

LiU-ITN-TEK-A--12/070--SE

# Framework for Calibration of a Traffic State Space Model

Magnus Fransson

Mats Sandin

2012-10-17



**Linköpings universitet**  
TEKNISKA HÖGSKOLAN

LiU-ITN-TEK-A--12/070--SE

# **Framework for Calibration of a Traffic State Space Model**

Examensarbete utfört i Transportsystem  
vid Tekniska högskolan vid  
Linköpings universitet

**Magnus Fransson  
Mats Sandin**

Handledare David Gundlegård  
Examinator Clas Rydergren

Norrköping 2012-10-17

## Upphovsrätt

Detta dokument hålls tillgängligt på Internet – eller dess framtida ersättare – under en längre tid från publiceringsdatum under förutsättning att inga extraordinära omständigheter uppstår.

Tillgång till dokumentet innebär tillstånd för var och en att läsa, ladda ner, skriva ut enstaka kopior för enskilt bruk och att använda det oförändrat för ickekommersiell forskning och för undervisning. Överföring av upphovsrätten vid en senare tidpunkt kan inte upphäva detta tillstånd. All annan användning av dokumentet kräver upphovsmannens medgivande. För att garantera äktheten, säkerheten och tillgängligheten finns det lösningar av teknisk och administrativ art.

Upphovsmannens ideella rätt innefattar rätt att bli nämnd som upphovsman i den omfattning som god sed kräver vid användning av dokumentet på ovan beskrivna sätt samt skydd mot att dokumentet ändras eller presenteras i sådan form eller i sådant sammanhang som är kränkande för upphovsmannens litterära eller konstnärliga anseende eller egenart.

För ytterligare information om Linköping University Electronic Press se förlagets hemsida <http://www.ep.liu.se/>

## Copyright

The publishers will keep this document online on the Internet - or its possible replacement - for a considerable time from the date of publication barring exceptional circumstances.

The online availability of the document implies a permanent permission for anyone to read, to download, to print out single copies for your own use and to use it unchanged for any non-commercial research and educational purpose. Subsequent transfers of copyright cannot revoke this permission. All other uses of the document are conditional on the consent of the copyright owner. The publisher has taken technical and administrative measures to assure authenticity, security and accessibility.

According to intellectual property law the author has the right to be mentioned when his/her work is accessed as described above and to be protected against infringement.

For additional information about the Linköping University Electronic Press and its procedures for publication and for assurance of document integrity, please refer to its WWW home page: <http://www.ep.liu.se/>

# Framework for Calibration of a Traffic State Space Model

Magnus Fransson, Mats Sandin

November 9, 2012

## Copyright

The publishers will keep this document online on the Internet – or its possible replacement – from the date of publication barring exceptional circumstances.

The online availability of the document implies permanent permission for anyone to read, to download, or to print out single copies for his/hers own use and to use it unchanged for non-commercial research and educational purpose. Subsequent transfers of copyright cannot revoke this permission. All other uses of the document are conditional upon the consent of the copyright owner. The publisher has taken technical and administrative measures to assure authenticity, security and accessibility.

According to intellectual property law the author has the right to be mentioned when his/her work is accessed as described above and to be protected against infringement.

For additional information about the Linköping University Electronic Press and its procedures for publication and for assurance of document integrity, please refer to its www home page: <http://www.ep.liu.se/>.

©Magnus Fransson & Mats Sandin

## Abstract

To evaluate the traffic state over time and space, several models can be used. A typical model for estimating the state of the traffic for a stretch of road or a road network is the cell transmission model, which is a form of state space model. This kind of model typically needs to be calibrated since the different roads have different properties. This thesis will present a calibration framework for the velocity based cell transmission model, the CTM-v.

The cell transmission model for velocity is a discrete time dynamical system that can model the evolution of the velocity field on highways. Such a model can be fused with an ensemble Kalman filter update algorithm for the purpose of velocity data assimilation. Indeed, enabling velocity data assimilation was the purpose for ever developing the model in the first place and it is an essential part of the Mobile Millennium research project.

Therefore a systematic methodology for calibrating the cell transmission is needed. This thesis presents a framework for calibration of the velocity based cell transmission model that is combined with the ensemble Kalman filter.

The framework consists of two separate methods, one is a statistical approach to calibration of the fundamental diagram. The other is a black box optimization method, a simplification of the complex method that can solve inequality constrained optimization problems with non-differentiable objective functions. Both of these methods are integrated with the existing system, yielding a calibration framework, in particular highways were stationary detectors are part of the infrastructure.

The output produced by the above mentioned system is highly dependent on the values of its characterising parameters. Such parameters need to be calibrated so as to make the model a valid representation of reality. Model calibration and validation is a process of its own, most often tailored for the researchers models and purposes.

The combination of the two methods are tested in a suit of experiments for two separate highway models of Interstates 880 and 15, CA which are evaluated against travel time and space mean speed estimates given by Bluetooth detectors with an error between 7.4 and 13.4 % for the validation time periods depending on the parameter set and model.

## Acknowledgements

We are truly indebted to our mentor, friend and group leader Anthony D. Patire for his help, patience and enthusiasm as well as for making our stay in Berkeley one to remember.

We owe thanks to our mentors, Andreas Allström and David Gundlegård for their support and especially their trust and the amount of freedom that followed with it.

For ever making this thesis possible we would like to thank our fellow interns and the staff at California Partners for Advanced Transportation Technology, especially Anastasios Kouvelas, Agathe Benoît, Boris Prodhomme, Elie Daou and Jérôme Thai. We hope that you have learned as much from us as we have from you.

We would like to thank our examiner, Clas Rydergren for his encouragement and for ever bringing the idea to do this thesis up to us.

Finally, our visit to University of California and California Partners for Advanced Transportation Technology was made possible by financial support from Sweco Infrastructure.

We have truly seen a new world of opportunities, thank you all.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Purpose and objectives . . . . .	2
1.2	Method . . . . .	2
1.3	Limitations . . . . .	2
1.4	Outline . . . . .	3
<b>2</b>	<b>Literature Review Part I: Highway Model</b>	<b>4</b>
2.1	Traffic flow model review . . . . .	4
2.1.1	Transformation of the Lightham-Whitham-Richard partial differential equation . . . . .	5
2.1.2	Transformation to the velocity domain . . . . .	7
2.1.3	Discretization . . . . .	8
2.1.4	Fundamental diagrams . . . . .	9
2.1.5	Cell transmission model for velocity . . . . .	11
2.1.6	Model representation of junctions . . . . .	12
2.1.7	Network algorithm . . . . .	13
2.2	Data assimilation: The ensemble Kalman filter . . . . .	14
2.2.1	Statistics . . . . .	14
2.2.2	Analysis schemes for combined model prediction and data assimilation . . . . .	20
2.2.3	Kalman filter . . . . .	26
2.2.4	Extended Kalman filter . . . . .	28
2.2.5	Ensemble Kalman filter . . . . .	30
2.3	Combining CTM-v and EnKF . . . . .	36
2.3.1	Implementation of the CTM-v and EnKF combination . . . . .	37
<b>3</b>	<b>Literature Review Part II: Calibration</b>	<b>38</b>
3.1	Calibration and validation of traffic models . . . . .	38
3.1.1	Literature survey . . . . .	39
3.2	Automatic empirical calibration method . . . . .	42
3.2.1	Regression analysis: Equality constrained least squares . . . . .	43
3.2.2	Step I . . . . .	45
3.2.3	Step II . . . . .	45
3.2.4	Step III . . . . .	46
3.2.5	Remarks . . . . .	48
3.3	Complex method . . . . .	48
3.3.1	Formal presentation . . . . .	49
3.3.2	Example: Visualization of the complex movement . . . . .	54

3.3.3	Example: Calibration of a single edge highway model . . .	54
<b>4</b>	<b>System Description</b>	<b>60</b>
4.1	Overview . . . . .	60
4.2	Measurement loader . . . . .	61
4.3	Highway model . . . . .	62
4.4	Ensemble Kalman filter revisited . . . . .	62
<b>5</b>	<b>Implementations of Calibration Procedures</b>	<b>64</b>
5.1	Implementation of the automatic empirical calibration method . .	64
5.1.1	Version I . . . . .	64
5.1.2	Version II . . . . .	67
5.1.3	Version III . . . . .	69
5.1.4	Implementation remarks . . . . .	70
5.2	Link assignment . . . . .	71
5.2.1	Forward link assignment . . . . .	73
5.2.2	Backward link assignment . . . . .	73
5.3	Complex method implementation . . . . .	73
5.3.1	Ground truth . . . . .	74
5.3.2	Complex method: Implementation concept . . . . .	76
5.3.3	Complex method: Implementation data flow . . . . .	80
5.3.4	Complex method: Simplification for problems with explicit constraints . . . . .	82
<b>6</b>	<b>Experiments</b>	<b>84</b>
6.1	Test site introduction . . . . .	84
6.1.1	Test site I: Interstate 880, CA . . . . .	84
6.1.2	Test site II: Interstate 15 Ontario . . . . .	85
6.2	Experiment layout: Fundamental diagram estimation . . . . .	86
6.3	Experiment layout: Automated calibration . . . . .	87
6.3.1	Local and source alternating calibration framework using standard fundamental diagrams . . . . .	88
6.3.2	Global and type-differentiated calibration framework using estimated fundamental diagrams . . . . .	90
6.3.3	Global calibration framework using estimated fundamental diagrams . . . . .	91
<b>7</b>	<b>Results</b>	<b>93</b>
7.1	Fundamental Diagram Calibration Results . . . . .	93
7.2	Complex method calibration results . . . . .	95
7.2.1	Local and source alternating calibration framework using standard fundamental diagrams: Interstate 880, CA . . . . .	95
7.2.2	Global and type-differentiated calibration framework using estimated fundamental diagrams: Interstate 880, CA . . . . .	102
7.2.3	Global calibration framework using estimated fundamental diagrams: Interstate 15, Ontario, CA . . . . .	106
<b>8</b>	<b>Analysis and Discussion</b>	<b>113</b>
8.1	Fundamental diagram estimation and link assignment procedure	113
8.1.1	Calibration framework . . . . .	114

