing the flow and the density calculated from the subset of known vehicle trajectories with the proportion of the known vehicle trajectories, 48 points of average network density and average network flow are calculated.

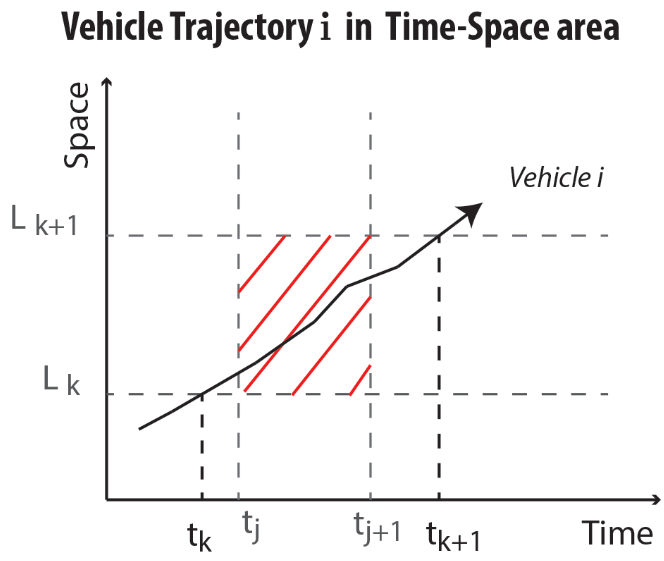


Figure 4.4: Schematic representation of a vehicle trajectory and the application of Edie's definitions

5. Construct the MFD of the network finding the mathematical function that best fits the data points of k_{net} and q_{net}.

   In order to acquire sufficient amount of points of network flow vs. network density to construct the MFD, the traffic demand is increased to cause maximum density in the network. The ideal situation would be to cause complete gridlock in the traffic network with maximum network density and zero network flow. In order to achieve that, the default demand profile presented in Figure 4.3, which corresponds to a demand rate of 100%, was increased to a rate up to 200% to simulate a gridlock situation.

   However, it was not made possible to capture the gridlock situation in the collected data. The reason for that is that Paramics generates the file with the vehicle trajectories at the end of the simulation time and it includes only the vehicles that completed their journey. When the simulated network is completely blocked, the vehicles are delayed for too many hours or are completely stopped. The realistic behaviour of the drivers in the case that their route is blocked would be to take an alternative route. However, this is not the case in the Paramics simulation environment and instead, the vehicles are waiting to follow their desired route at any cost and finally, do not manage to finish their journey within the simulation time. For the gridlock situation, the simulation time was increased up to even 24 hours, but even with that high margin, it was not made possible to produce the complete vehicle trajectories file.

   Nevertheless, it is not necessarily expected that the MFD of a real network will reach the gridlock point. This is because there will always be roads in the network that are not completely congested and the vehicle flow continues. Consequently, it is realistic to say that the MFD of a real network will not reach the point of maximum density and zero flow. For this reason, the demand rate of the simulation model was increased up to the point that it was still possible for the simulated vehicles to finish their journey within reasonable additional time (the simulation ran for two extra hours, so from 06:00 to 12:00).

   At this point, in order to know the demand rate at which the generated vehicle trajectories file is complete, the results of estimating the total number of vehicles were compared to the simulation statistics file. Figure 4.5 shows that when the demand rate is up to 120%, the difference

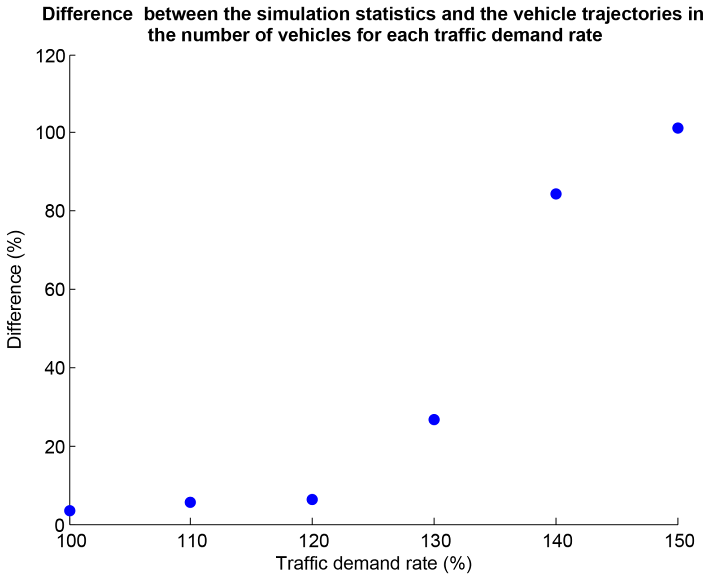


Figure 4.5: The effect of higher traffic demand on the accuracy of the total number of vehicles

is up to 6%. However, after the demand increases over 120%, the difference is significantly high (27% for 130% demand rate and 101% when the demand rate is 150%). This means that when the demand rate is more than 120%, the vehicle trajectories file is not accurate and there are too many vehicles that did not finish their journey within the simulation time. For this reason, the "trusted" demand rates are 100%, 110% and 120%.

After creating the data points of the MFD, the formula that fits the data best is searched. The resulting formula is the data fusion MFD of the network. In order to validate the data fusion MFD formula, the ground-truth MFD is also created using 100% vehicle trajectories. The results of the comparison are presented in the next chapter.

4.3.2. Step 2: Traffic state using the MFD and speed data

After executing the algorithm of the 1^{st} step, the data fusion MFD is used in combination with speed data to derive the traffic state in the 2^{nd} step. The points of the algorithm of the 2^{nd} step are given next, followed by the actions to apply them.

1. Given the obtained MFD formula, calculate the relationship between speed and density for the network from $q = kv$.

   The relationship between the speed and the density was constructed and plotted in order to reassure that there is a well-defined relationship between these two variables.

2. Given the relationship between speed and density for the network and speed measurements from any available data source, calculate the average network density.

   Speed measurements are retrieved from the vehicle trajectories with Equation 3.5. Since not all of the vehicle trajectories are expected to be known, a random sampling function was used to select low penetration rates that vary within the simulation time. The rates that were used are: 1%, 3%, 5%, 10%, 20%, 30%.

   Theoretically, speed data from the detectors could also be used applying Equations 3.3 and 3.4. However, this was not made possible, due to the location of the network detectors, which

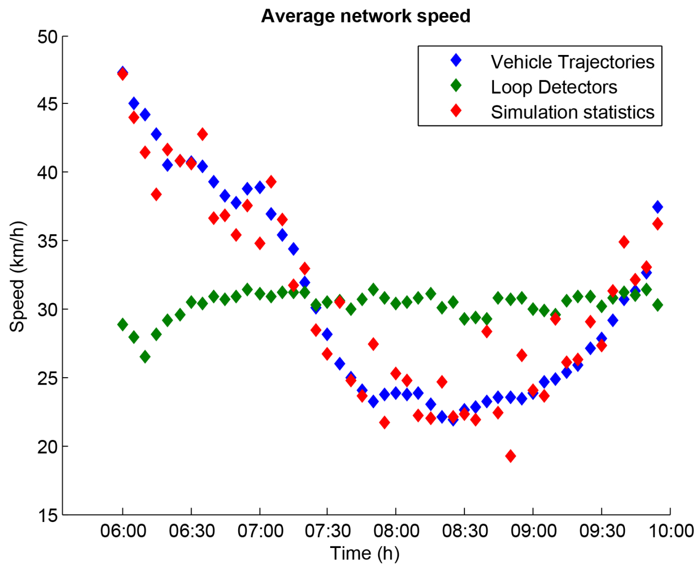


Figure 4.6: Average vehicle speed in the network during the time period of the simulation

is always near the stop line (see also Figure 4.2). The detector location directly upstream of the traffic light leads to the detectors capturing only very low vehicle speeds, and so the situation more upstream of the traffic light is not represented and the measured detector speed is not representative for the average network speed.

A speed comparison plot is shown in Figure 4.6 to depict the problem with the detector location. It shows the average network speed resulting from the vehicle trajectories, the detector data and the simulation statistics for the time period of the analysis. The average network speeds resulting from the vehicle trajectories and the simulation statistics are relatively similar (blue and red data points in Figure 4.6). In contrast, the speeds resulting from the loop detectors have a completely different behaviour being always about 30 km/h (green data points in Figure 4.6). In order to showcase further that the detector speeds are not representative of the situation, frequency bars of the speed variation at 4 different moments through the analysis period are presented in Figure 4.7. It can be observed here that, although the frequency increases when there are more vehicles (times 08:00-08:05 and 09:00-09:05), the range of the speeds is the same. Most of the vehicles have a speed around 20-30 km/h, which results in the biased average speed estimations presented in Figure 4.6.

Concluding from the speed comparison plot of the three data types, the loop detector speeds cannot be used for the speed measurements of the 2nd step. Thus, speed measurements from the vehicle trajectories were the only speed data that were used to derive the network density.

3. Use the derived average network density to indicate the traffic state on the MFD.

   The derived average network density is the indication of the traffic state. At this point, in order to validate the results of the traffic state estimation, the derived densities are compared to the real densities. The validation results are presented in the next chapter.

4.4. Data sample size

After following the data preparation steps that were presented in the previous section, 48 data points of network flow and network density, corresponding to the 48 5-minute time intervals of the simu-

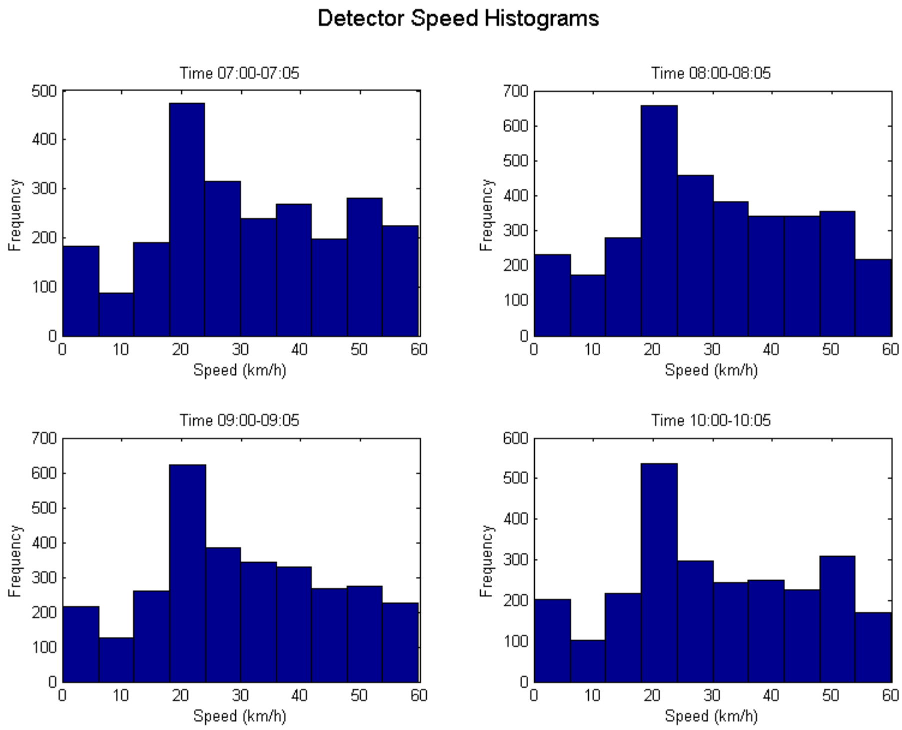


Figure 4.7: Frequency histograms of the average vehicle speed from the detector data during the time period of the simulation

lation time from 06:00 to 10:00, are available. In order to create the two desired MFDs, ground-truth MFD and the data fusion MFD, the same steps were performed for two sets of six runs of the simulation model. The reason that different data are used for these two MFDs is to ensure a reliable comparison environment for the data fusion MFD. Hence, in total, 12 simulation runs were used. For each one MFD, 288 data points were used resulting from the 48 points of each of the set of six runs (6 runs x 48 points).

The six runs of each set include two runs for each of the "trusted" traffic demand rates: 100%, 110% and 120%. There are two reasons that two runs were included from each traffic demand rate. The first reason is to tackle the randomness effect of the simulation environment by using multiple random seeds. The second reason is to include both of the two release algorithms that Paramics offers for the same demand data: precise and stochastic. The characteristics of each demand release algorithm are the following (Paramics manual):

1. Precise: The precise algorithm aims to release the exact amount of vehicles defined by the trip matrix and the demand profile and is the default setting of Paramics.

2. Stochastic: In the stochastic algorithm, the probability of a vehicle release each second is calculated and applied for every second of the demand profile interval and recalculated for the next interval. In this way, the average release for the same vehicle trip will tend to the required value but the actual release in each run will be around that average. Paramics software suggests the use of the stochastic algorithm when it is desired to investigate the reliability of results under variation.

Finally, it was discovered that the two algorithms produced almost the same outcome. This is because of the way that Paramics applies the stochastic algorithm resulting to the stochastic release

Table 4.2: Characteristics of the two sets of simulation runs

MFD-Set	Simulation run	Traffic demand rate	Release algorithm
A [Ground-truth]	1	100%	precise
	2		stochastic
	3	110%	precise
	4		stochastic
	5	120%	precise
	6		stochastic
B [Data fusion]	1	100%	precise
	2		stochastic
	3	110%	precise
	4		stochastic
	5	120%	precise
	6		stochastic

being very close to the precise release. Nevertheless, both of the algorithms were still used aiming at a more complete simulation depiction.

