Relative Flow Data: New Opportunities for Traffic State Estimation

van Erp, Paul

DOI
10.4233/uuid:4fa16188-7431-4a6e-8097-5bde9cabc466

Publication date
2020

Document Version
Final published version

Citation (APA)
van Erp, P. (2020). *Relative Flow Data: New Opportunities for Traffic State Estimation*. [Dissertation (TU Delft), Delft University of Technology]. TRAIL Research School. https://doi.org/10.4233/uuid:4fa16188-7431-4a6e-8097-5bde9cabc466

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Relative Flow Data

New Opportunities for Traffic State Estimation

Paul Bernardus Cornelis van Erp

Delft University of Technology, 2020

This PhD study was supported by the Netherlands Organisation for Scientific Research (NWO), grant-number: 022.005.030.

Cover illustration and design: Paul van Erp & Wouter Horstink

Relative Flow Data
New Opportunities for Traffic State Estimation

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology,
by the authority of the Rector Magnificus Prof.dr.ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates,
to be defended publicly on
Monday 3 February 2020 at 10:00 o'clock

by

Paul Bernardus Cornelis VAN ERP,
Master of Science in Econometrics and Management Science,
Erasmus University Rotterdam, the Netherlands
Master of Science in Civil Engineering, Delft University of Technology, the Netherlands
born in 's-Hertogenbosch, the Netherlands.

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus	chairperson
Prof.dr.ir. S.P. Hoogendoorn	Delft University of Technology, promotor
Dr. V.L. Knoop	Delft University of Technology, promotor
Independent members:	
Prof.dr.ir. C.F. Daganzo	University of California, Berkeley, United States
Prof.dr. L. Leclercq	Université de Lyon, IFSTTAR, France
Prof.dr. M. Menendez	NYU Abu Dhabi, United Arab Emirates
Prof.dr.ir. C.M.J. Tampère	Katholieke Universiteit Leuven, Belgium
Prof.dr.ir. J.W.C. van Lint	Delft University of Technology
Prof.dr.ir. B. De Schutter	Delft University of Technology, reserve member

TRAIL Thesis Series T2020/1, The Netherlands TRAIL Research School

TRAIL
P.O. Box 5017
2600 GA Delft
The Netherlands
E-mail: info@rsTRAIL.nl

ISBN: 978-90-5584-260-5

Keywords: Relative flow data; Traffic state estimation.

Printed in the Netherlands

Preface

This thesis marks the end of my PhD journey. A journey that started while working on my MSc. thesis under the supervision of Serge, Victor, Yufei and Kai. Serge and Victor asked me to consider the option of pursuing a PhD at the department where I was completing my MSc. degree (Transport & Planning at Delft University of Technology). At the moment, I was also working part-time at MICompany, and I was unsure whether I wanted to take the extra step in academia, work in industry or pursue a different path. However, at the same time, the unique opportunity arose to apply for a PhD-grant with a personal idea. I decided to apply for this NWO-grant (which was awarded by the research school TRAIL) and was accepted for the grant. The ability to work on a topic of my choosing, and having the flexibility to work four days a week on the PhD while pursuing other activities on the side, convinced me to start my PhD in September 2015.

In my PhD, I have explored the opportunity to estimate the traffic conditions with data that can be collected with automated vehicles. Here, I aim to explain which data are valuable to estimate the traffic conditions and how these data can be used for this purpose. Working on this topic allowed me to use my backgrounds in Transport & Planning (Delft University of Technology) and Econometrics (Erasmus University Rotterdam). Econometrics thought me to understand and work with data and estimation techniques. Transport & Planning taught me to understand different descriptions of traffic flow and traffic flow properties. The ability to combine this double-background and contribute to a topic that has clear practical applications (i.e., estimates of the traffic conditions are used for traffic management purposes and for navigational purposes) has motivated me throughout the past years.

Academic research should create building blocks for future academic research and practical applications. An academic study thus does not need to have direct practical applications; however, the possibility of practical applications (and societal benefits) should be there, even if the path leading to such applications (and benefits) is vague at the moment. As stated above, my field of study (i.e., estimating the traffic conditions) has direct practical applications (e.g., traffic management) and can lead to societal benefits (e.g., reduce congestion). However, my studies can be seen as rather theoretical (and fundamental). This is a conscious choice as I aim to create good building blocks for academia and practice. In my view, introducing new options and describing general principles yields more valuable building blocks than, for instance, algorithmic specifications and specific case studies. The case studies are a way to gain a deeper insight in the possibilities that are offered by developed methodologies and the principles on which these methodologies are based.

Academics or (broader) the academic process are valuable for society when used properly. When facing problems, the academic process can provide a structured approach to explore and find a solution. Within this approach, it is important to keep an open mind and explore all (foreseeable) options. Sometimes this requires taking the risk to explore a path that may not lead anywhere; however, in my view, this should not be seen as a (potential) waste. Without taking risks, it is unlikely that we (as a society) expose new opportunities that may in the end prove to be highly valuable. In my PhD studies, I have explored new opportunities. These studies mainly considered the option to use a specific type of data that can be collected using automated vehicles (i.e., relative flow data) to estimate the traffic state and reveal traffic flow properties. Automated vehicles will be able to collect more and new types of data that are currently not available on a large scale. The choice to work with and explore the (traffic state estimation) opportunities with relative flow data has been based on experiences with using other data types and an understanding of traffic flow theory (domain knowledge). My studies show why the relative flow data are valuable, and introduce new principles and methodologies that can be applied to these data (and some other specific combinations of data). Therefore, this thesis provides new opportunities for practitioners and thereby increases their choice set related to using new and/or different data and traffic state estimation methodologies. However, it will require an open-mind to collect these data and explore the practical use of the data, principles and methodologies. A presumption that may have to be relaxed by practitioners is that all we need is better speed estimates. This presumption restricts the option of estimating the cumulative flow over space and time, which provides a more complete macroscopic description of the traffic conditions. For me, this has similarities with a famous quote of Henry Ford (with a personal touch at the end): 'If I had asked people what they wanted, they would have said *better speed estimates*'.

Pursuing and completing a PhD will influence you for the rest of your life. It results in a critical mindset and trying to find the logic behind the things we encounter in our own lives and what we hear in the news. It can teach you to put things in perspective, but it can also make you highly skeptical about certain affairs. This also holds for me. I am skeptical that we can move from our current unsustainable society to a sustainable society. Not because of the technological challenges that we face, but because the major challenge lies in our economic system (and the related fields of sociology and politics). Of course, I hope that I am wrong and that we are able to tackle these major challenges ahead. Again, academics and having an academic attitude may play an important role in tackling these challenges. I hope that all people do not try to solely find that one piece of information (e.g., news article) that fits their current beliefs, but are open to new/other information. This is a lot to ask as nobody is immune to the natural tendency to respond positively to information that confirms existing beliefs, while dismissing opposite information. Recognizing this human problem (that goes against the academic attitude) is a first step.

I did not pursue my PhD on a full-time basis. As stated above, the ability to work four days a week and perform other activities on the side was a prerequisite for me to start the PhD. In this 'extra' time, I worked on real-estate projects, at Arane consultants in traffic and transportation and on developing own entrepreneurial ideas. For me, the variety in working activities was beneficial as it can be a welcome distraction and (in some cases) can provide synergies. My future is still uncertain. I like flexibility, I like variety, I like to explore new (intellectual) paths and prefer not to be restricted to much by factors that do not relate to the content. Academics may be in the mix, entrepreneurship may be in the mix, we will see.

Before, during (and probably after) my PhD, I have had a lot of support from family, friends and colleagues (at Delft University of Technology and Arane). To prevent that I forget to thank specific people, I will only name of few persons. As mentioned above, Victor and Serge played an important role prior to and during my PhD. Therefore, I would like to thank them for their support during this period. Furthermore, I am happy to have a strong relationship with my family and thankful for their support, so I would like to thank my dad & mum, Simone & Laurens, since last year Marie & Julie, grandparents (which unfortunately are no longer all among us to witness the completion of my PhD), uncles & aunts, and cousins. A PhD can be quite intense. It is your PhD and you alone can finish it. It requires working hard, being able to deal with 'failure', and getting the most energy out of your working activities and the occasional successes. The latter (i.e., successes) are occasional as you are working for a long period towards major goals (e.g., being accepted for conferences, getting a paper published and ultimately successfully defending your thesis). It is nice to have family, friends and colleagues to discuss challenges, and (to a greater extent) simply relax and have fun with. Therefore, I like to thank all my family, friends and colleagues!

Paul Bernardus Cornelis van Erp,
Delft, December 2019.

Summary

Traffic state information is crucial for different applications, e.g., in design and operation of road traffic networks, and in navigation services. Traffic sensing data, e.g., loop-detector data, may directly provide the desired information. Alternatively, the traffic state information may be estimated with data that only provides partial and noisy information. To apply this process, i.e., traffic state estimation, we have to make choices related to which data are collected and how these are processed.

The macroscopic traffic state can be described using the variables flow, density and mean speed, where flow is equal to the product of density and mean speed. Edie's generalized definitions of traffic flow define these three variables for spatial-temporal areas. Alternatively, traffic flow can be described using the three dimensions space, time and cumulative flow. The cumulative flow is defined as the cumulative number of vehicles that have passed a position at a specific time, which means that it is a discrete variable. However, the discrete function can be smoothed over space and time. In this case, the macroscopic variables flow and (negative) density can be determined for points in space-time by taking the derivatives to time and space of the smoothed cumulative flow function.

In this thesis, a distinction is made between microscopic and macroscopic traffic sensing data. Examples of microscopic traffic sensing data are probe individual speed data and spacing data. Macroscopic data can describe Edie's generalized definitions of traffic flow for spatial-temporal areas, e.g., probe mean speed data or aggregated double loop-detector data. Alternatively, macroscopic sensing data can describe the change in cumulative flow between points in space-time, e.g., detector count data or relative flow data.

The scientific gaps addressed in this thesis are subdivided in four parts that relate to each other. First, we evaluate the errors that are induced when estimating the mean speed for spatial-temporal areas based on error-free data. This provides insight in the errors that arise due to incomplete information and incorrect assumptions when estimating the mean speed. Second, the option to use probe data to mitigate the cumulative count error problem is considered. This problem occurs when estimating the cumulative flow curves based on (stationary) detector data. For this purpose, both probe mean speed and probe trajectory data are used. The probe mean speed data relates to the first part as they describe the mean speed for spatial-temporal areas. If relative flow observations are added to the probe trajectory data, relative flow data from moving observers that are part of the traffic flow are obtained. In the third part, these relative flow data are used to estimate the traffic state. In this part, different combinations of observers are used, which includes stationary observers, moving observers that are part of the traffic flow and moving observers that travel in opposing direction. To estimate the traffic state with relative flow data, streaming-data-driven and

model-driven estimation approaches are considered. In a model-driven estimation approach historical data are used to expose traffic flow models. Therefore, we address the possibility to use historical relative flow data to expose these model. The fourth and final part relates the option that road authorities collect personal traffic sensing data (e.g., probe trajectory and/or relative flow data) directly from road-users. In other parts of this thesis, we designed methodologies to use these data, which may be valuable for road authorities. Therefore, it is interesting to investigate how road authorities can gain access to these personal data. Below, all four parts are discussed individually.

Mean speed estimation with aggregated detector data and probe individual speed data

Edie's mean speed can be estimated for spatial-temporal areas with aggregated detector data and/or probe individual speed data. To gain insight in the estimation error characteristics, Edie's mean speeds are estimated based on error-free data. This evaluation shows that errors are induced by incomplete information and potentially incorrect assumptions. These estimates are sometimes presented as data, e.g., probe mean speed data. Understanding the estimation errors can be valuable in further applications of these estimates. For instance, within estimation methodologies that fuse information (in the form of data and models) based on its error characteristics.

Mitigation of the cumulative error problem with probe data

If the cumulative flow curves at two locations are known, the vehicle accumulation (i.e., number of vehicles) and mean travel time between these two locations can be determined. To estimate the cumulative flow curves at detector locations, detector count data can be used. The cumulative flow curves are perfectly estimated if the initial number of vehicles between the detector locations is known and the detector count data are error-free. However, in case of count errors, the cumulative flow estimates will become less accurate over time due to the cumulative nature of the variable. This problem is denoted as the cumulative error problem and needs to be mitigated to obtain accurate estimates of the cumulative flow curves and variables that can be derived from it (e.g., vehicle accumulations). In this thesis, we consider the option to mitigate this problem with probe mean speed and trajectory data.

A methodology is presented that mitigates the cumulative error problem and estimates the vehicle accumulation based on detector count data and probe mean speed data. In this methodology the probe data are used for two processes, i.e., recovery of cumulative error and learning the combined structural detector count error. We show that the methodology successfully prevents errors to build up over time. Combining the two types of data with the proposed methodology yields accurate estimates in a simulation case study, and is successful in observing peaks in the vehicle accumulation in an empirical study. However, we also show that the data provide incomplete information on the cumulative flow curves, and cannot expose the count errors relating to individual detectors. Therefore, data that contain more information, e.g., probe trajectory data, may be more valuable in mitigating the cumulative error problem.

Prior studies have proposed the idea to mitigate the cumulative error problem with probe trajectory data. These studies assume that the probe vehicles do not overtake other vehicles, which would mean that the cumulative flow value along the probe trajectory is constant.

However, this assumption is likely to be violated in multi-lane traffic. Improving on the assumption of 'no overtaking', while using probe trajectory data to mitigate the cumulative error problem is expected to beneficial. Therefore, we present a methodology to estimate the change in cumulative flow over a probe trajectory between detector locations based on disaggregated detector data. In this methodology, the probe-specific relative flow is estimated at the detector locations based on the individual probe speed and detector data-based macroscopic state estimates. Next, based on the probe-specific relative flows at the two detectors locations, the change in cumulative flow over the probe trajectory between these two locations is estimated. This approach clearly improves upon the assumption of 'no overtaking' in free-flow conditions, while the benefits are less clear in congested conditions. However, in both cases, the estimates are not perfect, which indicates that it is better to observe the change in cumulative flow over trajectories, i.e., collect relative flow data with moving observers.

Traffic state estimation with relative flow data

Relative flow data can be collected with stationary and moving observers. To assess whether these data provide valuable information on the traffic state, we presented multiple traffic state estimation methodologies and principles. Here, both streaming-data-driven and model-driven estimation approaches are considered. In the latter approach, the data are used in two processes, i.e., learning the (traffic flow) models based on historical data, and estimating the traffic state by assimilating real-time data and (traffic flow) models. Therefore, both processes are addressed in this thesis.

This thesis present a streaming-data-driven estimation methodology to estimate Edie's flow and density using relative flow data collected using stationary and moving observers. Combining relative flow data from stationary observers that are positioned at the link boundaries with moving observers that are part of the traffic flow, yields observations of the change in cumulative flow between points in space and time that lie on different observation paths. In the proposed methodology, we subdivide space-time in triangular areas for which the change in cumulative flow between the three corners are observed. Based on the change in cumulative flow, space and time, the flow and density are estimated for these triangular areas. Next, the state estimates related to these areas can be mapped to any estimation mesh (e.g., a rectangular estimation mesh that consists of road segments and time periods). This methodology is based on Edie's generalized definitions of traffic flow and only uses a single (non-crucial) parameter. In a simulation case study, we show that flow and density can be estimated accurately and are able to outperform the current Dutch loop-detector data at low penetration rates (i.e., between 2.5 % to 5.0 %).

Historical relative flow data collected with moving observers can be used to estimate the fundamental diagram of traffic flow. The paths over which the relative flow is observed are denoted as observation paths. If moving observers observe both their own and the opposite direction of traffic, the moving observers travel with positive and negative speeds through space and time. In this case, it is realistic that the observations paths of different observers enclose areas in space-time. For these areas the change in cumulative flow over all boundaries are observed, which allows us to apply Edie's definitions to estimate flow and density. We use these flow-density estimates to estimate the parameters of a triangular fundamental diagram. The methodology is able to accurately estimate the fundamental diagram for the

simple car-following models evaluated in the simulation case study. However, the designed methodology would need be extended to estimate the diagram for more complex driver behavior.

To assess which spatial-temporal relative flow data characteristics are valuable in model-driven traffic state estimation, principles are presented for the two processes that are part of this estimation approach. These processes are: (1) learn (traffic flow) models based on historical data and (2) estimate the traffic state by assimilating real-time data and models. These principles assume that traffic flow follows the LWR-model with triangular fundamental diagram. The principles show that it is valuable that the observers that collect relative flow data cross each other in space-time. To cross each other, observers need to travel at different speeds. Combining stationary observers positioned at the link boundaries (and potentially other locations) with moving observers is the most advantages scenario as the moving observers connect the observation paths of the stationary observers, and the stationary observers connect the observation paths of the moving observer. If we rely on moving observers alone, observers can still cross each other in space-time if they travel at different speeds, e.g., fast and slow observers, and/or observers that observe the opposing traffic flow. Furthermore, to learn the traffic flow model and to apply it for traffic state estimation, it is important that the different (potential) characteristics waves can be drawn from observation paths to a large spatial-temporal area. For this purpose, it is again beneficial to have observers that travel at different speeds. Therefore, collecting relative flow data with moving observers is a good addition or alternative to collecting relative flow data with stationary observers when estimating the traffic state using a model-driven estimation approach.

Traffic flow models describe the traffic flow properties. As stated above, relative flow data may be used to learn these traffic flow models. We presented an approach to reveal the traffic flow properties based on relative flow data collected by stationary observers and moving observers that are part of the traffic flow. In this approach, the relative flow data are used to obtain wave observations (which are a combination of the change in cumulative flow, space and time) from points in space-time to the link boundaries. Based on the wave observations different plots can be constructed. To interpret these plots and reveal the traffic flow properties, variational theory and its assumptions are considered. Therefore, we assume that the traffic flow properties can be modeled using the LWR-model. The conducted simulation study shows that the approach reveals a fundamental diagram based on relative flow data. Furthermore, it allows us to evaluate the non-linearity of and stochasticity related to fundamental diagram branches. The presented approach can be the basis for algorithms that learn the traffic flow model in an online setting.

Road authorities access to personal traffic sensing data

The methodologies, principles and approaches presented in this thesis can be applied by road authorities for traffic state estimation purposes in dynamic traffic management applications. This requires that road authorities have access to the traffic sensing data that they want to use for estimation. Probe trajectory and relative flow data are personal traffic sensing data as they describe features that relate to individual road users. Due to privacy concerns and legislation, road authorities may have difficulties to obtain these data via third parties. Alternatively, road authorities may gain direct access to these personal data from road-users. A stated preference study is presented to gain insight in the road users' willingness to share

their personal traffic sensing data with road authorities for the purpose of dynamic traffic management. This study is limited in number of respondents and representativity for the population (Dutch road users), and is therefore not generalized to whole population. However, it does provide initial insights in the features that affect the willingness to share personal data. The respondents show a natural preference to share personal data for application from which society can benefit (e.g., sharing data with road authorities, emergency services and research institutions). Sharing more detailed trip data is valued negatively, but respondents do not seem to make a difference between probe trajectory data and relative flow data. The respondents value being in control for which trips their data are shared. Furthermore, the desire to be in control of the use of their data is indicated by the unwillingness to share data with unnamed third parties. Finally, the survey indicates that respondents can be persuaded to share their data by offering a monetary compensation.

Conclusions and outlook

This thesis shows that relative flow data have preferable characteristics for traffic state estimation compared to other traffic sensing data. Relative flow data can be collected with stationary observers (which observe a fixed position over space-time) such as loop-detectors or road-side cameras. Furthermore, automated and/or other connected and equipped vehicles that serve as moving observers can collect relative flow data and move at different speeds than the stationary observers. The observed variable (i.e., the change in cumulative flow) describes the core macroscopic traffic flow variable. Combing relative flow data from observers that travel at different speeds makes it possible that observers cross each other in space-time. This allows us to relate the change in cumulative flow observations of different observers to each other and describe the change in cumulative flow between points in space-time. As shown in the different studies related to traffic state estimation with relative flow data, the combination of having observers traveling at different speeds and observing the change in cumulative flow is beneficial to estimate the traffic state. Relative flow data collected with moving observers provide more valuable information on the traffic state than probe trajectory data, which in multi-lane traffic flow miss observations of the change in cumulative flow.

The work presented in this thesis can have some important implications for practice. Manufacturers of automated and/or other equipped and connected vehicles can collect relative flow data. This requires them to develop the algorithms needed to observe passings and their directions with respect to their vehicles. They can use these data for their own applications, e.g., navigation services or new applications that may become valuable in the era of autonomous vehicles. Furthermore, they could share the data with road authorities. If road authorities gain access to the relative flow data and these data proof to be valuable for their applications, road authorities may change their data collection system. They may be able to reach their objectives with less road-side detectors.

Multiple methodologies, principles and approaches were presented in this thesis. These can be extended in future research. For instance, the principles and approaches (e.g., those that relate to revealing the traffic flow properties) can be extended to obtain methodologies that can learn traffic flow models in a online setting. Furthermore, it can be interesting to investigate other topics than estimating the traffic state on a link. For this purpose, we may investigate the ability to reveal node flow properties using relative flow data.

Samenvatting

Het schatten van de verkeerstoestand is cruciaal voor verschillende toepassingen, bijvoorbeeld bij het ontwerpen en opereren van wegverkeersnetwerken, en voor navigatie diensten. Verkeersdata, bijvoorbeeld lusdata, kan direct de gewenste informatie over de verkeerstoestand geven. Tevens kan de verkeerstoestand worden geschat op basis van onnauwkeurige data die de verkeerstoestand slechts gedeeltelijk beschrijft. In het proces om de verkeerstoestand te schatten moeten we keuzes maken met betrekking tot welke data worden verzameld en hoe deze worden verwerkt.

De macroscopische verkeerstoestand kan worden beschreven met de variabelen intensiteit, dichtheid en gemiddelde snelheid, waarbij de intensiteit gelijk is aan het product van de dichtheid en gemiddelde snelheid. Edie's gegeneraliseerde definities van verkeersstroom definiëren deze drie variabelen voor gebieden in ruimte en tijd. Tevens kan de verkeerstoestand worden beschreven met behulp van de drie dimensies ruimte, tijd en 'cumulatieve flow'. De cumulatieve flow is gedefinieerd als het cumulatief aantal voertuigen dat op een bepaald tijdstip een locatie is gepasseerd, wat betekent dat het een discrete variabele is. De discrete functie kan echter over tijd en ruimte worden afgevlakt. In dit geval kunnen de macroscopische variabelen intensiteit en (negatieve) dichtheid worden bepaald voor punten in ruimte en tijd door de afgeleiden naar tijd en ruimte van de afgevlakte cumulatieve flow functie te bepalen.

In dit proefschrift wordt een onderscheid gemaakt tussen microscopische en macroscopische verkeersdata. Voorbeelden van microscopische verkeersdata zijn individuele snelheid data en volgafstand data. Macroscopische verkeersdata kunnen Edie's gegeneraliseerde definities van verkeersstroom voor gebieden in ruimte-tijd beschrijven, bijvoorbeeld probe gemiddelde snelheid data of geaggregeerde (dubbele) lusdata. Tevens kunnen macroscopische verkeersdata de verandering in de cumulatieve flow tussen punten in ruimte-tijd beschrijven, bijvoorbeeld lus-tellingen of 'relatieve flow data'.

De (wetenschappelijke) hiaten die geadresseerd worden in dit proefschrift zijn onderverdeeld in vier delen die aan elkaar relateren. Ten eerste evalueren we de fouten die worden veroorzaakt bij het schatten van de gemiddelde snelheid voor gebieden in ruimte-tijd op basis van perfecte data (dus zonder meetfouten). Dit geeft inzicht in de fouten die ontstaan als gevolg van onvolledige informatie en onjuiste veronderstellingen bij het schatten van de gemiddelde snelheid. Ten tweede wordt de optie overwogen om probe data (welke ook bekend staan als floating car data) te gebruiken om het cumulatieve telfoutprobleem te reduceren. Dit probleem treedt op bij het schatten van de cumulatieve flow op basis van (stationaire) lusdata. Voor dit doel worden zowel probe gemiddelde snelheid data als probe trajectorie data gebruikt. Het gebruik van probe gemiddelde snelheid data relateert aan het

eerste deel omdat deze data de gemiddelde snelheid beschrijven voor gebieden in ruimte en tijd. In het derde deel worden deze relatieve flow data gebruikt om de verkeerstoestand te schatten. Als relatieve flow observaties worden toegevoegd aan de probe trajectorie data, worden relatieve flow data verkregen van bewegende waarnemers die deel uitmaken van de verkeersstroom. In dit deel worden verschillende combinaties van waarnemers gebruikt, waaronder stationaire waarnemers (bijvoorbeeld lussen), bewegende waarnemers die onderdeel zijn van de verkeersstroom en bewegende waarnemers die in tegengestelde richting reizen. Om de verkeerstoestand met relatieve flow data te schatten, worden streaming-data-driven en model-driven schattingsaanpakken in overweging genomen. In een model-driven schattingsaanpak worden historische gegevens gebruikt om verkeersstroommodellen bloot te leggen. Daarom bekijken we de mogelijkheid om historische relatieve flow data te gebruiken om zulke modellen bloot te leggen. Het vierde en laatste deel heeft betrekking op de optie dat wegbeheerders persoonlijke verkeersdata (bijvoorbeeld probe trajectorie data en/of relatieve flow data) rechtstreeks van weggebruikers verzamelen. In andere delen van dit proefschrift hebben we methodologieën ontwikkeld om deze data te gebruiken. Daarom is het interessant om te onderzoeken hoe wegbeheerders toegang kunnen krijgen tot deze persoonlijke data. Hieronder worden alle vier delen afzonderlijk besproken.

Gemiddelde snelheidsschatting met geaggregeerde lusdata en probe individuele snelheid data

De gemiddelde snelheid van Edie kan worden geschat voor gebieden in ruimte-tijd met geaggregeerde lusdata en/of probe individuele snelheid data. Om inzicht te krijgen in de schattingsfout-karakteristieken, wordt Edie's gemiddelde snelheid geschat op basis van perfecte verkeersdata (dus zonder meetfouten). Uit deze evaluatie blijkt dat fouten worden veroorzaakt door onvolledige informatie en mogelijk onjuiste veronderstellingen. De resulterende schattingen worden soms gepresenteerd als data, bijvoorbeeld probe gemiddelde snelheid data. Inzicht in de schattingsfouten kan waardevol zijn bij verdere toepassingen van deze schattingen. Bijvoorbeeld binnen schattingsmethoden die informatie (in de vorm van data en modellen) fuseren op basis van de foutkarakteristieken.

Reduceren van het cumulatieve fouten probleem met probe data

Als de cumulatieve flow curven op twee locaties bekend zijn, kan de voertuigaccumulatie (d.w.z. het aantal voertuigen) en de gemiddelde reistijd tussen deze twee locaties worden bepaald. Om de cumulatieve flow curven op lus-locaties te schatten, kunnen lus-tellingen worden gebruikt. De cumulatieve flow curven worden perfect geschat als het initiële aantal voertuigen tussen de lus-locaties bekend is en de lus-tellingsdata foutloos zijn. In het geval van telfouten worden de cumulatieve flow curven echter in de loop van de tijd minder nauwkeurig vanwege het cumulatieve karakter van de variabele en de fouten. Dit probleem wordt aangeduid als het cumulatieve fouten probleem en moet worden gereduceerd om nauwkeurige schattingen te verkrijgen van de cumulatieve flow curven en variabelen die daarvan kunnen worden afgeleid (bijvoorbeeld voertuigaccumulaties). In dit proefschrift bekijken we de mogelijkheden om dit probleem te reduceren met twee soorten probe data, i.e., probe gemiddelde snelheid data en probe trajectorie data.

We presenteren een methodologie die het cumulatieve fouten probleem vermindert en

de voertuigaccumulatie schat op basis van lus-tellingen en de probe gemiddelde snelheid data. In deze methodologie worden de probe data gebruikt voor twee processen, d.w.z., herstel van cumulatieve fouten en het leren van de gecombineerde structurele lus-telfout. We laten zien dat de methodologie met succes voorkomt dat fouten zich in de loop van de tijd opbouwen. Het combineren van de twee soorten data met de voorgestelde methodologie levert nauwkeurige schattingen op in een simulatie onderzoek en is succesvol in het waarnemen van pieken in de voertuigaccumulatie in een empirisch onderzoek. We laten echter ook zien dat de data onvolledige informatie bieden over de cumulatieve flow curven en dat de telfouten van individuele lussen niet kunnen worden blootgelegd. Daarom kunnen data die meer informatie bevatten, bijvoorbeeld probe trajectorie data, waardevoller zijn bij het verminderen van het cumulatieve fouten probleem.

Eerdere studies hebben het idee voorgesteld om het cumulatieve fouten probleem met probe trajectorie data aan te pakken. Deze studies gaan ervan uit dat de probe-voertuigen andere voertuigen niet inhalen, wat zou betekenen dat de cumulatieve flow waarde over hun trajectorie constant is. Het is echter waarschijnlijk dat deze aanname zal worden geschonden voor banen met meerdere rijstroken (waar voertuigen elkaar kunnen inhalen). Een verbetering ten opzichte van de aanname 'voertuigen halen elkaar niet in' zal ons in staat stellen om beter het cumulatieve tel probleem aan te pakken met probe trajectorie data. Daarom presenteren we een methodologie om de verandering in de cumulatieve flow over een probe trajectorie tussen lus-locaties te schatten op basis van individuele lus passages data. In deze methodologie wordt de probe-specifieke relatieve flow geschat op de lus-locaties met behulp van de individuele probe snelheid en op individuele lus passages data gebaseerde macroscopische toestandsschattingen. Vervolgens wordt op basis van de probe-specifieke relatieve flow op de twee lus-locaties de verandering in cumulatieve flow over probe trajectorie tussen deze twee locaties geschat. Deze benadering verbetert duidelijk de veronderstelling van 'voertuigen halen elkaar niet in' in free-flow condities, terwijl de voordelen minder duidelijk zijn in congestie. In beide gevallen zijn de schattingen echter niet perfect, wat aangeeft dat het beter is om de verandering in de cumulatieve flow over trajecten waar te nemen, d.w.z., relatieve flow data te verzamelen met bewegende waarnemers.

Verkeerstoestand-schatting met relatieve flow data

Relatieve flow data kunnen worden verzameld met stationaire en bewegende waarnemers. Om te beoordelen of deze data waardevolle informatie over de verkeerstoestand bevatten, hebben we meerdere methodologieën en principes om de verkeerstoestand te schatten gepresenteerd. Hiervoor worden zowel streaming-data-driven als model-driven schattingsaanpakken bekeken. In de laatste aanpak worden de data gebruikt in twee processen, d.w.z., het leren van de (verkeersstroom) modellen op basis van historische data, en het schatten van de verkeerstoestand door real-time data en (verkeersstroom) modellen te fuseren. Daarom worden beide processen in dit proefschrift behandeld.

Dit proefschrift presenteert een streaming-data-driven schattingsmethodologie om Edie's intensiteit en dichtheid te schatten met behulp van relatieve flow data verzameld met stationaire en bewegende waarnemers. Het combineren van relatieve flow data van stationaire waarnemers, die aan het begin (stroomopwaarts) en einde (stroomafwaarts) van de weg zijn gepositioneerd, met bewegende waarnemers die deel uitmaken van de verkeersstroom, levert waarnemingen op van de verandering in cumulatieve flow tussen punten in ruimte en

tijd op. In de voorgestelde methodologie delen we ruimte-tijd op in driehoekige gebieden waarvoor de verandering in de cumulatieve flow tussen de drie hoekpunten wordt waargenomen. Op basis van de verandering in cumulatieve flow, ruimte en tijd, worden de intensiteit en dichtheid geschat voor deze driehoekige gebieden. Vervolgens kunnen deze toestandsschattingen worden toegewezen aan elk gewenste opdeling van ruimte-tijd (bijvoorbeeld een rechthoekig opdeling die bestaat uit wegsegmenten en tijdsperioden). Deze methodologie is gebaseerd op de algemene definities van Edie en gebruikt slechts een enkele (niet-cruciale) parameter. In een simulatie case studie laten we zien dat intensiteit en dichtheid nauwkeurig kunnen worden geschat en dat we in staat zijn om de huidige Nederlandse lusdata te overtreffen bij lage penetratiegraden (d.w.z. tussen 2,5 % tot 5,0 %).

Historische relatieve flow data verzameld met bewegende waarnemers kunnen worden gebruikt om het fundamentele diagram van de verkeersstroom te schatten. De paden waarover de relatieve flow wordt waargenomen, worden aangeduid als observatiepaden. Als bewegende waarnemers zowel hun eigen als de tegenovergestelde rijrichting waarnemen, reizen de bewegende waarnemers met positieve en negatieve snelheden door ruimte en tijd. In dit geval is het realistisch dat de observatiepaden van verschillende waarnemers gebieden in ruimte-tijd omsluiten. Voor deze gebieden wordt de verandering in de cumulatieve flow over alle randen waargenomen, waardoor we de definities van Edie kunnen toepassen om de intensiteit en de dichtheid te schatten. We gebruiken deze intensiteit-dichtheid-schattingen om de parameters van een driehoekig fundamenteel diagram te schatten. De methodologie is in staat om het fundamentele diagram voor de eenvoudige voertuig-volgmodellen (die in de simulatie case studie zijn geëvalueerd) nauwkeurig te schatten. De ontworpen methodologie zou echter moeten worden uitgebreid om het diagram te schatten voor meer complex rijgedrag.

Om te beoordelen welke ruimtelijke en temporele relatieve flow data karakteristieken waardevol zijn bij model-driven verkeerstoestand-schattingen, worden principes gepresenteerd voor de twee processen die deel uitmaken van deze schattingsaanpak. Deze processen zijn: (1) blootleggen van de (verkeersstroom) modellen op basis van historische data en (2) verkeerstoestand schatten door real-time data en modellen te fuseren. Deze principes veronderstellen dat de verkeersstroom het LWR-model met driehoekig fundamenteel diagram volgt. De principes tonen aan dat het waardevol is dat de waarnemers die relatieve flow data verzamelen elkaar in kruisen in ruimte-tijd. Om elkaar te kruisen, moeten waarnemers met verschillende snelheden reizen. Het combineren van stationaire waarnemers gepositioneerd aan het begin en einde van de weg zijn gepositioneerd (en mogelijk andere locaties) met bewegende waarnemers is het scenario met de meeste voordelen, aangezien de bewegende waarnemers de observatiepaden van de stationaire waarnemers verbinden en de stationaire waarnemers de observatiepaden van de bewegende waarnemer verbinden. Als we alleen op bewegende waarnemers willen vertrouwen, kunnen waarnemers elkaar in nog steeds in ruimte-tijd kruisen als ze met verschillende snelheden reizen, bijvoorbeeld in het geval van snelle en langzame waarnemers, en/of waarnemers die ook de tegengestelde rijrichting waarnemen. Verder is het, om het verkeersstroommodel te bloot te leggen en toe te passen voor het schatten van de verkeerstoestand, van belang dat de verschillende (potentiële) karakteristieke-golven kunnen worden getrokken van observatiepaden naar een groot gebied in ruimte-tijd. Voor dit doel is het wederom voordelig om waarnemers te hebben die met verschillende snelheden reizen. Daarom is het verzamelen van relatieve flow data met bewegende waarnemers een goede toevoeging of alternatief voor het verzamelen van relatieve

flow data met stationaire waarnemers bij het schatten van de verkeerstoestand met behulp van een model-driven aanpak.

Verkeersstroommodellen beschrijven de verkeersstroomeigenschappen. Zoals hierboven vermeld, kunnen relatieve flow data worden gebruikt om deze verkeersstroommodellen bloot te leggen. We hebben een aanpak gepresenteerd om de verkeersstroomeigenschappen te onthullen op basis van relatieve flow data die zijn verzameld door stationaire waarnemers en bewegende waarnemers die deel uitmaken van de verkeersstroom. In deze aanpak worden de relatieve stroomgegevens gebruikt om golfwaarnemingen te verkrijgen (die een combinatie zijn van de verandering in cumulatieve flow, ruimte en tijd) van punten in ruimte-tijd naar de weggrenzen. Op basis van de golfwaarnemingen kunnen verschillende plots worden geconstrueerd. Om deze plots te interpreteren en de verkeersstroomeigenschappen te onthullen, wordt rekening gehouden met variationele theorie en de bijbehorende aannames. Daarom nemen we aan dat de verkeersstroomeigenschappen kunnen worden gemodelleerd met behulp van het LWR-model. Ons simulatieonderzoek toont aan dat de benadering een fundamenteel diagram onthult op basis van relatieve stroomgegevens. Bovendien stelt het ons in staat om de niet-lineariteit van en stochasticiteit met betrekking tot de verschillende fundamentele diagram takken (free-flow en congestie) te evalueren. De gepresenteerde aanpak kan de basis vormen voor algoritmen die het verkeersstroommodel leren.

De toegang van wegbeheerders tot persoonlijke verkeersdata

De methodologieën, principes en aanpakken die in dit proefschrift worden gepresenteerd, kunnen door wegbeheerders worden gebruikt om de verkeerstoestand te schatten. Vervolgens kunnen deze schattingen worden gebruikt voor dynamische verkeersmanagement toepassingen. Dit vereist dat wegbeheerders toegang hebben tot de verkeersdata die zijn bekeken in dit proefschrift. Probe trajectorie en relatieve flow data zijn persoonlijke data, omdat deze functies beschrijven die betrekking hebben op individuele weggebruikers. Vanwege privacy-kwesties en wetgeving kan het moeilijk zijn voor wegbeheerders om deze data via derden te verkrijgen. Als alternatief kunnen wegbeheerders directe toegang tot deze persoonlijke data krijgen van weggebruikers. Een ‘stated preference’ studie is gepresenteerd om inzicht te krijgen in de bereidheid van de weggebruikers om hun persoonlijke data te delen met wegbeheerders met het oog op dynamische verkeersmanagement. Dit onderzoek is beperkt in aantal respondenten en representativiteit voor de bevolking (Nederlandse weggebruikers) en is daarom niet representatief voor de gehele bevolking. Het biedt echter wel eerste inzichten in de attributen die van invloed zijn op de bereidheid om persoonlijke data te delen. De respondenten tonen een natuurlijke voorkeur om persoonlijke data te delen voor toepassingen waarvan de samenleving kan profiteren (bijvoorbeeld data delen met wegbeheerders, hulpdiensten en onderzoeksinstellingen). Het delen van meer gedetailleerde reisgegevens wordt negatief gewaardeerd, maar respondenten lijken geen verschil te maken tussen probe trajectorie data en relatieve flow data. De respondenten waarderen het om te bepalen voor welke reizen hun data worden gedeeld. Bovendien wordt de wens om controle te hebben over het gebruik van hun data aangegeven door de onwil om data te delen met niet nader genoemde derden. Ten slotte geeft de enquête aan dat respondenten kunnen worden overgehaald om hun gegevens te delen door een monetaire compensatie aan te bieden.

Conclusies en outlook

Dit proefschrift laat zien dat relatieve flow data waardevolle eigenschappen hebben voor het schatten van de verkeerstoestand in vergelijking met andere verkeersdata. Relatieve flow data kunnen worden verzameld met stationaire waarnemers (die een vaste positie in de ruimte-tijd observeren) zoals lussen of camera's langs de weg. Verder kunnen geautomatiseerde en/of andere 'connected' en met sensoren uitgevoerde voertuigen die dienst doen als bewegende waarnemers relatieve flow data verzamelen. Die bewegende waarnemers kunnen met andere snelheden bewegen dan de stationaire waarnemers. De geobserveerde variabele (d.w.z. de verandering in de cumulatieve flow) beschrijft de belangrijkste macroscopische verkeersstroomvariabele. Door relatieve flow data van waarnemers die met verschillende snelheden reizen te combineren, kunnen waarnemers elkaar kruisen in ruimte-tijd. Dit stelt ons in staat om waarneming van de verandering in cumulatieve flow van verschillende waarnemers aan elkaar te relateren en de verandering in cumulatieve stroom tussen punten in ruimte-tijd te beschrijven. Zoals aangetoond in de verschillende onderzoeken met betrekking tot de schatting van de verkeerstoestand met relatieve flow data, is de combinatie van het hebben van waarnemers die met verschillende snelheden reizen en het observeren van de verandering in de cumulatieve flow gunstig voor het schatten van de verkeerstoestand. Relatieve flow data verzameld met bewegende waarnemers bieden meer waardevolle informatie over de verkeerstoestand dan probe trajectorie data, die veranderingen in de cumulatieve flow missen, terwijl het zeer waarschijnlijk is dat deze wel voorkomen indien er meerdere stroken zijn.

Het werk gepresenteerd in dit proefschrift kan enkele belangrijke implicaties hebben voor de praktijk. Fabrikanten van geautomatiseerde en/of andere 'connected' en met sensoren uitgevoerde voertuigen kunnen relatieve flow data verzamelen. Dit vereist dat zij de algoritmen ontwikkelen die nodig zijn om passages en hun richting ten opzichte van hun voertuigen te observeren. Ze kunnen deze gegevens gebruiken voor hun eigen toepassingen, bijvoorbeeld navigatiediensten of nieuwe toepassingen die waardevol kunnen worden in het tijdperk van autonome voertuigen. Bovendien kunnen ze de data delen met wegbeheerders. Als wegbeheerders toegang krijgen tot de relatieve flow data en wordt bewezen dat deze data waardevol zijn voor hun toepassingen, kunnen wegbeheerders hun data verzamelingssystemen wijzigen. Dit leidt er misschien toe dat ze hun doelen kunnen bereiken met minder weg-kant data systemen (bijvoorbeeld lussen).

In dit proefschrift werden meerdere methodologieën, principes en aanpakken gepresenteerd. Toekomstig onderzoek kan hier verder op bouwen. De principes en aanpakken (bijvoorbeeld die welke betrekking hebben op het blootleggen van de verkeersstroomeigenschappen) kunnen bijvoorbeeld worden uitgebreid om methodologieën te verkrijgen die verkeersstroommodellen kunnen leren en deze modelschattingen continue te verbeteren. Verder kan het interessant zijn om andere onderwerpen te onderzoeken dan het schatten van de verkeerstoestand op wegvakken tussen discontinuïteiten van wegen. Voor dit doel kunnen we de mogelijkheid onderzoeken om knoopstroomeigenschappen te onthullen met behulp van relatieve flow data.

Contents

Preface **v**

Summary **ix**

Samenvatting **xv**

1 Introduction **1**
1.1 Background on traffic state estimation 2
1.1.1 Macroscopic description of traffic flow 2
1.1.2 Different estimation approaches 4
1.1.3 Traffic sensing data 4
1.2 Scientific gaps 6
1.3 Research objectives and questions 9
1.4 Contributions 11
1.4.1 Contributions to science 11
1.4.2 Contributions and relevance to practice 14
1.5 Thesis outline 16

2 Understanding traffic sensing data-based estimation errors **19**
2.1 Introduction 21
2.2 Variables used to describe the traffic conditions 22
2.3 Sensing data-based mean speed estimation 23
2.3.1 Traffic sensing data characteristics 23
2.3.2 Estimation approach 24
2.3.3 Estimation error distribution 24
2.4 Experimental set-up 24
2.4.1 Data collection 25
2.4.2 Ground truth 26
2.4.3 Evaluation of estimation error characteristics 26
2.5 Results and findings 30
2.5.1 Dependency on traffic conditions 30
2.5.2 Dependency on observed fraction of the traffic flow 32
2.6 Conclusions and discussion 34

3 Vehicle accumulation estimation with detector count and probe mean speed data 37
3.1 Introduction . . . 39
3.2 Recursive estimation of the vehicle accumulation and the cumulative error problem . . . 41
3.3 Methodology to estimate the vehicle accumulation . . . 41
3.3.1 Real-time high-frequency estimation . . . 43
3.3.2 Mean vehicle accumulation during a time-period . . . 44
3.3.3 Error recovery . . . 45
3.3.4 Online learning of the bias term . . . 47
3.4 Case studies . . . 50
3.4.1 Data collection and characteristics . . . 51
3.4.2 Vehicle accumulation estimation . . . 53
3.4.3 Evaluation of estimation performance . . . 54
3.5 Results . . . 56
3.5.1 Simulation study . . . 56
3.5.2 Empirical study . . . 61
3.6 Conclusions . . . 63

4 Est. of the change in cumulative flow over probe trajectories using detector data 65
4.1 Introduction . . . 67
4.2 Theoretical foundations . . . 69
4.2.1 Change in cumulative flow along a probe trajectory . . . 69
4.2.2 Differences in the change in cumulative flow between probe trajectories . . . 71
4.3 Methodology to estimate the change in cumulative flow between detectors . 72
4.4 Case study . . . 73
4.4.1 Traffic sensing data and traffic conditions . . . 73
4.4.2 Experimental set-up . . . 75
4.5 Results . . . 76
4.5.1 Simulation study . . . 76
4.5.2 Empirical study . . . 78
4.6 Conclusions and insights . . . 79

5 Traffic state estimation using relative flows from stationary and moving observers 83
5.1 Introduction . . . 85
5.2 Background on macroscopic traffic state estimation . . . 85
5.3 Point-observations of the cumulative vehicle number . . . 88
5.4 The PON estimation methodology . . . 89
5.4.1 Subdivide space-time into triangular areas . . . 89
5.4.2 Three-point traffic state estimation . . . 90
5.4.3 Space-time area traffic state estimation based on the mean boundary conditions . . . 91
5.4.4 Traffic state estimation in a pre-defined mesh . . . 95

5.5 Simulation study . . . 97
5.5.1 Microscopic simulation in FOSIM . . . 97
5.5.2 The reference traffic state estimator . . . 98
5.5.3 The PON traffic state estimator . . . 100
5.5.4 Evaluation of the estimators . . . 101
5.6 Results . . . 102
5.6.1 Estimation performance of the PON estimator . . . 102
5.6.2 Comparison between the PON and reference estimator . . . 103
5.7 Conclusions and discussion . . . 105

6 Estimating the fundamental diagram using moving observers 111
6.1 Introduction . . . 113
6.2 Background on fundamental diagram estimation . . . 113
6.3 Estimating flow and density for space-time areas using moving observers . 114
6.4 Estimating the fundamental diagram parameters . . . 115
6.5 Simulation study . . . 118
6.5.1 Microscopic simulation of traffic . . . 118
6.5.2 Evaluation of the proposed methodology . . . 120
6.6 Results . . . 121
6.7 Conclusions and outlook . . . 122

7 On the value of relative flow data 123
7.1 Introduction . . . 125
7.2 Collecting relative flow data with stationary and moving observers . . . 127
7.2.1 Potential data characteristics . . . 128
7.2.2 Data used in existing studies that estimate in or via the cumulative flow plane . . . 129
7.3 Model-based traffic state estimation in the cumulative flow plane . . . 130
7.3.1 Estimation using Newell's method and a triangular fundamental diagram . . . 130
7.3.2 Learning the link traffic flow model . . . 131
7.3.3 Traffic state estimation using relative flow data and a link traffic flow model . . . 135
7.3.4 Desirable spatial-temporal data characteristics . . . 141
7.4 Testing the principles using simulated data . . . 143
7.4.1 Learning the link traffic flow model parameters . . . 144
7.4.2 Estimating the boundary conditions based on moving observers alone 146
7.5 Conclusions and insights . . . 149

8 Using relative flow data to reveal traffic flow properties 151
8.1 Introduction . . . 153
8.2 Background on traffic flow modeling and estimation . . . 154
8.2.1 LWR-model with continuous concave fundamental diagram . . . 154
8.2.2 Newell's simplified kinematic wave theory and Daganzo's variational theory . . . 155
8.3 Obtaining wave observations from relative flow data . . . 157

8.3.1 Collecting relative flow data . . . 157
8.3.2 Obtaining the wave observations . . . 158
8.4 Methodology to expose traffic flow properties . . . 159
8.4.1 Finding characteristic wave speeds and passing rates from wave observations . . . 161
8.4.2 Stochasticity and non-linear fundamental diagram branches . . . 162
8.5 Case study . . . 164
8.5.1 Scenarios . . . 165
8.5.2 Data collection and observation points selection . . . 165
8.5.3 Exposing traffic flow properties . . . 166
8.6 Results . . . 167
8.6.1 Traffic conditions and assignment of the restrictive traffic phase to observation points . . . 168
8.6.2 Approximation of the branch-specific characteristic wave speeds and passing rates . . . 169
8.6.3 Stochasticity and non-linearity of fundamental diagram branches . . . 171
8.7 Conclusions . . . 174

9 Road-users participation in sharing personal data directly with road authorities 177
9.1 Introduction . . . 179
9.2 Stated preference experiment . . . 180
9.2.1 Survey design . . . 180
9.2.2 Data collection and sample representativity . . . 184
9.3 Discrete choice model . . . 185
9.3.1 Methodology . . . 185
9.3.2 Results . . . 186
9.4 Willingness to accept . . . 188
9.5 Conclusions and discussion . . . 189

10 Conclusions and outlook 191
10.1 Findings and conclusions . . . 191
10.1.1 Findings . . . 191
10.1.2 Overall conclusions . . . 195
10.2 Implications for practice . . . 196
10.3 Future research and implications for science . . . 197

Bibliography 201

About the author 209

TRAIL Thesis Series publications 213

Chapter 1

Introduction

Traffic state information is crucial for different applications. Road authorities are assigned with the task to design and manage the road infrastructure and traffic. This means that they have to make choices related to which infrastructure is constructed, which traffic management systems are installed and how these management systems are operated. We separate these tasks in *design*, *operation* and *validation*. All of these tasks require information related to the traffic state. The operation of traffic management systems requires real-time information, while for the other tasks it suffices to have historical information. Private companies such as Google, TomTom and car manufacturers provide traffic information (e.g., travel-times) and routing advice to their users.

To obtain the desired traffic state information, we rely on traffic sensing data. These data may directly provide the desired information. Alternatively, the information can be inferred from partially observed and noisy traffic data (Seo, Bayen et al., 2017). This process of inference is referred to as traffic state estimation (TSE). Observing all desired traffic states is often infeasible, i.e., this may require perfect (error-free) sensing equipment that covers the full road network all of the time, which is both a technical and financial challenge. TSE allows us to obtain the desired information at lower costs. However, it requires us to make choices related to which data are collected and how these are processed.

Collecting traffic sensing data is a form of estimation. It requires a procedure in which detector signals are processed to obtain information related to specific traffic features. As an example, let us consider inductive loop-detectors, which are the most widely deployed type of sensing equipment (Antoniou et al., 2011). These detectors are use to derive whether a vehicle is present at the detector location and potentially expose the vehicle signature based on the inductive signatures measures (Oh et al., 2002). In turn, this information can be used to derive speeds, flows and occupancies. Furthermore, the vehicle signatures may allow us to classify the vehicle types, e.g., passenger car or truck, and thereby provide class-specific speeds and flows. When talking about loop-detector data, it is common that we talk about the traffic flow features that are estimated using the loop-detector signals, e.g., individual vehicle speeds (disaggregated data) or mean speeds (aggregated data).

Traffic sensing data are collected using different types of sensing equipment and processing techniques. Road-side (i.e., stationary) sensors such as loop-detectors, cameras or radars, can be used to observe traffic for a fixed and small road segment. Observations from

sensors installed at different locations can potentially be linked to each other, e.g., when using camera's with license plate recognition. By linking observations, information on other traffic flow features can be obtained, e.g., route choice fractions and individual travel times (Antoniou et al., 2011). Furthermore, sensing equipment may be located in (e.g., smartphones) or installed in road vehicles. Such sensors can be used to collect probe data (which is also known as floating car data) that describe probe vehicle trajectories, speeds and/or travel-times. Furthermore, advances in vehicle automation lead to more sensing and processing equipment being installed in road vehicles. This yields opportunities in collecting new types of traffic sensing data, e.g., spacing between vehicles (Seo et al., 2015), (Seo & Kusakabe, 2015) and overtakings (Florin & Olariu, 2017). Depending on the road infrastructure and vehicle equipment, a vehicle may serve as a moving observer that does not only observe its own traffic flow, but also observes opposing or parallel traffic flows, e.g., Redmill et al. (2011).

This thesis introduction is organized as follows: First, a background on traffic state estimation is presented, see Section 1.1. This section discusses the topics that are important to identify the scientific gaps (Section 1.2) and define the research objectives and questions (Section 1.3). After these sections, the thesis contributions are summarized in Section 1.4. Finally, the outline of this thesis is presented in Section 1.5.

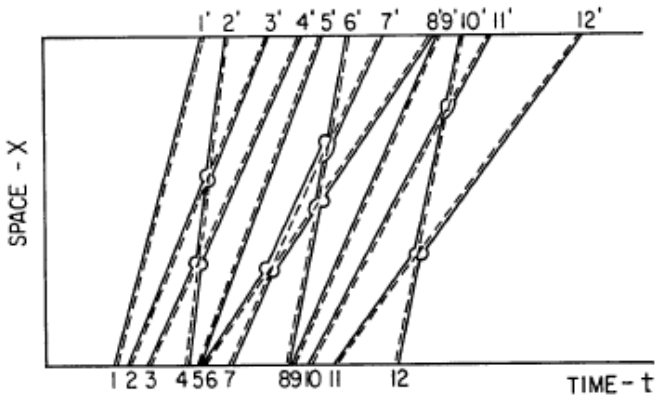
1.1 Background on traffic state estimation

This section is divided in three parts. First, it is explained how traffic flow is described on a macroscopic level, see Section 1.1.1. Second, different traffic state estimation approaches are discussed, see Section 1.1.2. Third, in Section 1.1.3, the key component in any estimation approach, i.e., traffic sensing data, is discussed. Here, traffic sensing data are categorized based on the microscopic or macroscopic traffic flow features that the data aim to describe.

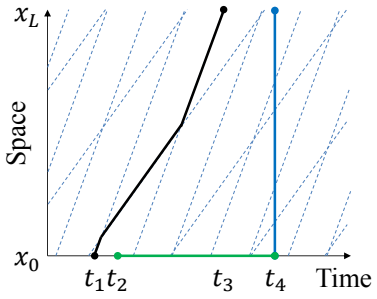
1.1.1 Macroscopic description of traffic flow

The traffic conditions can be described on different levels. On the microscopic level, we describe the characteristics of individual vehicles, e.g., the individual vehicle speed, time headway and space headway. On the macroscopic level, we describe the characteristics of traffic as a whole, e.g., the mean speed u, flow q and density k. Traffic state estimation aims to infer the macroscopic traffic state.

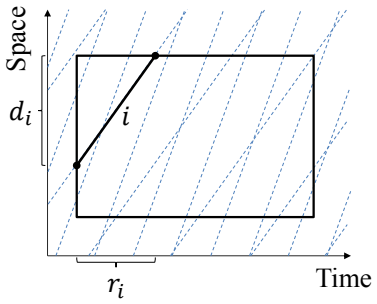
The three-dimensional representation of traffic flow (Makigami et al., 1971), i.e., space x, time t and cumulative flow N, provides a complete macroscopic description of traffic flow. Figure 1.1a shows the first figure presented in (Makigami et al., 1971). As stated by Makigami et al., the collection of trajectories describes 'everything (except lateral motion) that it relevant to the traffic behavior. In fact, it includes too much detail'. Tracking all individual vehicles is not needed to describe the traffic state and behavior on a macroscopic level. Therefore, vehicles can be renumbered when they pass each other such that the vehicle number always decreases with space at any time. In this case, the number assigned to vehicles bounce off when vehicle trajectories intersect in space-time, see Figure 1.1a. This yields the three-dimensional representation of traffic flow. In this study, we use the term



(a) Vehicle trajectories and renumbering of vehicles (Makigami et al., 1971).



(b) Macroscopic variables described by the change in cumulative flow between two points



(c) Macroscopic variables that relate to spatial-temporal areas

Figure 1.1: Three-dimensional representation of traffic flow and related macroscopic traffic flow variables.

'cumulative flow' to describe the number of vehicles that have passed x at t, which means that the cumulative flow always has an integer value. Other studies sometimes use different terms to describe the same feature, e.g., the cumulative number of vehicles, the cumulative vehicle number or the Moskowitz function.

The cumulative flow has to be initialized at one point in space-time. The value used in initialization is not important, because the change in cumulative flow over space and time suffices to obtain the features that are discussed below. However, it is important that the same set of vehicles is used to count the cumulative flow for all points in space-time. We make a distinction between two types of macroscopic traffic flow descriptions in the remainder of this study. These are (1) features described by a set of two points in space-time and the difference in cumulative flow between these points, and (2) features describing the macroscopic traffic state for an area in space-time.

The difference in cumulative flow between two points in space-time has in some cases a clear physical meaning. For instance, the link vehicle accumulation, i.e., the number of vehicles on the link, at time t is obtained by taking the difference between the cumulative flow

at the upstream and downstream boundary at t. For the example provided in Figure 1.1b, the vehicle accumulation at t_4 is equal to the change in cumulative flow over the blue line, i.e., $N(x_0,t_4) - N(x_L,t_4)$. Furthermore, the number of vehicles that entered the link between t_2 and t_4 is equal to the change in cumulative flow over the green line, i.e., $N(x_0,t_4) - N(x_0,t_2)$. Also, the (mean) travel-time related to a specific vehicle number is given by the time difference between equal values of the upstream and downstream cumulative flow curves, e.g., $TT = t_3 - t_1$ for the vehicle number related to the black line.

Edie's generalized definitions of traffic flow (Edie, 1965) allow us to describe the traffic state for areas in space-time. These definitions describe that the macroscopic traffic flow variables flow q and density k for the spatial-temporal area are respectively equal to the total travel distance TTD and total time spent TTS divided by the area size. Furthermore, the quotient of q and k yields the mean speed for the spatial-temporal area. The TTD and TTS can be determined based on the distance traveled d and time spent r within the area by all individual vehicles passing through the area, see Figure 1.1c. However, when we know the relative cumulative flow over the area boundaries, i.e., we know when vehicles entered and left the area, TTD and TTS can also be determined. Note that this process is irreversible. The change in cumulative flow over the area boundaries cannot be reconstructed based on a given flow and density for the spatial-temporal area.

1.1.2 Different estimation approaches

In traffic state estimation, different approaches may be followed. These approaches differ in terms of (1) whether historical data is used to learn models and (2) whether these models incorporate the physical properties (e.g., conservation of vehicles). To this extent, Seo, Bayen et al. (2017) define three estimation approaches, i.e., (1) streaming-data-driven, (2) data-driven and (3) model-driven. The graph used by Seo, Bayen et al. (2017) is shown in Figure 1.2.

This introduction does not provide a detailed description for each estimation approach. However, it is important to point out that traffic sensing data is not only a direct input for traffic state estimation, but can also be used to learn models that are used in traffic state estimation. Traffic data has played a crucial role in describing traffic flow behaviour leading to traffic flow models, e.g., the LWR-model (Lighthill & Whitham, 1955) and (Richards, 1956). As the traffic flow behaviour that these models aim to describe can differ over space and change over time, researchers have aimed to develop methodologies that can learn traffic flow models based on local historical data, e.g., Dervisoglu et al. (2009), Knoop & Daamen (2017) and Seo, Kusakabe & Asakura (2017). In addition to traffic flow models, other models, e.g., those used in data-driven estimation approaches or models describing observation error characteristics, may also be used based on historical data.

1.1.3 Traffic sensing data

Traffic sensing data describe different traffic features. In this study, we make a distinction between microscopic (Figure 1.3a) and macroscopic (Figures 1.3b and 1.3c) traffic sensing data. Furthermore, macroscopic traffic sensing data is separated in data describing the traffic state for a spatial-temporal area (Figure 1.3b) and data describing the change in cumulative flow between points in space-time (Figure 1.3c).

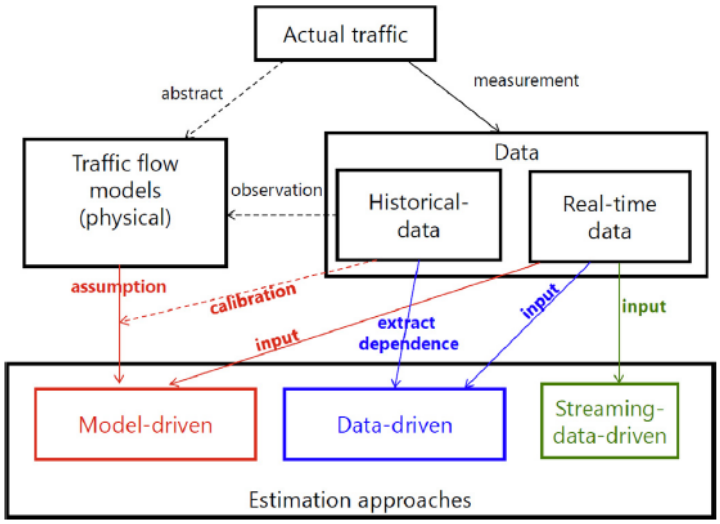
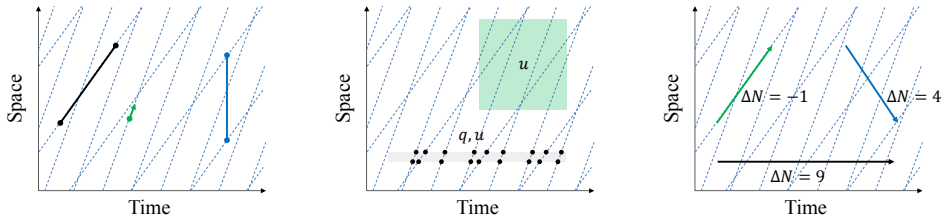


Figure 1.2: Categorization of estimation approaches by Seo, Bayen et al. (2017).



(a) Microscopic sensing data, i.e., probe trajectory data (black), individual vehicle speed (green) and lane-headway (blue).

(b) Macroscopic sensing data describing the traffic state for a spatial-temporal area.

(c) Macroscopic sensing data describing the change in cumulative flow between points in space-time.

Figure 1.3: Vehicle trajectories (blue dashed lines) and categorization of traffic flow data.

Microscopic traffic sensing data describe microscopic variables. Figure 1.3a shows three examples of microscopic sensing data, i.e., probe trajectory data (black), individual vehicle speed data (green) and lane-headway data (blue). These microscopic sensing data have been used in different studies that aim to estimate the macroscopic traffic state. For instance, Claudel & Bayen (2010b), Bhaskar et al. (2010), Van Lint & Hoogendoorn (2015) and Sun et al. (2017) use probe trajectory or vehicle re-identification data; Nanthawichit et al. (2003) and Herrera & Bayen (2010) use individual vehicle speed data; and Seo et al. (2015) and Seo & Kusakabe (2015) use lane-headway data. Working with microscopic sensing data has multiple disadvantages. These data contain information on the traffic state, but also on individual driving behavior (e.g., the headway may depend on the aggressiveness of the driver). Multiple studies have considered probe trajectory data to describe the change in cumulative flow between points in space-time, e.g., Claudel & Bayen (2010b), Bhaskar et

al. (2010), Van Lint & Hoogendoorn (2015) and Sun et al. (2017). For this purpose, it is often assumed that vehicles do not overtake each other and thus that the cumulative flow is constant over the probe trajectory. However, in multi-lane traffic this assumption is likely to be violated, e.g., in Figure 1.3a the cumulative flow is not constant over the black line as the probe vehicle is overtaken by a faster vehicle.

Traffic sensing data can describe Edie's generalized definitions of traffic flow for spatial-temporal areas. Figure 1.3b shows two examples of macroscopic traffic sensing data that describes the traffic state for an area in space-time. The grey area depicts an area for which the flow and mean speed can be estimated using a double loop-detector, i.e., two loop-detectors that are placed close to each other. These two detectors measure individual vehicle passings that can be used to describe the flow and mean speed for a slim spatial-temporal area (i.e., small road segment for a specified period). Furthermore, probe data (also known as floating car data) can be used to estimate the mean speed of travel-time for spatial-temporal areas, see the green area in Figure 1.3b. Due to privacy reasons, probe data are often only made available to traffic scientist in this aggregated form, e.g., Google BetterCities data (Eland, 2015) and NDW FCD (Uenk-Telgen, 2018). Although these data should aim to describe Edie's generalized definitions, incorrect assumptions and/or incomplete information may lead to a difference between the data and Edie's definitions. For instance, the double loop-detectors used in the Netherlands report the time-mean speed, which differs from Edie's definitions on the mean speed. This is a result of incorrect assumptions, as the issue could be solved by calculating the harmonic-mean instead of the time-mean of the individual speeds observed by the loop-detector. Probe mean speed data providers estimate the mean speed based on an (unknown) fraction of the vehicles, which means that the traffic state is estimated using incomplete information. Furthermore, the probe data providers often use differential privacy filters (Kargl et al., 2013), i.e., they add noise, to ensure privacy for its users.

Stationary and moving observers can be used to collect data that describes the change in cumulative flow between points in space-time. A stationary (road-side) detector, e.g., loop-detector or camera, can observe the flow (aggregated data) or individual passings (disaggregated data) over a horizontal line in space and time, see black line in Figure 1.3c. As argued by Redmill et al. (2011) and Florin & Olariu (2017), equipped and connected vehicles could serve as moving observers and collect these data for traffic flow in their driving direction (green line) and the opposite driving direction (blue line). In this study, these data are referred to as relative flow data. The relative flow data collected with stationary and/or moving observers describes the change in cumulative flow over a path in space-time. Relative flow data may contain observation errors, e.g., an observer may miss or double observe a passing vehicle. If these data are used to estimate the traffic state in or via the cumulative flow plane, the cumulative error problem has to be mitigated. Prior studies have proposed using different types of information and methodologies to mitigate this problem, e.g., Bhaskar et al. (2010), Van Lint & Hoogendoorn (2015) and Amini et al. (2016).

1.2 Scientific gaps

Multiple scientific gaps are addressed in this thesis. These relate to the following four topics. First, we look into estimating the traffic state for spatial-temporal areas based on

currently available traffic sensing data. It is valuable to gain a deeper understanding of this process as it can be used to construct traffic sensing data that describe the traffic state for spatial-temporal areas, e.g., probe mean speed data. Second, we evaluate the main problem when estimating the traffic state using detector data in the cumulative flow plane, i.e., the cumulative error problem. To mitigate this problem, different types of probe data, e.g., probe mean speed data and probe trajectory data, can be used. Third, we consider the option that automated and/or other connected and equipped vehicles collect relative flow data, and use these data for traffic state estimation in the cumulative flow plane. These data are seen as a better alternative to probe trajectory data because the relative flow (i.e., change in cumulative flow) to the connected and equipped vehicles is observed. Finally, we look at the way road authorities gain access to personal traffic sensing data (e.g., probe trajectory data and relative flow data from connected and equipped vehicles). The next four paragraphs describe the scientific gaps related to these four topics.

Traffic sensing data are often considered to describe the traffic state for spatial-temporal areas. Probe mean speed or travel-time data, e.g., Google BetterCities data (Eland, 2015) and NDW FCD (Uenk-Telgen, 2018), aim to describe the mean speed for spatial-temporal areas, i.e., for a road-segment during a period. Furthermore, it may be assumed that detector data, which relate to a specific location or short road segment, are representative for a larger road-segment. For instance, Wang & Papageorgiou (2005) use detector data to obtain mean speed estimates for road stretches of (approximately) 500 m. To determine the true (Edie's) traffic states for these spatial-temporal areas, we would need to observe the boundaries of these areas. However, both probe and detector data do not fully observe these boundaries and thus provide incomplete information to estimate the traffic state for these spatial-temporal areas. Therefore, the data and estimates will not perfectly describe the true traffic states for these areas, and there will thus be (estimation) errors. It is valuable to understand the potential error characteristics. In traffic state estimation, information can be fused based on the expected error characteristics, for instance, using a variant of the Kalman Filter (e.g., Nanthawichit et al. (2003) or Wang & Papageorgiou (2005)) or Particle Filter (e.g., Hegyi et al. (2007)). Furthermore, in a applications of traffic state information (e.g., dynamic traffic management), the potential error characteristics can be taken into account. However, there is only limited understanding on the characteristics of the errors induced by incomplete information and incorrect assumptions in estimating the traffic state for spatial-temporal areas.

Multiple studies aim to estimate the traffic state in the cumulative flow plane by levering detector and probe data, e.g., to estimate vehicle accumulation (Van Lint & Hoogendoorn, 2015) or travel-time (Bhaskar et al., 2010). In these studies, the detectors serve as stationary observers that provide data related to the inflow and outflow of a road segment. The probe trajectory data are used to provide information on the change in cumulative flow between two points in space-time that lie on the upstream and downstream boundary of the road segment. This information can be used to mitigate the cumulative error problem. To describe the change in cumulative flow over the probe trajectories, these studies assume 'no overtaking', i.e., the cumulative flow value does not change along the trajectory. This assumption is likely to be violated in multi-lane road segments, which yields estimation errors. To the best of our knowledge, this problem is not addressed using the detector data, while these data do provide information on the rate at which the cumulative flow value changes over the probe trajectory. Due to privacy concerns, in practice, probe mean speed or travel-time data, e.g.,

Google BetterCities data (Eland, 2015) and NDW FCD (Uenk-Telgen, 2018), are available on a wider-scale than probe trajectory data. However, no studies exist that use probe mean speed data (i.e., probe data that describe the mean speed for spatial-temporal areas) to mitigate the cumulative error problem. It is interesting to evaluate how these data can be used in combination with stationary observers to estimate the traffic state in the cumulative flow plane. Furthermore, it is interesting to evaluate which limitations there are when relying on this irreversible compressed traffic sensing data Seo, Bayen et al. (2017) instead of data that describes the change in cumulative flow between points in space-time.

The existing studies that consider the option to collect and use relative flow data for traffic state estimation, i.e., (Redmill et al., 2011) and (Florin & Olariu, 2017), propose to use (a variant of) the moving observer method (Wardrop & Charlesworth, 1954). This method follows a streaming-data-driven estimation approach and thus estimates the traffic state solely based on real-time data and 'weak' assumptions (Seo, Bayen et al., 2017). However, this methodology does not allow us to estimate the traffic state over the full domain of interest. Other studies propose methodologies that do allow us to estimate the change in cumulative flow over the full space-time domain, e.g., Newell's (three-detector) method (G. F. Newell, 1993a), (G. F. Newell, 1993c), (G. F. Newell, 1993b), (Laval et al., 2012), Claudel's method (Claudel & Bayen, 2010b), (Claudel & Bayen, 2010a) and Sun's method (Sun et al., 2017). These studies all use a model-driven estimation approach, i.e., they all require information in the form of a traffic flow model. Such a model can be learned using historical data, e.g., Dervisoglu et al. (2009), Knoop & Daamen (2017) and Seo, Kusakabe & Asakura (2017). However, no studies exist that learn the traffic flow model based on historical relative flow data. Sun et al. (2017) simultaneously estimate the traffic state and model parameters. Claudel & Bayen (2010b), Claudel & Bayen (2010a) and Sun et al. (2017) use probe trajectory (or vehicle reidentification) data in combination with the assumption that there is no overtaking, to describe the change in cumulative flow between points in space-time. However, they do not consider relative flow data collected by moving observers.

Probe trajectory data and relative flow data collected using moving observers allow us to follow individual road users over time and are therefore personal data. Due to potential commercial and legal issues, data providers are reluctant or not able to share these privacy-sensitive data. As stated above, the data may be masked, i.e., by aggregation and applying a differential privacy filter. This is for instance the case for the probe mean speed or travel-time data that are available from Google or NDW. However, in this and other studies, traffic state estimation methodologies are proposed that require personal traffic sensing data as input. To use such methodologies for traffic management purposes, road authorities may gain access to personal data directly from the road users themselves. In this way, they can create a data sharing system that satisfies the privacy regulations defined by the EU General Data Protection Regulation (GDRP) (EU, 2016). Whether road-users are willing to participate in a data-sharing system is expected to depend on the conditions that are proposed by the road-authorities, e.g., which data are shared, for which applications are the data used, and does the road-user get a (financial) compensation. Therefore, it is interesting to understand under which circumstances road users are willing to share their personal data with road authorities for the purpose of dynamic traffic management. However, to the best of our knowledge, this is not yet investigated.

1.3 Research objectives and questions

This thesis aims to show which advantages relative flow data that are collected by moving observers offer in addition or as an alternative to the currently widely used detector and probe data for the purpose of macroscopic traffic state estimation. For this purpose, the data are separated based on the type of macroscopic traffic feature that the data describe. As explained above, detector and probe data can describe the macroscopic traffic state for spatial temporal areas. Furthermore, relative flow data collected by stationary (e.g., detectors) and moving (e.g., automated vehicles) observers describe the change in cumulative flow between points in space-time.

It is interesting to understand the potential errors in these data or estimates, i.e., probe mean speed data and assuming that detector data is representative for a larger road stretch. For this purpose, detector and individual probe data that may be available in the Netherlands are used to estimate the mean speed for spatial-temporal areas. Here, we focus on estimating the mean speed for spatial-temporal areas as both data types can be used to estimate this variable without requiring additional information. Estimating the density based on speed observations is possible, but this requires additional information in the form of a calibrated fundamental diagram, e.g., Smulders fundamental diagram (Smulders, 1990). This yields the following research question:

Research question 1: What causes estimation errors when estimating Edie's mean speeds based on error-free probe or loop-detector data?