<b>9 Conclusion and Future Work</b>	<b>115</b>
<b>A Identified Parameters</b>	<b>119</b>
<b>B Bluetooth Deployment</b>	<b>126</b>
<b>C Fundamental Diagram Calibration Results</b>	<b>130</b>

# List of Figures

2.1	Speed-density and flow-density relationships. . . . .	9
2.2	Posterior probability. . . . .	16
3.1	Flow-density measurements. . . . .	43
3.2	Step I of Gunes' method. . . . .	45
3.3	Step II of Gunes' method. . . . .	46
3.4	Relationship between quartiles and mass of probability. . . . .	47
3.5	Data-bin representation. . . . .	47
3.6	Step III, and outcome, of Gunes' method. . . . .	48
3.7	Flowchart representation of the Complex method. . . . .	53
3.8	Complex movement during optimization effort. . . . .	54
3.9	A highway section. . . . .	56
3.10	Weak boundary conditions. . . . .	57
3.11	Example model output. . . . .	58
3.12	Complex movement during calibration, a 3-dimensional case. . . . .	59
4.1	Data flow in the Mobile Millennium system. . . . .	61
5.1	Step I of Version I of the modified automatic empirical calibration method. . . . .	65
5.2	Step II of Version I of the modified automatic empirical calibration method. . . . .	66
5.3	Step III of Version I of the modified automatic empirical calibration method. . . . .	66
5.4	Step I of Version II of the modified automatic empirical calibration method. . . . .	68
5.5	Step II of Version II of the modified automatic empirical calibration method. . . . .	68
5.6	Step III of Version II of the modified automatic empirical calibration method. . . . .	69
5.7	Step IV of Version III of the modified automatic empirical calibration method. . . . .	70
5.8	Link assignment procedure. . . . .	72
5.9	Ground truth data collection. . . . .	75
5.10	Ground truth visualization. . . . .	76
5.11	Layer view of the data processing. . . . .	78
5.12	Layer view of parameter evaluation. . . . .	79
5.13	Feedback loop merged with the Mobile Millennium system's data flow. . . . .	81

5.14	Flow chart of the simplified complex method. . . . .	83
6.1	Overview of test site I: I-880 . . . . .	85
6.2	Overview of test site II: I-15 Ontario . . . . .	86
7.1	Calibration framework 1: Initial validation of I-880 model. . . . .	96
7.2	Calibration framework 1: Validation of the initial calibration of the I-880 model. . . . .	97
7.3	Calibration framework 1: Initial validation of I-880 model, travel time perspective. . . . .	98
7.4	Calibration framework 1: Validation of the initial calibration of the I-880 model, travel time perspective. . . . .	99
7.5	Calibration framework 1: Final model validation. . . . .	100
7.6	Calibration framework 1: Final model validation, travel time perspective. . . . .	101
7.7	Calibration framework 1: Validation of the velocity field. . . . .	102
7.8	Calibration framework 2: Outcome of the calibration effort. . . . .	103
7.9	Calibration framework 2: Validation of the calibration effort. . . . .	103
7.10	Calibration framework 2: Validation of the calibrated model, travel time perspective. . . . .	105
7.11	Calibration framework 2: Validation of the velocity field. . . . .	106
7.12	Calibration framework 3: Model output prior to calibration. . . . .	107
7.13	Calibration framework 3: Validation of the calibration effort. . . . .	107
7.14	Calibration framework 3: Model output prior to calibration. . . . .	108
7.15	Calibration framework 3: Validation of the calibration effort. . . . .	108
7.16	Calibration framework 3: Validation of the calibration effort, travel time perspective. . . . .	110
7.17	Calibration framework 3: Validation of the calibration effort, travel time perspective. . . . .	111
7.18	Calibration framework 3: Validation of the velocity field. . . . .	112
7.19	Calibration framework 3: Validation of the velocity field. . . . .	112
B.1	Overview of the I-880 Bluetooth deployment area. . . . .	126
B.2	Detailed map over the I-880 Bluetooth deployment. . . . .	127
B.3	Overview of the I-15 Ontario Bluetooth deployment area. . . . .	128
B.4	Detailed map over the I-15 Ontario Bluetooth deployment. . . . .	129
C.1	Space mean speeds on I-880 for different fundamental diagrams. . . . .	130
C.2	Space mean speeds on I-880 for different fundamental diagrams. . . . .	131
C.3	Space mean speeds on I-880 for different fundamental diagrams. . . . .	131
C.4	Space mean speeds on I-880 for different fundamental diagrams. . . . .	132
C.5	Space mean speeds on I-880 for different fundamental diagrams. . . . .	132
C.6	Space mean speeds on I-880 for different fundamental diagrams. . . . .	133
C.7	Space mean speeds on I-880 for different fundamental diagrams. . . . .	133
C.8	Space mean speeds on I-880 for different fundamental diagrams. . . . .	134
C.9	Space mean speeds on I-880 for different fundamental diagrams. . . . .	134
C.10	Space mean speeds on I-880 for different fundamental diagrams. . . . .	135
C.11	Space mean speeds on I-880 for different fundamental diagrams. . . . .	135
C.12	Space mean speeds on I-880 for different fundamental diagrams. . . . .	136
C.13	Space mean speeds on I-880 for different fundamental diagrams. . . . .	136

C.14 Space mean speeds on I-15 Ontario for different fundamental diagrams. . . . .	137
C.15 Space mean speeds on I-15 Ontario for different fundamental diagrams. . . . .	138
C.16 Space mean speeds on I-15 Ontario for different fundamental diagrams. . . . .	138
C.17 Space mean speeds on I-15 Ontario for different fundamental diagrams. . . . .	139
C.18 Space mean speeds on I-15 Ontario for different fundamental diagrams. . . . .	139
C.19 Space mean speeds on I-15 Ontario for different fundamental diagrams. . . . .	140
C.20 Space mean speeds on I-15 Ontario for different fundamental diagrams. . . . .	140
C.21 Space mean speeds on I-15 Ontario for different fundamental diagrams. . . . .	141
C.22 Space mean speeds on I-15 Ontario for different fundamental diagrams. . . . .	141
C.23 Space mean speeds on I-15 Ontario for different fundamental diagrams. . . . .	142
C.24 Space mean speeds on I-15 Ontario for different fundamental diagrams. . . . .	142
C.25 Space mean speeds on I-15 Ontario for different fundamental diagrams. . . . .	143

# List of Tables

3.1	Summary of the calibration and validation literature survey, macroscopic models. . . . .	41
3.2	Summary of the calibration and validation literature survey, microscopic models. . . . .	42
3.3	Parameter values retrieved from a calibration example using Gunes' method. . . . .	48
4.1	Summary of the measurement loader's parameters. . . . .	61
4.2	Summary of the highway model's parameters. . . . .	62
4.3	Summary of the ensemble Kalman filter's parameters. . . . .	63
5.1	Parameter results of Version I of the modified automatic empirical calibration method. . . . .	67
5.2	Parameter results of Version II of the modified automatic empirical calibration method. . . . .	69
5.3	Parameter results of Version III of the modified automatic empirical calibration method. . . . .	70
6.1	Summary of the fundamental diagram calibration set-up. . . . .	87
7.1	RMSE-results for I-880 using different fundamental diagrams. . . . .	93
7.2	NRMSE-results for I-880 using different fundamental diagrams. . . . .	94
7.3	RMSE-results for I-15 Ontario using different fundamental diagrams. . . . .	94
7.4	NRMSE-results for I-15 Ontario using different fundamental diagrams. . . . .	95
7.5	MAPE of the travel time results after calibration of I-880. . . . .	100
7.6	MAPE of the travel time results after calibration of I-880. . . . .	104
7.7	MAPE of the travel time results after calibration of I-15 Ontario. . . . .	109

# Chapter 1

## Introduction

Today's traffic society has numerous methods available for estimating the traffic state on a highway. The methods can be based on different theories and approaches e.g. kinematic wave theory, statistical theory or queueing theory. The methods based on kinematic wave theory are often derived from Lightham-Whitham-Richards partial differential equation (the LWR PDE) and discretized with a Godunov scheme. This kind of highway representation is also known as the cell transmission model (CTM). Such macroscopic simulation models usually have numerous parameters. These parameters have to be calibrated so that the model can be validated and produce as accurate results as possible.