Concluding, two sets of six simulation runs were used, containing two runs, one precise and one stochastic, from each one of the three traffic demand rates. For clarity purposes, the set that was used for the ground-truth MFD will be called set A and the set for the data fusion MFD will be called set B. In Table 4.2, each set of simulation runs is enlisted with the corresponding traffic demand rate and the type of demand algorithm.

5

Results of the Traffic State Estimation

As it has been described in Chapter 3, the proposed process to derive the traffic state includes two steps. The 1st step is to obtain the MFD by fusing the available subset of vehicle trajectories and detector data and then, in the 2nd step, to use the obtained data fusion MFD and speed data to derive the network density, and thus the point on the MFD that the network is performing. In order to test the accuracy of the proposed process, the Paramics simulated network of Leidschendam-Voorburg, which was described in Chapter 4, was used as an application environment. The data that are necessary for the application of the traffic state estimation process, which are vehicle trajectories data and loop detector data, were collected from the simulated network and they were appropriately prepared, as it was also shown in Chapter 4.

In this chapter, the results of each step of the traffic state estimation on the simulated network of Leidschendam-Voorburg are presented. First, in Section 5.1, the process of constructing and fitting the MFD will be shown, followed by the resulting ground-truth MFD and the data fusion MFD. The ground-truth MFD is constructed in order to validate the accuracy of the data fusion MFD. Afterwards, in Section 5.2, the resulting network densities derived by the MFD and speed data are compared to the real densities, in order to check the accuracy of the traffic state estimation.

5.1. Step 1: Data fusion to obtain the MFD

As mentioned in Section 4.4, the required number of simulation runs to construct the MFD was determined to be six. The reason for this is to have a complete representation of the traffic situation in the simulated network and to tackle randomness effects. Simulation data are needed to construct both the data fusion MFD and the ground-truth MFD, which will be used as validation of the data fusion MFD. Different simulation data are used for each MFD in order to create a reliable comparison basis. Hence, 12 simulation runs (set A and B) of the simulated network of Leidschendam-Voorburg are used in total and their characteristics were presented in Table 4.2.

In this section, first, the way to fit a formula on the data points of the MFD will be presented. Then, the construction of the ground-truth MFD using 100% vehicle trajectories from the simulation runs of set A will be shown. Following, the construction of the data fusion MFD using different proportions of vehicle trajectories from the simulation runs of set B will be described. The resulting data fusion MFD will be compared to the ground-truth MFD in order to be validated.

5.1.1. MFD fitting

The MFD is approximated by a function relating the flow q with the density k and fitting the observed data. As mentioned by Knoop and Hoogendoorn (2013), the MFD can be described by many different formulas. The simplest formula is preferred, so as to be easier to explain. The requirements that the formula of the MFD needs to fulfil are:

1. The formula needs to follow a concave shape, due to the alternating trend of the two MFD branches: the free flow and the congested branch. In the free flow branch, the flow increases rapidly when the density increases and the slope of the curve is quite steep. In the congested branch, the flow decreases gradually with the density increase, so the slope is less steep.
2. The MFD has zero flow for zero density, so the fitting formula needs to have zero constant.
3. The derivative of the MFD function at zero density should be equal to the free flow speed.

Many functions were tested in order to fulfil the requirements and finally, a function with the general form of an exponential polynomial was selected, as in Equation 5.1. The reason that an exponential function is preferred over a polynomial is that the exponential term captures the alternating trend that the two MFD branches MFD have, as described in the first requirement.

$$q(k) = a_0 + \sum_{i=1}^{\infty} a_i \cdot (k \cdot e^{Ak})^i \tag{5.1}$$

where

q: flow

k: density

A: exponential coefficient

a_i: coefficients

According to the second requirement, the term a_0 is set equal to zero. In order to decide the order of the function, its simplest form for $i = 1$, thus $q(k) = a_1 \cdot k \cdot e^{Ak}$ is compared to its form for $i = 2$, so $q(k) = a_1 \cdot k \cdot e^{Ak} + a_2 \cdot k^2 \cdot e^{2Ak}$ in Figure 5.1. As it can be seen, the second-order function moves the MFD down and a bit to the right, leading to a smoother concave shape. Thus, from a traffic flow point of view, the term a_2 reassures that the flow decreases gradually and less steep in the congested branch, fitting the observed data better. Regarding the interpretation of the coefficient a_1, it can be easily found that the derivative of the function at zero density is equal to a_1, so this is the value of the free flow speed. As it will be seen in the next subsections, the resulting values of a_1 are in accordance with the observed values of the free flow speed. Hence, the final selected function (Equation 5.2) covers all the requirements.

$$\textbf{General MFD function}: q(k) = a_1 \cdot k \cdot e^{Ak} + a_2 \cdot k^2 \cdot e^{2Ak} \tag{5.2}$$

where

q: flow

k: density

A: exponential coefficient

a_1, a_2: coefficients

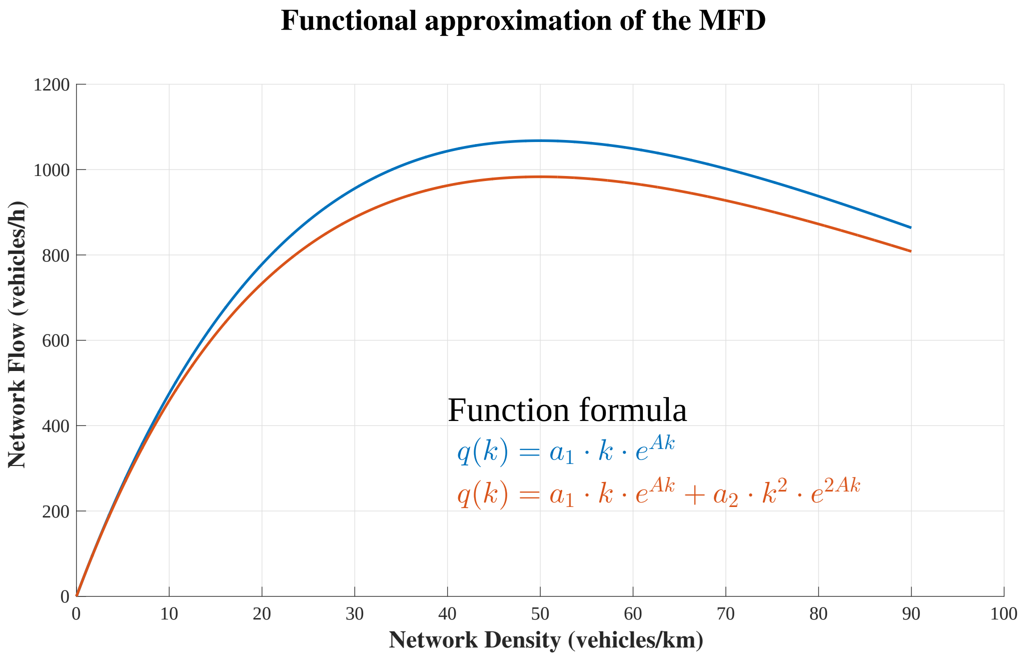


Figure 5.1: Different functions to approximate the MFD

5.1.2. Ground-Truth MFD

As it was seen in Table 4.2, the simulation runs from set A (number 1 to 6) are used for the construction of the ground-truth MFD. In order to obtain the ground-truth MFD, 100% vehicle trajectories data are used to calculate the network density (Equation 3.6) and the network flow (Equation 3.7) from Edie's definitions. The resulting data points are shown in Figure 5.2. The formula that fits the data points is plotted over them in the same figure (black curve). The resulting formula following the form of the general MFD function (Equation 5.2) is shown in Equation 5.3 with q being the flow and k the density.

$$\textbf{Ground-Truth MFD function}: q(k) = 54.4 \cdot k \cdot e^{-0.02k} + 0.19 \cdot k^2 \cdot e^{-0.04k} \tag{5.3}$$

$$Adjusted R^2 = 0.94$$

$$RMSE = 49.51$$

The derivative of the MFD formula at $k = 0$ gives the free flow speed. In the case of the chosen formula function, the derivative at $k = 0$ is equal to 54.4 km/h. This result is in full accordance with the observed speed data.

In order to check the goodness of the fit, the adjusted R^2 is calculated. The closer the value of adjusted R^2 is to 1, the better the fit is.The formula found to fit the ground-truth MFD data has an adjusted R^2 of 0.94, which means that it has a very good fit on the data. Another measure of accuracy that can be used is the root mean square error (RMSE), which compares the fitted to the observed values of the formula. The smaller the value of RMSE is, the better the fit of the formula is, so it can be used to compare the fit of the ground-truth MFD versus the fit of the data fusion MFD. It has the same units as the flow and for this case, it is equal to 49.51 vehicles/hour.

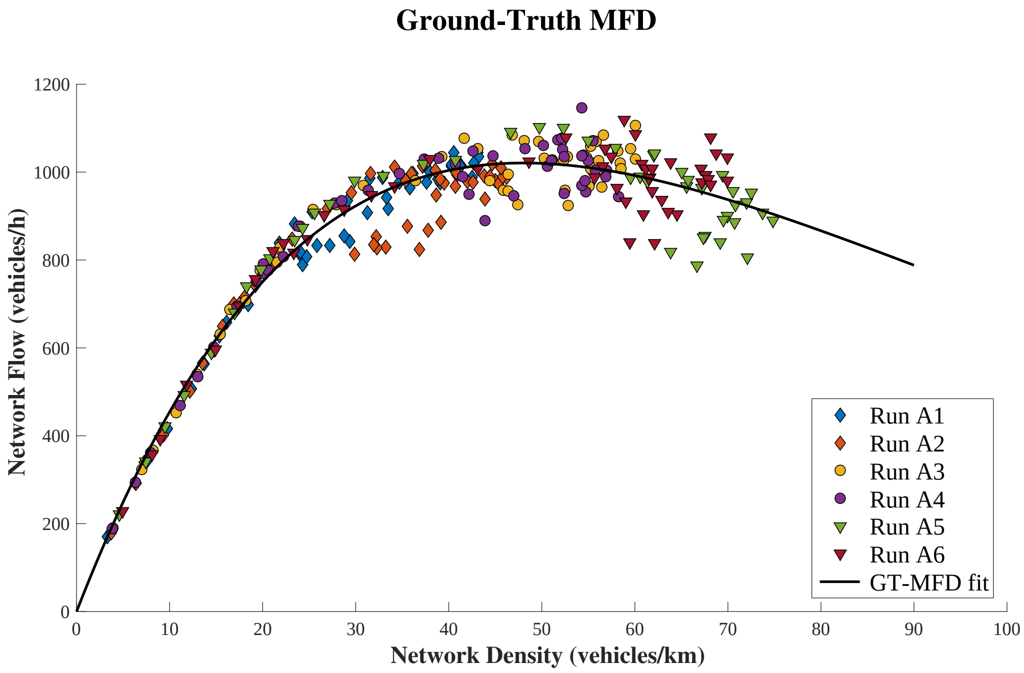


Figure 5.2: Ground-truth MFD of the simulated network of Leidschendam-Voorburg

5.1.3. Data Fusion MFD

The simulation runs from set B (number 1 to 6) (Table 4.2) are used for the construction of the data fusion MFD. As introduced in the proposed traffic state estimation process (Section 3.2.1), only fractions of the vehicle trajectories are used to obtain the data fusion MFD, dividing the proportion of the known vehicle trajectories with the total detector flows. The proportions of vehicle trajectories that were used are: 30%, 20%, 10%, 5%, 3% and 1%.

The resulting data points of network flow and network density produced from the data fusion process are presented in Figure 5.3. As expected, the data points derived from the data fusion process are much more scattered, compared to the data points of the ground-truth MFD, where 100% of the vehicle trajectories were used. The formula that fits the data fusion MFD is also plotted in Figure 5.3 (red curve) and is shown in Equation 5.4.

$$\textbf{Data fusion MFD function}: q(k) = 58.05 \cdot k \cdot e^{-0.02k} - 0.25 \cdot k^2 \cdot e^{-0.04k} \tag{5.4}$$

$$Adjusted R^2 = 0.90$$

$$RMSE = 69.20$$

In the case of the data fusion MFD, the derivative of the function at $k = 0$ is equal to 58.05 km/h. This result is again in full accordance with the observed speed data. The adjusted R^2 is 0.9, which is again very close to 1 and thus, the fit is very good. The RMSE value is equal to 69.20 vehicles/h. The meaning of this measure is understood better in the next section, where the two MFDs are compared.