As explained in the previous section, multiple studies aim to estimate the traffic state in the cumulative flow plane by fusing detector and probe data. Furthermore, in that section, two scientific gaps related to this estimation procedure were identified: (1) due to privacy concern probe mean speed data are available on a wider scale than probe trajectory data, but no studies exist that use these aggregated probe data to mitigate the cumulative error problem, and (2) the assumption of 'no overtaking' is likely to be violated in multi-lane roads and could be addressed using detector data. To address these gaps, two methodologies are designed and tested. This yields methodologies that (1) mitigate the cumulative error problem using probe mean speed data in on-ramp vehicle accumulation estimation, and (2) estimate the change in cumulative flow over probe trajectories using disaggregated detector data. For both studies an empirical dataset is available and used to test the methodologies. However, as these empirical dataset miss the ground truth, data obtained from microscopic simulation are also used for testing purposes. The research questions related to these studies are:

Research question 2: How can probe mean speed data be used to mitigate the cumulative error problem in on-ramp vehicle accumulation estimation?

Research question 3: How can disaggregated detector data be used to estimate the change in cumulative flow over probe trajectories between detector locations?

After addressing these research questions, the use of relative flow data collected by moving observers is investigated. For this purpose, we are interested in investigating how these data can be used in different estimation approaches and which characteristics make the data

valuable for traffic state estimation. First, we designed a traffic state estimation methodology that follows a streaming-data-driven estimation approach. This methodology is used to estimate the flow and density for spatial-temporal areas based on relative flow data collected by a combination of stationary and moving observers. The methodology and case study provide insight into the valuable information that the data contain and how this information can be extracted. Second, a model-driven estimation approach was considered. As shown in Figure 1.2, historical data can be used to calibrate the (physical) traffic flow models. Therefore, it is investigated how relative flow data can be used to find the traffic flow model parameters. Furthermore, we explore which spatial-temporal characteristics are valuable in model-based traffic state estimation using relative flow data and link this to the characteristics that can be expected when relying on a combination of stationary and moving observers or moving observers alone. For this purpose, principles are proposed that can be used to learn the model parameters based on historical data, and estimate the traffic state by combing real-time data and calibrated traffic flow model. These principles are based on Newell's principle (G. F. Newell, 1993a) and assume that traffic flow follows the LWR-model (Lighthill & Whitham, 1955), (Richards, 1956) with triangular fundamental diagram. However, the LWR-model with triangular fundamental diagram may not be the best description of traffic flow properties. Therefore, we investigate how relative flow data collected from stationary and moving observers can be used to expose the model parameters in the more general case of having a continues concave FD. This yields the following four research questions:

Research question 4: How can we estimate flow and density by solely relying on real-time relative flow data?

Research question 5: How can historical relative flow data and Edie's generalized definitions of traffic flow be used to estimate the fundamental diagram of traffic flow?

Research question 6: What spatial-temporal relative flow data characteristics are desired in model-based traffic state estimation?

Research question 7: How can relative flow data from stationary and moving observers be used to reveal traffic flow properties?

Data considered in research questions 1, 3, 4, 5, 6 and 7 can fall in the category 'personal traffic sensing data'. In the previous section, we conceptualized the idea that road authorities can set-up a data-sharing system to gain access to these personal data directly from road-users. This may allow road authorities to gain access to data that are otherwise not available to them or only available at high costs. Wether or not road users are willing to share their personal data is such a system will depend on the data-sharing conditions that are proposed by the road-authorities. Therefore, it is interesting to consider the following (and last) research question:

Research question 8: Under which conditions are road-users willing to share their personal traffic sensing data with road authorities for dynamic traffic management purposes?

1.4 Contributions

This section presents the contributions made in this thesis and why it is relevant for practice. For this purpose, the contributions to science, and contributions and relevance to practice are respectively presented in Sections 1.4.1 and 1.4.2.

1.4.1 Contributions to science

All studies contain one or multiple contributions to science. These are presented below in the order-of-appearance in this thesis:

- **Insight into the error characteristics when estimating the mean speed for a spatial-temporal area based on error-free aggregated detector data and individual probe speed data** (Chapter 2): Individual probe data are used to construct probe mean speed data, which describe the mean speed for road-segments during time periods (i.e., rectangular spatial-temporal areas). Furthermore, prior traffic state estimation studies use detector data to estimate the mean speed for similar spatial-temporal areas. This thesis provides insight into the error characteristics related to these mean speed estimates. These insights are valuable when designing methodologies that use these estimates as input.
- **Methodology to recover the cumulative count errors of stationary observers that observe the on-ramp arrivals and departures using probe mean speed data** (Chapter 3): The on-ramp vehicle accumulation can estimated recursively based on count data from detectors that are positioned at the upstream and downstream on-ramp boundaries. To obtain accurate vehicle accumulation estimates with this recursive estimation approach, the cumulative error problem needs to be mitigated. For this purpose, a methodology was designed that recovers the cumulative count errors that build up in prior periods. In this methodology, probe mean speed and detector count data are used to estimate the mean on-ramp vehicle accumulation during the probe data period. This reference estimate is compared with a mean on-ramp vehicle accumulation estimate that is based on the recursive detector count estimates. The methodology aims to recover the cumulative error based on the difference between these two estimates and a description of the error characteristics.
- **Methodology for online learning of the bias of stationary observers that observe the on-ramp arrivals and departures using probe mean speed data** (Chapter 3): In addition to 'error recovery', the mean on-ramp vehicles accumulation estimates obtained from probe mean speed and detector count data, can be used to learning the structural detector count error. The methodology designed for this application first exposes the detector count error based on estimates that relate to two consecutive probe data periods. Next, this exposed error is used to update a prior estimate of the structural detector count error.
- **Methodology to estimate the change in cumulative flow along probe trajectories between detector locations based on disaggregated detector data** (Chapter 4): In multi-lane road segments, probes can overtake or be overtaken by other road users. This means that the cumulative flow value can change over probe trajectories. A

methodology was designed to estimate the change in cumulative flow along probe trajectories between detector locations based on disaggregated detector data. In this methodology, disaggregated detector data that describe the times and speeds related to all observed individual vehicle passings and probe trajectory data, are used to estimate the probe-specific relative flow at detector locations. The probe-specific relative flow describes the rate at which the cumulative flow changes along the probe trajectory. Based on the probe-specific relative flow estimates and the probe travel-time between detectors, the change in cumulative flow along the probe trajectory is estimated.

- **Methodology to estimate flow and density for spatial-temporal areas using relative flow data collected using stationary and moving observers alone** (Chapter 5): Relative flow data can be collected with stationary observers (e.g., loop-detectors) and moving observers (e.g., automated vehicles). If observers cross each other, the change in cumulative flow can be determined between points along the spatial-temporal paths over which the observers travel (which are denoted as observation paths). A streaming-data-driven methodology (i.e., the PON-methodology) is designed that estimates flow and density for spatial-temporal areas based on these observations of the changes in cumulative flow between point along observation paths. In this methodology, we first subdivide space and time in triangular areas for which the change in cumulative flow between the corner points is observed. Next, flow and density are estimated for these triangular areas based on the changes in space, time and cumulative flow between the triangle corner points. Finally, the estimates related to the triangular areas are mapped to a freely definable estimation mesh (e.g., a rectangular estimation mesh the subdivides space and time in road segments and time-periods).

- **Methodology to estimate the parameters of a triangular fundamental diagram using relative flow data collected using moving observers** (Chapter 6): To estimate the parameters of a triangular fundamental diagram, we can rely on moving observers that can observe traffic in their own and opposite travel direction. The methodology designed for this application consists of two steps. First, flow and density are estimated for the spatial-temporal areas that are enclosed by the observation paths of the moving observers. For this purpose, we use Edie's generalized definitions of traffic flow. Second, based on the flow-density estimates the parameters of the triangular fundamental diagram are estimated. In this step, it is taken into account that traffic states that are part of different fundamental diagram branch can exist with a single spatial-temporal area for which flow and density are estimated.

- **Principles that can be used to learn the link traffic flow model in the cumulative-flow-plane using historical relative flow data** (Chapter 7): Learning the traffic flow model based on historical data is a process in a model-driven estimation approach. In this thesis, principles are derived and explained that can be applied to learn this model based on historical relative flow data. To derive these principles, it is assumed that the traffic flow properties follow the LWR-model with triangular fundamental diagram and that there are no internal constraints (e.g., moving bottleneck or accident) on the link. Combining relative flow data collected with stationary and moving observers, or moving observers alone, yields observations of the change in cumulative

flow between points in space-time. Based on Newell's principle, we derived which conditions have to be satisfied by the observed change in cumulative flow over time if the assumptions related to the traffic flow model hold. This insight can be used to learn the traffic flow model parameters based on a set of observations of the change in cumulative flow over time over different waves. This set of observations is denoted as the learning dataset.

- **Principles that can be used to partially estimate the traffic state in the cumulative-flow-plane using relative flow data collected by moving observers alone** (Chapter 7): In model-driven estimation, real-time data and (traffic flow) models are fused to estimate the traffic state. Newell's principle can be used to estimate the cumulative flow value for points in space-time based on the LWR-model with triangular fundamental diagram and the boundary conditions (i.e., cumulative flow curves related to the link boundaries). We explained that this principle also allows us to partially estimate the traffic state in the case that relative flow data are collected by moving observers alone. In this case, the link boundaries are not observed by stationary observers and the observation paths may be only partially connected via intersection. It is shown that the observation paths may be connected by drawing characteristics waves (which are dependent on the traffic flow model) between observation paths. Furthermore, by excluding that only of the two characteristic waves is restrictive, we can partially estimate the boundary conditions.

- **Explanation of the value of relative flow data in a model-driven traffic state estimation approach** (Chapter 7): Relative flow data can be collected for a combination of observers that travel at different speeds. For instance, relative flow data from stationary observers and moving observers can be combined, or we may be able to collect these data using moving observers that observe the own and opposing direction of travel. By collecting relative flow data with observers that travel a different speeds, observers can cross each other in space-time and waves traveling at the (potential) characteristics wave speeds can be drawn to a large number of points within space-time. This allows us to apply the two processes that are part of a model-driven estimation approach, i.e., 'learn models based on historical data' and 'state estimation by fusing real-time data with the exposed models'.

- **Approach to reveal traffic flow properties using relative flow data** (Chapter 8): Combining relative flow data from stationary observers positioned at the link boundaries and moving observers yields wave observations. Wave observations describe the change in space, time and cumulative flow. By considering wave observation that originate in a spatial-temporal point between the link boundaries to the link boundaries, we can reveal the traffic flow properties. For this purpose, wave observations and features derived from it are visualized in different plots. These plots provide insight in the shape and parameters of the traffic flow model that may be used to describe the traffic flow properties.

- **Insight into the road-users' willingness to share personal traffic sensing data with road authorities for the purpose of dynamic traffic management** (Chapter 9): Relative flow data and probe trajectory data are personal data. Due to potential privacy

concerns and legislation, road authorities may not be able to gain access to these data via third parties. Therefore, road authorities may want collect personal traffic sensing data directly from road-users. To gain insight in the road-users' willingness to share such data with road-authorities for the purpose of dynamic traffic management application a stated preference study was conducted and analyzed.

1.4.2 Contributions and relevance to practice

A distinction between contributions and relevance to practice is made. The contributions explain which elements of the work presented in this thesis are applicable to practice. The relevance explains why this work is relevant for practice.

Contributions to practice

The detector and probe data used in the studies presented in Chapters 2, 3 and 4 are currently available and used in practice. This means that the proposed methodologies that work with these data can be used in practical applications. Therefore, the first four contributions to science (which are presented above) are also contributions to practice. The first contribution is valuable for practitioners when they want to evaluate the potential uncertainty in data (e.g., probe mean speed data) that they use in their traffic analysis and/or when implementing traffic state estimation algorithms that fuse data and other types of information based on a description of their error characteristics (e.g., variants of the Kalman Filter). The second and third contributions describe how the on-ramp vehicle accumulation can be estimated based on detector count and probe mean speed data. In the Netherlands, these data are available to road authorities. Therefore, the presented methodologies can be implemented by road authorities and used in ramp metering applications. The fourth contribution has a less direct contribution to practice. To the best of our knowledge, road authorities do not yet work with estimates of the cumulative flow in their applications. By applying the methodology described by the fourth contribution, better estimates of the cumulative flow curves may be obtained, which in turn may be valuable for applications such as dynamic traffic management. However, the presented methodology is an intermediate step that addresses a specific problem (i.e., overtaking) in estimation of the cumulative flow curves. Practitioners still have to use other existing methodologies or design new ones to initialize the cumulative flow curves and correct for observation errors.

The other contributions to science relate to the use of relative flow data collected using moving observers and to the option that road authorities collect personal traffic sensing data from road-users. The relative flow data are not yet collected using automated and/or other equipped and connected vehicles that serve as moving observers. Therefore, the derived principles and designed methodologies that work with these data (see Chapters 5, 6, 7 and 8) cannot directly be implemented in practice. However, these studies show that relative flow data have favorable characteristics for traffic state estimation compared to the currently available data, which in turn is important for (practical) investment decisions related to the collection and use of traffic sensing data. The study conducted on the road-users' willingness to share personal traffic sensing data directly with road authorities (see Chapter 9), is worth more research if road authorities have difficulties gaining access to these kind of data via third parties. The stated-preference study that is presented provides initial in-

sights that may be valuable for practitioners; however, due to the limited scope of our study, more extensive research should be conducted by the practitioners prior to setting up a direct data-sharing system from road-users to road authorities.

Relevance to practice

The field of traffic state estimation is relevant for public authorities (e.g., road authorities) and private companies (e.g., navigation service providers). In time, relative flow data may be a better alternative to loop-detector and probe data, which are currently collected and used by these parties. Switching to another type of traffic sensing data impacts all steps from collecting the data to using the traffic state information in the different applications. This requires the public authorities and private companies to make decisions related to capital and intellectual investments. The insights obtained from this thesis should be taken into account in those decisions and are therefore highly relevant for this parties.

Below, it is explained why this PhD-study is relevant for car manufacturers (private companies) and road authorities:

- **Car manufacturers may collect and use relative flow data using automated and/or other equipped and connected vehicles**: Car manufacturers working on automated functions and/or autonomous vehicles may collect relative flow data. For this purpose, they can rely on the sensing and processing equipment that are installed in the vehicle for the automated or self-driving functions. Furthermore, they may partially rely on software that is developed for these functions (e.g., object recognition and tracking software). The collected relative flow data can be processed using the principles and methodologies proposed in this study. Furthermore, the data and/or estimates may be used for own applications or sold to other parties, e.g., road authorities.

- **The potential future collection and use of relative flow data may already influence the investment decisions of road authorities related to traffic sensing data collection systems**: Road authorities collect traffic sensing data for different applications, e.g., automated incident detection (AID), dynamic traffic management (DTM) and off-line monitoring reports that are used for future infrastructure planning purposes. These applications have different demands related to the required traffic state information, e.g., offline versus online (real-time) information, and the type of variables and related spatial-temporal characteristics. If relative flow data (or other data) proof to be a valuable source of information for these applications, and these data are (expected to be) accessible to road authorities, the collection of other types of traffic sensing data may be reduced. Loop-detectors, which are the dominant source of traffic sensing data for road authorities, are installed in the roads and have a long lifespan. When deciding how many and where these stationary detectors are installed in new or existing roads, road authorities should take into account which are data are expected to become available within the lifespan of the detectors and how these data can be used. It may be the case that road authorities decide to install less loop-detectors and free funds to gain access to other types of data such as probe trajectory data and/or relative flow data.

1.5 Thesis outline

Figure 1.4 shows the outline of this paper-based thesis. The research questions posed in Section 1.3 are addressed in eight separate chapters, which contain the articles that were written (as first author) during the PhD. These eight chapters form the body of this thesis. At the beginning of each chapter we state where the article was published or whether it is under review. To avoid potential citing-conflicts, no changes are made to the published articles with respect to the articles that can be found in the journals. In Chapter 10, the conclusions and implications of this thesis are discussed.

In line with Sections 1.2 and 1.3, the body of this thesis is separated in four parts. The first and last part relate to single chapters, i.e., Chapters 2 and 9. For the other two parts the relevant chapters are grouped using dashed black boxes in Figure 1.4. These boxes, i.e., 'mitigation of the cumulative count error problem with probe data' and 'traffic state estimation with relative flow data', respectively contain two and four chapters. The subdivision of chapters that a part of specific boxes is indicated by the white boxes. Chapters 3 and 4 use different types of probe data to mitigate the cumulative error problem. The chapters that relate to 'traffic state estimation with relative flow data' are subdivided based on the estimation approach. The methodology proposed in Chapter 5 follows a streaming-data-driven estimation approach, while the other chapters relate to model-driven estimation. When applying a model-driven estimation approach (in contrast to a streaming-data-driven estimation approach) data are used to expose the traffic flow properties. In this thesis, traffic flow properties are either are either expose in the flow-density plane (Chapter 6) or cumulative flow plane (Chapters 7 and 8). Chapters 6 and 8 solely relate to exposing these properties, while Chapter 7 also dives into 'real-time estimation'.

To describe other relations between chapters and/or the black dashed boxes, blue arrows are included in Figure 1.4. All arrows that describe relations within the thesis body are accompanied by a short description. Chapter 2 provides insight in the estimation errors of mean speed estimates, which, if obtained from disaggregated probe data, can be seen as probe mean speed data. Chapter 3 uses probe mean speed data to mitigate the cumulative error problem. Chapter 4 uses probe trajectory data. By adding relative flow observations with respect to the probe trajectory relative flow data are obtained, which are consider in the box 'traffic state estimation with relative flow data'. The chapters presented within that box mainly relate to different topics; however, Chapter 8 builds on a part of Chapter 7. Chapter 8 dives deeper in the topic of exposing the traffic flow properties in the cumulative flow plane, and relaxes assumption that are made in Chapter 7. Finally, arrows from Chapter 4 and these chapters that consider the use of relative flow data are drawn towards Chapter 9. Probe trajectory data and relative flow data are personal traffic sensing data. Chapter 9 relates to the road-users' willingness to share such privacy-sensitive data directly with road authorities.

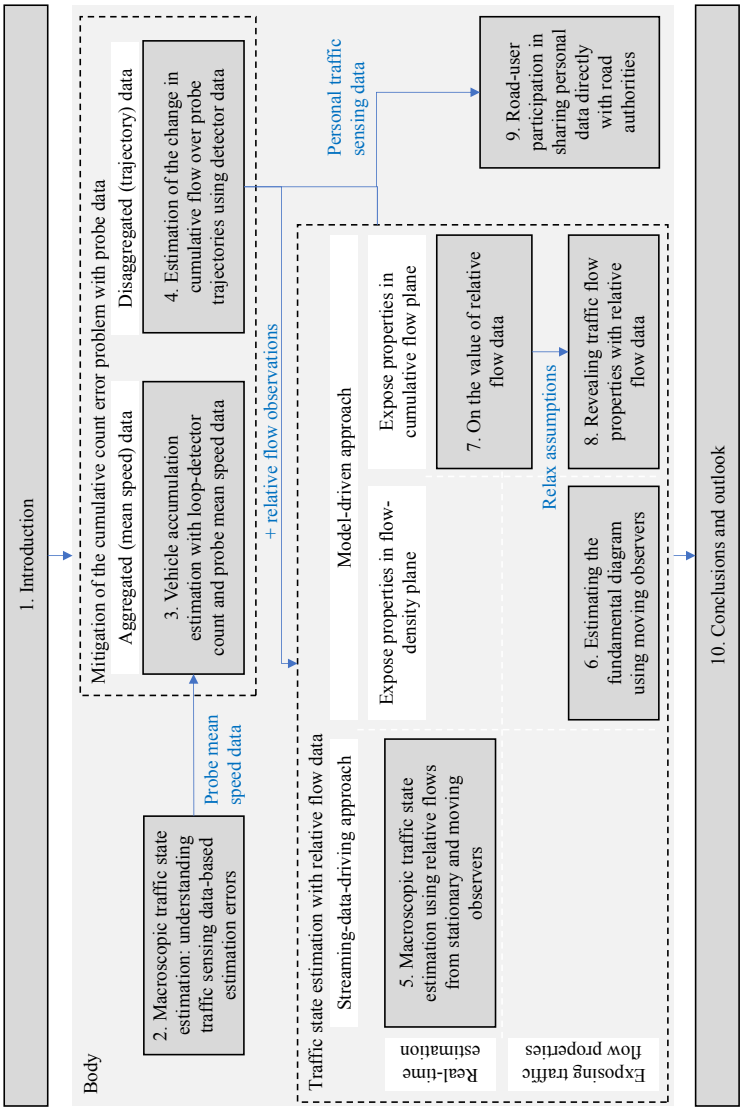


Figure 1.4: Thesis outline

Chapter 2

Macroscopic traffic state estimation: Understanding traffic sensing data-based estimation errors

This chapter is published as a journal article: van Erp, P.B.C., Knoop, V.L. & Hoogendoorn, S.P. (2017), Macroscopic traffic state estimation: Understanding traffic sensing data-based estimation errors. *Journal of Advanced Transportation*, 2017.

Traffic sensing data should describe a specific traffic feature, which is a combination of a variable and spatial-temporal characteristics, e.g., the mean speed on road segment during period. To construct these data, an (estimation) process is followed. This process starts with raw sensor signals and ends at the data, which may thus also be seen as estimates. The exact process depends on the available sensors and the data that one wants to construct.

Processing the raw sensor signals to the data/estimates causes errors due to incomplete information and incorrect assumptions. As the resulting data/estimates serve as input for other applications (e.g., a traffic state estimator that fuses different data and models or directly in traffic management systems), it is important to understand the error characteristics (e.g., mean and variance). To gain a deeper understanding in the errors that are induced in the data-construction-process, we investigate one potential step in this process. Here, we take traffic sensing data as input (i.e., detector and disaggregated probe data) as estimate the mean speed within an estimation mesh that covers the full space-time domain (i.e., road segments and time-periods). The resulting estimates describe the same traffic feature as is described by probe mean speed data, which can be available for traffic state estimation, e.g., NDW (Uenk-Telgen, 2018).

2.1 Introduction

Traffic state estimation is an important element in traffic management applications and traffic information services. The traffic state can be described on different levels (May, 1990). The microscopic traffic state describes the traffic on an individual vehicle level, thus using the time and space headways and individual vehicle speed. The macroscopic traffic state describes the traffic flow conditions using the mean speed, density and flow.

In traffic state estimation, different types of information can be used. Traffic sensing data is collected via different kind of sensors, e.g., loop-detectors (Hoogendoorn et al., 2013), (Wang & Papageorgiou, 2005), mobile-phones (probes) (Nanthawichit et al., 2003), (Herrera & Bayen, 2010). Additional to sensing data, information related to the traffic dynamics is captured in the form of traffic flow models, e.g., the LWR-model (Lighthill & Whitham, 1955), (Richards, 1956). Traffic flow models are based on physical laws and historical data. These information types, thus sensing data and traffic flow models, allow us to estimate the traffic state. The difference between the true and estimated traffic state is denoted as the estimation error. In this research the focus lies on these estimation errors.

It is valuable to have information related to the estimation errors for applications of the related estimates. For instance, in traffic state estimation different types of information, e.g., traffic sensing data-based estimates and traffic flow model-based prediction, can be fused. Examples of such applications are (Wang & Papageorgiou, 2005), (Nanthawichit et al., 2003), (Herrera & Bayen, 2010), (Van Hinsbergen et al., 2012). These applications all consider (amongst other) a variant of the Kalman Filter (KF) (Kalman, 1960) for information-fusion. KFs assume Gaussian distributed errors with a to-be-defined (co-)variance. Alternatively, a Particle Filter (PF) (Arulampalam et al., 2002) can be used for information-fusion, in which we are free to define any type of expected error distribution. In the provided references related to the KF, the researchers all assume constant error variances related to the data-based, i.e., loop-detector or probe data-based, estimates. They thus do not assume any dependency of the error characteristics on features such as the traffic conditions or varying sensing data characteristics. Additional to information fusion, information related to estimation or prediction errors can also have a more direct value for road users. For instance, in travel time prediction, a probability function can be provided instead of a single expected value (Woodard et al., 2017). This may be valuable for routing advise as the travel time variability may negatively influence the attractiveness of certain routes.

In this research the following way-of-thought is considered. If we have knowledge related to the relation between estimation error characteristics and potentially observable features, we may improve our understanding of the error characteristics related to an estimate. Examples of potentially observable features are traffic conditions and sensing data characteristics. Improved understanding of the estimation error characteristics is valuable in the applications discussed above, e.g., traffic state estimation using a variant of the KL (Wang & Papageorgiou, 2005), (Nanthawichit et al., 2003), (Van Hinsbergen et al., 2012).

The objective of this research is to expose the dependency of traffic sensing data-based estimation errors on traffic conditions and sensing data characteristics. We define traffic sensing data-based estimation as the estimation of a desired output based on traffic sensing data. Both the sensing data as the desired output have specific characteristics, e.g., type of variable and spatial/temporal characteristics. If these differ, we have to make assumptions to describe the relation between the two. In this research, we estimate the mean speed in

discrete time and space, i.e., time is discretized in periods and space in cells (road segments), similar to (Wang & Papageorgiou, 2005), (Nanthawichit et al., 2003), (Van Hinsbergen et al., 2012), (Daganzo, 1994), (Daganzo, 1995). The properties of the traffic sensing data-type we consider, i.e., loop-detector data and probe speed data, are based on existing research (Nanthawichit et al., 2003), (Knoop & Hoogendoorn, 2012).

This research focuses on specific combinations of traffic sensing data and estimation output. The findings can be used for applications which consider similar data-types and estimation output. However, more generally, we opt to show that the estimation error characteristics can depend on the traffic conditions and (varying) sensing data characteristics. Any application which requires defining the estimation error characteristics, e.g., information fusion using a KF, can take this into account. However, depending on the specific application this may require extra research.

In this article we first describe the (macroscopic) traffic conditions within a discrete space-time mesh. Next, we discuss traffic sensing data-based traffic state estimation and our focus related to this topic. After describing the conducted experiments, we present and discuss the results. Finally, the conclusions of this research are presented.

2.2 Variables used to describe the traffic conditions

The traffic conditions can be described as a function of space x and time t. For computational reasons, it is valuable to consider the traffic conditions in discretized space and time (Bellemans et al., 2002). To discretize the space x, the road stretch is subdivided into I cells, where i and Δ_i respectively denote the cell number and length of cell i. The number of lanes within i is given by λ_i. Furthermore, time t is discretized in time periods with duration T, which are denoted by $p = 1, \ldots, P$. Each combination of i and p corresponds with a discrete area in the space-time domain. In this discrete space-time mesh the macroscopic traffic variables mean speed, flow and density are respectively denoted by $u(i,p)$, $q(i,p)$ and $k(i,p)$.

In the literature, different methodologies are proposed to calculate the macroscopic variables in a discrete space-time mesh based on the microscopic variables. For instance, Wang & Papageorgiou (2005) propose to calculate each variable independently based on the downstream (flow) and end-of-period (mean speed and density) conditions. Alternatively, Edie (1965) proposed a generalized formulation of the macroscopic variables. In this research we follow Edie's formulation as it considers the conditions over the entire space-time area instead of only the end-of-period and downstream boundary conditions.

In traffic state estimation in a discrete space-time mesh, homogeneous conditions (defined as constant over space) and stationary conditions (defined as constant over time) are often assumed (Van Hinsbergen et al., 2012). Different vehicle classes (e.g., passenger cars and trucks) can co-exist in homogeneous and stationary conditions, namely if the conditions within these classes are homogeneous and stationary.

Assumptions related to homogeneity and stationarity can be important when applying a traffic flow model. For instance, the Cell Transmission Model (CTM) (Daganzo, 1994), (Daganzo, 1995) assumes a constant cell outflow during the entire period p. It is, however, also important in sensing data-based estimation. In non-homogeneous conditions, a loop-detector placed at the upstream cell boundary may observe different traffic conditions than

one placed at the downstream boundary. If the conditions are homogeneous, both loop-detectors observe the same conditions and a loop-detector placed at any location within the cell is representative for the conditions in the entire cell. Furthermore, the variation in individual vehicle speeds v can increase when the traffic conditions are non-homogeneous and non-stationary. This can lead to a larger estimation uncertainty when estimating the traffic conditions based on individual probe speeds.

In reality, traffic is non-homogeneous and non-stationary (Immers & Logghe, 2002). Such traffic conditions can still be expressed in the macroscopic traffic flow variables, but these variables may be incapable of uniquely describing the traffic conditions. For instance, in terms of the (traditional) macroscopic variables an area in which vehicles are decelerating due to downstream congestion (jam inflow) and in which vehicles are accelerating when leaving a jam may be the same, while in reality the conditions differ.

To capture the conditions which are non-homogeneous or non-stationary, extra traffic variables can be used. In this research, we add a single extra traffic variable related to the non-homogeneity of traffic, i.e., the change in speed over space. Although it is possible to add more variables, adding this single variable suffices for the analysis conducted in our experiments.

2.3 Sensing data-based mean speed estimation

The explanation of sensing data-based mean speed estimation is split into three parts. First, we discuss the traffic sensing data considered in this research. Second, the estimation approach to obtain the mean speed from the sensing data is presented. And third, we discuss how the estimation error distribution can be described.

2.3.1 Traffic sensing data characteristics

Seo (2015) states that we can regard traffic data collection as a special case of traffic state estimation. In the procedure to obtain traffic data from raw sensor signals exogenous assumptions are required. In this research, the starting point is traffic sensing data. We assume that these data do not contain errors. This assumption allows us to study the errors induced due to differences between the sensing data characteristics and desired estimate characteristics, and incomplete description of the relation between the two.

We consider two types of traffic sensing data, i.e., loop-detector data and probe speed data. The characteristics of the loop-detector data are based on the loop-detector data available in the Netherlands, i.e., lane-specific one-minute aggregated (time-mean) speeds u_l^T and flows q_l (Knoop & Hoogendoorn, 2012). Following (Wang & Papageorgiou, 2005), the loop-detectors are located at the downstream boundary of discrete road segments. In line with (Nanthawichit et al., 2003) and (Herrera & Bayen, 2010), we consider instantaneous individual vehicle speeds from probe vehicles, i.e., v_n where n describe the vehicle id. It is assumed that the probes are observed at the end of each period. These data can be collected from GPS-enabled mobile phones and navigation systems.

2.3.2 Estimation approach

The desired estimation output has specific characteristics, i.e., variable type and spatial/temporal characteristics. In this research, we estimate the mean speed u for a cell (discrete road segment) i and period p, i.e., $u(i,p)$. This desired output is estimated based on the two traffic sensing data-types discussed above.

The traffic sensing data (partly) and desired output differ in terms of variable type and spatial/temporal characteristics. Therefore, we have to define models to estimate $u(i,p)$ based on the sensing data. The models used in this research are taken from prior research, i.e., (Knoop & Hoogendoorn, 2012), (Nanthawichit et al., 2003).

The loop-detector data and probe speed data-based estimates are respectively denoted as $\hat{u}_{\text{ld}}$ and $\hat{u}_{\text{probe}}$. Based on the loop-detector data, the speed is estimated by taking the weighted harmonic mean of the lane-specific speeds (Knoop & Hoogendoorn, 2012). Here, we consider the loop-detector data which relates to the cell and period for which u is estimated.

$$\hat{u}_{\text{ld}} = \frac{\sum_{l=1}^{\lambda} q_l}{\sum_{l=1}^{\lambda} \frac{q_l}{u_l^T}} \tag{2.1}$$

We consider the mean v_n of the j number of vehicles observed in a specific cell and period as the probe data-based u estimate, i.e., $\hat{u}_{\text{probe}}$ (Nanthawichit et al., 2003).

$$\hat{u}_{\text{probe}} = \frac{1}{j} \sum_{n=1}^{j} v_n \tag{2.2}$$

2.3.3 Estimation error distribution

The traffic sensing data-based estimates may differ from the true $u(i,p)$. We denote this difference as the sensing data-based estimation error. The characteristics of these errors may be described using the error distribution.

We describe the estimation error distribution using four statistics, namely the mean, variance, skewness and kurtosis. The mean, variance, skewness and kurtosis respectively relate to the first, second, third and fourth standardized moments of a distribution (Ramsey et al., 2004). The skewness addresses the symmetry of a distribution and the kurtosis provides information related to the peakedness, or alternatively the 'fat tails', of a distribution (Ramsey et al., 2004). For perfect Gaussian distribution the skewness and kurtosis are respectively equal to 0 and 3. By means of the Jarque-Bera (JB) test (Jarque & Bera, 1987), we can test for normality. The null hypothesis of the JB-test is normality.

2.4 Experimental set-up

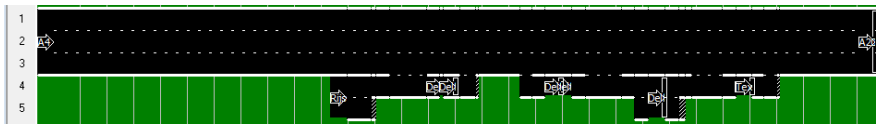
The objective of the experiments is to expose the characteristics of the data-based mean speed estimation errors. In this section, we discuss the data used in this research and explain the conducted experiments.

2.4.1 Data collection

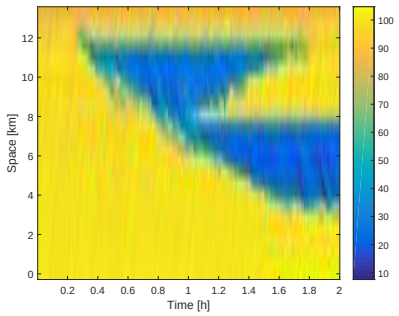
In this research we consider synthesized data collected using the microscopic simulation program FOSIM. The microscopic models and calibration used in FOSIM are described in (Dijker & Knoppers, 2006). Furthermore, it is validated for Dutch freeways (Minderhoud & Kirwan, 2001), (Henkens et al., 2017). FOSIM allows us to retrieve trajectory data for each individual vehicle. The trajectory data are used for two purposes. Firstly, we construct the traffic sensing data, i.e., loop-detector data and probe speed data, with the characteristics described in the previous section. Secondly, we construct the ground truth, as will be explained below. Combined, these allow us to obtain the traffic sensing data-based estimation errors and evaluate their characteristics.

Real trajectory data-sets are scarce and often limited in terms of spatial and temporal coverage. For instance, the NGSIM (Colyar & Halkias, 2007) trajectory dataset covers a study area of approximately 640 m for a 45 minute period. FOSIM allows us to simulate traffic for a much larger spatial and temporal coverage.

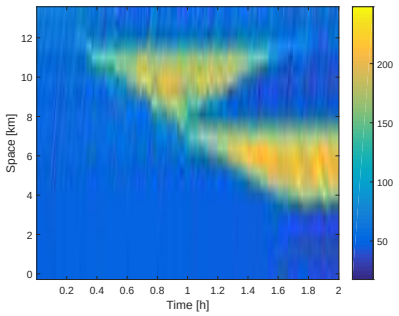
We consider a schematized road stretch of the Dutch A13 freeway from The Hague to Rotterdam which has a speed limit of 100 km/h in our experiments. The length of the road stretch is 13,878 m, has five on-ramps and four off-ramps. The road is discretized in 24 cells with lengths ranging from 520 to 770 m. We consider a two-hour time-domain which is discretized into periods of 15 seconds, $T = 15/3600$ h. This discretization is based on the approach followed by Wang & Papageorgiou (2005). The road layout and the traffic conditions in terms of u and k are shown in Figure 2.1[1]. Within this space-time domain, two standing queues are observed.



(a) Road lay-out in FOSIM



(b) Mean speed u km/h



(c) Density k veh/km

Figure 2.1: The schematized road stretch in FOSIM (above) and the true traffic conditions (below) in terms of the mean speed u and density k.

[1]The equations used to determine u and k are provided in Section 2.4.2.

2.4.2 Ground truth

The ground truth is important to describe the true traffic conditions in a discrete area in the space-time domain and determine the estimation errors by comparing the data-based estimates with the ground truth. As explained before, we describe the macroscopic traffic conditions by four variables, namely the mean speed, density, flow and change in speed over space. The former three variables are determined using Edie (1965) generalized formulation.

$$k = \frac{\sum_n r_n}{\Delta T} \tag{2.3}$$

$$q = \frac{\sum_n d_n}{\Delta T} \tag{2.4}$$

$$u = \frac{q}{k} \tag{2.5}$$

where the time spend and distance traveled by vehicle n within the space-time area are respectively denoted by r_n and d_n.

The change in speed over space is obtained by performing an OLS regression over the end-of-period individual vehicle speeds. For this purpose, we consider the following equation: $v_n = \theta_0 + \theta_1 x_n$. The change in speed over space is captured in θ_1.

Given the ground truth u and data-based estimate $\hat{u}$, the estimation error η can be determined.

$$\eta = u - \hat{u} \tag{2.6}$$

2.4.3 Evaluation of estimation error characteristics

In this research, we evaluate the dependency of loop-detector and probe data-based estimation error distributions on the traffic conditions. Here, the probe data-based estimates are based on individual probe observations. In reality we may have multiple probe observations for a given combination of cell and period. Therefore, we will also consider the dependency of the probe data-based estimation error characteristics on the observed fraction of the traffic flow. This sensing data-set characteristic is denoted by the number of probe observations.

Dependency on traffic conditions

The traffic conditions can be described using the traffic flow variables. The estimation errors may be directly explained based on the traffic flow variables using, for instance, linear regression. Problems with this approach are that the explanatory variables are correlated (as described by fundamental diagrams, e.g., (Smulders, 1990)) and non-linear relations may exist between the explanatory and dependent variables. Instead of considering the traffic flow variables as independent variables, we want to identify the different traffic conditions in the considered data-set based on the combination of these variables. For this purpose, the true traffic conditions are grouped into a discrete number of clusters using the K-means clustering algorithm. Next, the estimation error characteristics of the observations assigned to

each cluster can be described. This approach allows us to get an insight into the dependency of the estimation errors on the traffic conditions.

Using the K-means clustering algorithm (Tan et al., 2006) each area in time and space is assigned to one of the defined K number of clusters based on the traffic conditions. These conditions are described by three variables, namely the mean speed u, density k and change in individual vehicle speeds over space $\partial v/\partial x$. The vector $z^{(o)}$ contains these three variables for observation o, where o relates to a single area in space-time. The macroscopic traffic variable flow q is left out because u and k combined contain this information, i.e., $q = ku$, and in contrast to q are able to uniquely describe the homogeneous and stationary traffic conditions.

K-means clustering follows an iterative procedure to minimize a cost function J_{cost}. This iterative procedure is susceptible to local optima. This means that the algorithm can converge to different solutions. To overcome this problem, the algorithm is applied 10 times and the solution with the lowest cost J_{cost} is selected.

We consider the cost function given by (2.7). This function considers the mean squared difference between the individual observations $z^{(o)}$ and centroid of the assigned cluster $\mu_{c^{(o)}}$ for the total number of observations m. Here, the cluster centroid $\mu_{c^{(o)}}$ is defined as the mean values for each variable of all observations assigned to cluster c.

$$J_{cost} = \frac{1}{m} \sum_{o=1}^{m} \| z^{(o)} - \mu_{c^{(o)}} \|^2 \tag{2.7}$$

Before applying the clustering algorithm the explanatory variables (features) are scaled. If feature scaling is not applied it is possible that features with larger absolute difference will dominate the clustering. Therefore, for each feature the z-scores are considered. The z-score of an observation is equal to the difference between the observation and the mean of all observations divided by the standard deviation of all observations.

To find the optimal K, we plot J_{cost} as a function of K. Increasing K will decrease the cost since a more refined clustering is possible. However, this plot allows us to visually compare benefits in terms of J_{cost} of adding more clusters. The optimal K is selected by searching for a kink, which is referred to as the elbow, in the plot. Up till this point adding a cluster yields a relatively large benefit, thus decrease in J_{cost}, while the added value of increasing the number of clusters is limited. Therefore, the selected number of clusters is at the location of the elbow. This selection procedure is subjective as we have to define what a kink is and what is not. Therefore, it is also important to interpret the cluster characteristics and see if they make sense. As will be shown in the results, the selection procedure works well in our application.

After each observation has been assigned to a cluster, the estimation error distribution characteristics can be determined per cluster. We are specifically interested in the differences between clusters. Here the cluster characteristics, as described by the cluster centroids, can be compared with the error characteristics, as described by the error statistics.

We assume that the loop-detectors are located at the downstream boundary of each road segment. Based on this data a single u estimate is obtained for each combination of i and p, $\{i, p\}$. Furthermore, in this part of the research we consider every possible individual probe data-based speed estimate. The number of vehicles on road segment i at the end of period

p is defined as $N(i,p)$. Therefore, for each combination of $\{i,p\}$, one loop-detector and $N(i,p)$ probe data-based estimates are obtained.

Dependency on observed fraction of the traffic flow

The probe data-based estimate, see (2.2), and thus the probe data-based estimation error depends on the number of probe observations j. We are interested in the effect of j on estimation error characteristics. Here, we focus on the estimation error variance, as this is an important feature in traffic state estimation methodologies which apply a KF and which were discussed in the introduction, i.e., (Wang & Papageorgiou, 2005), (Nanthawichit et al., 2003), (Herrera & Bayen, 2010), (Van Hinsbergen et al., 2012).

Models to explain probe data-based estimation error variance We will compare two models which describe the influence of the number of probe observations j on the estimation error variance. The difference between these models relates to the expected correlation between estimation errors of individual probe observations.

The first model assumes that individual probe data-based estimates have the same error variance and the estimation errors are not correlated. In contrast, the second model does assume that there is a correlation between the individual estimation errors. The rationale behind the second model is as follows. The mean speed is dependent on the speeds of all vehicles within the considered area in space-time. In this case, the difference between an individual vehicle speed v_n of vehicle n with the mean speed u, i.e., the estimation error, is expected to be correlated with the difference between v_m and u, i.e., the estimation error based on v of vehicle m. For example, if we have two observations, the difference with respect to the mean (error) of the two observations has a correlation of minus one.

The assumptions discussed above are used to analytically derive the two models. For both models, it is assumed that v are Gaussian distributed with mean u and variance σ_v^2. Given this distribution, we can say that the estimation error distribution based on single probe observations is a zero-mean Gaussian distribution with variance σ_v^2. This variance is constant for a given area in the space-time domain, but can differ between areas, e.g., due to traffic conditions. In the first model each observation is seen as an independent observation with no relation to other observations, thus $E[\varepsilon_n \varepsilon_m] = 0$ for $n \neq m$, where ε_n denotes the difference between, i.e., error, u and v_n. In the second model we take into account that the estimation error based on different observations are (negatively) correlated. Each probe vehicle is given an equal probability of being observed, which means the observation sample is taken from a random draw. Based on this assumption we say that the expected covariance between two different probe observations, e.g., of vehicles n and m is constant, thus $E[\varepsilon_n \varepsilon_m] = c$ for $n \neq m$.

The difference between the two models is the assumption of the size of the sample for which we are interested in the mean. No correlation corresponds with the assumption that the observations are drawn from an infinity sample, while a finite sample yields results in a (negative) correlation between errors. Therefore, the two models will be denoted as Assumed Infinite Sample (AIS) and Assumed Finite Sample (AFS) models in the remainder of this article. Based on the assumptions stated above, the analytical derivations result in the AIS-model and AFS-model.

$$E\left[\eta^2\right]_{\text{AIS}} = \frac{1}{j}\sigma_v^2 \tag{2.8}$$

$$E\left[\eta^2\right]_{\text{AFS}} = \frac{N-j}{j\,(N-1)}\sigma_v^2 \tag{2.9}$$

The derivations of these models are shown in Appendix 2.A.

True probe data-based estimation error variance The two models which are proposed to describe the probe data-based estimation error variance are dependent on the observed number of vehicles j, total number of vehicles N and variance of individual vehicle speeds σ_v^2. To evaluate and compare the fit of the two models, we want to compare the model estimates with the true error variance. The true error variance is approximated using a Monte Carlo Experiment (MCE) and is therefore denoted as MCE. For each area in space and time, thus combination of i and p, we require a MCE-based error variance approximation for each potential value of j, i.e., σ_{ipj}^2. To obtain σ_{ipj}^2 the following procedure is followed for each potential value of j.

1. Random draw of j observations from $N(i,p)$ probes.
2. Calculate estimation error for each set using $\eta = u - 1/j\sum_{n=1}^{j} v_n$.
3. Repeat steps 1 and 2, 500 times.
4. Calculate error variance σ_{ipj}^2 of the set of 500 estimation errors.

The MCE-based error variance approximation, i.e., σ_{ipj}^2, is defined as the error variance observed in the data. Note that this is still an approximation based on the 500 sample sets which are randomly drawn. If the experiment is performed again with new random draws the error variances may slightly differ. It is expected that the MCE becomes more accurate when the number of random draws increases. For this paper, we consider the provided MCE over 500 runs as the representative ground truth.

Comparision of the models We want to compare the error variance observed in the data, i.e., the MCE approximation σ_{ipj}^2, with the two model-based estimates of the error variance, i.e., $\hat{\sigma}_{ipj}^2$. To get an initial insight into the differences between and the fit of the two models, two discrete areas in the space-time domain are selected. These are the representative areas, thus combinations of i and p, for the clusters with the lowest and highest k, which is related to N, are selected. To obtain a representative area, we consider the cost function given by (2.7). The contribution of individual areas can be computed using $z^{(o)}$ and $\mu_{c^{(o)}}$. Per cluster, we select the area, i.e., observation o, which has the smallest squared difference with the cluster centroid, i.e., $\mu_{c^{(o)}}$, as the representative area.

To get an overall insight, for both models, the Mean Absolute Percentage Error (MAPE) is calculated. The following equation is used to calculate the MAPE.

$$MAPE = \frac{1}{P}\sum_{p=1}^{P}\frac{1}{I}\sum_{i=1}^{I}\frac{1}{N(i,p)}\sum_{j=1}^{N(i,p)}\frac{|\sigma_{ipj}^2 - \hat{\sigma}_{ipj}^2|}{\sigma_{ipj}^2}\times 100\% \tag{2.10}$$

We will consider the overall MAPE together with the MAPE for bins of the penetration rate with a size of 10%. By considering the MAPE for different bins, we are able to discuss the effect the penetration rate on the accuracy of the two model-based estimates of the error variance. The choice for a bin-size of 10% is not crucial for this discussion. Both for the overall fit as for the fit per penetration rate bin, the penetration rate is capped at 90%. This limit is imposed to overcome the problem that the Percentage Error (PE), and thus the MAPE, goes to infinity when $j = N$. In this case σ^2_{ipj} will be equal to zero.

2.5 Results and findings

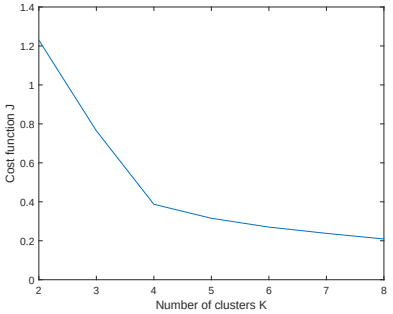
The results and findings of the experiments presented in the previous section are discussed here. We first consider the loop-detector and probe data-based estimation error dependency on the traffic conditions. Next, we zoom in into the error variance of the probe data-based estimates for different numbers of probe observations and traffic conditions.

2.5.1 Dependency on traffic conditions

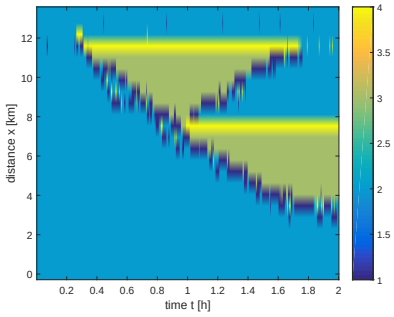
The relation between K and the cost is shown in Figure 2.2a. The elbow is observed for $K = 4$, which is therefore selected as the optimal K. In Figure 2.2b the cluster classification is shown in the space-time domain. Furthermore, the cluster centroids are provided in Table 2.1. Based on these centroids and the comparison between Figures 2.1 and 2.2b, the clusters are interpreted. For interpretation based on k and q it is important to know that we are considering a three-lane road stretch, i.e., $\lambda = 3$. In the cluster interpretation the $\partial v/\partial x$ plays an important role. A value close to zero corresponds with (near) homogeneous conditions, which can be either free-flow or congestion based on the other variables. Throughout the remainder of this paper we will refer to the free-flow and congested space-time areas as homogeneous traffic conditions. A negative value means that the speed decreases when moving downstream. This means that vehicles are (or have to) decelerating, which can correspond with the inflow of a congested area or jam. Vice versa, a positive value can correspond with jam outflow.

The clustering of areas in space and time based on the traffic conditions allows us to evaluate the sensing data-based estimation error dependency on traffic conditions. For each cluster, the loop-detector and probe data-based estimation error characteristics are described. By means of the mean, variance, skewness and kurtosis of the independent errors we gain insight into the distribution of these errors. Furthermore, we test for normality, thus whether the errors are Gaussian distributed, using the JB-test. The results are given in Table 2.1.

The first observation is that for each error distribution the null hypothesis of Gaussian distributed error is rejected, thus JB-test = 1. The non-homogeneous conditions, i.e., jam in- and outflow, yield challenges for data-based traffic state estimation in a discrete estimation mesh. Loop-detector data-based estimates can be biased in these conditions. This is caused by the combination of loop-detector location and change in speed over space. Following existing literature we placed the loop-detectors at the downstream boundaries on the cells. If the speed decreases over space, as is the case for jam inflow, the mean speed is underestimated using loop-detector data. Vice versa, if the speed increases over space, as is the



(a) Cost vs. number of clusters



(b) Space-time plot clusters $K = 4$

Figure 2.2: Cost function dependent on the number of clusters and visualization of the cluster in the space-time domain for the optimal number of clusters $K = 4$.

Table 2.1: Interpretation of the four clusters based on the cluster centroids and the error distribution characteristics for probe and loop-detector data-based estimates per cluster.

		Clusters			
		1	2	3	4
		Jam-inflow	Free-flow	Congestion	Jam-outflow
Cluster centroids	u [km/h]	58.5	96.0	23.8	47.7
	k [veh/km]	87.1	51.3	187.4	108.9
	q [veh/h]	4596	4878	4369	5049
	$\partial v/\partial x$ [km/(km h)]	-0.108	-0.001	-0.002	0.096
	Num. obs.	601	8424	1889	630
Probe data error char.	Mean	0.55	-0.04	0.01	0.05
	Variance	784.0	87.9	48.8	455.9
	Skewness	0.06	0.72	-1.48	-0.19
	Kurtosis	2.99	6.91	17.00	2.67
	JB-test	1	1	1	1
	Num. obs.	31187	248461	213184	41982
Loop-Detector data error char.	Mean	15.17	-0.08	-0.40	-31.23
	Variance	115.2	20.1	15.4	107.5
	Skewness	0.51	1.76	-0.17	1.03
	Kurtosis	3.43	19.50	8.75	5.25
	JB-test	1	1	1	1
	Num. obs.	601	8424	1889	630

case for jam outflow, the mean speed is overestimated using loop-detector data. Challenges also arise for probe data-based estimation, but these are a result of other effects. Caused by variation in speed over space in addition to the variation in speed due to vehicle and driver

heterogeneity, the total variance of the individual vehicle speeds increases with respect to homogeneous conditions. This yields an increased probe data-based estimation uncertainty. For the homogeneous traffic conditions, we make a distinction between free-flow and congested conditions. For probe data-based estimates, the uncertainty is larger in free-flow with respect to congestion. This is caused by the larger speed variation due to vehicle and driver heterogeneity in free-flow conditions. A similar relation is visible for loop-detector data-based estimates, however, here the relative difference between free-flow and congested conditions is smaller.

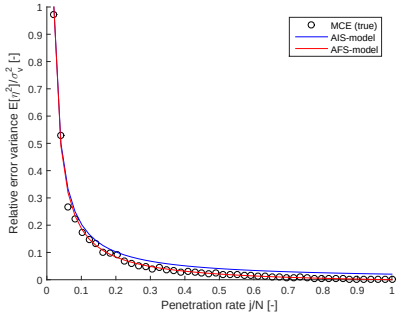
A direct comparison between the accuracy of loop-detector and probe data-based estimation in free-flow and congested conditions should not be made based on the results depicted in Table 2.1. These probe data-based estimates are based on individual vehicle speed data of a single probe, while we may observe more than one probe. For this reason, we will evaluate how the estimation error is affected when more than one probe observation is available below.

2.5.2 Dependency on observed fraction of the traffic flow

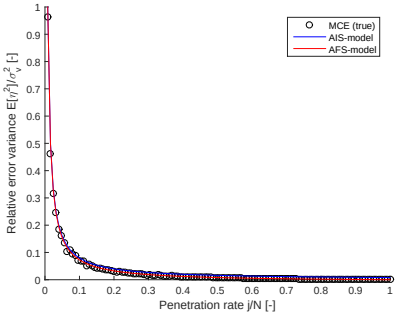
In this section we want to evaluate the performance of the two proposed model, i.e., the AIS-model and AFS-model, to describe the probe data-based estimation error variance. Figures 2.3a and 2.3b show representative areas in the space-time domain for free-flow and congested conditions, respectively $\{i,p\} = \{18,76\}$ and $\{i,p\} = \{20,205\}$. Keep in mind that these figures relate to a single area in the space-time domain. Considering a single area can help to understand the difference between the two models and how these relate with the true estimation error variance. In line with the findings from the previous section, the estimation error variance for an individual observation is larger in free-flow than in congestion. Furthermore, we can see that N is smaller in free-flow than in congestion. In the figures, it can be observed that the MCE sometimes (slightly) increases with the number of observations, while this is not expected for the true error variance. This can be explained be the limited set of random draws, i.e., 500, used to obtain the MCE-based error variance.

The AIS-model has the same shape based on j, but differs based on penetration rate. The relative accuracy of the AIS-model decreases when j increases. The reason may be that the AIS-model fails to capture the important (negative) correlation between estimation errors. The AFS-model does include these effects, yielding in a higher accuracy. It is clear that the AFS-model is dependent on the penetration rate j/N, as for $j = N$ the estimated error variance is zero. Furthermore, the effect of having more observations, thus when j increases, becomes also clear when comparing the two figures. In Figure 2.3b the AFS-model decreases more rapidly as a function of j/N then in Figure 2.3a. This can be explained by the fact that in the congested case j is larger than in the free-flow case for a given penetration rate j/N. As the MCE-based error variance follows this line, we may say that this effect seems to explain the true error variance. If N increases, the AIS-model seems to become more accurate. This makes sense as the AFS-model approaches the AIS-model for larger values of N and they are the same if $N \to \infty$. If one considers the AIS-model and for a given penetration rate, the probe data-based estimate thus becomes more accurate in terms of the point estimate and the error variance description.

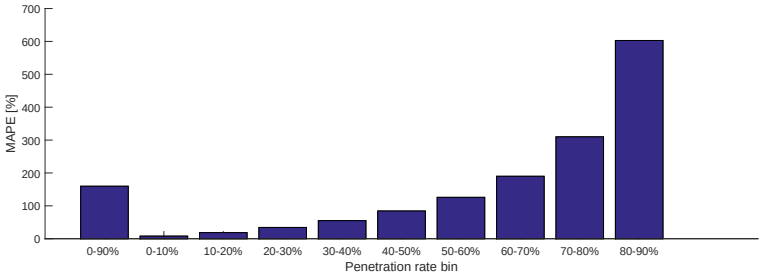
Up to now, we considered two observations to gain an insight in the two models and MCE-based error variance. However, we are interested in the overall fit and whether the



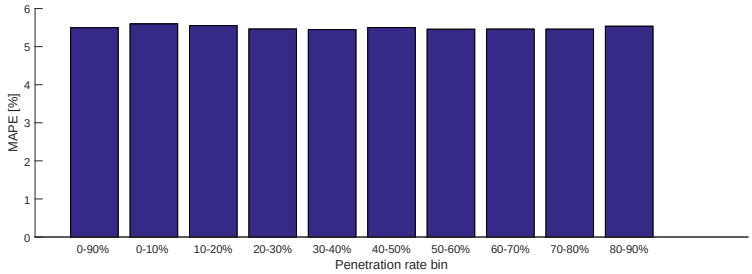
(a) Free-flow, $N = 49$ veh, $\sigma_v^2 = 52.9$ km^2/h^2



(b) Congestion, $N = 125$ veh, $\sigma_v^2 = 32.5$ km^2/h^2



(c) Mean Absolute Percentage Error (MAPE) for the AIS-model



(d) Mean Absolute Percentage Error (MAPE) for the AFS-model

Figure 2.3: Visualization of the AIS- and AFS-model fit and shape for an individual representative free-flow and congested area in the space-time domain. Furthermore, the total model fit, in terms of MAPE, is given for both models.

discussed characteristics hold for the entire sample. The MAPE is considered to describe the overall and penetration rate bin-specific fit. These are depicted in Figures 2.3c and 2.3d for respectively the AIS- and AFS-models.

Figures 2.3c and 2.3d shows that the AFS-model has a much better fit than the AIS-model. The AIS-model especially has problems describing the estimation error variance at higher penetration rates. At these penetration rates, the negative correlation between estimation error of individual probe observation becomes a more import factor. As the

AIS-model misses this factor while it is included by the AFS-model, the AIS-model is outperformed by the AFS-model. The MAPE of the AFS-model is relatively constant over the penetration rate bins.

2.6 Conclusions and discussion

In this research the traffic sensing data-based estimation error characteristics are evaluated by means of experiments. We focus on two combinations of a desired estimation output, type of traffic sensing data and estimation approach. Here, we estimate the mean speed in a discrete space-time mesh based on loop-detector data and probe speed data.

In the experiments, we observe a relation between the estimation error characteristics and traffic conditions and sensing data characteristics. We find that the extent to which traffic conditions are non-homogeneous negatively influences the estimation error characteristics. For instance, for both data-types we observe larger error variances in jam in- and outflow than in free-flow and congested conditions. Also, our loop-detector data-based estimates are biased in non-homogeneous traffic conditions. Furthermore, we show that it is valuable to take into account the correlation between estimation error to describe the probe data-based estimation error variance based on multiple probe observations.

Our experiments are conducted with the microscopic simulation program FOSIM. The results are thus influenced by the use of this simulation environment. However, we expect that the findings in this paper are also applicable to real-world applications as our explanations relate to real-world situations. As an example we can consider the effects of the level on non-homogeneity of the traffic conditions on the estimation errors in term of the mean and variance. We try to explain these effects based on the non-homogeneity itself which occurs both in real world as simulated environments. Furthermore, we think that the notion that the estimation errors can be explained based on certain features is of general importance. Even if one deals with different circumstances, e.g., other data-types or desired estimation output, it is valuable to consider this notion and try to explain the estimation errors.

Knowledge related to estimation errors is valuable when combining (fusing) different types of information or when weighing different alternatives. Examples of the latter are control decisions within a dynamic traffic management system or routing decisions of road users. This knowledge can lead to improved performance for these different applications without requiring additional (expensive) sensing data. Implementing knowledge related to estimation errors only marginally adds to the computation cost and does not require development of new, complex methodologies. However, before this can be put into practice, research is required to expose the added value for performance of applications in which the estimates are used as an input, e.g., traffic state estimation or control.

2.A. Derivation the AIS- and AFS-models

The individual vehicle speed within a cell and period are assumed to be Gaussian distributed with mean u and variance σ_v^2:

$$v_n = u + \varepsilon_n \qquad \varepsilon_n \sim N\left(0, \sigma_v^2\right) \tag{2.11}$$

The mean speed u is equal to the estimated mean speed $\hat{u}$ plus the estimation error η:

$$u = \hat{u} + \eta \tag{2.12}$$

We can estimate the u based on j number of individual vehicle speed observations:

$$\hat{u} = \frac{1}{j}\sum_{n=1}^{j} v_n \tag{2.13}$$

$$= u + \frac{1}{j}\sum_{n=1}^{j} \varepsilon_n \tag{2.14}$$

The estimation error variance, i.e., $E\left[\eta^2\right]$, is given by $E\left[(\hat{u} - E[\hat{u}])^2\right]$, thus:

$$E\left[\eta^2\right] = E\left[(\hat{u} - E[\hat{u}])^2\right] \tag{2.15}$$

$$= E\left[\left(u - u + \frac{1}{j}\sum_{n=1}^{j} \varepsilon_n\right)^2\right] \tag{2.16}$$

$$= E\left[\left(\frac{1}{j}\sum_{n=1}^{j} \varepsilon_n\right)^2\right] \tag{2.17}$$

Following the reasoning discussed in Section 2.4.3, we can make different assumptions related to (co)variance of the individual differences (errors) between the v_n and u. This results in the two models, i.e., Assumed Infinite Sample (AIS) and Assumed Finite Sample (AFS) models. Based on (2.11), for both models $E\left[\varepsilon_n \varepsilon_n\right] = \sigma_v^2$. However, the AIS assumes $E\left[\varepsilon_n \varepsilon_m\right] = 0$ for $n \neq m$, while AFS assumes $E\left[\varepsilon_n \varepsilon_m\right] = c$ for $n \neq m$.

Continuing from (2.17), for the AIS-model, the error variance becomes:

$$E\left[\eta^2\right]_{\text{AIS}} = \frac{1}{j^2} E\left[\sum_{n=1}^{j} \varepsilon_n^2\right] \tag{2.18}$$

$$= \frac{1}{j^2} j\sigma_v^2 \tag{2.19}$$

$$= \frac{1}{j}\sigma_v^2 \tag{2.20}$$

which thus yields (2.8).

Continuing from (2.17), for the AFS-model, the error variance becomes:

$$E\left[\eta^2\right]_{\mathrm{AFS}} = \frac{1}{j^2} E\left[\left(\sum_{n=1}^{j} \varepsilon_n\right)^2\right] \tag{2.21}$$

In contrast to the AIS-model, $E\left[\left(\sum_{n=1}^{j} \varepsilon_n\right)^2\right]$ does not simplify to $E\left[\sum_{n=1}^{j} \varepsilon_n^2\right]$. Instead the $\left(j^2 - j\right)$ number of terms of $E\left[\varepsilon_n \varepsilon_m\right]$ where $n \neq m$ are still of importance. Therefore, we obtain:

$$E\left[\eta^2\right]_{\mathrm{AFS}} = \frac{1}{j^2}\left(j\sigma_v^2 + (j^2 - j)c\right) \tag{2.22}$$

$$= \frac{1}{j}\sigma_v^2 + \frac{j-1}{j}c \tag{2.23}$$

The last step is to find c. For this purpose we say that the error variance is equal to zero when we observe all vehicles, i.e., $j = N$. This yields:

$$\frac{1}{N}\sigma_v^2 + \frac{N-1}{N}c = 0 \tag{2.24}$$

$$(N-1)\,c = -\sigma_v^2 \tag{2.25}$$

$$c = -\frac{1}{N-1}\sigma_v^2 \tag{2.26}$$

Next, we combine the relations above:

$$E\left[\eta^2\right]_{\mathrm{AFS}} = \frac{1}{j}\sigma_v^2 - \frac{j-1}{j}\frac{1}{N-1}\sigma_v^2 \tag{2.27}$$

$$= \frac{N-1}{j\,(N-1)}\sigma_v^2 - \frac{j-1}{j\,(N-1)}\sigma_v^2 \tag{2.28}$$

$$= \frac{N-j}{j\,(N-1)}\sigma_v^2 \tag{2.29}$$

Which yields (2.9).

Chapter 3

Vehicle accumulation estimation with detector count and probe mean speed data

This chapter is currently under review for journal publication.

In the remainder of this thesis, we consider the option to estimate the traffic state in or via the cumulative flow plane. As explained in the introduction, the cumulative flow is the governing macroscopic variable of traffic flow. Knowing the cumulative flow curves at two locations yields all information that is needed to obtain the (mean) travel-time and vehicle accumulation (i.e., number of vehicles between the two locations).

To estimate the traffic state in or via the cumulative flow plane, data that describe the change in cumulative flow (i.e., vehicle passings) are needed. These data can be collected using road-side sensing equipment such as loop-detectors. These data are expected to be imperfect, meaning that it is likely that there are observation errors (e.g., we may miss or double-count vehicle passings). In the scenario that we solely rely on road-side sensors (stationary observers) to collect the data, observation errors cause the so-called cumulative error problem, which needs to be mitigated. For this purpose, other data or information can be used.

Earlier studies proposed to use probe trajectory or vehicle re-identification data to mitigate the cumulative error problem. We will address a limitation of these disaggregated (probe) data in Chapter 4; however, first, we will consider the option to use aggregated probe data to mitigate the cumulative error problem. Due to privacy considerations, probe data providers often share aggregated data (e.g., mean speed or travel-time data) instead of disaggregated data (e.g., individual speed or trajectory data). Examples are the probe data provided by Google for their Better Cities program (Eland, 2015) or available in NDW (Uenk-Telgen, 2018), which both describe the mean speed for a road segment within a time-period. These or other data providers do not share details related to the disaggregated data and processing techniques that are used to construct the probe mean speed data. However, Chapter 2 has shown that incomplete information (e.g., only collect data from a fraction of the vehicles) cause potential errors in the resulting estimates, which in this case are the probe mean speed data.

3.1 Introduction

The number of vehicles in a queue and/or the queue length are important in traffic management applications. Intersection controllers can minimize delays based on real-time queue length estimates (Varaiya, 2013), (Le et al., 2015). Furthermore, a capacity drop can be avoided by limiting inflow into a freeway, and thus for instance storing vehicle on an on-ramp, e.g., (Papageorgiou, Hadj-Salem & Blosseville, 1991), (Hoogendoorn et al., 2013). This control approach requires real-time queue estimators to determine the remaining buffer-space in order to prevent spill-back from the buffers.

This study proposes a methodology that estimates the number of vehicles on an on-ramp, i.e., the vehicle accumulation, using detector count and probe mean speed data. Detectors positioned at the on-ramp boundaries observe the on-ramp inflow and outflow. These observations can be used to recursively estimate the vehicle accumulation. However, in this recursive approach, the current error is the sum of all previous errors, which causes the estimation error to grow over time (Van Lint & Hoogendoorn, 2015), (Amini et al., 2016). This is known as the cumulative error problem.

To mitigate the cumulative error problem, we can rely on different types of traffic sensing data. Stationary detectors can be used to obtain occupancies, e.g., Vigos et al. (2008). Furthermore, to mitigate the cumulative error probe trajectory data can be used, which in turn provide information on the travel-times between two locations, e.g., (Bhaskar et al., 2010) and (Van Lint & Hoogendoorn, 2015). The probe data used in those studies contains information on individual road users and are thus privacy-sensitive. To protect its users, probe data providers, e.g., Google, can process the data before sharing it with others, e.g., road authorities. Two common approaches to ensure privacy are aggregation and applying differential privacy filter (Cottrill, 2009), (Fries et al., 2012), (Kargl et al., 2013). It is more likely that these privacy-processed dataare shared with road-authorities than the privacy-sensitive probe trajectory data. Therefore, this study considers the use of aggregated probe speed data to mitigate the cumulative error problem.

Multiple studies have proposed methodologies to mitigate the cumulative error problem, e.g., Vigos et al. (2008), Bhaskar et al. (2010), Van Lint & Hoogendoorn (2015), Amini et al. (2016). Vigos et al. (2008) use an additional loop-detector that is positioned somewhere on a link and that provides time-occupancy measurements to counter the accumulation of flow observation errors. They apply a Kalman Filter (KF) (Kalman, 1960) to fuse the data. Bhaskar et al. (2010) and Van Lint & Hoogendoorn (2015) use probe trajectory data, i.e., the times at which a probe crosses the upstream and downstream boundary, to recover the cumulative errors. Both studies assume the cumulative flow value is constant over the probe trajectory, i.e., they assume 'no overtaking', which allows them to relate the upstream and downstream cumulative curves to each other. This potentially incorrect assumption is not needed if the change in cumulative flow over vehicle trajectories is observed (Van Erp, Knoop & Hoogendoorn, 2018b), (Van Erp et al., 2019). Amini et al. (2016) assume that they can determine whether the queue is empty or not. Based on this information they learn the detector bias and correct for the exposed bias in real-time estimation. We distinguish two processes that are applied in these studies to mitigate the cumulative error problem, i.e., (1) error recovery, e.g., Vigos et al. (2008), Bhaskar et al. (2010) and Van Lint & Hoogendoorn (2015), and (2) learning of the structural error, e.g., Amini et al. (2016). Both processes are applied in the methodology designed in this study. However, in contrast to the mentioned

studies, we use privacy-processed probe mean speed data. Working with other data requires new estimation methodologies, which are explained in the next section.

This article is organized as follows. Section 3.2 shows how the vehicle accumulation can be estimated recursively. Next, the proposed methodology is explained in Section 3.3. This is followed by an explanation of the case study (Section 3.4) and discussion of the results (Section 3.5). Finally, the conclusions are provided in Section 3.6.

Throughout this article different variables and notations are introduced. For the sake of clarity a table is included to provides an overview of all notations, see Table 3.1.

Table 3.1: Overview of symbols that are used in this article.

Description	Notation	Unit
Spatial-temporal properties:		
Position	x	m
Location upstream on-ramp boundary	x_0	m
Location downstream on-ramp boundary	x_L	m
On-ramp length	L	m
Time	t	s
Period	p	-
Short period	p_s	-
Duration of short (detector data) period	Δt	s
Long period	p_l	-
Duration of long (probe data) period	$\Delta\tau$	s
Short period within the long period	ι	-
Short period within two long periods	ν	-
Number of short periods within a long period	n	-
Traffic flow variables:		
Density	k	veh/m
Mean speed	u	m/s
Flow	q	veh/s
Cumulative flow	N	veh
Arrivals	A	veh
Departures	D	veh
Instantaneous vehicle accumulation	M	veh
Time-mean vehicle accumulation	$\overline{M}$	veh
Traffic sensing data and data-based estimates:		
Observed arrivals	$\hat{A}$	veh
Observed departures	$\hat{D}$	veh
Bias term estimate	b	-
Detector data-based estimates of Edie's flow	q_{detector}	veh/s
Probe mean speed	u_{probe}	m/s
Probe and Loop-detector (PL) based mean vehicle accumulation estimate	$\overline{M}_{PL}$	veh
Error recovery:		
Prior state vector	$\mathbf{x}^-$	veh
Posterior state vector	$\mathbf{x}^+$	veh
Individual state error variance	σ^2	veh^2
Induced error variance per departure	σ^2_{ind}	veh^2
Prior error covariance matrix	$\mathbf{P}^-$	veh^2
Posterior error covariance matrix	$\mathbf{P}^+$	veh^2
Error variance of $\overline{M}_{PL}$	$\mathbf{R}$	veh^2
Correction vector (Kalman Gain)	$\mathbf{K}$	-
Derivative of equation (3.9) to vector $\mathbf{x}$	$\mathbf{C}$	-
Online learning of the bias term:		
First-difference of $\overline{M}$	$\Delta\overline{M}$	veh
Individual estimate of b	$\tilde{b}$	-
Overall estimate of b	b	-
Update parameter	α	-

3.2 Recursive estimation of the vehicle accumulation and the cumulative error problem

The on-ramp vehicle accumulation M at time t, i.e., $M(t)$, is given by the difference between the cumulative flow N at the upstream and downstream on-ramp boundaries at time t:

$$M(t) = N(x_0,t) - N(x_L,t) \tag{3.1}$$

where x_0 and x_L respectively describe the location of the upstream and downstream on-ramp boundary. Note that the cumulative flow curves include the same set of vehicles, i.e., each vehicle that is counted at one of the boundaries should at one point in time also be counted at the other boundary.

Stationary detectors (e.g., loop-detectors) positioned at the on-ramp boundaries can observe the change in cumulative flow. These detectors can provide disaggregated (individual passings) or aggregated (number of passings) data. In this study, we use aggregated loop-detector data that describes the number of passings within a time period p that has a duration of Δt. The true arrivals $A(p)$ and departures $D(p)$ respectively denote the true inflow (at the upstream on-ramp boundary) and outflow (at the downstream on-ramp boundary) in p. The difference, i.e., $A(p) - D(p)$, describes the change in vehicle accumulation between the start and end of period p:

$$M(p) = M(p-1) + A(p) - D(p) \tag{3.2}$$

where $M(p-1)$ is the vehicle accumulation at the end of the period $p-1$ and thus at the start of period p. This recursive equation can be rewritten as a function of the initial M, i.e., $M(0) = N(x_0,0) - N(x_L,0)$, and the sum of the arrival and departure till and including p:

$$M(p) = M(0) + \sum_{i=1}^{p} (A(i) - D(i)) \tag{3.3}$$

If the detector count data are perfect, i.e., there are no observation errors, and we know $M(0)$, equation (3.3) yields perfect estimates. However, in reality, data will not be perfect, e.g., we may miss or double-count vehicles (Knoop et al., 2012). Due to the cumulative nature of equation (3.3), over time, errors in A and D lead to a large error in $M(p)$. This is known as the cumulative error problem.