Traffic model calibration and validation is a time consuming effort that has to be conducted for each individual model. The literature available on the subject is numerous but still does not define what calibration. Even the meaning of the central terminology might differ depending on which kind of source that is consulted and applications of model calibration, i.e. the method of choice, are often tailored for the individual model.

In 2008 a project called Mobile Century was driven by the California Center for Innovative Transportation (CCIT, later part of the Partners for Advanced Transportation Technology, or PATH), the Nokia Research Center (NRC) and the University of California, Berkeley [1]. The project's purpose was mainly a proof of concept; to prove that it is possible to collect data with GPS-equipped devices travelling with vehicles and use such observations for traffic state estimation, in real-time and in conjunction with a cell transmission model tailored for assimilation of velocity observations. The outcome was deemed successful in the controlled environment and the system from Mobile century evolved into the Mobile Millennium project for implementation on a more ambitious scale.

The purpose of the Mobile Millennium highway state estimation model (henceforth highway model) was to integrate and fuse data from different sources, such as stationary sensors and probes, combined with a cell transmission model in order to predict the traffic state on highways [1]. During the development and the research with the traffic state estimation model, the only calibration that was made, was ad-hoc.

Since there are no specified calibration methods for calibrating parameters in the highway model, this model will be the test subject for the thesis. Even though the literature points out that the usual approach is tailor made calibra-

tion methods for different models, this thesis will try to create a more general framework. This framework should be able to be applied to other models than the highway model. This thesis will also present two procedures for calibrating different types of parameters<sup>1</sup>.

## 1.1 Purpose and objectives

The purpose of this thesis is to develop a framework for calibrating a traffic state space model. Since the definition of calibration and validation can differ between articles, authors or other sources, it is required to define the concepts of calibration and validation in this thesis.

The framework should fulfil some criteria. The second criteria is to provide a general working process for how to calibrate and validate a traffic state space model. It should also provide some calibration procedures as examples of how to estimate parameters connected to the model. Another purpose is to provide some results from calibrated test cases using the calibration framework for a certain traffic state estimation model and thereby validate the framework.

The objective with this thesis, is to produce a calibration framework for a traffic state estimation model. The framework should contain certain parts. The first part should be a general working process that describes the procedure in which the calibration should be made. It should also contain a general calibration procedure as well as a parameter specific procedure. Another aim is to measure and evaluate the performance of the traffic model when using calibrated parameter sets for the test sites.

## 1.2 Method

The two major approaches for developing and creating the framework was a literature survey as well as a system analysis.

The literature survey was conducted as to bring in knowledge about the Mobile Millennium system, specifically the cell transmission model for velocity and data assimilation with the ensemble Kalman filter. In parallel, the literature survey also included the topics calibration, validation, empirical parameter estimation and black-box calibration so as to form the backbone of the calibration framework.

Even though the framework as such can be used for other systems, it was developed for a specific traffic state space model, the Mobile Millennium system, being in the development stage and a large research topic at that, an extensive system analysis was needed. The primary goals for doing the analysis was to understand and identify all issues in the system. The analysis pointed out that it was necessary to make modifications to the system, so that the system could be calibrated using the calibration framework.

## 1.3 Limitations

This thesis will mainly focus on two different calibration procedures for calibrating parameters connected to the highway model. The two methods consists

---

<sup>1</sup>These procedures are exchangeable in the framework.

of a black box calibration method called the complex method and the other is an empirical calibration method for calibration of the parameters connected to the fundamental diagram.

The data sources will be of three different kinds; probe data, inductive loop detector data (from Caltrans Performance Measurement System, PeMS) and Bluetooth data. Only limited data sets were available during the thesis. Processing of raw traffic data is outside the scope of this thesis although some quality assessment of filtered data will be conducted as part of the framework.

Due to lack of time, only two different calibration procedures was tested during the thesis in the frame work. Due to the factors mentioned, the performance of the calibration procedures will not be compared against other calibration methods.

Only off-line calibration of parameters is addressed in this thesis. This means that the framework presents calibration methods that are used with historical data when calibrating parameters.

## 1.4 Outline

The outline for the rest of this thesis is as follows; first a literature review will be presented in chapter. It consists of two chapters where the first chapter, chapter 2, will focus on the transformation of the CTM to the CTM-v, the network algorithm and the theory behind the ensemble Kalman filter. The second part of the literature review, chapter 2, will focus on the theories behind the calibration framework and calibration procedure. The purpose of the literature review is to present the theoretical foundation on which this thesis resides.

The following chapters will be the main contribution of this thesis. Chapter 4, presents the system description, which have the purpose of presenting a more detailed overview of the highway models and the relevant parameters connected to each part. It will also present the data flow of the highway model. After the system description, in 5, the implementation of the calibration framework together with the calibration procedures chapter are introduced. These two chapters system description and the implementation chapter, describes for the reader how the author developed and adapted the calibration procedures and integrated them into the highway model.

To prove that the calibration framework and the calibration procedures work, experiments are conducted. In chapter 6 the test sites, which is the subjects when conducting the experiments, as well as the layout are introduced and presented. Chapter 7 summarizes and concludes the results from conducting the experiments. The two last chapters in this thesis is the chapter 8 and chapter 9. Chapter 8 contains the analysis and discussion of the results from the work made. The final chapter which is chapter 9, concludes the work made in thesis and gives propositions for future work.

## Chapter 2

# Literature Review Part I: Highway Model

This chapter provides the literature review over the theoretical background for the highway model in the Mobile Millennium system [2]. This highway model is the traffic state space model that will be the experimental subject for the calibration framework presented in this thesis. The chapter describes the mathematics behind the highway model that lie within the scope of the thesis. The main focus lie on derivatio of the velocity based cell transmission model and the ensemble Kalman filter.

### 2.1 Traffic flow model review

This section will present a brief explanation to the Lightham-Whitham-Richard's (LWR) kinematic wave theory with a small introduction to cell transmission models (CTM), which is a recognized method for describing traffic states over time and space. A review of different flux functions and the Godunov scheme will be presented. It will also introduce an inversion of the CTM into a velocity based cell transmission model for which a network update algorithm is applied.

Macroscopic traffic models often use the CTM version from the Lightham-Whitham-Richard's partial differential equation (LWR PDE). The theory was developed by Lightham, Whitham (1955), see [3] and Richards (1956), see [4]. To express the flow as a function of density the PDE utilize a fundamental diagram. This diagram have evolved in many ways and there are different versions. This report will give a brief explanation for three different fundamental diagrams; Greenshields [5], Daganzo-Newell [6] and the Hyperbolic-Linear fundamental diagram and how they are connected to the LWR PDE.

The outline for the traffic model review section of this chapter is as follows. First a brief network representation to explain how a highway is simplified and modelled. Thereafter an introduction to the theories behind the traffic flow model. Next, a transformation from the density domain to the velocity domain will be introduced and explained. After a more detailed and link specific section, the cell transmission model for velocity (CTM-v) model is introduced. This kind of modelling are generally used to represent homogeneous stretches of road; a network representation these homogeneous stretches of road are defined as links.

To be able to extend the link modelling to a network representation, junctions are introduced. The junctions acts as connection between links and distributes the traffic flows over the outgoing links. Lastly, the network algorithm of how to estimate the state for the whole network is provided.

### 2.1.1 Transformation of the Lightham-Whitham-Richard partial differential equation

In the LWR theory there is a partial differential equation known as the LWR PDE that expresses how the density  $\rho$  evolves for a certain stretch of road with a length of  $L$ , for a time period  $T$ . The LWR PDE is (2.1).

$$\begin{cases} \frac{\partial \rho(x,t)}{\partial t} + \frac{\partial Q \rho(x,t)}{\partial x} = 0, & (x,t) \in (0,L) \times (0,T) \\ \rho(x,0) = \rho_0(x), \rho(0,t) = \rho_l(x), \rho(L,t) = \rho_r(x) \end{cases} \quad (2.1)$$

where  $Q(\cdot)$  is the fundamental digram<sup>1</sup> for  $\rho \in [0, \rho_{\max}]$ , where  $\rho_{\max}$  is the maximum density. Note that  $Q(\cdot)$  often is assumed to be concave and piecewise differentiable. The initial condition, left boundary condition and right boundary condition are expressed by the terms  $\rho_0(\cdot)$ ,  $\rho_l(\cdot)$  and  $\rho_r(\cdot)$ . It is assumed that the flow can be described as (2.2).

$$Q = \rho V(\rho) \quad (2.2)$$

where the velocity  $V(\cdot)$  is a function of the density  $\rho$ .