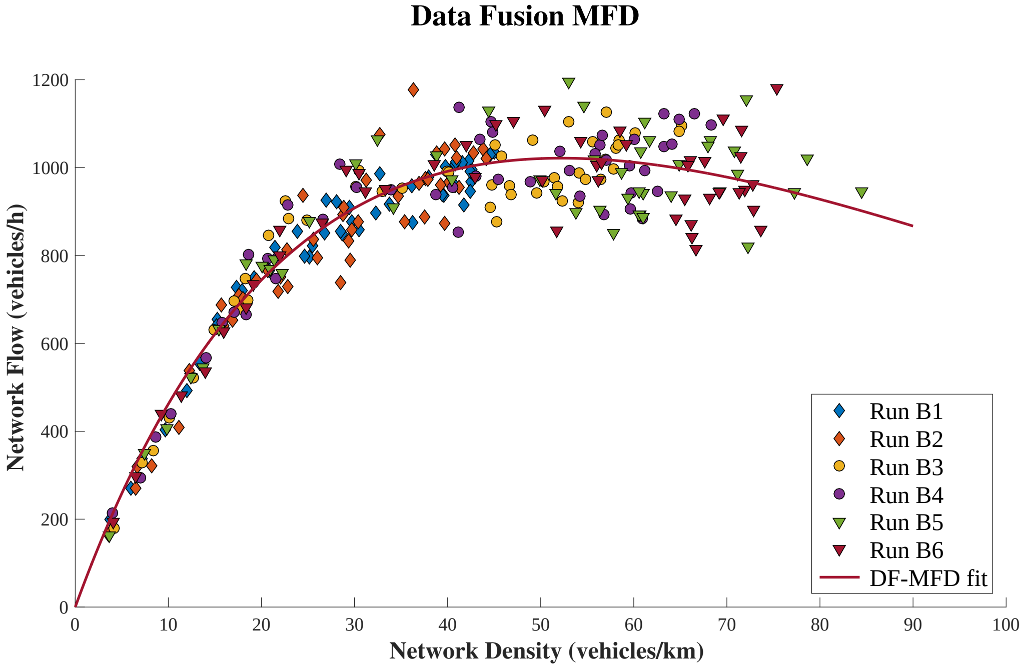


Figure 5.3: Data fusion MFD of the simulated network of Leidschendam-Voorburg

5.1.4. Comparison of the MFDs

Comparing the two MFD formulas (Equation 5.3 and Equation 5.4), the formula for the ground-truth MFD has a higher adjusted R^2 and a lower root mean square error (RMSE) than the data fusion MFD. This means that the ground-truth MFD has a better fit, which was expected since the data for it do not deviate a lot. Whereas, the data for the data fusion MFD have higher dispersion, due to the random sampling of the known vehicle trajectories during the time period of the analysis.

Nevertheless, as it can be seen in Figure 5.4, the two MFDs have very similar behaviour and even share many similar points. In the free flow part, when the flow is low, the data fusion MFD (red curve) is slightly lower than the ground-truth (black curve), while moving to higher flows, the opposite occurs. In the congested branch, the data fusion MFD is higher than the ground truth MFD. This is because the fused data deviate much more in the high flows compared to the low-flow part. This is an indication that it is more difficult to represent the congested traffic state with the fused data compared to the free-flow state.

Regarding the characteristic points of the MFD, the maximum capacity point occurs at $q = 1022$ vehicles/ hour and at $k = 52$ vehicles/ km for the data fusion MFD and at $q = 1021$ vehicles/ hour and at $k = 48$ vehicles/ km for the ground-truth MFD. The speed in the maximum capacity point is equal to $v = 19.7$ km /h and $v = 21.3$ km /h for the data fusion and the ground-truth MFD, respectively. From the derivative of the MFD formulas at $k = 0$ vehicles/ km, the free flow speed was found equal to $v = 58.05$ km/ h for the data fusion MFD and $v = 54.45$ km/ h for the ground-truth MFD. All the function parameters and the characteristic points for each MFD are collectively presented in Table 5.1.

Concluding from the comparison of the two MFDs, the random sampling of the vehicle trajectories leads to high variation in the resulting flow and density data points, but the resulting data fusion MFD fit is still very good and similar to the ground-truth MFD fit. Also, the characteristic points are very similar between the two MFDs, indicating that the data fusion MFD is a very good approximation of the real traffic situation of the network.

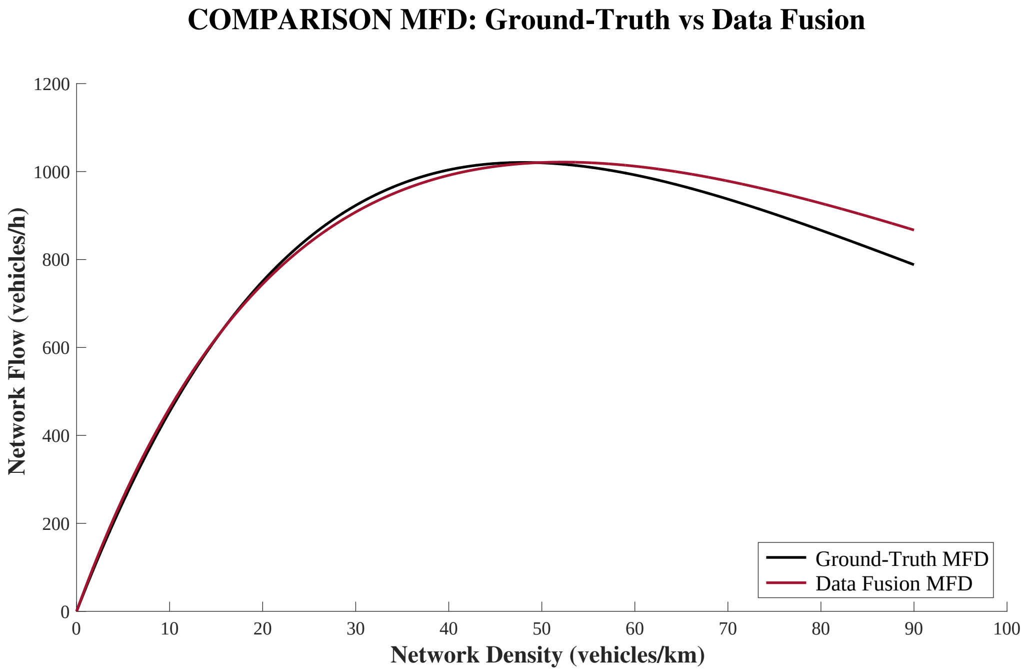


Figure 5.4: Comparison between the Data fusion MFD and the Ground-truth MFD

Table 5.1: Comparison between the parameters of the Data Fusion MFD and the Ground-Truth MFD

	Parameters	Ground-truth MFD	Data Fusion MFD	Units
Function	A	-0.02	-0.02	-
	a1	54.4	58.05	-
	a2	0.19	-0.25	-
	Adjusted R^2	0.94	0.90	-
	RMSE	49.51	69.20	vehicles/h
Points	Free Flow Speed	54.4	58.05	km/h
	Maximum Capacity	1021	1022	vehicles/h
	Critical Density	48	52	vehicles/km
	Critical Speed	21.3	19.7	km/h

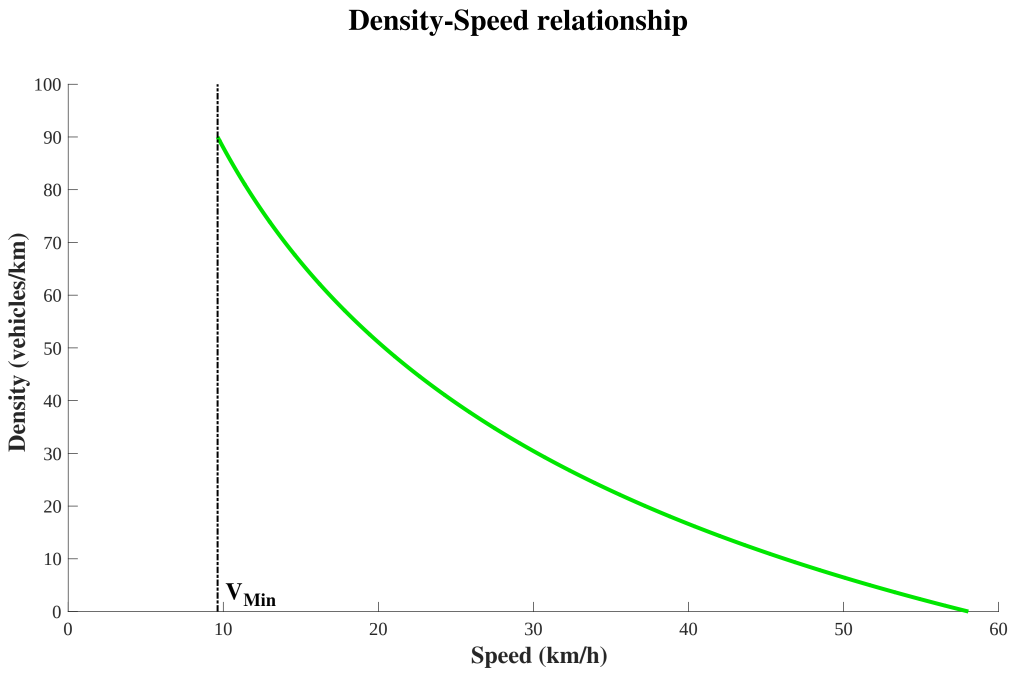


Figure 5.5: Density-Speed relationship of the simulated network of Leidschendam-Voorburg

5.2. Step 2: Traffic state using the MFD and speed data

Continuing with the proposed traffic state estimation process, after obtaining the data fusion MFD, the next step is to estimate the relationship between the speed and the density from the relationship $q = kv$. The obtained relationship between the network density and the network speed of the simulated network of Leidschendam-Voorburg is shown in Figure 5.5.

Substituting the average speed taken from a fraction of vehicle trajectories in the relationship of Figure 5.5, results in the calculation of the network density. The data from the simulation runs A1-A6 are used to derive speed measurements from different fractions of vehicle trajectories. The fractions used are: 30%, 20%, 10%, 5%, 3% and 1%. The resulting network densities are compared with the real densities taken from 100% vehicle trajectories to validate their accuracy. The resulting deviation of the estimated densities from the real values can be seen in Figure 5.6.

As it can be seen in Figure 5.6, the error seems to be rather constant throughout the entire density domain. It would be expected that the errors would increase relatively to the increase in density, because of the error magnifying as the density increases and also because in the high densities, the slope of the Density-Speed relationship (Figure 5.5) is more steep. This means that it is more difficult to estimate the exact density, because the differences in speed at that part will lead to large differences in the estimated densities. Nevertheless, when the density is low, it means that there are not many vehicles to represent accurately the traffic situation and maybe they are very spread, so the final estimated density may be biased. Thus, finally, the error follows a rather constant tendency as densities increase.

In Table 5.2, the mean percentage error and the standard deviation for each fraction of vehicle trajectories is calculated. What can be observed is that the higher the fraction of the known vehicle trajectories is, the lower the mean error and the standard deviation are. This means that the estimated densities are more similar to the reality when more vehicle trajectories are known, which is an expected outcome. More precisely, for a fraction of 30% of known vehicle trajectories, the error is 7.6% with a standard deviation of 0.15 vehicles/km and for the lowest fraction tested, which is

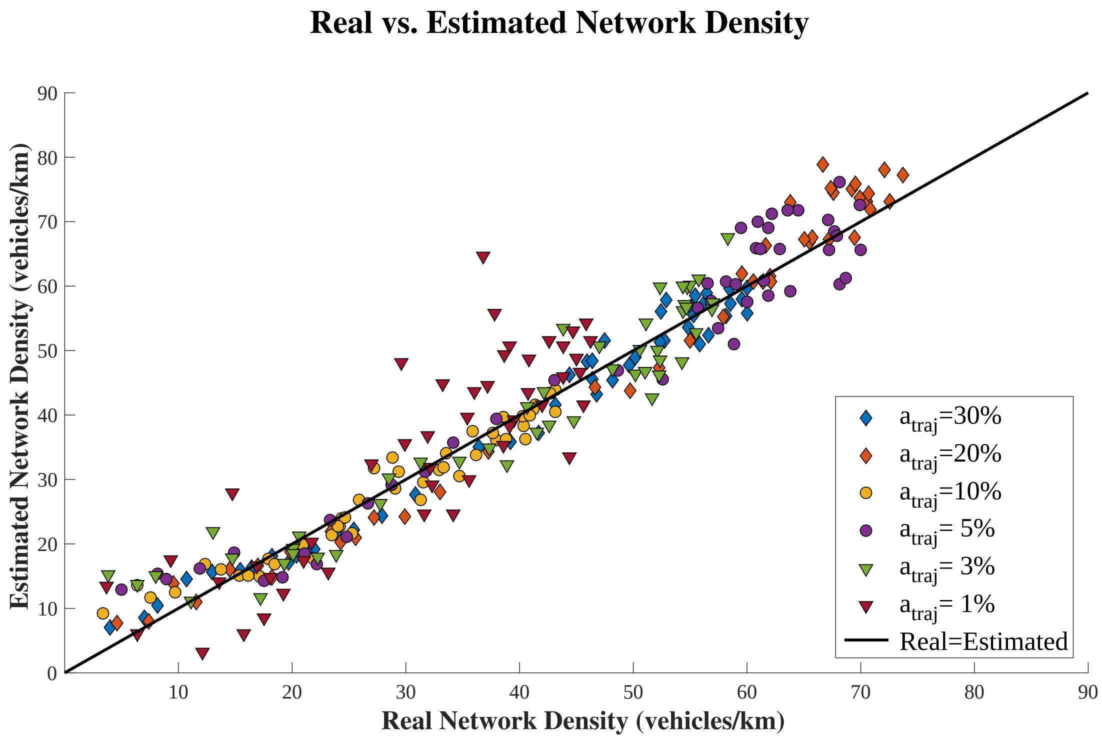


Figure 5.6: Comparison between the estimated and the real network densities

Table 5.2: Mean percentage error and standard deviation of the Densities estimated from various proportions of known vehicle trajectories a_{traj}

Known Vehicle Trajectories a_{traj}	Mean error (%)	Standard deviation
30%	7.6	0.15
20%	10.8	0.25
10%	11.7	0.30
5%	12	0.33
3%	14.3	0.50
1%	26	0.52

1%, the error is almost three times higher and equal to 26% with a standard deviation of 0.52 vehicles/km. In the case of 5%-10% of known vehicle trajectories, the error is 12% and deviates 0.3 vehicles/km.

The presented results can offer a first glance at how the different penetration rates influence the accuracy of the estimation of the traffic state. Nevertheless, these results should not be perceived as nothing more than a validation test on how well the proposed process works under different conditions. If it is desired to extract conclusions on the required penetration rate of floating car data, additional information is needed on whether an exact traffic state estimation is desired or it is sufficient to only know whether the network is in the congested state or not. The desired accuracy level can be determined depending on the purpose that the traffic state estimation will be used for.