3.3 Methodology to estimate the vehicle accumulation

In this section, the methodology to estimate the on-ramp vehicle accumulation based on detector count and probe mean speed data is explained. Below, we first provide a description of the data characteristics. Next, we explain the different elements of the proposed methodology and how these fit together. Finally, in the subsections, these elements are explained in detail.

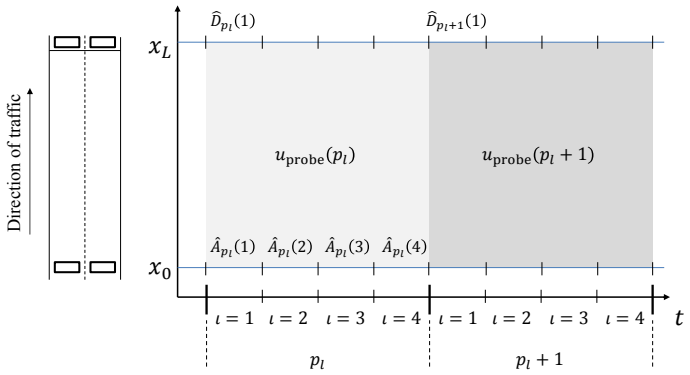


Figure 3.1: Illustration of the detector count and probe mean speed data characteristics and notations.

To explain the data characteristics, we consider Figure 3.1. This figure shows a two-lane on-ramp on the left side. Here, the detectors used to collect detector count data are illustrated with four rectangular boxes. These are positioned at the upstream and downstream on-ramp boundaries (x_0 and x_L). As the presented methodology can be used in a ramp-metering application, for which we want to estimate the number of vehicles in the queue, x_L is positioned directly downstream of the ramp-metering stop-line.

The detector and probe data are available at different frequencies. Probe data are available for long periods p_l with duration $\Delta\tau$. These data describe the mean speed during a long periods p_l for the segment between x_0 and x_L. To denote the probe mean speed during long period p_l, we use $u_{\text{probe}}(p_l)$. Detector data relate to short periods with duration Δt. The number of short periods within a long period is given by $n = \Delta\tau/\Delta t$. In the proposed methodology, it is assumed that n is an integer. Within a long period p_l, ι is used to describe the short periods starting from $\iota = 1$. In the example shown in Figure 3.1, n is equal to four. The detector counts are either observations of the arrivals A (upstream boundary) or departures D (downstream boundary). To denote the observed arrivals and departures in short period ι in long period p_l, we respectively use $\hat{A}_{p_l}(\iota)$ and $\hat{D}_{p_l}(\iota)$. In Figure 3.1, the observed departures are shown for the first short periods within long periods p_l and p_l+1, i.e., $\hat{D}_{p_l}(1)$ and $\hat{D}_{p_l+1}(1)$. Furthermore, all observed arrivals within long period p_l are shown, i.e., $\hat{A}_{p_l}(1)$, $\hat{A}_{p_l}(2)$, $\hat{A}_{p_l}(3)$ and $\hat{A}_{p_l}(4)$.

The basis of the presented methodology is a high-frequency recursive estimator that uses detector count data. Due the recursive nature of this estimator, we have to mitigate the cumulative error problem. For this purpose, low-frequency non-recursive estimates are used, which are based on detector count and probe mean speed data. These reference estimates are used for two processes: (1) recovery of the cumulative error and (2) online learning of the bias in the high-frequency recursive estimator. In error recovery, we recover the cumulative error build up in previous periods. This influences the initial vehicle accumulation used to estimate the vehicle accumulation for the next periods. In online learning of the bias term, we opt to learn the size of the structural error. We expect that the future detector data also contains this structural error and take it into account in the recursive estimator.

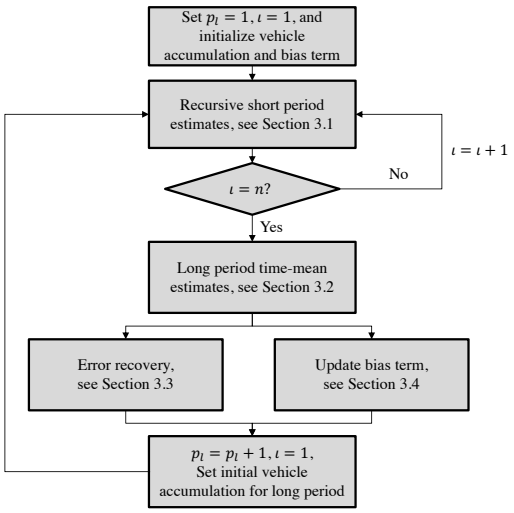


Figure 3.2: Flowchart of the estimation approach.

Figure 3.2 provides a flowchart that illustrates how the different elements of the methodology fit together. In the flowchart two loops are observed. These loops relate to (1) the short periods within a long period and (2) the long periods. In the first loop, the vehicle accumulation is recursively estimated using detector count data, see Section 3.3.1. At the end of the long period, i.e., $\iota = n$, the two processes aimed at mitigating the cumulative error problem are applied. Both processes require long-period time-mean vehicle accumulation estimates based on the recursive detector count estimates and non-recursive detector count and probe mean speed estimates. Section 3.3.2 explains how these estimates are obtained. Next, the two processes, which are explained in Sections 3.3.3 and 3.3.4, are applied. After recovering the cumulative error and updating the bias term, we move to the next long period, i.e., $p_l = p_l + 1$ and $\iota = 1$.

3.3.1 Real-time high-frequency estimation

The basic recursive equation to estimate the vehicle accumulation is given by equation (3.2) in Section 3.2. This equation is extended in two ways. First, we take a physical limitation of the vehicle accumulation into account. Negative vehicle accumulations are physically impossible, i.e., M should always be greater or equal to zero. Second, we include a term that corrects for the structural error. Section 3.3.4 explains how this term that is exposed in previous long periods. For this purpose, we hypothesize that the structural error (bias) is proportional to the traffic flow. The rationale behind this hypothesis is that an error caused by missing or double counting vehicles is small if the traffic flow is small. Furthermore, lane-changing behavior, which influences the number of misses and double-counts, can change dependent on the traffic flow. Therefore, the estimation bias is equal to $-\hat{D}b$, where b can be interpreted as the number of vehicles that are underestimated leaving and/or overestimated entering the on-ramp per vehicle observed leaving the on-ramp. However, note that the description of the structural error, i.e., $-\hat{D}b$, may be changed. Although the

equations provided in the remainder of this paper use this description, where needed, our approach can be used to derive the alternative representations of the structural count error.

The two extensions yield the following recursive estimator:

$$M^{-}_{p_l}(\iota) = \max\left[M^{-}_{p_l}(\iota-1) + \hat{A}_{p_l}(\iota) - \hat{D}_{p_l}(\iota) - \hat{D}_{p_l}(\iota)b(p_l-1), 0\right] \tag{3.4}$$

where every long period p_l starts with the final estimates of the previous period, i.e., $M^{-}_{p_l}(0) = M^{+}_{p_l-1}(n)$ and $b(p_l-1)$. Throughout this section, the superscript $-$ is added to define that we are dealing with prior estimates. In Section 3.3.3, we update the estimates based on the low-frequency reference estimate, which yields the posterior estimates. However, as these posterior estimates will only be available at the end of the long period, in real-time applications we need to rely on the prior estimates.

3.3.2 Mean vehicle accumulation during a time-period

The methodology discussed in Section 3.3.1 estimates the instantaneous on-ramp vehicle accumulation, i.e., the number of vehicles on the on-ramp at a time instant. Alternatively, the mean vehicle accumulation can be described for a time period. The mean vehicle accumulation $\overline{M}$ in period p is equal to Edie's density (Edie, 1965) times the on-ramp length:

$$\overline{M}(p) = k(p)L \tag{3.5}$$

where $k(p)$ describes the mean density on the on-ramp during period p and L is the on-ramp length. Furthermore, k is equal to the quotient of Edie's flow q and mean speed u:

$$k(p) = \frac{q(p)}{u(p)} \tag{3.6}$$

Here, Edie's flow $q(p)$ and mean speed $u(p)$ relate to the same spatial-temporal areas as $k(p)$, i.e., the on-ramp during period p.

Mean vehicle accumulation estimate based on detector flow and probe mean speed data

Equations (3.5) and (3.6) can be used to estimate the mean vehicle accumulation based on detector count data and probe mean speed data. The probe mean speed data should describe Edie's mean speed for a road segment during a defined period. The exact spatial-temporal characteristics of these data, i.e., road segments and time-periods, will depend on the probe data provider. Furthermore, these providers estimate the mean speed based on data collected from a fraction of the road users and often use an unknown approach to process the data. Both can cause estimation errors. Furthermore, the detector count data can be used to approximate Edie's flow. For this purpose, the mean of the flow observed by the two detectors (which are positioned at the upstream and downstream on-ramp boundaries) during p_l is taken, i.e., $q_{\text{detector}}(p_l)$. Similar to the probe data, the estimates of Edie's flow can contain estimation errors due to incomplete information (i.e., only the on-ramp boundaries

are observed) and observation errors (e.g., missed or double counted vehicles). This yields a reference estimate for the mean vehicle accumulation, which is denoted as the Probe mean speed and Loop-detector count data (PL) estimator.

$$\overline{M}_{PL}(p_l) = \frac{q_{\text{detector}}(p_l)}{u_{\text{probe}}(p_l)} L \tag{3.7}$$

As Edie's flow and mean speed may not perfectly be described by respectively $q_{\text{detector}}(p_l)$ and $u_{\text{probe}}(p_l)$, the estimate mean vehicle accumulation $\overline{M}_{PL}(p_l)$ may not be perfectly estimated. The uncertainty in this reference estimate is taken into account in the 'error recovery' and 'online learning of the bias term' processes that are presented in Sections 3.3.3 and 3.3.4.

Based on high-frequency estimates

The time-mean vehicle accumulation is approximated based on the high-frequency estimates that are obtained using equation (3.4). Here, we assume that M changes linearly between consecutive estimates. The M at the start and end of short period ι are respectively given by $M(\iota-1)$ and $M(\iota)$. Based on the assumption stated above, the time-mean vehicle accumulation within this period, i.e., $\overline{M}(\iota)$, is given by:

$$\overline{M}(\iota) = \frac{1}{2}\left(M(\iota-1) + M(\iota)\right) \tag{3.8}$$

This $\overline{M}$ relates to the short period ι. To estimate $\overline{M}$ for the long period p_l, i.e., $\overline{M}(p_l)$, the average of all relevant $\overline{M}(\iota)$ within p_l is taken. This is defined as the time-mean vehicle accumulation for period p_l based on the high-frequency estimates, i.e., $\overline{M}^{-}(p_l)$:

$$\overline{M}^{-}(p_l) = \frac{1}{2}\frac{\Delta t}{\Delta \tau}\sum_{\iota=1}^{n}\left(M^{-}_{p_l}(\iota-1) + M^{-}_{p_l}(\iota)\right) \tag{3.9}$$

where Δt and $\Delta\tau$ are respectively the duration of the short and long period, e.g., 10 s and 300 s for the data used in our case study. Furthermore, the number of short periods within a long period is denoted by n, i.e., $n = \Delta\tau/\Delta t$.

3.3.3 Error recovery

Error recovery is important to limit the size of the estimation error. In this section, we explain how the PL-estimates are used to correct the high-frequency prior estimates. Within this process zero-mean Gaussian errors are assumed. Therefore, to describe the estimation error uncertainty we use the error (co)variance. The methodology used for error recovery resembles the use of a Kalman Filter (Kalman, 1960). However, instead of having a model-based prior estimate, our prior estimate is obtained using equation (3.4).

Vigos et al. (2008) also applied a Kalman Filter to estimate the vehicle accumulation. However, they use different data, i.e., detector occupancy data instead of probe mean speed data. These data are used to estimate the vehicle accumulation. This reference estimate and

the inflow and outflow data are available for the same time-periods, while in our case the reference estimates are obtained for longer periods. The differences in reference estimates and the periods for which these are available result in a different application of the Kalman Filter.

Prior estimate of the vehicle accumulation

Section 3.3.1 explains how the individual prior estimates of the vehicle accumulation $M^-_{p_l}(\imath)$ are obtained. This yields prior estimates for each of the n short periods within a long period. The $(n+1)\times 1$ vector $\mathbf{x}^-(p_l)$ contains the prior initial vehicle accumulation, i.e., $M^-_{p_l}(0)$ and the n vehicle accumulations at the end of each short period that are estimated using equation (3.4):

$$\mathbf{x}^-(p_l) = \left[M^-_{p_l}(0)\cdots M^-_{p_l}(n)\right] \tag{3.10}$$

In addition to $M^-_{p_l}(\imath)$, we also recursively estimate the error variance $\sigma^2_{p_l,-}(\imath)$ to each $M^-_{p_l}(\imath)$-estimate. Due to the recursive nature of equation (3.4), the error variance related to $M^-_{p_l}(\imath)$ is the error variance related to $M^-_{p_l}(\imath-1)$ plus an induced error variance. It is assumed that this induced error variance is proportional to the departures in the related period, i.e., $\hat{D}_{p_l}(\imath)$:

$$\sigma^2_{p_l,-}(\imath) = \sigma^2_{p_l,-}(\imath-1) + \hat{D}_{p_l}(\imath)\sigma^2_{ind} \tag{3.11}$$

where σ^2_{ind} is a parameter of the estimator that needs to be defined.

The error covariance matrix $\mathbf{P}$ describes the error variance of each individual estimate and the error covariance between estimates. The error variance of the individual estimates, i.e., the diagonal elements of $\mathbf{P}$, are obtained using equation (3.11). Due to the recursive nature of the high-frequency estimator, the off-diagonal elements of $\mathbf{P}$ will be non-zero. It is assumed that the errors induced in $\imath$ are independent of prior errors. Given this and previous assumptions, the prior error covariance matrix $\mathbf{P}^-(p_l)$ is given by:

$$\mathbf{P}^-(p_l) = \begin{bmatrix} \sigma^2_{p_l,-}(0) & \sigma^2_{p_l,-}(0) & \cdots & \sigma^2_{p_l,-}(0) \\ \sigma^2_{p_l,-}(0) & \sigma^2_{p_l,-}(1) & \cdots & \sigma^2_{p_l,-}(1) \\ \vdots & \vdots & \vdots & \vdots \\ \sigma^2_{p_l,-}(0) & \sigma^2_{p_l,-}(1) & \cdots & \sigma^2_{p_l,-}(n) \end{bmatrix} \tag{3.12}$$

Posterior estimate of the vehicle accumulation

The mean vehicle accumulation estimates for the long period based on $\mathbf{x}^-$, i.e., $\overline{M}^-(p_l)$, are obtained using equation (3.9). For the same period, $\overline{M}(p_l)$ is estimated by combining probe speed and loop-detector data. This estimate is denoted as $\overline{M}_{PL}(p_l)$. Based on the difference between $\overline{M}_{PL}(p_l)$ and $\overline{M}^-(p_l)$, $\mathbf{x}^-(p_l)$ is corrected. This yields the posterior estimate of the vehicle accumulation $\mathbf{x}^+(p_l)$:

$$\mathbf{x}^+(p_l) = \mathbf{x}^-(p_l) + \mathbf{K}(p_l)\left[\overline{M}_{PL}(p_l) - \overline{M}^-(p_l)\right] \tag{3.13}$$

The correction vector $\mathbf{K}(p_l)$, also known as the Kalman Gain when using a Kalman Filter, plays an important role. It depends on the prior error covariance matrix $\mathbf{P}^-(p_l)$ which relates to $\overline{M}^-(p_l)$ and the error variance of $\overline{M}_{PL}(p_l)$, i.e., $\mathbf{R}$:

$$\mathbf{K}(p_l) = \frac{\mathbf{P}^-(p_l)\mathbf{C}^T}{\mathbf{C}\mathbf{P}^-(p_l)\mathbf{C}^T + \mathbf{R}} \tag{3.14}$$

where $\mathbf{C}$ is the derivative of equation (3.9) to vector $\mathbf{x}$, which is thus a $1 \times (n+1)$ vector:

$$\mathbf{C} = \frac{\partial \overline{M}^-}{\partial \mathbf{x}} \tag{3.15}$$

which yields $\mathbf{C} = \left[\frac{\Delta t}{2\Delta\tau}\ \frac{\Delta t}{\Delta\tau}\ \frac{\Delta t}{\Delta\tau} \cdots \frac{\Delta t}{\Delta\tau}\ \frac{\Delta t}{\Delta\tau}\ \frac{\Delta t}{2\Delta\tau}\right]$.

Finally, the error covariance matrix which relates to $\mathbf{x}^+(p_l)$, i.e., $\mathbf{P}^+(p_l)$, is obtained by:

$$\mathbf{P}^+(p_l) = [I - \mathbf{K}(p_l)\mathbf{C}]\,\mathbf{P}^-(p_l) \tag{3.16}$$

The recovered final estimates

Error recovery at the end of long period p_l yields the initial values for the next long period, i.e., $p_l + 1$. The final estimates are given by $M^+_{p_l}(n) = \mathbf{x}^+(n)$ and $\sigma^2_{p_l,+}(n) = \mathbf{P}^+(p_l, n, n)$. These are used as initial values for $p_l + 1$, i.e., $M^-_{p_l+1}(0) = M^+_{p_l}(n)$ and $\sigma^2_{p_l+1,-}(0) = \sigma^2_{p_l,+}(n)$.

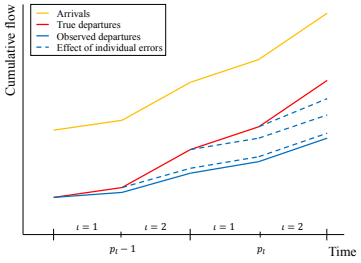
3.3.4 Online learning of the bias term

Information on the structural error (bias) in the detector count data can be used in recursive estimation of the on-ramp vehicle accumulation, see equation (3.4). Amini et al. (2016) proposed a methodology to learn this structural error (bias) by looking at the moments that a link is empty, i.e., the vehicle accumulation is zero. That methodology can only be applied when the link is empty, which may not occur regularly. We present a methodology to expose the structural errors in the detector count data using probe mean speed data. As the probe mean speed data is available on a regular basis (i.e., every probe data period), the estimated structural error (bias) can be updated every probe data period.

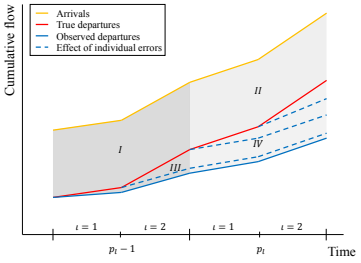
In this section, we explain how we learn the bias term b, which is part of equation (3.4, based on the low-frequency time-mean vehicle accumulation estimates. In this process, two steps are taken. First, we describe the approach used to obtain an individual estimate of b at the end of long period p_l, i.e., $\tilde{b}(p_l)$. Second, we explain how this individual estimate is used to update our prior belief of b, thereby yielding a combined estimate of b, i.e., $b(p_l)$.

Individual estimate of the bias term

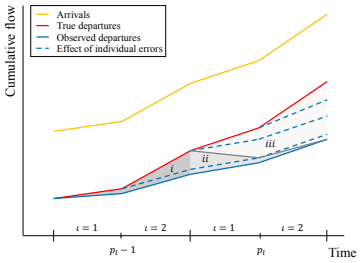
To explain our approach to obtain an individual estimate of the bias term a simplified situation is considered, see Figure 3.3. In this example, we only have two short periods per long period, i.e., $n = 2$.



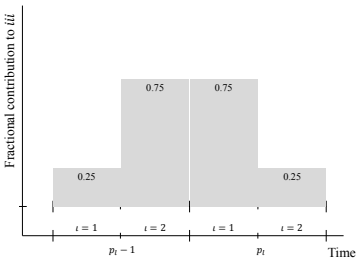
(a) Cumulative curves of the arrivals and departures



(b) Areas relevant for $\overline{M}_{PL}(p_l - 1)$ and $\overline{M}_{PL}(p_l)$



(c) Areas relevant for the contributions of the cumulative error



(d) fraction of p_l for which the individual errors contribute to area *iii*

Figure 3.3: Visualizations of the example used to clarify the procedure used for online learning of b.

Figure 3.3a illustrates the influence of errors induced in individual short periods for two long periods. The true cumulative arrival and departure curves are respectively shown by the orange (top) and red (middle) solid lines. The blue (bottom) line denotes the observed departures. Taking the difference between the red and blue line yields the cumulative error. In this example, it is assumed that there is no initial error. However, if there would be an initial error the blue line would be lowered with a fixed value over time. The cumulative error is the sum of the error induced in the different short periods. To visualize the error contributions of these individual periods blue dashed lines are used. Here, the difference between two consecutive blue dashed lines denotes the contribution of an individual short period.

The size of the dark and light gray areas in Figure 3.3b are respectively equal to time-mean vehicle accumulation based on the low-frequency estimator for $p_l - 1$ and p_l, i.e., $\overline{M}^-(p_l - 1)$ and $\overline{M}^-(p_l)$, times the duration of the long period $\Delta\tau$. Let us consider the difference between these area, i.e., $\left[\overline{M}^-(p_l) - \overline{M}^-(p_l - 1)\right]\Delta\tau = \Delta\overline{M}^-(p_l)\Delta\tau$. In Figure 3.3b, we distinguish four areas that are marked with *I*, *II*, *III* and *IV*. These areas either lie in the light or dark gray areas and are separated by the solid red line. The difference between areas *I* and *II* relates to the true first difference in time-mean vehicle accumulation,

i.e., $II - I = \Delta\overline{M}(p_l)\Delta\tau$. The difference between areas *III* and *IV* relates to the cumulative error.

The reference estimates, i.e., $\overline{M}_{PL}$, are not subjected to cumulative errors. The first difference of these estimates, i.e., $\Delta\overline{M}_{PL}(p_l)$, is equal to the actual difference in $\overline{M}$, i.e., $\Delta\overline{M}(p_l)$, and the resultant of the random error in the two periods. A structural fixed error in the reference estimate drops out as this is equal for both periods. As the expected value of the random errors is equal to zero, $E\left[\Delta\overline{M}_{PL}(p_l)\right] = \Delta\overline{M}(p_l)$. Therefore, we expect the difference between areas *I* and *II* to be equal to $\Delta\overline{M}_{PL}(p_l)\Delta\tau$. The contribution of the cumulative error in Figure 3.3b, i.e., $IV - III$, is thus expected to be equal to $\left[\Delta\overline{M}^{-}(p_l) - \Delta\overline{M}_{PL}(p_l)\right]\Delta\tau$.

Figure 3.3c focuses on the contribution of the cumulative error. In this visualization the error contribution in p_l, i.e., area *IV* in Figure 3.3b, is subdivided into areas *ii* and *iii*. To obtain *ii*, we draw straight lines in the short periods within p_l, which move one level down, i.e., to the first lower dashed line. This yields an area with an equal size as area *i*. Therefore, area *iii* is the difference between the areas *IV* and *III*, i.e., $IV - III$, which was given by $\left[\Delta\overline{M}^{-}(p_l) - \Delta\overline{M}_{PL}(p_l)\right]\Delta\tau$.

To obtain an estimate of the bias term b, we need to expose the contributions of error induced in individual short periods to area *iii*. Following equation (3.4), the expected induced error in a short period ν is equal to $b\hat{D}(\nu)$. The contributions of the errors in individual short periods within the two long periods is visualized in Figure 3.3 by the vertical distance between the two consecutive blue lines, e.g., the solid blue line and the bottom dashed blue line for the first short period. The fraction of long period p_l for which an error induced in an individual short period to area *iii*, which is denoted using o, is shown in Figure 3.3d. This allows us to describe the expected contribution to *iii* for an individual period ν, i.e., $b\hat{D}(\nu)o(\nu)\Delta\tau$. Combining all $2n$ short periods in $p_l - 1$ and p_l yields an estimate for area *iii*, i.e., $b\Delta\tau\sum_{\nu=1}^{2n}\hat{D}(\nu)o(\nu)$.

To obtain o, we consider the construction of the line between areas *ii* and *iii*. The approach used to construct this line is specifically chosen to explain o. As an example, let us consider the first short period. This error exists in the full long period p_l. However, the fractions for which it contributes to areas *ii* and *iii* are respectively 0.75 and 0.25. In other words, an error increase in the first period of 1 veh yields an increase of 0.25 veh times $\Delta\tau$ for area *iii*. The fractional contributions of all four short periods are shown in Figure 3.3d. The general function of o for all short periods ν within long periods $p_l - 1$ and p_l, i.e., $\nu = 1, \ldots, 2n$, is given by:

$$o(\nu) = \min\left(1, 2 + \frac{1/2 - \nu}{n}\right) - \max\left(0, 1 + \frac{1/2 - \nu}{n}\right) \tag{3.17}$$

This allows us to estimate the bias term b based on the mean vehicle accumulation estimates, the departures and fractional contribution of the individual errors to area *iii*:

$$\tilde{b}(p_l) = \frac{\Delta\overline{M}^{-}(p_l) - \Delta\overline{M}_{PL}(p_l)}{\sum_{\nu=1}^{2n}\hat{D}(\nu)o(\nu)} \tag{3.18}$$

where $\tilde{b}(p_l)$ denotes the individual estimate of b at the end of period p_l.

The sign of $\tilde{b}(p_l)$ is positive if $\Delta\overline{M}^{-}(p_l) > \Delta\overline{M}_{PL}(p_l)$. In this case, there is a positive drift, i.e., $\overline{M}^{-}$ is growing over time. In line with equation (3.4) a positive value of b is needed to compensate this positive drift.

Updating combined estimate of the bias term

Each first difference $\Delta\overline{M}(p_l)$, yields an estimate of b, i.e., $\tilde{b}(p_l)$. These estimates are combined into one estimate $b(p_l)$. For this purpose a recursive equation similar to Amini et al. (2016) is used, where we update our previous estimate, i.e., $b(p_l - 1)$, based on the current observation:

$$b(p_l) = b(p_l - 1) + \alpha \left[\tilde{b}(p_l) - b(p_l - 1)\right] \tag{3.19}$$

where α is the update parameter. This parameter may be defined based on the expected characteristics of the bias term. Amini et al. (2016) use an update parameter which is a positive decreasing sequence with specific properties as they assume b does not change over time, i.e., $b(t) = b$. In the empirical results provided later in this paper we observe that the bias changes over time. Therefore, our estimator assumes that b can change over time. Accordingly, more emphasis is placed on more recent observations. In this research, we will not investigate different schemes to define α as the resulting estimates differ too little to evaluate in case study. Therefore, we simply take a fixed value of α, which does satisfy the condition of placing more emphasis on more recent observations.

3.4 Case studies

The proposed methodology is tested in a simulation and an empirical data case study. The reason behind conducting these two different case studies lies in the individual strengths and weaknesses of the studies. The proposed methodology should be able to mitigate the cumulative error problem. For this purpose, real errors are desired over synthesized errors. Therefore, empirical data provides an unique possibility to assess the performance of the methodology. However, in the empirical data study, the ground truth (both in terms of on-ramp vehicle accumulation as in terms of the on-ramp flows and mean speed) is not available. Instead, we rely on reference data (i.e., the radar speed data) to evaluate the performance of the estimator. In the simulation study, all information is available, which allows us to evaluate the estimation performance (i.e., compare estimates with the ground truth). However, we need to define the (data) observation error characteristics, which may differ from the real-life error characteristics.

Both studies aim to estimate the vehicle accumulation on an on-ramp that is equipped with a ramp metering installation. The two-lane on-ramp considered in the empirical study is located in the Metropolitian Region Amsterdam and connects the urban arterial S106 to the freeway A10-S. The distance from the upstream on-ramp boundary to the ramp-metering stop bar is 214 m. A lane-drop is located 30 m downstream of the ramp-metering stop bar, which forces vehicles traveling on the left lane to change to the right lane. For the simulation study, traffic is simulated using VISSIM (Fellendorf & Vortisch, 2001). To represent an on-ramp with a ramp metering installation, we consider a single-lane road with a controllable

(a) Empirical study *(b)* Similation study

Figure 3.4: Lay-out of the on-ramps considered in the simulation and empirical case studies.

traffic signal. The length of the considered on-ramp, i.e., from the upstream detector to the ramp-metering stop bar and downstream detector, is 200 m. Figure 3.4 schematically shows the test sites for the two case studies.

The empirical and simulation case studies respectively consider four-hour and two-hour periods. In the empirical study, data are available for the evening peak period, i.e., 15:00-19:00h, on July 6th, 2016. Figure 3.5a shows the space-time plot of the radar speeds. Two periods in which lower speeds are observed stand out in this plot, i.e., (1) 15:55-16:00h and (2) 16:15-16:30h. The largest delays are observed for the second period. Therefore, within the full 15:00-19:00h period, we expect that the vehicle accumulations are largest between 16:15-16:30h. In the simulation study, traffic is simulated for a two-hour period. The demand (average inflow) during the full period is 800 veh/h. The metering installation is active during a 30 minute period, i.e., from $t = 3600$ s till $t = 5400$ s. This causes the vehicle accumulation to increase during this period, see Figure 3.5b. After metering has stopped, the on-ramp queue dissolves and the vehicle accumulation returns to the levels before metering.

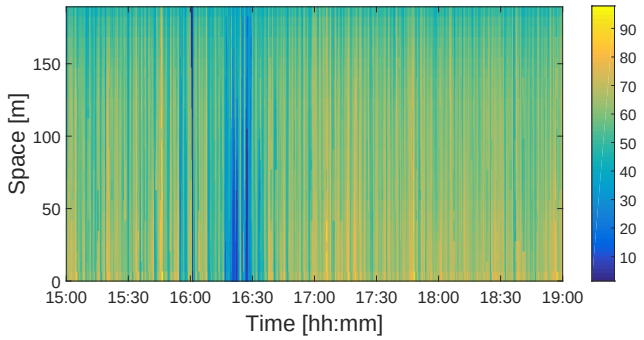
Below, we first discuss the data collection procedures and data characteristics, see Section 3.4.1. Next, Sections 3.4.2 and 3.4.3 respectively present the parameter values that are used for estimation and the procedure taken to evaluate the estimation performance.

3.4.1 Data collection and characteristics

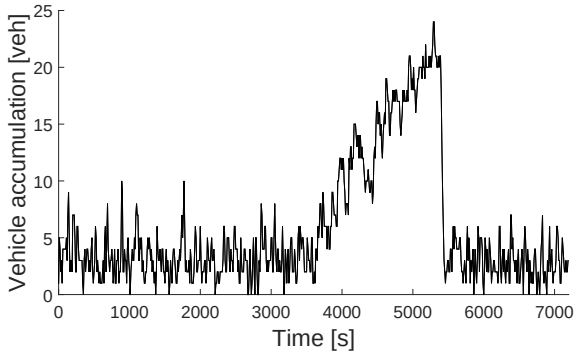
The two studies use the same data to estimate the vehicle accumulation, i.e., detector count data and probe mean speed data. For the empirical study, a third data-type is available, i.e., radar speed data (see Figure 3.5a). Unfortunately, no direct information on the true vehicle accumulation is available in the empirical case study; however, the radar speed data serve as reference data to evaluate the vehicle accumulation estimation performance.

Empirical study

The data used in the empirical study are obtained from Google and the Dutch road authorities (Rijkswaterstaat). In addition to the detector count and probe mean speed data, which are used to estimate the on-ramp vehicle accumulations, we have access to radar speed data, which is used for evaluation purposes. We assume that there is no data latency, i.e., the data is available directly at the end of the related time-period. However, the proposed methodologies can be adjusted to cope with a potential data latency in the probe data.



(a) Radar speeds in km/h (empirical study)



(b) True vehicle accumulation (simulation study)

Figure 3.5: Evaluation data (empirical study) and ground truth (simulation study) for the two case studies.

Probe mean speed data are provided by Google. This data describes the mean speed for road segments during 300 s periods. The on-ramp considered in the empirical study consist of two Google road segments. The lengths of these segments are 120 and 94 m, which yields a total on-ramp length of 214 m, see Figure 3.4a. Google has provided the probe speed data as part of its Better Cities program (Eland, 2015). The purpose of this program is to investigate the added value of Google data in tackling mobility problems. The source of these data are mobile devices using Google services, e.g., Android or Google Maps, for which the users explicitly agreed to share their location data with Google. To ensure privacy, Google aggregates the data and applies a differential privacy filter. The effect of this filter is limited in high traffic demand, e.g., peak, periods as studied in this research.

Detector count and radar speed data are provided by the Dutch road authorities (Rijkswaterstaat). These data are collected as part of the Praktijkproef (field operational test) Amsterdam ("Praktijkproef Amsterdam (PPA)", 2016). The radar data provide speed information for road segments of 7 m during 10 s periods. The process of obtaining this speed information is designed and applied by Rijkswaterstaat and related third parties. Details related to the accuracy of this process are not available to us; however, for our qualitative

evaluation, it suffices to observe periods where the speed is reduced due to the build-up of a queue. As shown in Figure 3.5a, these periods can be distinguished. The original detector data is known as VLOG data, which is used in practice to operate the ramp metering installations as well as the intersection controllers. This data-set describes changes in occupation of, i.e., whether a vehicle is located above, the detectors. Based on this information we derived the passings (flow) at the loop-detector locations. The period duration is selected in such a way that it coincides with the available evaluation (radar) data, i.e., 10 s. The detector positioned at the upstream boundary of the on-ramp was out-of-order in this period. Instead we use the relevant outflows of the upstream intersections, which cover all legal inflows and are separated from other flows.

Simulation study

The trajectory data obtained using VISSIM is used to construct loop-detector count and probe mean speed data. The temporal data characteristics are chosen such that they align with the data that is used in the empirical case study. This means that the detector count data and probe mean speed data are respectively collected for 10 s and 300 s periods.

Detector count data can be constructed by counting the number of vehicles passing the detector location with the 10 s periods. Count errors can be caused by missing or double-counting passing vehicles. To reconstruct these observation errors, we use a miss-probability p_0, double-count-probability p_2 and correct-probability p_1, where $p_1 = 1 - p_0 - p_2$. These probability are used in combination with a random draw to determine whether the individual passings are missed, correctly observed or double counted.

The probe mean speed data should aim to describe Edie's mean speed between the upstream and downstream on-ramp detectors for 300 s periods. The ability to do so will depend on the data availability and data processing algorithms. Probe data providers do not share the privacy and IP sensitive information related to these two factors, which makes it difficult to define the observation error characteristics that can be expected in practice. In this simulation study, we assume that 10 % of the vehicles serves as a probe vehicle and use Edie's definitions on those vehicles to estimate the mean speed. The resulting probe mean speed errors are thus a result of not always observing a representative subset of all vehicles.

3.4.2 Vehicle accumulation estimation

The methodology explained in Section 3.3 is applied to the data discussed above. As explained in Section 3.4.1, the empirical detector count and probe mean speed data respectively relate to 10 s and 300 s periods. The same data periods are chosen for the simulation study. Therefore, the short and long period durations are Δt = 10 s and $\Delta\tau$ = 300 s. Furthermore, the number of short periods within a long period is n = 30.

The error recovery process relies on the relative error (co)variances of the different estimates. Therefore, we need to define error variance for each departure, i.e., σ^2_{ind}, and error variance of the reference estimate, i.e., $\mathbf{R}$. As σ^2_{ind} relates to an individual departure and $\mathbf{R}$ relates to an estimate of $\overline{M}$ for a long-period, we assign a lower value to σ^2_{ind} than to $\mathbf{R}$, i.e., $\sigma^2_{ind} < \mathbf{R}$. The selected values are: $\sigma^2_{ind} = 0.125$ veh^2 (same for the two studies), $\mathbf{R} = 5.0$ veh^2 (empirical study) and $\mathbf{R} = 2.0$ veh^2 (simulation study). Different values of $\mathbf{R}$ are taken due to the different number of on-ramp lanes. Varying these parameters, while preserving

that σ^2_{ind} is an order of magnitude smaller than $\mathbf{R}$, leads to the same results in terms of the observed peak periods in vehicle accumulation. The update parameter α, which is used in online learning, is set to $\alpha = 0.25$. Varying this parameter has no real effects in the observed peak periods in vehicle accumulation. Evaluating other values of α and other schemes of updating the bias term b is left for future research due to space constraints.

The initial vehicle accumulation and related error variance, and the initial bias term have to be defined. The initial vehicle accumulation is respectively chosen to be $M_1^-(0) = 5$ veh and $M_1^-(0) = 0$ veh for the empirical and simulation studies. Different values are selected as the simulation study starts with an empty road, while this is not expected for the empirical study. The initial vehicle accumulation error variance is set to $\sigma^2_{1,-}(0) = 5.0$ veh^2 (empirical study) and $\sigma^2_{1,-}(0) = 2.0$ veh^2 (simulation study). Finally, in all studies, an initial bias term of zero, i.e., $b(0) = 0.00$, is chosen.

3.4.3 Evaluation of estimation performance

A more extensive evaluation of the estimation performance is possible in the simulation study compared to the empirical study due to the availability of the ground truth and the ability the define the observation error characteristics. In the simulation study the ground truth is available for the feature that we aim to estimate, i.e., the instantaneous vehicle accumulation, but also for intermediate features, e.g., the mean speed and on-ramp inflows and outflows. This allows us to visualize and quantify the errors through the full estimation process.

Simulation study

Seven scenarios that differ in terms of the detector count error probabilities are considered in the simulation study. Table 3.2 presents the miss and double-count probabilities of the upstream and downstream detectors for the seven scenarios. Furthermore, it describes the expected bias term for each scenario, which can be calculated with:

$$E[b] = \frac{p_0^D - p_2^D - \left(p_0^A - p_2^A\right)}{1 + p_2^D - p_0^D} \tag{3.20}$$

The first three scenarios all have an expected bias of zero. However, they differ in terms of whether the flow is expected to be correctly estimated (scenario 1), underestimated (scenario 2) or overestimated (scenario 3). In the fourth and fifth scenario, the upstream detector (which relates to the on-ramp inflow) is unbiased, but the downstream detector underestimates (scenario 4) or overestimates (scenario 5) the on-ramp outflow. This respectively results in an overestimation and underestimation of the vehicle accumulation and therefore should be corrected by a positive and negative bias term, see Table 3.2. In the last two scenarios, i.e., scenario 6 and 7, the upstream detector is biased, while the downstream detector is unbiased.

To understand how the proposed methodology works and gain initial insights in the estimation performance, we will analyze different time-series plots. Here, we look at (1) the difference in the upstream and downstream cumulative curves, (2) the reference estimator and its elements, (3) the estimated bias terms and (4) the vehicle accumulation estimates.

Table 3.2: The seven detector count error scenarios.

Scenario	p_0^A	p_2^A	p_0^D	p_2^D	$E[b]$
1	0.025	0.025	0.025	0.025	0.00
2	0.00	0.05	0.00	0.05	0.00
3	0.05	0.00	0.05	0.00	0.00
4	0.025	0.025	0.05	0.00	0.053
5	0.025	0.025	0.00	0.05	-0.048
6	0.05	0.00	0.025	0.025	-0.05
7	0.00	0.05	0.025	0.025	0.05

Due to space constraints, only three out of the seven scenarios are considered, i.e., scenario 1, 4 and 5. These scenarios have the same upstream detector error probabilities, but differ in terms of the downstream detector error probabilities. This allows us to study (1) an unbiased detector setting ($p_0^D = 0.025$ and $p_2^D = 0.025$), (2) a detector setting with positive bias ($p_0^D = 0.05$ and $p_2^D = 0.00$) and (3) a detector setting with negative bias ($p_0^D = 0.00$ and $p_2^D = 0.05$).

In addition to evaluating the time-series plots related to the three scenarios, the error statistics Root Mean Square Error (RMSE) and Mean Error (ME) are compared and discussed for all seven scenarios. The RMSE statistic is often considered in judging the estimation accuracy and provides insight in the uncertainty of the estimates. The ME statistic shows whether the vehicle accumulation is on average overestimated or underestimated. To study the effect of using the reference estimates for online learning the bias term, for each scenario, the vehicle accumulation is estimated with and without applying this process. As the online learning process requires time to learn the bias term, the RMSE and ME statistics are determined for all estimated after $t = 1800$ s.

Empirical study

In contrast to the simulation study, the estimation errors cannot be quantified in the empirical study due to the absence of a ground truth. However, the estimates throughout the estimation process are discussed and compared to the radar speed data (see Figure 3.5a). For this purpose, we first evaluate the different elements of the long-period reference estimates, i.e., probe mean travel-times, detector mean flows and the estimated mean vehicle accumulations. Second, we look at the difference between the cumulative inflow and outflow, which provides an initial insight into the potential detector count bias. Third, the process of learning the bias term is evaluated by showing the individual and overall bias term estimated. Finally, we compare the short-period vehicle accumulation estimates where the reference estimates are used for error recovery. Here, two options are considered, i.e., one with and one without online learning of the bias term.

Two outflow-detector scenarios are evaluated in this empirical study, i.e., a scenario where solely the right-lane detector is used to describe the on-ramp outflow (scenario 1) and a scenario in which both detectors are used to describe the on-ramp outflow (scenario 2). This choice is made because initial analysis indicate that the downstream detectors suffer

from a large number of double-counts, i.e., the two detectors both observe the same vehicle. This seems to be caused by a lane-drop (from two to one lane) downstream of the two detectors, see Figure 3.4a. Therefore, vehicles are likely to perform a lane-changing maneuver at the location of the detector. To overcome this problem, scenario 1 solely includes the counts from the lane that continues, i.e., the right lane.

The detector count data used in this study has a low quality as both the upstream and downstream detectors are subjected to limiting factors. The upstream detectors were out-of-order. This problem is solved by taking all relevant outflows of the upstream intersection, which induces additional uncertainties. Furthermore, the downstream detectors are located just upstream of a lane-drop, which is expected to result in many double-counts. We recommend that the proposed methodology is applied to situations in which the detector data have a higher quality. However, at the same time, we believe that it is interesting to show that the methodology still yields valuable insights when working with such low-quality data.

3.5 Results

This section presents the results of the simulation study (Section 3.5.1) and empirical study (Section 3.5.2).

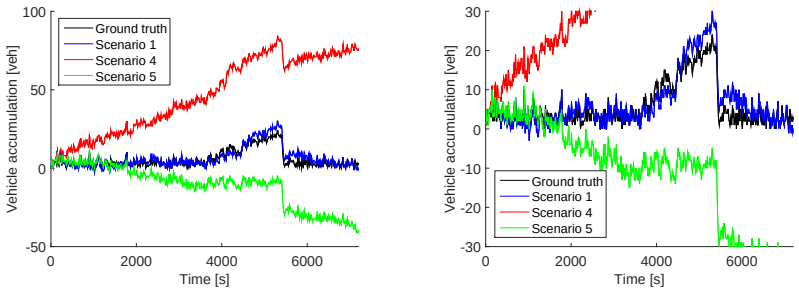
3.5.1 Simulation study

To understand how the proposed methodology works and gain initial insights in the estimation performance, we look at different time-series plots and overall error statistics. In the figures three scenarios are considered, which differ in terms of the detector observation error probabilities. In these scenarios, the upstream detector error probabilities are constant ($p_0^A = 0.025$ and $p_2^A = 0.025$), while the downstream detector error probabilities are differed to have (1) an unbiased detector ($p_0^D = 0.025$ and $p_2^D = 0.025$), (2) a detector with negative bias ($p_0^D = 0.05$ and $p_2^D = 0.00$) and (3) a detector with positive bias ($p_0^D = 0.00$ and $p_2^D = 0.05$). After discussing the figures related to these scenarios, the error statistics RMSE and ME are compared and discussed for the wider range of scenarios that were presented in Table 3.2.

Below, we will first evaluate the overall vehicle accumulation estimation performance (Section 3.5.1). Next, we dive into the intermediate steps of the estimation process, see Section 3.5.1.

Overall vehicle accumulation estimation performance

Both random and structural errors cause estimation errors when recursively estimating vehicle accumulations based on the cumulative inflow and outflow. This indicates that both processes used in the estimation methodology, i.e., error recovery and online learning of the structural error (bias), are important. Figure 3.6 shows the difference in upstream and downstream cumulative curves for three detector error scenarios. In this figure, the differences in detector biases can clearly be distinguished. The scenario in which both detectors are unbiased (i.e., scenario 1), no drift is observed. Figure 3.6b shows that there are still random errors, which cause the observed difference in cumulative detector counts (blue line) to temporarily move away from the ground truth (black line). For scenarios 4 and 5, we



(a) All detector error scenarios and the ground truth

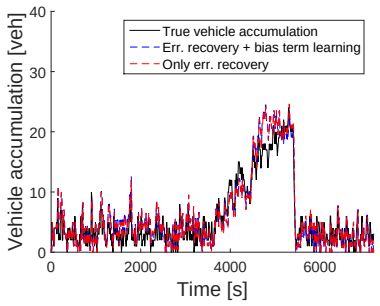
(b) Zoom to compare unbiased detector scenario to ground truth

Figure 3.6: Difference in upstream and downstream cumulative curves for three detector count error scenarios.

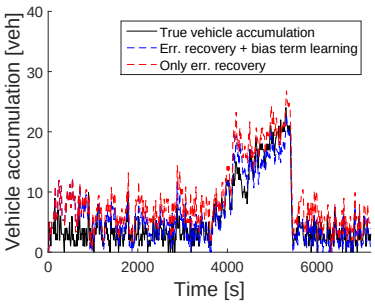
respectively see a systematic positive and negative drift. In case of a positive drift, the upstream detector counts more passings than the downstream detector. This holds for scenario 4, as the downstream detector has a negative bias, which means that it observes less than the actual number of passings.

Error recovery is valuable in any scenario, while online learning of the bias term is valuable in case of structural errors. Figures 3.7a, 3.7b and 3.7c show the true vehicle accumulation (black line) together with the estimates if we solely use the reference estimates for error recovery (dashed red line) and the estimates if we also use the reference estimates for online learning of the bias term (dashed blue line). The influence of learning the bias term is limited for scenario 1 as the expected bias is zero, i.e., the blue and red lines show similar patterns. For scenarios 4 and 5, online learning of the bias term improves the estimation accuracy after an initialization period. For scenario 4, not learning the bias term results in overestimation of the vehicle accumulation, see Figure 3.7b. For scenario 5, the effect is less clear as the estimator does not allow for negative vehicle accumulations, see equation (3.4); however, during the ramp metering period, i.e., between $t = 3600$ s and $t = 5400$ s, the red line underestimates the vehicle accumulation. Furthermore, the figures show that the proposed methodology recovers random errors at the end of the long periods. For instance, in scenario 1, some random errors result in an overestimation of the vehicle accumulation (e.g., in the period between $t = 4500$ s and $t = 4800$ s). In Figure 3.7a these errors are recovered shortly after, while in Figure 3.6b the random errors remain.

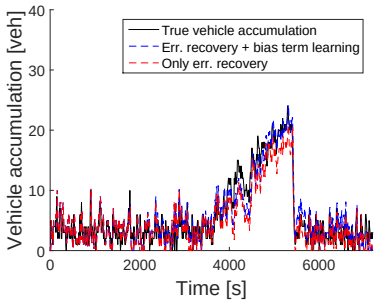
The error statistics RMSE and ME show similar patterns for the seven scenarios as the figures discussed above, Table 3.3. In scenarios 1 till 3, for which the bias term is expected to be zero, estimates are (in terms of RMSE) slightly better without online learning of the bias term as random errors result in non-zero bias term estimates, see Section 3.5.1 for more detailed analysis. For the other scenarios, online learning of the bias term does improves the estimation accuracy for the bias scenarios. As explained above, overestimation is not corrected by not allowing negative vehicle accumulations, while underestimation is. Therefore, the improvement is largest for the scenarios with a positive expected bias term (scenarios 4 and 7). In these scenarios working with a bias term of zero results in an overestimation of the vehicle accumulation, e.g., see Figure 3.7b.



(a) Scenario 1



(b) Scenario 4



(c) Scenario 5

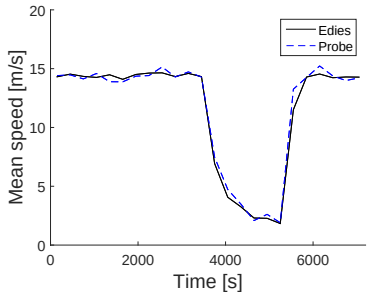
Figure 3.7: Vehicle accumulation estimates for three detector count error scenarios.

Table 3.3: Error statistics for the seven detector count error probabilities scenarios.

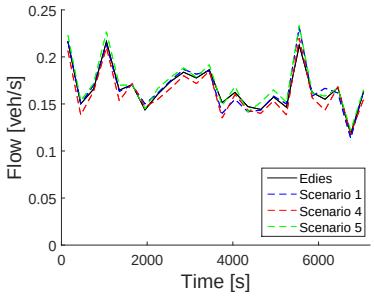
	Only error recovery		Error recovery and bias term learning	
Scenario	RMSE [veh]	ME [veh]	RMSE [veh]	ME [veh]
1	2.36	0.27	2.37	0.21
2	2.29	0.44	2.45	0.50
3	2.36	0.31	2.39	-0.01
4	3.88	3.00	2.34	0.14
5	2.74	-1.17	2.38	0.04
6	3.51	-1.80	2.89	-0.01
7	3.76	2.62	2.49	0.36

Assessment of individual parts in the estimation process

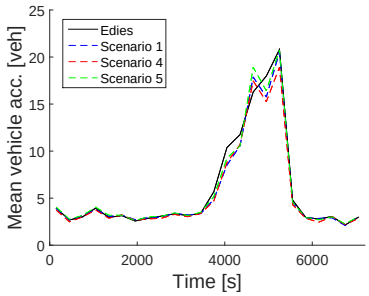
To gain a deeper understanding in the estimation approach, we dive deeper into two parts of the estimation process. First, we look at the mean vehicle accumulation estimates that are obtained using detector and probe data. Next, we assess how these reference estimates are used to learn the bias term.



(a) Edie's and probe mean speeds



(b) Edie's and detector mean flows

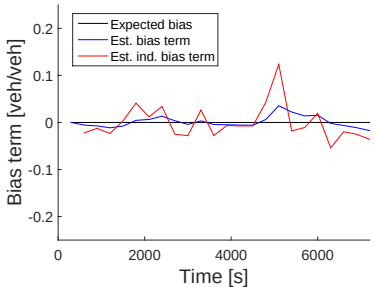


(c) True and estimated mean vehicle accumulation

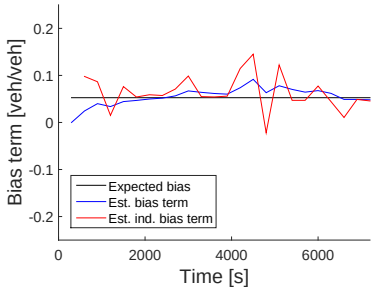
Figure 3.8: Mean speed, mean flow and mean vehicle accumulation estimates based on detector count and probe mean speed data.

The reference estimator is most sensitive to probe mean speed errors in the periods for which the mean speed is low. Figure 3.8 shows that the mean vehicle accumulation during long periods is estimated accurately using probe mean speed data and detector data in the three scenarios. The largest errors in this reference estimates is observed for the period between $t = 4500$ s and $t = 5100$ s. In the first 300 s period, the mean vehicle accumulation is overestimated, while in the second period it is underestimated. Let us considered the second period (i.e., between $t = 4800$ s and $t = 5100$ s) for explanatory purposes. As shown in Figure 3.8a Edie's mean speed is low during this period, but is overestimated in the probe mean speed data (i.e., 2.28 m/s versus 2.61 m/s). This results in an underestimation of the mean vehicle accumulation during this period, see Figure 3.8c. Figure 3.8b shows that a positive and negative bias of the downstream detector counts (i.e., scenarios 4 and 5) respectively yield an overestimation and underestimation of Edie's flow. This also results in and overestimation and underestimation of the mean vehicle accumulation; however, as can be seen in Figure 3.8c, this effect is limited. Therefore, we do not expect that the detector errors have a large effect on the error recovery and online learning of the bias term processes.

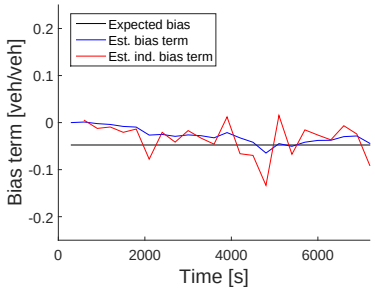
The proposed methodology to learn the bias term yields bias terms estimates that move towards the expected bias term; however, random errors cause the bias term estimates to



(a) Scenario 1



(b) Scenario 4

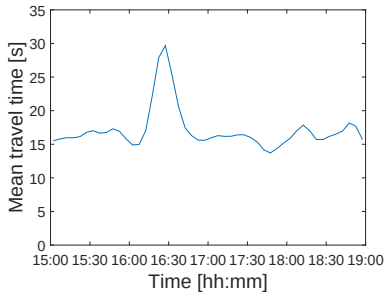


(c) Scenario 5

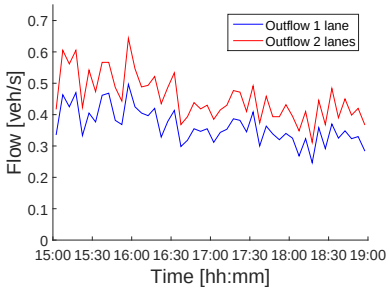
Figure 3.9: Bias term estimates for three detector count error scenarios.

vary. Figures 3.9a, 3.9b and 3.9c show the bias term estimates together with the expected bias term, see equation (3.20), for the three scenarios. In all scenarios, the initial bias term is zero, i.e., $b(0) = 0.00$. Therefore, the estimated bias terms (blue lines) all start at zero. The individual bias term estimates (red lines) all fluctuate around the expected bias term, which causes the estimated bias terms to gradually approach and fluctuate around these expected values. Variations in the individual bias term estimates seem to be caused by two factors: (1) inaccuracies in the reference estimator (Figure 3.8c) and (2) random detector count errors (e.g., Figure 3.6b). For example, the above-discussed errors in the reference estimator cause relatively large fluctuations in the individual bias term estimates between $t = 4800$ s and $t = 5400$ s. Furthermore, the random errors that we observed in scenario 1 between $t = 4500$ s and $t = 4800$ s yield an increase in individual bias term estimates.

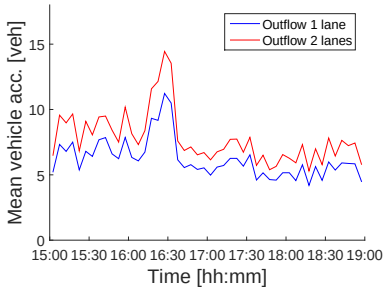
Note that the effect of the individual bias term estimates $\tilde{b}$ (red lines) on the overall bias term estimates b (blue lines), which is described by equation (3.19), can be altered. The individual bias term estimates should be leading in deciding how the overall bias term estimates are updated. For instance, in the case that $\tilde{b}$ seem to vary around a single value, which is the case in this simulation study, it may be logical to weigh all previous $\tilde{b}$ instead of assigning more weight to more recent $\tilde{b}$, which is the case in equation (3.19). As it depends on the actual detector count errors and the main contribution lies in estimating the individual bias term estimates $\tilde{b}$, it is left to researchers or practitioners that apply the methodology to define the best procedure to update b for their specific situation.



(a) Probe mean travel-time



(b) Detector mean flows



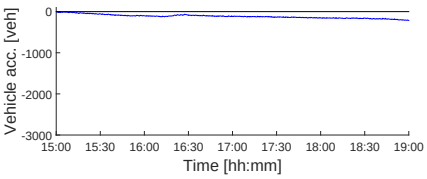
(c) Estimated mean vehicle accumulation

Figure 3.10: Reference estimates for the 300s probe mean speed periods.

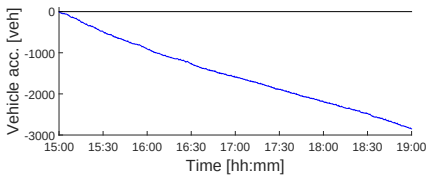
3.5.2 Empirical study

Two outflow-detector scenarios are evaluated in this empirical study, i.e., a scenario where solely the right-lane detector is used to describe the on-ramp outflow (scenario 1) and a scenario in which both detectors are used to describe the on-ramp outflow (scenario 2). This findings discussed below show that in both scenarios we can distinguish peaks in vehicle accumulation that coincide periods in which lower speeds were observed by the radar (see Figure 3.5a). Furthermore, the different estimates provide insight in the cause of the detector count errors and its characteristics depending on the on-ramp traffic conditions.

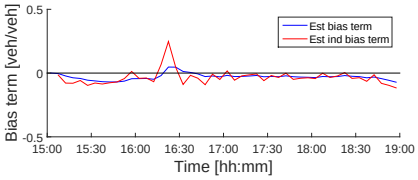
For each long-period (300 s), the mean travel-time $\overline{TT}$, mean flow $\bar{q}$ and mean vehicle accumulation $\overline{M}$ are estimated. These estimates are shown in Figure 3.10. In line with the radar speed plot, see Figure 3.5a, the probe speed data clearly describe the decrease in speed and thereby increase in travel-time between 16:15-16:30h. However, it seems that the probe speeds are smoothed over time. Figure 3.10a shows a smooth increase before and decrease after the peak travel-time in 16:25-16:30h, while the radar data shows a more abrupt decrease in travel-time after 16:30h. The mean flow estimates (which directly rely on the detector counts) fluctuate more than the mean travel-times, see Figure 3.10b, which results in more fluctuations in the mean vehicle accumulation estimates. As an additional (downstream) detector is used, the mean flow and mean vehicle accumulations are highest for scenario 2.



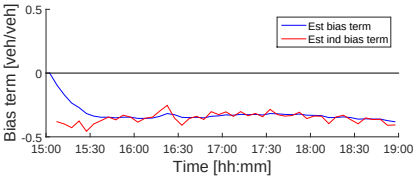
(a) Difference cumulative inflow and outflow, scenario 1



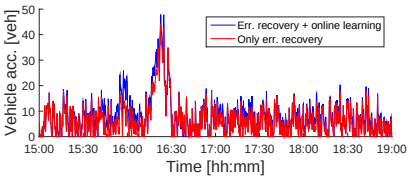
(b) Difference cumulative inflow and outflow, scenario 2



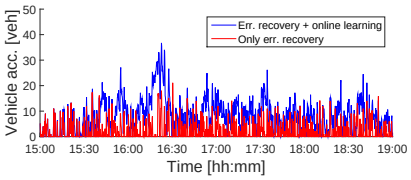
(c) Bias term estimates, scenario 1



(d) Bias term estimates, scenario 2



(e) Vehicle accumulation, scenario 1



(f) Vehicle accumulation, scenario 2

Figure 3.11: Difference in cumulative flows, and bias and vehicle accumualation estimates for the two scenarios considered in the empirical case study.

The difference in cumulative inflow and outflow, see Figure 3.11a and 3.11b, provides evidence of a structural error. In these figures the difference (i.e., $\sum_p A(p) - D(p)$) grows relatively constant over time for both scenarios. For scenario 1 the final difference in cumulative flows is -213 veh, while for scenario 2 this difference is -2849 veh. However, it does not indicate whether this structural error is caused by the upstream detectors, downstream detectors or both. Therefore, the data itself does not indicate that the mean flow (see Figure 3.10b) is underestimated or overestimated.

Figures 3.11c and 3.11d, which show the bias term estimates, also find negative bias terms, where the bias term in scenario 1 is much smaller than the bias term in scenario 2. In both scenarios, a peak is observed in the individual bias term estimates (red line) in the period 16:15-16:30h. Let us consider the bias term estimate for scenario 1. Here, the sign of the individual bias term estimates changes. Negative bias terms are estimated for nearly all other long-periods, which indicates that the departures are overestimated with respect to the arrivals. As discussed above, we expect that the probe mean speed data are smoothed and that the decrease in speed is not fully captured. In turn this yields an underestimation of the mean vehicle accumulation and an underestimation of the change in vehicle accumulation ($\Delta\overline{M}_{PL}$). This yields overestimation of the individual bias term, which is an artificial (i.e., not correct) increase of the bias term estimate as it is caused by an

error in the reference estimate. An error in the probe mean speed influences the bias term estimates for both scenarios. However, it does not explain, why the peak is much higher in scenario 1 compared to scenario 2 (i.e., Figure 3.11c versus 3.11d). A potential cause lies in the combination of on-ramp traffic conditions, road lay-out and detector placement. The detectors are located upstream of a lane drop. In free-flow conditions vehicles traveling on the left lane are expected to be in the process of lane-changing at the detector location, while in congested conditions vehicles can change lanes downstream of the detector. In this case, departures are missed when we solely consider the right-lane outflow detector, which is the case in scenario 1. An underestimation of the departures have a positive effect on the true bias term and can therefore contribute to the peak in Figure 3.11c.

The short-period instantaneous vehicle accumulation estimates are shown in Figures 3.11e and 3.11f. In these figures the effect of online-learning of the bias term is larger in scenario 2. Without online learning of the bias term, the vehicle accumulation estimates remain close to zero in scenario 2 and no clear peaks are observed (see red line in Figure 3.11f). For scenario 1, the differences between the estimates with or without online learning are small. However, while in both cases a clear peak is observed between 16:15-16:30h, the estimates without online learning do not show a peak between 15:55-16:00h, which is shown by the estimates with online learning and indicated by the radar speed plot. Two major differences are observed for the estimates with online learning of the bias term. Overall, the estimates in scenario 2 are larger, which is in line with the higher values of the reference mean vehicle accumulation estimates (see Figure 3.10c). However, a higher peak is reached in scenario 1 between 16:15-16:30h. As discussed above, in this period, the departures may be underestimated, which results in an overestimation of the vehicle accumulation.

3.6 Conclusions

This paper proposes methodologies to estimate the on-ramp vehicle accumulation by fusing detector measurements and probe mean speed information. Estimating the vehicle accumulation solely based on detector measurements is sensitive to the cumulative error problem. The probe mean speeds can be used to periodically learn the structural error, i.e., bias, and recover the cumulative error.

The proposed methodologies are evaluated in an empirical and a simulation study. These studies show that adding information in the form of probe mean speed to detector count data successfully mitigates the cumulative error problem. The estimates no longer drift away from feasible values. Both applications of the mean speed information, i.e., error recovery and online learning of the bias term, contribute to improving the quality of the vehicle accumulation estimates. Random (i.e., non-structural) count errors are not immediately be observed, but the resulting cumulative error can be recovered once the probe mean speed data is available.

The process of updating the bias term can be altered depending on the error characteristics which are exposed by the individual bias term estimates. Random count errors also influence the process of learning the bias term, thereby causing the estimated bias term to move around the true bias term. As this phenomena is exposed by the individual bias term estimates time-series, one may take it into account in updating the bias term. It is therefore recommended to adjust this part of the methodology for the specific application.

This study shows that probe mean speed data provide valuable information to estimate the vehicle accumulation and learn the structural error. The data provide information on the relative count error between the detectors positioned at the upstream and downstream on-ramp boundaries. Other data may be more suitable to expose the individual detector count errors, e.g., probe trajectory data, but may not be available for road-authorities.

Improvements in the real-time vehicle accumulation estimation is beneficial for the effectiveness of dynamic traffic management systems. The estimates are, for example, important in traffic management systems which temporarily store traffic on buffers to temporarily reduce the inflow to a downstream bottleneck. Accurate estimates can prevent spill-back or underusing the buffer capacity.

The mean speed information used in the empirical study is provided by Google as part of their Better Cities program. The results are thus dependent on both the quality of the Google data and the proposed methodologies. However, if other providers are able to provide reliable mean speed information, the proposed methodologies can be applied with their data.

Chapter 4

Estimation of the change in cumulative flow over probe trajectories using detector data

This chapter is currently under review for journal publication.

The study presented in Chapter 3 shows that aggregated probe data can be used to mitigate the cumulative error problem, but that valuable information is lost in the aggregation process. Disaggregated (probe) data, such as probe trajectory and vehicle re-identification data, provide a more direct link between the cumulative flow curves at detector locations; however, these data do not describe how the cumulative flow changes along a probe trajectory. Therefore, a commonly used assumption is that the cumulative flow value is constant along the trajectory, i.e., there is no overtaking. Although this is a valid assumption in single-lane traffic, it is likely to be violated in multi-lane traffic.

In this chapter, we investigate the option to use disaggregated detector data to estimate the change in cumulative flow over probe trajectories. These detector data describe the change in cumulative flow at the detector locations, but can also be used to estimate the macroscopic variables flow, speed and density at these locations. Based on these macroscopic variables and the probe data, the probe specific relative flow (i.e., the rate at which the cumulative flow changes along the probe trajectory) can be estimated at the detector locations. In turn, these estimates can be used to estimate the probe specific relative flow and thereby the change in cumulative flow along the full probe trajectory between detector locations. If these estimates are more accurate than the ‘no overtaking assumption, it is beneficial to use the proposed approach to mitigate the cumulative error problem and describe the cumulative flow along probe trajectories.

4.1 Introduction

Traffic state estimation (TSE) is important in dynamic traffic management (DTM) applications (Seo, Bayen et al., 2017). TSE aims to infer the traffic state (which may be described using different variables) from incomplete and inaccurate information, e.g., partially observed and noisy traffic sensing data and traffic flow models. The traffic state estimates can be used as input for different types of DTM applications, e.g., local ramp-metering (Papageorgiou, Hadj-Salem & Blosseville, 1991) or network-wide traffic management (Hoogendoorn et al., 2013).

Throughout this study, road segments without discontinuities (which are denoted as links) are considered. In these links, conservation-of-vehicles condition holds. Vehicles enter (flow in) the link at the upstream boundary and leave (flow out) the link at the downstream boundary.

Traffic can be described using three dimensions: space x, time t and cumulative flow N (Makigami et al., 1971). The cumulative flow $N(x,t)$ denotes the number of vehicles that have passed position x at time t. Here, it is important that the same set of vehicles is used for the counts at all locations. As the number of vehicles are counted, the cumulative flow is discrete. However, by smoothing the discrete function, we can obtain a continuously differentiable cumulative flow function. Taking the derivatives to time and space respectively yield flow and (-) density, i.e., $\partial_t N(x,t) = q(x,t)$ and $\partial_x N(x,t) = -k(x,t)$, which in turn can be used to obtain the mean speed $u(x,t) = q(x,t)/k(x,t)$.

Multiple methodologies have been proposed to estimate the traffic state in or via the cumulative flow plane. For instance, Newell's (three-detector) method (G. F. Newell, 1993a), (Laval et al., 2012), Claudel's method (Claudel & Bayen, 2010b), (Claudel & Bayen, 2010a), Sun's method (Sun et al., 2017) and Van Erp's principles (Van Erp et al., 2019), all apply variational theory (Daganzo, 2005a), (Daganzo, 2005b) to estimate the cumulative flow over space and time, i.e., $N(x,t)$. Other studies estimate the macroscopic traffic states flow and density (e.g., (Van Erp, Knoop & Hoogendoorn, 2018b)), vehicle accumulations (e.g., (Van Lint & Hoogendoorn, 2015), (Van Erp, Knoop & Hoogendoorn, 2018a)) or (mean) traveltimes (e.g., (Bhaskar et al., 2010)) based on observations related to the cumulative flow curves over paths in space-time.

To obtain the cumulative flow curves, we rely on traffic sensing data. Stationary observers such as a loop-detectors can be used to observe the flow with respect to a fixed position and can thus be used to construct the cumulative flow curves at these locations. However, these curves should be initialized, i.e., we want to know the number of vehicles that are between the detectors at the initial time, and we need to address the cumulative error problem. Over time, error in the flow observations accumulate causing a drift in the cumulative flow estimation error. If the problem is not mitigated, the traffic state estimates that are based on the cumulative curves will become highly inaccurate. Therefore, multiple studies propose to use other data to periodically recover the cumulative error, e.g., (Bhaskar et al., 2010), (Van Lint & Hoogendoorn, 2015), (Sun et al., 2017) and (Van Erp, Knoop & Hoogendoorn, 2018b).

Bhaskar et al. (2010), Van Lint & Hoogendoorn (2015) and Sun et al. (2017) use probe trajectory or vehicle re-identification data to mitigate the cumulative error problem. In these studies, it is assumed that there is no overtaking, i.e., the cumulative flow value is constant over the probe trajectory ($\Delta N = 0$). This is a valid assumption for single-lane links, but it is

likely to be violated in multi-lane links, which is also mentioned by Sun et al. (2017) as a limitation. One may think of different mitigation techniques to address this limitation: (1) deal with the uncertainty in ΔN in error correction, (2) observe ΔN over probe trajectories (i.e., collect relative flow data with moving observers (Van Erp, Knoop & Hoogendoorn, 2018b)) or (3) estimate ΔN over probe trajectories based on alternative data. The first technique will always be valuable, because observations or estimates of ΔN along probe trajectories are still subjected to uncertainties (potential errors). Out of the latter two techniques, observing ΔN is expected to be most accurate; however, this would require that probe vehicles are equipped with sensors that observe the vehicles that are overtaken or overtake the probe vehicle. The trend of vehicle automation is expected to make it possible to collect this so-called relative flow data (Van Erp, Knoop & Hoogendoorn, 2018b), (Van Erp et al., 2019) and (Takenouchi et al., 2019), but it will take time before these data can be collected on a wide-scale. Therefore, in this study, we evaluate mitigation technique (3), where we explore the option to use disaggregated detector data (which are currently widely available in The Netherlands and in other countries) to estimate ΔN over probe trajectories.

To study the possibility to estimate the change in cumulative flow (ΔN) over probe trajectory using lane-specific detector data, we make use of simulated and real datasets. Both datasets have their strength and limitation and therefore it is interesting to consider both. Real probe trajectory and disaggregated lane-specific detector data are available for a road stretch in The Netherlands. These data relate to real-life traffic behavior and are subjected to real-life observation errors. However, we do not have a ground truth for the real data. We can still evaluate the ability to estimate the probe relative flow using detector data, but these evaluations are less thorough than are possible with a ground truth. The simulated data, for which the microscopic simulation tool FOSIM (Dijker & Knoppers, 2006) is used, allows us to construct the two data-types and the ground truth. This thus gives us the opportunity to compare the estimated and true changes in cumulative flow. However, in contrast to the real data, traffic behavior may be less realistic and we do not consider observation errors.

The main contribution of this paper is the design and evaluation of a methodology to estimate the change in cumulative flow along probe trajectories based on detector data. This methodology estimates the probe relative flow at the detector locations and uses these relative flows to estimate the relative flow over the full probe trajectory between detector locations. Evaluation using real and simulated data shows that in most cases estimation of relative flow using detector data is an improvement over the assumption that the relative flow is zero. However, changes in traffic conditions (e.g., when a probe encounters a traffic jam) negatively affect the estimation performance.

This article is structured as follows: First, the theoretical foundations that are relevant for this study are explained. Next, a methodology to estimate the change in cumulative flow along probe trajectories between detector locations is presented. The performance methodology is testing using simulated and real data. After explaining how these data are used to assess the estimation performance of the methodology, we present the results. Finally, the conclusions and insights of this study are presented.

4.2 Theoretical foundations

As explained in the introduction, this study aims to estimate the change in cumulative flow along probe trajectories between two detector locations. This describes the number of vehicles that have overtaken the probe vehicle minus the number of vehicles that are overtaken by the probe vehicle. This section provides the theoretical foundations that are relevant in this study. First, we explain that the change in cumulative flow along a probe trajectory depends on the individual probe speed and macroscopic traffic flow variables. Second, we explain how the change in cumulative flow along two probe trajectories relates to detector passing observations. The former is important to design the methodology that is proposed in the next section, while the latter is important to evaluate that methodology in an empirical case study.

In this section, we assume that the macroscopic traffic state is described using continues functions of the variables flow $q(x,t)$, density $k(x,t)$ and mean speed $u(x,t)$. As explained above, these functions can be obtained by taking derivatives to time and space of the smoothed cumulative flow functions. In reality, the cumulative flow function is discrete, which means that the theoretical foundations provided below are not perfect in reality. However, they do allow us to derive the relations that are valuable in the next steps of this study.

4.2.1 Change in cumulative flow along a probe trajectory

The position of vehicle j at time t is described by $X_j(t)$. Furthermore, the probe vehicle speed can be obtained by taking the derivative to time of the probe vehicle position, i.e., $V_j(t) = \partial_t X_j(t)$. Figure 4.1a shows the trajectory, i.e., position over time, of probe vehicle j.