To be able to characterize the behaviour of the LWR PDE correctly, under the assumption that the desired density or measured density is not always equal to the modelled density gives that boundary conditions are needed to solve the PDE. They can be either strong or weak, depending on how the network<sup>2</sup> representation is formulated. The boundary conditions will be described more closely in the next part. Notations for the initial conditions as well as for the boundary conditions is as described above.

The initial boundary condition is directly corresponding to the density along the stretch of the modelled highway at time  $t = 0$ , this gives (2.3).

$$\rho(x,0) = \rho_0(x), x \in [0, L] \quad (2.3)$$

where  $x$  is a specific stretch of road. In general, when trying to estimate the state of a traffic system on a stretch of road, it is generally difficult to estimate the initial conditions. To be able to cope with this problem, a system warm up can be used. By letting the system run over a sufficiently long time period, the significance of the initial conditions will be negligible. This is also called the flush out effect [2]. However, to solve (2.1) for the left and right boundary conditions, [2] states that at least weak boundary conditions are used<sup>3</sup>. By introducing the weak boundary conditions, it is meant that the modelled density

<sup>1</sup>The fundamental digram express the relation between flow and density.

<sup>2</sup>The network is a representation of the modelled highway, with junctions e.g. on and off ramps.

<sup>3</sup>By weak boundary conditions, it is meant that they are no longer required to hold absolutely. In this way it is possible to use concepts from linear algebra to solve PDE:s. In this case it used to transform the PDE into an entropy function to be able to prove that a solution exists.

and the desired density which is represented by  $\rho(l, t) = \rho_l(t)$  and  $\rho(r, t) = \rho_r(t)$  not always have to be absolute for all  $t$ .

According to [7] a simplification of the weak boundary conditions can be expressed as (2.4) and (2.5). However for the mathematical specifics of the hyperbolic conservation law, which the LWR PDE is, the reader is referred to [8].

$$\left. \begin{array}{l} \text{(I) } \rho(l, t) = \rho_l(t) \text{ and } Q'(\rho_l(t)) \geq 0 \text{ or} \\ \text{(II) } Q'(\rho(l, t)) \leq 0 \text{ and } Q'(\rho_l(t)) \leq 0 \text{ or} \\ \text{(III) } Q'(\rho(l, t)) \leq 0 \text{ and } Q'(\rho_l(t)) \geq 0 \text{ and } Q(\rho(l, t)) \leq Q(\rho_l(t)) \end{array} \right\} \text{ for all } t \quad (2.4)$$

and

$$\left. \begin{array}{l} \text{(I) } \rho(r, t) = \rho_r(t) \text{ and } Q'(\rho_r(t)) \leq 0 \text{ or} \\ \text{(II) } Q'(\rho(r, t)) \geq 0 \text{ and } Q'(\rho_r(t)) \geq 0 \text{ or} \\ \text{(III) } Q'(\rho(r, t)) \geq 0 \text{ and } Q'(\rho_r(t)) \leq 0 \text{ and } Q(\rho(r, t)) \leq Q(\rho_r(t)) \end{array} \right\} \text{ for all } t \quad (2.5)$$

where  $Q'(\cdot) = \frac{dq}{d\rho}$ . Case (I) represents that the desired density is equal to the modelled density. Case (II) represents the traffic state where both the desired inflow as well as the modelled inflow, therefore the boundary condition is pushed by the current state. Case (III) represents that the boundary condition is not able to push the inflow. This due to that the modelled density is in congestion and larger than the uncongested desired density. According to [7] this formulation ensures that it is possible to find an existing and unique entropy solution for a bounded domain.

So to be able to transform the density based CTM to a velocity based CTM- $v$  it is needed to consider the density entropy solution, see (2.6), everything in line with [2]. This is also needed due to discontinuity that can appear in (2.1). The weak entropy solution for the density evolution model, the LWR PDE, can be written as (2.6). Note that earlier in the report, it is mentioned that the fundamental diagram is generally differential, piecewise linear and concave. However it is not possible to differentiate the Daganzo-Newell fundamental diagram, which in turn rends it impossible to use in the entropy solution. It is a necessity that the fundamental diagram used in the entropy function can be twice differentiable (which the Daganzo-Newell digram is not) and is strictly concave with super linear growth. This is the motivation that [2] uses to introduce the hyperbolic-linear fundamental diagram.

$$\begin{aligned} & \int_0^L \int_0^T (|\rho(x, t) - k| \frac{\partial}{\partial t} \varphi(x, t) + \text{sgn}(\rho(x, t) - k)(Q(\rho(x, t)) - Q(k)) \\ & + \frac{\partial}{\partial x} \varphi(x, t)) dt dx + \int_0^L \int_0^T \text{sgn}(k)(Q(\Upsilon \rho(x, t)) - Q(k)) \cdot n \varphi(x, t) dt dx \geq 0 \end{aligned} \quad (2.6)$$

$$\forall \varphi \in C_c^2([0, L] \times [0, T]; \mathbb{R}_+), \forall k \in \mathbb{R}$$

where  $\Upsilon$  is the trace operator and  $n$  is the normal vector to the domain. To transform the solution of the density based LWR PDE, which means that additional boundary conditions for the initial conditions, left boundary conditions

and right boundary conditions is needed to be reformulated for this specific case. Not only in the weak sense but in a strong sense as well.

$$\sup_{k \in D(u(0,t), u_l(t))} \text{sgn}(u(0,t) - u_l(t))(F(u(0,t)) - F(k)) = 0, a.e. t > 0. \quad (2.7)$$

where  $D(x, y) = [\inf(x, y), \sup(x, y)]$ . 2.7 is the proper weak description for the trace of the solution for  $u(0, t)$  and  $u_l(t)$  for the left boundary condition of 2.1.

### 2.1.2 Transformation to the velocity domain

To be able to extend the LWR-v to a network model with proper boundary conditions, it is required to formulate them as strong boundary conditions instead of weak conditions. In [2], they prove that the strong boundary conditions for this case can be formulated as (2.8)-(2.9).

$$\left. \begin{array}{l} Q'(u(0, t)) \geq 0 \text{ and } Q'(u_l(t)) \geq 0 \\ Q'(u(0, t)) \leq 0 \text{ and } Q'(u_l(t)) \leq 0 \text{ and } u(0, t) = u_l(t) \\ Q'(u(0, t)) \leq 0 \text{ and } Q'(u_l(t)) > 0 \text{ and } Q(u(0, t)) > Q(u_l(t)) \end{array} \right\} \begin{array}{l} \text{xor} \\ \text{xor} \\ \text{xor} \end{array} \left. \begin{array}{l} \\ \\ \end{array} \right\} \text{a.t. } t > 0 \quad (2.8)$$

and

$$\left. \begin{array}{l} Q'(u(L, t)) \leq 0 \text{ and } Q'(u_r(t)) \leq 0 \\ Q'(u(L, t)) \geq 0 \text{ and } Q'(u_r(t)) \geq 0 \text{ and } u(L, t) = u_r(t) \\ Q'(u(L, t)) \geq 0 \text{ and } Q'(u_r(t)) < 0 \text{ and } Q(u(L, t)) > Q(u_r(t)) \end{array} \right\} \begin{array}{l} \text{xor} \\ \text{xor} \\ \text{xor} \end{array} \left. \begin{array}{l} \\ \\ \end{array} \right\} \text{a.t. } t > 0 \quad (2.9)$$

Note that  $u(\cdot)$  is measured data,  $l$  and  $r$  is the boundaries,  $Q$  is the fundamental diagram and  $u(\cdot, \cdot)$  is the solution.

Lastly, to be able to show that it is possible to recreate the LWR PDE for the velocity domain with the same attributes as the density based LWR PDE, [2] introduces a modified velocity based entropy function that solves the PDE in its weak sense<sup>4</sup>, see (2.10).

$$\int_0^L \int_0^T \left( P(v(x, t)) \frac{\partial \varphi}{\partial t}(x, t) + Q(P(v(x, t))) \frac{\partial \varphi}{\partial x}(x, t) \right) dx dt + \int_0^L P(v_0(x)) \varphi(x, 0) dx = 0, \quad \forall \varphi \in C_c^2([0, L] \times [0, T]) \quad (2.10)$$

The first step to formulate the CTM-v, [2] wants to formulate the LWR-v PDE conservation law in the velocity based domain according to (2.11).

$$\begin{cases} \frac{\partial}{\partial t} v(x, t) + \frac{\partial}{\partial x} R(v(x, t)) = 0 \\ v(x, 0) = v_0(x) \end{cases} \quad (2.11)$$

<sup>4</sup>By its weak sense, it is meant that a solution for the PDE cannot be guaranteed for the whole domain.

where  $R(v)$  is a convex, velocity based fundamental diagram.  $Q = \rho v$  is invertible, due to a strictly linear relationship.