6
Uncertainty of the Traffic State Estimation

In Chapter 5, the results of applying the 2-step traffic state estimation on a simulated network were presented and validated. The data fusion MFD obtained in the 1st step was compared to the ground-truth MFD of the simulation and the estimated traffic densities of the 2nd step were compared to the real densities. However, in reality, neither the ground-truth MFD nor the real density of the network are known. Therefore, a different way is necessary to express the confidence that we have in the validity of the traffic state estimation.

In this chapter, a detailed error analysis is performed aiming to assess the confidence level of the estimated traffic state. Hence, the third sub-question, that was formed to reach the research objective will be answered:

What is the effect of the uncertainty that the obtained Macroscopic Fundamental Diagram encompasses on the derived traffic state?

First, the parts that constitute the uncertainty of the estimated traffic state will be discussed in Section 6.1. Next, the concept and the necessary calculations to determine the uncertainty of the estimated traffic state will be described in Section 6.2, followed by the respective results. Finally, in Section 6.3, examples of traffic state estimation cases with their confidence levels will be given.

6.1. Uncertainty Components

The proposed 2-step traffic state estimation of Chapter 3 includes two components of uncertainty, one at each step. At the 1st step, there is uncertainty in the data used to find the best MFD fit. At the 2nd step, the speed data that are used to indicate the traffic state on the MFD include errors as well, since data are erroneous in most cases. The two components contributing to the final uncertainty of the estimated traffic state will be further analysed in this section.

6.1.1. Uncertainty of the Data fusion MFD

In the 1st step of the traffic state estimation process, the data fusion of subsets of vehicle trajectories with detector data provided data points of network density and network flow. Fitting a formula to these data points resulted in the function of the data fusion MFD presented in Equation 5.4. In order to assess how well the formula fits the data, the root mean square error was calculated, or else called the standard error of the estimate σ, as in Equation 6.1. In order to correct for the bias due to the

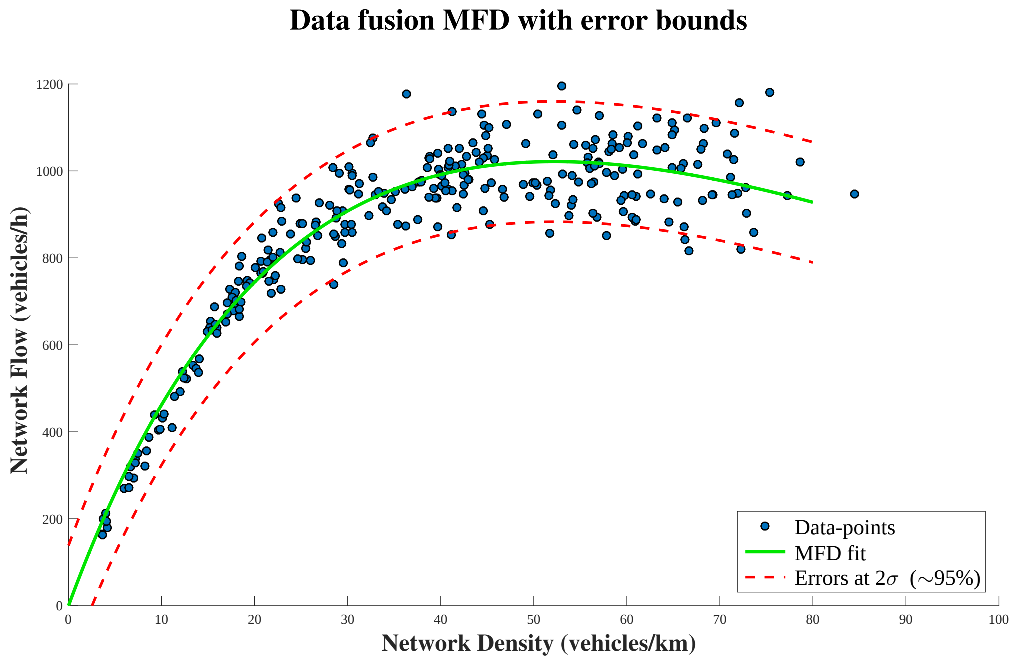


Figure 6.1: Data fusion MFD with 95% confidence interval

fact that our observed values are only a sample of the entire population of possible values, Bessel's correction is used, so dividing with $n-1$ instead of n.

$$\sigma = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{6.1}$$

where

σ = standard error of the estimate

y_i = observed value

$\hat{y}_i$ = estimated value

n = number of observed values

If we assume that the errors of the estimates (the differences of the estimated from the observed values) follow a Gaussian distribution, 95% of the observations lie within $\pm 2\sigma$. In this way, a confidence interval of 95% of certainty can be drawn around the MFD function at $\pm 2\sigma$, as in Figure 6.1.

Nevertheless, for the purposes of the 2$^{\text{nd}}$ step to estimate the traffic state, the Density-Speed version of the MFD is used. For this reason, it is more practical to present the confidence interval of the relationship between the density and the speed. Thus, in the same way as calculated before, the confidence bounds at 95% level of certainty are drawn for the Density-Speed relationship as seen in Figure 6.2.

If we plot the errors of the estimates, or as they are else called the residuals, the plot of Figure 6.3 is obtained. In this figure, we can see that the errors take both positive and negative values and are randomly spread, thus following the expected behaviour when the prediction model is good.

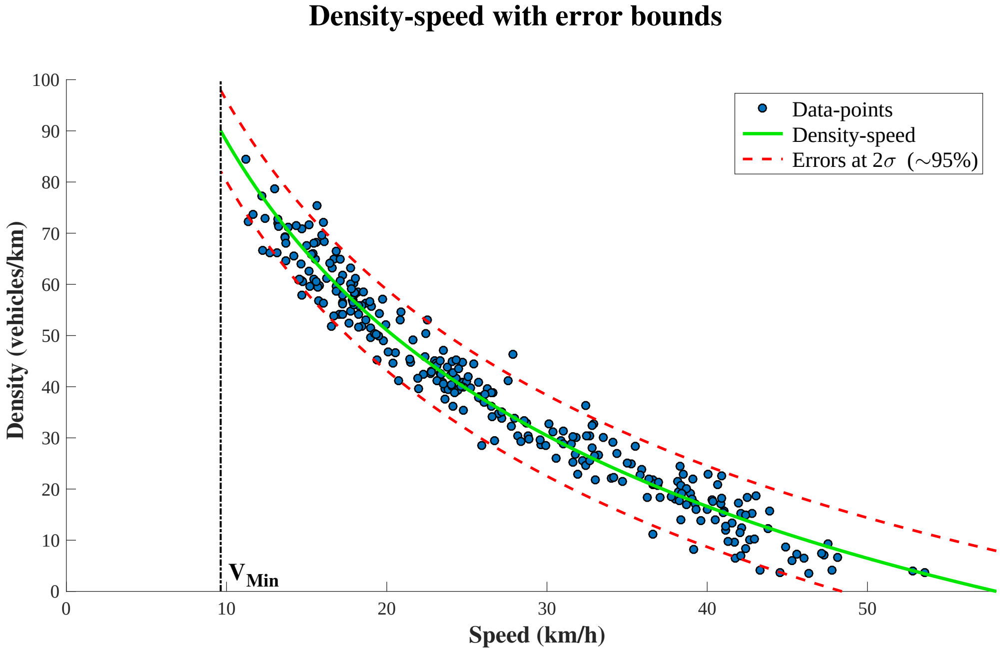


Figure 6.2: Density-Speed Relationship with 95% confidence interval

For this reason, we can assume that the errors follow a Gaussian distribution. In this way, the probability density function presented in Figure 6.4 can be constructed. As seen, the resulting error distribution for the 288 values (48 observed data points from each one of the 6 simulation runs) has a mean value of −0.4 vehicles/km and a standard deviation of 4.0 vehicles/km. This probability density function describes the probability for the error of the density to have a specific value given the real speed value, so it will be noted as $f(k|v_{real})$.

6.1.2. Uncertainty of the Speed data

Regarding the uncertainty of the speed data at the 2$^{\text{nd}}$ step of the estimation, it should be reminded here that the measured speeds are derived from different levels of low fractions of vehicle trajectories, whereas the real speeds are derived from 100% of vehicle trajectories. Following the same approach as before, the errors between the measured and the real speeds are calculated and expressed with the standard error of the estimate σ and plotted in Figure 6.5.

The errors of the measured speed data with regard to the real speed data can be seen in Figure 6.6. In this case, again, the errors spread homogeneously around the mean error value and take both positive and negative values. This means that the assumption can be made again that the errors of the speed measurements follow a Gaussian distribution.

In Figure 6.7, the probability density function of the errors of the measured speeds with regard to the real speeds is given. The resulting error distribution has a mean of 0.2 km/h and a standard deviation of 2.2 km/h. This probability density function describes the probability for the error of the measured speed to have a specific value given the real speed value, so it will be noted as $f(v|v_{real})$.