The cumulative flow N along the probe trajectory is given by $N(X_j(t),t)$, which we will denote as $N_j(t)$. In multi-lane traffic, where overtaking is possible, $N_j(t)$ can change over time. The rate at which the cumulative flow along a probe trajectory changes over time is denoted as the probe relative flow and described by $q_j^{\text{rel}}(t) = \partial_t N_j(t)$. This probe-specific relative flow can be described as a function of the probe speed $V_j(t)$ and the macroscopic variables, i.e.:

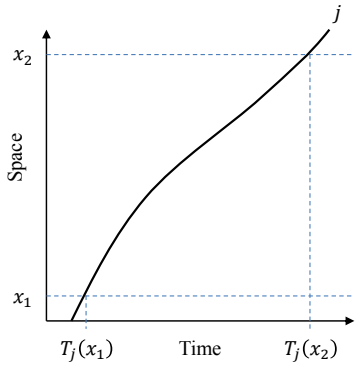
$$q_j^{\text{rel}}(t) = k(X_j(t),t)\,[u(X_j(t),t) - V_j(t)] \tag{4.1}$$

A positive value indicates that N_j increases over time, which means that probe j is overtaken by more vehicles than it overtakes, i.e., it is a relatively slow vehicle.

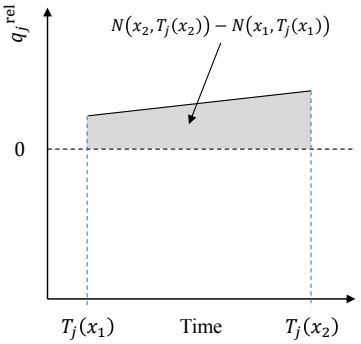
The change in cumulative flow ΔN along the probe trajectory between two locations (x_1 and x_2) can be described as a function of the relative flow. For this purpose, we take the integral of the probe relative flow between the period that is considered:

$$N(x_2,T_j(x_2)) - N(x_1,T_j(x_1)) = \int_{T_j(x_1)}^{T_j(x_2)} q_j^{\text{rel}}(t) \tag{4.2}$$

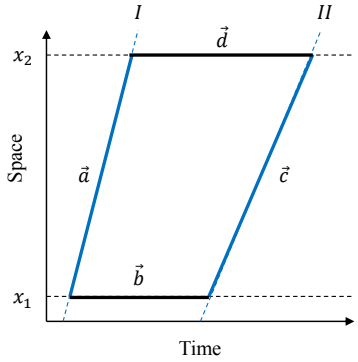
where $T_j(x_1)$ and $T_j(x_2)$ respecitively denote the times at which probe j passes locations x_1 and x_2.



(a) Probe trajectory $X_j(t)$



(b) Probe relative flow and change in cumulative flow



(c) Area enclosed by two observation paths of stationary observers and two probe trajectories

Figure 4.1: Visualizations related to the theoretical foundations.

Figure 4.1b visually shows how equation (4.2) can be interpret. The probe relative flow $q_j^{\text{rel}}(t)$ is indicated by the solid black line. The area under this line describes the change in cumulative flow over the probe trajectory between x_1 and x_2, i.e., $N(x_2, T_j(x_2)) - N(x_1, T_j(x_1))$.

Equations (4.1) and (4.2) state that the change in cumulative flow along a probe trajectory depends on the probe speed and the macroscopic variables along this trajectory. In this study, probe trajectory data are considered that do not contain observations of the relative flow. However, the relations provided in this section show that other data related to the macroscopic variables (i.e., detector data) can be used to estimate the probe relative flow. Therefore, in the next section, a methodology is proposed to estimate the change cumulative flow along probe trajectories using detector data.

4.2.2 Differences in the change in cumulative flow between probe trajectories

Equation (4.2) shows that it is possible to evaluate the accuracy of $q_j^{\text{rel}}(t)$-estimates if the cumulative flow curves at x_1 and x_2 are known, i.e., if we know $N(x_1, T_j(x_1))$ and $N(x_1, T_j(x_1))$. However, as explained in the introduction, real detector data alone does not provide accurate information on these curves as they need to be initialized and we need to mitigate the cumulative error problem. Below, we propose an approach to evaluate the ΔN-estimates along two (consecutive) probe trajectories based on detector data. This approach does not require initialization of cumulative flow curves and is less sensitive to the cumulative error problem as it only uses the detector passings in a short time-period.

Let us consider a combination of two (consecutive) detectors and two (consecutive) probe vehicles, see Figure 4.1c. The black lines in this figure show the observation paths of the detectors, while the blue lines show two probe trajectories. In case we combine detector and probe trajectory data, the change in cumulative flow ΔN is observed over the black lines ($\overrightarrow{b}$ and $\overrightarrow{d}$), but not over the blue lines ($\overrightarrow{a}$ and $\overrightarrow{c}$). Conservation-of-vehicles determines that the net flow over the enclosed area boundary needs to be equal to zero:

$$\Delta N_{\overrightarrow{b}} + \Delta N_{\overrightarrow{c}} = \Delta N_{\overrightarrow{a}} + \Delta N_{\overrightarrow{d}} \tag{4.3}$$

where the elements in the left part ($\Delta N_{\overrightarrow{b}}$ and $\Delta N_{\overrightarrow{c}}$) and the right part ($\Delta N_{\overrightarrow{a}}$ and $\Delta N_{\overrightarrow{d}}$) of the equation are respectively positive when there is an inflow into or outflow out of the area. The individual changes in cumulative flow are given by:

$$\Delta N_{\overrightarrow{a}} = N(x_2, T_I(x_2)) - N(x_1, T_I(x_1)) \tag{4.4}$$
$$\Delta N_{\overrightarrow{b}} = N(x_1, T_{II}(x_1)) - N(x_1, T_I(x_1)) \tag{4.5}$$
$$\Delta N_{\overrightarrow{c}} = N(x_2, T_{II}(x_2)) - N(x_1, T_{II}(x_1)) \tag{4.6}$$
$$\Delta N_{\overrightarrow{d}} = N(x_2, T_{II}(x_2)) - N(x_2, T_I(x_2)) \tag{4.7}$$

In equation (4.3), $\Delta N_{\overrightarrow{a}}$ and $\Delta N_{\overrightarrow{c}}$ describe the changes in cumulative flows over the probe trajectories, which in turn depend on the probe relative flows, see equations (4.1) and (4.2):

$$\Delta N_{\overrightarrow{a}} = \int_{T_I(x_1)}^{T_I(x_2)} q_I^{\text{rel}}(t) \tag{4.8}$$
$$\Delta N_{\overrightarrow{c}} = \int_{T_{II}(x_1)}^{T_{II}(x_2)} q_{II}^{\text{rel}}(t) \tag{4.9}$$

The other parts of equation (4.3), i.e., $\Delta N_{\overrightarrow{b}}$ and $\Delta N_{\overrightarrow{d}}$, describe the true changes in cumulative flow over detectors. In case of observation errors ε, the observed changes in cumulative flow $\Delta\hat{N}$ can differ from the true changes in cumulative flow ΔN, i.e.,:

$$\Delta N_{\overrightarrow{b}} = \Delta\hat{N}_{\overrightarrow{b}} + \varepsilon_{\overrightarrow{b}} \tag{4.10}$$
$$\Delta N_{\overrightarrow{d}} = \Delta\hat{N}_{\overrightarrow{d}} + \varepsilon_{\overrightarrow{d}} \tag{4.11}$$

Therefore, we can describe the sum of the (net) detector observation errors ($\varepsilon_{\overrightarrow{b}}$ and $\varepsilon_{\overrightarrow{d}}$) and (net) change in cumulative flow over probe trajectories ($\Delta N_{\overrightarrow{a}}$ and $\Delta N_{\overrightarrow{c}}$) based on the observed changes in cumulative flow ($\Delta\hat{N}_{\overrightarrow{b}}$ and $\Delta\hat{N}_{\overrightarrow{d}}$):

$$\Delta\hat{N}_{\overrightarrow{d}} - \Delta\hat{N}_{\overrightarrow{b}} = \Delta N_{\overrightarrow{c}} - \Delta N_{\overrightarrow{a}} + \varepsilon_{\overrightarrow{b}} - \varepsilon_{\overrightarrow{d}} \tag{4.12}$$

This relation is used in the empirical case study to evaluate the accuracy of estimates related to the change in cumulative flow along probe trajectories (i.e., $\Delta N_{\overrightarrow{a}}$ and $\Delta N_{\overrightarrow{c}}$). The left side of this equation is observed using detectors, while $\Delta N_{\overrightarrow{a}}$ and $\Delta N_{\overrightarrow{c}}$ are estimated. By comparing $\Delta\hat{N}_{\overrightarrow{d}} - \Delta\hat{N}_{\overrightarrow{b}}$ and the estimates of $\Delta N_{\overrightarrow{c}} - \Delta N_{\overrightarrow{a}}$, we gain insight in the combination of observation and estimation errors.

4.3 Methodology to estimate the change in cumulative flow between detectors

To estimate the change in cumulative flow over probe trajectories between detectors, we will rely on disaggregated detector data. Disaggregated lane-specific detector data is collected using double loop-detectors. These data describe each individual passing n of detector d in lane l. This passing is described by the passing time $T_l^d(n)$ and passing speed $V_l^d(n)$.

The disaggregated data can be used to calculate the macroscopic traffic states. Within a defined period p with duration Δt, the flow and mean speed are respectively calculated by dividing the number of passing by the period duration and by taking the harmonic mean speed of the individual speeds related to the relevant passings. This yields $q(x_d, p)$ and $u(x_d, p)$ where x_d denotes the location of detector d. In line with Edie's generalized definitions of traffic flow, the harmonic mean speed is taken instead of the arithmetic mean speed. Based on the flow and mean speed, we calculate the density, i.e., $k(x_d, p) = q(x_d, p)/u(x_d, p)$. Note that it is possible to use different approaches to estimate mean speed, flow and density. In this study, we do not evaluate the effect of using different approaches for this purpose and leave this for potential future research.

Equation (4.1) shows how the probe relative flow relates to the macroscopic traffic flow variables and the individual probe speed. This equation is applied to estimate the probe relative flow at the times that it passes the detector locations. For this purpose, a time-window of length Δt around the time at which the probe passes detector d is selected, i.e., the detector passings between $t = T_j(x_d) - \Delta t/2$ and $t = T_j(x_d) + \Delta t/2$ are used to estimate the macroscopic variables.

The detector data provide probe relative flow $q_j^{\text{rel}}(t)$ estimates for the times at which the probe passes the upstream and downstream detector locations x_1 and x_2, i.e., $T_j(x_1))$ and $T_j(x_2))$. However, in order to estimate the change in cumulative flow between x_1 and x_2, we need estimates for the full period between $T_j(x_1)$ and $T_j(x_2)$.

Depending on probe vehicle driving behavior and the traffic conditions that are encountered by the probe vehicle, the probe relative flow can change along its trajectory. The probe trajectory data contains information on the probe speed $V_j(t)$ along the full probe trajectory. Changes in the probe speed indicate that the relative flow has changed; however, a decrease in probe speed does not mean that the relative flow has to increase. A relatively fast probe

may reduce its speed because it encounters congestion, but may (still) be a relatively fast vehicle. Furthermore, in congested conditions the density is higher than in free-flow conditions. This means that, following equation (4.1), the probe relative flow may even decrease when the probe speed decreases, due to decreasing mean speed and increasing density.

As the probe speed does not provide sufficient information to estimate $q_j^{\text{rel}}(t)$, we solely rely on the estimates of $q_j^{\text{rel}}(T_j(x_1))$ and $q_j^{\text{rel}}(T_j(x_2))$ to estimate the probe relative flow along the full trajectory. For this purpose, we consider a simple scheme in which the probe-specific relative flow linearly changes between detector locations:

$$q_j^{\text{rel}}(t) = \phi(t) q_j^{\text{rel}}(T_j(x_1)) + (1-\phi(t)) q_j^{\text{rel}}(T_j(x_2)) \tag{4.13}$$

where x_1 and x_2 denote the upstream and downstream detector locations and the weight factor $\phi(t)$ is given by:

$$\phi(t) = \frac{T_j(x_2) - t}{T_j(x_2) - T_j(x_1)} \tag{4.14}$$

If we solely want to estimate the change in cumulative flow along the probe vehicle between the detector locations, i.e., $N(x_2, T_j(x_2)) - N(x_1, T_j(x_1))$, equations (4.13) and (4.14) simplify to:

$$N(x_2, T_j(x_2)) - N(x_1, T_j(x_1)) = \frac{q_j^{\text{rel}}(T_j(x_1)) + q_j^{\text{rel}}(T_j(x_2))}{2} \left[T_j(x_2) - T_j(x_1)\right] \tag{4.15}$$

4.4 Case study

In the case study, both real and simulated traffic data are used. Real data has the advantage of real traffic behavior and real observation errors. However, we miss a ground truth for real data. Therefore, we also use microscopic simulation to construct traffic sensing data with similar characteristics, while having access to a ground truth.

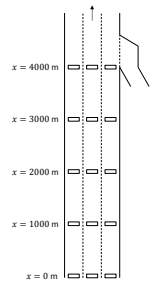
Below, we first explain which traffic sensing data are collected for the two studies and which traffic conditions occur in the study period and road stretch. Next, we explain which experiments will be conducted and which insights these experiments should provide.

4.4.1 Traffic sensing data and traffic conditions

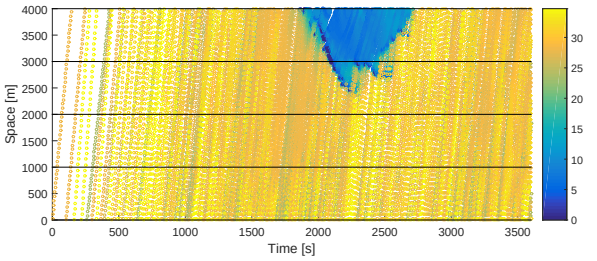
Figure 4.2 shows the road lay-outs and traffic conditions for the two case studies. For both studies, we will discuss which data are available and why certain study periods are selected.

Simulation study

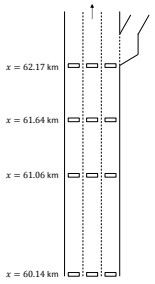
In the simulation study, the microscopic simulation program FOSIM (Dijker & Knoppers, 2006) is used. The model used in this program is validated for Dutch freeway traffic (Henkens et al., 2017). Figure 4.2a shows that we consider a three-lane road segment with



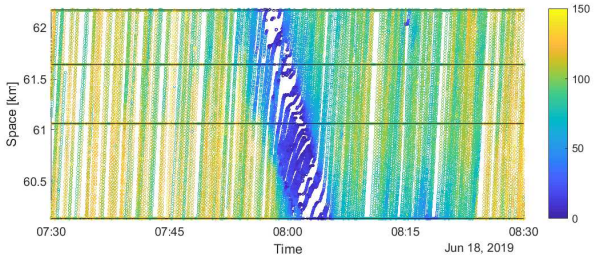
(a) Road layout, simulation study (x=3000 toevoegen)



(b) Individual vehicle speeds (m/s) (every 10th vehicle), simulation study



(c) Road layout, empirical study



(d) Individual probe vehicle and detector passing speeds (km/h), empirical study

Figure 4.2: Road layouts and traffic conditions for the two case studies.

an on-ramp that is located at $x = 4000$ m. In the simulation, the on-ramp traffic causes congestion that spills back on the link upstream of the on-ramp, see Figure 4.2b. Detectors are located at $x = 0, 1000, 2000, 3000, 4000$ m, which provide disaggregated detector data.

Empirical study

Real disaggregated detector and probe trajectory data are available for a test section on the A9 in the Netherlands on June 18th, 2019. These data are respectively made available by the Dutch road authority (RWS) and BeMobile as part of a project that aims to evaluate the value of fusing these two data-types in order to gain more accurate traffic state estimates and potentially reduce the required road-side sensing equipment. BeMobile provides high-frequency (1 s) probe GPS-data, which are map-matched by Modelit. This yields probe trajectory data that describe the position over time (i.e., trajectory) of a subset of the vehicles (i.e., the probe vehicles).

The lay-out (including detectors locations) of the road-segment that is considered in the empirical study is shown in Figure 4.2c. An off-ramp is located directly downstream of the considered segment. Two one-hour peak-periods are selected, i.e., 07:30h-08:30h and 16:00h-17:00h. These periods are selected to study the effect of changing traffic conditions on the ability to correctly estimate the change in cumulative flow over probe trajectories

between detector locations. In the first period, some probes experience congested conditions, see Figure 4.2d. This figure shows that a stop-and-go wave propagates upstream. The cause of this jam lies downstream of the considered segment. In the second period, solely free-flow conditions are observed.

4.4.2 Experimental set-up

Multiple experiments are performed that provide insight in the estimation accuracy. The aim of the experiments is to evaluate the accuracy of estimating the change in cumulative flow over probe trajectories between two detector locations based on disaggregated detector data. In all estimates of the probe relative flow we consider a period of one-minute around the time that the probe passes the detector, i.e., $\Delta t = 60$ s.

In both the empirical and simulation studies, we compare the estimates for different detector spacings. In the simulation study, the considered detector spacings are 1000, 2000, 3000 and 4000 m. In the empirical study, the considered detector spacings are 920, 1500 and 2030 m (which includes all detectors installed on the test section).

The availability of the ground truth in the simulation study allows us to visualize and quantify the estimation errors. As this provides a detailed insight into the estimation performance, we will first perform the simulation study. The resulting insights help to analyze the results of the empirical study.

In the simulation study, two steps are taken. First, we compare the estimated and true changes in cumulative flow (ΔN) for all vehicles. For this purpose, a scatter plot is constructed with the estimated and true ΔN respectively on x-axis and y-axis. Furthermore, the vehicle travel-time is indicated by the color of the dots. In these figures, two ΔN errors can be distinguished: (1) The error if it is assumed that the cumulative flow does not change along the probe trajectory (as assumed by Bhaskar et al. (2010), Van Lint & Hoogendoorn (2015) and Sun et al. (2017)) is indicated by the vertical distance between the dots and the true change in cumulative flow is zero. For this purpose, a horizontal dashed black line is drawn for 'true $\Delta N = 0$'. (2) The error that remains after correction based detector data is indicated by the horizontal/vertical difference between the dots and the diagonal black line. These errors are also represented using the error statistic Root Mean Squared Error (RMSE), where we make a difference between vehicles that experience congestion between detectors and those that do not. Second, to gain a deeper understanding of the underlying factors that influence the estimation accuracy, the estimates related to four individual vehicles are studied in more detail. For these vehicles, time-series plots are constructed of the true and estimated changes in cumulative flow, and the vehicle speed.

Due to absence of the ground truth, it is not possible to directly compare the estimated and true changes in cumulative flow over probe trajectories for the empirical study. Therefore, an alternative comparison is considered in the empirical study; we analyze the difference in ΔN, i.e., the difference in the net number of (-) overtakings, between two consecutive probe vehicles. As explained in Section 'Theoretical foundations', this difference can be observed using detector data. The observed difference (which is still subjected by observation errors) can be compared with the probe specific relative flow estimates. In line with the simulation study, we construct a scatter plot the shows the estimated (x-axis) and observed (y-axis) difference in ΔN between two consecutive probe vehicles. As the observations relate to two probes, the color of the dots is based on the mean travel-time. The ability to

explain the difference in N based on the detector data is indicated by the differences between the dots on the black diagonal line. However, it is important to note that these differences can also be a factor of detector count errors, see equation (4.12).

4.5 Results

This section presents the results from the simulation study and the empirical study. The simulation study is discussed first because it yields insights that are valuable in analysis the empirical study results.

4.5.1 Simulation study

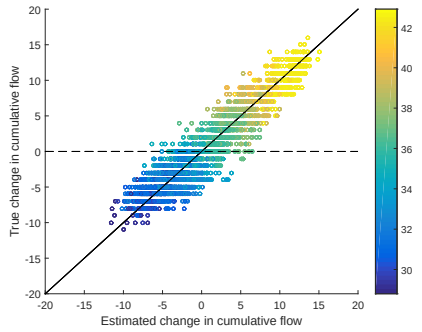
Figure 4.3 shows the true and estimated change in cumulative flow over vehicle trajectories between two detector locations. In these figures the vehicle travel-time between detectors is indicated by the color. Furthermore, Figure 4.4 shows time-series of the true and estimated change in cumulative flow together with the individual speed for four vehicles. These time-series are used to provide more detailed explanations on the features that are observed in Figure 4.3.

Table 4.1: Error statistics for the four detector spacings. A travel-time threshold is used to distinguish vehicles that solely experienced free-flow (FF) conditions and those that experienced congested (CG) conditions.

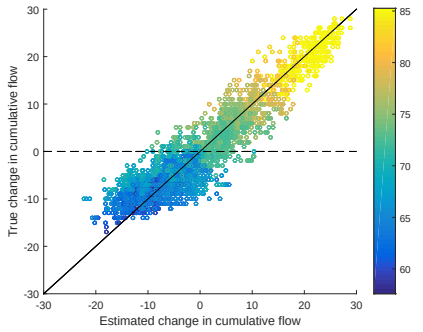
Detector spacing [m]	Travel-time threshold [s]	Correction using detector data		Assume no overtaking	
		RMSE FF [veh]	RMSE CG [veh]	RMSE FF [veh]	RMSE CG [veh]
1000	45	1.65	n.a.	5.11	n.a.
2000	90	3.18	n.a.	9.25	n.a.
3000	135	7.92	29.55	13.29	30.58
4000	180	17.85	28.51	18.16	44.02

In free-flow conditions, estimating the change in cumulative flow over vehicle trajectories is a clear improvement over assuming that there is no overtaking. As the congestion does not spill back upstream of $x = 2500$ m, Figures 4.3a and 4.3b solely include vehicle trajectories in free-flow conditions. These figures show that the solid black line is a better fit than the dashed black line, which indicates that estimation is an improvement over assuming 'no overtaking'. This is also indicated by an improvement in the RMSE, i.e., it improves from 5.11 veh to 1.65 veh for the detector spacing of 1000 m, and from 9.25 veh to 3.18 veh for the detector spacing of 2000 m, see Table 4.1. Figures 4.4a till 4.4d show the vehicle speed and lane together with the estimated and true change in cumulative flow between $x = 0$ m and $x = 4000$ m for a relatively slow and fast vehicle that solely travel in free-flow conditions. These figures show that the change in cumulative flow is estimated relatively accurate along the vehicle trajectory.

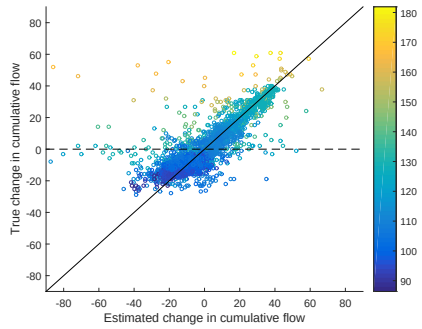
If vehicles experience congested conditions, the estimated changes in cumulative flow are less accurate than for vehicle solely experiencing free-flow conditions, i.e., the black diagonal line is a better fit for Figures 4.3a and 4.3b than for Figures 4.3c and 4.3d; however, also in these cases it is still more accurate to estimate the change in cumulative flow than



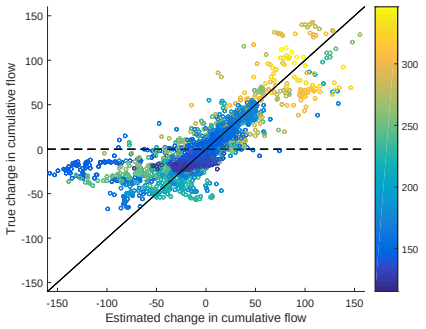
(a) Between detectors located at $x = 0$ m and $x = 1000$ m (solely free-flow conditions)



(b) Between detectors located at $x = 0$ m and $x = 2000$ m (solely free-flow conditions)



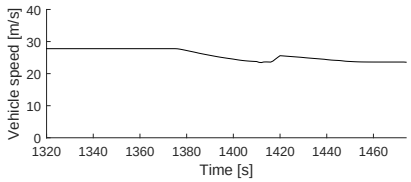
(c) Between detectors located at $x = 0$ m and $x = 3000$ m (free-flow and congested conditions)



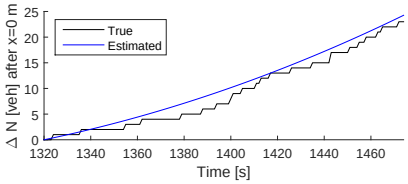
(d) Between detectors located at $x = 0$ m and $x = 4000$ m (free-flow and congested conditions)

Figure 4.3: Simulation study: True and estimated change in cumulative flow over vehicle trajectories between two detector locations. The color of the dots indicates the vehicle travel-time between the considered detector locations.

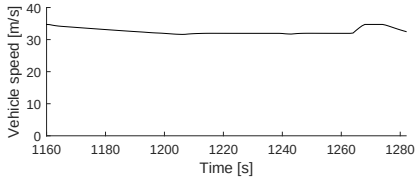
assume 'no overtaking', see Table 4.1. In this table for a detector spacing of 4000 m, the RMSE of vehicles solely experiencing free-flow conditions is relatively large with respect to the other spacings, i.e., it jumps from 7.92 veh to 17.85 veh for detector spacings of 3000 m and 4000 m respectively. In Figure 4.3d there a quite some observations that how a travel-time below the threshold (i.e., smaller than 180 s) and for which the change in cumulative flow is highly underestimated (i.e., estimates around $\Delta N = -125$ veh, while the true ΔN are approximately -25 veh). Let us look at Figures 4.4e and 4.4f to study this in more detail. Figure 4.4f shows that ΔN is highly underestimated for this vehicle. It also shows that the true ΔN reduces sharply in the last second that the vehicle is between the detectors and Figure 4.4e shows that at the same time the vehicle speed decreases. At this last period, the vehicle experiences congested conditions, but is relatively fast, which results in a short highly negative relative flow. In this case the relative flow observed at the downstream detector is not representative for the full trajectory, which causes a large error in the estimated change in cumulative flow. The large underestimation of ΔN in Figure 4.3c



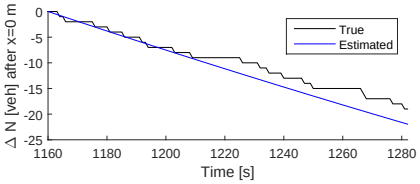
(a) Slow vehicle solely in free-flow



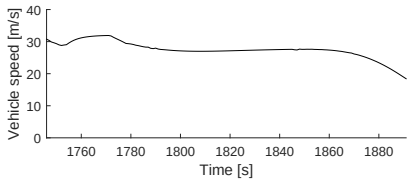
(b) Slow vehicle solely in free-flow



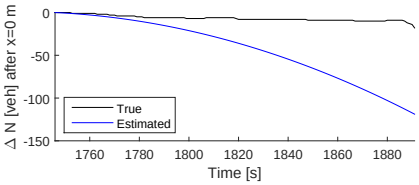
(c) Fast vehicle solely in free-flow



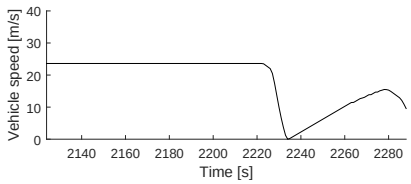
(d) Fast vehicle solely in free-flow



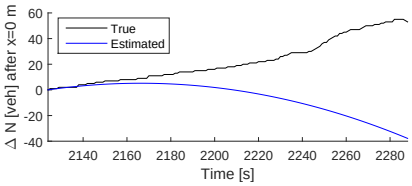
(e) Fast vehicle shortly experiencing congestion at downstream detector



(f) Fast vehicle shortly experiencing congestion at downstream detector



(g) Slow vehicle shortly experiencing negative relative flow



(h) Slow vehicle shortly experiencing negative relative flow

Figure 4.4: Simulation study: True and estimated changes in the cumulative flow over probe trajectories together with the probe speeds.

(large negative estimates, while the true values are positive) can be explained using the same principle, see Figures 4.4g and 4.4h. In these figures, we show a vehicle between $x = 0$ m and $x = 3000$ m for the estimated ΔN is negative due to a large negative vehicle relative flow at the downstream detector.

4.5.2 Empirical study

Figure 4.5 shows the results of the empirical case study. The axis are the same for all subfigures, i.e., Estimated $\Delta N_{\vec{c}} - \Delta N_{\vec{d}}$ (x-axis) and Observed $\Delta N_{\vec{d}} - \Delta N_{\vec{b}}$ (y-axis). The section 'Theoretical foundations' explains the details behind comparing these two features.

In short, both axis relate to the difference in the number of vehicles that overtake two consecutive probes between the detector locations. The y-axis describes this feature based on the detector passing observations, while the x-axis is a result of the estimated change in cumulative flow along the two probe trajectories.

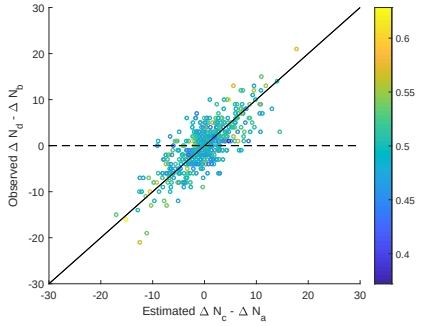
In line with the simulation study, the empirical results indicate that estimation of the change in cumulative flow between detectors is relatively accurate in free-flow conditions, see Figures 4.5a, 4.5c and 4.5e. Also here, the estimates are better than assuming that $\Delta N = 0$, hence the points are better aligned with the diagonal line. To get an insight in the potential mean error of assuming 'no overtaking' during the considered, we may look at the mean estimated change in cumulative flow along individual trajectories between detectors. For the three detector spacings, i.e., 920, 1500 and 2030 m, these mean ΔN are respectively equal to 3.26, 5.45 and 7.49 veh.

In congestion, the computed relative flow is not very accurate in estimating the real number of vehicles passed, see Figures 4.5b, 4.5d and 4.5f. These figures do not show a good relation between the two axis. The largest differences are observed for probes that have a high mean travel-time, which means that these probes are affected by the stop-and-go wave. The figures indicate that the probe relative flows estimated at the detectors are not representative for the full probe trajectory between detectors. The simulation study also showed that the estimation performance decreased when probes experience congestion; however, the estimation performance in congestion for the simulation study seems to be better than for the empirical study. This can partially be explained by the different features that we compare. Figure 4.5 uses the estimates related two probes. If these errors have the same sign (+ or -), the total absolute error increases, which leads to large positive or negative values of $\Delta N_{\vec{c}} - \Delta N_{\vec{d}}$. Another potential reason for the low estimation performance, lies in the lane-drop directly upstream of the first detector (which is used in all estimates. The stop-and-go wave that propagates upstream in the considered period causes a standing queue at this lane-drop. Figure 4.2d show that the vehicles passing the detector at $x = 60.14$ km are affected by this queue (which is indicated by the lower speeds at the upstream detector between 08:00h and 08:45h). This effect can result in probe relative flow estimates that are not representative for the full probe trajectory, and thereby cause errors in the ΔN-estimates.

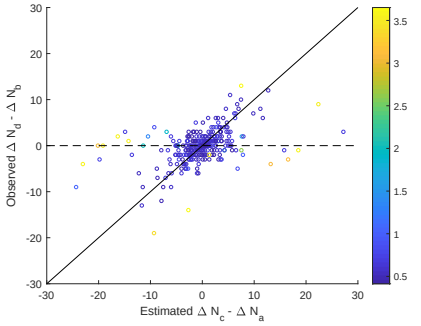
4.6 Conclusions and insights

Probe trajectory or vehicle re-identification data can be used for initialization and error correction of cumulative flow curves constructed using stationary detectors. Studies that use these data for this purpose often assume that there is no overtaking, which would mean that the cumulative flow value is constant along a probe trajectory. However, in multi-lane traffic, this assumption is often violated. This study investigates the option to estimate the change in cumulative flow along probe trajectories based on disaggregated detector data, and in this way improve upon the 'no overtaking' assumption.

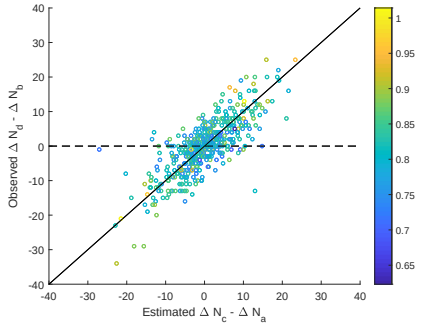
In this study, both simulated as real data are used to investigate the changes in cumulative flow along probe trajectories and the ability to expose this using detector data. By means of a case study we show that the probe relative flow estimated at two detector locations is representative for the full trajectory between these locations in free-flow conditions. Therefore, in these conditions it is a clear improvement to describe the change in cumu-



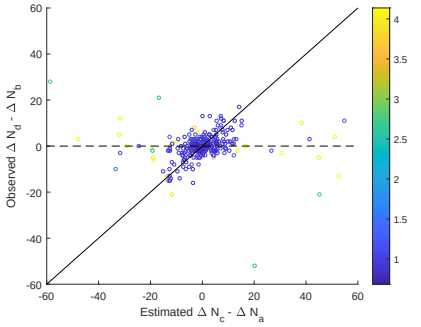
(a) Between detectors at $x = 60.14$ km and $x = 61.06$ km, free-flow period



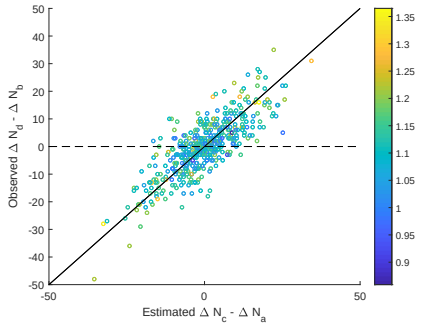
(b) Between detectors at $x = 60.14$ km and $x = 61.06$ km, period with congestion



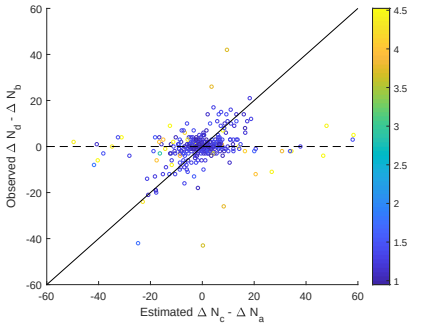
(c) Between detectors at $x = 60.14$ km and $x = 61.64$ km, free-flow period



(d) Between detectors at $x = 60.14$ km and $x = 61.64$ km, period with congestion



(e) Between detectors at $x = 60.14$ km and $x = 62.17$ km, free-flow period



(f) Between detectors at $x = 60.14$ km and $x = 62.17$ km, period with congestion

Figure 4.5: Empirical study: The difference in the number of vehicles that overtake two consecutive probes between the detector locations based on detector passing observations (y-axis) and probe relative flow estimates (x-axis). The color of the dots indicates the mean travel-time between the considered detector locations of the two consecutive probes.

lative flow based on detector data instead of assuming 'no overtaking'. If probe vehicles experience congestion between detectors, the probe relative flows estimated at the detector locations are less representative for the rest of the trajectory. In the simulation study (where the estimation accuracy can be quantified), using disaggregated detector data still yields more accurate estimates than assuming 'no overtaking'. However, in the empirical study, these benefits are not observed. Changing traffic conditions along the probe trajectory (which are related to ability to estimate the change in cumulative flow based on detector data) can be observed using the probe speed. This means that the probe speeds observed between detectors could and should be used to assign an uncertainty to the estimates of the change in cumulative flow along probe trajectories.

More complex schemes may be used that may include information such as the probe speed and the traffic states along the trajectory between detector locations. Probe trajectory data provides information on the probe speeds between detector locations; however, the data do not contain exact information on the traffic state between these locations. Estimating the traffic states between detector locations is the intended application and is the reason for estimating the relative flows along probe trajectories. The circular relation between estimating the probe-specific relative flows and estimating the traffic state indicates that an (iterative) optimization approach to estimate both features is potentially interesting. However, in this study, we focused on the first step and evaluate how accurate the probe-specific relative flows can be estimated without estimating the macroscopic traffic states along the full probe trajectory.

Chapter 5

Macroscopic traffic state estimation using relative flows from stationary and moving observers

This chapter is published as a journal article: van Erp, P.B.C., Knoop, V.L. & Hoogendoorn, S.P. (2018), Macroscopic traffic state estimation using relative flows from stationary and moving observers. *Transportation Research Part B: Methodological*, 114, 281-299.

The previous chapter shows that detector data can be used to estimate the change in cumulative flow over probe trajectories. In that study, the detector data are used to estimate the probe-specific relative flow (i.e., the rate at which the cumulative flow changes along a probe trajectory) at the detector locations. However, as shown in the study, these relative flows are not always representative for the full probe trajectory between detectors. Therefore, it would be better if we are able to observe the changes in cumulative flow along probe trajectories (i.e., observe when and in which direction other road users pass the probe). It is expected that such data (which we denote as relative flow data) can be collected using automated or other equipped and connected vehicles. In this and the next three chapters, we will investigate different potential uses of relative flow data that are collected using stationary and moving observers.

As explained in the introduction, different approaches may be used to estimate the traffic state, e.g., a streaming-data-driven or a model-driven estimation approach. The first approach does not use historical data to derive relations (i.e., learn models) that can be used in real-time estimation, while the second does. In this chapter, we present a streaming-data-driven methodology that can estimate the flow and density for spatial-temporal areas using relative flow data that is collected with stationary and moving observers. Given that this methodology solely relies on real-time data, the estimation accuracy provides a direct insight into the valuable information that is available in the data.

In this chapter, the term 'cumulative vehicle number' is used, which describes the same variable as the term 'cumulative flow'. The second term is used in all other parts of this thesis. However, in our view, it is better to use the term that was used in publication and therefore have not changed the term 'cumulative vehicle number' in 'cumulative flow' for this thesis chapter. Furthermore, in the case study of this chapter, we consider a freeway segment with a lane-drop. Here, the term 'link' is used to refer to the entire freeway segment. However, in other chapters a road segment around the lane-drop would be defined as a node. Defining a lane drop as a node becomes important when traffic flow is modeled using a traffic flow model, which is not the case in this chapter.

5.1 Introduction

This paper addresses macroscopic traffic state estimation. Estimates of the macroscopic traffic flow variables, i.e., flow q, density k and speed u, can be used as input for control decisions within dynamic traffic management applications (Papageorgiou, Hadj-salem & Blosseville, 1991), (Smaragdis et al., 2004).

The estimation procedure introduced in this paper allows us to estimate the macroscopic traffic flow variables within a pre-defined space-time mesh using stationary and moving observers. This procedure consists of two main (independent and consecutive) processes. These are: (1) estimate the cumulative vehicle number N for points along the observed paths in space-time (we will call these point-observations) using traffic sensing data from stationary and moving observers and (2) estimate q and k in a pre-defined mesh based on point-observations of N.

This paper has two important contributions. The first and more generic contribution is the full estimation procedure. We propose to use equipped and/or automated vehicles that observe the relative flow with respect to their trajectory in combination with stationary observers. The traffic sensing data from these observers can be fused on the cumulative vehicle number level (first process) and can be used to estimate the macroscopic traffic conditions (second process). Both processes are explained in this paper. The second and more specific contribution is the methodology designed for the second process. This methodology is called the Point-Observation N (PON) estimation methodology. The two processes require independent methodologies and should in our view not be discussed in detail in a single paper. Therefore, in this paper, we explain the second process in detail, while we solely explain the principles behind the first process.

This paper is organized as follows. Section 5.2 provides the existing methodological basis for our work and positions this work within the research field. Next, we explain the two processes in Sections 5.3 and 5.4. The PON estimation methodology (explained in Section 5.4) is evaluated using a simulation study in Sections 5.5 and 5.6. We conclude with the conclusions and discussion in Section 5.7.

5.2 Background on macroscopic traffic state estimation

In this section, we discuss the topic of macroscopic traffic state estimation and explain how the proposed methodology differs from existing work. First, Edie's generalized definitions of the macroscopic traffic flow variables and the three-dimensional representation of traffic flow are provided. Second, the categorization discussed by Seo, Bayen et al. (2017) is used to categorize the proposed estimation approach. Third, we elaborate on different types of traffic sensing data used for macroscopic traffic state estimation and which data are used in the proposed estimation procedure. And finally, we discuss our estimation output and how this relates to existing work.

The generalized definitions of flow q, density k and speed u, for an area $\mathcal{D}$ in space-time are provided by Edie (1965):

$$q_{\mathcal{D}} = \frac{\sum_i d_i}{A_{\mathcal{D}}} \tag{5.1}$$

$$k_{\mathcal{D}} = \frac{\sum_i r_i}{A_{\mathcal{D}}} \tag{5.2}$$

$$u_{\mathcal{D}} = \frac{q_{\mathcal{D}}}{k_{\mathcal{D}}} \tag{5.3}$$

where d_i and r_i respectively denote the distance traveled and time spent by vehicle i within the area $\mathcal{D}$ and $A_{\mathcal{D}}$ denotes the surface of $\mathcal{D}$. $\sum_i d_i$ and $\sum_i r_i$ respectively denote the Total Travel Distance (TTD) and Total Time Spent (TTS) in $\mathcal{D}$.

Makigami et al. (1971) proposed the three-dimensional representation of traffic flow. The three dimensions are space, time and the cumulative vehicle number, where $N(x,t)$ denotes the cumulative vehicle number at location x and time instant t. As vehicles are discrete, N can be represented as a discrete variable. Here, $N(x,t)$ increases instantly by one vehicle at the time instant t when a vehicle passes location x.

We want to describe traffic flow on a macroscopic level. For this purpose, the discrete N can be smoothed (Makigami et al., 1971). For the smoothed and continuously differentiable N, the macroscopic traffic flow variables can be described based on the three dimensions. The macroscopic variables for a point in space-time, i.e., (x,t), are given by the time and space derivatives of $N(x,t)$:

$$q(x,t) = \frac{\partial N(x,t)}{\partial t} \tag{5.4}$$

$$k(x,t) = -\frac{\partial N(x,t)}{\partial x} \tag{5.5}$$

$$u(x,t) = \frac{q(x,t)}{k(x,t)} \tag{5.6}$$

Seo, Bayen et al. (2017) categorizes the estimation approach into three categories (i.e., model-driven, data-driven and streaming-data-driven) based on information input. Following this categorization, the methodology presented in this study can be categorized as a streaming-data-driven traffic state estimation methodology. Examples of other streaming-data-driven methodologies are (Wardrop & Charlesworth, 1954), (Seo & Kusakabe, 2015), (Florin & Olariu, 2017). A streaming-data-driven methodology does not depend on information in the form of a traffic flow model, fundamental diagram or historical data, but solely relies on real-time data and 'weak' assumptions such as the conservation-of-vehicles. Therefore, 'it is robust against uncertain phenomena and unpredictable incidents' (Seo, Bayen et al., 2017). At the same time, Seo, Bayen et al. (2017) denotes two limitations of streaming-data-driven methodologies: (1) additional information (e.g., a traffic flow model) is needed to predict the (future) traffic state and (2) massive streaming data may be needed to obtain accurate estimates. However, the combination of sensing data and estimation methodology presented in this study can accurately estimate the traffic state without needing massive streaming data. Although we cannot solve the former limitation, the methodology results in estimates that can be used as inputs for a traffic flow model, e.g., the Cell Transmission Model (CTM) (Daganzo, 1994), (Daganzo, 1995). Furthermore, the methodology may be extended by adding information in the form of a traffic flow model. However, in contrast to methodologies such as Newell's three-detector method (G. F. Newell, 1993a), (G. F. Newell, 1993c), (G. F. Newell, 1993b), Laval et al. (2012) and Claudel's approach (Claudel & Bayen, 2010b), (Claudel & Bayen, 2010a), it does not need this information.

The fundamental input of traffic state estimation is traffic sensing data. These data are collected using sensing equipment and observe specific traffic variables for specific space-time properties. In existing work, different types of sensing data are used for macroscopic traffic state estimation. Treiber & Helbing (2002), Wang & Papageorgiou (2005) use detectors to observe macroscopic variables, i.e., mean speed and flow. The speed can be aggregated over time (as is done by Dutch detectors), thereby yielding a time-mean speed, which differs from the desired space-mean speed. As a result, estimating the density based on the aggregated detector variables is biased (Knoop et al., 2009). Alternatively, vehicle-based sensing equipment can be used to observe microscopic traffic variables, i.e., individual vehicle speed (Nanthawichit et al., 2003), (Herrera & Bayen, 2010), space-headway (Seo & Kusakabe, 2015). These data have a potential bias-problem (Seo, Bayen et al., 2017) as equipped vehicles can have a largely consistent different driving behavior than the average road user. For instance, we may collect speed data for a fleet of trucks, which will a lower speed than average. Furthermore, we may collect headway data from vehicles equipped with headway sensors and automated driving technologies, which may act different than manual vehicles.

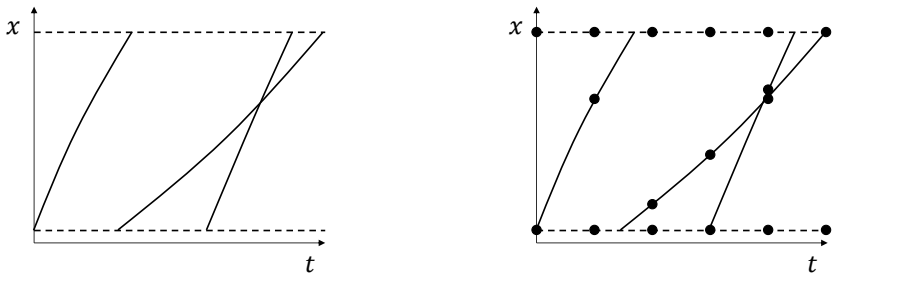
The estimation procedure proposed in this study solely uses observations of ΔN, i.e., flow data from stationary observers and relative flow data from moving observers. Stationary sensing-equipement that observe flow (e.g., detectors or cameras) are widely applied in practice (e.g., PeMS). As proposed by Redmill et al. (2011) and Florin & Olariu (2017), equipped (e.g., automated) vehicles can serve as moving observers that record the flow relative to their position over time. These vehicles observe ΔN along their trajectory. Earlier studies (e.g., (Claudel & Bayen, 2010b)) have used trajectory data of vehicles that cannot observe the relative flow in the same manner. In this case, they often assume that $\Delta N = 0$, i.e., no overtaking, along the vehicle trajectory. Although this assumption is expected to hold for single-lane traffic, it is likely to violated in multi-lane traffic. Therefore, these methodologies could also benefit from using moving observers that observe ΔN.

In terms of estimation output, we can differentiate various types of estimations. These can differ both in the variable types that are estimated as the spatial-temporal characteristics related to the estimates. The estimation methodology proposed in this study estimates the flow and density in a pre-defined space-time mesh. A potential space-time mesh is a discritisation of space in road-segments (cells) and time in periods, e.g., (Nanthawichit et al., 2003), (Wang & Papageorgiou, 2005), (Herrera & Bayen, 2010). In contrast to these methodologies, our methodology is free to work with any other pre-defined mesh. In the remainder of this paper, our focus lies on the two direct output of the methodology, i.e., flow and density. However, as we can obtain the speed from the flow and density, all three macroscopic variables can be estimated. Other methodologies exist that only estimate the speed using data from individual vehicles (e.g., Work et al. (2010), Del Arco et al. (2011)). Newell's three-detector method (G. F. Newell, 1993a), (G. F. Newell, 1993c), (G. F. Newell, 1993b), (Laval et al., 2012) and Claudel's approach (Claudel & Bayen, 2010b), (Claudel & Bayen, 2010a) estimate the cumulative vehicle number for different points in space-time. Although this is related to our estimation output, comparing the outputs would require an additional step for one of the estimation methodologies.

5.3 Point-observations of the cumulative vehicle number

We combine traffic sensing data from stationary and moving observers to estimate N for points in space-time, i.e., $N(x,t)$. In this section, we explain the principles behind this process.

We define observations paths as paths in space-time over which we observe ΔN. Figure 5.1 shows a combination of five observation paths. For stationary observers (dashed lines), the observation path is a horizontal line in space-time, i.e., a fixed location over time is observed. For moving observers (solid lines), the observation path is the trajectory of the connected automated vehicle. The relative flow observed between points along these boundaries is the change in N between these points. Individual observers thus provide the relative N along a single boundary. However, we are interested in having the relative N for all combinations of point-observations in space-time.



(a) Observation boundaries, stationary sensing (dashed lines) and moving observations (solid lines)

(b) Point-observations along the observation boundaries

Figure 5.1: Visualization of observation boundaries and point-observations along these boundaries.

Combining sensing data from stationary and moving observers allows us to relate the data from different observers and to deal with potential observation errors. As stationary and moving observers move with different speeds through space-time, the observation paths intersect. Observation paths that have not yet been initialized (e.g., a moving observer entering the link) can be initialized based on their first interaction (e.g., with a stationary observer at the upstream link boundary). An intersection of two already-initialized observation paths can be used for error correction. Observation errors can lead to a discrepancy between N at the intersection point on the two observation paths. As both paths should have the same N-value at the intersection point, we can correct for observation errors based on the difference in N. This principle is simple; however, in designing an error correction methodology, we need to define how the difference in N translates in to a correction in ΔN over the observation paths. As explained in the introduction, designing this methodology is not the focus of this paper. However, we believe it is valuable to show that the full estimation procedure is robust to observation errors. For this purpose, we design a simple

methodology and evaluate the estimation performance under different observation errors in Appendix 5.A.

To apply the PON-estimation methodology (which will be explained in the next section), all point-observations of N have to be defined in a single framework. Therefore, it should be possible to travel from an observation path to any other observation path. This holds for the combination of observation paths presented in Figure 5.1. Each moving observer is initialized by intersecting with the stationary observer at the upstream boundary and the moving observers cross all stationary observers. Therefore, we have a ΔN for each combination of two points that lie on an observation path.

5.4 The PON estimation methodology

Three steps are taken to estimate flow q and density k in a pre-defined mesh using point-observations of N. These are (1) subdivide the space-time domain in triangular areas, where the three corners are point-observations of N, i.e., $N(x,t)$, (2) estimate q and k for each triangular area $\mathcal{T}$, i.e., $q_{\mathcal{T}}$ and $k_{\mathcal{T}}$, based on the three point-observations and (3) estimate q and k for each area $\mathcal{D}$, i.e., $q_{\mathcal{D}}$ and $k_{\mathcal{D}}$, in a pre-defined mesh based on $q_{\mathcal{T}}$ and $k_{\mathcal{T}}$. In the former step, we use an existing methodology (Delaunay triangulation). The latter two steps are derived based on the three-dimensional representation of traffic flow (Makigami et al., 1971) and Edie's generalized definitions of the macroscopic traffic flow variables (Edie, 1965) .

This section contains four sub-sections. Sections 5.4.1 and 5.4.4 respectively explain the first and third steps. Sections 5.4.2 and 5.4.3 both relate to the second step (estimating $q_{\mathcal{T}}$ and $k_{\mathcal{T}}$). In Section 5.4.2, we derive the equations used for this step based on the three-dimensional representation of traffic flow Makigami et al. (1971). However, there is a subtle difference between the interpretation of the estimates obtained using the derived equations and estimates for a triangular area in space-time. In Section 5.4.3, we use Edie's generalized definitions (Edie, 1965) to show that we can use the equations to estimate $q_{\mathcal{T}}$ and $k_{\mathcal{T}}$ are quantify the related estimation errors.

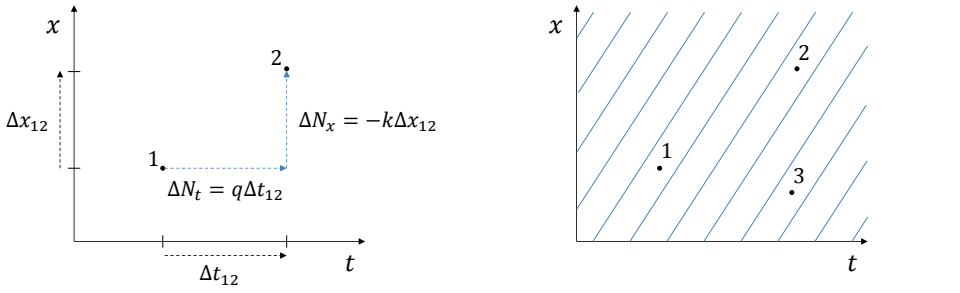
5.4.1 Subdivide space-time into triangular areas

Given a set of point-observations, we want to subdivide the space-time domain into triangular areas, where the three corners are point-observations. We use Delaunay triangulation, which avoids sliver, i.e., narrow, triangular areas, to subdivide space-time based on coordinates of the corner points.

The triangular areas are defined on two different dimensions, i.e., space and time. In order to define three interior angles of triangular areas, we need to define the relation between distances in space and time. This yields the sole parameter of our traffic state estimator, i.e., the space-time ratio υ used in Delaunay triangulation. By defining υ we define a desired dimensions of the triangular areas. For instance, reducing υ yields triangular areas with a smaller space-time ratio, i.e., wider in terms of time and more sliver in terms of space. Therefore, υ influences which information is used to estimate the traffic conditions within the individual areas of the pre-defined mesh.

5.4.2 Three-point traffic state estimation

We propose equations to estimate the macroscopic traffic conditions based on three point-observations of the cumulative vehicle number in space-time, i.e., $N(x,t)$. Let us consider two points in the space-time domain, i.e., (x_1,t_1) and (x_2,t_2), for which the difference in N, i.e., $\Delta N_{12} = N(x_2,t_2) - N(x_1,t_1)$ is known.



(a) Two point observations in space-time and the considered path between these points

(b) Estimated homogeneous and stationary traffic conditions

Figure 5.2: Visualizations related to three-point traffic state estimation.

Figure 5.2a shows two points in space-time and the considered path between these points. Using (5.4) and (5.5) a relation between q and k based on ΔN_{12}, Δx_{12} and Δt_{12} is derived. To derive the relation, homogeneous and stationary traffic conditions are assumed, i.e., $q(x,t) = q$ and $k(x,t) = k$ for any (x,t) in the space-time domain. Therefore, over space and time the changes in N are respectively equal to $-k\Delta x_{12}$ and $q\Delta t_{12}$. This yields:

$$\Delta N_{12} = q\Delta t_{12} - k\Delta x_{12} \tag{5.7}$$

Note that any path between the two points will yield the same equation for ΔN_{12} as we assume homogeneous and stationary traffic conditions. Re-arranging (5.7) allow us to write q as a function of k and vice versa:

$$q = \frac{\Delta N_{12}}{\Delta t_{12}} + \frac{\Delta x_{12}}{\Delta t_{12}} k \tag{5.8}$$

$$k = -\frac{\Delta N_{12}}{\Delta x_{12}} + \frac{\Delta t_{12}}{\Delta x_{12}} q \tag{5.9}$$

An extra (third) point is added to estimate q and k as a function of Δx, Δt and ΔN. In combination with the two initial points, the third point provides ΔN_{13}, Δx_{13}, Δt_{13} and ΔN_{23}, Δx_{23}, Δt_{23}. To incorporate the information of all three points, it suffices to consider two combination of points. In this case, we consider combinations 12 and 23. For both combinations, the relations (5.8) and (5.9) are obtained. Below, we show the steps taken to derive the equation for q, i.e., (5.14):

$$q = \frac{\Delta N_{12}}{\Delta t_{12}} + \frac{\Delta x_{12}}{\Delta t_{12}} \left(-\frac{\Delta N_{23}}{\Delta x_{23}} + \frac{\Delta t_{23}}{\Delta x_{23}} q \right) \tag{5.10}$$

$$q \left(1 - \frac{\Delta x_{12}}{\Delta t_{12}} \frac{\Delta t_{23}}{\Delta x_{23}} \right) = \frac{\Delta N_{12}}{\Delta t_{12}} - \frac{\Delta N_{23}}{\Delta x_{23}} \frac{\Delta x_{12}}{\Delta t_{12}} \tag{5.11}$$

$$q \left(\frac{\Delta t_{12}}{\Delta t_{12}} \frac{\Delta x_{23}}{\Delta x_{23}} - \frac{\Delta x_{12}}{\Delta t_{12}} \frac{\Delta t_{23}}{\Delta x_{23}} \right) = \frac{\Delta N_{12}}{\Delta t_{12}} \frac{\Delta x_{23}}{\Delta x_{23}} - \frac{\Delta N_{23}}{\Delta x_{23}} \frac{\Delta x_{12}}{\Delta t_{12}} \tag{5.12}$$

$$q \left(\Delta t_{12} \Delta x_{23} - \Delta t_{23} \Delta x_{12} \right) = \Delta N_{12} \Delta x_{23} - \Delta N_{23} \Delta x_{12} \tag{5.13}$$

$$q = \frac{1}{\Delta t_{12} \Delta x_{23} - \Delta t_{23} \Delta x_{12}} \left(\Delta N_{12} \Delta x_{23} - \Delta N_{23} \Delta x_{12} \right) \tag{5.14}$$

Similar steps can be taken to derive the equation for k, which results in the following equation:

$$k = \frac{1}{\Delta t_{12} \Delta x_{23} - \Delta t_{23} \Delta x_{12}} \left(\Delta N_{12} \Delta t_{23} - \Delta N_{23} \Delta t_{12} \right) \tag{5.15}$$

The resulting q and k can be interpreted as the homogeneous and stationary traffic conditions that satisfy the cumulative vehicle number for three points in the space-time domain. Figure 5.2b provides a visual interpretation of the estimate homogeneous and stationary conditions. Here, the blue lines denote trajectories satisfying q and k.

If the conditions are indeed homogeneous and stationary, any three observations of N are sufficient to estimate the traffic conditions for the full space-time domain. However, one conditions has to hold. The three points should not lie on a straight line in space-time, i.e., $\Delta x_{12}/\Delta t_{12} \neq \Delta x_{23}/\Delta t_{23}$. If $\Delta x_{12}/\Delta t_{12} = \Delta x_{23}/\Delta t_{23}$ the denominator will be zero. In this case the third point does not provide any additional information. It is furthermore important to note that (5.14) and (5.15) are invariant to the numbering of the three points, i.e., the same results are obtained if the same three points are numbered differently.

5.4.3 Space-time area traffic state estimation based on the mean boundary conditions

We consider Edie's generalized definitions (Edie, 1965) to relate the equations derived above to a triangular area in space-time and explain the related estimation errors. Edie's generalized definitions (5.1) and (5.2) can be rewritten on an aggregated level:

$$q = \frac{I\bar{d}}{A} \tag{5.16}$$

$$k = \frac{I\bar{r}}{A} \tag{5.17}$$

where $\bar{d}$ and $\bar{r}$ are respectively the mean distance traveled and time spent by the I number of vehicles. $\bar{d}$ and $\bar{r}$ can be described as function of the mean entry $(\bar{x}_{\text{in}}, \bar{t}_{\text{in}})$ and exit $(\bar{x}_{\text{out}}, \bar{t}_{\text{out}})$ points, i.e., $\bar{d} = \bar{x}_{\text{out}} - \bar{x}_{\text{in}}$ and $\bar{r} = \bar{t}_{\text{out}} - \bar{t}_{\text{in}}$.

Let us consider areas in the space-time domain with straight boundaries, e.g., a triangle or pentagon. Each boundary b has a net flow ΔN_b. Here outflow and inflow are respectively denoted by positive and negative values of ΔN_b. Furthermore, each b has a mean intersection point $(\bar{x}_b, \bar{t}_b)$, which reflects the mean point in space-time where vehicles intersect with the boundary. As we consider straight boundaries, this point lies on the boundary. In the case that traffic conditions are homogeneous and stationary and N is considered continuous, $(\bar{x}_b, \bar{t}_b)$ lie at the middle point of boundary b.

For the specific case of having an area with a finite number of straight boundaries, (5.16) and (5.17) can rewritten as a function of the boundary conditions:

$$q = \frac{\sum_b \Delta N_b \bar{x}_b}{A} \tag{5.18}$$

$$k = \frac{\sum_b \Delta N_b \bar{t}_b}{A} \tag{5.19}$$

Let us consider a triangular area for which all three ΔN_b are known. This resembles the case presented in Section 5.4.2, in which we know ΔN for three combinations of two points in space-time. Drawing a straight line (boundary) between the combinations of points in space-time results in a triangle.

Theorem 5.1 *For triangular areas in which the mean intersection of each boundary coincides with the middle point of the boundary, (5.18) and (5.19) are equal to respectively (5.14) and (5.15).*

Proof : We prove that (5.18) and (5.19) are equal to respectively (5.14) and (5.15) in the case a triangular area is considered and the mean intersection of each boundary coincides with the middle point of the boundary, i.e.,:

$$\frac{\sum_b \Delta N_b \bar{x}_b}{A} = \frac{1}{\Delta t_{12}\Delta x_{23} - \Delta t_{23}\Delta x_{12}} \left(\Delta N_{12}\Delta x_{23} - \Delta N_{23}\Delta x_{12}\right) \tag{5.20}$$

$$\frac{\sum_b \Delta N_b \bar{t}_b}{A} = \frac{1}{\Delta t_{12}\Delta x_{23} - \Delta t_{23}\Delta x_{12}} \left(\Delta N_{12}\Delta t_{23} - \Delta N_{23}\Delta t_{12}\right) \tag{5.21}$$

For triangular areas we have three corner points, i.e., 1, 2 and 3, and three boundaries, i.e., 12, 23 and 31. For this situation, the left part of the equation can be rewritten as

$$\frac{\sum_b \Delta N_b \bar{x}_b}{A} = \frac{1}{A} \left(\Delta N_{12}\bar{x}_{12} + \Delta N_{23}\bar{x}_{23} + \Delta N_{31}\bar{x}_{31}\right) \tag{5.22}$$

where the conservations of vehicles condition depicts that $\Delta N_{12} + \Delta N_{23} + \Delta N_{31} = 0$, thus $\Delta N_{31} = -\left(\Delta N_{12} + \Delta N_{23}\right)$. Furthermore, for case in which the mean intersection of each boundary coincides with the middle point of the boundary, $\bar{x}_b$ can be described as a function of the corner points, e.g., $\bar{x}_{12} = \left(x_1 + x_2\right)/2$. This allows us to rewrite (5.22):

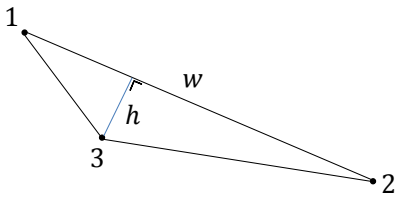
$$\frac{\sum_b \Delta N_b \bar{x}_b}{A} = \frac{1}{A} \left(\Delta N_{12}\frac{x_1 + x_2}{2} + \Delta N_{23}\frac{x_2 + x_3}{2} - \left(\Delta N_{12} + \Delta N_{23}\right)\frac{x_1 + x_3}{2}\right) \tag{5.23}$$

$$= \frac{1}{A} \left(\Delta N_{12}\frac{x_2 - x_3}{2} + \Delta N_{23}\frac{x_2 - x_1}{2}\right) \tag{5.24}$$

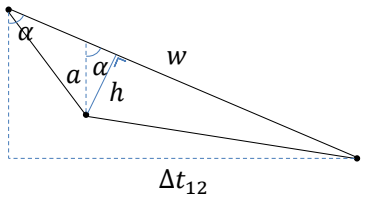
where $x_2 - x_3 = -\Delta x_{23}$ and $x_2 - x_1 = \Delta x_{12}$, thus:

$$\frac{\sum_b \Delta N_b \bar{x}_b}{A} = -\frac{1}{2A}\left(\Delta N_{12}\Delta x_{23} - \Delta N_{23}\Delta x_{12}\right) \tag{5.25}$$

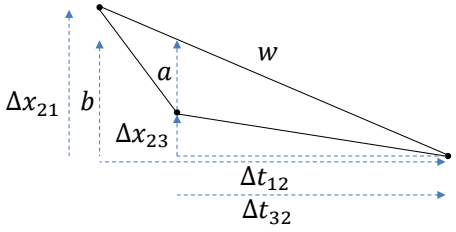
Next, we consider the surface of the triangular area, i.e., A. To calculate the surface of a triangular area we can use $A = \frac{1}{2}wh$, where w is the width and h is the height. Figure 5.3a provides the triangular area and definitions of w and h which we consider to find an equation for A. Note that we want to describe A as a function of the boundaries 12 and 23 as these boundaries are also considered for the right sides of equations 5.20 and 5.21. Furthermore, note that there are multiple correct approaches to find the desired function for A.



(a) Height h and width w



(b) Relevant dimensions and angles to describe h



(c) Relevant dimensions to find a

Figure 5.3: Important dimensions to find an equation for A.

To find h we first consider Figure 5.3b. In this figure, we observe two triangles which have a common angle α. This allows us to relate the ratio of different known and unknown triangle sides:

$$\sin(\alpha) = \frac{h}{a} = \frac{\Delta t_{12}}{w} \tag{5.26}$$

Thus h is given by:

$$h = \frac{a\Delta t_{12}}{w} \tag{5.27}$$

This allows us to simplify the equation for A:

$$A = \frac{1}{2} a \Delta t_{12} \tag{5.28}$$

As Δt_{12} is known we only need to find a. Figure 5.3c shows the relevant dimensions used to find a. This figure provides two basic relations, which we use to find a, namely:

$$a = b - \Delta x_{23} \tag{5.29}$$

$$\frac{b}{\Delta t_{32}} = \frac{\Delta x_{21}}{\Delta t_{12}} \tag{5.30}$$

where $\Delta t_{32} = -\Delta t_{23}$ and $\Delta x_{21} = -\Delta x_{12}$, thus:

$$b = \frac{\Delta t_{23} \Delta x_{12}}{\Delta t_{12}} \tag{5.31}$$

$$a = \frac{\Delta t_{23} \Delta x_{12}}{\Delta t_{12}} - \Delta x_{23} \tag{5.32}$$

This allows us to find A given (5.28):

$$A = \frac{1}{2} \left(\frac{\Delta t_{23} \Delta x_{12}}{\Delta t_{12}} - \Delta x_{23} \right) \Delta t_{12} \tag{5.33}$$

$$= \frac{1}{2} \left(\Delta t_{23} \Delta x_{12} - \Delta t_{12} \Delta x_{23} \right) \tag{5.34}$$

Substituting this equation into (5.25) yields:

$$\frac{\sum_b \Delta N_b \bar{x}_b}{A} = -\frac{1}{2A} \left(\Delta N_{12} \Delta x_{23} - \Delta N_{23} \Delta x_{12} \right) \tag{5.35}$$

$$= -\frac{1}{2 \cdot \frac{1}{2} \left(\Delta t_{23} \Delta x_{12} - \Delta x_{23} \Delta t_{12} \right)} \left(\Delta N_{12} \Delta x_{23} - \Delta N_{23} \Delta x_{12} \right) \tag{5.36}$$

$$= \frac{1}{\Delta x_{23} \Delta t_{12} - \Delta t_{23} \Delta x_{12}} \left(\Delta N_{12} \Delta x_{23} - \Delta N_{23} \Delta x_{12} \right) \tag{5.37}$$

Similar steps can be taken to prove equality for k, i.e., (5.21).

□

The comparison made in Theorem 5.1 is important for two reasons. Firstly, it shows that (5.14) and (5.15) provide a proxy for the TTS and TTD within the triangular area enclosed by the three points. Secondly, it allows us to describe the estimation error in (5.14) and (5.15) based on the mean level of inhomogeneity and non-stationarity of the traffic conditions over the three boundaries. The mean level of inhomogeneity and non-stationarity of the traffic conditions over b affects $(\bar{x}_b, \bar{t}_b)$. The difference between this point and the middle point is denoted as $\Delta\bar{x}_b$ and $\Delta\bar{t}_b$. As both the mean intersection point and middle point lie on the boundary, the inhomogeneous and non-stationary traffic conditions can be described by a single variable, i.e., the fractional difference between the two points $\Delta\bar{\mu}_b$, where $\Delta\bar{x}_b = \Delta\bar{\mu}_b x_b$ and $\Delta\bar{t}_b = \Delta\bar{\mu}_b t_b$. This allows us to quantify the error in (5.14) and (5.15), i.e., ε_q and ε_k:

$$\varepsilon_q = \frac{\sum_b \Delta N_b \Delta\bar{\mu}_b x_b}{A} \tag{5.38}$$

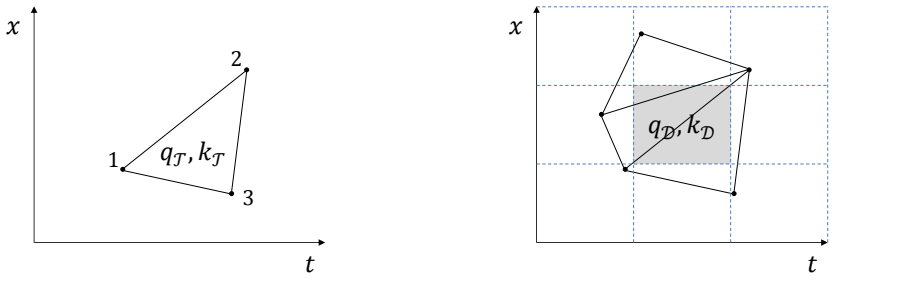
$$\varepsilon_k = \frac{\sum_b \Delta N_b \Delta\bar{\mu}_b t_b}{A} \tag{5.39}$$

These equations provide insight in the estimation error dependent on the area dimensions, i.e., A, x_b and t_b, and traffic conditions, i.e., N_b and $\Delta\bar{\mu}_b$. They show that the estimator estimates the traffic conditions perfectly, i.e., $\varepsilon_q = 0$ and $\varepsilon_k = 0$, in homogeneous and stationary conditions, i.e., $\Delta\bar{\mu}_b = 0$ for all b. Furthermore, the equations provide insight into the estimation errors of two triangular areas which share a boundary, i.e., adjacent triangular areas. Inhomogeneous and non-stationary traffic conditions over the shared boundary result in an estimation error in both triangular areas. There are, however, two differences, i.e., (1) the flow is opposite (inflow vs. outflow) and (2) the error is scaled by the related triangular area surface. Therefore, the error induced by a shared boundary is inversely proportional to the surfaces of the adjacent triangular areas.

5.4.4 Traffic state estimation in a pre-defined mesh

The relations introduced in previous sections allow us to estimate q and k for triangular areas in space-time, see Figure 5.4a. However, depending on the application, we may want to select the dimensions of the estimation area, i.e., define the estimation mesh. As an example, in this research, we consider a mesh which subdivides space-time in rectangular areas, see Figure 5.4b. The estimates for the rectangular areas can be interpreted as the mean traffic conditions for a road segment (cell) during a time period. The selected mesh, i.e., discretizing space and time, allows for model-based prediction with the Cell Transmission Model (CTM) (Daganzo, 1994), (Daganzo, 1995). Therefore, our estimates can be used as input for traffic state estimation using the CTM. Prior proposed estimators (Wang & Papageorgiou, 2005), (Herrera & Bayen, 2010) and control algorithms (Smaragdis et al., 2004) also consider a discrete space and time mesh.

Let us consider the case in which the complete space-time domain is subdivided in triangular areas. For each triangular area $\mathcal{T}$ we know the flow $q_{\mathcal{T}}$ and density $k_{\mathcal{T}}$. To estimate the traffic conditions for a defined area in space-time, the conditions within $\mathcal{T}$ are assumed to be homogeneous and stationary. This means that the TTS and TTD within a subarea of $\mathcal{T}$ is proportional to the relative size of the subarea. Based on Edie (1965), we



(a) Three-point estimation

(b) Estimation in desired space-time mesh, in gray: the triangle subareas which are relevant for the rectangular area

Figure 5.4: Visualisation of three-point traffic state estimation and estimation in a defined mesh.

known that $TTD_\mathcal{T} = q_\mathcal{T} A_\mathcal{T}$ and $TTS_\mathcal{T} = k_\mathcal{T} A_\mathcal{T}$. Combined with the assumption stated above, the TTD and TTS for a subarea $\mathcal{S}$ of the triangular area $\mathcal{T}$ are given by:

$$TTD_\mathcal{S} = TTD_\mathcal{T} \frac{A_\mathcal{S}}{A_\mathcal{T}} = q_\mathcal{T} A_\mathcal{S} \tag{5.40}$$

$$TTS_\mathcal{S} = TTS_\mathcal{T} \frac{A_\mathcal{S}}{A_\mathcal{T}} = k_\mathcal{T} A_\mathcal{S} \tag{5.41}$$

Let us consider a rectangular area $\mathcal{D}$ with a surface $A_\mathcal{D}$. The surface of the relevant subarea $A_\mathcal{S}$ is the surface of triangle $\mathcal{T}$, i.e., $A_\mathcal{T}$, in the desired area $\mathcal{D}$, i.e., $A_\mathcal{T} \in A_\mathcal{D}$. Based on the traffic conditions in all relevant areas, i.e., $q_\mathcal{T}$ and $k_\mathcal{T}$, and their contribution, i.e., $(A_\mathcal{T} \in A_\mathcal{D}) / A_\mathcal{D}$, the traffic conditions in the rectangular area can be estimated:

$$q_\mathcal{D} = \frac{\sum_\mathcal{T} (A_\mathcal{T} \in A_\mathcal{D})\, q_\mathcal{T}}{A_\mathcal{D}} \tag{5.42}$$

$$k_\mathcal{D} = \frac{\sum_\mathcal{T} (A_\mathcal{T} \in A_\mathcal{D})\, k_\mathcal{T}}{A_\mathcal{D}} \tag{5.43}$$

Figure 5.4b provides a visualization of the relevant triangle subareas for a single rectangular area.