This is done with the same approach as earlier, by introducing weak boundary conditions for the entropy function (2.10). The boundary conditions then needs to be reformulated, see (2.12) and (2.13) for (2.10).

$$\left. \begin{array}{l} u(0, t) = u_l(t) \\ Q'(u(0, t)) \leq 0 \text{ and } Q'(u_l(t)) \leq 0 \text{ and } u(0, t) \neq u_l(t) \\ Q'(u(0, t)) \leq 0 \text{ and } Q'(u_l(t)) > 0 \text{ and } Q(u(0, t)) \geq Q(u_l(t)) \end{array} \right\} \begin{array}{l} \text{xor.} \\ \text{xor} \end{array} \left. \vphantom{\begin{array}{l} u(0, t) = u_l(t) \\ Q'(u(0, t)) \leq 0 \text{ and } Q'(u_l(t)) \leq 0 \text{ and } u(0, t) \neq u_l(t) \\ Q'(u(0, t)) \leq 0 \text{ and } Q'(u_l(t)) > 0 \text{ and } Q(u(0, t)) \geq Q(u_l(t)) \end{array}} \right\} \text{a.e. } t > 0 \quad (2.12)$$

and

$$\left. \begin{array}{l} u(L, t) = u_r(t) \\ Q'(u(L, t)) \geq 0 \text{ and } Q'(u_r(t)) \geq 0 \text{ and } u(L, t) \neq u_r(t) \\ Q'(u(L, t)) \geq 0 \text{ and } Q'(u_r(t)) < 0 \text{ and } Q(u(L, t)) \geq Q(u_r(t)) \end{array} \right\} \begin{array}{l} \text{xor.} \\ \text{xor} \end{array} \left. \vphantom{\begin{array}{l} u(L, t) = u_r(t) \\ Q'(u(L, t)) \geq 0 \text{ and } Q'(u_r(t)) \geq 0 \text{ and } u(L, t) \neq u_r(t) \\ Q'(u(L, t)) \geq 0 \text{ and } Q'(u_r(t)) < 0 \text{ and } Q(u(L, t)) \geq Q(u_r(t)) \end{array}} \right\} \text{a.e. } t > 0 \quad (2.13)$$

where  $u_l(\cdot)$ ,  $u_r(\cdot)$  are of non differentiable functions. The functions  $u_l(\cdot)$  and  $u_r(\cdot)$  strong boundary conditions applied to the left and right boundaries<sup>5</sup>.

### 2.1.3 Discretization

To discretize (2.13), [2] uses the Godunov scheme. This is a method to discretize the LWR PDE and is commonly used in the traffic society. The Godunov scheme is a numerical approximation to the weak solution of the PDE in its conservative form. It discretize the PDE in both time and space. In other words, by applying the Godunov scheme it is possible to reformulate the LWR-v PDE as a non-linear, dynamic system that is discrete in space and time. In other words, the cell transmission model. To ensure numerical stability, it is required that  $v_{\max} \frac{\Delta T}{\Delta x} \leq 1$ , where  $v_{\max}$  is the maximum modelled velocity. In the descretization, the space step  $i \in \{0, \dots, i_{\max}\}$  with length  $\Delta x$  is introduced as well as the time step  $n \in \{0, \dots, n_{\max}\}$  with length  $\Delta T$ . (2.14) is the cell transmission function for the LWR PDE.

$$\rho_i^{n+1} = \rho_i^n - \frac{\Delta T}{\Delta X} (G(\rho_i^n, \rho_{i+1}^n) - G(\rho_{i-1}^n, \rho_i^n)) \quad (2.14)$$

where  $\rho$  is the density and  $G(\rho_1, \rho_2)$  is the Godunov density flow function, where the Godunov velocity flow function is defined as (2.15).

$$G(\rho_1, \rho_2) = \begin{cases} Q(\rho_2) & \text{if } \rho_{cr} \leq \rho_2 \leq \rho_1 \\ Q(\rho_{cr}) & \text{if } \rho_2 \leq \rho_{cr} \leq \rho_1 \\ Q(\rho_1) & \text{if } \rho_2 \leq \rho_1 \leq \rho_{cr} \\ \min(Q(\rho_1), Q(\rho_2)) & \text{if } \rho_1 \leq \rho_2 \end{cases} \quad (2.15)$$

where  $Q(\cdot)$  is the fundamental diagram,  $\rho_1$  is the upstream density and  $\rho_2$  is the downstream density.

Since (2.14) is for estimating the traffic state for a certain cell in a certain time step using density, a transformation to the velocity domain is needed. The

<sup>5</sup>This to ensure that a solution to the PDE can be found.

inversion is made by stating that  $Q(\rho) = \tilde{Q}(v) = V^{-1}(v)v$ . It is possible if  $V(\cdot)$  is strictly decreasing and  $\rho_1 \leq \rho_2$  while  $v_1 = V(\rho_1)$  and  $v_2 = V(\rho_2)$  and  $v_1 \leq v_2$ , where  $v_1$  is the upstream velocity and  $v_2$  is the downstream velocity. It is also required that  $V(\cdot)$  is monotonically decreasing and invertible. By using this statement, the relationship  $Q(\rho) = \tilde{Q}(v) = V^{-1}(v)v$  can be assumed. By inverting (2.14) and transforming (2.15), results in the velocity based cell transmission model, the CTM-v, see (2.16) and (2.17). Note that if the fundamental diagram is not affine, the inversion is needed to be made after the discretization to yield the CTM-v.

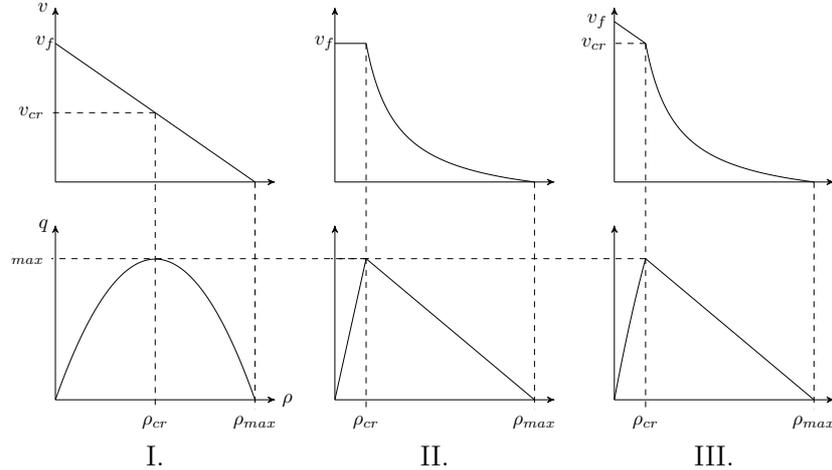
$$v_i^{n+1} = V \left( V^{-1}v_i^n - \frac{\Delta T}{\Delta x} (\tilde{G}(v_i^n, v_{i+1}^n) - \tilde{G}(v_{i-1}^n, v_i^n)) \right) \quad (2.16)$$

where the Godunov velocity flow  $\tilde{G}(v_1, v_2)$  is given by (2.17).

$$\tilde{G}(v_1, v_2) = \begin{cases} \tilde{Q}(v_2) & \text{if } v_{cr} \leq v_2 \leq v_1 \\ \tilde{Q}(v_{cr}) & \text{if } v_2 \leq v_{cr} \leq v_1 \\ \tilde{Q}(v_1) & \text{if } v_2 \leq v_1 \leq v_{cr} \\ \min(\tilde{Q}(v_1), \tilde{Q}(v_2)) & \text{if } v_1 \leq v_2 \end{cases} \quad (2.17)$$

#### 2.1.4 Fundamental diagrams

This section will provide an introduction to the Greenshields [5], Daganzo-Newell [6] and the Hyperbolic-Linear fundamental diagram. It will also provide the inversion of the Hyperbolic-Linear fundamental diagram [2], which is later used in the CTM-v.



**Figure 2.1:** Three different fundamental diagrams is pairwise shown in this figure. The first pair I is the Greenshields fundamental diagram where the upmost diagram is represented in the velocity density domain and the lower is represented in the flow density domain. The second pair is the Daganzo-Newell fundamental diagram and the third pair is the Hyperbolic-Linear fundamental diagram.

### Greenshields fundamental diagram

The Greenshields fundamental diagram was invented by Bause D. Greenshields in 1934. It was observed that the speed depended on the density. In [5], they state the affine velocity function as (2.18). It expresses a linear relationship between speed and density.

$$v = V_G(\rho) = v_{\max} \left( 1 - \frac{\rho}{\rho_{\max}} \right) \quad (2.18)$$

where  $v$  is the estimated velocity and  $\rho$  is the density,  $G$  is a notation for Greenshields fundamental diagram, in graphical representation, see figure 2.1.

### Daganzo-Newell fundamental diagram

In July 1993, Carlos F. Daganzo presented a cell transmission model in [9], as a dynamic representation of highway traffic. In this article, instead of using the Greenshields fundamental diagram, the author presents another fundamental diagram, see figure 2.1. A triangular where the differentiation between the congested regime and the uncongested regime is strengthened. The fundamental diagram is expressed by (2.19).

$$v = V_{DN}(\rho) \begin{cases} v_{\max}, & \text{if } \rho \leq \rho_{cr} \\ -w_f \left( \frac{\rho_{\max}}{\rho} \right), & \text{otherwise} \end{cases} \quad (2.19)$$

where  $v_{\max}$  is the maximum velocity (free flow velocity),  $\rho_{\max}$  is the maximum density,  $w_f$  is the backward propagating shock wave velocity and  $DN$  is a notation for Daganzo-Newell fundamental diagram. The Daganzo-Newell fundamental diagram is not invertible since it is not strictly monotonic in free flow.