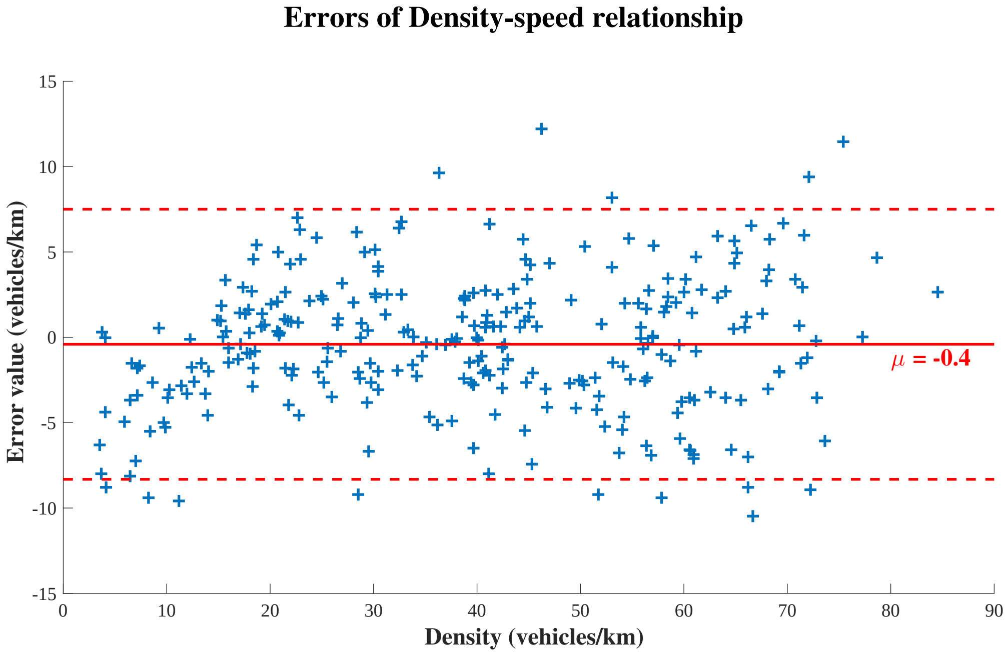


Figure 6.3: Errors of the estimates of the Density-Speed relationship

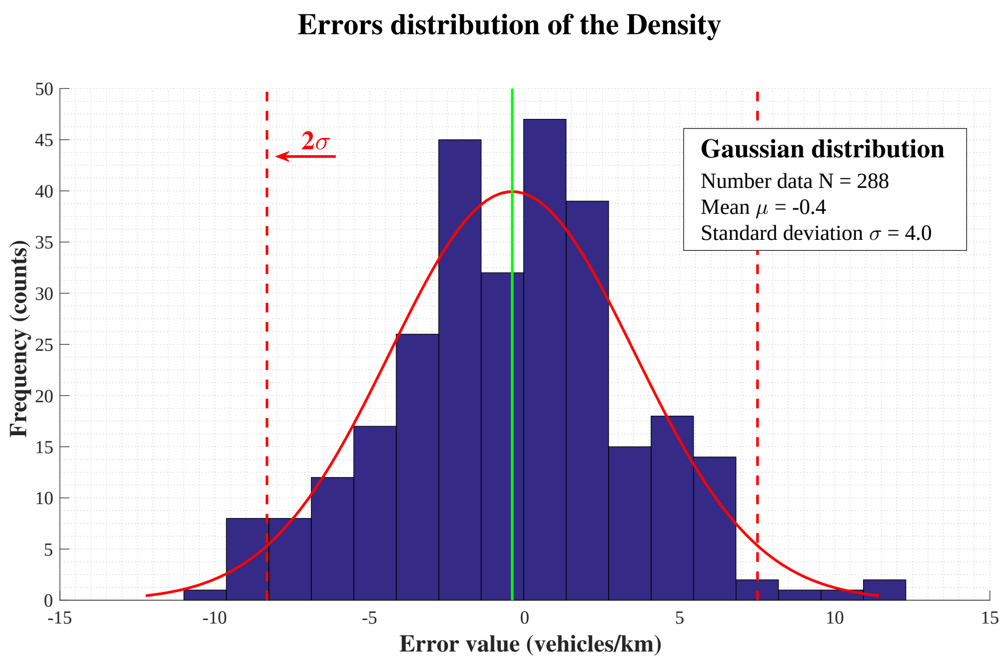


Figure 6.4: Distribution of the errors of the estimated network densities

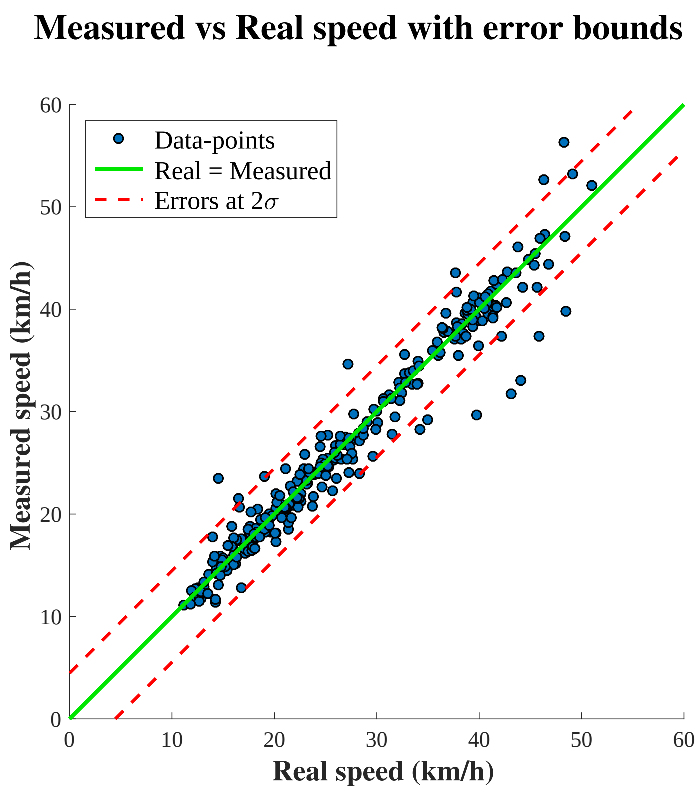


Figure 6.5: Measured vs. Real Speeds with 95% confidence interval

Figure 6.6: Errors of the measured speeds

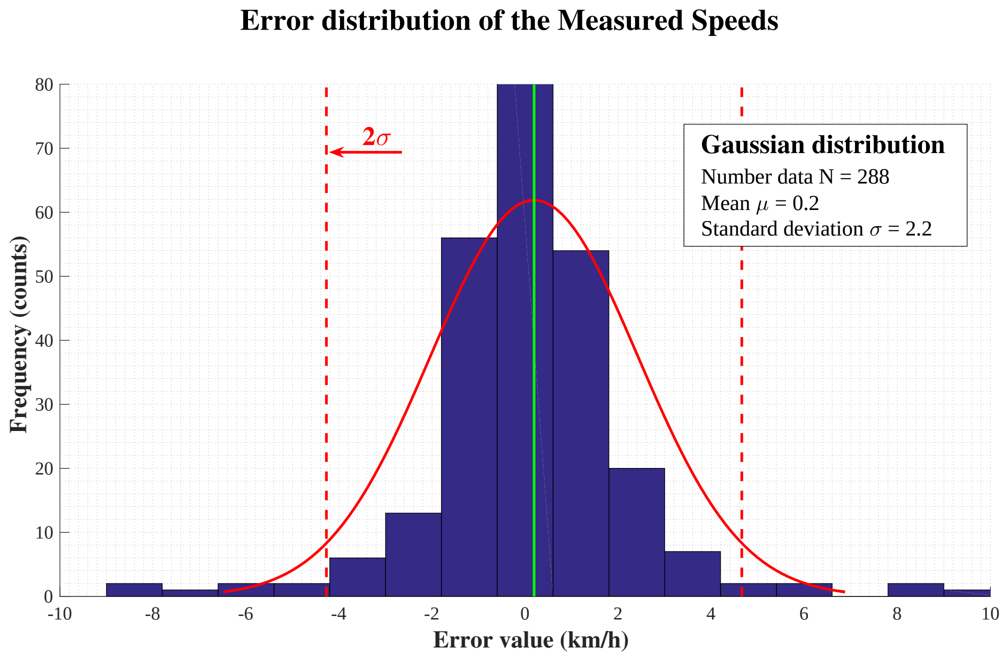


Figure 6.7: Distribution of the errors of the measured speeds

6.2. Evaluation of the Traffic State Uncertainty

The concept to assess the uncertainty of the final traffic state estimation and the mathematical calculations to apply it are described in this section. The resulting evaluation of the uncertainty level of the estimated traffic state is also presented.

6.2.1. Concept to estimate the Uncertainty

As it is indicated by Drosg (2009), scientific knowledge always encompasses some level of uncertainty. Sometimes, uncertainties can be slight, but there is no scientific truth without uncertainty at all. In this sense, we need to accept that it is impossible to determine the exact traffic state and there will always be uncertainty in the validity of the estimated traffic state. In order to deal with the uncertainty, a way is needed to express the confidence level of the estimated value. A decision on whether the estimated traffic state is sufficiently reliable or not can then be taken.

According to the previous section, there are two components causing the final estimated traffic state to be uncertain. These components are:

1. the errors of the density-speed version of the MFD

2. the errors of the speed data

These two error components follow the already described probability density functions (Figure 6.4, Figure 6.7), indicating the probability of the error to have a certain value. More analytically, the probability density function $f(k|v_{real})$ of the density-speed version of the MFD describes the probability for the error of the density to have a specific value given the real speed value (Figure 6.4). The probability density function $f(v|v_{real})$ of the speed data describes the probability for the error of the measured speed to have a specific value given the real speed value (Figure 6.7). These two probability density functions can be considered known

The combination of the above-mentioned probability density functions can lead to the joint prob-

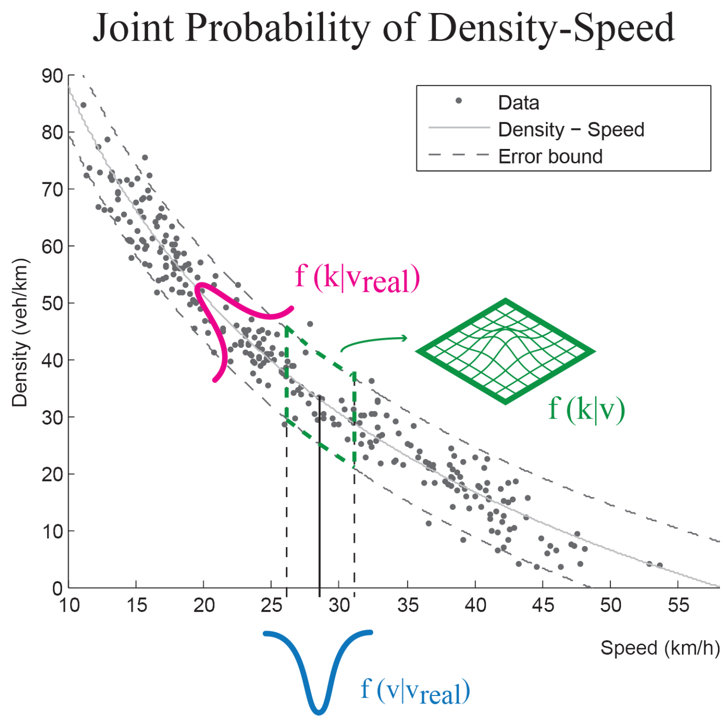


Figure 6.8: Schematic representation of the combination of the probabilities

ability of the final error, indicating the error level of the final estimated traffic state. What is needed for this purpose is to calculate the probability density function of the error of a density value occurring given a speed measurement v. This wanted quantity is represented as $f(k|v)$ and will result from the combination of the areas shaped by the two probability density functions $f(k|v_{real})$ and $f(v|v_{real})$. This concept can be seen schematically in Figure 6.8.

6.2.2. Mathematical calculations

Given a speed measurement v, we want to estimate the probability density function $f(k|v)$ of the error level of a density occurring given this speed measurement. What we have is the probability density function of the densities given the real speed measurements $f(k|v_{real})$ and the probability density function of the speeds given the real speeds $f(v|v_{real})$.

We consider the density k and the speed values v as dependent events, so the conditional probability of k given v is defined as (Kolmogorov definition):

$$f(k|v) = \frac{f(k,v)}{f(v)} \tag{6.2}$$

Thus, in order to calculate the wanted variable of $f(k|v)$, the probabilities $f(k,v)$ **(1)** and $f(v)$ **(2)** need to be calculated:

1. Starting with $f(k,v)$, the probability is bounded within the subset of the possible real speed values, which are according to our measured speeds from 0 km/h to 60 km/h. So, the probability $f(k,v)$ of a speed and a density occurring can be expressed using the term of the marginal probability:

$$f(k,v) = \int_0^{60} f(k,v,v_{real})dv_{real} \tag{6.3}$$

Using again the conditional probability as in Equation 6.2, this can be written as:

$$f(k,v) = \int_0^{60} f(k,v|v_{real})f(v_{real})dv_{real} = \int_0^{60} f(k|v,v_{real})f(v|v_{real})f(v_{real})dv_{real} \tag{6.4}$$

In this probability, if v_{real} is given, the information about v is not necessary to estimate the probability of k, so this can be simplified to:

$$f(k,v) = \int_0^{60} f(k|v_{real})f(v|v_{real})f(v_{real})dv_{real} label \tag{6.5}$$

2. Next, $f(v)$ can be calculated using the term of the marginal probability as before, leading to:

$$f(v) = \int_0^{60} f(v,v_{real})dv_{real} \tag{6.6}$$

Using the conditional probability leads to:

$$f(v) = \int_0^{60} f(v|v_{real})f(v_{real})dv_{real} \tag{6.7}$$

Hence, combining Equations 6.5 and 6.7, the wanted quantity $f(k|v)$ is equal to:

$$f(k|v) = \frac{\int_0^{60} f(k|v_{real})f(v|v_{real})f(v_{real})dv_{real}}{\int_0^{60} f(v|v_{real})f(v_{real})dv_{real}} \tag{6.8}$$

$f(v_{real})$ is a uniform distribution and is independent of the derivation at dv_{real}, so it can be simplified from both the nominator and the denominator, resulting in:

$$f(k|v) = \frac{\int_0^{60} f(k|v_{real})f(v|v_{real})dv_{real}}{\int_0^{60} f(v|v_{real})dv_{real}} \tag{6.9}$$

In this equation, all the variables are known, so it can be solved. The wanted probability density function is calculated over a discretized grid of speed values. The resulting probability density function can be seen in the next section.

6.2.3. Resulting Uncertainty levels

Combining the probability density functions of the Density-Speed version of the MFD $f(k|v_{real})$ and of the speed data $f(v|v_{real})$, the uncertainty level of the estimated density given a speed value was calculated. The resulting probabilities $f(k|v)$ of the estimated density given the speed can be seen in Figures 6.9 and 6.10 from two different perspectives.

The resulting probabilities show that the error bound of the estimated density, so the value of $\pm 2\sigma$, increases as the density increases and the speed decreases. The red curves in Figure 6.10, which

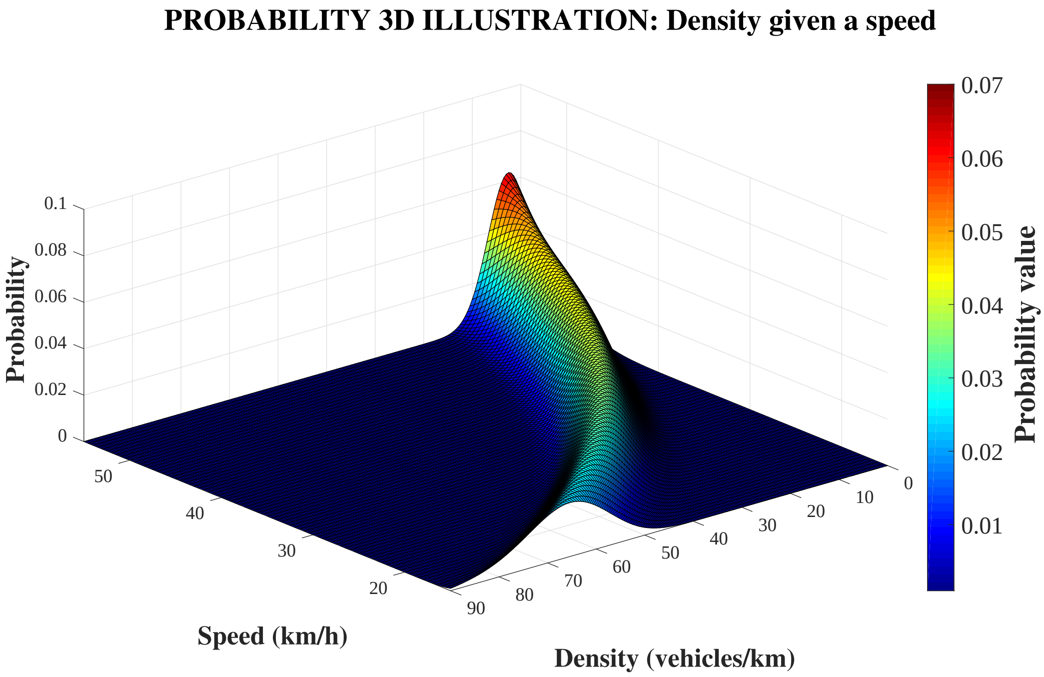


Figure 6.9: Probability of the estimated density values given speed measurements: View a

indicate the probability at 1% (or 0.01), are further from each other on the high-density and low-speed part, so in the congested state. This can also be observed by the fact that in the low-density and high-speed part, so in the free flow state, the probability value is higher, so the 95% confidence bounds are narrower.

This indicates that when the traffic is in the congested state, the estimated density has lower confidence levels, compared to the higher speeds. Thus, the estimated densities are more accurate when the network is in the free-flow state. This should be expected, because there are more variations in the speed throughout the network when it is congested leading to more difficulties in the network-wide traffic state estimation. Another reason that this occurs is because the obtained data fusion MFD resulted from data points with much higher variation in the congested part than in the free-flow part.

A more detailed representation of the probabilities of the density errors given specific speed values can be seen in the next section.