The accuracy of estimates obtained with (5.42) and (5.43) depend on two factors. Firstly, if $q_\mathcal{T}$ and $k_\mathcal{T}$ are estimates, they may contain estimation errors. If these estimates are obtained using three-point estimation, the estimation errors of adjacent areas induced by the shared boundary are negatively correlated. In this case the errors are partially corrected by combining adjacent areas. Secondly, we assumed that conditions within $\mathcal{T}$ are homogeneous and stationary. If $\mathcal{T}$ partly falls outside the desired area $\mathcal{D}$, inhomogeneity and non-stationarity within $\mathcal{T}$ can induce estimation error for $\mathcal{D}$.

5.5 Simulation study

As explained before, the estimation procedure consists of two processes: (1) estimating $N(x,t)$ based on measurements from stationary and moving observers and (2) estimating q and k in a pre-defined space-time mesh based on $N(x,t)$. These processes were consecutively discussed in Sections 5.3 and 5.4. In the simulation study we focus on the second process, for which the PON estimation methodology is designed. The simulation study is explained in this section and the results are discussed in Section 5.6.

To investigate and explain the working of the methodology in more detail, it is assumed that the first process is perfect, so we correctly know $N(x,t)$ for the point-observations along the observations paths of the available stationary and moving observers. This requires that the observation paths are correctly initialized and there are no (or fully corrected) observation errors. In practice, there will be count-errors in stationary (and moving) observers. Therefore, error correction is required for methodologies that only require cumulative counts, e.g., (Van Lint et al., 2014). In our study, as explained in Section 5.3, the cumulative drift problem is less of an issue. We will therefore further elaborate on the second process, assuming no errors, in this and the next section. Additionally, Appendix 5.A shows how observations errors might affect the performance.

5.5.1 Microscopic simulation in FOSIM

The simulation study is conducted with the microscopic simulation program FOSIM (Dijker & Knoppers, 2006). FOSIM is calibrated and validated for Dutch freeways (Minderhoud & Kirwan, 2001), (Henkens et al., 2017). However, the simulated traffic conditions may still deviate from real traffic conditions for similar road infrastructure properties and traffic demand. In estimation we want to reconstruct the conditions based on a limited data-set. In our case study, this means that we want to reconstruct the simulated traffic conditions using the proposed methodologies and data. It therefore is essential that the full ground truth, which should be reconstructed, is know, which is the case for simulation. Therefore, we collect two synthesized data-sets; (1) a data-set of all vehicle trajectories to obtain the true traffic conditions, i.e., the ground truth, and (2) a limited data-set of traffic sensing data to reconstruct the true traffic conditions.

We consider a 10 km freeway link with a bottleneck at $x = 7$ km. The number of lanes drops from three to two at this location and goes back to three at $x = 8.5$ km. This road lay-out, as visualized in FOSIM, is shown in Figure 5.5.

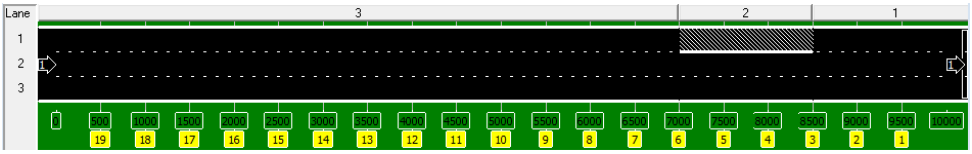


Figure 5.5: Considered road lay-out in the microscopic simulation program FOSIM.

The traffic conditions are described by the macroscopic traffic flow variables in a rectangular mesh, i.e., discrete space and time. Here, the freeway is subdivided in 500 m long road segments and 15 s time-periods. Given the speed limit of 120 km/h this combination of

road segment length and time-period duration satisfies the Courant-Friedrichs-Lewy conditions (Courant et al., 1928). Although this condition is not a prerequisite for the proposed method, it is important for numerical stability in model-based estimation and prediction. As the estimates may be used as initial conditions for model-based traffic state prediction, we decided to evaluate the methodology in this mesh.

Traffic is simulated for two one-hour periods, one with solely free-flow conditions and one with free-flow and congested conditions. In the remainder of this paper these two cases are respectively referred to as the free-flow and congested case. Figures 5.6a and 5.6b show the traffic demand that is used as input into the microscopic simulation program is shown. In both cases traffic is composed of 90 % passenger cars and 10 % trucks. More details about the microscopic models and parameters used in FOSIM can be found in the users manual (Dijker & Knoppers, 2006). The true $q_{\mathcal{D}}$, $k_{\mathcal{D}}$ and $u_{\mathcal{D}}$, are obtained from the trajectory information for all vehicles using (5.1), (5.2) and (5.3) (Edie, 1965). This ground truth is shown in Figure 5.6.

Figures 5.6d, 5.6f and 5.6h, i.e., the congested case, show patterns in the congested area. To understand these patterns, we determine the ground truth for a finer mesh, i.e., smaller road segments and shorter time periods. Figure 5.7 shows the ground truth in terms of density and speed for road segments of 100 m and 5 s time-periods. This figure shows that there are multiple waves which move upstream with an approximate speed of 25 km/h. The speed within these waves is close to 0 km/h, while the speed between the waves reaches 40 km/h. These patterns are thus stop-and-go waves in congestion. As explained above, our objective is to reconstruct the true simulated traffic conditions. Therefore, we will not address the realism of the observed patterns. However, we are interested in the ability of our estimator to reconstruct these patterns.

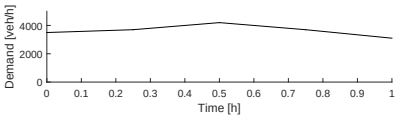
In addition to the free-flow and the congested case, we simulated and evaluated the estimation performance for an incident situation. As the results show large similarities with the congested case, we shortly discuss the estimation performance, but do not provide detailed information in this and the next section.

5.5.2 The reference traffic state estimator

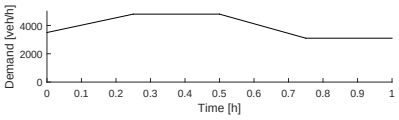
Existing types of traffic sensing data and estimation methodologies can be used to estimate the macroscopic traffic flow variables. To evaluate the added value of our work, we compare the estimation performance with a reference estimator which uses loop-detector data. It is assumed that a loop-detector is installed in the middle of each cell, i.e., the loop-detector spacing is equal to 500 m (thereby following the Dutch state-of-the-art). The choice of considering loop-detectors installed in the middle of each cell is beneficial for its estimation performance compared to other locations, e.g., upstream or downstream boundaries.

The loop-detector data characteristics are based on the loop-detectors installed on the Dutch freeways, i.e., lane-specific one-minute aggregated speed u_l^T and flow q_l. The flow for all lanes q is obtained by summing the lane-specific flows q_l. To approximate the mean speed, the following equations are considered (Knoop & Hoogendoorn, 2012):

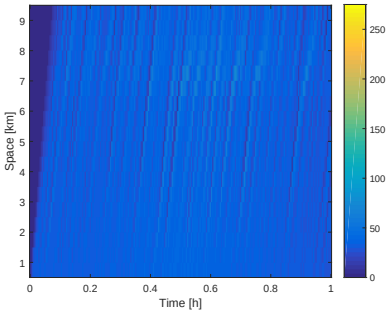
$$q = \sum_{l=1}^{\lambda} q_l \tag{5.44}$$



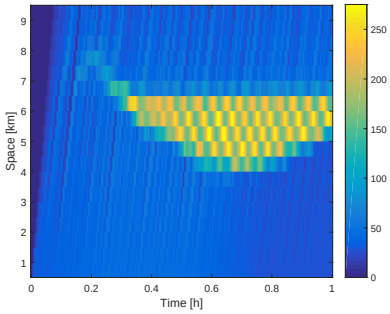
(a) Demand, free-flow case



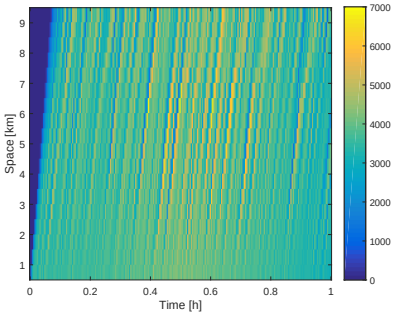
(b) Demand, congested case



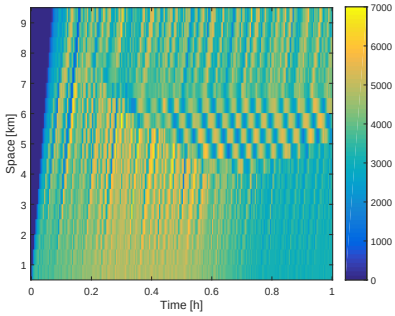
(c) Density, free-flow case



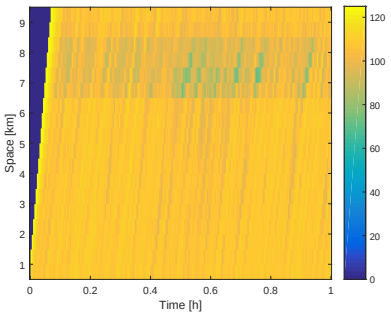
(d) Density, congested case



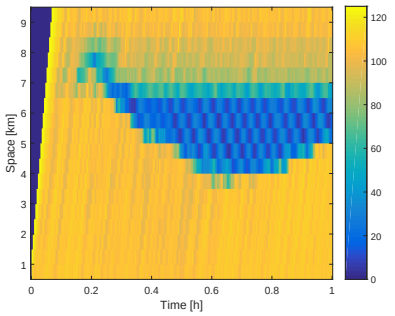
(e) Flow, free-flow case



(f) Flow, congested case

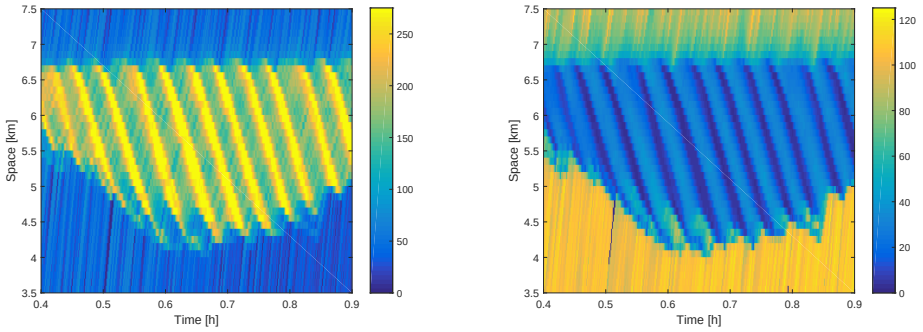


(g) Speed, free-flow case



(h) Speed, congested case

Figure 5.6: Demand and ground truth for free-flow and congested cases.



(a) Density *(b)* Speed

Figure 5.7: Ground truth of the congested case for a finer mesh.

$$u = \frac{\sum_{l=1}^{\lambda} q_l}{\sum_{l=1}^{\lambda} \frac{q_l}{u_l^T}} \tag{5.45}$$

$$k = \frac{q}{u} \tag{5.46}$$

where λ denotes the number of lanes.

The resulting estimates relate to the one-minute loop-detector data period, i.e., four consecutive 15 s periods in the considered estimation mesh. The estimates are assigned to each 15 s period with the one-minute period. As the conditions obtained using Edie (1965) differ within this one-minute period, the one-minute data aggregation period is a cause of the estimation errors.

Loop-detector data characteristics differ across the world. A downside of the Dutch state-of-the-art is that time-mean speeds are observed. These data overestimate the speed (and thus underestimate the density) is congested conditions (Knoop et al., 2009). This bias-problem can be addressed by collecting harmonic mean speeds instead of time-mean speed. As some systems outside the Netherlands can collect harmonic mean speeds, we will also discuss this type of mean speeds in the results section.

5.5.3 The PON traffic state estimator

The data used by the proposed traffic state estimator is retrieved from sensing equipment installed on fixed locations (stationary observers) and vehicle-based sensing equipment (moving observers). In our study, all observers have a fixed sampling period of 15 s. Every 15 s, i.e., $t = 0$ s, 15 s, 30 s, ..., the observers share their position, which yields a set of observation-points along the observation path. For each period between these points ΔN is observed, i.e., the number of vehicles that passed the observer minus the number of vehicles that are passed by the observer within this period.

We fix three observation paths, which are thus used for all evaluated PON estimators. These are: the upstream, i.e., $x = 0$ km, and downstream, i.e., $x = 10$ km, link boundaries, and the first vehicle entering the link. Additional to this fixed information, a fraction of the

vehicles serve as moving observers. In this set-up all observation paths can be connected, thereby yielding a ΔN for each combination of two point-observations. Furthermore, as we observe the upstream and downstream boundaries, all corner points of space-time domain until the current time T are observed, i.e., $(0,0)$, $(10,0)$, $(0,T)$ and $(10,T)$. Observing the corner points is important as this means that the complete space-time domain will be subdivided into triangular areas.

Initial evaluations have shown that the space-time ratio used in Delaunay triangulation can affect the estimation performance. However, these evaluations also indicated that the relations are not straightforward. Therefore, to keep this article concise and to-the-point, we have decided to leave additional analysis related to subdivision of space-time in triangular areas for future research. This also means that there still is room for improvement in the results shown in the next section. In this study, the space-time ratio used in Delaunay triangulation is selected to be equal to the space-time ratio of the defined estimation mesh, i.e., $\upsilon = 120$ km/h. When the triangular area dimensions are in line with estimation mesh dimensions, the PON estimator is expected to use the observations nearest (in space-time) to the desired estimation area.

The PON estimator is used at the same time interval as the reference estimator, i.e., every minute. For both estimators it is assumed that there is no data availability latency. Therefore, at the end of each minute both estimators estimate the traffic conditions for the four 15 s periods within this minute. In contrast to the reference estimator, the PON estimator is able to utilize information prior to this minute.

5.5.4 Evaluation of the estimators

To evaluate the overall performance of the different estimators we consider the bias (where a positive bias indicates an underestimation) and Root Mean Squared Error (RMSE). These statistics are used to evaluate the performance in terms of the flow and density, and for the two cases, i.e., the free-flow and the congested case. Although we focus on flow and density (which are the direct outputs of the PON estimation methodology), we will briefly discuss the estimation performance in terms of speed. We use a warm up period of 15 min, which ensures all observation paths are connected.

At first we will focus solely on the PON estimator. As explained in Section 5.4, the homogeneity and stationarity of the traffic conditions is expected to influence the estimation accuracy. The two cases shown in Figure 5.6, i.e., free-flow and congested, allow us to evaluate the estimation performance for different levels of inhomogeneity and non-stationarity. Furthermore, as the constructed triangular areas depend on the data-availability it is interesting to vary this factor and discuss the related estimation performance for the two cases. In terms of the data-availability we solely vary the penetration rate. Here, we will depict the estimation performance for seven different penetration rates, i.e., $p = 0.10, 0.25, 0.50, 1.0, 2.5, 5.0, 10.0\%$.

The main goal of the simulation study is to show the working of the proposed methodology. Additionally, to discuss the added value of the PON methodology and considered data, the PON estimator is compared with the reference estimator. Here, it is specifically interesting to see how the relative estimation performance differs between the two cases and the macroscopic traffic flow variables.

5.6 Results

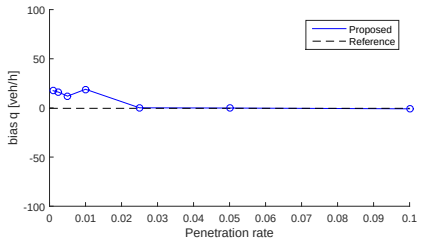
Figure 5.8 shows the estimation performance of the estimators in terms of the bias and RMSE. The eight sub-figures show the macroscopic traffic flow variables flow and density for the two cases, i.e., free-flow and congested. Each sub-figure shows two lines, i.e., one for the PON estimator (solid blue line) and one for the reference estimator (dashed black line). For all penetration rates p the upstream and downstream boundary are observed, i.e., we have more information than solely traffic sensing data from the moving observers.

5.6.1 Estimation performance of the PON estimator

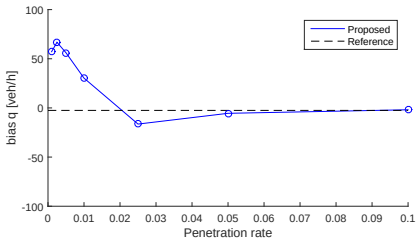
The estimation performance for the different penetration rates depends on the homogeneity and stationarity of the traffic conditions. The lower the penetration rate, the larger the constructed triangular areas. Therefore, for lower penetration rates these triangular areas are expected to individually be part of more rectangular areas in the defined estimation mesh. If the traffic conditions are inhomogeneous and non-stationary within the triangular area, the macroscopic traffic conditions are not representative for all related rectangular areas yielding estimation errors. This explains the clear difference between the RMSEs in the free-flow and the congested case observed in Figure 5.8. At low penetration rates, the RMSEs are larger in the congested case; however, at a penetration rate of 10.0% the performance in terms of RMSE is similar for both cases. In terms of flow estimates, the bias improves with the penetration rate (similar to the RMSE). However, in terms of density estimates, we observe a near-zero bias for all penetration rates.

To illustrate the importance of localizing changes in traffic conditions over space-time, we look at Figure 5.9. This figure shows that the estimates for PON estimator at three different penetration rates, i.e., p = 0.10, 1.0 and 10.0%, and the reference estimator. These estimates can be compared to the ground truth given in Figure 5.6d. The pattern at $p =$ 0.10% is caused by three individual vehicles of which the trajectories are shown using the black lines. These vehicles do not provide sufficient information to localize the congestion (delay). Although the congested pattern is more in line with the ground truth at $p = 1.0\%$, the identification of this pattern clearly improves when we move to $p = 10.0\%$. When estimates for $p = 10.0\%$, i.e., Figure 5.9c, with the ground truth, i.e., Figure 5.6d, we see that the estimator is able to approximate both the correct values of density and the pattern caused by the stop-and-go waves in congestion.

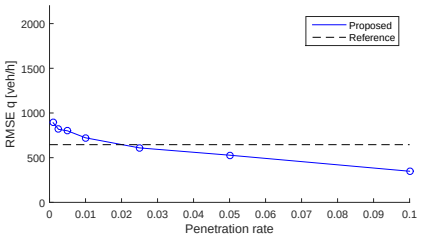
In addition to the free-flow and congested case, the PON estimator was applied to an incident case. This represents a form of non-recurrent congestion and a situation in which the traffic flow behavior temporarily changes due to a lane closure. The resulting estimation performance was similar to the congested case and therefore we do not present the detailed results in this paper. Similar results to the congested case were expected as the PON estimator does not use any information in the form of a traffic flow model or historical information. Therefore, the performance of the PON estimator should and did not differ between recurrent and non-recurrent congestion, or be effected by using an inaccurate description of the traffic flow behavior.



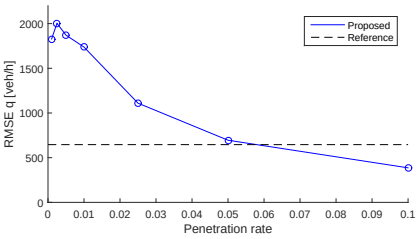
(a) Bias flow, free-flow case



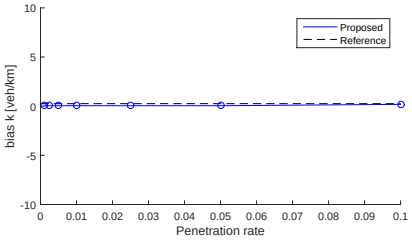
(b) Bias flow, congested case



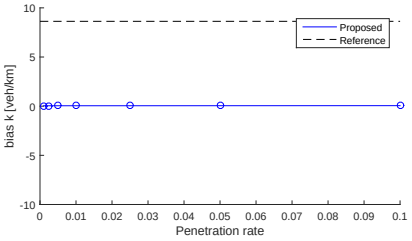
(c) RMSE flow, free-flow case



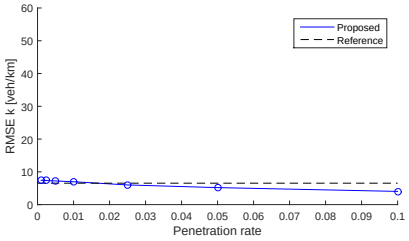
(d) RMSE flow, congested case



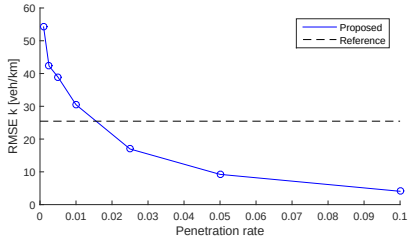
(e) Bias density, free-flow case



(f) Bias density, congested case



(g) RMSE density, free-flow case

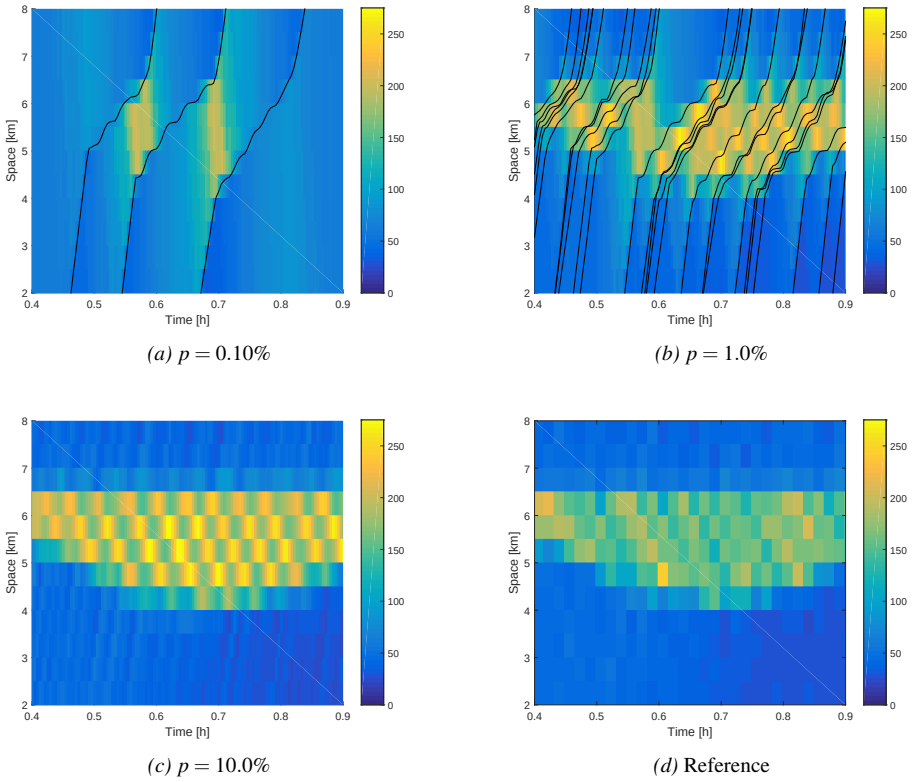


(h) RMSE density, congested case

Figure 5.8: Estimation performance (bias and RMSE) of the PON and reference (loop-detector) estimator for the free-flow and congested case.

5.6.2 Comparison between the PON and reference estimator

The simulation study allows us to compare the estimation performance of the PON estimator with the reference (loop-detector data) estimator. We discuss two elements of the relative



(a) $p = 0.10\%$ *(b)* $p = 1.0\%$

(c) $p = 10.0\%$ *(d)* Reference

Figure 5.9: Density estimation in space-time for the PON estimator at $p = 0.10$, 1.0 *and* 10.0%, *and the reference estimator. The moving observers trajectories (black lines) are shown for* $p = 0.10\%$ *and* 1.0%.

performance, i.e., (1) the relation to the inhomogeneity and non-stationarity in traffic conditions and (2) the differences between the macroscopic variables.

The relative performance of the two estimator depends on the level of inhomogeneity and non-stationarity of the traffic conditions. As discussed above, the PON estimator requires a higher data-availability for traffic phase identification if the inhomogeneity and non-stationarity of traffic conditions increases. The reference estimator uses observations from the middle of each cell, i.e., sensing equipment is installed every 500 m. As a result, the reference estimator accurately locates the different traffic phases in space-time for the congested case, see Figure 5.9. This explains the larger difference between the PON and reference estimators in the congested with respect to the free-flow case at low penetration rates.

With respect to the reference estimator, we observe a better relative performance of the PON estimator for density estimates than for flow estimates. In terms of flow, the relative estimation performance is better in the free-flow case than the congested case, i.e., the PON estimator outperforms the reference estimator in terms of RMSE respectively at penetration

rates higher than approximately 2.5% and 5.0%. The bias of the PON estimator at penetration rates higher than 5% and the reference estimator are both (near) zero. In terms of density, the PON estimator outperforms the reference estimator in terms of RMSE starting from a penetration between 1.0% and 2.5% for both cases. As shown in Figure 5.8f and by comparing Figures 5.6d and 5.9d, the reference estimator has a large bias in congested conditions. This bias is reduced if the loop-detectors would observe harmonic mean speeds instead of lane-specific time-mean speeds (ceteris paribus). In this case the PON estimator outperforms the reference estimator in terms of density RMSE starting from a penetration rate between 2.5% and 5.0% for the congested case.

In contrast to density and flow estimation, we do not recommend using the PON estimator for speed estimation. Existing methodologies that use trajectory data or probe speed data (e.g., (Work et al., 2010), (Del Arco et al., 2011)) are expected to outperform speed estimates with the PON estimator. However, our approach to combine stationary and moving observers is valuable for (the speed-related) dynamic link travel-time estimation. This approach allows us to estimate the cumulative curves at the upstream and downstream link boundaries, which in turn can be used to estimate the link travel-time, e.g., (Van Lint et al., 2014).

5.7 Conclusions and discussion

In this study, we propose an estimation procedure to estimate the macroscopic traffic conditions in a pre-defined space-time mesh using traffic sensing data collected by stationary and moving observers. This procedure consists of two processes: (1) obtain point-observations of the cumulative vehicle number N using stationary and moving observers and (2) estimate the macroscopic traffic conditions based on the point-observations of N. For the second, we designed a full methodology, which we denote as the Point-Observations N (PON) methodology. This methodology does not use any information related to traffic behavior (e.g., a fundamental diagram) or historical data and is thus a streaming-data-driven methodology.

The PON estimation methodology assumes homogeneous and stationary traffic conditions. If this assumption holds the methodology perfectly estimates the macroscopic traffic conditions in space-time. If this assumption is violated, errors are induced. Here, we make a difference between errors induced when estimating the conditions for the triangular areas that form the basic estimation unit and the errors induced when going from the triangular areas to the desired areas. The relation between these assumptions and the estimation errors explain that a positive relation exists between the level of inhomogeneity and non-stationarity of the traffic conditions and the required data-availability to reach a similar estimation performance. If the traffic conditions change highly over space and time, e.g., from free-flow to congestion or stop-and-go waves, having sufficient point-observations is important to localize the different traffic phases. Nevertheless, in the conducted case-study we still only needed to observe 1.0 to 2.5% of the vehicles in combination with the upstream and downstream 10 km link boundary to reach the same density estimation performance in terms of RMSE as having loop-detectors installed every 500 m. In this case, the flow estimation performance is similar for a penetration rate of approximately 5%. The estimation performance for the same link was also evaluated for solely free-flow conditions. Here, the PON estimation methodology outperforms loop-detector data-based estimated in terms of

both flow and density estimates at a penetration rate of 1.0 to 2.5%. At lower penetration rates the relative estimation performance is clearly better than in the congested case. The PON estimation methodology does not have advantages when estimating speed in a discrete space-time mesh. However, the approach to combine observations of stationary and moving observers can be used to accurately estimate the cumulative vehicle number at the upstream and downstream link boundaries, which in turn can be used to accurately estimate the link travel-times.

The estimation performance seems to be largely determined by the ability to localize the changes in traffic conditions in space and time. Especially, miss-localizing large changes in traffic conditions, i.e., from free-flow to congestion or stop-and-go waves, results in large errors. The ability to localize the changes in traffic conditions depends on the sampling (in space and time) of the point-observations of N. In the simulation study, the influence of the penetration rate (defined as the fraction of vehicles that is observed) on the estimation performance was discussed. However, the penetration rate is not the only factor affecting the sampling of the moving observers. The number of moving observers in space and time respectively depend on the penetration rate in combination with the density and flow. Furthermore, a fixed number of available observers can be spread differently in space and time, which can (depending on the changes in traffic conditions) affect the estimation performance.

In addition to the PON estimation methodology, the more general estimation procedure is an important contribution of this paper. Equipped and/or automated vehicles can be used to collect traffic sensing data on the relative flow with respect to their trajectories. This describes the change in the cumulative vehicle number ΔN over a path in space-time. These data (together with data from stationary observers) can be fused on the N-level using the principles discussed in this paper. This is highly valuable information as N is the core macroscopic traffic flow variable. Knowing N over space and time allows for deriving all three macroscopic traffic flow variables. The PON estimation methodology can be used to estimate the flow and density in a pre-defined space-time mesh; however, different (existing) methodologies may also be used to estimate macroscopic traffic conditions, e.g., N at non-observed points in space-time or the travel-time.

5.A Performance under observation errors

This paper proposes and evaluates new methodology, i.e., the PON estimation methodology. Up till now, we assumed that the data allow us to obtain error-free point-observations of N. However, in reality, it is likely that the data is not perfect. To investigate the effects of observation errors, we relax the prior assumptions, i.e., observation errors in N over the observation paths are introduced. As explained in Section 5.3 the data, i.e., a combination of sensing data from stationary and moving observers, allows us to deal with observation errors; however, we need to design an error correction methodology to do so.

The objective of this appendix is to show that the combination of sensing data and the PON estimator also yields a good estimation output when having to deal with observation errors. For this purpose, a simple error-correction methodology is designed. We opt to show that even a simple methodology suffices to deal with observation errors. Similar to the TSE methodology discussed in Section 5.4, we do not want to rely on parameters that have to be calibrated, as this is a way of incorporating additional information in the form of historical data.

5.A.1 Methodology

Section 5.3 explains the basic concept behind the error correction methodology: *An intersection of two already-initialized observation paths can be used for error correction. Observation errors can lead to a discrepancy between N at the intersection point on the two observation paths. As both paths should have the same N-value at the intersection point, we can correct for observation errors based on the difference in N. This principle is simple; however, in designing an error correction methodology, we need to define how the difference in N translates in to a correction in ΔN over the observation paths.*

Let us consider (as we do in the simulation study) a link with stationary observers at the upstream and downstream link boundaries, and more than one moving observer. To initialize N over the observation paths, the upstream link boundary ($x = 0$ m) is taken as the reference point and $N(0,0) = 0$. Upon entering the link, thus when the moving observer interacts with the upstream boundary, each moving observer is assigned an initial N. The downstream link boundary ($x = L$) is initialized by the first moving observer exiting the link. Initialization of the downstream boundary does not occur at $t = 0$ s. Therefore, in the simulation study, a warm up period of 15 min is used.

Any other intersection of observation paths is *an intersection of two already-initialized observation paths* and can thus *be used to for error correction* (see above). In our simple methodology, for each of these intersections, we define a leading and following path. The leading path has the most recent information from the upstream link boundary. In case of interaction between the downstream link boundary and a moving observer, the moving observer is leading. In case of an interaction between two moving observers, the overtaking (i.e., fastest) moving observer is leading.

A difference in N in the intersection point between the two observers is accounted for by the following observer in the period between its last and current interaction. In this way every observation is at most corrected one time. To account for the difference, which is the error, the ΔN are altered. the difference is spread out evenly over the absolute number of overtakings registered by the observer. For instance, if there is an error of +4 veh and

the following vehicle registered 16 overtakings, we add -0.25 veh to each of the 16 registered overtakings. Here we assume that the connected vehicle provides information on each overtaking it observes. For instance, if the connected vehicle registers that it overtakes one vehicle and is overtaken by another vehicle in one data-period, we want to know that two overtakings were registered and that $\Delta N = 0$ veh within this period. If the following observer did not register any overtakings, the difference is spread out evenly over the time periods.

This simple error correction methodology does not require any parameters. Furthermore, for each observed point in space-time there are maximally two values of N, that is, before and after correction. If the intersecting of the two relevant observations paths (which are used for correction) falls within the available data period, we use the corrected value.

5.A.2 Implementation in the simulation study

The performance under observation errors is evaluated for the cases that were discussed in Sections 5.5 and 5.6. The difference between the simulation study in these sections and this appendix lies in the observations and implementation of an error correction methodology (see above). Instead of assuming that we know N for specific points in space-time, we consider potentially erroneous observations of ΔN and estimate N based on these observations.

Stationary and moving observers should observe vehicles that they pass or that pass them. However, these observers can miss or incorrectly define vehicle-passings. In this study, we consider the following errors: each passing can either be (1) correctly observed, (2) missed or (3) double-counted. Furthermore, we assign equal probabilities to events (2) and (3).

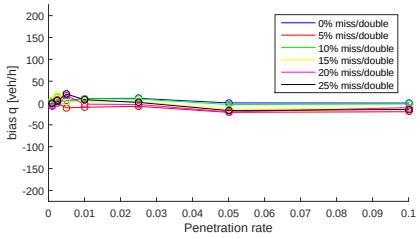
In theory, it is possible that with one data-period (5 s) a vehicle overtakes another vehicle and is directly overtaken again by the same vehicle. These overtakings are missed in our study, as we expose the passings based on the difference between the vehicle positions in consecutive time-instances, which are 5 s apart. However, as vehicle speeds are not expected to show large short-term fluctuations, we believe that we only miss a very small fraction of the overtakings.

In line with the main simulation study of this article, we evaluate the RMSE of density, flow and speed estimates for different penetration rates. However, in this simulation study, we also consider different probabilities for missing or double-counting overtakings, i.e., 0, 5, 10, 15, 20, 25%. A miss/double-count probability of 10% means that the probabilities of correct, missed and double-counted observations are respectively 90%, 5% and 5%.

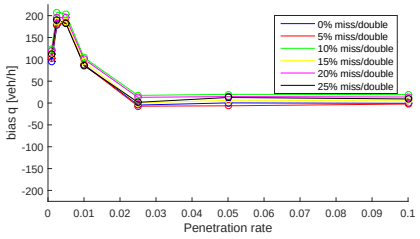
5.A.3 Results

Figure 5.10 shows the estimation performance for different miss/double-count probabilities. As expected, observation errors lead to a lower estimation performance. However, even with large observation errors (e.g., 25.0%), we observe a good estimation performance.

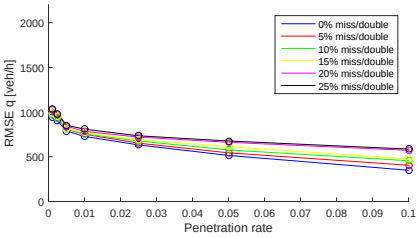
The estimation performance statistics, i.e., bias and RMSE, under observation errors (Figure 5.10) have a similar shape to those without observation errors. The RMSE error reduces with increasing penetration rate in both cases (free-flow and congested) and for both variables (flow and density). However, the influence of the observation errors, i.e., the difference between the lines with different miss/double-count probabilities, is smallest



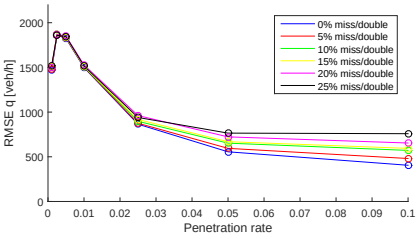
(a) Bias flow, free-flow case



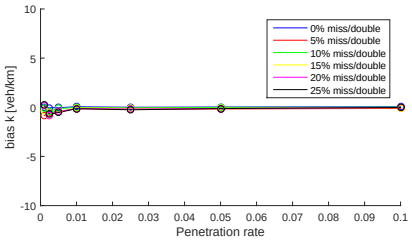
(b) Bias flow, congested case



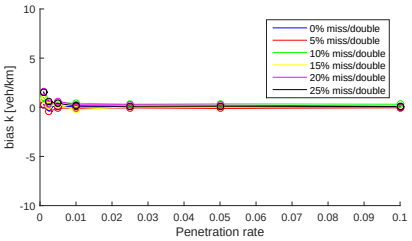
(c) RMSE flow, free-flow case



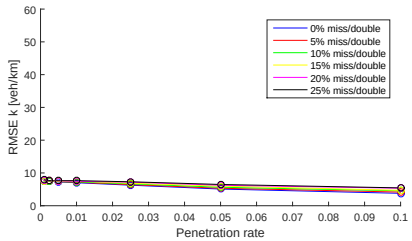
(d) RMSE flow, congested case



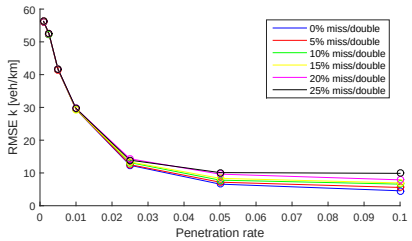
(e) Bias density, free-flow case



(f) Bias density, congested case



(g) RMSE density, free-flow case



(h) RMSE density, congested case

Figure 5.10: Estimation performance (bias and RMSE) of the PON estimator for the free-flow and congested case under observation errors.

for low penetration rates. At these low penetration rate, the influence of incorrect ΔN values between points in space-time may be of lower significance compared to the existing estimation errors (which are discussed in Section 5.6). Furthermore, the errors caused by

incorrect observations of ΔN can become more local, e.g., an underestimation and overestimation of the density respectively upstream and downstream of a moving observer can come hand-in-hand and both contribute to a larger RMSE.

For each observation error probability the bias moves towards a level that depends on the errors in the upstream stationary detector. However, depending on these errors this level lies at a certain non-zero value. This makes sense, as our error correction methodology does not try to correct all errors (e.g., no corrections are performed on the upstream link boundary), but opts to limit the errors between point observations in near proximity in space and time. This may be a reason to design a better error correction methodology; however, a stated before, this is out of the scope of this appendix. In contrast to the RMSE, the bias of one observation error probability can be better (i.e., closer to zero) than for a lower observation error probability. Again, this makes sense, as we consider a zero-mean error distribution. Therefore, additional errors may actually compensate earlier errors.

We discussed that the intersection between observation paths is important for initialization and error correction. However, the observers considered in this study, do not directly observe these intersection point. Instead, in the error correction methodology we needed to estimate the intersection points based on the trajectory observation points that have an interval of 15 s. Therefore, it would be beneficial if the stationary and moving observers observe the exact intersection points. However, as this puts an extra strain on the required data characteristics, in this study, we want to keep this as a recommendation for practice.

Chapter 6

Estimating the fundamental diagram using moving observers

This chapter is published as a conference paper: van Erp, P.B.C., Thoen, S., Knoop, V.L. & Hoogendoorn, S.P. (2018), Estimating the fundamental diagram using moving observers. *21st International Conference on Intelligent Transportation Systems (ITSC), IEEE*, Maui, Hawaii, USA, November 4-7, 2018. Initial work on this topic is included in the additional thesis 'Using relative flow data to expose the fundamental diagram', which is written by S. Thoen under the supervision of V.L. Knoop and P.B.C. van Erp. This additional thesis does consider the option to use Edie's definitions to estimate the flow and density for spatial-temporal areas, but does not include the methodology proposed in this chapter to estimate the fundamental diagram parameters.

Chapter 5 shows that relative flow data from stationary and moving observers can be used to estimate flow and density within a spatial-temporal estimation mesh. It shows that flow and density are estimated relatively accurately without adding information in the form of a traffic flow model. However, there are also indications that the methodology would benefit from using relations that can be exposed using historical data. For instance, previous studies have shown that traffic states propagate in space-time over certain characteristics speeds, and have used this information for the purpose of traffic state estimation, e.g., the Adaptive Smoothing Filter (Treiber & Helbing, 2002).

The fundamental diagram (FD) of traffic flow describes the relation between the macroscopic variables flow and density when traffic is in equilibrium. Combined with conservation-of-vehicles, this diagram yields the first-order LWR traffic flow model, which can be used in model-driven estimation and prediction. This study explores the option to expose the fundamental diagram using relative flow data collected using moving observers. For this purpose, two steps are taken. First, the data are used to estimate flow and density for spatial-temporal areas using Edie's definitions. Second, based on the $\{$flow,density$\}$-estimates, the parameters of the fundamental diagram are estimated.

6.1 Introduction

The fundamental diagram (FD) describes the relation between the macroscopic traffic flow variables flow q and density k when traffic is in equilibrium[1]. In combination with $q = ku$, where u is the mean speed, it describes the relation between all three macroscopic variables in equilibrium. The (parameters of the) FD are valuable for traffic state estimation and prediction (Seo, Bayen et al., 2017), and in traffic control measures, e.g., (Papageorgiou, Hadj-salem & Blosseville, 1991).

This paper proposes an alternative approach to estimate the FD based on traffic sensing data collected by moving observers. This method consist of two main steps. First, we use the individual passings observed by moving observers to estimate q and k for a set of areas in space-time using Edie's definitions (Edie, 1965). Using these data in combination with Edie's definitions is new to FD estimation. Secondly, we estimate the FD based on the $\{q,k\}$-estimates. Our methodology takes into account the representativeness of the $\{q,k\}$-estimates for the FD and does not require prior knowledge or assumptions related to any parameters of the FD. In this way it differs from existing methodologies that are discussed in the next section.

This paper is organized as follows. In Section 6.2, a literature overview is provided that serves as background information for the proposed approach. Sections 6.3 and 6.4 respectively explain the methodologies designed for the first and second step of the approach. Next, in Section 6.5, the simulation study to test the proposed approach is explained, after which the results of the study are reported in Section 6.6. Conclusions and an outlook are provided in Section 6.7.

6.2 Background on fundamental diagram estimation

This section covers the important topics related to exposing the fundamental relation between the macroscopic traffic flow variables, i.e., the fundamental diagram (FD). For this purpose, we discuss: (1) the macroscopic description of traffic flow, (2) the traffic sensing data that is used to estimate the fundamental diagram and (3) the important characteristics of the proposed FD estimation methodology.

Traffic flow can be described on a macroscopic level using flow q and density k. Edie (1965) provided the generalized definitions for these two variables for an space-time area:

$$q = \frac{\sum_i d_i}{A} \tag{6.1}$$

$$k = \frac{\sum_i ts_i}{A} \tag{6.2}$$

where the sum of the distance traveled and time spent by all vehicles with the area, i.e., $\sum_i d_i$ and $\sum_i ts_i$ are respectively the total travel distance (TTD) and total time spent (TTS). Furthermore, size of the space-time area is given by A. Given that flow and density are known, the mean speed for the space-time area is also known given the relation $u = \frac{q}{k}$.

[1] This chapter considers the situation in which the FD is a property of the entire road-stretch for which the estimates are obtain, i.e., we consider a homogeneous road stretch for which the FD function does not change over space.

Edie's definitions provide the true aggregated traffic conditions for the full space-time area. However, within this area different traffic states can be present. Cassidy (1998) argues that there is no reason to expect Edie's q, k and u to lie on a reproducible function, as the values will be merely a weighted average of the different traffic states. If traffic behaves according to a concave FD (e.g., the triangular FD) the weighted average of different traffic states will lie on (if both states are on the same linear function) or below (otherwise) the FD. Furthermore, if one traffic state holds for the full space-time area, the traffic state will lie on the FD.

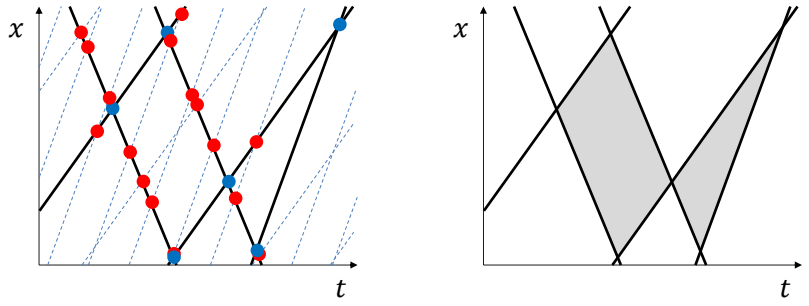
To estimate the FD we rely of observations of traffic flow, i.e., traffic sensing data. The dominant source of sensing data that are used to estimate the FD are stationary detectors that observe flow and speed aggregated over a time-period. For instance, Dervisoglu et al. (2009) and Knoop & Daamen (2017) use loop-detector data to fit/calibrate the FD. However, there are multiple important problems related to using stationary detectors for FD estimation: For instance, (1) density is underestimated (even when observing harmonic mean speeds) if traffic is non-stationary during the aggregation period, (2) the detectors do not detect traffic at standstill (Knoop & Daamen, 2017), and (3) FD estimation is limited to those road sections where detectors are installed (Seo, Bayen et al., 2017). Alternatively, it has been proposed to use probe data for FD estimation, e.g., Seo, Kusakabe & Asakura (2017). The general downsides of probe data are that include individual driving behavior, e.g., aggressive vs. timid driver are likely to behave differently in the same situation, and that it is difficult to accurate estimate flow and density. To overcome the latter problem, Seo, Kusakabe & Asakura (2017) assume that the jam density is known.

Due to technological advances, an increasing number of vehicles is equipped with sensing equipment that can observe other road-users (Florin & Olariu, 2017). These vehicles can be used to collect new types of traffic sensing data. For instance, they can serve as moving observers that observe passings of other road-users with respect to their own position over time (trajectory) (Florin & Olariu, 2017), (Van Erp, Knoop & Hoogendoorn, 2018b). Such new data may be used to estimate the FD.

The important characteristics of the proposed methodology are as follows. Our methodology does not require prior knowledge or assumptions related to any parameters of the FD, e.g., (Dervisoglu et al., 2009), (Knoop & Daamen, 2017) and (Seo, Kusakabe & Asakura, 2017) respectively define a minimum value for the free-flow speed, the value for the wave speed and the value of jam density. Furthermore, we estimate q and k for areas in space-time based on Edie's definitions. Thereby we overcome the discussed downsides of stationary detector and probe data. However, as discussed above, some estimates, which include multiple traffic states, may not be representative for the FD. Therefore, in the proposed methodology, we use the expectation that traffic behaves according to a concave FD and thus that these estimates lie on or below the FD.

6.3 Estimating flow and density for space-time areas using moving observers

In this study, we use relative flow data that are collected by moving observers. The path in space-time over which the moving observer travels is denoted as the observation path. We assume that each individual vehicle passing (and its direction) is observed.



(a) Vehicle trajectories on a road stretch (blue dashed lines) and moving observers (black lines)

(b) Two space-time areas (grey) that are enclosed by moving observers

Figure 6.1: Moving observers and relevant space-time areas.

Figure 6.1a shows the observation paths of five moving observers (black lines) and the observed individual passings (dots). In this example, respectively three and two moving observers travel in the driving and in the opposite-driving direction of the observed traffic flow. There are multiple advantages of observing the opposite-driving direction. For instance, these moving observers report more vehicle passings. However, for this paper the most important advantage is that there are more intersections (indicated with the blue dots) of the observation paths, as will be discussed below.

A set of intersecting observation paths can form areas in space-time for which all boundaries are observed, i.e., all boundaries are part of an observation path. Figure 6.1b shows the areas for which this holds given the same set of observation paths that are shown in Figure 6.1a. Given that we observe all passings over the observation paths, we observe each individual vehicle that enters or leaves the space-time area.

Based on the individual passings and their direction (in- or outflow of the area), we can determine the TTD and TTS. To determine the TTD and TTS, the trajectories of the observers traveling in driving direction have to be taken into account. A weight of 0.5 is assigned to these trajectories as the boundaries separate two adjacent areas. Furthermore, the size of the space-time area, i.e., A, can be determined based on the spatial-temporal characteristics of the observation paths. This provides the required information to estimate q and k using Edie's definitions, i.e., equations (6.1) and (6.2), thereby yielding a $\{q,k\}$-estimate for each enclosed area.

6.4 Estimating the fundamental diagram parameters

The methodology explained in this section can be used to fit a triangular FD to $\{q,k\}$-estimates obtained in Section 6.3. A triangular FD is a simple two-variate FD that consists of two connected linear branches, which we refer to as the free-flow and congested branch. According to Seo, Bayen et al. (2017) the triangular FD is popular due to its simplicity, theoretically preferable features and some empirical evidence. Figure 6.2 schematically visualizes a triangular FD with the important parameters, i.e., free-flow speed v^f, wave

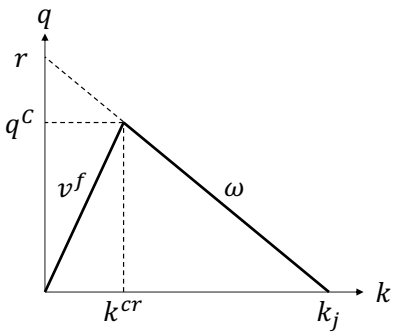


Figure 6.2: Triangular fundamental diagram with important parameters.

speed w, capacity q^C, critical density k^{cr}, jam density k_j and passing rate r. The triangular FD is described by the following function:

$$\begin{aligned} q &= v^f k \quad \text{if } k \leq k^{cr}, \\ q &= r - wk \quad \text{if } k > k^{cr}. \end{aligned} \tag{6.3}$$

The free-flow and congested branch are connected at $\{k^{cr}, q^C\}$, i.e., $q^C = v^f k^{cr} = r - wk^{cr}$. Therefore, three parameters suffice to define the triangular fundamental diagram (e.g., v^f, w and k^{cr}).

To estimate the FD based on combined estimates of q and k, i.e., $\{q,k\}$-estimates, we need to know which $\{q,k\}$-estimates are representative for the free-flow branch and which are representative for the congested branch. As explained before, the $\{q,k\}$-estimates, that are obtained using Edie's definitions, will be the weighted average of the different traffic states in the area. Therefore, if traffic is deterministic and its behavior is correctly described by a triangular FD, the $\{q,k\}$-estimate either lie on or below the FD.

The critical density k^{cr} is the location of a structural break in describing q based on k. Estimates for which $k \leq k^{cr}$ can be representative for the free-flow branch, but not for the congested branch. Alternatively, estimates for which $k > k^{cr}$ be representative for the congested branch. However, in both cases, the observation might still include traffic states from both traffic phases, which are expected to lie below the line (branch) of the FD. If we known the location of the structural break, i.e., if we know k^{cr}, we can choose the remaining two parameters (defining the FD) such that *all* observations lie on or below the FD. To this end, we need an approach to find this unknown structural break.

Finding unknown structural breaks is studied extensively in the field of time-series analysis. The Quandt-Andrews breakpoint test (Quandt, 1960), (Andrews, 1993) is commonly applied approach to (1) find the location of the 'largest' structural break and (2) test if this structural break is significant. This test examines the different possible locations of the structural break and applies a Chow test (Chow, 1960) for each location. For each Chow-break-test, the F-statistic is calculated and the location with the maximum (supreme) F-statistic is selected as the structural break location. Depending on the F-statistic at this location and in the case that all assumptions related to the error distribution hold, we can say whether there is a significant structural break.

The proposed methodology is based on the Quandt-Andrews breakpoint test; however, there are two important differences. Firstly, in contrast to time-periods (which are used as explanatory variables in time-series analysis), k is a continuous and will (most-likely) have varied intervals. This limits the accuracy in finding the correct k^{cr}, as the correct k^{cr} may lie in between two consecutive k^{cr} that are considered. However, by reducing the step-size in the k^{cr} that are evaluated, we can improve the accuracy in finding the structural break. Secondly, the assumptions related to the error distribution will not hold as all 'errors' will be negative (lie below the fitted line). The term 'errors' is used as there can still be differences between the fitted FD and $\{q,k\}$-estimate while the fitted FD is correct. Therefore, in the remainder of this article we refer to differences instead of errors. However, similar to an error we still minimize a difference statistic (see FD estimation sequence) as an erroneous FD induces larger differences. Dealing with differences instead of errors prevents us from testing whether the structural break is significant. However, this is not the purpose of the proposed methodology, as we always assume that there is a structural break in the FD.

Given the explanations and assumptions discussed above, the following FD estimation sequence is designed (note that steps 2b and 2d are a result of the expectation that all $\{q,k\}$-estimate either lie on or below the FD). In the estimation sequence, the full set of $\{q,k\}$-estimates is given by O. To find the location of the structural break we minimize a difference statistic between the estimated FD and $\{q,k\}$-estimates. The choice for the sum of squared differences (see below) in not crucial. Other statistics, e.g., the mean absolute difference, may be chosen without having a large effect on the final result.

FD estimation sequence

1. Define a set of n to-be evaluated k^{cr}, i.e., $\mathbf{k}^{cr} = [k_1^{cr} \cdots k_n^{cr}]$
2. For each k_i^{cr}:
   (a) Define subsets of free-flow observations, i.e., $O_{\mathcal{F}} = \{o \in O | k_o \leq k_i^{cr}\}$, and congested observations, i.e., $O_C = \{o \in O | k_o > k_i^{cr}\}$.
   (b) Find the free-flow speed v_i^f:

$$v_i^f = \max_{o \in O_{\mathcal{F}}} \left(\frac{q_o}{k_o} \right) \tag{6.4}$$

   (c) Calculate capacity q_i^C:

$$q_i^C = v_i^f k_i^{cr} \tag{6.5}$$

   (d) Find the wave speed w_i:

$$w_i = \max_{o \in O_C} \left(\frac{q_o - q_i^C}{k_o - k_i^{cr}} \right) \tag{6.6}$$

   (e) Calculate passing rate r_i:

$$r_i = q_i^C - w_i k_i^{cr} \tag{6.7}$$

(f) Calculate the Sum of Squared Differences SSD_i:

$$SSD_i = \sum_{o \in [k_o \leq k_i^{cr}]} \left(q_o - v_i^f k_o \right)^2 + \sum_{o \in [k_o > k_i^{cr}]} \left(q_o - (r_i + w_i k_o) \right)^2 \tag{6.8}$$

3. Find the location of the structure break, i.e., find i for which $\min_i = SSD_i$, and select the related parameters as the estimate for the FD.

6.5 Simulation study

A simulation study is conducted to evaluate the performance of the proposed FD estimation approach. This section contains the following elements: explanation of (1) the two applied (microscopic) car-following models and the simulation traffic conditions and data characteristics and (2) the evaluations performed to test and understand the methodology.

6.5.1 Microscopic simulation of traffic

Traffic is simulated using two different car following models: Newell's car following model (G. Newell, 2002) and the Intelligent Driver Model (IDM) (Treiber, Hannecke & Helbing, 2000). The main difference between the models lies in the presence of transient phases. In Newell's model speed changes occur instantaneously and hence no transient phases are found. The IDM includes acceleration and deceleration towards a desired speed and hence includes transient phases. Table 6.1 contains the parameters that we use for simulating traffic using Newell's model and IDM.

Both models are deterministic, which means that the traffic flow simulated with both models results in a deterministic FD. However, the shape of the FD differs for the two models. When using Newell's model, the FD is triangular and is given by:

$$\begin{array}{ll} q = v^f k & \text{if } k \leq k^{cr}, \\ q = \frac{1}{\tau} - \frac{s_{jam}}{\tau} k & \text{if } k > k^{cr}\,. \end{array} \tag{6.9}$$

When using IDM, the FD is a smooth continuous function. We can describe the spacing s as a function of speed u for equilibrium conditions, i.e., the acceleration is zero:

$$s = \sqrt{\frac{(s_0 + l + uT)^2}{1 - \left(\frac{u}{v_0}\right)^{\delta}}} \tag{6.10}$$

Based on this function we can plot the FD, where $k = 1/s$ and $q = u/s$.

Table 6.1: Parameters used for Newell's model and IDM

Model	Parameter	Value	Unit
Newell's	v^f	33.33	m/s
	s_{jam}	6.00	m
	τ	0.90	s
IDM[1]	v_0	33.33	m/s
	T	1.20	s
	δ	4.00	-
	a	0.80	m/s^2
	b	1.25	m/s^2
	s_0	1.00	m
	l	5.00	m

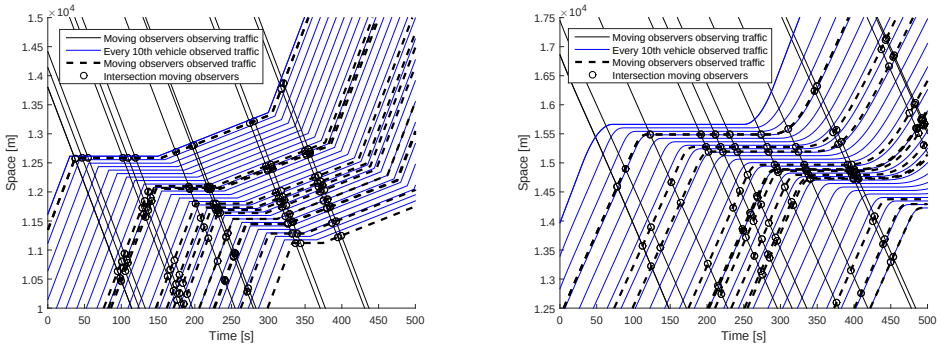
[1]) Default values from Treiber, Hennecke & Helbing (2000)

In this simulation study, we want to estimate the FD for a road. For this purpose, traffic is simulated in two directions and data is collected from moving observers that are driving on this road and those that are driving in opposite direction. In both directions 250 veh are simulated over a period of 600 s with time-steps of 0.1 s. We only estimate the FD for one direction. Traffic in this direction is denoted as the 'observed traffic', while the opposing direction is denoted as the 'observing traffic'. The initial positions of the most upstream vehicle is respectively 0 m and 30,000 m for the observed and observing traffic. To determine the spacing between each combination of consecutive vehicles, we take a random draw from an exponential function with mean 10 m plus $s_{jam} + u\tau$ (for Newell's model) or $s_0 + l + uT$ (for IDM).

As we need to observe both free-flow as congested states in order to estimate the FD, a bottleneck is simulated in the observed traffic flow. In the simulation with Newells model, a bottleneck is simulated on the observed stream by reducing the speed of the first vehicle to 0 m/s for 30 s $\leq t <$ 150 s, to 5 m/s for 150 s $\leq t <$ 300 s, and increasing it back to v^f at $t = 300$ s until the end of the simulation. In the IDM simulation, the first vehicle decelerates with a m/s^2 to standstill starting at t = 30 s, and accelerates back to v^f with a m/s^2 starting at t = 250 s. Congested states are found upstream of this vehicle. No bottlenecks are included in the observing traffic flow, which thus has a constant speed of v^f.

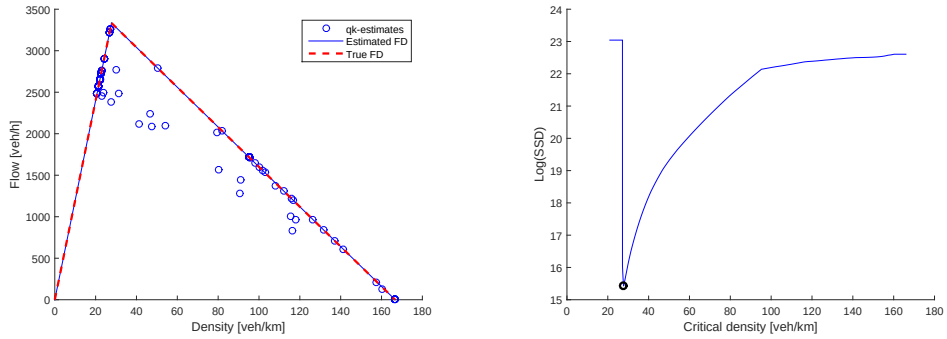
It is assumed that the moving observers are able to observe all vehicles that they pass in opposite direction, i.e., the moving observers that are part of the 'observing traffic' observe the 'observed traffic'. For both directions 5 % of the vehicles is a moving observer. Following the principles explained in Section 6.3, the trajectories of these vehicles are used to construct areas in space-time for which we can obtain $\{q,k\}$-estimates.

Figure 6.3 shows the trajectories of the observed traffic and the moving observers in space-time for Newell's model and IDM.



(a) Newell's model *(b)* IDM

Figure 6.3: Simulated traffic flow and moving observers.



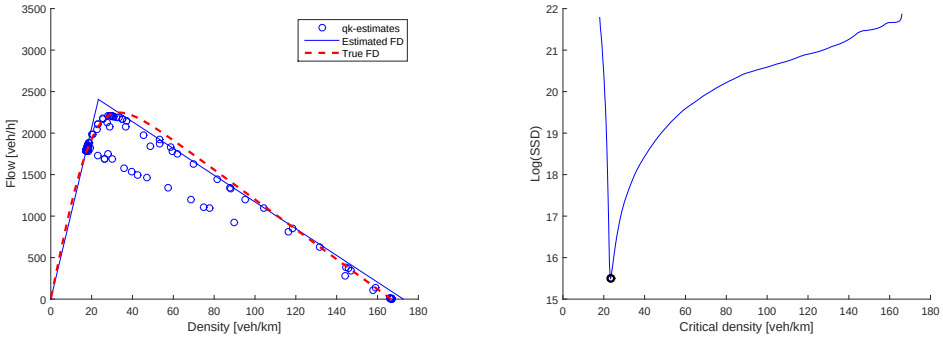
(a) Estimated FD *(b) SSD* for different k^{cr}

Figure 6.4: Estimated FD for Newell's model.

6.5.2 Evaluation of the proposed methodology

The objective of this simulation study is to evaluate the performance of the proposed methodology in correctly estimating the FD. Therefore, we compare the estimated FD with the true FD.

For both traffic models (Newell's model and IDM) we show (1) the $\{q,k\}$-estimates, (2) the estimated triangular FD and (3) the true FD. Furthermore, for each of the evaluated k^{cr} the Sum of Squared Differences, i.e., *SSD* is reported. Due to the extreme variation in *SSD*, we report the $\log(SSD)$. These plots allow us to evaluate the performance of the proposed methodology and discuss in more detail how it works.



(a) Estimated FD *(b) SSD* for different k^{cr}

Figure 6.5: Estimated FD for IDM.

6.6 Results

Figures 6.4 and 6.5 respectively show the estimated triangular FDs for traffic that is simulated using Newell's model and the Intelligent Driver Model (IDM). Furthermore, they depict the sum of squared differences between the estimated FD and $\{q,k\}$-estimates. For both car following models, we observe $\{q,k\}$-estimates that lie approximately on a line between a traffic state on the free-flow branch and the jam density traffic state, see Figures 6.4a and 6.5a. These estimates are weighted averages of the free-flow and jam density traffic states within the space-time area related to these estimates. Below, we separately discuss the accuracy of the estimated FDs for traffic simulated with both models.

The proposed methodology accurately estimates the FD that holds for the traffic simulated with Newell's model. The estimated (blue lines) and true (red dotted line) FD almost perfectly overlap in Figure 6.4a. In case a k^{cr} is considered that is below the true k^{cr}, there might be observations in the free flow branch which are now (incorrectly) assigned to the congested branch. This changes the direction of the congested branch, giving a positive wave speed w. This explains the instant decrease in $\log(SSD)$ in Figure 6.4b between $k^{cr} = 27.77$ veh/km and $k^{cr} = 27.78$ veh/km.

The estimation of the IDM FD is not as accurate as for Newell's triangular FD. The proposed methodology makes two important assumptions, i.e., (1) the FD has a triangular shape and (2) all $\{q,k\}$-estimates lie on or below the FD. As shown in Figure 6.5a the true FD (red dotted line) does not have a triangular shape. However, given the restrictive shape, the estimated FD (blue line) still seems to be a good fit. The sum of squared differences *SSD* (Figure 6.5b) shows a less extreme change prior to the optimum k^{cr} than for the Newell's model-based FD. This is a result of the smoothed top of the true IDM-FD. The combination of the two assumptions may lead to a larger difference between the true and estimated FD if other states on the FD are observed. For instance, at lower density k a larger speed u can be observed. If such a state is part of the set of $\{q,k\}$-estimates, the proposed methodology would estimate a larger v^f. Therefore, instead of the maximum observed u, we may want to consider all $\{q,k\}$-estimates that are relevant for the free-flow branch. In this case, the challenge lies in defining which estimates are relevant.

6.7 Conclusions and outlook

This paper proposes an approach to estimate the fundamental diagram (FD) based on relative flow data collected with moving observers. This approach consists of two steps: (1) estimate flow and density based on the sensing data for areas in space-time using Edie's definitions and (2) estimate a triangular FD based on the flow and density estimates, and their theoretical relation to the FD.

The proposed approach works well. The flow and density estimates, i.e., $\{q,k\}$-estimates, provide valuable information to find the FD. With a simple algorithm, which assumes that the $\{q,k\}$-estimates lie on or below the FD, we were able to accurately expose the FD.

In reality the $\{q,k\}$-estimates can lie above the desired FD. For instance, if traffic behaves stochastic and we may observe states that lie above the desired mean FD. Furthermore, we may have to deal with observation errors, which yield $\{q,k\}$-estimates that do not perfectly describe the true traffic conditions. In this case, the approach proposed in this work can still be followed, i.e., obtain $\{q,k\}$-estimates using Edie's definitions and estimate the FD based on these $\{q,k\}$-estimates. However, one should alter the second step and estimate each branch based on the $\{q,k\}$-estimates relevant for that branch. This is more complex than selecting the maximum value for each branch and will therefore lead to a more extensive methodology[2].

[2] A more extensive methodology should use the fundamental principles introduced in this chapter. The simple methodology introduced in this chapter takes into account that multiple traffic states may be present within the spatial-temporal area for which the $\{q,k\}$-estimates are obtained. It remains important to take this into account when estimating the FD for traffic that behaves stochastic. Whether different traffic states are present within the spatial-temporal area for which the $\{q,k\}$-estimates are available may be determined based on the observed traffic states at the area boundaries. Based on this information, we can trust certain $\{q,k\}$-estimates more (i.e., estimates that relate to areas with largely homogeneous and stationary traffic conditions) than others (i.e., estimates that relate to areas with states that lie on different branches of the FD).

Chapter 7

On the value of relative flow data

This chapter is presented (podium) at the 23rd International Symposium on Transportation and Traffic Theory (ISTTT23) in Lausanne, Switzerland, and published as a journal article: van Erp, P.B.C., Knoop, V.L. & Hoogendoorn, S.P. (2019), On the value of relative flow data. *Transportation Research Part C: Emerging Technologies*, in press.

The previous chapter shows that relative flow data can be used to estimate the parameters of the fundamental diagram (FD) of traffic flow. However, it solely focuses on estimating the FD parameters and does not consider model-driven traffic state estimation using relative flow data as a whole.

This chapter takes the perspective of model-based traffic state estimation using relative flow data and aims to present the value of relative flow data for the two processes that take data as input. These processes are: (1) learning models that can be used in real-time estimation based on historical data and (2) fuse real-time data and models to estimate the traffic state. Before discussing the value of data, we present principles that can be used for both processes (which can be seen as foundations for full methodologies). Based on these principles, we explain why it is valuable to collect relative flow data and which spatial-temporal data characteristics are desired. These spatial-temporal data characteristics are related to the type of observers that collect the data (i.e., stationary observers have a fixed location over time, while moving observers can travel at different speeds). Therefore, they provide insight in the value of collecting relative flow data with moving observers in addition or as an alternative to collecting these data with stationary observers.

7.1 Introduction

Traffic state estimation (TSE) is an important element in road traffic operations management and planning (Seo, Bayen et al., 2017). It is, for instance, important in infrastructure planning, dynamic traffic management (DTM) systems and navigation services. For each application the desired estimation output can differ in terms of accuracy, reliability and semantics. For instance, navigation services and DTM systems require real-time information, while this time-constraint is not there for infrastructure planning applications. Furthermore, the required spatial resolution can differ, e.g., a local DTM system requires less spatial coverage than a navigation service that needs information for all elements on each potential route.

A road traffic network can be represented as a set of links (roads), nodes (intersections or discontinuities like lane drops) and network boundaries (source and sink nodes) (Seo, Bayen et al., 2017). This paper focuses on TSE on links; however, as will be explained in this paper, information that propagates over a network of links and nodes can also be valuable for link TSE. Therefore, we will discuss how information can propagate over such a network.

On a link, traffic flow can be described using three dimensions, i.e., space x, time t and cumulative flow N (Makigami et al., 1971). The cumulative flow $N(x,t)$ is defined as the number of vehicles that have passed location x at time t, where it is important that we include the same set of vehicles at all locations. If $N(x,t)$ is smoothed and continuously differentiable, the macroscopic traffic flow variables flow q, density k and mean speed u can be obtained by taken the derivatives of $N(x,t)$ to time and space. Furthermore, overall link properties, i.e., the link travel-time and link vehicle accumulation (number of vehicles on the link), can be obtained from the cumulative flow at the link boundaries, i.e., x_0 (upstream boundary) and x_L (downstream boundary). Due to large informative value of the cumulative flow, in this study, we focus on estimating $N(x,t)$ within a space-time domain.

This study presents the value of relative flow data for TSE. This variable (relative flow) describes the change in cumulative flow, i.e., ΔN, over a path in space-time (which we denote as an observation path). Automated or other equipped and connected vehicles could serve as moving observers that collect these data by observing the vehicles that they overtake and that overtake them (Redmill et al., 2011), (Florin & Olariu, 2017), (Van Erp, Knoop & Hoogendoorn, 2018b). Furthermore, stationary observers, e.g., loop-detectors or road-side camera-systems, can be used to observe the flow at a fixed position over time. For stationary observers the observation path is a horizontal line in space and time. Although both describe the flow relative to the observation path, only stationary observers consistently observe the macroscopic variable flow q.

To present the *value of data*, we take the perspective of a model-based estimation approach, see Figure 7.1. In this approach, the data are used for two processes (which are highlighted using the blue boxes): (1) Real-time data and models are fused, which we denote as information assimilation (IA), to estimate the traffic state. And, (2) historical data are used to learn the models that are used in IA. This study focuses on explaining the principles that can be used for both processes and how relative flow data can be transformed into valuable information using these principles. Therefore, the principles help us to understand why the characteristics of relative flow data (i.e., the combination of the observed variable and spatial-temporal characteristics) have a large value in traffic state estimation. Design-

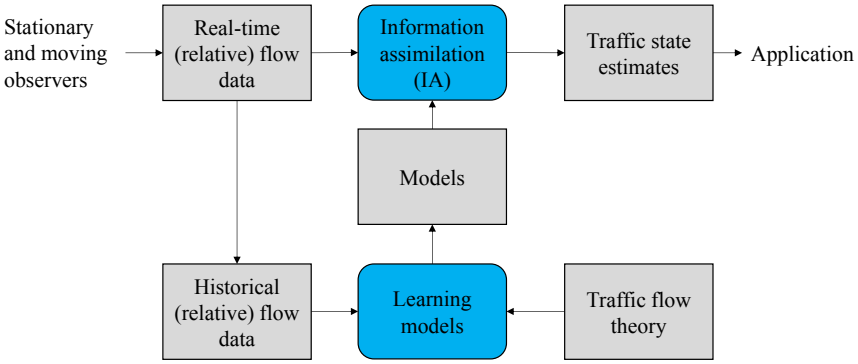


Figure 7.1: Schematic representation of the considered model-based estimation approach.

ing and testing full methodologies for the two processes, which are aimed at extracting the potential information in the data, is beyond the scope of this study.

Throughout this study, we consider data scenarios that rely on a combination of stationary and moving observers, and that rely on moving observers alone. Positioning stationary observers at the link boundaries aligns with existing studies, e.g., Claudel & Bayen (2010b) and Sun et al. (2017), as the resulting boundary observations play a crucial role in the estimation of $N(x,t)$. However, multiple studies opt to limit the dependence of these stationary observers due to the high installation and operation costs (Seo, Bayen et al., 2017). Therefore, we will focus on scenarios in which we solely rely on moving observers to collect the relative flow data.

The main contributions of this paper are: (1) We show which valuable information relative flow data provide to learn the link traffic flow model. For this purpose, we first explain which learning dataset can be constructed using relative flow data and then explain which informative value this dataset has to find the model parameters. (2) We show how relative flow data can be fused with the link traffic flow model to connect observation paths, estimate the link boundary conditions, and estimate the link supply and demand. These principles can be used to estimate the traffic state based on moving observers alone. Furthermore, we explain that, in absence of stationary observers that are positioned at the link boundaries, it may be valuable to consider information propagation over nodes and that the estimates (i.e., link boundary conditions, supply and demand) may be used for this purpose. (3) Based on these principles, we explain why it is valuable to observe the change in cumulative flow using observers that travel at different speeds with respect to each other and with respect to the characteristic wave speeds of the link traffic flow model.