### Hyperbolic-Linear fundamental diagram

The main reason to why the Hyperbolic-Linear fundamental diagram is introduced is that the Daganzo-Newell fundamental diagram is not invertible. So the Hyperbolic-Linear fundamental diagram is a combination between the Greenshields and the Daganzo-Newell fundamental diagram. For the mathematical expression of the Hyperbolic-Linear, see (2.20) and for graphical presentation see 2.1.

$$v = V_{HL}(\rho) = \begin{cases} v_{\max} \left( 1 - \frac{\rho}{\rho_{\max}} \right), & \text{if } \rho \leq \rho_{cr} \\ -w_f \left( 1 - \frac{\rho_{\max}}{\rho} \right), & \text{otherwise} \end{cases} \quad (2.20)$$

where  $\rho_{\max}$  is the maximum density,  $\rho_{cr}$  is the critical density,  $w_f$  is the backward propagating shock wave velocity,  $v_{\max}$  is the free flow velocity and HL stands for that this fundamental diagram is the Hyperbolic-Linear fundamental diagram.

Since the system requires that everything should be expressed in the velocity domain, an inversion of the Hyperbolic-Linear fundamental diagram is needed.

To be able to invert the fundamental diagram, continuity is required, especially where  $\rho = \rho_{cr}$ . To achieve this, the authors in [2] introduce the continuity constraint 2.21.

$$\frac{\rho_{cr}}{\rho_{\max}} = \frac{w_f}{v_{\max}} \quad (2.21)$$

After the continuity constraint is introduced, it is possible to invert (2.21), and the inverted fundamental diagram is formulated as 2.22.

### 2.1.5 Cell transmission model for velocity

This section presents a summary of the CTM-v model proposed by [2] as well as all the reasoning previously made. Just as for the network representation, all declarations are in line with those of [2]. To express the density as a function of speed (2.22) is used.

$$\rho = V_{\text{HL}}^{-1}(v) = \begin{cases} \rho_{\max} \left( 1 - \frac{v}{v_{\max}} \right), & \text{if } v \geq v_{cr} \\ \rho_{\max} \left( \frac{1}{1 + \frac{v}{w_f}} \right), & \text{otherwise} \end{cases} \quad (2.22)$$

In (2.22),  $V_{\text{HL}}^{-1}(v)$  denotes the inverted Hyperbolic-Linear velocity function, which is the inverse function of the approximated Daganzo-Newell velocity function (see e.g. [2] for details).  $\rho$  is the vehicle density,  $v$  is the velocity,  $\rho_{\max}$  is the maximum, or jam, density,  $v_{\max}$  is the free flow speed,  $v_{cr}$  is the critical velocity corresponding to the critical density  $\rho_{cr}$  and  $w_f$  is the backwards propagating shock wave velocity.

The predicted velocity in the next time step for cell  $i$  is calculated via (2.23). This expression is what is called the Cell Transmission Model for velocity, or CTM-v.

$$v_i^{n+1} = V \left( V^{-1}(v_i^n) - \frac{\Delta T}{\Delta x} \left( \tilde{G}(v_i^n, v_{i+1}^n) - \tilde{G}(v_{i-1}^n, v_i^n) \right) \right) \quad (2.23)$$

where  $\tilde{G}$  is the transformed Godunov velocity flux.

Transforming  $\tilde{G}$  and using (2.22), it is possible to yield the Hyperbolic-Linear model (2.24).

$$\tilde{G}(v_1, v_2) = \begin{cases} \tilde{Q}(v_2) = v_2 \rho_{\max} \left( \frac{1}{1 + \frac{v_2}{w_f}} \right), & \text{if } v_{cr} \geq v_2 \geq v_1 \\ \tilde{Q}(v_{cr}) = v_{cr} \rho_{\max} \left( 1 - \frac{v_{cr}}{v_{\max}} \right), & \text{if } v_2 \geq v_{cr} \geq v_1 \\ \tilde{Q}(v_1) = v_1 \rho_{\max} \left( 1 - \frac{v_1}{v_{\max}} \right), & \text{if } v_2 \geq v_1 \geq v_{cr} \\ \min \left( V_{\text{HL}}^{-1}(v_1) v_1 V_{\text{HL}}^{-1}(v_2) v_2 \right), & \text{if } v_1 \geq v_2 \end{cases} \quad (2.24)$$

For the first and last cell however, the velocity at the next time step is calculated with (2.25) instead.

$$\begin{aligned}
v_0^{n+1} &= V\left(V^{-1}(v_0^n) - \frac{\Delta T}{\Delta x}\left(\tilde{G}(v_0^n, v_1^n) - \tilde{G}(v_{-1}^n, v_0^n)\right)\right) \\
v_{i_{\max}}^{n+1} &= V\left(V^{-1}(v_{i_{\max}}^n) - \frac{\Delta T}{\Delta x}\left(\tilde{G}(v_{i_{\max}}^n, v_{i_{\max}+1}^n) - \right. \right. \\
&\quad \left. \left. \tilde{G}(v_{i_{\max}-1}^n, v_{i_{\max}}^n)\right)\right)
\end{aligned} \tag{2.25}$$

So at the boundaries of an edge, (2.23) will reference to points  $v_{-1}^n$  and  $v_{i_{\max}+1}^n$ . These points are not within the physical domain, but are given by boundary conditions. (2.24) can result in situations where the boundary conditions are not imposed on the physical domain.

### 2.1.6 Model representation of junctions

Physical restrictions require that the flow distributions at the junctions are solved. The first two restrictions (vehicle conservation and an imposed routing scheme) can be represented according to (2.26) and (2.27).

$$\sum_{e_{\text{in}} \in \mathcal{I}_j} \tilde{Q}_{e_{\text{in}}}(v_{e_{\text{in}}}(L_{e_{\text{in}}}, t)) = \sum_{e_{\text{out}} \in \mathcal{O}_j} \tilde{Q}_{e_{\text{out}}}(v_{e_{\text{out}}}(0, t)) \tag{2.26}$$

$$\sum_{e_{\text{out}} \in \mathcal{O}_j} \alpha_{j, e_{\text{out}}, e_{\text{in}}} = 1 \tag{2.27}$$

where  $\alpha_{j, e_{\text{out}}, e_{\text{in}}}$  is the allocation parameter and  $\mathbf{A}_j \in [0, 1]^{|\mathcal{O}_j| \times |\mathcal{I}_j|}$  is the allocation matrix for junction  $j$ . We also have the relations  $\mathbf{A}_j(e_{\text{out}}, e_{\text{in}}) = \alpha_{j, e_{\text{out}}, e_{\text{in}}}$  and  $\alpha_{j, e_{\text{out}}, e_{\text{in}}} \geq 0$ . The purpose of the allocation parameters in the allocation matrix is to distribute the traffic flow from the incoming edges to the outgoing edges according to routing information.

The restrictions expressed by (2.26) and (2.27) cannot always be fulfilled since they combined impose strong boundary conditions at the end of the edges (note the equality). To solve this issue the third restriction is included (traffic flow is maximized over the junction) and the strong boundary conditions represent upper bounds for that problem. The result is an optimization procedure (a LP-problem). The LP-problem solves the exiting flow on each incoming edge for the junction  $j$ . The dummy vector variable  $\xi \in \mathbb{R}^{|\mathcal{I}_j|}$  is introduced for the incoming edges  $e_{\text{in}}$  of junction  $j$ . It is possible to conclude the three restrictions into the following LP-problem.

$$\begin{aligned}
&\text{maximize: } \mathbf{1}^T \xi \\
&\text{Subject to: } \mathbf{A}_j \xi \leq \gamma_{\mathcal{O}_j}^{\max} \\
&\quad \mathbf{0} \leq \xi \leq \gamma_{\mathcal{I}_j}^{\max}
\end{aligned} \tag{2.28}$$

where the upper bounds are set by  $\gamma_{\mathcal{O}_j}^{\max} := (\gamma_{e_{\text{out},1}}^{\max}, \dots, \gamma_{e_{\text{out},|\mathcal{O}_j|}}^{\max})$  and  $\gamma_{\mathcal{I}_j}^{\max} := (\gamma_{e_{\text{in},1}}^{\max}, \dots, \gamma_{e_{\text{in},|\mathcal{I}_j|}}^{\max})$ . The LP-problem expressed by (2.28) will have an optimal solution  $\xi^*$  for junction  $j$ . After estimating the maximum admissible outgoing flux, it is possible to determine the state for  $\tilde{G}(v_{-1}^n, v_0^n)$  and  $\tilde{G}(v_{i_{\max}-1}^n, v_{i_{\max}}^n)$  in (2.25) by (2.29).