6.3. Examples of Traffic State Estimation

In order to analyse more thoroughly the resulting uncertainty levels of the previous section, some example probabilities of the density values resulting given specific speed measurements are drawn in Figure 6.11.

As it can be seen in the label of Figure 6.11, if the given speed is 45 km/h, the derived network density is 11 ± 9 vehicles/km at 95% confidence level. If the given speed is 35 km/h, the traffic density estimation is at 22 ± 10 vehicles/km. Whereas if the given speed is lower, at 15 km/h, the derived network density is 65 ± 17 vehicles/km at 95% level of confidence.

If we focus only on the confidence intervals of the estimated densities (± values), we conclude that the traffic state estimation is more trusted in the free flow state, because the interval is smaller. How-

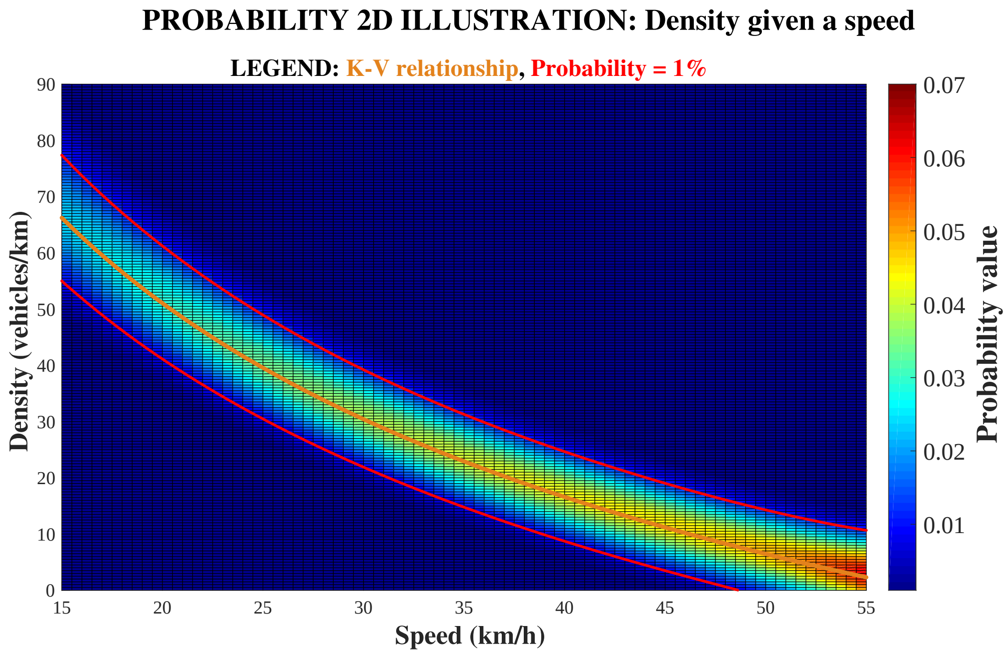


Figure 6.10: Probability of the estimated density values given speed measurements: View b

ever, if we compute the relative errors for each speed case, we see that the relative error is smaller in the congested state compared to the higher speeds (26% error for a speed of 15 km/h and 82% error for a speed of 45 km/h). However, in free flow the density values are relatively low, so even if the highest error occurs, it will still be possible to know that the network is in the free flow state.

What is important to realise here is that the interpretation of the errors depends on the purpose that the traffic state estimation will be used for. The level of accuracy and the type of error that is needed depends on the case that the traffic is estimated for. If it is only desired to know whether the network is in free flow or in congested state, the confidence intervals can provide the necessary information. If it is needed to know the exact amount of vehicles in the network, then the relative errors are more important. Nevertheless, the fact that the confidence interval of the estimated density given any speed can be expressed, is a very useful way to express the resulting traffic state estimation and the certainty that it encompasses.

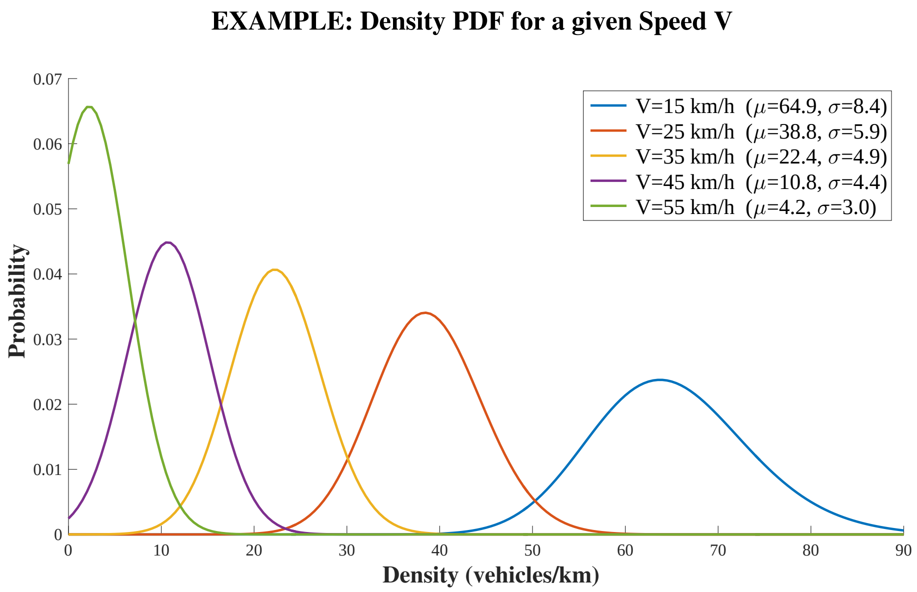


Figure 6.11: Examples of probabilities of estimated densities given the speed measurements

7

Traffic State Estimation in the case of an Incident

In Chapter 6, detailed examples were given showing that the proposed traffic state estimation can provide accurate results. The traffic state estimation was examined in cases with various demand levels, up to the maximum possible demand level. Nevertheless, the question that rises here is how well the proposed process would work in the case that something special happens in the network, for example an accident, a festival or any other kind of incident. In order to investigate this matter, incident cases are simulated and the estimation process is tested again. Within this perspective, the following research question was formed and will be answered in this chapter:

How is it possible to know if the Macroscopic Fundamental Diagram is not valid in the case of an incident?

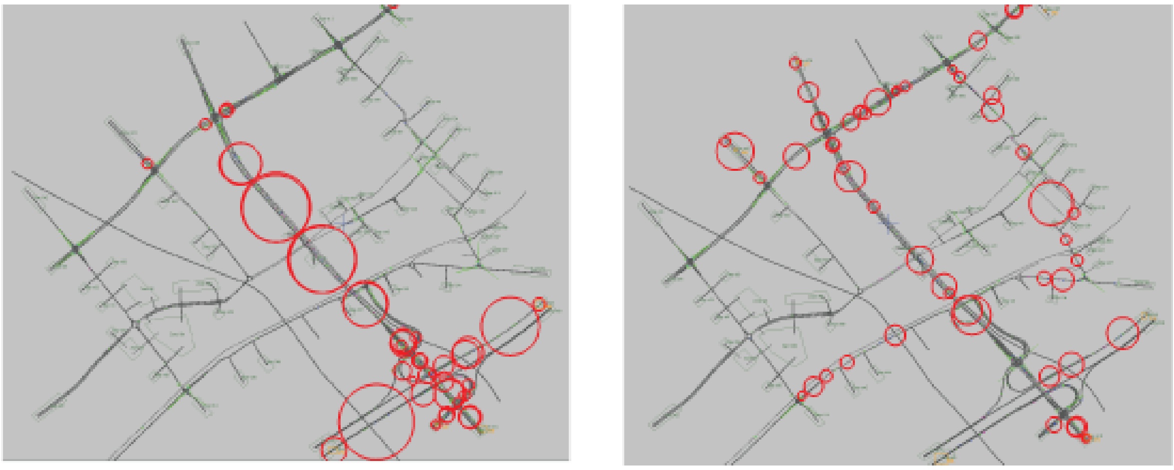
In order to answer this last sub-question, the simulation of the incident cases and the traffic situation that they cause, will be described in Section 7.1. Then, in Section 7.2, the way to know that the estimated traffic state is not valid in the special case of an incident will be described.

7.1. Traffic situation due to the incident

In order to investigate the practical use of the traffic state estimation in the case of an incident, seven spot incidents are simulated. The locations of the incidents in the network can be seen in Figure 7.1. The road incidents are simulated for one hour from 07:00-08:00. During the road incidents, one lane is completely blocked, reducing the capacity of the respective road segment. The incidents are placed in spots that are already highly congested (intersection with the four incidents A, C, D and E), in the highway (G), in the network center (B) and in a smaller road of less significance (F).

The resulting traffic situation in the case of the incident at spot G can be seen in comparison to the regular traffic condition in Figure 7.2. As it can be seen, in the case of the incident, there is very high congestion (the red circles indicate the size of the queue) around the incident location. Whereas, in the no incident case, congestion is spread more homogeneously throughout the network. In Figure 7.3,the resulting traffic state due to each one of the simulated incidents can be seen on the data fusion MFD that we obtained in the previous chapter (Figure 6.1). As we can see, the incidents in the locations A, C, E and G have data points that lie outside the 95% error bounds of the MFD. This means that these incidents have caused a situation in the network with really high congestion levels. These cases cannot be described sufficiently from the proposed traffic state estimation process,

Figure 7.1: Location of the seven simulated incidents A-G in the network



Incident Spot G No Incident

Figure 7.2: Comparison between the traffic situation in the case of incident and in regular conditions

because it will always give traffic states on the points of the MFD (black solid line in Figure 7.3). However, the real traffic states are quite different in some points of these cases, since they lie much lower than the MFD curve.

For example, in the case of the incident at spot A (dark blue line in Figure 7.3), there is a data point (indicated with magenta asterisk) with density of 61 vehicles/ km and flow of 560 vehicles/ hour. According to the proposed process, the way to estimate the traffic state is to use the speed from the available vehicle trajectories, which in that case is 9.7 km/h. So, if we follow the dashed magenta line, it indicates the point on the MFD that we receive, which has a density of 90 vehicles/km and a flow of 868 vehicles/hour. Thus, the network flow would be overestimated at a 1.6 times higher value (868:560 vehicles/hour). The false traffic state that we estimate is still in the congested part, but the flow is much lower than the real one. In case that the traffic state estimation is used for a purpose where high accuracy is required, this invalid estimation would be undesired. For this reason, the next section will look into a way to know that the MFD estimation of the traffic state is not valid.

7.2. Speed data during the incident

As it was seen in the previous section, the estimated traffic state in the cases of incidents A, C, E and G is off the reliable limits, because the estimated traffic state is not within the 95% confidence interval. This means that the proposed traffic state estimation process using the MFD is not valid for these cases.

As described in the 2nd step of the traffic state estimation process of Chapter 3, the traffic state is derived from the speed data of low fractions of vehicle trajectories. In Figure 7.4, a time series of the measured speed values during the simulation time is presented for every 5 minutes from 06:00 to 10:00. What we can see is that when the incident starts, at 07:00, there is a considerable decrease in the speeds of the spots A, C, E and G. The speeds of the rest of the incidents have a gradually decreasing behaviour, similar to the no-incident case (time series in magenta colour in Figure 7.4). The gradual decrease of the speed shows that there is, of course, higher congestion in the network, since the speed is decreasing, but nothing extraordinary happens. The speed decrease is most probably due to the peak hour traffic as in the no-incident case. For this same reason, the

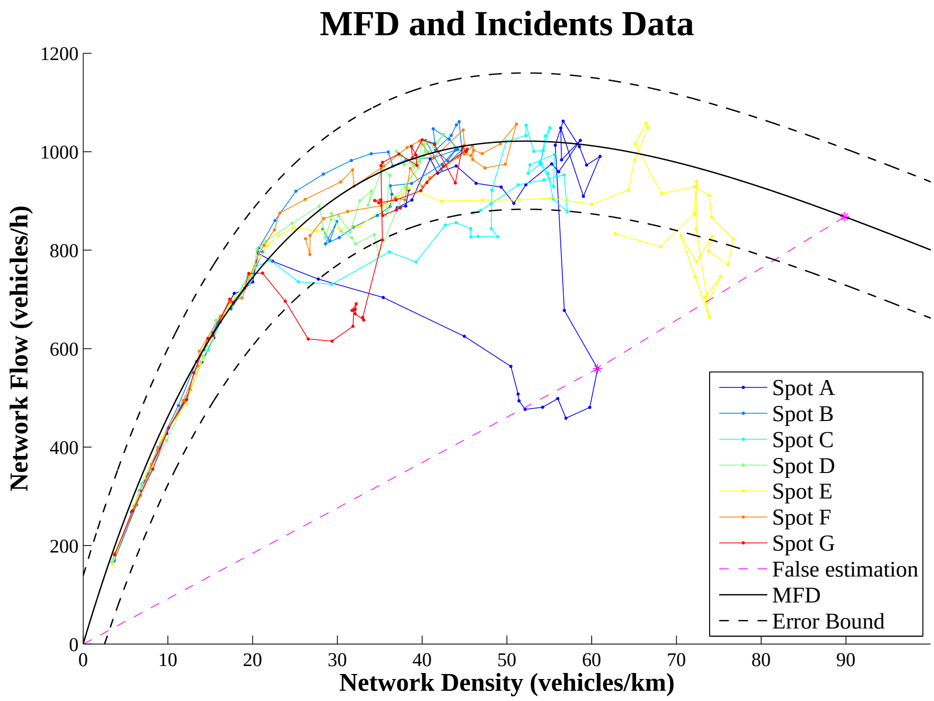


Figure 7.3: Data fusion MFD of Leidschendam-Voorburg with the incidents data

data points of the incidents B, D and F are within the MFD bounds in Figure 7.3.