This paper is organized as follows. In Section 7.2, the potential relative flow data characteristics are discussed are compared to other data types. Section 7.3 explains why the relative flow data are valuable in the two processes that are part of the model-based estimation approach. In that section, different principles are explained that can be used to learn the link traffic flow model and estimate the cumulative flow over space and time. To test and provide a better understanding of these principles, a simulation (VISSIM) case study was conducted and is presented in Section 7.4. Finally, the conclusions are presented in Section 7.5.

Throughout this paper many different symbols are used. A description of these symbols is provided in Table 7.1.

Table 7.1: Description of symbols that are used in this article.

Description	Notation	Unit
Position	x	m
Time	t	s
Location upstream link boundary	x_0	m
Location downstream link boundary	x_L	m
Flow	q	veh/s
Density	k	veh/m
Mean speed	u	m/s
Cumulative flow	N	veh
Position of observer o at time t	$X_o(t)$	m
Speed of observer o at time t	$V_o(t)$	m/s
Cumulative flow on the observation path of observer o at time t	$N_o(t)$	veh
Time when observer o is at position x	$T_o(x)$	s
Triangular fundamental diagram:		
Free-flow speed	v^f	m/s
Wave speed	w	m/s
Passing rate	r	veh/s
Capacity	q^C	veh/s
Critical density	k^{cr}	veh/m
Jam density	k^j	veh/m
Newell's method:		
Boundary over which the cumulative flow is known	$\mathcal{B}$	
Space-time point for which the cumulative flow is estimated	P	
Change in cumulative flow over v^f-wave from $\mathcal{B}$ to P	$\Delta N^{\text{ff}}_{\mathcal{B}P}$	veh/s
Change in cumulative flow over w-wave from $\mathcal{B}$ to P	$\Delta N^{\text{cg}}_{\mathcal{B}P}$	veh/s
Speed of potential characteristic waves	v^c	m/s
Model parameter p guess	$\hat{p}$	*Depends on parameter*
Error in model parameter guess	ε	*Depends on parameter*
Error in element of equation (7.4)	Υ	m/s
Supply	S	veh/s
Demand	D	veh/s
Upper bound restriction of the cumulative flow	$\tilde{N}^+$	veh
Lower bound restriction of the cumulative flow	$\tilde{N}^-$	veh

7.2 Collecting relative flow data with stationary and moving observers

This section discusses the potential characteristics of relative flow data. We make a distinction between data collected with stationary observers and data collected with moving observers. The former are collected with road-side sensing equipment, e.g., loop-detectors (Treiber & Helbing, 2002), (Wang & Papageorgiou, 2005), (Van Hinsbergen et al., 2012). The latter may be collected with vehicles equipped with sensors that observe the other road-users (Redmill et al., 2011), (Florin & Olariu, 2017), (Van Erp, Knoop & Hoogendoorn, 2018b), (Van Erp, Thoen et al., 2018).

Below, we first discuss the potential characteristics of relative flow data that are collected with stationary or moving observers (Section 7.2.1). Next, Section 7.2.2 discusses which

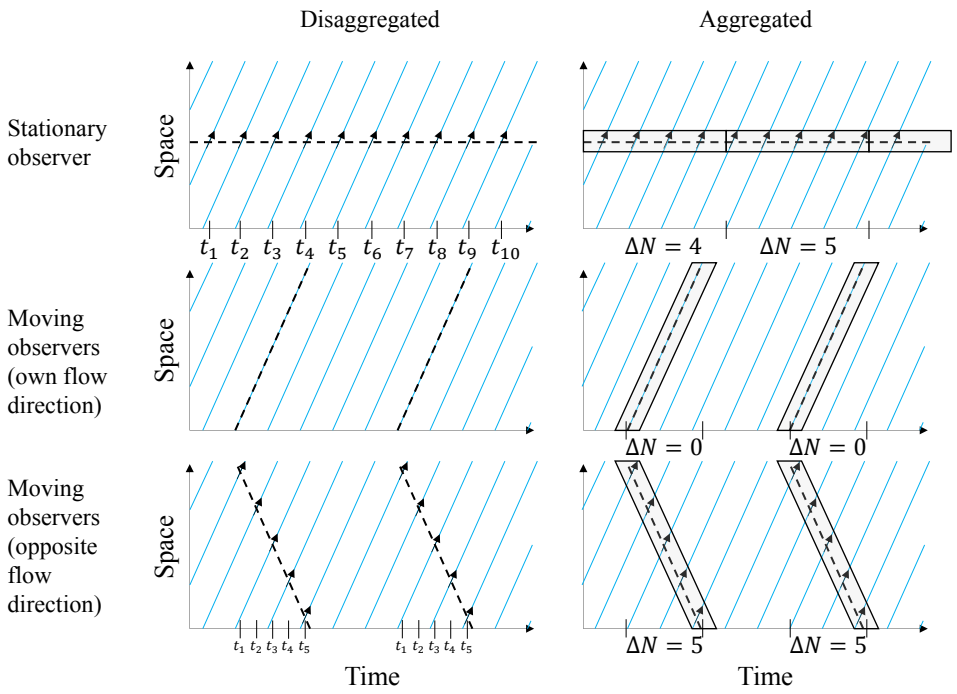


Figure 7.2: Aggregation level and spatial-temporal characteristics of (relative) flow data collected by different observers.

types of data are used in studies that propose methodologies that estimate the traffic state in or via the cumulative flow plane.

7.2.1 Potential data characteristics

Stationary and moving observers both observe the flow relative to a path in space-time. We refer to this path as the observation path. The position and speed of observer o at time t are respectively given by $X_o(t)$ and $V_o(t)$, where $V_o(t) = \partial_t X_o(t)$. The observation path of a stationary observer is a horizontal line in the space-time plane, i.e., it observes the flow with respect to a fixed position over time. Moving observers observe the flow with respect to their position over time, i.e., their trajectory. Depending on the sensing equipment and the infrastructure moving observers may also observe the flow for the traffic flow in the other direction. In this case, the moving observer is not part of the traffic in the direction of interest and moves with a negative speed through space-time. Figure 7.2 shows the observation paths for stationary observers, and moving observers that are part of the traffic flow or move in opposing direction. The observer speeds of these types of observers will respectively be $V_o(t) = 0$, $V_o(t) \geq 0$ and $V_o(t) \leq 0$.

Traffic sensing data can be disaggregated or aggregated (Seo, Bayen et al., 2017). We solely use this categorization to define the flow data. Disaggregated flow data describe all

individual passings, i.e., time and direction. For stationary observers the direction is always the same, i.e., the vehicle overtakes the stationary observer. In this case the cumulative flow N increases by one, i.e., +1. Moving observers that travel in flow direction can be overtaken by another vehicle (N increases by one) or overtake another vehicle (N decreases by one). Aggregated flow data describe the sum of all passings within a time-period. Figure 7.2 visually shows the difference between disaggregated or aggregated relative flow data for the different observers.

The observation path for stationary observers is fixed. Therefore, information related to the passing (disaggregated) or flow (aggregated) suffices. However, spatial-temporal information, i.e., its position over time, is needed for moving observers. This information may be shared with a fixed time-interval (temporal-sampling), with a fixed space-interval (spatial-sampling), in case of an event, e.g., passing, or a combination of these options. If the flow data is disaggregated it may be logical that the spatial-temporal information is provided in case of an event. However, even if the cumulative flow does not change along an observation path, i.e., there is no event, it is still valuable to known the paths along which N is constant.

The observed cumulative flow is a discrete variable. However, to understand the relation of the relative flow to the macroscopic variables flow q, density k and mean speed u, we can consider the smoothed and continuously differential cumulative flow. Following (Makigami et al., 1971) the three macroscopic variables can be determined for points in space-time by taking the derivatives to space and time, where $q(x,t) = \partial_t N(x,t)$, $k(x,t) = -\partial_x N(x,t)$ and $u(x,t) = q(x,t)/k(x,t)$. The smoothed cumulative flow along an observation path is given by $N(X_o(t),t)$, which we denote as $N_o(t)$. This allows us to describe the relative flow along an observation path, which is the derivative to time of the cumulative flow along this path, i.e., $\partial_t N_o(t) = k(X_o(t),t)\,[u(X_o(t),t) - V_o(t)]$.

7.2.2 Data used in existing studies that estimate in or via the cumulative flow plane

Multiple methodologies have been proposed to estimate the traffic state in the cumulative flow plane. For instance, Newell's (three-detector) method (G. F. Newell, 1993a), (G. F. Newell, 1993c), (G. F. Newell, 1993b), (Laval et al., 2012), Claudel's method (Claudel & Bayen, 2010b), (Claudel & Bayen, 2010a) and Sun's method (Sun et al., 2017) can be used to estimate the cumulative flow for points in space-time. Other methodologies have been proposed to estimate traffic flow features that can be directly derived from the cumulative flow curves at link boundaries, e.g., the vehicle accumulation (Van Lint & Hoogendoorn, 2015), (Amini et al., 2016) and the travel-time (Bhaskar et al., 2010).

All these studies work with stationary observers that are positioned at the link boundaries. Furthermore, some studies, e.g., Claudel & Bayen (2008), Claudel & Bayen (2010b), Claudel & Bayen (2010a), Bhaskar et al. (2010), Van Lint & Hoogendoorn (2015) and Sun et al. (2017), consider probe trajectory (or vehicle reidentification) data to describe ΔN over a path (the probe trajectory) or between specific points in space-time. Here, it is assumed that the probe vehicles do not overtake other vehicles or are overtaken by other vehicles, i.e., they assume $\Delta N = 0$ over the probe trajectory. The reason for adding the probe trajectory data to the stationary observers is twofold. Firstly, it yields internal conditions (constraints) that can be used in estimation, e.g., Claudel & Bayen (2010b) and Claudel & Bayen (2010a).

Secondly, intersecting observations paths can be used to initialize the cumulative flow over observation paths and mitigate the cumulative error problem, e.g., Bhaskar et al. (2010) and Van Lint & Hoogendoorn (2015).

However, there are important differences between probe trajectory data and relative flow data collected by moving observers. In multi-lane traffic, i.e., if overtaking is possible, the cumulative flow can change over the probe or moving observer trajectory (Sun et al., 2017). This change is not observed by the probes, but is observed by the moving observers. Therefore, moving observers that are part of the traffic flow provide a more accurate representation of the change in cumulative flow over the observation path. Furthermore, depending on the road infrastructure and equipment installed in the vehicle, moving observers can observe the change in cumulative flow for other traffic flows, e.g., opposite traffic flows (Redmill et al., 2011) or parallel traffic flows. For these reasons relative flow data is a good addition and/or alternative to the data considered in Claudel & Bayen (2008), Claudel & Bayen (2010b), Claudel & Bayen (2010a), Bhaskar et al. (2010), Van Lint & Hoogendoorn (2015) and Sun et al. (2017).

7.3 Model-based traffic state estimation in the cumulative flow plane

This section discusses the principles that can be used for the two processes that are part of a model-based estimation approach, i.e., (1) learning the models that are used in estimation based on historical data and (2) real-time estimation by assimilating real-time data and models, and why the characteristics of relative flow data are valuable for applying these principles. Prior to discussing the two processes, Section 7.3.1 explains the considered link traffic flow model (L-TFM) and Newell's method (G. F. Newell, 1993a), (G. F. Newell, 1993c), (G. F. Newell, 1993b), which is the foundation for the principles introduced in this section. Both processes take part in the cumulative flow plane. For learning the L-TFM, see Section 7.3.2, this means that we remain in the cumulative flow plane to learn the model parameters. For traffic state estimation, see Section 7.3.3, this means that we want to estimate $N(x,t)$. Section 7.3.4 discusses which spatial-temporal data characteristics are desired to apply the principles explained in Sections 7.3.2 and 7.3.3.

7.3.1 Estimation using Newell's method and a triangular fundamental diagram

Newell's method (G. F. Newell, 1993a), (G. F. Newell, 1993c), (G. F. Newell, 1993b) can be used to estimate the cumulative flow value for points in space-time. This method requires a LWR-model with a concave and continuous fundamental diagram (FD). For the sake of simplicity, we will illustrate the principles in this paper based on a triangular FD (see Figure 7.3a). However, the principles can be extended to other concave and continuous FDs, e.g., piecewise-linear FDs (triangular FD is a two-piece-linear FD), Greenshields FD and Smulders FD. In Section 7.4, we will apply the principles on data from a stochastic microscopic simulation, where drivers are not bound to a FD, let alone a triangular FD. Hence, Section 7.4 will also show the accuracy of the results if the assumptions behind the LWR-model with a triangular FD are partially violated.

The conventional LWR-model (Lighthill & Whitham, 1955), (Richards, 1956) can be re-written where the cumulative flow is the state variable (G. F. Newell, 1993a), (G. F. Newell, 1993c), (G. F. Newell, 1993b):

$$\partial_t N(x,t) - Q\left(-\partial_x N(x,t)\right) = 0 \tag{7.1}$$

where $Q(\cdot)$ is the flow-density FD of traffic flow.

The triangular FD is widely used in traffic state estimation 'because of its simplicity, theoretically preferable features and some empirical evidence' (Seo, Bayen et al., 2017). This FD can be described by three parameters: free-flow speed v^f, wave speed w and passing rate r, where r describes the relative flow observed by a moving observer that travels at w through any state on the congested branch of the FD. The FD defines the minimum (lower bound) and maximum (upper bound) ΔN between two points in space-time, which are important the variational theory presented by Daganzo (2005a). The lower and upper bounds depend on the speed of the line between these points. In some studies this line is referred to as a path, e.g., Daganzo (2005a); however, to avoid confusion with observation paths, we use the term 'wave'. The triangular FD has two characteristic wave speeds, i.e., v^f and w, which yield two restrictive waves with respectively upper bounds for ΔN of 0 and $r\Delta t$. Following G. F. Newell (1993a), G. F. Newell (1993c), G. F. Newell (1993b), we can estimate the cumulative flow in a point P given a boundary $\mathcal{B}$ over which $N_{\mathcal{B}}$ is known and from which we can draw a v^f-wave and a w-wave to P. In this special situation, the minimum of the two restrictions yields N for P. Figure 7.3b provides an example in which N is known for $\mathcal{B}$ (solid black lines) and we want to estimate N for P. The restrictive paths are obtained by drawing lines starting from $\mathcal{B}$ with the characteristics speeds. Next, the cumulative flow for P, i.e., $N(x_P,t_P)$, is estimated using:

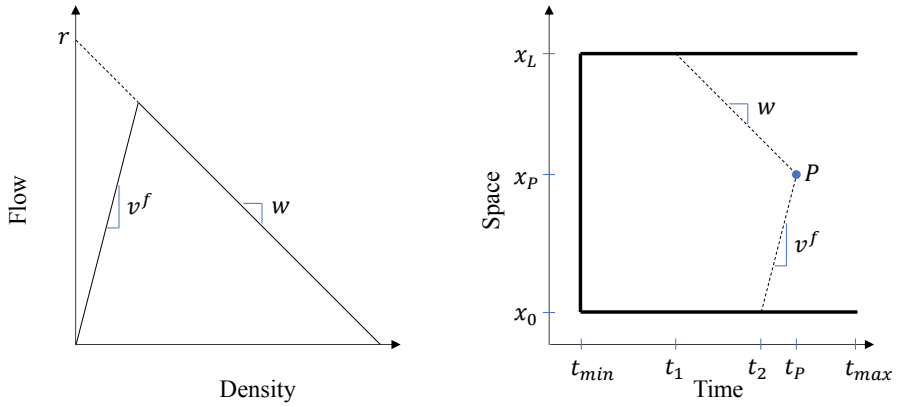
$$N(x_P,t_P) = \min\left[N(x_0,t_2), N(x_L,t_1) + r\cdot(t_P - t_1)\right] \tag{7.2}$$

This approach can be followed to estimate $N(x,t)$ for the full space-time domain if we have the initial condition (IC) and boundary conditions (BCs). The IC and BCs describe the cumulative flow curves over the solid black lines that are shown in Figure 7.3b, i.e., the IC describes $N(x,t_{\min})$ and the upstream and downstream BCs respectively describe $N(x_0,t)$ and $N(x_L,t)$.

7.3.2 Learning the link traffic flow model

The aim of this section is to explain how relative flow data can be used to learn the link traffic flow model (L-TFM) parameters. For this purpose, we explain how a learning dataset can be constructed using relative flow data that contains combinations of the change in cumulative flow and time over potential characteristic waves, see Section 7.3.2. This part yields fundamental insights that can be used to explain which spatial-temporal (relative) flow data characteristics are valuable to learn the model parameters, see Section 7.3.4. Next, in Section 7.3.2, we provide the theoretical relations between the Δt- and ΔN-observations over potential characteristics waves and the model parameters.

To learn the L-TFM parameters, an approach is considered for which we remain in the cumulative flow plane. This approach differs from the approaches followed by Dervisoglu



(a) Triangular fundamental diagram with known parameters v^f, w and r.

(b) Estimation of $N(x_P, t_P)$.

Figure 7.3: Estimation using Newell's method.

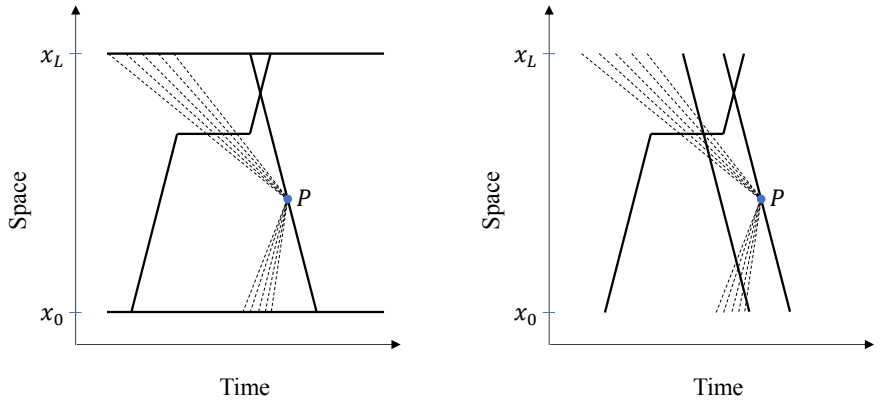
et al. (2009), Knoop & Daamen (2017) and Van Erp, Thoen et al. (2018), which use flow and density estimates, i.e., $\{q,k\}$-estimates, to learn the triangular FD. These approaches fit the free-flow and congested branch in the $\{q,k\}$-plane based on the difference between the fitted FD and the $\{q,k\}$-estimates. Although this may yield the 'best' fit in the $\{q,k\}$-plane, it may not be the most accurate L-TFM for traffic state estimation in the cumulative flow plane.

As explained in the introduction, this study does not design full methodologies as our focus lies on understanding the value of relative flow data. However, the principles explain in this section provide the basis for methodologies that can learn the model parameters based on relative flow data. It is left to future research to design and test such full methodologies and algorithms.

Construct a learning dataset that contains cumulative flow and time over potential characteristic waves

As explained in Section 7.3.1, a L-TFM can be used to estimate the cumulative flow N for a point P in space-time given that we know N over a boundary $\mathcal{B}$, where all potential restrictive paths can be drawn between $\mathcal{B}$ and P. Following earlier assumptions, we define the model structure of the L-TFM that we want to learn: a LWR-model with a triangular fundamental diagram (FD), see Figure 7.3a. In this case, this model has three unknown parameters, i.e., the passing rate r, free-flow speed v^f and wave speed w.

To learn the L-TFM, we want to obtain a learning dataset that describes the change in cumulative flow and time over all potential restrictive waves for points in space-time. Here, all potential restrictive waves are straight lines in space-time that travel with a certain speed, i.e., the potential values of v^f and w. Stationary and moving observers can yield a set of observations paths that allow us to construct a learning dataset that combines ΔN and Δt over potential characteristic waves. If observation paths intersect, the ΔN-observations



(a) Example that combines stationary and moving observers.

(b) Example that relies on moving observers alone.

Figure 7.4: Examples of sets of observation paths (solid lines) that allow us to draw potential backwards (in time) propagating waves from a point P (dashed lines) and obtain combinations of the change in cumulative flow and time over the different waves.

can be related to each other (Van Erp, Knoop & Hoogendoorn, 2018b). In this way, we can obtain ΔN between two points that lie on the same or different observation paths. Depending on the spatial-temporal characteristics of individual observation paths, we can have a set of intersecting observation paths for which we can draw backwards (in time) propagating waves from a point P that all intersect with an observation path that is part of the set. Figures 7.4a and 7.4b illustrate two situations in which it is possible to find ΔN and Δt for the potential values of v^c, where v^c is used to describe the speed of a potential characteristic wave. In this figure, we only show a selection of potential values of v^c around v^f and w. However, it is possible to draw waves for other values of v^c.

Learn traffic flow model parameters

This section explains the theoretical relations between the learning dataset, i.e., combinations of ΔN, Δt and v^c, and the L-TFM model parameters. Below, we first explain which values of ΔN and Δt are expected given that the traffic flow follows a known triangular FD. Next, we describe the rate errors, i.e., Υ, over the characteristic waves that are induced by error in the model parameters.

The learning dataset contains the information that is needed to learn the traffic flow model parameters, i.e., v^f, w and r, following Newell's method. For this purpose, equation (7.2) can be rewritten to:

$$0 = \min\left[\Delta N^{\text{ff}}_{P\mathcal{B}}, \Delta N^{\text{cg}}_{P\mathcal{B}} - r\Delta t^{\text{cg}}_{P\mathcal{B}}\right] \tag{7.3}$$

Note that this relation includes the change in cumulative flow and time from P to $\mathcal{B}$. Let us define $\mathcal{B}^{\text{cg}}$ as the point-of-intersection of the w-wave and $\mathcal{B}$, which means that $\Delta N^{\text{cg}}_{P\mathcal{B}} = N(x_{\mathcal{B}^{\text{cg}}}, t_{\mathcal{B}^{\text{cg}}}) - N(x_P, t_P)$ and $\Delta t^{\text{cg}}_{P\mathcal{B}} = t_{\mathcal{B}^{\text{cg}}} - t_P$.

This function can be rewritten to describe the rates that are observed instead of the counts, i.e., $\Delta N/\Delta t$ instead of ΔN:

$$0 = \max\left[\frac{\Delta N^{\text{ff}}_{P\mathcal{B}}}{\Delta t^{\text{ff}}_{P\mathcal{B}}}, \frac{\Delta N^{\text{cg}}_{P\mathcal{B}}}{\Delta t^{\text{cg}}_{P\mathcal{B}}} - r\right] \tag{7.4}$$

In the remainder of this section, we consider the second relation, i.e., equation (7.4). This relation can be interpreted as follows: The cumulative flow in point P is restricted by the v^f-wave, the w-wave or both (at capacity). We use the variable R to describe the restrictive wave, where $R = 1$, $R = 2$ and $R = 3$ respectively denote that the v^f-wave, w-wave or both waves are restrictive. Therefore, for each P the restrictive wave is given by R_P:

$$\begin{aligned}
R_P = 1 \quad &\text{if } \tfrac{\Delta N^{\text{ff}}_{P\mathcal{B}}}{\Delta t^{\text{ff}}_{P\mathcal{B}}} = 0 \text{ and } \tfrac{\Delta N^{\text{cg}}_{P\mathcal{B}}}{\Delta t^{\text{cg}}_{P\mathcal{B}}} - r < 0,\\
R_P = 2 \quad &\text{if } \tfrac{\Delta N^{\text{ff}}_{P\mathcal{B}}}{\Delta t^{\text{ff}}_{P\mathcal{B}}} < 0 \text{ and } \tfrac{\Delta N^{\text{cg}}_{P\mathcal{B}}}{\Delta t^{\text{cg}}_{P\mathcal{B}}} - r = 0,\\
R_P = 3 \quad &\text{if } \tfrac{\Delta N^{\text{ff}}_{P\mathcal{B}}}{\Delta t^{\text{ff}}_{P\mathcal{B}}} = 0 \text{ and } \tfrac{\Delta N^{\text{cg}}_{P\mathcal{B}}}{\Delta t^{\text{cg}}_{P\mathcal{B}}} - r = 0.
\end{aligned} \tag{7.5}$$

Equation (7.4) holds if traffic flow follows a triangular FD and the model parameters are known or correctly estimated. However, without having the correct parameter values, it becomes more challenging to find the restrictive wave.

Let us explore the influence of error in the model parameters on the two elements of equation (7.4), i.e., $\Delta N^{\text{ff}}_{P\mathcal{B}}/\Delta t^{\text{ff}}_{P\mathcal{B}}$ and $\Delta N^{\text{cg}}_{P\mathcal{B}}/\Delta t^{\text{cg}}_{P\mathcal{B}} - r$. In the remainder of this work, $\hat{v^f}$, $\hat{w}$ and $\hat{r}$ are used to denote the (potentially incorrect) parameter-guesses. Furthermore, ε is used to denote the parameter errors, where the error in v^f is given by $\varepsilon_{v^f} = v^f - \hat{v^f}$. The observations related to the v^f-wave, i.e., $\Delta N^{\text{ff}}_{P\mathcal{B}}$ and $\Delta t^{\text{ff}}_{P\mathcal{B}}$, are affected by ε_{v^f}, while the observations related to the w-wave, i.e., $\Delta N^{\text{cg}}_{P\mathcal{B}}$ and $\Delta t^{\text{cg}}_{P\mathcal{B}}$, are affected by ε_w. The errors related to the two elements of equation (7.4) are denoted by Υ, where Υ^{cg} describes the error related to $\Delta N^{\text{cg}}_{P\mathcal{B}}/\Delta t^{\text{cg}}_{P\mathcal{B}} - r$. The parameter errors ε_{v^f} and ε_w influence which backwards propagating waves are drawn from point P. The errors Υ^{ff} and Υ^{cg} can be quantified if the mean state through which this wave travels still lies on the restrictive traffic state, e.g., Υ^{cg} can be quantified if the $\hat{w}$-wave still fully travels through congested (including capacity) states:

$$\Upsilon^{\text{ff}} = \varepsilon_{v^f}\bar{k} \tag{7.6}$$

$$\Upsilon^{\text{cg}} = \varepsilon_w\bar{k} + \varepsilon_r \tag{7.7}$$

These equations provide insight in relations between $\Delta N_{P\mathcal{B}}/\Delta t_{P\mathcal{B}}$ for different potential values of the characteristics wave speeds, i.e., $\hat{v^f}$ and $\hat{w}$. Figure 7.5 provides a visual interpretation of the expected relations given that traffic flow follows a triangular FD for points P that are restricted by the w-wave, i.e., the black line (jam state) and the orange line (queue

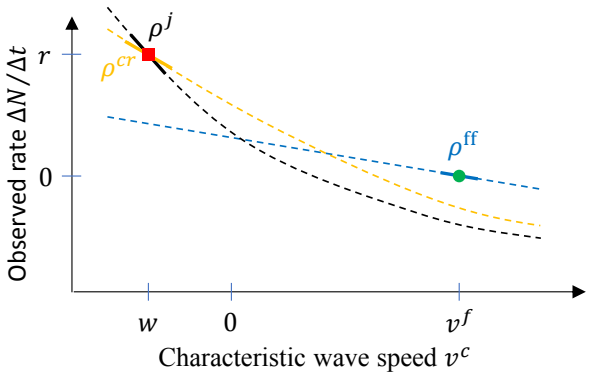


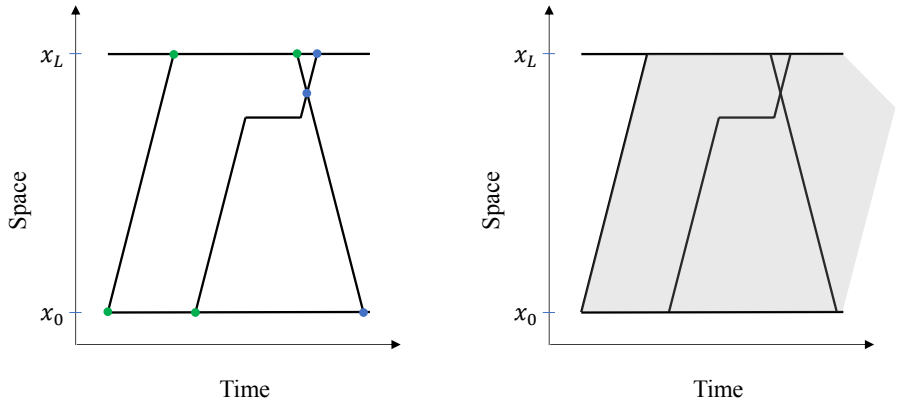
Figure 7.5: Expected relations between the observed rates dependent and the considered characteristic wave speed given that traffic flow follows a triangular FD for points that are restricted by the w-wave (black and orange lines) and v^f-wave (blue line).

discharge state), and that are restricted by the v^f-wave, i.e., the blue line. The exact shapes of the lines also depend on the non-restrictive traffic conditions and can therefore differ. However, a line related to a point that is restricted by the w-wave should cross (r,w), which is indicated by the red square, and a line related to a point that is restricted by the v^f-wave should cross (0,v^f), which is indicated by the green dot. As every P is restricted by one or both waves, see equation (7.5), every line should intersect with one or both characteristic points, i.e., the red square and/or green dot. Furthermore, if P is not restricted by the characteristic wave related to that characteristic point, it should lie below the characteristic point. As depicted by equations (7.6) and (7.7), the slope of the lines depends on the average density. The wave traveling at the wave speed that is restrictive for a point P travels through a single state, e.g., the jam state where $k = k^j$. Therefore, the slope of the lines in the related characteristic point depends on the density at P. This means that slopes of the black, orange and blue lines will be equal to the jam density k^j, the critical density k^{cr} and a free-flow density k^{ff} at the characteristic point that relates to their restrictive wave. These slopes and densities are indicated in Figure 7.5.

7.3.3 Traffic state estimation using relative flow data and a link traffic flow model

This section explains how relative flow data combined with the L-TFM can be used to (partially) estimate the traffic state for different data availability scenarios. This allows us to evaluate the potential value of relative flow data. The difference in these scenarios lies in the inclusion of stationary observers positioned at the link boundaries.

Figure 7.6 shows a data scenario in which all observation paths can be related to each other through intersection. In this scenario, stationary observers positioned at the link boundaries are combined with a set of moving observers. The dots shown in Figure 7.6a are used to illustrate the intersection points between the observation paths. The green dots illustrate the initial intersection points of the observers that can be used to define the cumulative flow over all observation paths in one framework. As explained in Van Erp, Knoop



(a) Observation paths and intersection points.

(b) Space-time domain for which $N(x,t)$ can be estimated using Newell's method.

Figure 7.6: Data scenario that combines stationary observers positioned at the link boundaries and a set of moving observers.

& Hoogendoorn (2018b), this is also a favorable scenario to mitigate the cumulative error problem (CEP), which is also referred to as the cumulative drift problem (Van Lint & Hoogendoorn, 2015), due to the intersection of already initialized observation paths (indicated by the blue dots in Figure 7.6a). As all observation paths are connected, Newell's method can be applied to estimate the cumulative flow for points that do not lie on an observation path. The space-time domain for which $N(x,t)$ can be estimated using this method is indicated with a grey area in Figure 7.6b.

In the remainder of this section, we will consider a more restrictive data scenario that solely includes moving observers. Compared to the data configuration shown in Figure 7.6, no stationary observers are positioned at the link boundaries. By removing the stationary observers at the link boundaries, the observation paths are no longer connected at these boundaries and there are no observation paths that fully cover the link boundaries. As a result, the boundary conditions (BCs) are not obtained and the remaining observation paths can be (partially) unconnected.

However, by combining these observation paths with information in the form of traffic flow models, we may obtain information on the link BCs and connect observation paths. The principles that can be used for this purpose are discussed below. First, Section 7.3.3 discusses which information remains without the availability of stationary observers on an individual link. Next, we explain why it is beneficial to estimate the traffic state in a network of links and nodes, where information can propagate over a node, i.e., from link to link (Section 7.3.3).

The principles explained below can be used to design a full methodology that assimilates relative flow data and the traffic flow model to estimate the traffic state on a single or set of links. Such a to-be-developed full methodology should be able to exploit the valuable interactions of information (i.e., data and models) in space-time and thereby estimate 'the

most probable state'. For this purpose, different assimilation techniques can be considered (Seo, Bayen et al., 2017).

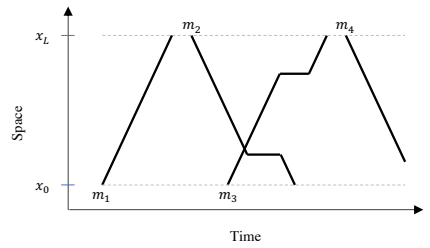
Information propagation on an individual link when relying on moving observers only

This section explains how relative flow data and the link traffic flow model (L-TFM) can be combined to connect observation paths and estimate the BCs. Furthermore, we explain how the demand D (downstream boundary) and supply S (upstream boundary) can be estimated, which are valuable for information propagation in a network, see Section 7.3.3. To explain which information is available on an individual link, we often refer to Figure 7.7. This figure shows the observation paths of four moving observers (solid black lines) that are numbered at the point they enter the link in Figure 7.7a. Again, a LWR-model with a known triangular FD is considered.

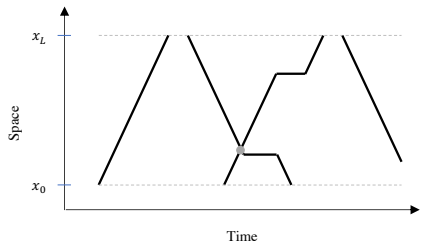
Intersecting observation paths can form sets that allow us to describe the relation in ΔN between certain points on the link boundaries (see Figure 7.7b). A moving observer that crossed both boundaries describes $N(x_0, T_m(x_0)) - N(x_L, T_m(x_L))$, where $T_m(x_0)$ and $T_m(x_L)$ respectively denote the times at which moving observer m crossed the upstream and downstream link boundary. Furthermore, if observation paths intersect, more points on the boundaries can be related to each other. Out of the four observation paths shown in Figure 7.7b only those of moving observers m_2 and m_3 intersect. The cumulative flow over these paths can thus be related to each other based on their intersection point.

From one set of connected observation paths, using Newell's method we may be able to estimate N for points that lie observation paths that are not connected via intersection (see Figure 7.7c). For instance, for the case shown in Figure 7.7c, we can estimate $N(x_P, t_P)$ in the N-framework used for moving observers m_2 and m_3. As point P lies on the observation path of moving observer m_4, the cumulative flow over that observation path can be described in the same framework. However, we still miss information to describe the observation path of moving observer m_1 in the same framework. With this information the link outflow between $T_{m_1}(x_L)$ and $T_{m_2}(x_L)$ and link inflow between $T_{m_1}(x_0)$ and $T_{m_3}(x_0)$ remain unknown. However, the two unknowns are related, i.e., if information related to $N(x_L, T_{m_2}(x_L)) - N(x_L, T_{m_1}(x_L))$ is obtained we can describe $N(x_0, T_{m_3}(x_0)) - N(x_0, T_{m_1}(x_0))$.

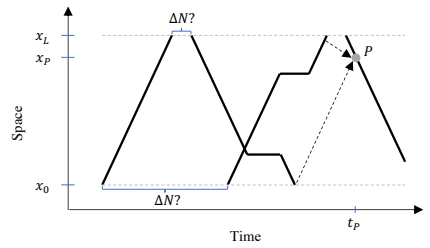
In the case of a triangular FD (as is assumed here), Newell's method relies on two forward (in time) propagating waves, i.e., one traveling at free-flow speed v^f and one traveling at wave speed w. The FD describes the maximum change in N as a function of time over these waves, i.e., 0 for the v^f-wave and $r\Delta t$ for the w-wave. Combined with the knowledge that one of these two upper bounds is restrictive, we can estimate N for the point where these two waves intersect. Newell's method (in this form) cannot be used to estimate N for a point in space-time on the link boundary as only one of the two time-forward propagating waves intersects with the link boundary, i.e., v^f-wave and w-wave respectively intersect with the downstream and upstream boundary. However, the individual waves do yield restrictions on ΔN, where forward and backward (in time) propagating waves respectively yield an upper and lower bound for ΔN. Figure 7.7d shows a point (grey dot) in space-time together with two backwards (blue dashed arrow lines) and forward (red dashed arrow lines) propagating waves. Given that we know N for the point indicated by the grey dot, the waves provide upper and lower bounds for points on the link boundaries. We focus on these four paths as they play an important role in the remainder of this section. However, we can find upper



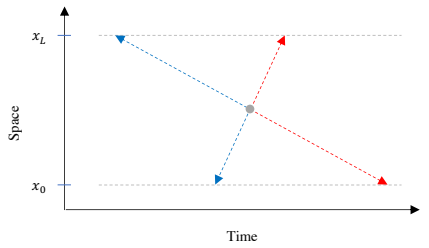
(a) Space-time diagram showing the observation paths of four moving observers (solid black lines).



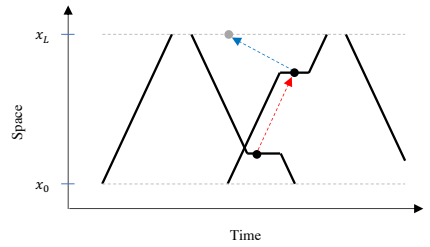
(b) Connecting observation paths via intersection.



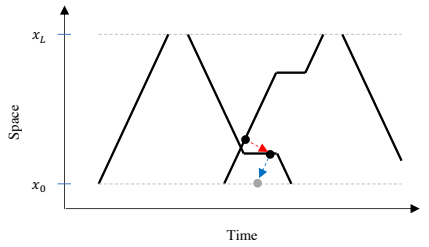
(c) Connecting observation paths via L-TFM estimates.



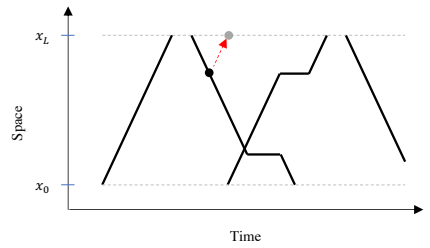
(d) Restrictive paths from a point in space-time to the link boundaries.



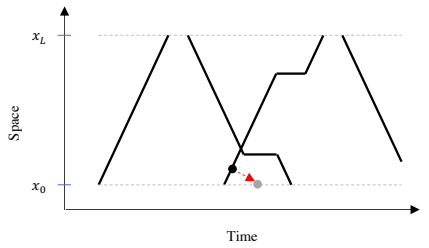
(e) Estimate grey point on $N(x_L,t)$.



(f) Estimate grey point on $N(x_0,t)$.

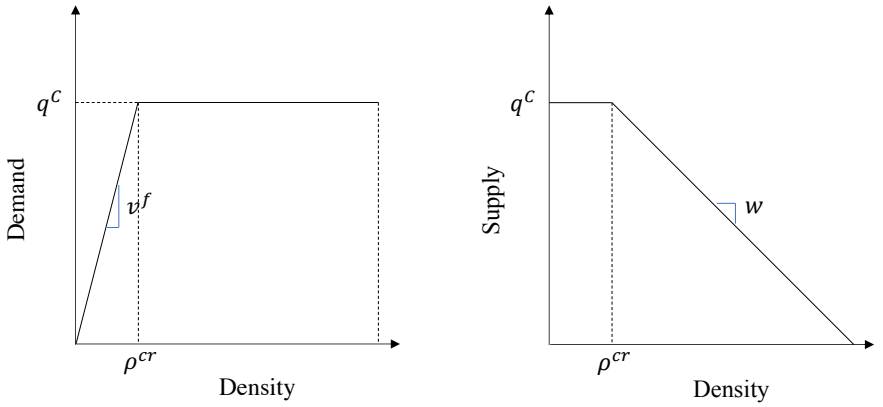


(g) Find direct upper restriction for grey point on $\tilde{N}^+(x_L,t)$.



(h) Find direct upper restriction for grey point on $\tilde{N}^+(x_0,t)$.

Figure 7.7: Example of information that is available on an individual link.



(a) Demand as a function of density at the downstream link boundary.

(b) Supply as a function of density at the upstream link boundary.

Figure 7.8: Demand and supply curves.

and lower bounds for ΔN for any two points in space-time. Therefore, if we have a set of observation paths for which N is initialized (i.e., N is defined in the same framework for all paths that are part of the set), we can use all point that lie on these observation paths to find upper and lower bounds for $N(x_0,t)$ and $N(x_L,t)$.

By exposing the restrictive wave we can (partially) estimate the boundary conditions. As stated above, we cannot reach a point on the link boundaries with a forward propagating v^f-wave and w-wave that start from a point on the link. However, in some cases, we can find that a forward propagating wave originating from the boundary is restrictive. The forward propagating waves from the upstream and downstream boundaries are respectively the v^f-wave and w-wave. A forward propagating wave starting from a point on the link boundary (with unknown N) can intersect with the other forward propagating wave starting from a point and in a point for which N is known. If the forward propagating wave between the two points with known N is not restrictive, the forward propagating wave from the link boundary has to be restrictive. This information allows us to estimate N at the link boundaries. To explain this we consider the examples shown in Figures 7.7e and 7.7f, which respectively relate to estimating N for a point on the downstream and upstream boundary. In these figures the two black dots indicate points in space-time for which we known ΔN (as they both lie on a set of connected observation paths). If this wave (indicated by the red dashed arrow line) is not restrictive, i.e., $\Delta N < 0$ for Figure 7.7e and $\Delta N < r\Delta t$ for Figure 7.7f, the wave from the link boundary has to be restrictive. Therefore, we can use a backwards propagating wave to the link boundary (indicated by the blue dashed arrow line) to estimate N on the link boundary.

Supply S and demand D can be partially estimated based on the BC estimates and direct upper bound restrictions of the boundary conditions (see Figures 7.7g and 7.7h). Direct upper bound restrictions are obtained using by drawing forward propagation waves with the speeds v^f and w, i.e., the red dashed arrow lines in Figure 7.7d, from points on the

observation paths to the link boundaries. We use the notations $\tilde{N}^+(x_0,t)$ and $\tilde{N}^+(x_L,t)$ are used for these upper bound restrictions, which are respectively obtained using w-waves and v^f-waves. Demand and supply depend on the density at the link boundaries, see Figures 7.8a and 7.8b. The critical density k^{cr} plays an important role in estimating D and S. As can be seen in Figures 7.8a and 7.8b, D and S are respectively equal to the capacity q^C in over-critical ($k \geq k^{cr}$) and under-critical ($k \leq k^{cr}$) densities. Furthermore, D and S are respectively described by the free-flow and congested FD branch in under-critical ($k \leq k^{cr}$) and over-critical ($k \geq k^{cr}$) densities. We include $k = k^{cr}$ both in under- and over-critical densities as D and S are equal to q^C and are correctly described by the FD branches in this point. By comparing the link BC estimates and direct upper bound restrictions, we can find whether the density at the link boundary is under-critical or over-critical. For the upstream boundary, $\tilde{N}^+(x_0,t) > N(x_0,t)$ and $\tilde{N}^+(x_0,t) = N(x_0,t)$ respectively describe under-critical and over-critical densities. For the upstream boundary, $\tilde{N}^+(x_L,t) > N(x_L,t)$ and $\tilde{N}^+(x_L,t) = N(x_L,t)$ respectively describe over-critical and under-critical densities. For the cases in which the link boundary condition estimates and direct upper bound restrictions are equal, D and S are equal to respectively the outflow and inflow, i.e., $\partial_t N(x_L,t)$ and $\partial_t N(x_0,t)$.

In some cases, we may not be able to determine the boundary conditions, while we can determine whether the link boundary density is under- or over-critical. If we known that the upstream and downstream boundary densities are respectively under-critical and over-critical, we known that supply and demand are equal to the capacity q^C. As an example, we consider a moving observer that travels in the opposing flow direction with free-flow speed. Starting on this observation path, we can draw forward propagating v^f-waves and w-waves that allow us to estimate ΔN with respect to the observation path for an area in space-time. If a v^f-wave is not the restrictive wave for at least one point that it crosses, $D = q^C$ at the intersection of this v^f-wave and the downstream boundary. Similarly, if a w-wave is not the restrictive wave for at least one point that it crosses, $S = q^C$ at the intersection of this w-wave and the upstream boundary.

The techniques discussed in this section allow us to partially determine the boundary conditions, demand and supply, based on the information that is available on an individual link. Below, we explain why it is beneficial to use information that is available in a network of links and nodes to estimate the traffic state. Furthermore, we explain why the link boundary estimates from one link can help to estimate the boundary conditions on an other, connected link.

Information within a network of links and nodes

As explained above, the information available on individual links may yield estimates of the link inflow, outflow, demand and supply. However, the moving observers combined with the L-TFM may not provide sufficient information to estimate the full link boundary conditions. To overcome this problem, it is valuable to add information that is available for other parts of the road network.

The road network can be represented using links and nodes, where the nodes connect multiple in-flowing and out-flowing links. There are many different types of nodes in the road traffic network, e.g., inhomogeneous nodes, merge nodes, diverge nodes, cross nodes (Yperman, 2007). In general, motorway intersection nodes are less complex than urban intersection nodes. For instance, motorway nodes will often connect less links and there is no

turn delay. The latter means that a vehicle that exits an in-flowing link will instantaneously enter an out-flowing link.

The conservation-of-vehicles condition can be used for information propagation over nodes. Especially, for simple nodes (e.g., those that are used to describe a lane drop, on-ramp or off-ramp) this condition can suffice to (partially) estimate the flow over the node and in this way estimate the link boundary conditions from the node. For instance, in case of a lane drop (i.e., an inhomogeneous node), we may be able to estimate the outflow of the in-flowing link during a period, while this is not possible for the inflow of the out-flowing link. By applying the conservation-of-vehicles conditions, the inflow of the out-flowing link can be estimated.

However, additional information may be added in the form of node traffic flow models (N-TFMs), which can be used to describe the traffic behavior over a node. These models can describe the flow over a node as a function of the demand and supply. As explained in Section 7.3.3, relative flow data combined with a L-TFM can provide (partial) estimates for the link boundary conditions, i.e., $N(x_0,t)$ and $N(x_L,t)$. Furthermore, the demand D (which relates to the downstream link boundary) and supply S (which relates to the upstream link boundary) can be (partially) estimated. Partial information (i.e., if we are able to estimate a part of the flows, demands and supplies) in combination with a N-TFM may suffice to estimate the other variables (i.e., the flows, demands and/or supplies that we could not estimate on the individual links). Furthermore, for some periods we may be able to estimate all variables (potentially by involving the conservation-of-vehicles conditions). This yields a learning dataset that combines the dependent and explanatory variables of a N-TFM, which therefore may be used to learn the N-TFM.

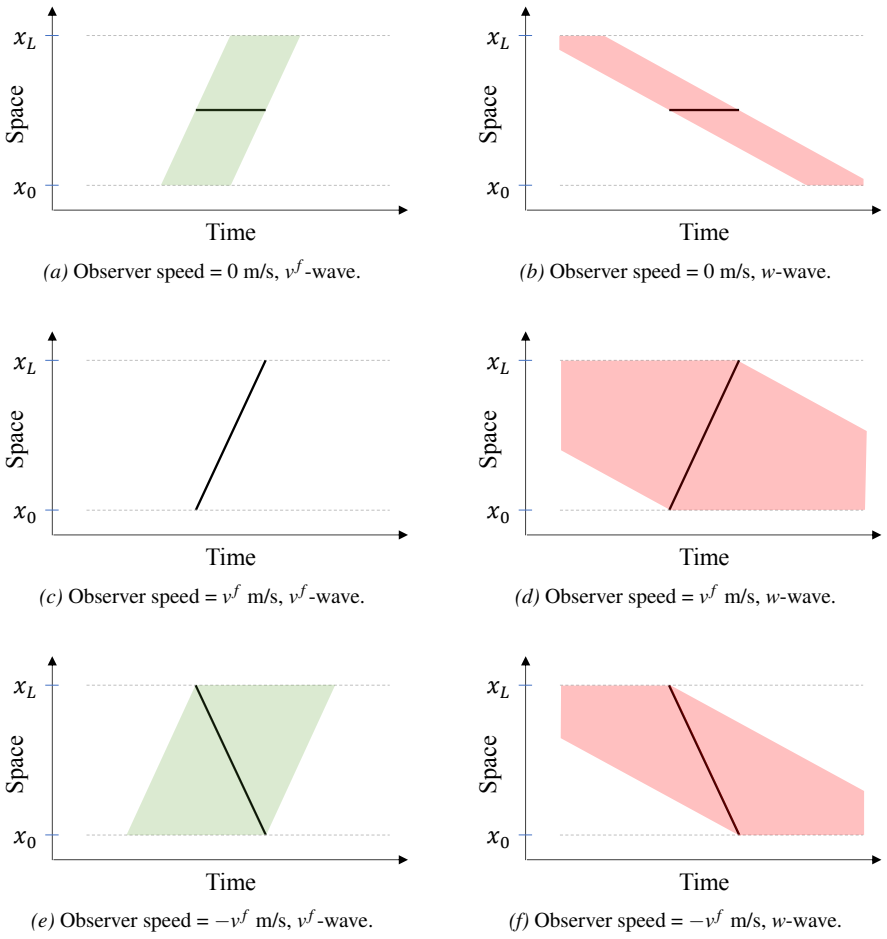
Learning N-TFMs and estimating the traffic state in a network of links and nodes is left for future research. It is important to understand that estimation in a network can be valuable to overcome the limitation of not observing the link boundaries using stationary observers, and that estimates from individual links provide the information valuable to apply and learn N-TFMs.

7.3.4 Desirable spatial-temporal data characteristics

The spatial-temporal characteristics of the relative flow data affect the intersections of observation paths and the space-time areas for which it is possible to draw forward or backwards propagating characteristics waves from the observation paths. Both play an important role in the two processes that were discussed in Sections 7.3.2 and 7.3.3.

Observation paths only intersect if they have different speeds. A set of moving observers that travel at the same speed or a set of stationary observers (which all travel at 0 m/s) will not intersect. Therefore, combining stationary observers with moving observers or moving observers that observe both their own and opposite travel direction is beneficial to learn traffic flow models.

Furthermore, the area in space-time for which it is possible to draw forward or backwards propagating characteristics waves depends on the observer speed, see Figure 7.9. In this figure, we consider three observers that travel at 0, v^f and $-v^f$ m/s, and which collect relative flow data for the same time period. These three observers illustrate a stationary observer (or moving observer at standstill), and moving observers traveling freely in the



(a) Observer speed = 0 m/s, v^f-wave.

(b) Observer speed = 0 m/s, w-wave.

(c) Observer speed = v^f m/s, v^f-wave.

(d) Observer speed = v^f m/s, w-wave.

(e) Observer speed = $-v^f$ m/s, v^f-wave.

(f) Observer speed = $-v^f$ m/s, w-wave.

Figure 7.9: Visualization of the spatial-temporal area for which forward or backwards propagating characteristic waves can be drawn.

direction of interest and in opposite direction. The figures show that an increase in the absolute difference between the observer and characteristics wave speeds yields an increase in the spatial-temporal area for which the characteristic wave can be drawn.

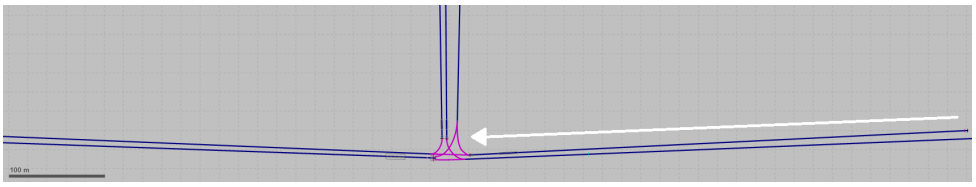
For a certain set of observers it may be possible to (partially) estimate the traffic state while it does not provide observations for the learning dataset. The link traffic flow model can be learned based on historical data. This means that we can rely on relative flow data that is collected over a longer period, e.g., months or years. Therefore, is does not have to be a problem that real-time data provide insufficient observations to learn the link traffic flow model.

To obtain observations for the link traffic flow model learning dataset, we need to be able to draw both backwards (in time) propagating waves for the potential values of the characteristic wave speeds between points that lie on observation paths that are connected

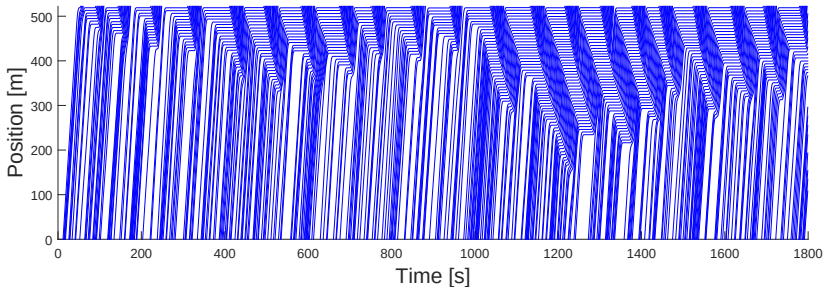
through intersection. In estimating the traffic state, we work with a single value instead of a potential range for the characteristic wave speeds. Furthermore, each principle presented in Section 7.3.3 puts less requirements on the data than obtaining observations for the learning dataset does: (1) To connect observation paths using Newell's method, we need to draw both backwards moving characteristic waves from a point P to points that lie on the same or connected observation paths, but do not have to be connected to the observation path of point P. (2) To judge whether a characteristic wave is not restrictive for a point on an observation path (which is the basis for estimating the boundary conditions), we need to draw one backwards moving characteristic wave to a point on a connected observation path. And, (3) To estimate supply and demand, we need to draw one forward propagating characteristic wave to points on the link boundary conditions or intermediate points that can be estimated.

7.4 Testing the principles using simulated data

In this section, the principles explained in Sections 7.3.2 and 7.3.3 are tested on simulated data that is collected using the microscopic simulation program VISSIM, which is empirically validated (Fellendorf & Vortisch, 2001). Figure 7.10a shows the road network for which traffic is simulated. To test the principles, we consider a single one-lane link, which is highlighted with a white arrow in Figure 7.10a. The outflow of this link is constricted by a traffic light, which causes queues to build up on the link. Figure 7.10b shows the trajectories on the considered link during the 1800 s simulation period.



(a) Road network for which traffic is simulated.



(b) Vehicle trajectories during the 1800 s simulation period.

Figure 7.10: VISSIM road network and simulated vehicle trajectories.

Different examples are considered to understand and to test the principles and thereby

show how relative flow data can be used in model-based estimation. In these examples, it is assumed that a set of moving observers, which are part of the observed link and opposing link, collect disaggregated relative flow data. It is thus assumed that all individual passings are observed by the moving observers, which allows us to find the change in cumulative flow between any two points that lie on the same or different intersecting observation paths.

The principles used to extract the valuable information from the relative flow data depend on the process, i.e., learning the link traffic flow model parameters (see Section 7.4.1) or estimating the traffic state (see Section 7.4.2). Therefore, each section starts with an explanation of the steps that will be taken in that section.

7.4.1 Learning the link traffic flow model parameters

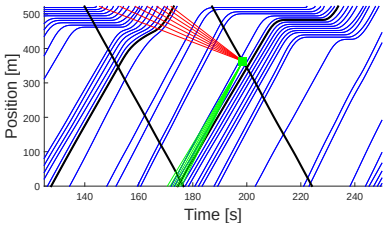
To learn the link traffic flow model parameters the following steps are taken: (1) Define the points P. (2) Define the potential characteristics wave speeds and draw the backwards propagating potential characteristic waves. (3) Determine $\Delta N_{P\mathcal{B}}$ and $\Delta t_{P\mathcal{B}}$ for each combination of P and potential characteristics wave speeds. (4) Visualize and interpret the observed rates, i.e., $\Delta N_{P\mathcal{B}}/\Delta t_{P\mathcal{B}}$. And, (5) Approximate the model parameters free-flow speed v^f, wave speed w and passing rate r. In the last step, we approximate the model parameters based on the considered examples. These approximated model parameters will be used to estimate the traffic state in Section 7.4.2.

Figures 7.11a to 7.11d show space-time plots in which the moving observer observation paths (thick black lines) and other vehicle trajectories (thin blue lines) are visualized. Six points P that lie on an observation path of a moving observer traveling on the opposing road are considered, which are indicated with six different shapes and colors. From the trajectory plots, we know that the restrictive wave for the green square and the cyan hexagram (see Figures 7.11a and 7.11d) is the v^f-wave, while the others (see Figure 7.11b, 7.11c and 7.11d) are restricted by the w-wave.

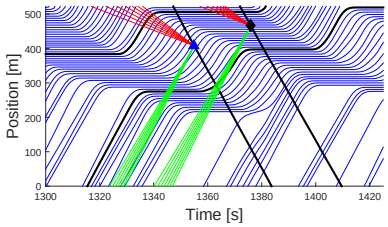
Backwards traveling potential v^f-waves and w-waves can be drawn that originate from the points P. In our example, seven potential values are considered for both characteristic waves. The potential values selected for the $\hat{v^f}$-wave are 13.0, 13.5, 14.0, 14.5, 15.0, 15.5 and 16.0 m/s and the potential values selected for the $\hat{w}$-wave are -3.0, -3.5, -4.0, -4.5, -5.0, -5.5 and -6.0 m/s. These are visualized in Figures 7.11a to 7.11d using green ($\hat{v^f}$-wave) and red ($\hat{w}$-wave) lines.

In these examples, all potential characteristic waves intersect with the observation paths of connected moving observers, which allows us to determine $\Delta N_{P\mathcal{B}}$ and $\Delta t_{P\mathcal{B}}$ for each wave. Figures 7.11e and 7.11f show the observed rates, i.e., $\Delta N_{P\mathcal{B}}/\Delta t_{P\mathcal{B}}$, for the potential characteristic wave speeds. Here, the shape and color describes the point P that relates to the $\Delta N_{P\mathcal{B}}/\Delta t_{P\mathcal{B}}$-observation.

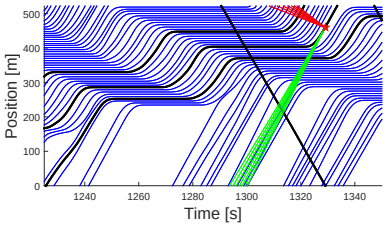
The figures provide three main insights: (1) The restrictive characteristic wave is clear for each point. The points for which the v^f-wave is restrictive, i.e., the green square and the cyan hexagram, show near-zero values of $\Delta N_{P\mathcal{B}}/\Delta t_{P\mathcal{B}}$ in Figure 7.11e, while the rates of the potential v^f-waves for the other points show clear negative values. In Figure 7.11f the smallest $\Delta N_{P\mathcal{B}}/\Delta t_{P\mathcal{B}}$-value is observed for the points for which the v^f-wave is restrictive. (2) The spread in the rates observed by the different potential characteristic wave speeds is proportional to the density. For instance, the spread in the rates related to the considered $\hat{w}$-wave speeds is larger for waves that travel through jam density (i.e., the black diamond



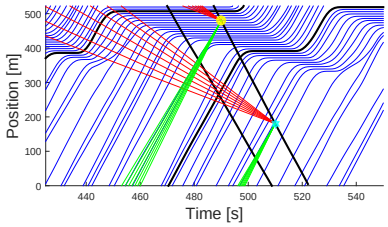
(a) Space-time plot with green square.



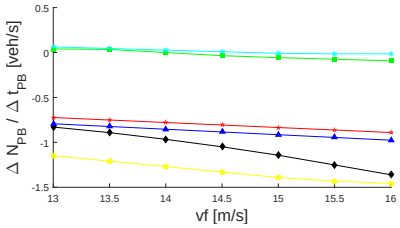
(b) Space-time plot with blue triangle and black diamond.



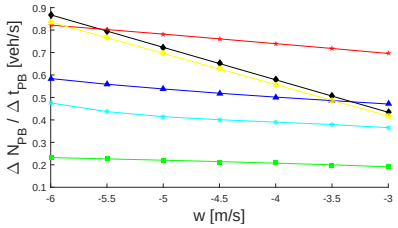
(c) Space-time plot with red pentagram.



(d) Space-time plot with yellow circle and cyan hexagram.



(e) Plot of observed rates and potential values selected for the $\hat{v^f}$-wave.



(f) Plot of observed rates and potential values selected for the $\hat{w}$-wave.

Figure 7.11: Examples that are considered to learn the link traffic flow model parameters.

and the yellow circle) than for those that travel through a queue discharge state (i.e., the blue triangle and the red pentagram), see Figure 7.11f. (3) There seem to be two different queue discharge states, which occur dependent on whether or not vehicles are able to exit the link within that green period. This difference can be observed in Figure 7.11f, where the $\Delta N_{P\mathcal{B}}/\Delta t_{P\mathcal{B}}$-value indicated by the red pentagram are higher than those indicated by the blue triangle. When comparing the space-time plots related to these points, we observe that the waves related to the red pentagram cross the trajectories of vehicle that exit the link in that green period. These plots also show that these vehicles have a smaller headway than the vehicles that need to wait at least one more red period. It seems that the vehicles 'anticipate' that they will not be able to catch the green light and therefore accelerate slower.

The first two insights are in line with the principles explained in Section 7.3; however, the third insight indicates that Newell's method in combination with a triangular FD will not yield perfect estimates. As explained in Section 7.3.2: If all points that are restricted

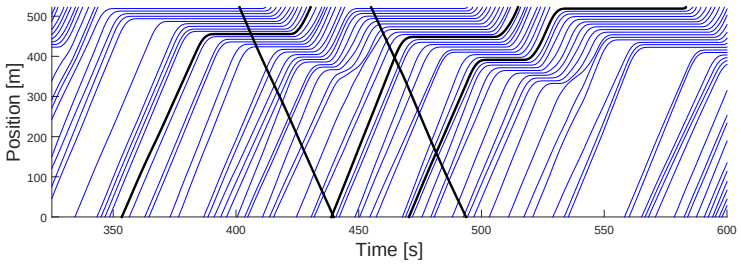
by the w-wave would lie on the congested branch of a triangular FD, the lines related these points would all intersect in the same point in Figure 7.11f. In the ($\Delta N_{P\mathcal{B}}/\Delta t_{P\mathcal{B}}$,$\hat{w}$)-plane this intersection point would lie at (r,w). Instead, we approximate two intersection points in Figure 7.11f, i.e., (0.815 veh/s,-5.70 m/s) and (0.485 veh/s,-3.40 m/s), which respectively include the queue discharge states related to vehicles that exit the link within that green period and related to the vehicles that have to wait at least one more red period. Therefore, we will respectively refer to these parameter sets as 'queue-to-queue congested wave parameters' and 'link outflow congested wave parameters'. In Section 7.4.2 both combinations of $\hat{w}$ and $\hat{r}$ will be used to estimate the traffic state. This allows us to discuss the effect on the estimation accuracy of selecting one of the two combination. Furthermore, for traffic state estimation, we need to define the free-flow speed. Therefore, based on Figure 7.11e, $\hat{v^f}$ is set to 14.5 m/s because the $\Delta N_{P\mathcal{B}}/\Delta t_{P\mathcal{B}}$ for the green square and cyan hexagram are closest to zero for this value.

7.4.2 Estimating the boundary conditions based on moving observers alone

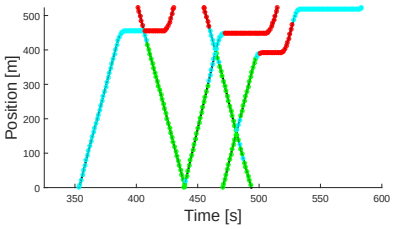
The principles explained in Section 7.3.3 are applied to estimate the traffic state. In this case study, the focus lies on partially estimating the link boundary conditions based on moving observers alone. For this purpose the following two steps are taken: (1) finding the restrictive wave for points along the observation paths and (2) estimating the link boundary conditions, i.e., $N(x_0,t)$ and $N(x_L,t)$. Figure 7.12a shows the example case that is considered in this section. Again, the moving observers and the other vehicle trajectories are respectively plotted with thick black and thin blue lines. In this case, all observation paths can be related to each other through intersection.

The restrictive wave for points along the observation paths are found by excluding that the other wave is restrictive. For this purpose, characteristics waves are drawn between points of connected observation paths. Given that the change in cumulative flow and time are known for waves drawn between points of connected observation paths, it is possible to judge whether these waves are not restrictive for the point furthest in time. As traffic behavior has a stochastic component, we use a threshold value to judge whether a wave is not restrictive. A v^f-wave and w-wave are respectively said to be non-restrictive if $\Delta N^{\text{ff}}_{P\mathcal{B}}/\Delta t^{\text{ff}}_{P\mathcal{B}} < -0.05$ veh/s and $\Delta N^{\text{cg}}_{P\mathcal{B}}/\Delta t^{\text{cg}}_{P\mathcal{B}} - \hat{r} < -0.05$ veh/s. In these cases, the other backwards propagating wave is said to be restrictive. Figures 7.12b and 7.12c show the results of this approach for the two different parameter sets that we found in Section 7.4.1. In these figures, the colors green, red and cyan respectively denote that the v^f-wave is restrictive, the w-wave is restrictive or that the restrictive wave is not found. The differences in $\hat{w}$ and $\hat{r}$ cause a few differences in the points for which the v^f-wave is found to be restrictive. These points are indicated using a black circle in Figure 7.12c.

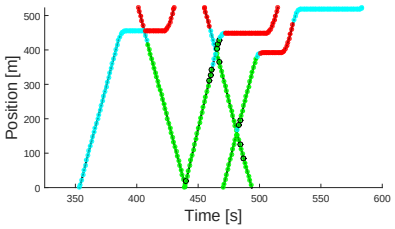
To estimate the upstream and downstream boundary conditions, i.e., $N(x_0,t)$ and $N(x_L,t)$, we respectively draw backwards propagating v^f-waves and w-waves origination from the points on the observation paths for which these waves were found to be restrictive. The estimated change in cumulative flow over these waves depends on the characteristic wave, i.e., $\Delta N^{\text{ff}} = 0$ veh and $\Delta N^{\text{cg}} = -\hat{r}\Delta t^{\text{cg}}$ veh. An additional assumption that follows from the assumed traffic flow model is used in estimation. After a vehicle that is part of the observed flow is initially restricted by the w-wave, it remains restricted by this wave for the remaining



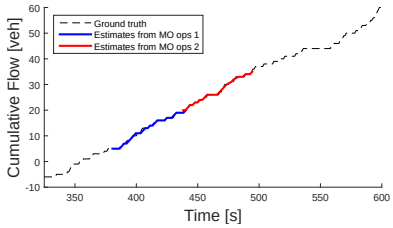
(a) Trajectories and moving observers.



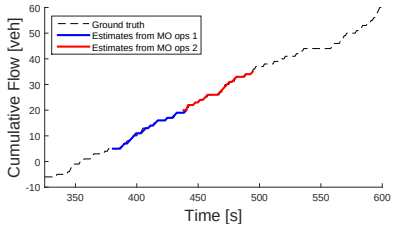
(b) Restrictive waves that are found by excluding the other characteristic wave (queue-to-queue congested wave parameters).



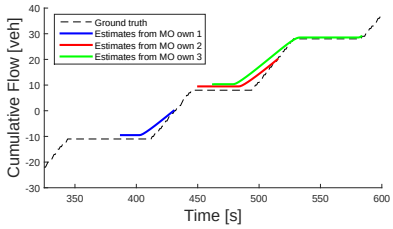
(c) Restrictive waves that are found by excluding the other characteristic wave (link outflow congested wave parameters).



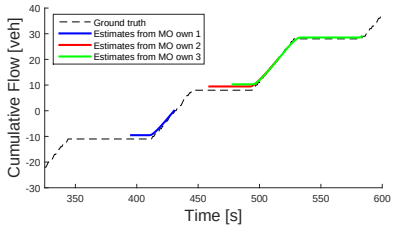
(d) True and partially estimated link inflow cumulative curves (queue-to-queue congested wave parameters).



(e) True and partially estimated link inflow cumulative curves (link outflow congested wave parameters).



(f) True and partially estimated link outflow cumulative curves (queue-to-queue congested wave parameters).



(g) True and partially estimated link outflow cumulative curves (link outflow congested wave parameters).

Figure 7.12: Explanation of principles that can be used to estimate the traffic state.

time it is on the link. Again, we consider the two model parameter sets. This yields four figures that depict the estimated (upstream or downstream) boundary conditions related to one of the parameter sets, see Figures 7.12d, 7.12e, 7.12f and 7.12g. The figures show the true cumulative curves together with the estimates related to the observation paths of individual moving observers that travel in opposite direction than the related characteristic wave. For instance, for the upstream boundary (which relates to the v^f-wave) the moving observers (MOs) that travel with a negative speed in space-time are shown. In all figures the cumulative flow is set to zero for the point in space-time at which the first moving observer enters the link, i.e., $N(x_0, 354) = 0$.

Figures 7.12d and 7.12e show that the upstream boundary conditions are accurately estimated for the periods that information is available. These figures are equal as both parameter sets use the same $\hat{v^f}$ and in both cases the observation paths are found to be restricted by the v^f-wave starting from the same time instances. Figures 7.12f and 7.12g show a larger difference between the estimates related to the two model parameters sets. The downstream boundary conditions are more accurately estimated for the second set, see Figure 7.12g. In Figure 7.12f, both the $\hat{w}$ and $\hat{r}$ seem to contribute to the errors in found in the boundary condition estimates. As a result, the slope during the green periods (link outflow) and the shockwave-speed (which according to kinematic wave theory should be w in this case) are respectively underestimated and overestimated (i.e., $\hat{w}$ should have a larger negative value). The overestimation of $\hat{w}$ leads to an earlier estimated start of the green period. Although the model parameters set used in Figure 7.12g lead to more accurate estimates of the link outflow cumulative curves, it seems that $\hat{w}$ is underestimated. This may be caused by stochasticity in the traffic behavior in combination with the fact that we only used a limited set of observations to approximate the model parameters.

The case study presented in this section shows that relative flow data collected with moving observers provides information that is valuable to learn the traffic flow model parameters and estimate the unobserved boundary conditions. Furthermore, we observe that the traffic flow simulated using VISSIM has two different queue outflow states, i.e., one related to vehicles that are able to leave the link within this green period and one related to vehicles that have to wait at least one more red period. Therefore, the assumptions behind the LWR-model with triangular FD are partially violated, which negatively influences the estimation accuracy. The voilation of the model assumptions was exposed using relative flow data, see Figures 7.11e and 7.11f. We may use the same learning dataset to expose which traffic phenomena occur and find the model that best describes traffic flow behavior. As stated in Section 7.3.1, the principles explained in this paper may be extended the LWR-model with other concave and continuous fundamental diagrams since this still allows us to apply Newell's method. However, the anticipation-behavior exposed in this simulation cannot be captured by the first-order LWR-model. This anticipation-behavior is included in second-order models such as the PW-model. Traffic behavior captured by such higher-order traffic flow models may still be captured in the cumulative flow plane using relative flow data. Further research into this topic is beyond the scope of the current paper.

7.5 Conclusions and insights

This paper presents the value of relative flow data in estimating the cumulative flow in space-time. To present the value of data, we take the perspective of a model-based estimation approach. In model-based estimation two processes are important: (1) information assimilation (IA) of real-time data and models and (2) learning models that are used in IA based on historical data. This study shows how relative flow data can be used for both processes and why the data characteristics (i.e., the observed variable in combination with the spatial-temporal characteristics) are valuable in traffic state estimation.

This study shows that relative flow data can be used to learn the link traffic flow model and to estimate the traffic state in the cumulative flow plane. If traffic flow follows the LWR-model with a triangular fundamental diagram (FD), relative flow data can be used to find the model parameters. Furthermore, relative flow data can be used to expose that the assumptions behind the LWR-model with a triangular FD are violated. The case study showed that these model assumptions were partially violated for the simulated (using VISSIM) traffic. Furthermore, this study explains that to estimate the traffic state, it is valuable to rely on a combination of moving observers and stationary observers positioned at the link boundaries. However, it also presents principles that can be used to relate the observation paths to each other, estimate the boundary conditions, and estimate the link supply and demand, for the scenario in which we rely on moving observers alone. The simulation study shows that it is possible to partially estimate the link boundary conditions using relative flow data that is collected by moving observers alone. The accuracy of these boundary conditions estimates depends on the ability to learn the model parameters that best describe the traffic flow behavior.

The principles explained in this study yield insights for the valuable characteristics of the relative flow data in model-based traffic state estimation. The valuable characteristics are: (1) Relative flow data provide information on the change in cumulative flow over paths in space-time. For this purpose the observer needs to observe all passings (overtakings), which means that all lanes should be observed, i.e., the full width of the road lie in the sensors range. (2) The observers that collect relative flow data should travel at different speeds through space and time. This is valuable as the observation paths of observers that travel at different speeds can intersect, which allows us to relate the cumulative flow observation to each other. Furthermore, the speeds at which observers travel influence to which points in space-time the characteristic waves can be drawn. Here, differences in the observer and characteristics wave speeds are beneficial. Therefore, it is desirable to have both relatively slow (e.g., trucks) and fast (e.g., passenger cars) vehicles that collect relative flow data, have observers that observe the opposite traffic flow, and combine stationary and moving observers.