$$\begin{aligned}\bar{G}_{e_{\text{in}}}(v_{i_{\text{max}}}^n, v_{i_{\text{max}}+1}^n) &= \xi_{e_{\text{in}}}^*, \\ \bar{G}_{e_{\text{out}}}(v_{-1}^n, v_0^n) &= \sum_{e_{\text{in}} \in \mathcal{I}_j} \alpha_{j, e_{\text{out}}, e_{\text{in}}} \xi_{e_{\text{in}}}^*\end{aligned}\quad (2.29)$$

The maximum admissible outgoing and incoming maximum flow in (2.28) is defined via (2.30) and (2.31) respectively.

$$\gamma_{e_{\text{out}}}^{\text{max}}(v_{e_{\text{out}}}(0, t)) = \begin{cases} \rho_{\text{max}} \left(1 - \frac{v_{cr, e_{\text{out}}}}{v_{\text{max}}}\right) v_{cr, e_{\text{out}}} & \text{if } v_{e_{\text{out}}}(0, t) \in [v_{cr, e_{\text{out}}}, v_{\text{max}, e_{\text{out}}}] \\ \rho_{\text{max}} \left(\frac{1}{1 + \frac{v_{e_{\text{out}}}(0, t)}}{w_f}}\right) v_{e_{\text{out}}}(0, t) & \text{if } v_{e_{\text{out}}}(0, t) \in [0, v_{cr, e_{\text{out}}}] \end{cases}\quad (2.30)$$

and

$$\gamma_{e_{\text{in}}}^{\text{max}}(v_{e_{\text{in}}}(L_{e_{\text{in}}}, t)) = \begin{cases} \rho_{\text{max}} \left(1 - \frac{v_{e_{\text{in}}}(L_{e_{\text{in}}}, t)}{v_{\text{max}}}\right) v_{e_{\text{in}}}(L_{e_{\text{in}}}, t) & \text{if } v_{e_{\text{in}}}(L_{e_{\text{in}}}, t) \in [v_{cr, e_{\text{in}}}, v_{\text{max}, e_{\text{in}}}] \\ \rho_{\text{max}} \left(\frac{1}{1 + \frac{v_{cr, e_{\text{in}}}}{w_f}}\right) v_{cr, e_{\text{in}}} & \text{if } v_{e_{\text{in}}}(L_{e_{\text{in}}}, t) \in [0, v_{cr, e_{\text{in}}}] \end{cases}\quad (2.31)$$

### 2.1.7 Network algorithm

The authors of [2], apart from deriving the CTM-v from the LWR PDE, defined an algorithm that progresses the velocity field in the network in time. The algorithm basically applies the CTM-v for each individual network edge and solves the LP-problem for each junction. The velocity field for the entire network, that is, for every cell  $i \in \{0, \dots, i_{\text{max}}\}$  on all edges, is

$$v^n := \left[ v_{0, e_0}^n, \dots, v_{i_{\text{max}}, e_0}^n, \dots, v_{0, e_{|\varepsilon|}}^n, \dots, v_{i_{\text{max}}, e_{|\varepsilon|}}^n \right]$$

The velocity at time  $t = (n+1)\Delta T$  is given by

$$v^{n+1} = \mathcal{M}[v^n]\quad (2.32)$$

Where  $\mathcal{M}[v]$  denotes the update algorithm presented below.

Step 1. For all junctions  $j \in \mathcal{J}$ :

Compute  $\gamma_{i_{\text{max}}, e_{\text{in}}}^n(v_{i_{\text{max}}, e_{\text{in}}}^n) \forall e_{\text{in}} \in \mathcal{I}_j$  and

$$\gamma_{0, e_{\text{out}}}^n(v_{0, e_{\text{out}}}^n) \forall e_{\text{out}} \in \mathcal{O}_j \text{ using (2.30)-(2.31).}$$

Solve the LP-problem (2.28) and update

$$\tilde{G}_{e_{\text{in}}}(v_{i_{\text{max}}}^n, v_{i_{\text{max}}+1}^n) \text{ and}$$

$$\tilde{G}_{e_{\text{out}}}(v_{-1}^n, v_0^n) \text{ using (2.29).}$$

Step 2. For all edges  $e \in \varepsilon$ :

Compute  $v_{i, e}^{n+1} \forall i \in \{1, \dots, i_{\text{max}, e}\}$

according to the CTM-v (2.23) and (2.25).

## 2.2 Data assimilation: The ensemble Kalman filter

One of the more noticeable features of the Mobile Millennium system is its ability to assimilate data in real time and offline mode from multiple sources together with the highway model. This increases model performance in the sense that it becomes more knowledgeable about the true velocity state on the road. This section will highlight the filtering technique that is used for data assimilation in the Mobile Millennium system.

Evensen [10] presented a new method for sequential data assimilation, later called the ensemble Kalman filter, a variation of the Kalman filter which originates from the 1960's [11]. Evensen's contribution was said to be a solution to unwanted unbounded error growth in the extended Kalman filter, another version of the Kalman filter, due to the simplifications in the error covariance estimation [10]. Evensen used Monte Carlo methods to forecast the error statistics since the error covariance approximation in the extended Kalman filter was deemed to be too costly from a computational viewpoint but also because the new approach would eliminate the unbounded error growth in the extended Kalman filter [10].

The main purpose of this section is to present the ensemble Kalman filter as well as some background in statistics, data assimilation, the original Kalman filter and the extended Kalman filter. This should make the reader aware of what these data assimilation techniques are, what an analysis scheme is, why neither the Kalman filter nor the extended Kalman filter can be used for data assimilation with the highway model update algorithm for velocity presented earlier and highlight some important simplifications made in the extended Kalman filter.

The outline of the Data Assimilation: The ensemble Kalman filter section is as follows; the next section includes some background to statistics that is carried throughout the description of Kalman filtering techniques. The best possible estimate of a state is then presented for circumstances where a prior state and measurements of the true state are known. After the introduction to state estimates with data assimilation the original Kalman filter is presented, followed by the extended Kalman filter and the ensemble Kalman filter. The usual pattern in this section is that the presentation is done in a scalar case first and then moved into a spatial domain.

### 2.2.1 Statistics

Before the Kalman filter or any of its variations are presented some statistical terms and definitions upon which the data assimilation technique is founded will be presented.

Assume a random variable  $\Psi$  that is continuous over its domain. The random variable has an associated function  $F(\psi)$  which is known as a distribution function. This distribution function is defined as (2.33).

$$F(\psi) = \int_{-\infty}^{\psi} f(\psi') d\psi' \quad (2.33)$$

where  $f(\psi)$  is a probability density function.  $f(\psi')$  is therefore the change in

probability (density) and the distribution function  $F(\psi)$  states the cumulative probability that  $\Psi$  takes a value less than, or equal to  $\psi$ . As (2.33) is defined the probability density function  $f(\psi)$  must be the first derivative of the distribution function  $F(\psi)$ :

$$f(\psi) = \frac{\partial F(\psi)}{\partial \psi} \quad (2.34)$$

The function  $f(\psi)$  expresses the probability of the random variable  $\Psi$  being equal to  $\psi$ . Note that  $f(\psi) \geq 0$  always holds, that the probability of  $\Psi$  taking a value within a very small interval is equal to  $f(\psi)d\psi$  on the one hand and on the other hand that the probability of  $\Psi$  to take a value in  $]-\infty, \infty[$ , is equal to one. The probability of  $\Psi$  being in an arbitrary interval  $[a, b]$  is defined by (2.35).

$$\Pr(\Psi \in [a, b]) = \int_a^b f(\psi) d\psi \quad (2.35)$$

The probability distribution called the Gaussian (or normal) distribution is commonly referred to in this thesis. The probability density functions for these kinds of distributions are characterized by their variance  $\sigma^2$ , mean value  $\mu$  and are defined by (2.36).

$$f(\psi) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\psi - \mu)^2}{2\sigma^2}\right) \quad (2.36)$$

### Bayesian statistics in 2-dimensional spaces

Consider a case with two random variables  $\Psi$  and  $\Phi$ . The joint probability density function expresses the probability of  $\Psi$  and  $\Phi$  occur together, this function is denoted as  $f(\psi, \phi)$ . The conditional probability density function expresses the probability of occurrence  $\Psi$  if occurrence  $\Phi$  already took place, this function is denoted  $f(\psi | \phi)$  and is defined by (2.37).

$$f(\psi | \phi) = \frac{f(\psi, \phi)}{f(\phi)} \Leftrightarrow f(\psi, \phi) = f(\psi | \phi)f(\phi) \quad (2.37)$$

where

$$f(\phi) = \int_{-\infty}^{\infty} f(\psi, \phi) d\psi \quad (2.38)$$

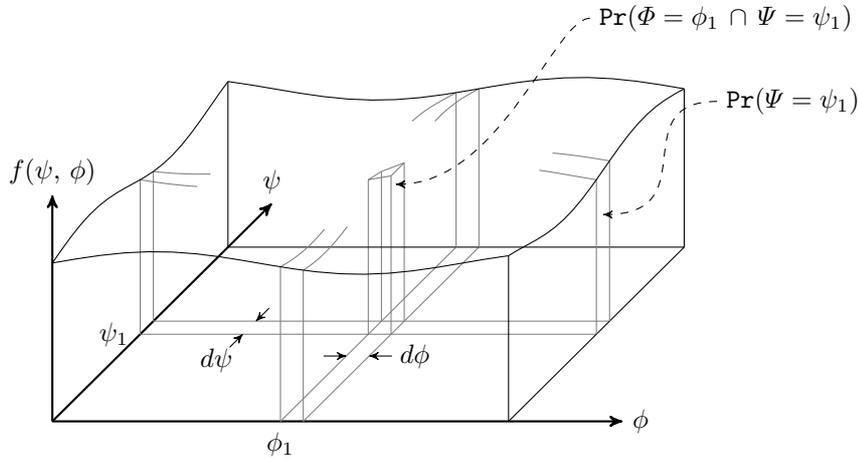
Equation (2.37) states that  $f(\psi, \phi) = f(\psi | \phi)f(\phi)$  where the right hand side interpreted as the likelihood  $\Psi$  given  $\Phi$  times the probability of  $\Phi$ . Note that if the occurrence  $\Psi$  is independent of the occurrence  $\Phi$  and vice versa is the joint probability density function  $f(\phi, \psi) = f(\phi)f(\psi)$ .