The observation that the decrease of the speed is different between the various incidents means that the speed drop can be an indication of the validity of the estimated traffic state. In Figure 7.5, the percentage difference of the speed after every 5 minutes can be seen for each one of the incidents. As it can be seen, when the incident starts, there is a very sudden decrease in the speed for the incidents A, C and G. Specifically, the speed decrease is greater than or equal to 20% (indicated with dashed magenta line).

However, in the case of incident E (yellow colour), the speed drop does not show any extreme behaviour. A bit later in time, at 09:00, there is a high speed drop, but this may not be associated with the incident. Nonetheless, the estimated traffic state in the case of incident E is not as much outside of the MFD limits as the other cases (Figure 7.3). The moment that the estimation is off limits in the MFD is also the moment that the high speed drop occurs in case E.

Consequently, it was shown that when an incident happens in spots with low demand, the proposed process is still able to predict the traffic state. However, when the incident is located in roads with higher demand, the estimated traffic state is not accurate within the considered confidence level. In that case, it was found that a speed drop of more than 20% can indicate that there is an incident occurring in the network.

It should be noted that this conclusion cannot be used as a strict rule, since it is only based on the observation of the speed drop of seven incidents. Nevertheless, it can, undoubtedly, be considered as a rough indication that the MFD traffic state is not valid and further data are necessary to confirm the accuracy of the estimated traffic state.

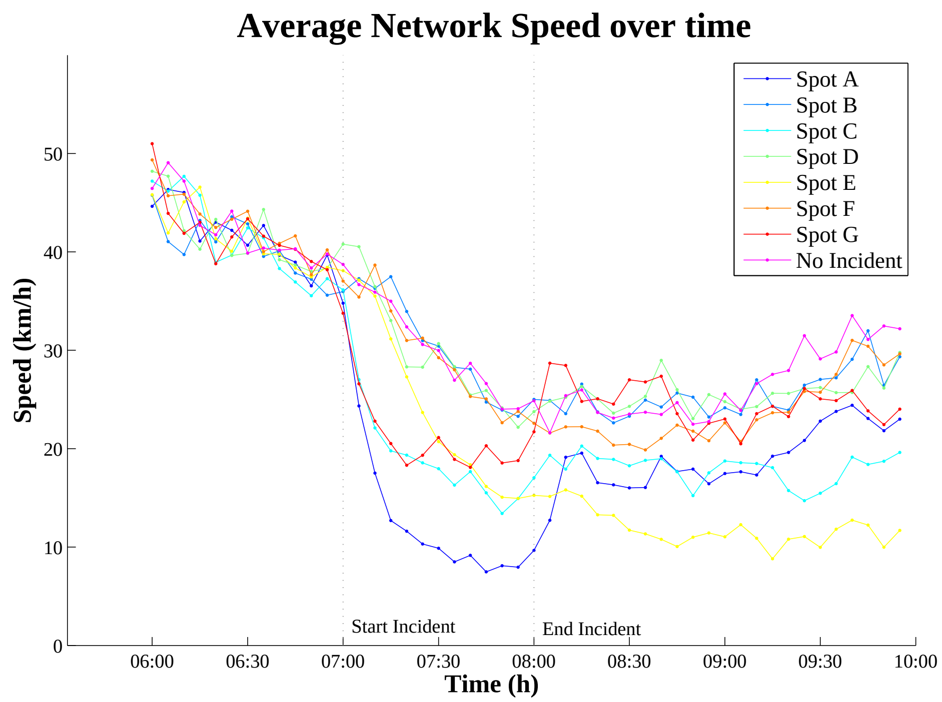


Figure 7.4: Speed measurements during the simulation of the incidents

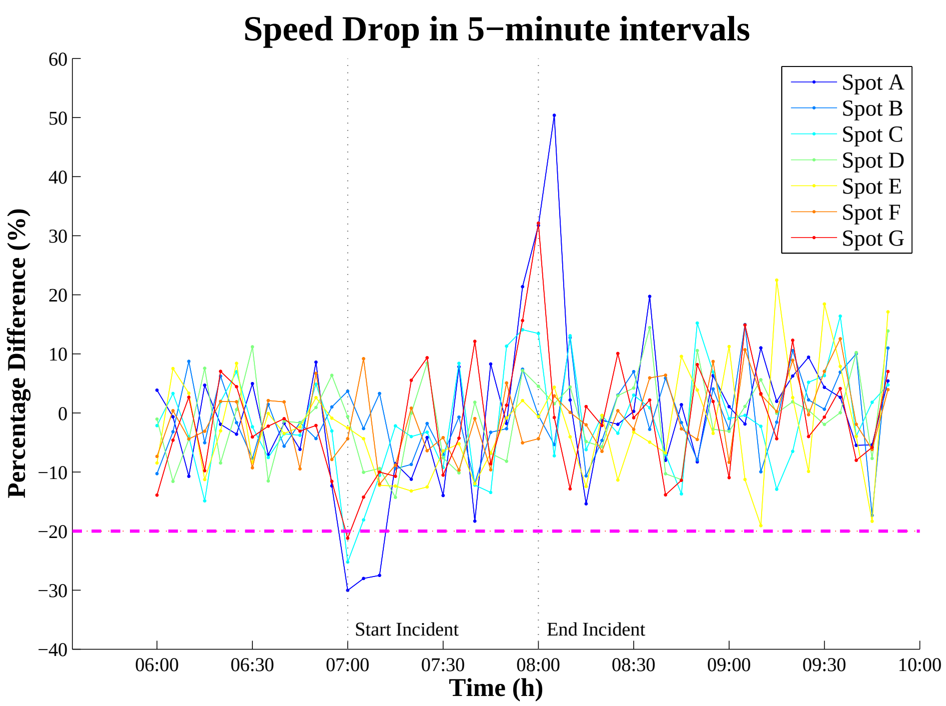


Figure 7.5: Percentage difference of the Speed drop every 5 minutes during the simulation of the incidents

8
Discussion

In this chapter, a discussion will be presented on the content of the thesis. In order to apply the proposed traffic state estimation, some simplifications were made or assumptions were taken that led to certain implications or drawbacks. These will be thoroughly discussed in Section 8.1. The same matters coming up while deriving the results, analysing the errors and testing the process will be discussed next, in Section 8.2.

8.1. Discussion of the Application

In Chapter 3, we proposed a 2-step traffic state estimation process and then, applied it in a simulated road network in Chapter 4. Traffic simulation is preferred over reality as a testing environment, because real traffic data are not easily available, and even when they are, they have errors and malfunctions that create interferences. In order to investigate the required types and amounts of traffic data for the MFD, flexibility over the nature of the data is necessary. Simulation allows such flexibility, since the desired traffic data can be produced and moreover, the demand and congestion levels in the network can be easily alternated. Furthermore, the assumption that the simulation model is the ground-truth, provides an easy way to validate the results of the traffic state estimation process. As it was also seen in the literature review (Chapter 2), many traffic studies have also used simulation to investigate and support their methodologies. Nevertheless, since only simulation data are used to test the proposed traffic state estimation process, its applicability on real networks might be argued. This is definitely a matter that needs to be considered and it is highly recommended to also test the process on real networks, as it will also be elaborated in Section 9.2.

An aspect that can be considered as a simplification is the selection of Leidschendam-Voorburg as the test network, which was presented in section 4.1. This network has significant congestion levels causing serious issues to the area. However, it does not have such a large size. Therefore, the applicability of the process in larger networks might be questioned. Nonetheless, the main focus of this research is on the data requirements of the network-wide traffic state estimation using the MFD. Hence, a reasonable-sized network, as the one that was chosen, is considered sufficient to fulfil the requirements of this project.

As it has been presented in section 4.3, the loop detector data of the simulated network were used only to extract the flow and not the speed and the density. This is due to the fact that the detectors are right upstream of the traffic light and hence, cannot capture adequately the speed and density levels more upstream in the road. Regarding the 1st step of the process, the data fusion approach

that was applied to obtain the MFD, requires only flow data from the detectors and not the density. In the 2nd step, detector speeds could have theoretically been used to indicate the traffic state on the MFD, but this was not made possible because of the issue with the detectors' location.

At this point, it was considered to move the detectors of the simulation model more upstream. However, this was regarded out of the scope of the research, because then, the question would rise on where to place the detectors. As literature has already suggested in Buisson and Ladier (2009) and in Courbon and Leclercq (2011), the detectors' location can have a significant impact on the network speed estimation and it is not easy to determine the ideal location to place the detectors. This is a problem that could have also occurred in real networks and in that case it would not be considered to just change the location of the detectors. For this reason, it seemed more appropriate to focus on using the reliable speed from the vehicle trajectories and only the flows from the detectors. Undoubtedly, the number of the detectors needs to be sufficient and in key locations of the network, so that the detector flows are including, if not all, at least almost all of the vehicles in the network.

On the one hand, the fact that it was not made possible to use the detector speeds can be seen as a drawback, since detector data are available in almost every traffic network and have been used for many years in traffic research. On the other hand, the fact that the proposed traffic state estimation needs only the detector flows that can be fully trusted is a very significant advantage of the process. This advantage of the proposed process should be emphasized here, because in this way, we utilize the detector data and fallacies are avoided.

8.2. Discussion of the Results and the Error Analysis

The application of the proposed traffic state estimation resulted in the MFD presented in Chapter 5. The MFD formula (Equation 5.4) that was found to fit the data seems a bit complicated to explain at first glance. Many functions were tested in order to find the best fit. The main reason that this formula type was chosen is because of the different trend that the two branches-the free flow and the congested branch-of the MFD have. The free flow branch increases steeply whereas the congested branch is more concave. A polynomial function is not possible to capture this alternating trend that the MFD curve has, but the addition of the exponential term managed to do so.

After obtaining the MFD formula, an extensive error analysis was performed in Chapter 6. The uncertainty of the traffic state estimation was investigated by combining the uncertainties that the MFD and the speed data encompass. At this point, some assumptions were made in order to be able to calculate the uncertainty of the final estimated traffic state. The uncertainty of the data fusion MFD and the speed data were expressed with the probability density functions of the errors of the respective variables. Both of these probability density functions were assumed to follow a Gaussian distribution. The resulting histograms of the errors showed that their shape is not too far from the bell shape of the Gaussian distribution. Hence, it can be considered as a logical assumption.

The resulting traffic state estimation was further tested in the case of an incident in Chapter 7. For this chapter, seven locations in the network were chosen to simulate incidents. The choice of the locations was mainly based on the congestion levels and the road types. Of course more locations could have been chosen and different types of incidents could have been simulated. However, that would be time consuming and out of the scope of the research subquestion that was investigated. The point was to create extreme traffic situations falling out of the estimated MFD bounds. As it can be seen in Figure 7.3, the requested point was succeeded. Hence, it was considered unnecessary to simulate other incident types or other locations.

9

Conclusions and Recommendations

This last chapter presents the conclusions and the recommendations of this thesis project. In Section 9.1, all the main findings and conclusions that were drawn during the thesis project are described. Next, recommendations both for practical use and for further research are presented in Section 9.2.

9.1. Conclusions

The objective of this thesis project was to estimate the traffic state network-wide using the Macroscopic Fundamental Diagram (MFD). In order to fulfil the objective, a 2-step traffic state estimation process was proposed and applied to the simulated network of Leidschendam-Voorburg in Paramics. In the 1st step of the process, the MFD is obtained fusing vehicle trajectories and loop detector data. In the 2nd step of the process, speed data are used to indicate the traffic state on the MFD. With the proposed process, it was managed to estimate the traffic state of the network using only a low amount of vehicle trajectories data and loop detector data.

In order to obtain the MFD, the subsets of vehicle trajectories were divided with the detector flows to calculate the fraction of the known vehicle trajectories to the total number of vehicles. The average network density and the average network flow were calculated dividing the density and the flow from the vehicle trajectories to the fraction of the known vehicle trajectories. In this way, the valuable and reliable information that even a small fraction of vehicle trajectories can offer was adjusted to represent the situation in the entire network. Accordingly, answering also to the first sub-question, the types of the required data to obtain the MFD are vehicle trajectories and detector flows. The necessary amount of vehicle trajectories can be any available subset of them, as long as the total detector flows are available.

In the next step of the process, the obtained data fusion MFD was combined with the fundamental relationship of the flow to derive the relationship between the density and the speed for the network. The density-speed relationship is a one-to-one function, which means that there is no density value that can be paired with more than one speed value. This property was taken into advantage to estimate the network density given only the speed at the network. The speed data that were used were collected from different penetration rates of vehicle trajectories varying from 1% to 30%. Consequently, answering also to the second sub-question, speed data extracted from the available number of vehicle trajectories were sufficient to indicate the traffic state that the network is performing on the MFD.

The results from the application of the traffic state estimation on the simulated network were used to validate the proposed process. Starting with the 1^{st} step of the process, the obtained data fusion MFD was compared to the ground-truth MFD, obtained from 100% vehicle trajectories. The comparison of the two MFDs showed that they were very similar and even shared many common points in the free flow branch. Next, the estimated network densities of the 2^{nd} step of the process were compared to the real densities that were observed in the network. The comparison showed that the higher the fraction of the known vehicle trajectories was, the lower the error was. However, even in the case that only 1% of vehicle trajectories was used, the estimation error was only 26%. When the fraction of vehicle trajectories was 30%, which was the higher fraction that was tested, the error was very low at 8%.