There are multiple interesting future research directions. A first is a systematic (and analytical) extension of the effects of traffic properties, e.g., other FD shapes or higher order effects such as anticipation-behavior. Furthermore, based on the principles presented in this study, full methodologies and algorithms that exploit the informative value of the relative flow data can be developed. An important addition for this would be to apply the provided insights to information propagation over nodes.

Chapter 8

Using relative flow data to reveal traffic flow properties

This chapter is currently under review for journal publication.

Chapter 7 presented principles that can be used to learn the link traffic flow model and estimate the traffic state using relative flow data. In that study, the LWR-model with triangular fundamental diagram is considered. However, in Section 7.3.1, it is also noted that the principles can be extended to other (more complex) traffic flow models.

In this chapter, we investigate the opportunity to reveal the traffic flow properties using relative flow data. The term 'traffic flow properties' is used to describe which traffic states (i.e., combination of flow, density and speed) occur and how these change over space and time. The ability to reveal traffic flow properties is potentially beneficial for traffic state estimation and to gain a deeper understanding of traffic flow (i.e., contribute to traffic flow theory). To reveal these properties with relative flow data, we look at the observed changes in cumulative flow between points in space-time and the boundary conditions (cumulative flow curves at the link boundaries). Based on these observations different plots are constructed that provide insight in the traffic flow properties. The presented work builds on Chapter 7, but considers a more general description of traffic flow properties (i.e., LWR-model with concave continuous fundamental diagram), presents additional steps to reveal the traffic flow properties and conducts a more comprehensive case study. The ability to reveal traffic flow properties is potentially beneficial for traffic state estimation and to gain a deeper understanding of traffic flow (i.e., contribute to traffic flow theory).

8.1 Introduction

In this study, we use the term 'traffic flow properties' to describe which traffic states (i.e., combination of flow, density and speed) occur and how these change over space and time. These properties can be modeled using traffic flow models, which can be used in traffic state estimation and design of traffic management systems. The traffic flow properties depend on the infrastructure and driving characteristics, and can depend on other (external) conditions (e.g., weather). Differences in infrastructure, traffic regulations, vehicle characteristics and driver training make that the traffic flow properties are not the same everywhere and change over time. Therefore, it is important that traffic sensing data are collected that allow us to reveal the traffic flow properties in a local and online setting.

Traffic can be described on a macroscopic level using the traffic variables flow q, density k and mean speed u, where $q = ku$. Edie (1965) introduced the generalized definitions of traffic flow, which can be used to determine flow, density and speed for areas in space and time. Makigami et al. (1971) proposed to describe traffic using the cumulative flow N over space and time, i.e., $N(x,t)$. This variable is discrete by nature, but may be smoothed to obtain a continuous cumulative flow plane in space-time. The derivatives to space and time of this continuous function yield the macroscopic variables (minus) density and flow for points in space-time, i.e., $\partial N(x,t)/\partial x = -k(x,t)$ and $\partial N(x,t)/\partial t = q(x,t)$.

The influential works of G. F. Newell (1993a) and Daganzo (2005a) showed that traffic flow properties can be described in the cumulative flow plane. Multiple studies use these influential works to design methodologies for traffic state estimation, e.g., (Laval et al., 2012), (Claudel & Bayen, 2010b), (Claudel & Bayen, 2010a), (Sun et al., 2017) and (Takenouchi et al., 2019), and traffic control, e.g., (Wada et al., 2018). The methodologies designed in these studies assume that the traffic flow properties can be described by the LWR-model (Lighthill & Whitham, 1955), (Richards, 1956) with a continuous concave fundamental diagram (FD). For this purpose, Laval et al. (2012), Sun et al. (2017), Wada et al. (2018) and Takenouchi et al. (2019) consider a triangular FD, which is often used in traffic state estimation due to its simplicity, theoretical preferable characteristics (e.g., concaveness) and some theoretical evidence (Seo, Bayen et al., 2017).

In Van Erp et al. (2019), we showed that relative flow data provide information on the change in cumulative flow between points in space-time, and have explained that this information can be used to expose traffic flow properties. Relative flow data contain observations on the change in cumulative flow over paths in space and time, i.e., observation paths. These data can be collected using moving (e.g., automated or other equipped and connected vehicles) and stationary observers (Van Erp, Knoop & Hoogendoorn, 2018b), (Van Erp et al., 2019), (Takenouchi et al., 2019). Depending on the infrastructure and vehicle characteristics, moving observers may be able to collect relative flow data for opposing or parallel roads. However, this option is less likely for freeway traffic compared to urban traffic. Therefore, in this study, we solely collect relative flow data with stationary observers and moving observers that are part of the considered traffic flow, i.e., which move in the flow direction.

This article extends on the work in Van Erp et al. (2019). However, we relax the assumptions that we made in that study related to the FD shape (i.e., triangular shape), propose additional steps to expose the traffic flow properties and conduct a more extensive case study for freeway traffic flow. In this study, freeways are defined as uninterrupted multi-lane

flow facilities. In multi-lane road segments, overtaking is possible, which means that the cumulative flow value can change over vehicle trajectories.

The main contribution of this study is that we show which information relative flow data provide on traffic properties. For this purpose, different plots are constructed based on relative flow data. To interpret the plots we turn to Newell's simplified kinematic wave theory (G. F. Newell, 1993a), (G. F. Newell, 1993c), (G. F. Newell, 1993b) and Daganzo's variational theory (Daganzo, 2005a), (Daganzo, 2005b). Therefore, when interpreting the constructed plots, we take the traffic flow model assumptions related to these theories into account. The constructed plots provide insight in the ability to model the traffic flow properties using the LWR-model with a continuous concave fundamental diagram, but can also expose features that are not captured by this model. Constructing and interpreting the proposed plots is an important first step in (online) learning of traffic flow models. It shows which features can be estimated based on relative flow data and which extent model assumptions are valid. This can be used to design algorithm for (online) learning of traffic flow models, which is beyond the scope of the current article.

This article is structured as follows. In Section 8.2, we provide the important background for this study. Section 8.3 describes the relative flow data characteristics and how these data can be processed to obtain wave observations. Next, in Section 8.4, the theoretical relations between the wave observations and the FD characteristics are derived. In a simulation case study, which is presented in Section 8.5 and its results are shown in Section 8.6, we evaluate how suitable relative flow data are to expose traffic flow properties. Finally, Section 8.7 provides the conclusions of this work.

The mathematical symbols used throughout this article are presented in Table 8.1.

8.2 Background on traffic flow modeling and estimation

This section provides the background that is relevant for this study. First, it presents the LWR-model with continuous concave fundamental diagram. Second, variational theory is discussed, which is important for the approach proposed in this study to expose traffic flow properties.

8.2.1 LWR-model with continuous concave fundamental diagram

The basis for all considered macroscopic traffic flow models is the conservation-of-condition:

$$\frac{\partial k(x,t)}{\partial t} + \frac{\partial q(x,t)}{\partial x} = 0 \tag{8.1}$$

For steady-state conditions the relation between density k and flow q can be described by the function $Q(k)$. This function is known as the fundamental diagram (FD).

The LWR-model (Lighthill & Whitham, 1955), (Richards, 1956) combines the conservation-of-vehicles conditions with an FD, which yields:

$$\frac{\partial k}{\partial t} + \frac{\partial Q(k)}{\partial x} = 0 \tag{8.2}$$

Table 8.1: Overview of symbols that are used in this article.

Description	Notation	Unit
Position	x	m
Time	t	s
Location upstream link boundary	x_0	m
Location downstream link boundary	x_L	m
Flow	q	veh/s
Density	k	veh/m
Mean speed	u	m/s
Cumulative flow	N	veh
Wave observations (Section 8.3):		
Change in space from P to B	Δx_{PB}	m
Change in time from P to B	Δt_{PB}	s
Change in cumulative flow from P to B	ΔN_{PB}	veh
Wave speed	v_W	m/s
Wave passing rate	q_W^{rel}	veh/s
Fundamental diagram (FD) (Section 8.2 and 8.4):		
FD flow function	Q	veh/s
FD flow function free-flow branch	Q^{ff}	veh/s
FD flow function congested branch	Q^{cg}	veh/s
Critical density	k_{cr}	veh/m
Jam density	k_j	veh/m
Minimum density congested branch	k_o	veh/m
Set of potential k-values for the FD function	$\mathcal{K}$	veh/m
Characteristic wave speed	w	m/s
Set of potential w-values for the FD function	$\mathcal{W}$	m/s
Characteristic passing rate	r	veh/s

This model assumes that all possible traffic states lie on the FD and that the traffic state can change instantly over space and time. Therefore, in this model, vehicle speeds change instantly when moving from one state to another.

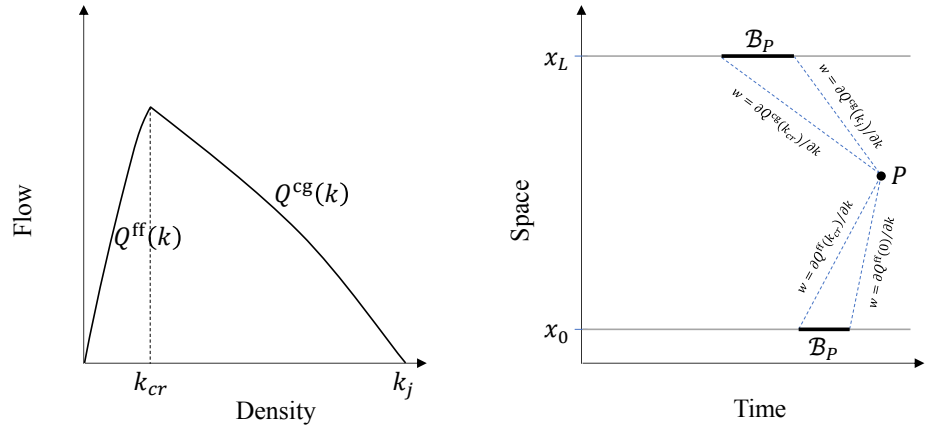
Throughout this study, we consider homogeneous freeways. This means that the same flow function (FD) hold for all locations, i.e., $\partial Q/\partial x = 0$. Furthermore, the FD is assumed to be concave. The FD is concave if its second-derivative to density is always smaller or equal to zero, i.e., $\partial^2 Q/\partial k^2 \leq 0$. It is strictly concave if its second-derivative to density is always smaller than zero, i.e., $\partial^2 Q/\partial k^2 < 0$.

Figure 8.1a show a general continuous concave FD, which is considered for explanatory purposes. This FD has two branches, i.e., a free-flow and a congested branch. Therefore, the FD function $Q(k)$ is described by two separate functions, i.e., $Q^{\text{ff}}(k)$ and $Q^{\text{cg}}(k)$, where:

$$\begin{aligned} Q(k) &= Q^{\text{ff}}(k) \quad \text{if } k < k_{cr}, \\ Q(k) &= Q^{\text{cg}}(k) \quad \text{if } k \geq k_{cr}\ . \end{aligned} \tag{8.3}$$

8.2.2 Newell's simplified kinematic wave theory and Daganzo's variational theory

G. F. Newell (1993a), (G. F. Newell, 1993c), (G. F. Newell, 1993b) presented a simplified theory of kinematic waves in highway traffic. Daganzo (2005a), Daganzo (2005b) extended on Newell's theory, leading to 'a variational formulation of kinematic waves'. Daganzo demonstrated that:



(a) Continuous concave FD with a free-flow and a congested branch

(b) Domain of dependence for a point P

Figure 8.1: The FD flow function and domain of dependence for a point P.

$$N_P = \min_{B \in \mathcal{B}_P} \{N_B + \Delta_{BP}\} \tag{8.4}$$

where P and B are points in space-time, $\mathcal{B}_P$ is the set of all points on a boundary within the domain of dependence of P and Δ_{BP} is the maximum change in cumulative flow over a path between B and P (Laval et al., 2012).

In Daganzo (2005b), fact 3, it states that the cost between two points is minimized by a straight path. Throughout this study, we will use the term wave describe a straight path. The speed of this wave (denoted as the wave speed w, where $w = \partial x_W(t)/\partial t$) is thus constant, i.e., $\partial^2 x_W(t)/\partial t^2 = 0$.

In line with Van der Gun et al. (2017), we assume that the considered link is initially empty and extend the boundary conditions sufficiently far into the past such that $\mathcal{B}_P$ lies on the upstream and downstream boundaries. The $\mathcal{B}_P$ depends on the flow function (FD) $Q(k)$. As stated above, we solely consider continuous concave Q. For continuous concave Q, $\mathcal{B}_P$ depends on the set of characteristic wave speeds $\mathcal{W}$. This set can be constructed by finding the characteristic wave speeds w of Q for all potential density values $\mathcal{K}$. The characteristic wave speed related to a density k is given by:

$$w(k) = \frac{\partial Q(k)}{\partial k} \tag{8.5}$$

For the continuous concave FD shown in Figure 8.1a all w related to the free-flow branch are positive, i.e., $\partial Q^{ff}(k)/\partial k > 0 \forall k \in [0, k_{cr}]$, while the w related to the congested branch are all negative, i.e., $\partial Q^{cg}(k)/\partial k > 0 \forall k \in [k_{cr}, k_j]$. Therefore, the characteristic waves that relate to the free-flow and congested branches respectively originate from the upstream and

downstream boundaries. Therefore, $\mathcal{B}_P$ can be described using time-ranges on the upstream (x_0) and downstream (x_L) boundaries (see Theorem 1 in (Van der Gun et al., 2017)):

$$\begin{aligned} t &\in \left[t_P - \tfrac{x_P - x_0}{\partial Q^{\text{ff}}(k_{cr})/\partial k}, t_P - \tfrac{x_P - x_0}{\partial Q^{\text{ff}}(0)/\partial k}\right] && \text{for } x_0 \text{ and,} \\ t &\in \left[t_P - \tfrac{x_P - x_L}{\partial Q^{\text{cg}}(k_{cr})/\partial k}, t_P - \tfrac{x_P - x_L}{\partial Q^{\text{cg}}(k_j)/\partial k}\right] && \text{for } x_L. \end{aligned} \tag{8.6}$$

Figure 8.1b visualizes these points on the link boundaries and shows the waves that need to be drawn to find N_P.

To find the maximum change in cumulative flow over a path between B and P, i.e., Δ_{BP}, we should again look at Q. The maximum wave passing rate r, i.e., maximum rate at which N changes over a wave, depends on the flow function Q and wave speed w (Daganzo, 2005b):

$$r(w) = R(w) = \sup_k \{Q(k) - kw\} \tag{8.7}$$

For concave Q this reduces to:

$$r(w) = Q(k(w)) - k(w)w \tag{8.8}$$

For strictly concave Q, every value of $w \in \mathcal{W}$ yields an unique value of $k(w)$. However, for concave Q that are not strictly concave, there are multiple potential k related to w. For instance, the triangular FD has two characteristic wave speeds (v^f and w), which respectively relate to $k \in [0, k_{cr}]$ and $k \in [k_{cr}, k_j]$. Still, any value of $w \in \mathcal{W}$ yields a unique value of r, i.e., respectively 0 and r.

Given that the speed along a wave from B to P is w_{BP}, the maximum change in cumulative flow over the wave is:

$$\Delta_{BP} = \int_{t_B}^{t_P} r(k(x_B + w_{BP}(t - t_B), t_B + t))dt \tag{8.9}$$

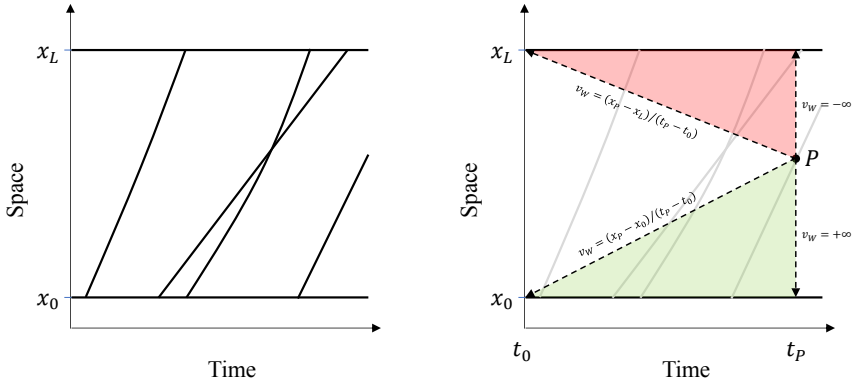
$$= r(w_{BP})\Delta t_{BP} \tag{8.10}$$

8.3 Obtaining wave observations from relative flow data

This section explains the relative flow data characteristics and how wave observations can be obtained from these data.

8.3.1 Collecting relative flow data

In line with Florin & Olariu (2017), Van Erp, Knoop & Hoogendoorn (2018b), Van Erp et al. (2019) and Takenouchi et al. (2019), this study assumes that relative flow data can be collected using stationary and moving observers. These data describe the relative flow over the paths in space-time, i.e., the observation paths. Stationary observers are installed at a



(a) The considered data configuration with stationary and moving observers

(b) Set of wave observations that can be drawn from point P to the link boundaries traveling in the direction of decreasing time

Figure 8.2: Data configuration and wave observations.

fixed position and thus observe the flow with respect to this this fixed position over time. For moving observers that are part of the traffic flow, the observation path is the trajectory of the observer. Relative flow data collected using these moving observers describe the overtakings, i.e., +1 if the observer is overtaken by another vehicle and -1 if the observer overtakes another vehicle.

The relative flow equals the change in the cumulative flow. If observers cross each other in space and time, i.e., if observation paths intersect, we can describe the change in cumulative flow between any two points on the intersecting observation path (Van Erp, Knoop & Hoogendoorn, 2018b). In this study, we consider a set of stationary and moving observers for which the change in cumulative flow between any combination of two points along any observation path can be determined.

Throughout this paper, we consider a data configuration in which relative flow data are collected using stationary observers positioned at the link boundaries (x_0 and x_L), and using moving observers that are part of the traffic flow. This data configuration is visualized in Figure 8.2a. As the observation paths (i.e., the black lines) intersect, we observe the change in cumulative flow between any two points that lie on the black lines shown in Figure 8.2a.

8.3.2 Obtaining the wave observations

A wave is defined as a straight line between two points in space-time. The spatial-temporal wave characteristics are given by the change in space and time between the two points, i.e., Δx_{PB} and Δt_{PB} for a wave traveling from P to B. Furthermore, the change in cumulative flow over the wave is given by ΔN_{PB}.

Throughout this paper, we solely consider waves that travel from points on the observation paths of the moving observers to the link boundaries (i.e., observation paths of the stationary observers) in the direction of decreasing time. The points on the observation

paths are denoted using the letter P, the observed link boundaries are denoted using $\mathcal{B}$ and points on these boundaries are denoted using the letter B. We only consider points P that lie between the boundaries, i.e., $x_0 < x_P < x_L$. Furthermore, as we only consider wave traveling in the direction of decreasing time, i.e., $t_B < t_P$ and thus $\Delta t_{PB} < 0$.

Based on the spatial-temporal and cumulative flow differences between points P and B, we can obtain the wave speed v_W and wave (mean) passing rate q_W^{rel}, i.e., $v_W = \Delta x/\Delta t$ and $q_W^{\text{rel}} = \Delta N/\Delta t$. Given that $\Delta t_{PB} < 0$ and $x_0 < x_P < x_L$, waves traveling from P to the downstream boundary (x_L) will travel with a negative speed, while the waves traveling to the upstream boundary (x_0) will travel with a negative speed.

The set of waves that can be drawn from P to the boundaries depends on the initial time of collecting relative flow data using stationary observers (t_0) and the spatial temporal characteristic of P. As shown in Figure 8.2b, the waves can be constructed for all waves speeds within $[-\infty, (x_P - x_L)/(t_P - t_0)]$ and $[(x_P - x_0)/(t_P - t_0), +\infty]$. The set of the wave speeds that can be drawn from P to $\mathcal{B}$ is denoted using $\mathbf{O}_P$.

8.4 Methodology to expose traffic flow properties

In Van Erp et al. (2019), we explained that relative flow data can be used to find the model parameters if it is assumed that traffic flow follows the LWR-model with triangular fundamental diagram (FD). In this study, we consider a specific combination of observers that collect relative flow data (see Section 8.3). Furthermore, the assumption of a triangular FD is relaxed and derive the equations that can be used to expose the traffic flow properties that is described using the LWR with any continuous concave FD.

Throughout this section, three FDs are considered for explanatory purposes. First, a general concave and continuous FD, which is considered for the theoretical explanations in the remainder of this section. (2) The triangular FD, which is commonly considered in traffic state estimation (Seo, Bayen et al., 2017). This FD is concave, but not strictly concave. And third, Smulders' FD, which has a strictly concave free-flow branch and linear congested branch. Figures 8.3a, 8.3b and 8.3c show these three FDs.

All three FDs have two branches, i.e., the free-flow and congested branch. The branch-specific flow functions for the triangular FD are:

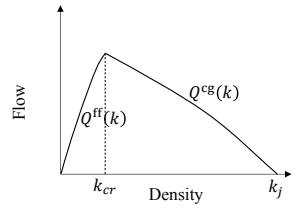
$$\begin{aligned} Q(k) &= w^{\text{ff}}k && \text{if } k < k_{cr}, \\ Q(k) &= r^{\text{cg}} + w^{\text{cg}}k && \text{if } k \geq k_{cr}. \end{aligned} \tag{8.11}$$

where w^{ff} denotes the free-flow speed (which has a positive value), w^{cg} the congested wave speed (which has a negative value) and r^{cg} is the passing rate observed by a moving observer traveling at w^{cg} through congested states.

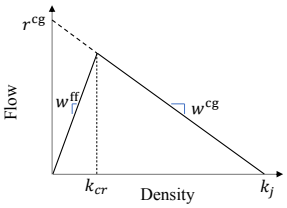
The FD of Smulders (1990) is described by:

$$\begin{aligned} Q(k) &= w_0^{\text{ff}}k - \tfrac{w_0^{\text{ff}} - u(k_{cr})}{k_{cr}}k^2 && \text{if } k < k_{cr}, \\ Q(k) &= r^{\text{cg}} + w^{\text{cg}}k && \text{if } k \geq k_{cr}. \end{aligned} \tag{8.12}$$

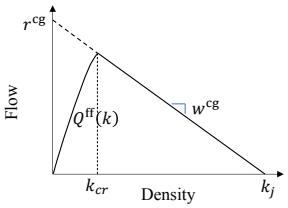
where the mean speed u always decreases with increasing density. At the extremes of the free-flow branch the mean speed is respectively $u(0) = w_0^{\text{ff}}$ and $u(k_{cr})$. Between these points the speed decreases linearly with density.



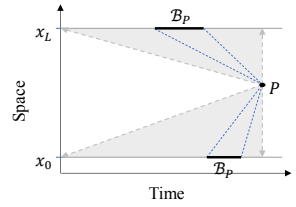
(a) General continuous concave FD



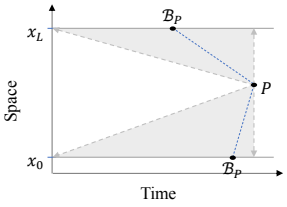
(b) Triangular FD



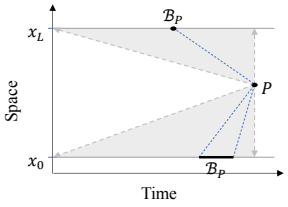
(c) Smulders' FD



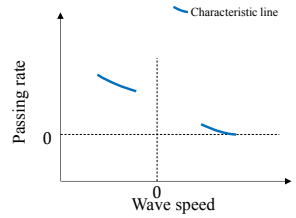
(d) Domain of dependence for general continuous concave FD



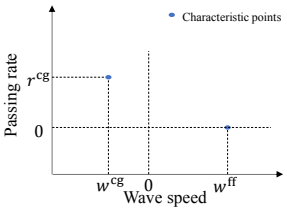
(e) Domain of dependence for triangular FD



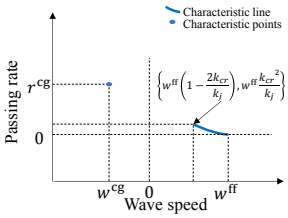
(f) Domain of dependence for Smulders' FD



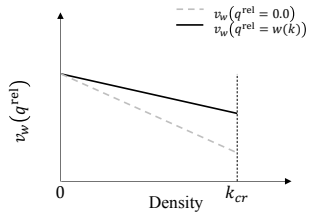
(g) Characteristic lines for the general continuous concave FD



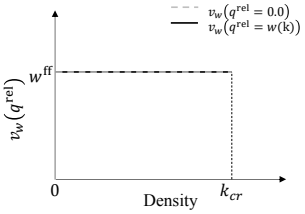
(h) Characteristic points for the triangular FD



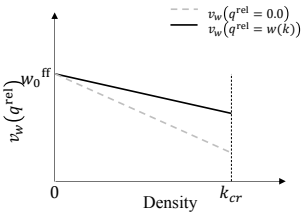
(i) Characteristic point and line for Smulders' FD



(j) Free-flow wave speeds related to specific passing rates for the general continuous concave FD



(k) Free-flow wave speeds related to specific passing rates for the triangular FD



(l) Free-flow wave speeds related to specific passing rates for Smulders' FD

Figure 8.3: The different features explained in Section 8.4 for the three different fundamental diagrams.

For each of the three FDs shown in Figures 8.3a, 8.3b and 8.3c, the set of characteristic wave speeds can be determined using equation (8.5). This yields:

$$\begin{array}{ll} \mathcal{W} = \left\{ \left[\frac{\partial Q^{\text{ff}}(0)}{\partial k}, \frac{\partial Q^{\text{ff}}(k_{cr})}{\partial k} \right], \left[\frac{\partial Q^{\text{cg}}(k_{cr})}{\partial k} \right] \right\} & \text{for the general concave FD} \\ \mathcal{W} = \left\{ w^{\text{ff}}, w^{\text{cg}} \right\} & \text{for the triangular FD} \\ \mathcal{W} = \left\{ \left[w_0^{\text{ff}}, 2u(k_{cr}) - w_0^{\text{ff}} \right], w^{\text{cg}} \right\} & \text{for Smulders' FD} \end{array} \tag{8.13}$$

Figures 8.3d, 8.3e and 8.3f show the domain of dependence $\mathcal{B}_P$ (which depends on the set of characteristic wave speeds) for the three FDs.

8.4.1 Finding characteristic wave speeds and passing rates from wave observations

As explained in Section 8.3, relative flow data can be used to obtain wave observations that travel from points P in the direction of decreasing time to the boundary $\mathcal{B}$. The set of wave speeds for which these wave observations can be drawn is denoted by $\mathbf{O}_P$. In Figures 8.3d, 8.3e and 8.3f, this set is indicated by the grey areas. Depending on $\mathbf{O}_P$ and Q, $\mathbf{O}_P$ can contain all or a part of the characteristic wave speeds related to Q. For the example shown in the figures, wave observations are available that relate to all characteristic wave speeds. This means that $\mathbf{O}_P$ is a superset of the set of characteristic waves $\mathcal{W}$, i.e., $\mathbf{O}_P \supset \mathcal{W}$.

The unknown Q constraints the wave observations. Therefore, by evaluating the observed passing rates over different wave speeds, we may expose the traffic flow properties that is described by the LWR-model and Q. For this purpose, we will use equations (8.4) and (8.10). Combined these yields:

$$N_P = \min_{B \in \mathcal{B}_P} \{ N_B + r(w_{BP}) \Delta t_{BP} \} \tag{8.14}$$

By moving the left part of the equation to the right we obtain:

$$0 = \min_{B \in \mathcal{B}_P} \{ N_B + r(w_{BP}) \Delta t_{BP} - N_P \} \tag{8.15}$$

where $N_B - N_P = \Delta N_{PB}$, and $\Delta t_{BP} = -\Delta t_{PB}$. This yields:

$$0 = \min_{B \in \mathcal{B}_P} \{ \Delta N_{PB} - r(w_{BP}) \Delta t_{PB} \} \tag{8.16}$$

which in turn can be rewritten to:

$$0 = \max_{B \in \mathcal{B}_P} \left\{ \frac{\Delta N_{PB}}{\Delta t_{PB}} - r(w_{BP}) \right\} \tag{8.17}$$

Note that the waves considered in equation (8.14) are traveling with a fixed speeds in the direction of increasing time. As we turned the time-direction, in equation (8.17), we are now only considering waves that travel at fixed speeds in the direction of decreasing time.

Relation (8.17) states that for any characteristic waves speed, i.e., $\mathcal{W} \in \mathcal{W}(\mathcal{K})$, the rate observed over a wave traveling at this speed is equal or smaller to the characteristic rate

related to that wave speed, i.e., $r^{\mathcal{W}}$. Furthermore, it states that the rate observed over at least one characteristic wave speed in $\mathcal{W}$ should be equal to the characteristic rate related to that wave speed, which is given equation (8.8).

As shown in Figures 8.3g, 8.3h and 8.3i, the characteristic wave speeds and passing rates are dependent on the shape of the FD branches. If the branch is linear (e.g., triangular FD), each branch yields a single combination of $\{w,r\}$ (e.g., $\{w^{\text{ff}},0\}$ and $\{w^{\text{cg}},r^{\text{cg}}\}$ for the triangular FD). If the branch is non-linear and smooth (e.g., the free-flow branch of Smulders' FD), the branches both yield a characteristic line.

If a FD branch is described by a non-linear function, i.e., Q^{ff} or Q^{cg}, the characteristic branch-specific wave speeds will lie within a certain range. For specific branches this range is expected to be small with respect to the range between the minimum and maximum characteristic wave speeds for the full FD, i.e., Q. A point that is restricted by a wave related to a specific branch, needs to touch the branch-specific characteristic line. Therefore, as this wave speed range of these lines is expected to be limited, the range of observed passing rates within the wave speed range is expected to be limited. This means that we are able to gain a better insight in the traffic flow model shape and parameters if we make a separation between points that are restricted by free-flow waves and those that are restricted by congested waves. To make this separation the individual speeds of the moving observers can be used. This approach will be followed and explained in more detail in the case study.

8.4.2 Stochasticity and non-linear fundamental diagram branches

In reality, traffic is stochastic, which means that it will not be perfectly described by a traffic flow model. Due to this stochasticity the observed wave speeds and passing rates may not perfectly follow the explanations above. For instance, even if the LWR-model with triangular FD is the best description of traffic flow properties, the observed wave passing rates at the characteristic speeds, i.e., $q_W^{\text{rel}}(w^{\text{ff}})$ and $q_W^{\text{rel}}(w^{\text{cg}})$, can differ slightly from the related characteristics passing rates, i.e., 0 and r^{cg}. However, if the LWR-model with non-linear FD branches perfectly describe traffic flow (i.e., there is no stochasticity), the same effect is observed.

To assess whether the FD branch is best described using a linear or non-linear function, we look at the changes in wave passing rate when the wave speed changes for specific origin points P. The equations presented below, assume that traffic conditions are homogeneous and stationary over the considered waves. This assumption is valid for restrictive wave, as kinematic wave theory prescribes that the traffic state is constant along restrictive waves in case the LWR-model with concave and continuous FD holds, i.e., $k(x_W(t),t) = k_W$ and $q(x_W(t),t) = q_W$. Due to stochastic traffic flow properties and because the precise shape and parameters of the FD are still unknown, it is difficult to find the restrictive wave. However, if we are able to approximate the characteristic lines and/or points related to FD, we can find waves that close to the restrictive wave. Furthermore, if the changes in demand and supply are limited, the traffic conditions are expected to be approximately homogeneous and stationary, and are expected to be representative for the state along the restrictive wave. However, in using the relations presented below to assess the stochasticity and non-linearity of FD branches, one should keep in mind that heterogeneity and non-stationarity over the considered waves can affect the results.

Let us consider a case in which traffic is homogeneous and stationary, i.e., the traffic state

is constant over space and time. In this case, waves from a point P to the link boundaries travel to a single traffic state, which we will denote as $\{q_W, k_W\}$. The observed passing rate q_W^{rel} over wave traveling at speed v_W is given by:

$$q_W^{\text{rel}}(v_W) = q_W - k_W v_W \tag{8.18}$$

The density along the wave is not directly observed, but can be obtained by taking the derivative of equation (8.18) to the wave speed:

$$\frac{\partial q_W^{\text{rel}}(v_W)}{\partial v_W} = -k_W \tag{8.19}$$

which allows us to estimate the density based on different wave observations. For this purpose, the left hand side of equation (8.19) can be approximated based on the wave observations. The approach applied in this study to approximate density is discussed in more detail in Section 8.5.2.

Equation (8.18) can be rewritten to describe the wave speed as a function of the passing rate:

$$v_W\left(q_W^{\text{rel}}\right) = \frac{q_W - q_W^{\text{rel}}}{k_W} \tag{8.20}$$

$$= u_W - \frac{q_W^{\text{rel}}}{k_W} \tag{8.21}$$

If the potential traffic states $\{q_W, k_W\}$ are described by the fundamental diagram (FD) $Q(k_W)$, equation (8.20) can be rewritten to:

$$v_W\left(q_W^{\text{rel}}\right) = \frac{Q(k_W) - q_W^{\text{rel}}}{k_W} \tag{8.22}$$

Let us look at a single branch on the FD, i.e., the free-flow branch $Q^{\text{ff}}(k_W)$. Equations (8.19) and (8.22) allow us to describe the relation between the wave passing rate and density for a specific wave speed, i.e., $v_W\left(q_W^{\text{rel}}\right)$ and k_W. As the flow is zero is the density is zero, the free-flow FD branch is restricted at the origin, i.e., $Q^{\text{ff}}(0) = 0$. Therefore, $q_W^{\text{rel}} = 0$ has to be a characteristic passing rate that relates to the free-flow branch; however, we do not know which characteristic wave speed relates to this characteristic passing rate.

If we take $q_W^{\text{rel}} = 0$, equation (8.22) simplifies to:

$$v_W(0) = \frac{Q(k_W)}{k_W} \tag{8.23}$$

which is the speed - density FD. If we only consider points that are restricted by free-flow waves and the free-flow branch is linear (e.g., triangular FD), this results in a horizontal line in the $\{k_W, v_W(0)\}$ from $k_W = 0$ till $k_W = k^{cr}$. This FD branch is concave, but not strictly

concave. In case of a strictly concave free-flow branch (e.g., Smulders' FD), this results in a strictly decreasing function from $k_W = 0$ till $k_W = k^{cr}$. Figures 8.3j, 8.3k and 8.3l show the resulting functions that we would find for homogeneous and stationary traffic in the density - $v_W\left(q_W^{\text{rel}} = 0\right)$ plane. In the figure that relates to the general continuous concave FD, we show a linear decreasing wave speed (similar to Smulders' FD); however, as the function for the general continuous concave FD is not defined, other shapes are also possible.

If instead of $q_W^{\text{rel}} = 0$, we plot $q_W^{\text{rel}} = r(k)$ as a function of density, we obtain the relation between the characteristic wave speed and density. As Q is concave and $Q(0) = 0$, $r \geq 0$ for all $k > 0$. Therefore, $v_W(0) \geq v_W(r)$. If Q is strictly concave, $r > 0$ for all $k > 0$ and $v_W(0) > v_W(r)$. This difference is also shown in Figures 8.3j and 8.3l, where the dashed grey line (which describes $v_W(r)$) is lower than the solid black line (which describes $v_W(0)$).

For the congested FD branch, we do not know with certainty that a specific passing rate is a characteristics passing rate. However, we could use different passing rates and find the related wave speeds as a function of density. If the congested branch is described by a linear function, the characteristic passing rate will result in a horizontal line in the density - $v_W\left(q_W^{\text{rel}} = r\right)$ plane.

As a final step, we plot estimates of flow and density, which provide insight in the fundamental diagram flow function $Q(k)$. To construct this plot, flow is estimated based on density and $v_W\left(q_W^{\text{rel}} = r\right)$. For this purpose, equation 8.18 can be rewritten to:

$$q_W = q_W^{\text{rel}} + k_W v_W\left(q_W^{\text{rel}}\right) \tag{8.24}$$

Multiple factors influence the accuracy of the flow and density estimates, i.e., (1) the separation of point P in points that are restricted by free-flow or congested waves, (2) the approximation of the left hand side of equation (8.19) and (3) the selection of the characteristics passing rates. These factors should be taken into account when interpreting the plots and may explain outliers that are observed.

The relations presented in Sections 8.4.1 and 8.4.2 can be used to expose traffic flow properties that is described by the LWR-model with a continuous concave FD Q based on wave observations obtained from relative flow data. The relations show that data transformations and visualizations can be used to expose this properties. In the next section, a case study is presented that uses the relations presented in this section.

8.5 Case study

A simulation case study is conducted to evaluate whether relative flow data and the methodology proposed in Section 8.4 can be used to expose traffic flow properties.

In this study, traffic is simulated using the FOSIM (Dijker & Knoppers, 2006). FOSIM is a microscopic simulation program that is designed, calibrated and validated to simulate traffic for Dutch freeways (Henkens et al., 2017). Microscopic simulation provides the freedom to test specific scenarios. It allows us to design the infrastructure and define the traffic demand. Furthermore, a dataset describing all trajectories can be collected. This dataset provides insight into the true traffic conditions and it allows us to construct the relative flow data. This combination, i.e., freedom to test the specific scenarios and have

access to a trajectory dataset, is not found in empirical datasets. For instance, the widely used NGSIM trajectory dataset describes a relative short and complex road segment, i.e., five mainline lanes with an approximate length of 640 m, an off-ramp and an on-ramp (Colyar & Halkias, 2007).

Below, we explain which scenarios are consider in our case study. Furthermore, we explain which data are collected and how these are used to obtain wave observation points. Finally, the different steps taken to expose traffic flow properties are presented.

8.5.1 Scenarios

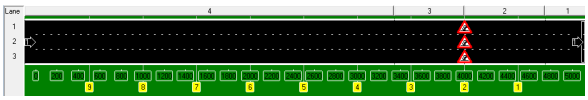
The methodology proposed in the previous section assumes a homogenous freeway. This means that the FD function is assumed to be constant over space, i.e., $\partial Q/\partial x = 0$. Therefore, in this case study, we apply the proposed methodology to scenarios for which this assumption is expected to be valid.

The validity of assuming a constant FD function for a road segment depends on the infrastructure of the road segment and the infrastructure upstream and downstream of the road segment. The road segment itself should have a homogenous infrastructure, which means that the number of lanes and traffic rules (e.g., maximum speed) are constant over space. Furthermore, traffic behavior should not be significantly influenced by strategic choices for downstream restrictions or route-choices. Examples of such significant influences are merging effects (e.g., merging of on-ramp and through-going traffic) or strategic choices that relate to downstream infrastructure (e.g., pre-sorting for a downstream off-ramp). Therefore, in our case study, we consider scenarios in which traffic does not pre-sort for downstream directional choices and there are no directional downstream supply restrictions. Following this definition, links upstream of an on-ramp can be considered, while links upstream of off-ramps should not.

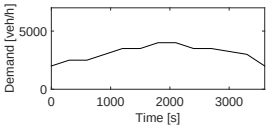
The three scenarios considered in this case study, i.e., (1) a downstream temporary blockage, (2) a downstream on-ramp and (3) a downstream lane drop, are simulated using FOSIM. For all scenarios the link length is 4000 m and the number of lanes is three. Figure 8.4 shows the three road layouts in FOSIM together with the traffic demand. These layouts show a road segment with a length of 5000 m. Out of this segment, the first 4000 m is considered for the evaluated links, i.e., the upstream and downstream link boundaries are respectively located at $x_0 = 0$ m and $x_L = 4000$ m. The on-ramp and temporary bottleneck are positioned at the downstream link boundary, while the lane drop is located at $x = 4750$ m. We choose to position the lane drop further downstream to account for the lane changing (converging) behavior directly upstream of the lane drop.

8.5.2 Data collection and observation points selection

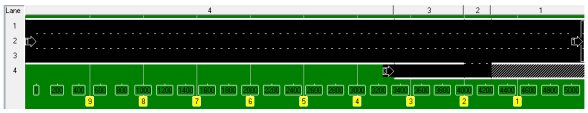
Disaggregated relative flow data is collected by stationary observers that are located at the link boundaries, i.e., $x_0 = 0$ m and $x_L = 4000$ m. Furthermore, 1 % of the road users (which are randomly selected) serves as moving observer. Along the moving observer paths, observation points are selected with a 15 s interval. These observation points are used as the point-of-origin P of waves that travel in the direction of decreasing time to the upstream and downstream boundaries.



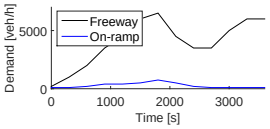
(a) Temporary blockage, road layout



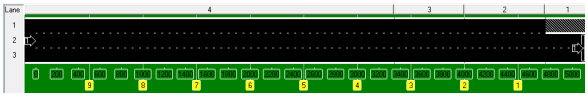
(b) Temporary blockage, demand



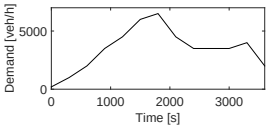
(c) On-ramp, road layout



(d) On-ramp, demand



(e) Lane drop, road layout



(f) Lane drop, demand

Figure 8.4: Road layout in FOSIM and demand (inflow) for the three scenarios.

8.5.3 Exposing traffic flow properties

Multiple data transformations and visualizations are used to expose the traffic flow properties using relative flow data. First, we construct the (slanted) cumulative curves for the link boundaries. These provide insight in the time-variation of the traffic states that occur at these boundaries and that propagate in increasing direction of time. This is interesting, as the extent to which the traffic state fluctuates over time influences our ability to asses the non-linearity of the FD branches, see explanation in Section 8.4.2.

Next, for analysis purposes, see below, a traffic phase (i.e., free-flow or congestion) is assigned to each observation point based on the individual vehicle speeds of the moving observers. For this purpose, we first visualize the moving observer speeds in space-time. Next, we set a speed threshold value, which is used to define when the traffic phase of a moving observer changes from free-flow to congestion. Once the traffic phase switches to the congested traffic phase, we assign all later observation points of the specific moving observer to the congested state. This mean that a queue discharge state is assigned to the congested phase. For the next steps, the points P are separated into two groups depending on their estimated traffic phase.

Two steps are taken to expose the traffic flow properties. The first step provides a first approximation of the characteristic wave speeds and passing rates that relate to the two FD branches. The second step uses this first approximation and aims to get more detailed insights in the shape and parameters of the FD branches. In both steps, we do consider points that are close to the boundaries because we may observe extreme passing rate values for short waves, which are not representative for the traffic flow properties. For thus purpose, all points P within 250 m of the link boundaries are not used to construct the plots shown in the next section.

To approximate the branch-specific characteristic wave speeds and passing rates the waves are drawn from all points P that relate to the specific branch to points B on the link boundaries. Here, all time instances that a vehicle passes the link boundaries are used as points B. Furthermore, for the temporary bottleneck scenario, an additional point B is added every second during the period that there is no outflow. Without adding these additional points, we are limited in drawing potentially restrictive congested waves from the points for which these are important. Each wave drawn from a point P to a point B yields a wave speed and passing rate. Note that, as explained in Section 8.3, that we only consider waves that travel in the direction of decreasing time, i.e., $t_P > t_B$. This yields a large set of wave speeds and related passing rates per branch. Next, we select all passing rates that fall with wave speed bin of 0.5 m/s and determine the 5, 50 (median) and 95 percentiles. Here, bins are selected from wave speed is -25 m/s till 50 m/s, which is expected to be a wide enough range to include characteristic wave speeds. This yields the branch-specific passing rate ranges for different wave speeds for all three scenarios, which in turn allow us to approximation the characteristic speeds and passing rates.

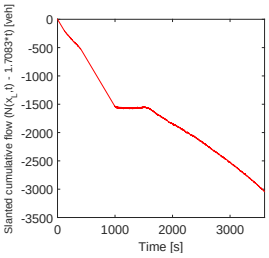
To gain more detailed insights in the FD branch characteristics, we follow the analysis presented in Section 8.4.2. As explained in that section, we can find the wave speed that relates to a certain passing rate, i.e., $v_W(q_W^{\text{rel}})$, and can estimate the density over a wave traveling at that speed, see equation (8.19). For the points that are expected to be restricted by a free-flow wave, we select $v_W(q_W^{\text{rel}} = 0)$ as $q_W^{\text{rel}} = 0$ needs to be a characteristic passing rate for the free-flow FD branch. For the other points, a passing rate is selected based on the branch-specific passing rate ranges for different wave speeds for all three scenarios, see above. To estimate density over the wave, the derivative of equation (8.19) needs to be approximated. To this extent, we find the passing at the boundary for which the wave passing rate is closest to the selected passing rate, i.e., $q_W^{\text{rel}} = 0$ for points that are expected to be restricted by a free-flow wave. Next, we take the passing times related to the tenth passing before and after the 'closest passing', which means that we consider a range of $\Delta N = 20$. For this range the change in wave passing rate ($\Delta q_W^{\text{rel}}(v_W)$) and the change in wave speed (Δv_W) are determined and density is estimated.

8.6 Results

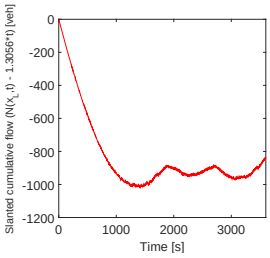
This section presents the results of the case study. As explained in the previous section, three-lane links with three different downstream nodes are considered.

All figures that are shown in this section are constructed using the relative flow data from moving observers and stationary observers positioned at the link boundaries. This means that the same figures and analysis can be performed in other scenarios for which these data are available. Therefore, empirical relative flow data can be used in the same way to observe and model traffic flow properties.

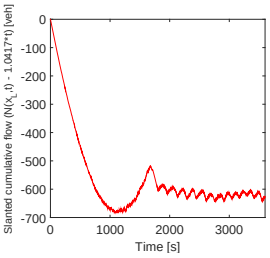
Below, we will first analyze the traffic conditions and divide the observation points in those that are expected to be restricted by free-flow and congested waves. Next, in two steps, we gain insight in the traffic flow properties based on the wave observations obtained from relative flow data.



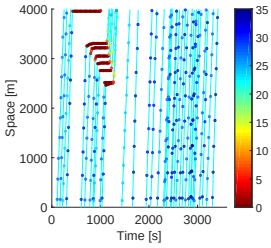
(a) Temporary bottleneck, slant factor = 6150/3600 veh/s



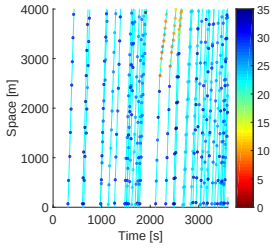
(b) On-ramp, slant factor = 4700/3600 veh/s



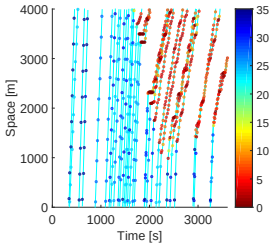
(c) Lane drop, slant factor = 3750/3600 veh/s



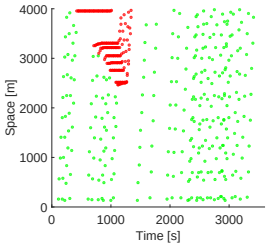
(d) Temporary bottleneck, moving observer trajectories and speeds



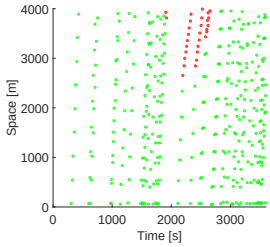
(e) On-ramp, moving observer trajectories and speeds



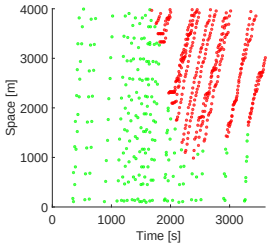
(f) Lane drop, moving observer trajectories and speeds



(g) Temporary bottleneck, traffic phase assignment



(h) On-ramp, traffic phase assignment



(i) Lane drop, traffic phase assignment

Figure 8.5: Slanted downstream boundary cumulative curve, moving observer speeds and observation points traffic phase assignment for the three scenarios.

8.6.1 Traffic conditions and assignment of the restrictive traffic phase to observation points

Prior to exposing the traffic flow properties using wave observations, we want to gain insight in the traffic conditions and assign the restrictive traffic phase to observation points. Gaining insight in the traffic conditions is important to discuss the results in the next sections, and assigning the restrictive traffic phase is needed to be able to independently study the characteristics related to the two branches of Q (the FD).

The slanted cumulative curves (Figures 8.5a, 8.5b and 8.5c) show that the downstream constraints differ for the three scenarios. In the temporary bottleneck scenario, the outflow

is zero during the temporary bottleneck (i.e., from $t = 400$ s till $t = 1000$ s). As shown in Figure 8.5d, this results in a standing queue. The queue discharge rate is observed between $t = 1000$ s and $t = 1600$ s and is approximately equal to the slant factor (= 6150/3600 veh/s). In the on-ramp scenario, a breakdown occurs at approximately $t = 1850$ s, which results in outflow restrictions till approximately $t = 2750$ s. During this period, the link outflow slowly increases after the initial drop at $t = 1850$ s and is on average 4700/3600 veh/s. This results in congested states that have speeds that relate to states between the jam state and queue discharge states observed for the temporary bottleneck scenario, see Figure 8.5e. In the lane-drop scenario, a breakdown is observed at approximately $t = 1700$ s. After this time, the outflow fluctuates between approximately zero and a queue discharge flow. On average, the outflow between $t = 1800$ s and $t = 3600$ s is approximately 3750/3600 veh/s. The fluctuations in link outflow case stop-and-go waves and different congested states on the link, see Figure 8.5f.

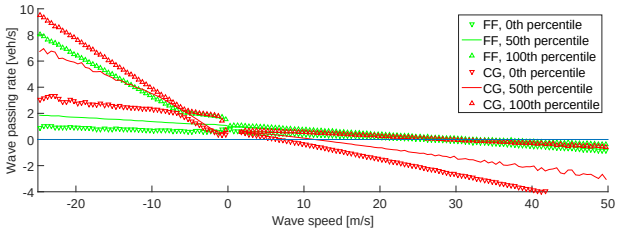
The restrictive traffic phases are assigned to the observation points using a threshold speed of 20 m/s. This results in the division of observation points that is shown in Figures 8.5g, 8.5h and 8.5i. Using different threshold speeds that are close to the selected threshold speed yields small differences at the transition between free-flow and congested states. The points that are effect are thus in a transition state, and often relate the outliers that are observed in the analysis presented below. We do not exclude these outliers in our plots, but one may choose to exclude them when applying our principles to learn the traffic flow model parameters.

8.6.2 Approximation of the branch-specific characteristic wave speeds and passing rates

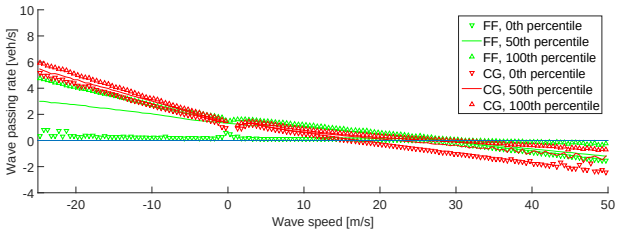
The branch-specific characteristic wave speeds and passing rates are approximated by analyzing the wave observations that related to the free-flow or congested points. Figure 8.6 shows ranges in passing rates for wave speed bin sizes of 0.5 m/s between wave speeds between -25 and 50 m/s. For each of the three scenarios an individual plot is made that contains the ranges related to the points that are estimates to be restricted by a free-flow or congested wave, see Figures 8.6a, 8.6b and 8.6c. Furthermore, the independent branch-specific ranges are plotted for the three scenarios combined, see Figure 8.6d and 8.6e.

For the points that are expected to be restricted by free-flow characteristic waves, we observe evidence of a non-linear FD branch and a relatively wide range in free-flow speeds in low density periods. The narrowing observed in Figure 8.6e lies at approximately $\{w^{\text{ff}}, r^{\text{ff}}\} = \{27, 0.1\}$. A narrowing at a passing rate that is larger than zero indicates that the speed variance is lower at densities that relate to these passing rates in the data set used in this study and that the free-flow FD branch is non-linear. However, if we would have more observations that relate to low demand periods, i.e., density is close to zero, we could not have made this observation, as the lower bounds would lie closer to passing rates of zero for all wave speeds. The range in wave speed for which the lines intersect with the line $q_W^{\text{rel}} = 0$ veh/s, is approximately 27 till 32 m/s, which provides an approximate range of the expected mean speeds in free-flow conditions.

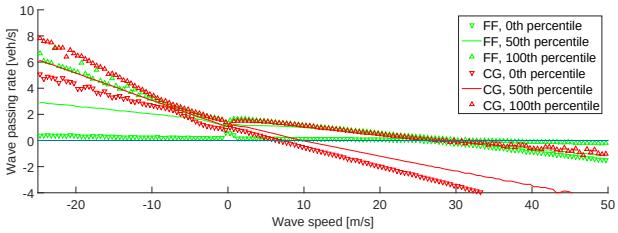
For the points that are expected to be restricted by congested characteristic waves, we observe a much clearer narrowing for the 'temporary bottleneck' scenario than for the other two scenarios, see Figure 8.6a versus Figures 8.6b and 8.6c, and the figure that combines



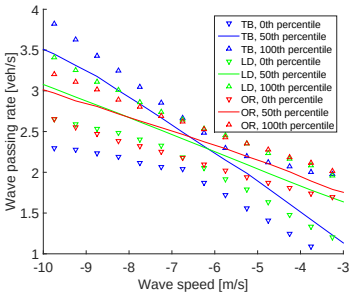
(a) Temporary bottleneck, wave observation ranges



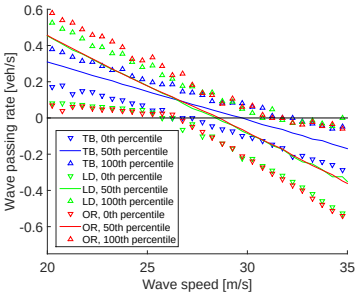
(b) On-ramp, wave observation ranges



(c) Lane drop, wave observation ranges



(d) Congested wave observation ranges for all three scenarios



(e) Free-flow wave observation ranges for all three scenarios

Figure 8.6: Wave observation ranges per scenario, and all scenarios combined for points expected to be restricted by free-flow (FF) or congested (CG) waves. In the legend, the scenarios are abbreviated as follows TB = temporary bottleneck, OR = on-ramp and LD = lane drop.

the congested ranges for all three scenarios (Figure 8.6d). For the 'temporary blockage' scenario, the narrowing is clearly observed as a result of waves fully traveling through the jam state or queue discharge state with different negative speeds. For wave speed lower (more negative) than characteristics congested wave speed(s), waves fully traveling through the jam state and queue discharge state respectively yield the upper bound and lower of the observed wave passing rate. At higher (less negative) wave speed, the waves fully traveling through the jam state yield the lower bound, while those traveling through the queue discharge state yield the upper bound. That these extremes are not observed in the other two scenarios has two reasons. First, as explained above, the congested states occurring in these scenarios are less extreme. Second, the heavy congested states, i.e., (near) jam states, are short and have a relatively constant frequency (in the 'lane drop' scenario). Therefore, waves traveling at speeds that are different than the characteristics congested wave speed(s), are likely to travel through multiple congested traffic states and observe a mean state that lies somewhere in between.

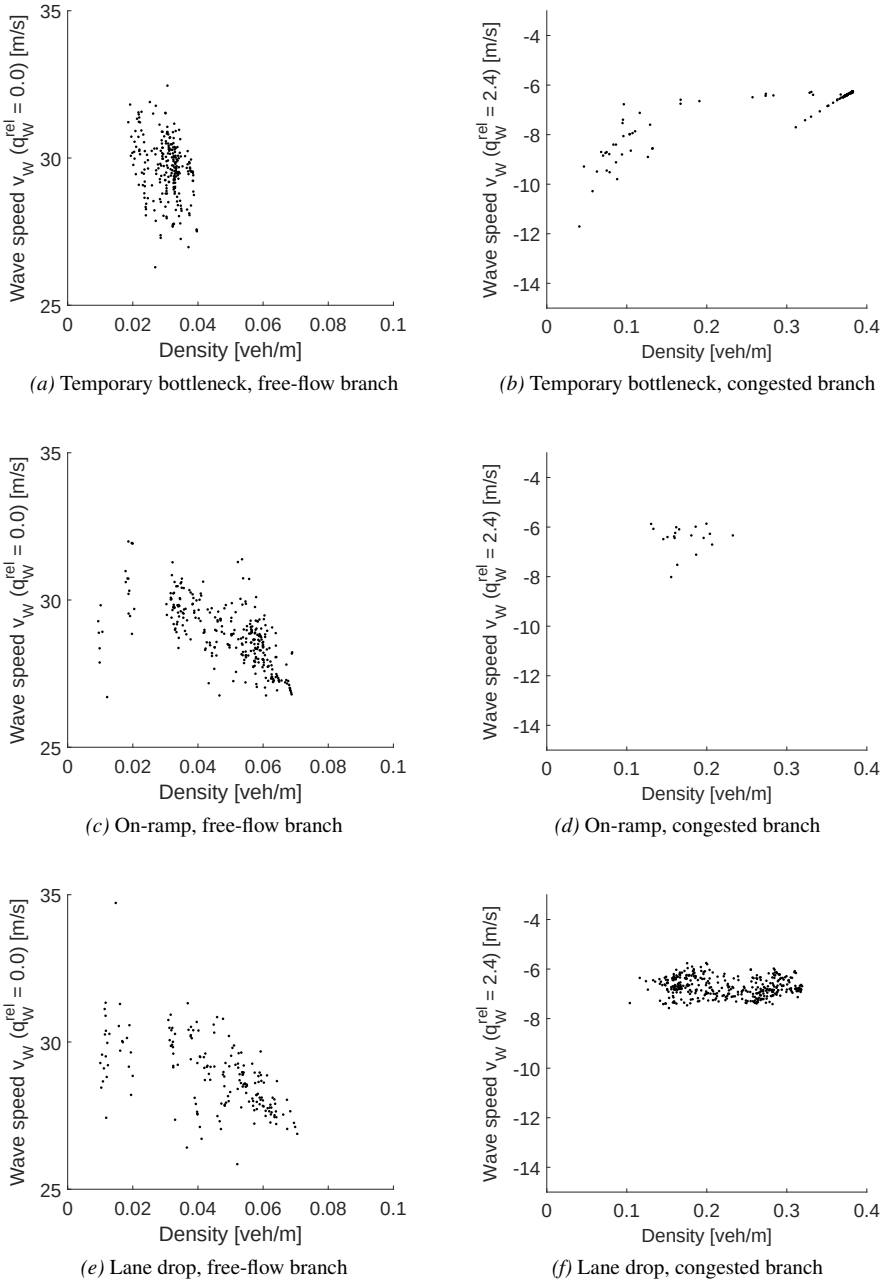
The figure that combines the congested ranges for all three scenarios, i.e., Figure 8.6d, provides an indication that the congested branch is concave, but still allows us to approximate the characteristic wave speeds and passing rates that relate to this branch. The narrowing observed for the 'temporary bottleneck' scenario lies approximately at $\{w^{\mathrm{cg}}, r^{\mathrm{cg}}\} = \{-5.5, 2.2\}$; however, for the other two scenarios higher passing rates are observed at this wave speed and a narrow may be observed at higher wave speeds and passing rates. A reason for observing this pattern may be a concave congested FD branch. In the temporary bottleneck scenario the congested area solely consist of the extremes of the congested branch, i.e., the jam state and queue discharge state, while in the other scenarios congested state between these extremes are observed. In case of a concave congested FD branch, the wave speed between the two extremes of the branch yields a higher passing rate for states that lie in between the extremes, which we observe in Figure 8.6d.

8.6.3 Stochasticity and non-linearity of fundamental diagram branches

As a final step of this study, we asses the stochasticity and non-linearity of the FD branches using Figure 8.7. In this figure, the x-axis is an approximation of the density, which is obtained from taking the derivative of the passing rate to the wave speed for a specified passing rate, see equation (8.19). The y-axis describes the wave speed that relates to a specified passing rate. In line with previous analysis, we make a distinction between points that are expected to be restricted by free-flow and congested waves.

The specified passing rate that is used for points that relate to the free-flow FD branch is zero, as by definition, this is a characteristic passing rate of this branch. To select the specified passing rate related to the congestion branch, we consider Figure 8.6d. As explained above, due to the differences in congested states that occur for the three scenarios, we do not observe a consistent narrowing for all scenarios. Still, to compare scenarios, we select a single specified passing rate. The selected value is 2.4 veh/s, which approximately aligns with the narrowing of all scenarios combined.

Figures 8.7c and 8.7e provide evidence of a concave free-flow branch. This is indicated by decreasing wave speeds for increasing density. As the specified passing rate is 0.0 veh/s, these wave speeds represent the macroscopic variable mean speed. Furthermore, the variance in speeds also reduces with increasing density. These features are not observed for low



(a) Temporary bottleneck, free-flow branch

(b) Temporary bottleneck, congested branch

(c) On-ramp, free-flow branch

(d) On-ramp, congested branch

(e) Lane drop, free-flow branch

(f) Lane drop, congested branch

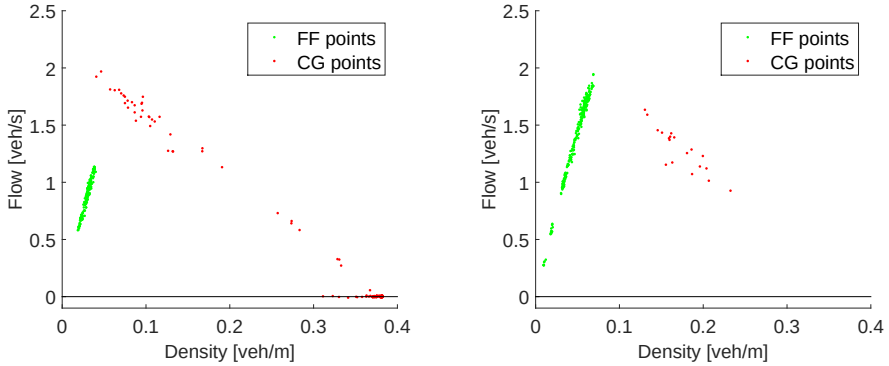
Figure 8.7: Approximated density (x-axis) versus wave speeds that relate to specified passing rates (y-axis) to assess the stochasticity and non-linearity of the fundamental diagram branches.

densities (smaller than 0.04 veh/m) and are therefore also not observed in Figure 8.7a. Such a shape of the free-flow branch may be caused by heterogeneous driving behavior. Heterogeneity in driver behavior (e.g., desired speeds) cause variances in the speed. At higher densities road users that desire to drive at higher speeds are more likely to be restricted by drivers that desire to drive at lower speeds, which causes the mean speed to decrease with increasing density.

Figures 8.7b, 8.7d and 8.7f indicate that the specific passing rate is reasonable for the 'on-ramp' and 'lane-drop' scenarios, but that it is on overestimation for the passing rate that related to the congested states occurring in the 'temporary bottleneck' scenarios. Reducing the specified passing rate moves all point towards lower absolute values of the wave speeds; however, this effect is larger for the points in the lower cloud. If there is a single characteristic passing rate and this one is used to construct the figures, all points will align horizontally. For the temporary bottleneck scenario, we could align the points by the specified passing rate. However, for the other scenarios the points already seem to align horizontally. This is especially clear for the lane drop scenario, see Figure 8.7f.

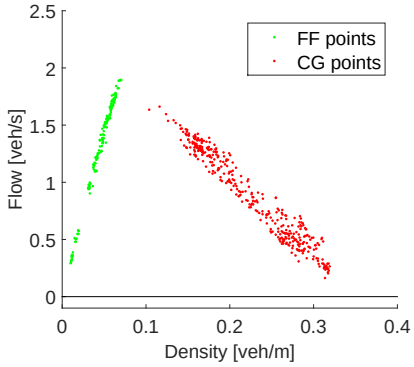
All points shown in Figure 8.7 can be plotted in the flow-density plane by applying equation (8.24), see Figure 8.8. Figures 8.8a, 8.8b and 8.8c show flow-density plots for the individual scenarios, while Figure 8.8d combines all scenarios. In these figures, the estimates that relate to points that are expected to be restricted by free-flow or congested waves are respectively plotted in green and red. The points related both free-flow and congestion could be approximated by a line. However, we also see some flow-density estimates related to congested waves that are expected to be inaccurate. For instance, for the temporary bottleneck scenario, some estimates are located in the left top corner (i.e., have a high flow and low density). As said above, the selected specified passing rate of 2.4 veh/s causes inaccurate estimates for this scenario. Furthermore, the procedure that is used to approximate the derivative in equation (8.19), which is explained in Section 8.5.3, may not be optimal.

The analysis performed in study solely rely on relative flow data obtained from stationary observers positioned at the link boundaries and some moving observers. Based on plots constructed with these data, we gain insight in the traffic flow properties. The plots presented in this study allow us approximate the shape and parameters of the fundamental diagram. Future work can use these insights to design algorithms that can calibrate models using relative flow data. These algorithms may follow an iterative procedure as the approximated model parameters can be used to estimate the restrictive wave for observation points (which is currently done based on individual vehicle speeds), and in assessment of stochasticity and non-linearity (for which we currently select a single specified passing rate per fundamental diagram branch). Designing and implementing such an algorithm may yield a more accurate description of traffic flow properties and is thus an interesting topic to pursue in future research.

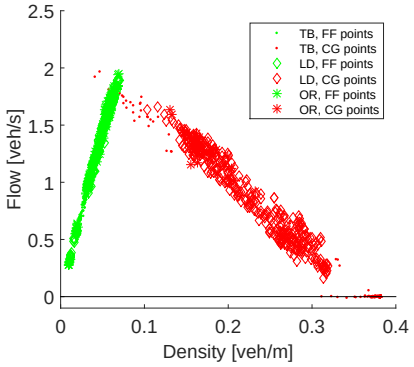


(a) Temporary bottleneck

(b) On-ramp



(c) Lane drop



(d) All scenarios combined

Figure 8.8: Flow-density estimates for the individual scenarios and all scenarios combined. In the legend, the scenarios are abbreviated as follows TB = temporary bottleneck, OR = on-ramp and LD = lane drop.

8.7 Conclusions

In this study we investigate how relative flow data collected with stationary and moving observers can be used to reveal traffic flow properties on freeway links. For this purpose, the data from different observers are fused, which allows us to observe the change in cumulative flow over different wave, i.e., obtain wave observations. Next, these wave observations are used to construct different plots that provide information on the traffic flow properties. To interpret these plots, we turn to Newell's simplified kinematic wave theory and Daganzo's variational theory, and aim to describe traffic flow using the LWR-model with a continuous concave fundamental diagram (FD).

The study shows that relative flow data can be used to reveal traffic flow properties. To test whether these visualization can be used to expose the traffic flow properties, a simulation case study (FOSIM) is conducted. Here, three scenarios are considered, i.e., one-directional (multi-lane) freeway links with three different downstream supply restrictions, i.e., (1) a

temporary blockage, (2) an on-ramp and (3) a lane drop. This case study shows that a limited quantity of data (in this case stationary observers located at the link boundaries together with 1 % of the road users serving as moving observers) suffices to expose the characteristics of both FD branches. However, the ability to do so depends on the presence of different traffic states within the considered space-time domain. For individual scenarios the density varies over time, which results in visualizations that indicate that the free-flow branch is concave and that stochasticity reduces with density. For individual scenarios, the congested states that occur only relate to a limited part of the congested branch. However, combining observations related to the different scenarios (which relate to different parts of the congested branch), provides evidence of a concave congested branch.

Revealing traffic flow properties is valuable to understand why certain traffic conditions occur and how we potentially can prevent or reduce undesirable traffic conditions such as congestion. For this purpose, dynamic traffic management can be applied, in which real-time traffic state estimation is required. In a model-based estimation approach, the insights from this study are valuable to define and calibrate the link traffic flow models that can be used in real-time estimation. In defining the model structure, a trade-off should be made between accuracy and other features that are important in real-time estimation, e.g., computational efficiency.

Relative flow data may also be used to obtain a more detailed description of traffic flow properties. For instance, the data may be used to expose traffic flow properties for multi-directional link, e.g., link with downstream off-ramp. Furthermore, we may distinguish different vehicle classes and describe the multi-class traffic flow properties.

Chapter 9

Road-users participation in sharing personal data directly with road authorities

This chapter is specifically written for this thesis and will be refocused, rewritten and extended for journal publication. It is based on a Master thesis study conducted by A.M. de Jong, which is available in the repository of Delft University of Technology under the name 'Road-user participation in vehicle-data sharing systems'. The topic of that Master thesis and this chapter was originally conceptualized by P.B.C. van Erp (the author of this PhD thesis). Next, A.M. de Jong investigated the topic under the supervision of Prof. dr. ir. C. Chorus, Dr. E.J.E Molin, Dr. H. Asghari and Ir. drs. P.B.C. van Erp.

Chapters 4 till 8 propose methodologies, principles and approaches to use probe trajectory data and relative flow data to estimate the traffic state. These studies and other studies that work with similar data show that these data provide valuable information to estimate the traffic state. Therefore, it can be beneficial for road authorities to gain access to these data and use them for applications such a dynamic traffic management.

Probe trajectory data and relative flow data are personal traffic sensing data as they contain information on individual road-users. As noted in Chapter 3, road authorities may have difficulties to collected these data via third parties due to privacy concerns and legislation. Instead of collecting the data via third parties, road authorities may gain direct access to the personal data from road-users. For this purpose, road authorities can design a data sharing system that meets the privacy legislation that is in place. With this legal bounds, it depends on the road-users' willingness to share their personal data whether road authorities gain access to sufficient data for their applications. To investigate under which conditions the road-users are willing to share their data with road authorities, a stated preference study is conducted.

9.1 Introduction

Dynamic traffic management (DTM) systems can be used to reduce road traffic delays by preventing of postponing the onset of congestion (Hoogendoorn et al., 2013). Examples of DTM systems are ramp-metering (e.g., (Papageorgiou, Hadj-Salem & Blosseville, 1991)), variable speed limits (e.g., (Hegyi et al., 2005)) and dynamic routing advise (e.g., (Wang et al., 2002)). Multiple DTM systems can be combined to obtain an integrated network management (INM) system (Hoogendoorn et al., 2013).

Traffic sensing data are crucial in DTM systems. These data are needed to provide real-time information on the traffic state. Depending on the DTM system, control decisions are based on unprocessed or processed traffic sensing data. For instance, a ramp-metering system (Papageorgiou, Hadj-Salem & Blosseville, 1991) can directly respond on the flows or occupancies that are observed by sensors installed in the road. Alternatively, data can be processed using traffic state estimation (and prediction) methodologies. These methodologies aim to infer traffic state variables (e.g., flow, density and mean speed) based on partially observed and noisy traffic data (Seo, Bayen et al., 2017). By applying these methodologies more information can be extracted from the available data, which in turn may allow road-authorities to collect less data, leading to cost savings.

Traditionally, road authorities rely mostly on road-side sensing equipment to collect the traffic sensing data that is used in DTM. Installing and maintaining these road-side sensors is expensive. Furthermore, these data have technical limitations as they only observe traffic on specific locations and do not provide accurate descriptions of the traffic state in case of slow moving traffic (Knoop & Daamen, 2017).

In addition to or as alternative to road-side equipment, vehicle-based sensing equipment can be used to collect different types of traffic sensing data. For instance, floating car data (FCD), which is also known as probe data, can be collected using smart-phones and in-vehicle systems. These data can be used to estimate the traffic state, e.g., (Nanthawichit et al., 2003), (Herrera & Bayen, 2010), (Van Lint & Hoogendoorn, 2015), (Van Erp, Knoop & Hoogendoorn, 2018a). NDW (the Dutch notational databank for road traffic information) has partnered with BeMobile to obtain FCD travel-time data (Uenk-Telgen, 2018). These data can be used by road authorities in DTM systems. Furthermore, technological advances in vehicle-based sensing equipment allow us to collect new types of data. These data are sometimes referred to as extended FCD (xFCD) (Seo, Bayen et al., 2017). Examples of these data are spacing (headway) data (Seo & Kusakabe, 2015) and relative flow data (Florin & Olariu, 2017), (Van Erp, Knoop & Hoogendoorn, 2018b), (Van Erp et al., 2019), (Takenouchi et al., 2019).

FCD or xFCD can be personal or non-personal data. The data available from NDW (Uenk-Telgen, 2018) and used in (Van Erp, Knoop & Hoogendoorn, 2018a), are aggregated and/or differential-privacy-filtered FCD. These data are thus masked prior to sharing, which is a dominant approach to protect privacy (Cottrill, 2009), (Fries et al., 2012). However, many studies, e.g., (Herrera & Bayen, 2010), (Seo & Kusakabe, 2015), (Van Erp et al., 2019), propose methodologies that require disaggregated data (Seo, Bayen et al., 2017). These data are personal data as they provide information on individual road-users. Although studies show the value of personal data for the purpose of traffic state estimation (which is crucial for DTM), privacy concerns limit the access of road authorities to these data. Therefore, road authorities may be unable to use personal traffic sensing data to increase

throughput, and decrease delays and emissions (which are potential objective of DTM) on road networks.

This study considers the option that road authorities directly collect traffic sensing data from road-users. A direct data-exchange between road-users and road authorities can be beneficial for different reasons. As stated above, privacy legislation (e.g., the EU General Data Protection Regulation (GDPR) (EU, 2016)) and other privacy-related motives may restrict access to the valuable personal traffic sensing data via third-parties. A direct personal data-sharing between road-users and road authorities for the purpose of dynamic traffic management can fall within this legislation. Therefore, such a system can provide access to (otherwise inaccessible) traffic sensing data. Furthermore, in some cases road authorities may already have access to personal traffic sensing data via third parties. In this case, the question arises whether this way of collecting data is legally sound, and if so, whether road authorities have any (additional) moral issues with it. In case of any (potential) legal or moral issues, road authorities may not want to collect the traffic sensing data via third-parties. Also, having an alternative way of gaining access to these traffic sensing data may drive down the data collection cost. This additional alternative may be cheaper than the current alternatives and/or make potential data providers reduce their prices.

The stated preference study presented in this chapter aims to provide initial insights in the road-users' willingness to share personal traffic sensing data with road authorities for the primary objective of dynamic traffic management. Although we do not design the full data-sharing system, the general principles should fall within the general principles of the privacy legislation that is in place, which is the EU General Data Protection Regulation (GDPR) (EU, 2016). In such a system, road authorities should provide clear information on the processing of their data, i.e., it is used for DTM applications. Detailed design of the system should take into account all other elements of the GDRP, but such a design falls outside the scope of this study.

The remaining parts of this article is structured as follows. Section 9.2 presents the stated preference experiment. That section discusses the survey design, data collection and sample representativity. The data are used in Section 9.3 to estimate and interpret a discrete choice model. Next, in Section 9.4 the willingness to accept is determined based on the estimated discrete choice model. And finally, the conclusions and a discussion are provided in Section 9.5.

9.2 Stated preference experiment

This section presents the stated preference experiment. First, the survey design (stated preference experiment) is presented, see Section 9.2.1. Second, in Section 9.2.2, we explain how the survey respondents are reached and evaluate the representativity of the sample with respect to the population (potential Dutch road-users).

9.2.1 Survey design

The survey is targeted at potential Dutch road-users. Therefore, it is chosen to use the local language (Dutch) as this expected to lead to a more representative sample (i.e., we do not want that people do not participate because they are not proficient in English).

The survey starts with an introduction into the survey. This introduction explains what kind of questions respondents will encounter. There are two versions of the introduction that differ in term of a single additional slide. This additional slide describes the potential social benefits of sharing personal traffic sensing data with road authorities (i.e., more effective traffic management is possible which leads to less delays). A part of the respondents is exposed to this slide, which allows us to study the effect of communicating the social benefits to respondents.

The remainder of the survey consist of choice, attitude and socio-demographic questions. Below, we present the survey questions that are used in the analysis conducted in this study. This analysis is kept simple, and focuses on gaining overall insights that are relevant for this thesis and to conduct future research.

Choice questions

The core part of the survey consist of twelve binary choice questions where participants are asked whether they a willing to share their data under a specific set of attributes. The base set, as explained in the survey introduction, is that location data are solely shared per road with the road authorities for dynamic traffic management (DTM) applications without any monetary compensation or control over during which trips data are shared.

Table 9.1 shows all attributes that are part of choice set. These attributes are presented using icons, see Figure 9.1. The first attribute (trip registration) has two options: (1) the position of the road-user is solely tracked per road and not assigned to a personal account or (2) the position of the road-user is assigned to a personal account. In all choice sets one of this two options is presented. For all other attributes the level 0 exists (i.e., No or 0 euro/month), in which case no icon presented.

The first three attributes shown in Table 9.1 relate to the characteristics of the traffic sensing data that are shared. These data characteristics influence which traffic features can be estimated and thereby for which applications the data can be used. Most of the traffic state estimation (TSE) literature focuses on estimating the traffic state on individual links (Seo, Bayen et al., 2017). For this purpose, it suffices to collect traffic data per road. Due to its complexity, the concept of subdividing the road network in links and nodes is not explained to the survey participants; however, the participants are expected to think of a link or multiple connected links and nodes when being presented with the attribute ‘share data per road’. Depending on the proposed methodology, different types of traffic sensing data are needed. For the traditional FCD, which is used in the methodologies proposed by Nanthawichit et al. (2003) and Herrera & Bayen (2010), solely the vehicle position over time (which is part of all choice sets) is needed. Other TSE methodologies require more information, which are captured by the attributes ‘number of passed vehicles’, e.g., (Van Erp et al., 2019), and ‘accelerometer data’. Assigning the trip to a personal account is valuable to gain insight in routing choices (or turn fractions), travel-time between different origins and destinations, and traffic demand (estimate OD-matrices), e.g., (Yang et al., 2017).

The fourth till sixth attributes in Table 9.1 relate to the option that road authorities can share road-users’ data with other parties. It may be beneficial to share the personal data with the public emergency services and researchers, and unnamed commercial third parties. Public parties may use the data for societal benefits, e.g., for emergency services or research purposes. It is explicitly noted that data that are shared with emergency services will not

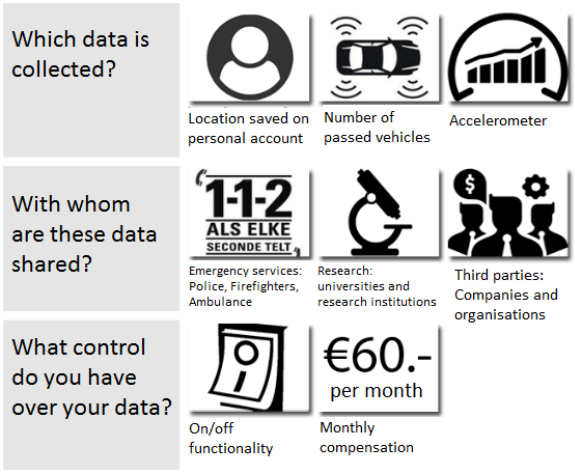


Figure 9.1: Presentation of a choice set to the survey participants. The language used in the survey is Dutch; however, the attribute descriptions are translated in this figure to make it readable for non-Dutch individuals. In this example, the most complete choice set (i.e., choice set 4, see Table 9.2) is shown. For all other choice sets, a part of the attributes is not active. In this case, a grey box is shown instead of the icon.

be used by the police for the purpose of fining drivers that exceed the speed limit. Third parties and their potential applications are not defined to the participant. Sharing potentially identifiable data with unnamed third parties is not compatible with the principles of the GDRP (EU, 2016), but it is included for experimental purposes.

In addition to the data which are shared and with whom it is shared, an attribute controlling for which trips the data are shared is added, i.e., an on/off functionality (where data-sharing must be active during at least 80 % of the trips). Road-users may have reasons for not wanting to share their data all the time. In this case, having the opportunity to selectively share their data may lead to a larger number of participants in the data-sharing system. At the same time, selective data-sharing may may not have a large impact on the ability for road authorities to collect the relevant traffic sensing data (especially if it is distributed over space and time).

The final attribute in Table 9.1 is a monetary compensation per month. The levels of monthly compensation that are included are 0, 20, 40 and 60 euro per month. Offering a monetary compensation to road-users for sharing their data is expected to lead to a larger number number of participants in the data-sharing system. Whether it is needed to offer a monetary compensation depends on the willingness to share data without compensation and the amount of data that is needed for the intended application. The willingness to share data without compensation relates to monthly compensation level of 0 euro per month. As shown in Figure 9.1, this level is presented as a grey box.

Table 9.2 shows the twelve choice situations that are part of the survey. As evaluating all potential combinations of the choice set would lead to a large amount of choice situations, an orthogonal design was made using Ngene.

Table 9.1: Attributes that are part of the choice set.

Factor	Attribute	Levels
Data characteristics	Trip registration	Per road / personal account
Data characteristics	Number of passed vehicles	No / Yes [0 / 1]
Data characteristics	Accelerometer data	No / Yes [0 / 1]
Sharing with other parties	Sharing with emergency services	No / Yes [0 / 1]
Sharing with other parties	Sharing with researchers	No / Yes [0 / 1]
Sharing with other parties	Sharing with third parties	No / Yes [0 / 1]
Control	On/off functionality	No / Yes [0 / 1]
Monetary compensation	Compensation per month	[0 / 20 / 40 / 60 Eur]

Table 9.2: Experimental design, choice situations.

Attribute	Levels	Choice situation											
		1	2	3	4	5	6	7	8	9	10	11	12
Trip registration	0 (per road) / 1 (personal account)	0	1	0	1	0	0	0	1	1	1	0	1
Number of passed vehicles	0 (not present) / 1 (present)	1	0	0	1	0	1	0	0	1	1	1	0
Accelerometer data	0 (not present) / 1 (present)	1	1	0	1	0	0	1	1	0	0	1	0
Sharing with emergency services	0 (not present) / 1 (present)	0	1	1	1	0	1	1	0	1	0	0	0
Sharing with researchers	0 (not present) / 1 (present)	0	1	1	1	1	0	0	0	0	1	1	0
Sharing with third parties	0 (not present) / 1 (present)	1	0	1	1	0	0	1	0	0	1	0	1
On/off functionality	0 (not present) / 1 (present)	1	0	1	1	1	0	0	1	1	0	0	0
Compensation per month	Amount (euro)	0	0	0	60	60	40	60	40	20	40	20	20

Attitude and socio-demographic questions

Once participants completed the binary choice questions, they were presented with three privacy and three trust attitude questions followed by socio-demographic questions. As explained above, we do not use all of these questions for further analysis. We solely consider a part of the socio-demographic questions to assess the representativity of the collected sample for the population, see Section 9.2.2. These socio-demographic factors are presented in Table 9.3.

Table 9.3: Socio-demographic factors used to assess the sample representativity

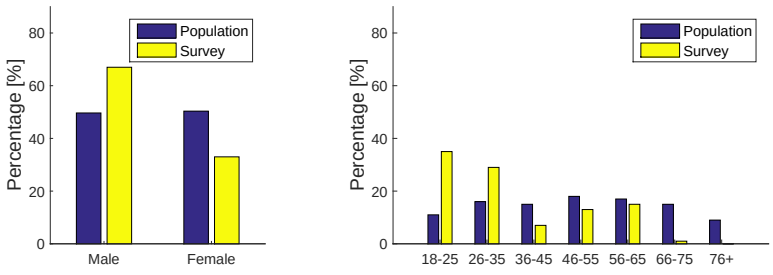
Attribute	Levels
Year of birth	[YYYY]
Gender	[Male / Female / Other]
Education level*	[Basisonderwijs / VMBO / HAVO / VWO / MBO / HBO / WO]

* Explanation on the Dutch education levels: 'Basisonderwijs' is the basic form of education that needs to be followed by everyone. It takes approximately 8 years between ages 4 and 12. After 'basisonderwijs', there are three education levels, which take 4 till 6 years. In increasing difficulty these are VMBO (4 years), HAVO (5 years) and VWO (6 years). These education level respectively qualify students for MBO, HBO and WO. Completing HBO yields a bachelor degree and completing WO yields a bachelor and a master degree.

9.2.2 Data collection and sample representativity

Two data collection approaches were applied to reach respondents. First, the link to the online survey was shared via mail and LinkedIn with the professional and social networks of the researchers involved in this study. Second, surveys were collected using a tablet at several public locations in Delft. These two approaches let to 128 completed surveys. Out of these surveys, 28 respondents stated that do not have access to a car. Their completed surveys are not used for further analysis as they are not representative for Dutch road-users. Therefore, 100 completed surveys are used in the evaluations presented below.

Figure 9.2 shows differences between the Dutch population aged 15 and older, and the survey sample. The population data are retrieved from StatLine, which is the online database of the Dutch Central Bureau of Statistics (CBS). In the comparison, we selected the age groups that are available in StatLine and that contain potential road-users. The comparisons for gender and age are respectively shown in Figures 9.2a and 9.2b. A comparison for education level is also made, but not visualized as the levels used in the survey (see Table 9.3) and StatLine do not completely overlap. This comparison shows that our sample is relatively highly educated, as 83 % of the respondents has either HBO or WO (see description below Table 9.3 for a explanation on the Dutch education levels) as the highest level of education against on 30 % of the Dutch population of 15 and older. Furthermore, the comparisons in Figure 9.2 show that there are biases gender (relatively large amount of male respondents)



(a) Gender *(b)* Age

Figure 9.2: Comparison between the survey sample and the Dutch population aged 15 and older (CBS StatLine, 2019).

and age (respondents are relatively young). Due to these biases the results cannot be generalized to the whole population, which should be taken into account in the discussion and interpretation of the results. Another potential bias lies in the participation in the stated preference experiment. By participating in the experiment, the participant show a willingness to share data for research purposes. However, the results can still provide valuable (initial) insights related to the road-users' preferences related to sharing their personal traffic data with road authorities for the purpose of dynamic traffic management. Furthermore, in future research a more extensive and representative sample may be collected and studied. In such future studies, we may use the survey presented in this study and/or improve the survey design based on the finding presented in this study.

9.3 Discrete choice model

In this section, the procedure to estimate and the estimated discrete choice model are presented.

9.3.1 Methodology

In the survey respondents were offered the binary choice to participate (1) or not to participate (0) in the data sharing scheme. Random utility theory (Manski, 1977) is considered to evaluate the binary choices of respondents. In this theory, it is assumed that users are rational decision-makers, but have incomplete information. Following Ben-Akiva & Bierlaire (1999), utility U is modeled as a random variable, where the utility related to participating in the data sharing scheme in case of choice set C_i is given by:

$$U_i = V_i + \varepsilon_i \tag{9.1}$$

where V_i and ε_i respectively denote deterministic and random utility terms. The random utility term describes the 'incomplete information' part and captures the uncertainty Ben-

Akiva & Bierlaire (1999). The deterministic utility term is modelled as the sum of the product of the attributes x_{ik} times attribute-specific coefficients β_k.

$$V_i = \sum_k \beta_k x_{ik} \tag{9.2}$$

For this binary choice scenario, the influence of different attributes on the respondents' choices can be modeled using a binary logit model. This model aims to describe the probability that respondents state that they are willing to participate in the data-sharing scheme given the set of attributes $\mathbf{x}_i$, which is denoted by $P(S_i = 1|\mathbf{x}_i)$.

$$P(S_i = 1|\mathbf{x}_i) = \frac{e^{V_i}}{1 + e^{V_i}} \tag{9.3}$$

The set $\mathbf{x}_i$ contains the eight attributes presented in Table 9.1 plus a constant and an attribute describing whether respondents were shown the additional introductory slides explaining the social benefits of using their data for dynamic traffic management applications. Only one attribute is a non-binary attribute, i.e., *Compensation per month*, which can take the values 0, 20, 40 or 60 euros per month. The coefficients related to the binary attributes describe the effect of including that attribute on the probability that road-users are willing to share their personal data. A positive or negative coefficient respectively indicates that including the attribute leads to an increase or decrease in the data sharing probability. To describe the (level of) significance of the coefficients, the t-test is applied, which yields a probability p that the coefficient has a specific sign (i.e., is positive or negative).

As our sample is not representative for the population (potential Dutch drivers), see Section 9.2.2, we will keep our analysis simple and will not generalize the findings to the whole population. In the analysis, the focus will lie on assessing which coefficients are significant, which signs they have (positive or negative) and how the order of magnitudes relate to each other.

9.3.2 Results

Table 9.4 shows the coefficients and their significance for the estimated binary logit model. A positive coefficient in Table 9.4 indicates a positive effect on the willingness to share data. To describe the (level of) significance of the coefficients in p, we distinguish three levels, i.e., $p < 0.05$, $p < 0.01$ and $p < 0.001$, which are respectively indicated by 1 to 3 asterisks (*, ** and ***) behind the coefficient.

Table 9.4: Estimated binary logit model.

Attribute	Coefficient	Standard error
Constant	1.011***	(0.225)
Trip registration	-0.676***	(0.147)
Number of passed vehicles	-0.044	(0.146)
Accelerometer data	-0.243	(0.146)
Sharing with emergency services	0.294*	(0.146)
Sharing with researchers	0.396***	(0.146)
Sharing with third parties	-2.459***	(0.150)
On/off functionality	0.314*	(0.143)
Monthly compensation	0.019***	(0.003)
Communication of social benefits	-0.057	(0.139)

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The *constant* relates to case in which respondents are willing to share the basic form of data with the sole application of dynamic traffic management: only location data are shared, trajectories are registered per road and not linked to a personal account, the data is not shared with any other parties, there is no on/off function, no monthly (monetary) compensation and no communication of benefits. The positive and highly significant coefficient indicates that the respondents have an estimated probability of 73 % of stating that they are willing to share the basic form of data with the sole application of dynamic traffic management. This estimated probability is determined using equations (9.2) and (9.3).

The three attributes that relate to different traffic sensing data characteristics (i.e., *trip registration*, *number of vehicles passed* and *accelerometer data*) all have a negative coefficient. Negative coefficient could be expected as road-users are asked to same more detailed information related to their personal behavior. Out of the three attributes, only *trip registration* is significant. However, this absolute value of this coefficient is smaller than the constant, which indicates that the respondents are still more likely to participate in the proposed data-sharing system if all of their trips are assigned to one account, without on/off functionality, monthly compensation and sharing with other parties. It is also interesting to note that respondents do not seem to attribute a large negative utility to sharing the *number of vehicles passed*, i.e., sharing relative flow data. This is interesting as we investigated the use of these data and explained that these data can be valuable in macroscopic traffic state estimation (Van Erp, Knoop & Hoogendoorn, 2018b), (Van Erp et al., 2019).

Sharing with other parties is valued in different ways. *Sharing with emergency services* and *sharing with researchers* have positive and significant coefficients, i.e., respondents are more willing to share their personal data when it is shared with (and used by) emergency services and research institutions. This aligns with the positive constant, which indicates that the respondents have a natural preference for sharing the with road-authorities for the purpose of dynamic traffic management. *Sharing with unnamed third parties* is assigned a highly negative and highly significant coefficient. This indicates that road-users assign a large dis-utility to giving away control over their data as the third parties and the use of data

were not defined in the survey. Control over which personal data are shared is possible by having an *on/off function*. The coefficient related to this attribute is positive and significant, which means that respondents are more likely to state that they will participate in the data-sharing system if they are given control over the trips for which they share their data.

Monetary incentives for data sharing is the form of a *compensation per month* is highly significant and positive. This indicates that road-authorities may benefit from offering a monetary compensation in order to gain access to personal traffic sensing data.

Communication of social benefits was done using an additional introductory slide related to this topic. The coefficient related to this attribute is negative, but not significant. Other coefficients, i.e., the constant, and those related to sharing data with emergency services and research institutions, indicate that respondents are willing to share their personal data in they contribute to social benefits. This may indicate that the social benefits of sharing their data are already clear to respondents based on prior information.

9.4 Willingness to accept

To gain a better insight in the estimated model coefficients, we will estimate the willing to accept (WTA). The WTA is defined as 'the minimum amount an individual would be willing to accept to forgo it' (Hanemann, 1991). In this study, 'it' can be one of the attributes that road-users that give up. For instance, they can give up 'not sharing their accelerometer data' or 'not having an on/off functionality'.

We can estimate the WTA based on the estimated attribute coefficients β. The coefficient that relates to a monetary compensation per month provides a monetary proxy that can be used to estimate the WTA. By taking the negative quotient of an attribute β_k and the monetary coefficient β_m, we obtain the WTA related to that attribute WTA_k:

$$WTA_k = -\frac{\beta_k}{\beta_m} \tag{9.4}$$

The unit of the WTA is the same as the unit of the monetary attribute, i.e., euro/month.

Table 9.5 presents the estimated WTA for all attributes presented in Table 9.1 except for the monetary attribute. A positive WTA indicates that respondent require a monetary compensation in order for the attribute to be active in the choice set.

As shown in equation (9.2) the influence of all attributes is assumed to be independent and a linear relation is assumed to describe the change in utility given a change in monetary compensation per month. The extent to which these assumptions are correct is not taken into account in the WTA. Since the offered monetary compensation in the survey ranges between 0 and 60 euro/month, it is not possible to assess whether a linear relation between utility and monetary compensation holds for all potential monetary compensations. For instance, road-users may respond differently to losses (having to pay) than gains (obtaining a monetary compensation). For this reason, the WTA is mainly considered to gain a rough insight in monetary equivalents on the coefficients presented in Table 9.4, which are easier to interpret than the coefficients.

Table 9.5: Estimated willingness to accept (WTA) for the attributes included in the choice set.

Attribute	WTA [euro/month]
Trip registration	35.07
Number of passed vehicles	2.29
Accelerometer data	12.61
Sharing with emergency services	-15.23
Sharing with researchers	-20.56
Sharing with third parties	127.56
On/off functionality	-16.28

9.5 Conclusions and discussion

Road authorities require traffic sensing data for dynamic traffic management applications. Personal traffic sensing data (e.g., probe trajectory data or relative flow data), which are often collected by commercial companies, can provide valuable information for this application. However, due to privacy concerns and legislation, these personal data are not always accessible to road authorities. Instead of gaining access to these personal data via commercial companies, road authorities may get direct access to personal traffic sensing data from road-users. In this way road authorities and road-users may make explicit agreements related to the data which is shared and for which applications it is used. To gain an initial insight in the road-users' preferences related to the willingness to share personal traffic sensing data with road-users, a stated preference survey is conducted.

The analysis of the stated preference experiment is kept simple, but still provides interesting insights. To model the binary choice of sharing data (1) or not sharing data (0), a binary logit model is estimated. This model includes a constant and nine attributes that describe the data characteristics, with whom the data needs to be shared, a control attribute, a monetary compensation attribute and whether respondents were exposed to an additional introductory slide that describes the social benefits of sharing data with road authorities. Seven out of the ten coefficient were found to be significant ($p < 0.05$). The constant is significant and positive, which indicates that the respondents have a preference to opt for sharing the basic form of personal data (i.e., location and time per road) with road authorities without any compensation or control. This indicates that respondents may be willing to share their personal data because of the social benefit of improving traffic efficiency and safety by means of dynamic traffic management. The coefficients related to sharing data with emergency services and researchers are also positive and significant, which again points towards that the dis-utility of sharing personal data is out-weight by the utility of contributing to social benefits. Two coefficients are negative and significant, i.e., sharing data with unnamed third parties and register all trips to a personal account (which allows road authorities to track road-users during their complete trip). By sharing data with unnamed third parties, road-users give away control of which parties have access to their data and for which purposes they are used. It thus seems that the respondents value being in control of

their data. This is also indicated by the positive and significant control variable 'on/off functionality', which allows road-users to control for which trips their data are shared. Finally, the coefficient related to giving road-users a monthly monetary compensation is positive and significant, which indicates that it road authorities may gain access to more personal traffic sensing data when offering monetary incentives.

The stated preference study provides insight into the conditions under which road-users are willing to share their personal traffic sensing data with road authorities for dynamic traffic management purposes. However, the study is limited in terms of number of respondents and representativeness for the population (potential Dutch road-users). Furthermore, biases may occur because respondents that were willing to participate in the stated preference experiment respondents were by definition willing to share their data for research purposes. Therefore, the survey responses should be interpreted with care and should not be generalized to the whole population.

Even-though the survey responses should not be generalized to the whole population, the study points towards some interesting insights for the (Dutch) road authorities. It is for instance interesting that the sample used to estimate the binary logit model shows a natural preference to share the basic data (i.e., position over time per road segment) with road authorities for the purpose of dynamic traffic management. Prior studies show that data from only a small fraction of road-users is needed to accurately estimate the traffic state (e.g., (Herrera & Bayen, 2010) and (Van Erp, Knoop & Hoogendoorn, 2018b)). Therefore, the stated preference of a sub-sample of the road-users to share their personal data with road authorities points towards the possibility that road authorities may already gain access to sufficient personal data without needing to offer monetary compensation to road-users for sharing these data.

This study should be seen as a first step in exploring the option that road authorities directly collect traffic sensing data from road-users. The conclusions that can be drawn from our and any stated preference study can be questioned, as there can be large difference between saying that you are willing to share your data and that you actually share your data. It may therefore be better to move from academics to practice in order to explore this option in more detail. For this purpose, road authorities should follow a lean approach to develop and test the proposed system in practice.

Chapter 10

Conclusions and outlook

This section presents the conclusions and an outlook. To present the conclusions, Section 10.1 provides the findings and conclusions. Here, first, each of the eight research questions (which relate to Chapters 2 till 9) is answered individually and then the overall conclusions are presented. Sections 10.2 and 10.3 respectively discuss the 'implications for practice' and 'future research and implications for science'. These two sections provide an outlook for potential implications and future research that relates to the studies presented in this thesis.

10.1 Findings and conclusions

Below, we will first provide answers to the individual research questions. Next, the overall conclusions of this thesis are presented.

10.1.1 Findings

Research question 1: What causes estimation errors when estimating Edie's mean speeds based on error-free probe or loop-detector data?

To answer this question, Edie's mean speed is estimated for rectangular (i.e., time periods and road segments) spatial-temporal areas based on lane-specific detector time-mean speed and flow data (the standard on Dutch freeways), and individual probe speed data. To study the errors that are induced in the step from traffic sensing data to mean speed estimates, we work with error-free data. We show that, even with error-free data, estimation errors arise due to incomplete information and incorrect assumptions. This is relevant because the estimates that are evaluated in the study related to this research question can be presented as data, e.g., probe mean speed data. It is important to understand that data may be based on incomplete information and incorrect assumptions, which in turn limits the reliability of these data. In turn this is important when comparing the value of different types of traffic sensing data.

Research question 2: How can probe mean speed data be used to mitigate the cumulative error problem in on-ramp vehicle accumulation estimation?

Probe mean speed data can be used (1) to recover the cumulative error that has build up in previous periods and (2) to learn the structural detector count error. Both processes are valuable in mitigation of the cumulative error problem that arises in recursive estimation of the vehicle accumulation based on detector count data. We show that using probe mean speed data for both processes yields the most accurate vehicle accumulation estimates and is thus most effective for mitigation of the cumulative error problem. However, we also show that the data provide incomplete information on the cumulative flow curves at the on-ramp boundary. For instance, the combined upstream and downstream detector structural count error can be exposed using probe mean speed data. However, as shown in the empirical case study, it does not allow us to find the structural error for the individual detectors. Therefore, we do not know whether the reference vehicle accumulation estimates (which are obtained by fusing detector count and probe mean speed data) have a structural error. A structural error in the reference estimates affects the recursive on-ramp vehicle accumulation estimates through the error recovery process, which thereby reduces the estimation accuracy. Despite this data issue, mitigation of the cumulative error problem using probe data is still valuable as it yields feasible estimates and the variation in the estimates (e.g., peaks in vehicle accumulation) can still be distinguished.

Research question 3: How can disaggregated detector data be used to estimate the change in cumulative flow over probe trajectories between detector locations?

Disaggregated detector data (which describe the speed and time related to individual detector passings) allow us to estimate the relative flow with respect to probe vehicles that pass the detector locations. Based on the probe-specific relative flow estimates related to two consecutive detectors, the change in cumulative flow over probe trajectories between detector locations can be estimated. We show that this approach is a clear improvement over assuming 'no overtaking' (i.e., assuming that the change in cumulative flow over a probe trajectory is zero) in free-flow conditions. In congested conditions the advantages are less clear. As probe trajectory data provides information on the encountered traffic conditions, we can use this information to assign an error variance to the change in cumulative flow estimates. Improving upon the 'no overtaking' assumption is valuable for mitigating the cumulative error problem with probe trajectory data. However, as the changes in cumulative flow are not perfectly estimated, it would be better to observe this variable, i.e., collect relative flow data with moving observers.

Research question 4: How can we estimate flow and density by solely relying on real-time relative flow data?

To answer this research, we consider the option that relative flow data are collected with stationary observers positioned at the link boundaries and with a fraction of the road-users that serve as moving observers. The considered combination of observers is beneficial as all relative flow observations can be related to each other. Therefore, the data provide information on the change in cumulative flow between points in space-time that lie on the same or

different observation paths. The study shows that a combination of three points for which the change in cumulative flow, space and time are known, allow us to estimate flow and density for triangular areas. These triangular areas can then be mapped to any estimation mesh and thereby allow us to estimate flow and density for the independent spatial-temporal areas that are part of the estimation mesh. In a simulation case study, it is shown that the proposed methodology and data can outperform loop-detector data starting at low penetration rates, i.e., between 2.5 and 5.0 %. Furthermore, it is shown that relative flow observation errors do not have a major negative effect on the estimation accuracy. As the methodology solely relies on real-time data (i.e., it follows a streaming-data-driven estimation approach), it shows that relative flow data collected with stationary and moving observers alone provide sufficient information to accurately estimate the traffic state. Additional information, e.g., in the form of a traffic flow model, is thus not needed. However, as the data do not contain complete information on the traffic features that are estimated, including additional information may still improve upon the current estimation performance.

Research question 5: How can historical relative flow data and Edie's generalized definitions of traffic flow be used to estimate the fundamental diagram of traffic flow?

Edie's definitions can be applied to estimate flow and density for areas that are enclosed by relative flow observation paths of moving observers. These enclosed areas take different and uncontrollable shapes as they depend on the observation paths that are available. Therefore, the physical meaning of these estimates are difficult to interpret and the applications may seem limited. However, we show that these estimates can be used to estimate the fundamental diagram of traffic flow. The proposed methodology takes into account that different traffic states can be present within the enclosed areas and for continuous concave fundamental diagrams the combined state is expected to lie on or below the fundamental diagram. This study shows that relative flow data yield valuable information to estimate the fundamental diagram in the flow-density plane. This procedure works well for the non-stochastic car-following models considered in the study. However, these methodology should be extended to deal with more complex traffic flow properties.

Research question 6: What spatial-temporal relative flow data characteristics are desired in model-based traffic state estimation?

To analyze the value of relative flow data in model-based traffic state estimation, we propose principles that can be used (1) to learn the traffic flow model based on historical relative flow data and (2) to estimate the traffic state by fusing a calibrated traffic flow model and relative flow data. For this purpose, it is assumed that traffic flow is described by the LWR-model with triangular fundamental diagram. We show that having observers that travel at different speeds and that observe the change in cumulative flow are valuable for both processes. Having access to observers that travel at different speeds is beneficial as this allows for intersecting observation paths. Furthermore, the study shows that differences between the observer speeds and characteristics speeds of the traffic flow model are beneficial as affects the size of the space-time domain for which the characteristics waves originating from the observation path can be drawn. Again, different observers speeds are desired such that the different characteristic waves, e.g., the free-flow and wave speed for the LWR-model with

triangular fundamental diagram, can be drawn for large (overlapping) space-time domains. Using a combination of stationary and moving observers is beneficial as the observers will cross each other in space-time and the link boundaries can be used using the stationary observers. If we want to rely on moving observers alone (e.g., for financial reasons), it is valuable that the moving observers observe that own and opposing direction of traffic.

Research question 7: How can relative flow data from stationary and moving observers be used to reveal traffic flow properties?

Relative flow data collected using stationary and moving observers can be used to obtain wave observations. These wave observations describe the combination of the change in time, change in space and change in cumulative flow between two points in space-time. Variational theory depicts which conditions have to be satisfied by these wave observations if traffic flow can be described by the LWR-model with continuous concave fundamental diagram. This theory allows us to interpret different plots that can be created using the wave observations. Visualizing the wave observations in the 'wave speed - wave passing rate' plane allows us to approximate the parameters related to the free-flow and congested branches of the fundamental diagram. However, this plot does not allow us to make a distinction between non-linearity of the branch and stochasticity. To gain more insight in these features, the density along a wave and wave speed related to a specified passing rate can be visualized and analyzed. Finally, based on the density, wave speed and passing rate, flow can be estimated. This allows us to plot combined estimates of flow and density. In a simulation study, we find that the mean speed decreases with density in free-flow conditions, which indicates that the free-flow branch of the fundamental diagram should be a non-linear concave function. Furthermore, there is evidence of a concave congested branch.

Research question 8: Under which conditions are road-users willing to share their personal traffic sensing data with road authorities for dynamic traffic management purposes?

To answer this question a stated preference study is conducted. The collected sample in not representative for the population (Dutch road-users) as it is biased towards younger highly educated males. However, it still provides valuable insights. This study indicates that the respondents have a natural preference to contribute to traffic management applications by sharing the most basic form of personal data (i.e., trajectory data per road segment) and that they do not have a significant dis-utility from sharing relative flow data. The preferences related to allowing road authorities to share the road-users' data with other parties differ per party. Sharing data with emergency services and/or research institutions is valued positively (i.e., it makes respondents more willing to share their personal data with road authorities), while sharing data with unnamed third parties is valued highly negative. Offering road-users to option to selectively share their data (on/off option) and giving road-users financial compensation for their data are valued positively by the respondents. However, if a system is created for direct sharing of personal traffic sensing data from road-users to road authorities, offering financial compensation may not be needed as many road-users may already be willing to share their data without this compensation.

10.1.2 Overall conclusions

The studies conducted in this thesis show that relative flow data have preferable characteristics for traffic state estimation compared to the other traffic sensing data. The main advantage of relative flow data lies in the observed variable and spatial-temporal data characteristics. The variable, i.e., relative flow, describes the core macroscopic traffic flow variable, i.e., the change in cumulative flow. In contrast to data that describe the traffic state for spatial-temporal areas it is not subjected to incomplete information and irreversible compressed (i.e., it is possible to determine Edie's traffic states based on cumulative flow curves over the area boundaries, but not possible to reverse this process). Stationary observers (e.g., loop-detectors) are already widely used to collected relative flow data with respect to a fixed position over time. However, if we solely rely on these existing detectors and if we want to estimate the traffic state in or via the cumulative flow plane, estimates will be highly inaccurate due to the cumulative error problem. We show that existing types of probe data (i.e., probe mean speed and probe trajectory data) can be used to mitigate the cumulative error problem. However, these data also have their limitation in mitigating this problem: (1) probe mean speed data provide incomplete information on the cumulative flow curves at on-ramp boundaries and (2) using disaggregated detector data to estimate the change in cumulative flow along probe trajectories is an improvement over assuming 'no overtaking', but it does not yield perfect estimates. These limitations can be overcome by using automated and/or other connected and equipped vehicles to collect relative flow data. Such vehicles can serve as moving observers that are part of the traffic flow, or travel on opposing or parallel roads. Combining the existing stationary observers with moving observers or relying on moving observers alone, opens up multiple opportunities to estimate the traffic state and expose traffic flow properties (which in turn can be valuable for state estimation). In four studies, we show that: (1) Relative flow data collected using stationary observers positioned at the link boundaries and a limited number of moving observers allow us to accurately estimate Edie's flow and density. (2) Relative flow data from moving observers traveling in different direction allow us to determine Edie's flow and density for certain spatial-temporal areas, which in turn can be used to estimate the fundamental diagram of traffic flow in the flow-density plane. (3) For both processes in a model-based estimation approach, i.e., learning models and information assimilation, it is favorable to collect relative flow data with observers that travel at different speeds (i.e., stationary and moving observers or moving observers alone that travel at different speeds). (4) Traffic flow properties can be revealed by constructing different plots based on relative flow data from stationary observers positioned at the link boundaries and a limited number of moving observers.

In addition to analyzing the suitability of the characteristics of different data-types to obtain traffic state information, we propose the option that road authorities directly go to the road-users to gain access to personal traffic sensing data. Different types of probe data and relative flow data fall in this category, and may therefore be difficult to obtain via third parties. The stated preference study conducted on this topic indicates that road-users have a natural preference to contribute to traffic management applications by sharing the most basic form of personal data (i.e., trajectory data per road segment) and that they do not have a significant dis-utility from sharing relative flow data. Therefore, setting up a system for direct exchange of personal traffic sensing data from road-users to road authorities, is an idea that is worth further research.

10.2 Implications for practice

The most important implications for practice are: Road authorities and/or commercial companies (1) can use the methodologies designed during this thesis in practice, (2) with access to a fleet of automated or other equipped and connected vehicles can investigate collecting and using relative flow data, and (3) should take the potential uses of relative flow data into account when making decisions related to future data collection and processing. These three implications for practice are consecutively discussed in the next three paragraphs.

Chapters 3 to 8 (partially) design methodologies that can be used to estimate the traffic state and/or learn traffic flow models. These methodologies may be used in practice or offer the basis for future methodologies. The methodology designed to estimate the vehicle accumulation (see Chapter 3) works with readily-available data and is therefore closest to a practical application. However, prior to using the methodology in practice, a more extensive validation using real-data and comparison to the alternatives should be performed.

Relative flow data may be collected using automated and/or other equipped and connected vehicles. However, access to the sensing and processing equipment in such vehicles is needed to have the possibility to collect these data. Commercial car companies, e.g., Waymo, Tesla and Mercedes, drive the development of automated or other equipped and connected vehicles, and are thus the most likely players to potentially collect relative flow data. Furthermore, these companies desire traffic state information for routing (and potentially other) purposes or to sell traffic state information to other players (e.g., road authorities). This should offer sufficient incentives for the commercial car companies to investigate the option to collect relative flow data with their vehicles and further develop the methodologies to use these data to obtain valuable traffic state information.

In time, obtaining traffic state information using relative flow data may become an alternative to existing approaches to obtain traffic state information. Nowadays, road authorities heavily rely on loop-detectors to collect traffic sensing data to design and manage the road infrastructure and traffic. Installing and maintaining these detectors on a wide-scale requires large investments and requires temporary road closures. Relative flow data are collected using sensing equipment that has a different primary function (i.e., automated vehicle operational control). Therefore, the marginal costs of collecting these data are expected to be limited. If road authorities can obtain these data at limited costs (either directly from road-users, see Chapter 9, or via private companies), the data may be financially a good alternative to loop-detector and other data. Furthermore, as show in the studies presented in this thesis, relative flow data offer technical advantages over relying on the stationary loop-detectors alone. Although relative flow data are not yet available on a wide scale, they may be within the lifetime of loop-detectors that are installed in the near future. To prevent unnecessary investments, road authorities should investigate when relative flow data and/or other data become a good alternative or addition to loop-detector data. This requires them to investigate when and at what costs the data alternatives are likely to become available, and how these data can be used for the applications related to design and management of the infrastructure and traffic.

10.3 Future research and implications for science

This thesis shows that relative flow data offer theoretically preferable characteristics for traffic state estimation. For this purpose, different potential approaches to use relative flow data within the field of traffic state estimation were studied. These studies focused on estimating the traffic state for links (i.e., road segments without discontinuities) and exposing the traffic flow properties (which can be described using link traffic flow models such as the LWR-model with specific fundamental diagrams). Also, in Chapter 3, we presented a methodology to learn the structural detector count error based on probe mean speed data. Furthermore, in Chapter 7, we state that it may be beneficial to use information that is available in a network of links and nodes to estimate the traffic state. This information can propagate over nodes from one node to other nodes. We may use node traffic flow models to describe the properties of traffic flow over nodes, which in turn may be exposed by levering relative flow data and link traffic flow models.

The important implications for science are that, in our view, future studies related to traffic state estimation should focus more on using relative flow data. Although these data are not yet collected on a wide scale, this study shows that the data have theoretically preferable characteristics for traffic state estimation. More research is needed to explore the opportunities these data offer and proof to practitioners that it is valuable to collect and use these data.

Below, potential future research directions are presented. Combining these directions works in favor of designing network-wide traffic state estimation methodologies and using such estimates for traffic management purposes. However, the topics and proposed directions are also interesting as independent studies.

- **(Online) learning of link traffic flow properties**: Chapters 6, 7 and 8 describe different approaches to gain insight in the link traffic flow properties. The latter two chapters explain that relative flow data can be used to expose these properties in the cumulative flow plane by using observations of the change in cumulative flow over different waves. Within this field there are multiple future research directions. For instance, the studies presented in Chapters 7 and 8 show which information the relative flow data provide on the link traffic flow properties, but they do not present methodologies for online learning of models that describe these properties. Such methodologies may be developed in future research. Furthermore, assumption related to the description of traffic flow and models used to describe the traffic flow properties may be relaxed. For instance, instead of describing traffic flow as one commodity (which can be a valid assumption for some parts of the road network), it may be described using multiple commodities (e.g., based on a downstream direction). Furthermore, we may exploit techniques brought forward by the machine learning domain (e.g., neural networks) to describe the propagation of traffic states over space-time in the cumulative flow plane. Whether such an approach is a good alternative to the LWR-model with a specific fundamental diagram can be explored in future research.
- **(Online) learning of node traffic flow properties**: As stated above, it may be valuable expose the node traffic flow properties and use these to describe the information propagation over nodes. In Chapter 7, we explained that relative flow data combined

with a link traffic flow model can be used to estimate the link boundary flows, supply and demand. Historical estimates that describe these features for all inflowing and outflow links related to a node may be used to expose the node traffic flow properties.

- **(Online) learning of observation error characteristics**: In Chapter 3, we present a methodology to learn the structural detector count error based on probe mean speed data. The general concept of learning the observation error characteristics based on data, can be extended to applications of relative flow data. Collecting relative flow data with moving observers that travel at different speeds creates a system of 'checks and balances'. The observation paths (i.e., the paths in space-time over which we observe the relative flow) can enclose spatial-temporal areas. The physical condition of conservation-of-vehicles states that the area inflow should equal the area outflow, i.e., the net flow should be equal to zero. This condition allows us to correct observation errors, but also to expose the observation error characteristics (e.g., bias and error variance), which in turn may be used in traffic state estimation.
- **(Online) learning of traffic demand (origin-destination estimation) and route choice behavior**: Accurate traffic state in a network of links and nodes may require less data than accurate traffic state estimation for all individual links. In this way, it may be possible to accurately estimate the traffic state when relying on moving observers alone to collect relative flow data. If we do not rely on stationary observers to observe all link boundaries, in addition to exposing node traffic flow properties (see above), it is beneficial to expose the traffic demand and route choice behavior. This provides information on the network boundary conditions and turn fractions at nodes. Traffic demand (origin-destination) estimation and route choice modelling are topics which are widely studied. For both topics, methodologies exist that rely on flow observations (from stationary observers) and probe data (from probes that can be followed from their origin to their destination). Therefore, future studies can explore whether existing methodologies can be used for origin-destination estimation and route-choice modelling with relative flow data, and/or whether it is beneficial to design new methodologies that better fit the data characteristics.
- **Design of methodologies for traffic state estimation that assimilate relative flow data and different models**: To assimilate real-time relative flow data and other information (e.g., in the form of the models discussed above), we need to design new traffic state estimation methodologies. All pieces of information provide partial and noisy information. The to-be-designed traffic state estimation methodologies should describe which information the data and models provide and how these relate to each other. To assimilate the information, different techniques may be used. Bayesian techniques like the Kalman Filter and its variants, and the Particle Filter may be suitable assimilation techniques as these allow us to use information related to the error characteristics of the data and model estimates, which in turn may be exposed based on relative flow data (see above).
- **Validation of methodologies using real relative flow data**: In Chapters 3 and 4, empirical case studies are conducted. However, in the studies that consider the use of relative flow data (Chapters 5, 6, 7 and 8) the case studies rely on microscopic simulation. Microscopic simulation is interesting as it allows us to (1) create relative flow data, (2) obtain the ground truth and (3) simulate a wide range of traffic scenarios,

and do so at limited cost. However, real data also have their specific upsides, i.e., they describe real traffic behavior and contain real errors. Therefore, it is interesting to collect empirical relative flow data in combination with other data that may be used validation purposes, to validate the methodologies presented in this thesis and other methodologies that use these data (e.g., related to the future work that is proposed in this section).

- **Design traffic control methodologies that use estimates obtained from assimilating relative flow data and different models**: The studies conducted in this thesis and which relate the future research directions presented above all relate to traffic state estimation. The output of methodologies that relate to that topic are traffic state estimates. Obtaining such traffic state information is not an objective on its own, but is wanted as input for different applications. For instance, road authorities can use these estimates to design and operate dynamic traffic management systems, e.g., ramp-metering. The topic of 'traffic control' is widely studied, and the state estimates obtained using relative flow data may directly be used as input for these existing control methodologies. However, existing methodologies may be designed from the perspective of the currently widely used traffic sensing data, e.g., it may solely use detector data. It may beneficial to design new control methodologies that make better use of the traffic state estimates that can be obtained using relative flow data. Therefore, it should be investigated whether methodologies that are designed from the perspective of relative flow data may improve upon the traffic control objectives. This may involve the use of Artificial Intelligence (AI) techniques.

Bibliography

Amini, Z., Pedarsani, R., Skabardonis, A. & Varaiya, P. (2016). Queue-Length Estimation Using Real-Time Traffic Data. In *2016 ieee 19th international conference on intelligent transportation systems (itsc)* (pp. 1–6).

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, *61*(4), 821–856.

Antoniou, C., Balakrishna, R. & Koutsopoulos, H. N. (2011). A Synthesis of emerging data collection technologies and their impact on traffic management applications. *European Transport Research Review*, *3*, 139–148.

Arulampalam, M. S., Maskell, S., Gordon, N. & Clapp, T. (2002). A Tutorial on Particle Filters for Online Nonlinear / Non-Gaussian Bayesian Tracking. *IEEE Transactions on Signal Processing*, *50*(2), 174–188.

Bellemans, T., De Schutter, B. & De Moor, B. (2002). Models for traffic control. *Journal A*, *43*(3-4), 13–22.

Ben-Akiva, M. & Bierlaire, M. (1999). Discrete choice methods and their applications to short term travel decisions. In *Handbook of transportation science* (pp. 5–34). Kluwer.

Bhaskar, A., Chung, E. & Dumont, A.-G. (2010). Fusing loop detector and probe vehicle data to estimate travel time statistics on signalized urban networks. *Computer-Aided Civil and Infrastructure Engineering*, *26*(6), 433–450.

Cassidy, M. (1998). Bivariate relations in nearly stationary highway traffic. *Transportation Research Part B: Methodological*, *32*(1), 49–59.

CBS StatLine. (2019). *Dutch Central Bureau of Statistics (CBS).* Op 2019-09-30 verkregen van `opendata.cbs.nl/statline{#}/CBS/nl/`

Chow, G. C. (1960). Tests of Equality Between Sets of Coefficients in Two Linear Regressions. *Econometrica*, *28*(3), 591–605.

Claudel, C. G. & Bayen, A. M. (2008). Guaranteed bounds for traffic flow parameters estimation using mixed. In *Forty-sixth annual allerton conference* (pp. 636–645).

Claudel, C. G. & Bayen, A. M. (2010a). Lax-Hopf Based Incorporation of Internal Boundary Conditions Into Hamilton-Jacobi Equation. Part II: Computational Methods. *IEEE Transactions on Automatic Control*, *55*(5), 1158–1174.

Claudel, C. G. & Bayen, A. M. (2010b). Lax-Hopf Based Incorporation of Internal Boundary Conditions Into Hamilton-Jacobi Equation. Part I: Theory. *IEEE Transactions on Automatic Control*, *55*(5), 1142–1157.

Colyar, J. & Halkias, J. (2007). *Fact sheet US Highway 101 Dataset.* Op 2017-09-19 verkregen van https://www.fhwa.dot.gov/publications/research/operations/07030/index.cfm

Cottrill, C. D. (2009). Approaches to Privacy Preservation in Intelligent Transportation Systems and Vehicle Infrastructure Integration Initiative. *Transportation Research Record: Journal of the Transportation Research Board*(2129), 9–15.

Courant, R., Friedrichs, K. & Lewy, H. (1928). Uber die partiellen Differenzengleiehungen der mathematischen Physik. *Math. Ann.*, *100*(1), 32–74.

Daganzo, C. F. (1994). The Cell Transmission Model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B: Methodological*, *28B*(4), 269–287.

Daganzo, C. F. (1995). The Cell Transmission Model, Part II: Network traffic. *Transportation Research Part B: Methodological*, *29B*(2), 79–93.

Daganzo, C. F. (2005a, feb). A variational formulation of kinematic waves: basic theory and complex boundary conditions. *Transportation Research Part B: Methodological*, *39*(2), 187–196.

Daganzo, C. F. (2005b, dec). A variational formulation of kinematic waves: Solution methods. *Transportation Research Part B: Methodological*, *39*(10), 934–950.

Del Arco, E., Morgado, E., Ramiro-Barguene, J., More-Jimenez, I. & Caamano, A. (2011). Vehicular Sensor Networks in Congested Traffic: Linking STV Field Reconstruction and Communications Channel. In *14th international ieee conference on intelligent transportation systems* (pp. 606–613). Washington, DC, USA.

Dervisoglu, G., Gomes, G., Kwon, J., Muralidhuran, A., Varaiya, P. & Horowitz, R. (2009). Automatic Calibration of the Fundamental Diagram and Empirical Observations on Capacity. In *Transportation research board 88th annual meeting* (pp. 1–13).

Dijker, T. & Knoppers, P. (2006). *FOSIM 5.1 Gebruikershandleiding (Users Manual)* (Rapport). Technische Universiteit Delft.

Edie, L. C. (1965). Discussion of traffic stream measurements and definitions. In *2nd int. symp. on the theory of traffic flow.* Paris: OECD.

Eland, A. (2015). Tackling Urban Mobility with Technology. *https://europe.googleblog.com/2015/11/tackling-urban-mobility-with-technology.html (accessed on 2017-07-26)*. Verkregen van https://europe.googleblog.com/2015/11/tackling-urban-mobility-with-technology.html

EU. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016: on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General D. *Official Journal of the European Union*, 1–88.

Fellendorf, M. & Vortisch, P. (2001). Validation of the Microscopic Traffic Flow Model VISSIM in Different Real-World Situations. *Transportation Research Board 80th Annual Meeting*, 1–9.

Florin, R. & Olariu, S. (2017). On a Variant of the Mobile Observer Method. *IEEE Transactions on Intelligent Transportation Systems*, *18*(2), 441–449.

Fries, R. N., Reisi, M., Chowdhury, M. & Conway, A. J. (2012). Meeting privacy challenges while advancing intelligent transportation systems. *Transportation Research Part C: Emerging Technologies*, *25*, 34–45.

Hanemann, W. M. (1991). Willingness to Pay and Willingness to Accept: How Much Can They Differ? *The American Economic Review*, *81*(3), 635–647.

Hegyi, A., Mihaylova, L., Boel, R. & Lendek, Z. (2007). Parallelized particle filtering for freeway traffic state tracking. In *Proceedings of the european control conference 2007* (pp. 2442–2449). Kos, Greece.

Hegyi, A., Schutter, B. D. & Hellendoorn, H. (2005). Model predictive control for optimal coordination of ramp metering and variable speed limits. *Transportation Research Part C: Emerging Technologies*, *13*, 185–209.

Henkens, N., Mieras, W. & Bonnema, D. (2017). *Validatie FOSIM* (Rapport). De Bilt: Sweco.

Herrera, J. C. & Bayen, A. M. (2010). Incorporation of Lagrangian measurements in freeway traffic state estimation. *Transportation Research Part B: Methodological*, *44*(4), 460–481.

Hoogendoorn, S. P., Landman, R., Van Kooten, J. & Schreuder, M. (2013). Integrated Network Management Amsterdam: Control approach and test results. In *16th international ieee conference on intelligent transportation systems (itsc 2013)* (pp. 474–479).

Immers, L. & Logghe, S. (2002). *Traffic Flow Theory* (Rapport nr. May). Katholieke Universiteit Leuven.

Jarque, C. M. & Bera, A. K. (1987). A Test for Normality of Observations and Regression Residuals. *Internation Statistical Review / Revue Internationale de Statistique*, *55*(2), 163–172.

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*(March), 35–45.

Kargl, F., Friedman, A. & Boreli, R. (2013). Differential Privacy in Intelligent Transportation Systems. In *Proceedings of the sixth acm conference on security and privacy in wireless and mobile networks* (pp. 107–112).

Knoop, V. L. & Daamen, W. (2017). Automatic fitting procedure for the fundamental diagram. *Transportmetrica B: Transport Dynamics*, *5*(2), 129–144.

Knoop, V. L. & Hoogendoorn, S. P. (2012). Empirics of a Generalized Macroscopic Fundamental Diagram for Urban Freeways. In *92th annual meeting of the transportation research board* (pp. 1–16).

Knoop, V. L., Hoogendoorn, S. P. & Van Zuylen, H. (2009). Empirical Differences between Time Mean Speed and Space Mean Speed. In C. Appert-Rolland, F. Chevois, P. Gondret, S. Lassarre, J.-P. Lebacque & M. Schreckenberg (red.), *Proceedings of traffic and granular flow 07* (pp. 351–356). New York: Springer.

Knoop, V. L., Wilson, R. E., Buisson, C. & Van Arem, B. (2012). Number of Lane Changes Determined by Splashover Effects in Loop Detector Counts. *IEEE Transactions on Intelligent Transportation Systems*, *13*(4), 1525–1534.

Laval, J. A., He, Z. & Castrillon, F. (2012). Stochastic extension of Newells three-detector method. *Transportation Research Record: Journal of the Transportation Research Board*, *2315*, 73–80.

Le, T., Kovács, P., Walton, N., Vu, H. L., Andrew, L. L. H. & Hoogendoorn, S. S. P. (2015). Decentralized signal control for urban road networks. *Transportation Research Part C: Emerging Technologies*, *58*, 431–450.

Lighthill, M. J. & Whitham, G. B. (1955, may). On Kinematic Waves. II. A Theory of Traffic Flow on Long Crowded Roads. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *229*(1178), 317–345.

Makigami, Y., Newell, G. F. & Rothery, R. (1971). Three-Dimensional Representation of Traffic Flow. *Transportation Science*, *5*(3), 302–313.

Manski, C. (1977). The structure of random utility models. *Theory and Decision*, *8*, 229–254.

May, A. D. (1990). *Traffic Flow Fundamentals*. Prentice Hall.

Minderhoud, M. M. & Kirwan, K. (2001). *Validatie FOSIM voor asymmetrische weekvakken - CAPWEEF fase 1* (Rapport). Delft: Laboratorium voor Verkeerskunde, Faculteit Civiele Techniek en Geowetenschappen, Technische Universiteit Delft.

Nanthawichit, C., Nakatsuji, T. & Suzuki, H. (2003). Application of probe-vehicle data for real time traffic state estimation and short term travel time prediction on a freeway. *Transportation Research Record*, *5890*(1855), 49–59.

Newell, G. (2002). A simplified car-following theory: a lower order model. *Transportation Research Part B: Methodological*, *36*, 195–205.

Newell, G. F. (1993a). A simplified theory of kinematic waves in highway traffic, Part I: General theory. *Transportation Research Part B: Methodological*, *27*(4), 281–287.

Newell, G. F. (1993b). A simplified theory of kinematic waves in highway traffic, Part III: Multi-destination flows. *Transportation Research Part B: Methodological*, *27*(4), 305–313.

Newell, G. F. (1993c). A simplified theory of kinematic waves in highway traffic, Part II: Queueing at freeway bottlenecks. *Transportation Research Part B: Methodological*, *27*(4), 289–303.

Oh, S., Ritchie, S. G. & Oh, C. (2002). Real time traffic measurement from single loop inductive signatures. In *Proceedings of the 81st annual meeting of the transportation research board.* Washington, DC, USA.

Papageorgiou, M., Hadj-salem, H. & Blosseville, J. (1991). ALINEA : A Local Feedback Control Law for On-Ramp Metering. *Transportation Research Record*, *1320*, 58–64.

Papageorgiou, M., Hadj-Salem, H. & Blosseville, J. (1991). ALINEA: A Local Feedback Control Law for On-ramp Metering; A Real-life Study. *Transportation Research Record: Journal of the Transportation Research Board*, *1320*, 58–64.

Praktijkproef Amsterdam (PPA). (2016). *www.praktijkproefamsterdam.nl (accessed on 2017-07-26)*. Verkregen van `www.praktijkproefamsterdam.nl`

Quandt, R. E. (1960). Tests of the Hypothesis that a Linear Regression System Obeys Two Separate Regimes. *Journal of the American Statistical Association*, *55*(290), 324–330.

Ramsey, J. B., Newton, H. J. & Harvill, J. L. (2004). Chapter 4: Moments and the Shape of Histograms. In *The elements of statistics: With applications to economics and the social sciences* (pp. 77–119). Duxbury/Thomson Learning.

Redmill, K. A., Coifman, B., Mccord, M. & Mishalani, R. G. (2011). Using Transit or Municipal Vehicles as Moving Observer Platforms for Large Scale Collection of Traffic and Transportation System Information. In *14th international ieee conference on intelligent transportation systems.*

Richards, P. I. . (1956). Shock Waves on the Highway. *Operations Research*, *4*(1), 42–51.

Seo, T. (2015). *Traffic Estimation with Vehicles Observing Other Vehicles*. Academisch proefschrift, Tokyo Institute of Technology.

Seo, T., Bayen, A. M., Kusakabe, T. & Asakura, Y. (2017). Traffic state estimation on highway: A comprehensive survey. *Annual Reviews in Control*, *43*, 128–151.

Seo, T. & Kusakabe, T. (2015). Probe vehicle-based traffic state estimation method with spacing information and conservation law. *Transportation Research Part C*, *59*, 391–403.

Seo, T., Kusakabe, T. & Asakura, Y. (2015). Estimation of flow and density using probe vehicles with spacing measurement equipment. *Transportation Research Part C*, *53*, 134–150.

Seo, T., Kusakabe, T. & Asakura, Y. (2017). Calibration of fundamental diagram using trajectories of probe vehicles: Basic formulation and heuristic algorithm. In *Transportation research procedia* (Dl. 21, pp. 6–17).

Smaragdis, E., Papageorgiou, M. & Kosmatopoulos, E. (2004, mar). A flow-maximizing adaptive local ramp metering strategy. *Transportation Research Part B: Methodological*, *38*(3), 251–270.

Smulders, S. (1990). Control of freeway traffic flow by variable speed signs. *Transportation Research Part B: Methodological*, *24B*(2), 111–132.

Sun, Z., Jin, W.-L. & Ritchie, S. G. (2017). Simultaneous estimation of states and parameters in Newell's simplified kinematic wave model with Eulerian and Lagrangian traffic data. *Transportation Research Part B: Methodological*, *104*, 106–122.

Takenouchi, A., Kawai, K. & Kuwahara, M. (2019). Traffic state estimation and its sensitivity utilizing measurements from the opposite lane. *Transportation Research Part C*, *104*(October 2018), 95–109.

Tan, P.-N., Steinbach, M. & Kumar, V. (2006). Cluster Analysis: Basic Concepts and Algorithms. In *Introduction to data mining* (pp. 487–568).

Treiber, M., Hannecke, A. & Helbing, D. (2000). Congested traffic states in empirical observations and microscopic simulations. *Physical Review E*, *62*(2), 1805–1824.

Treiber, M. & Helbing, D. (2002). Reconstructing the Spatio-Temporal Traffic Dynamics from Stationary Detector Data. *Cooper@tive Tr@nsport@tion Dyn@mics*, *1*, 3.1–3.24.

Treiber, M., Hennecke, A. & Helbing, D. (2000). Microscopic Simulation of Congested Traffic. In *Traffic and granular flow '99* (pp. 365–376).

Uenk-Telgen, M. (2018). *Vergelijkings- onderzoek Floating Car Data* (Rapport). Utrecht: NDW: Nationale Databank Wegverkeergegevens.

Van der Gun, J. P. T., Pel, A. J. & Van Arem, B. (2017). Extending the Link Transmission Model with non-triangular fundamental diagrams and capacity drops. *Transportation Research Part B: Methodological*, *98*, 154–178.

Van Erp, P. B. C., Knoop, V. L. & Hoogendoorn, S. P. (2018a). Estimating the vehicle accumulation: Data-fusion of loop-detector flow and floating car speed data. In *Transportation research board 97th annual meeting.*

Van Erp, P. B. C., Knoop, V. L. & Hoogendoorn, S. P. (2018b). Macroscopic traffic state estimation using relative flows from stationary and moving observers. *Transportation Research Part B: Methodological*, *114*, 281–299.

Van Erp, P. B. C., Knoop, V. L. & Hoogendoorn, S. P. (2019). On the value of relative flow data. *Transportation Research Part C: Emerging Technologies (in press)*, *00*(00), 1–21.

Van Erp, P. B. C., Thoen, S., Knoop, V. L. & Hoogendoorn, S. P. (2018). Estimating the fundamental diagram using moving observers. In *The 21st ieee international conference on intelligent transport systems* (pp. 1–7).

Van Hinsbergen, C. P. I. J., Schreiter, T., Zuurbier, F. S., Van Lint, J. W. C. & Van Zuylen, H. J. (2012, mar). Localized Extended Kalman Filter for Scalable Real-Time Traffic State Estimation. *IEEE Transactions on Intelligent Transportation Systems*, *13*(1), 385–394.

Van Lint, J. W. C., Bertini, R. L. & Hoogendoorn, S. P. (2014). Data fusion solutions to compute performance measures for urban arterials. In *Celebrating 50 years of traffic flow theory* (pp. 1–5). Portland, Oregon.

Van Lint, J. W. C. & Hoogendoorn, S. P. (2015). A generic methodology to estimate vehicle accumulation on urban arterials by fusing vehicle counts and travel times. In *94th annualmeeting of the transportation research board.*

Varaiya, P. (2013). Max pressure control of a network of signalized intersections. *Transportation Research Part C: Emerging Technologies*, *36*, 177–195.

Vigos, G., Papageorgiou, M. & Wang, Y. (2008). Real-time estimation of vehicle-count within signalized links. *Transportation Research Part C: Emerging Technologies*, *16*, 18–35.

Wada, K., Usui, K., Takigawa, T. & Kuwahara, M. (2018). An optimization modeling of coordinated traffic signal control based on the variational theory and its stochastic extension . *Transportation Research Part B*, *117*, 907–925.

Wang, Y. & Papageorgiou, M. (2005, feb). Real-time freeway traffic state estimation based on extended Kalman filter: a general approach. *Transportation Research Part B: Methodological*, *39*(2), 141–167.

Wang, Y., Papageorgiou, M. & Messmer, A. (2002). A Predictive Feedback Routing Control Strategy for Freeway Network Traffic. In *Proceeding of the american control conference* (pp. 3606–3611). Anchorage.

Wardrop, J. & Charlesworth, G. (1954). A Method of Estimating Speed and Flow of Traffic from a Moving Vehicle. *Proceedings of the Institution of Civil Engineers*, *3*(1), 158–171.

Woodard, D., Nogin, G., Koch, P., Racz, D., Goldszmidt, M. & Horvitz, E. (2017). Predicting travel time reliability using mobile phone GPS data. *Transportation Research Part C*, *75*, 30–44.

Work, D. B., Blandin, S., Tossavainen, O.-P., Piccoli, B. & Bayen, A. M. (2010). A Traffic Model for Velocity Data Assimilation. *Applied Mathematics Research eXpress*, *1*, 1–35.

Yang, X., Lu, Y. & Hao, W. (2017). Origin-Destination Estimation Using Probe Vehicle Trajectory and Link Counts. *Journal of Advanced Transportation*, *2017*, 1–18.

Yperman, I. (2007). *The Link Transmission Model for Dynamic Network Loading*. Academisch proefschrift, Katholieke universiteit Leuven.

About the author

Paul Bernardus Cornelis van Erp was born in 's-Hertogenbosch, The Netherlands, on May 18th, 1988. After spending his early-life in s-Hertogenbosch, he moved to Delft in 2006 to study Aerospace Engineering at Delft University of Technology. After obtaining his BSc. degree in Aerospace Engineering, he set out to explore other fields and simultaneously pursued MSc. degrees in Civil Engineering (Transport & Planning) at Delft University of Technology and Econometrics and Management Sciences at Erasmus University Rotterdam. These MSc. degrees were respectively obtained (both with distinction) in January 2014 and June 2015.

While working on his MSc. thesis at the department of Transport & Planning, Paul applied for a personal NWO grant at TRAIL Research School. After obtaining this grant, he started his PhD at Transport & Planning in September 2015. In this PhD, he studied the field of (road) traffic state estimation, and more specifically looked at the opportunity to use relative flow data for traffic state estimation.

Paul pursued his PhD on a part-time basis (four days a week). Alongside his PhD, he worked at Arane consultants in traffic and transportation, on real-estate projects and explored other opportunities.

In line with his diverse background, Paul is interested in a wide range of topics. His academic work focusses on the domain of (road) traffic flow; however, he also likes to learn about and work on other domains, such as education, economics and politics. In any endeavor, Paul wants to gain a deeper understanding of the topic and wants to create value in an efficient manner.

List of publications

The journal and conference publications are listed below. These are separated into contributions that relate to this PhD thesis and other contributions.

Contributions related to this PhD thesis

The contributions related to this PhD thesis are grouped based on the related thesis chapters. Some chapters have lead to a conference publication and a related journal publication, which are both listed below.

Chapter 2

- Van Erp, P.B.C., Knoop, V.L. & Hoogendoorn, S.P. (2017), Evaluation of data-based estimation error characteristics in macroscopic traffic state estimation. In *Proceedings of the 96th Annual Meeting of the Transportation Research Board*.
- Van Erp, P.B.C., Knoop, V.L. & Hoogendoorn, S.P. (2017), Macroscopic traffic state estimation: Understanding traffic sensing data-based estimation error. *Journal of Advanced Transportation*.

Chapter 3

- Van Erp, P.B.C., Knoop, V.L. & Hoogendoorn, S.P. (2018), Estimating the vehicle accumulation: Data-Fusion of loop-detector flow and floating car speed data. In *Proceedings of the 97th Annual Meeting of the Transportation Research Board*.
- Van Erp, P.B.C., Knoop, V.L. & Hoogendoorn, S.P. (*under journal review*), Vehicle accumulation estimation with detector count and probe mean speed data.

Chapter 4

- Van Erp, P.B.C., Knoop, V.L., Smits, E.-S., Tampere, C. & Hoogendoorn, S.P. (2020), Estimation of the change in cumulative flow over probe trajectories using detector data. *Accepted for the 99th Annual Meeting of the Transportation Research Board* and *under journal review*.

Chapter 5

- Van Erp, P.B.C., Knoop, V.L. & Hoogendoorn, S.P. (2018), Macroscopic traffic state estimation using relative flows from stationary and moving observers. *Transportation Research Part B: Methodological*, 114, 281-299.

Chapter 6

- Van Erp, P.B.C., Knoop, V.L., Thoen, S., & Hoogendoorn, S.P. (2018), Estimating the fundamental diagram using moving observers. In *Proceedings of the 21st International Conference on Intelligent Transportation Systems (ITSC)*.

Chapter 7

- Van Erp, P.B.C., Knoop, V.L. & Hoogendoorn, S.P. (2019), On the value of relative flow data. In *Proceedings of the 23rd International Symposium on Transportation and Traffic Theory (ISTTT23)*.
- Van Erp, P.B.C., Knoop, V.L. & Hoogendoorn, S.P. (2019), On the value of relative flow data. *Transportation Research Part C: Emerging Technologies*, in press.

Chapter 8

- Van Erp, P.B.C., Knoop, V.L. & Hoogendoorn, S.P. (*under journal review*), Using relative flow data to reveal traffic flow properties.

Other contributions prior and during PhD

- Marzuoli, A., Boidot, E., Colomar, P., Guerpillon, M., Feron, E., Van Erp, P.B.C., Ucko, A., Bayen, A. & Hansen, M. (2015), Multimodality in a metroplex environment: A case study in the San Francisco Bay area. In *Proceedings of the AIAA SciTech Forum*.
- Marzuoli, A., Boidot, E., Feron, E., Van Erp, P.B.C., Ucko, A., Bayen, A. & Hansen, M. (2015), Multimodal impact analysis of an airside catastrophic event: A case study of the Asiana crash. *IEEE Transactions on Intelligent Transportation Systems*, 17 (2), 587-604.
- Knoop, V.L., Van Erp, P.B.C., Leclercq, L. & Hoogendoorn, S.P. (2018), Empirical MFDs using Google Traffic Data, In *Proceedings of the 21st International Conference on Intelligent Transportation Systems (ITSC)*.
- Knoop, V.L., Taale, H., Meulenberg, M., Van Erp, P.B.C. & Hoogendoorn, S.P. (2018), Ramp Metering with Real-Time Estimation of Parameters, In *Proceedings of the 21st International Conference on Intelligent Transportation Systems (ITSC)*.

TRAIL Thesis Series

The following list contains the most recent dissertations in the TRAIL Thesis Series. For a complete overview of more than 250 titles see the TRAIL website: www.rsTRAIL.nl.

The TRAIL Thesis Series is a series of the Netherlands TRAIL Research School on transport, infrastructure and logistics.

Erp, P.B.C. van, *Relative Flow Data: New Opportunities for Traffic State Estimation*, T2020/01, February 2020, TRAIL Thesis Series, the Netherlands

Zhu, Y., *Passenger-Oriented Timetable Rescheduling in Railway Disruption Management*, T2019/16, December 2019, TRAIL Thesis Series, the Netherlands

Chen, L., *Cooperative Multi-Vessel Systems for Waterborne Transport*, T2019/15, November 2019, TRAIL Thesis Series, the Netherlands

Kerkman, K.E., *Spatial Dependence in Travel Demand Models: Causes, implications, and solutions*, T2019/14, October 2019, TRAIL Thesis Series, the Netherlands

Liang, X., *Planning and Operation of Automated Taxi Systems*, T2019/13, September 2019, TRAIL Thesis Series, the Netherlands

Ton, D., *Unravelling Mode and Route Choice Behaviour of Active Mode Users*, T2019/12, September 2019, TRAIL Thesis Series, the Netherlands

Shu, Y., *Vessel Route Choice Model and Operational Model Based on Optimal Control*, T2019/11, September 2019, TRAIL Thesis Series, the Netherlands

Luan, X., *Traffic Management Optimization of Railway Networks*, T2019/10, July 2019, TRAIL Thesis Series, the Netherlands

Hu, Q., *Container Transport inside the Port Area and to the Hinterland*, T2019/9, July 2019, TRAIL Thesis Series, the Netherlands

Andani, I.G.A., *Toll Roads in Indonesia: transport system, accessibility, spatial and equity impacts*, T2019/8, June 2019, TRAIL Thesis Series, the Netherlands

Ma, W., *Sustainability of Deep Sea Mining Transport Plans*, T2019/7, June 2019, TRAIL Thesis Series, the Netherlands

Alemi, A., *Railway Wheel Defect Identification*, T2019/6, January 2019, TRAIL Thesis Series, the Netherlands

Liao, F., *Consumers, Business Models and Electric Vehicles*, T2019/5, May 2019, TRAIL Thesis Series, the Netherlands

Tamminga, G., *A Novel Design of the Transport Infrastructure for Traffic Simulation Models*, T2019/4, March 2019, TRAIL Thesis Series, the Netherlands

Lin, X., *Controlled Perishable Goods Logistics: Real-time coordination for fresher products*, T2019/3, January 2019, TRAIL Thesis Series, the Netherlands

Dafnomilis, I., *Green Bulk Terminals: A strategic level approach to solid biomass terminal design*, T2019/2, January 2019, TRAIL Thesis Series, the Netherlands

Feng, Fan, *Information Integration and Intelligent Control of Port Logistics System*, T2019/1, January 2019, TRAIL Thesis Series, the Netherlands

Beinum, A.S. van, *Turbulence in Traffic at Motorway Ramps and its Impact on Traffic Operations and Safety*, T2018/12, December 2018, TRAIL Thesis Series, the Netherlands

Bellsola Olba, X., *Assessment of Capacity and Risk: A Framework for Vessel Traffic in Ports*, T2018/11, December 2018, TRAIL Thesis Series, the Netherlands

Knapper, A.S., *The Effects of using Mobile Phones and Navigation Systems during Driving*, T2018/10, December 2018, TRAIL Thesis Series, the Netherlands

Varotto, S.F., *Driver Behaviour during Control Transitions between Adaptive Cruise Control and Manual Driving: empirics and models*, T2018/9, December 2018, TRAIL Thesis Series, the Netherlands

Stelling-Konczak, A., *Cycling Safe and Sound*, T2018/8, November 2018, TRAIL Thesis Series, the Netherlands

Essen, van M.A., *The Potential of Social Routing Advice*, T2018/7, October 2018, TRAIL Thesis Series, the Netherlands

Su, Zhou, *Maintenance Optimization for Railway Infrastructure Networks*, T2018/6, September 2018, TRAIL Thesis Series, the Netherlands

Cai, J., *Residual Ultimate Strength of Seamless Metallic Pipelines with Structural Damage*, T2018/5, September 2018, TRAIL Thesis Series, the Netherlands

Ghaemi, N., *Short-turning Trains during Full Blockages in Railway Disruption Management*, T2018/4, July 2018, TRAIL Thesis Series, the Netherlands