The comparison of the results with the real density values showed that the estimated traffic state is valid. Nevertheless, it is not possible to always validate the estimated traffic state in this way, due to data limitations. For this reason, a detailed error analysis was performed to estimate the confidence level of the estimated traffic state and express its validity. Both the obtained data fusion MFD and the speed measurements have errors that affect the estimation of the traffic state. In order to quantify this effect, the probability density function of the errors of the MFD was combined with the probability density function of the errors of the speed data to produce the joint probability that a density will occur in the network.

For any given speed measurement, the density value was estimated with a confidence interval at 95% level of certainty. The resulting probabilities showed that the error bound of the derived density is higher in the congested state compared to the error bound when the speed is in the free flow state. For instance, if the given speed is 45 km/h, the derived network the derived network density is 11 $\pm$ 9 vehicles/km at 95% confidence level of certainty. Whereas, if the given speed is 15 km/h, the derived network 65 $\pm$ 17 vehicles/km at 95% confidence level of certainty. A possible explanation that the estimation of the traffic state is more accurate when the network is in the uncongested state is that there are more speed variations in the network when it is congested. However, even in that case, the error is not very high, but whether it is acceptable or not depends on the purpose that the MFD will be used for. Thus, answering to the third sub-question, despite the uncertainties that both the MFD and the speed data encompass, the proposed process offers a robust and valid traffic state estimation.

Nonetheless, the accuracy of the traffic state estimation was only tested in cases with higher demand. In order to further examine the precision of the process, it was also tested in the case that an incident occurred in the network. Seven locations within the network were chosen to simulate incidents by blocking one lane for one hour during the simulation period. The four out of the seven incidents caused a traffic state outside the error bounds of the data fusion MFD. These incidents were mainly located in spots with high demand. The incidents that occurred in roads with low demand caused a less severe situation that was still within the 95% confidence interval of the MFD. In the three out of the four cases that the real traffic state was outside the error bounds, it was observed that the speed drop after 5 minutes was over 20%. Thus, answering also to the fourth sub-question, it was concluded that a sudden speed drop can act as a sign that an incident has happened in the network and the estimations on the MFD are not valid anymore.

Overall, this thesis project successfully managed to estimate the traffic state using the MFD. A simple data fusion process with low data requirements was used to obtain the MFD. In this way, a reliable foundation was created which in combination with solely speed data was used to estimate the traffic state at any desired moment. The results of this project are highly encouraging to further utilise the proposed traffic state estimation process.

9.2. Recommendations

At points of the thesis project, additional directions to follow were considered. Moreover, ideas were generated on how this project could further continue. All these can act as suggestions for practical use of the traffic state estimation or for further scientific research and they will be elaborated in this section.

Recommendations for Practice

1. **Input to traffic control mechanisms:** the initial idea on why traffic state estimation is important brings us back to the general cycle of the traffic system. The utmost goal is to optimise the traffic system, thus we recommend to use the proposed traffic state estimation as an input to the traffic control mechanisms of the network. This recommendation can also indicate how efficiently the proposed traffic state estimation serves its main purpose.

2. **Evaluation of traffic control strategies:** another way that the proposed traffic estimation can be beneficial to practice is as an evaluation tool of applied traffic control strategies. The comparison of the traffic state on the MFD before and after the application of specific strategies can offer insight on their effectiveness. In this way, traffic control strategies can be compared and validated and this can act as a decision tool for traffic managers.

3. **Evaluation of network changes:** significant network changes, such as the addition of new lanes or the closing of roads for construction can have critical impact on the traffic levels of a network. In that case, it is necessary to evaluate the effect of the change and assess the drawbacks that may occur. The MFD of the network may change due to the network alterations, so also the optimal capacity point will increase or decrease depending on the change. The effect on the maximum capacity on the MFD can be used as the necessary evaluation tool.

4. **Support of policy-making decisions:** in the case that a new policy is considered, information is needed on the effect of the policy on traffic. Usually, the main scope of the policy is to improve the safety or the liveability of a municipality or a city, but the impact on the traffic also needs to be taken into consideration. The MFD can be used to provide the necessary information about not only whether the policy will have a positive or a negative impact on the traffic flow, but also on how large the effect will be.

Recommendations for Further Research

1. **Application to other networks:** a suggestion that could contribute to the potential generalization of the process is to apply the traffic state estimation to other urban networks. In the case that the networks are very large, network partitioning techniques could be used to produce more than one MFD for the parts that the network is divided. The application of the proposed process to other networks can also provide an indication on whether the MFD can be well defined in any urban network.

2. **Application with real data:** the use of real traffic data could offer a very strong hypothesis on the applicability of the MFD traffic state estimation with any data types. The application of the proposed process with simulation data offered very promising results. Nevertheless, the quality of the simulation data is better than the real traffic data. For this reason, it is recommended to use loop detector data and vehicle trajectories from a real network to apply the same traffic state estimation process.

3. **Use of other types of speed data**: another recommendation regarding the traffic data, corresponds to the use of speed data at the 2nd step of the traffic state estimation. Within this project, the necessary speed data were derived from the vehicle trajectories. Nevertheless,

theoretically, other data sources such as loop detectors or camera data could also provide the required speed data. Such data could be used to validate their ability to indicate the traffic state on the MFD.

4. **Data fusion with Kalman filter:** a more advanced data fusion technique, such as the application of Kalman filter could offer a more accurate estimation and a mathematically stronger way to take into consideration the uncertainty of the estimated quantities. In order to apply Kalman filter techniques, alterations are required to the formulation of the problem in successive time steps. The resulting estimation could offer further insights on the robustness of the traffic state with the MFD.

5. **Inclusion of the deviation of density:** as literature has suggested (Knoop and Hoogendoorn, 2013), the spatial dispersion of the density could offer further insights on the traffic state in the network. Within this project, this aspect was investigated when detecting the incidents and it was concluded that indeed, the spatial variation of density could prove to be a beneficial addition in the accuracy of the MFD. Therefore, it is suggested to further research the spatial attribute of the network density. This could be achieved by incorporating the variation of the density throughout space with k-means clustering methods.

6. **Effect of OD-matrix and signal timing:** within the literature review that was performed, various characteristics that can affect the shape of the MFD were found. One of these was the format of the traffic signals. This parameter was not examined within the application that we performed in this project. Hence, it is suggested as an additional test of the shape definition of the MFD. Furthermore, the OD-matrix was not alternated and this is another aspect that could affect the shape of the MFD. Checking the effect of such parameters could promote the establishment of the MFD as a solid and reliable foundation for traffic state estimation.

Bibliography

Buisson, C. and Ladier, C. Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams. *Transportation Research Record: Journal of the Transportation Research Board*, (2124):127–136, 2009.

Cassidy, M., Jang, K., and Daganzo, C. Macroscopic fundamental diagrams for freeway networks: theory and observation. *Transportation Research Record: Journal of the Transportation Research Board*, (2260):8–15, 2011.

Courbon, T. and Leclercq, L. Cross-comparison of macroscopic fundamental diagram estimation methods. *Procedia-Social and Behavioral Sciences*, 20:417–426, 2011.

Daamen, W., Buisson, C., and Hoogendoorn, S. P. *Traffic Simulation and Data: Validation Methods and Applications*. CRC Press, 2014.

Daganzo, C. F. Urban gridlock: Macroscopic modeling and mitigation approaches. *Transportation Research Part B: Methodological*, 41(1):49–62, 2007.

Daganzo, C. F. and Geroliminis, N. An analytical approximation for the macroscopic fundamental diagram of urban traffic. *Transportation Research Part B: Methodological*, 42(9):771–781, 2008.

De Jong, D., Knoop, V. L., and Hoogendoorn, S. P. The effect of signal settings on the macroscopic fundamental diagram and its applicability in traffic signal driven perimeter control strategies. In *Intelligent Transportation Systems-(ITSC), 2013 16th International IEEE Conference on*, pages 1010–1015. IEEE, 2013.

Drosg, M. *Dealing with uncertainties: a guide to error analysis*. Springer Science & Business Media, 2009.

Du, J., Rakha, H., and Gayah, V. V. Deriving macroscopic fundamental diagrams from probe data: Issues and proposed solutions. *Transportation Research Part C: Emerging Technologies*, 2015.

Edie, L. C. *Discussion of traffic stream measurements and definitions*. Port of New York Authority, 1963.

Gayah, V. V. and Dixit, V. V. Using mobile probe data and the macroscopic fundamental diagram to estimate network densities: Tests using microsimulation. *Transportation Research Record: Journal of the Transportation Research Board*, (2390):76–86, 2013.

Geroliminis, N. and Boyacı, B. The effect of variability of urban systems characteristics in the network capacity. *Transportation Research Part B: Methodological*, 46(10):1607–1623, 2012.

Geroliminis, N. and Daganzo, C. F. Macroscopic modeling of traffic in cities. In *TRB 86th annual meeting*, number 07-0413, 2007.

Geroliminis, N. and Daganzo, C. F. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological*, 42(9):759–770, 2008.

Geroliminis, N. and Levinson, D. M. Cordon pricing consistent with the physics of overcrowding. In *Transportation and Traffic Theory 2009: Golden Jubilee*, pages 219–240. Springer, 2009.

Geroliminis, N. and Sun, J. Properties of a well-defined macroscopic fundamental diagram for urban traffic. *Transportation Research Part B: Methodological*, 45(3):605–617, 2011.

Geroliminis, N., Haddad, J., and Ramezani, M. Optimal perimeter control for two urban regions with macroscopic fundamental diagrams: A model predictive approach. *Intelligent Transportation Systems, IEEE Transactions on*, 14(1):348–359, 2013.

Godfrey, J. The mechanism of a road network. *Traffic Eng. Control*, 11:323–327, 1969.

Ji, Y. and Geroliminis, N. On the spatial partitioning of urban transportation networks. *Transportation Research Part B: Methodological*, 46(10):1639–1656, 2012.

Ji, Y., Daamen, W., Hoogendoorn, S. P., Hoogendoorn-Lanser, S., and Qian, X. Investigating the shape of the macroscopic fundamental diagram using simulation data. *Transportation Research Record: Journal of the Transportation Research Board*, (2161):40–48, 2010.

Keyvan-Ekbatani, M., Kouvelas, A., Papamichail, I., and Papageorgiou, M. Exploiting the fundamental diagram of urban networks for feedback-based gating. *Transportation Research Part B: Methodological*, 46(10):1393–1403, 2012.

Knoop, V. L. *Road Incidents and Network Dynamics: Effects on driving behaviour and traffic congestion.* TU Delft, Delft University of Technology, 2009.

Knoop, V. L. and Hoogendoorn, S. P. Empirics of a generalized macroscopic fundamental diagram for urban freeways. *Transportation Research Record: Journal of the Transportation Research Board*, (2391):133–141, 2013.

Knoop, V. L., Hoogendoorn, S. P., and Van Lint, J. W. C. Routing strategies based on macroscopic fundamental diagram. *Transportation Research Record: Journal of the Transportation Research Board*, (2315):1–10, 2012.

Knoop, V. L., Van Lint, J. W. C., and Hoogendoorn, S. P. Traffic dynamics: Its impact on the macroscopic fundamental diagram. *Physica A: Statistical Mechanics and its Applications*, 438:236–250, 2015.

Laval, J. A. and Castrillón, F. Stochastic approximations for the macroscopic fundamental diagram of urban networks. *Transportation Research Part B: Methodological*, 81:904–916, 2015.

Leclercq, L. and Geroliminis, N. Estimating mfds in simple networks with route choice. *Transportation Research Part B: Methodological*, 57:46–484, 2013.

Leclercq, L., Chiabaut, N., and Trinquier, B. Macroscopic fundamental diagrams: A cross-comparison of estimation methods. *Transportation Research Part B: Methodological*, 62:1–12, 2014.

Mazloumian, A., Geroliminis, N., and Helbing, D. The spatial variability of vehicle densities as determinant of urban network capacity. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 368(1928):4627–4647, 2010.

Nagle, A. and Gayah, V. Accuracy of networkwide traffic states estimated from mobile probe data. *Transportation Research Record: Journal of the Transportation Research Board*, (2421):1–11, 2014.

Ortigosa, J., Menendez, M., and Tapia, H. Study on the number and location of measurement points for an mfd perimeter control scheme: a case study of zurich. *EURO Journal on Transportation and Logistics*, 3(3-4):245–266, 2014.

Paramics. *S-Paramics 2014 Reference Manual*. Paramics Microsimulation, SIAS Limited, 37 Manor Place Edinburgh EH3 7EB UK.

Rijkswaterstaat. *Bereikbaarheidsmonitor hoofdwegennet 2006*, 2007.

Simoni, M. D., Pel, A. J., Waraich, R. A., and Hoogendoorn, S. P. Marginal cost congestion pricing based on the network fundamental diagram. *Transportation Research Part C: Emerging Technologies*, 56:221–238, 2015.

Smeed, R. J. Road capacity of city centers. *Traffic Engineering and Control*, 8(7):455–458, 1966.

Taale, H. *Integrated anticipatory control of road networks: a game-theoretical approach*. TU Delft, Delft University of Technology, 2008.

Tsubota, T., Bhaskar, A., and Chung, E. Macroscopic fundamental diagram for brisbane, australia: empirical findings on network partitioning and incident detection. *Transportation Research Record: Journal of the Transportation Research Board*, (2421):12–21, 2014.

Van Lint, J. W. C. *Integrated and Coordinated Network management*. TU Delft, 2014.

Yildirimoglu, M., Ramezani, M., and Geroliminis, N. Equilibrium analysis and route guidance in large-scale networks with mfd dynamics. *Transportation Research Part C: Emerging Technologies*, 59:404–420, 2015.

Zahavi, Y. Traffic performance evaluation of road networks by the α-relationship. *Traffic Engineering and Control*, 14(5-6), 1972.