Equation (2.37) gives Bayes' theorem as (2.39) which states the conditional (after  $\Phi$ ) probability of  $\Psi$  which is more knowledgeable than the probability of  $\Psi$  alone since the data  $\Phi$  now is taken into account.

$$f(\psi | \phi) = \frac{f(\psi)f(\phi | \psi)}{f(\phi)} \quad (2.39)$$

Consider Figure 2.2 below for example, before any information of the event  $\Phi$  is  $\Pr(\Psi = \psi_1)$  a volume that stretches the entire  $\phi$ -space. This probability, or volume, considers the marginal distribution of  $\Psi$  alone and follows what is called a prior probability distribution. Assume now that  $\Psi$  and  $\Phi$  are depending events and that information about the event  $(\Phi = \phi_1)$  is known. This results in a conditional probability of  $\Psi$  given  $\Phi$  expressed by a new intersection volume  $\Pr(\Phi = \phi_1 \cap \Psi = \psi_1)$ , a new probability following a posterior probability distribution. According to the figure, the conditional probability can be expressed by (2.40).

$$\Pr(\Psi = \psi_1 | \Phi = \phi_1) = \frac{\Pr(\Phi = \phi_1 \cap \Psi = \psi_1)}{\Pr(\Phi = \phi_1)} \quad (2.40)$$



**Figure 2.2:** Visualization of Bayes' theorem with a marginal distribution of only  $\Psi$  called a prior and a conditional distribution of  $\Psi$  given  $\Phi$  called the posterior.

This knowledge about a prior and a posterior is important during the understanding of the Kalman filtering techniques that are to be presented in this chapter. A good thought to keep track of is the idea of a model estimate as a prior while information about the true state through a field measurement allows increased knowledge of the true state.

### Bayesian statistics in n-dimensional spaces

Since models which operates on  $n$ -dimensional spaces often are referred to in this thesis, Bayesian statistics will be presented in a case more general than the 2-dimensional one. Consider an event  $\psi \in \mathbb{R}^n$ , it has an associated function  $F(\psi)$  and a probability density function  $f(\psi)$ , which is also called the joint probability density function for  $(\psi_1, \dots, \psi_n)$ . The associated function and the probability density function are related according to (2.41) and the probability

density function is still the derivative of the associated function similar to (2.34).

$$F(\psi_1, \dots, \psi_n) = \int_{-\infty}^{\psi_1} \dots \int_{-\infty}^{\psi_n} f(\psi'_1, \dots, \psi'_n) d\psi'_1 \dots d\psi'_n \quad (2.41)$$

(2.35) is also true in the  $n$ -dimensional case, which means that the probability of  $\psi$  lying somewhere in  $\mathfrak{R}^n$  is equal to one.

Consider the random variable  $\psi \in \mathfrak{R}^n$  to be a model state which has an associative function as well as a probability density function. In line with data assimilation, introduce a measurement vector  $\mathbf{d}$  holding measurement of the true (real world) state and let the likelihood of  $\mathbf{d}$  given  $\psi$  be expressed by  $f(\mathbf{d} | \psi)$ . The joint probability density function of model state and measurements then becomes

$$f(\psi, \mathbf{d}) = f(\psi)f(\mathbf{d} | \psi) = f(\mathbf{d})f(\psi | \mathbf{d}) \quad (2.42)$$

The relationships given by (2.42) can, again, be used to express Bayes' theorem

$$f(\psi | \mathbf{d}) = \frac{f(\psi)f(\mathbf{d} | \psi)}{f(\mathbf{d})} \quad (2.43)$$

which now expresses the model state probability density function with measurements, as a proportion to the probability density function of the model state alone times the likelihood for a certain set of measurements.

## Two important statistical moments and covariance

Probability density functions have so far been introduced as functions that defines the probability of a random variable to take a certain value (see (2.35)). The probability density function contains information, such as its own expected value  $\mu$ , standard deviation  $\sigma$  and variance  $\sigma^2$ .

The expected value for a function  $h(\Psi)$  where  $\Psi$  is a random variable is given by (2.44).

$$E[h(\Psi)] = \int_{-\infty}^{\infty} h(\psi)f(\psi)d\psi \quad (2.44)$$

and the expected value of a random variable  $\Psi$  is then given by (2.45).

$$\mu_{\Psi} = E[\Psi] = \int_{-\infty}^{\infty} \psi f(\psi)d\psi \quad (2.45)$$

The expected value given by equations (2.44)-(2.45) expresses the expected average value of  $\Psi$  if one performs an infinite number of realizations from that distribution.

The variance of a random variable  $\Psi$  is expressed by (2.46). This value states the spread of a probability distribution around the expected value.

$$\sigma^2 = E[(\Psi - E[\Psi])^2] = \int_{-\infty}^{\infty} (\psi - E[\Psi])^2 f(\psi)d\psi = E[\Psi^2] - E[\Psi]^2 \quad (2.46)$$

(2.46) states that the variance is the expected square deviation of  $\Psi$  from the expected value  $E[\Psi]$ , also known as the mean squared deviation.

Before moving on to estimation of statistics we present a measure, the covariance, which states how much two random variables  $\Psi$  and  $\Phi$  follow one another. If a small  $\Psi$  corresponds to a small  $\Phi$  this would give a positive covariance, a small  $\Psi$  corresponding to a big  $\Phi$  would give a negative covariance. The covariance for two random variables, for which the joint probability density function is given by (2.37), is defined as (2.47).

$$\begin{aligned}
\text{Cov}(\Psi, \Phi) &= E\left[(\Psi - E[\Psi])(\Phi - E[\Phi])\right] \\
&= E\left[(\Psi\Phi - E[\Psi]\Phi - E[\Phi]\Psi + E[\Psi]E[\Phi])\right] \\
&= E[\Psi\Phi] - E[\Phi E[\Psi]] - E[\Psi E[\Phi]] + E[\Psi]E[\Phi] \\
&= E[\Psi\Phi] - E[\Psi]E[\Phi] - E[\Psi]E[\Phi] + E[\Psi]E[\Phi] \\
&= E[\Psi\Phi] - E[\Psi]E[\Phi] \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi\phi f(\psi, \phi) d\psi d\phi - E[\Psi]E[\Phi]
\end{aligned} \tag{2.47}$$

Note in the last line of (2.47) that the probability density function  $f(\psi, \phi) = f(\psi)f(\phi)$  if  $\Psi$  and  $\Phi$  are independent. In that case (2.47) is reduced to 0 according to (2.48).

$$\begin{aligned}
\text{Cov}(\Psi, \Phi) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi\phi f(\psi, \phi) d\psi d\phi - E[\Psi]E[\Phi] \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi\phi f(\psi)f(\phi) d\psi d\phi - E[\Psi]E[\Phi] \\
&= \int_{-\infty}^{\infty} \phi f(\phi) \left( \int_{-\infty}^{\infty} \psi f(\psi) d\psi \right) d\phi - E[\Psi]E[\Phi] \\
&= E[\Psi] \int_{-\infty}^{\infty} \phi f(\phi) d\phi - E[\Psi]E[\Phi] \\
&= E[\Psi]E[\Phi] - E[\Psi]E[\Phi] = 0
\end{aligned} \tag{2.48}$$

### Estimations from samples

As pointed out by [12], there is no practical way to make integrations as the ones presented earlier in this chapter using a computer (performing numerical integrations) if the dimensionality of the probability functions is high (i.e.  $> 3$ ).

There is an alternative to numerical integrations called Markov Chain Monte Carlo methods in which a large number of realizations (draws),  $N$ , of the distribution  $f(\psi)$  are known. Statistical moments can be estimated from such samples. In this section it will be shown how the expected value,  $\mu$ , the variance,  $\sigma^2$ , and the covariance can be estimated from samples.

Assume a set  $\{\psi_i\}$  with  $i = 1, \dots, N$  where  $\psi_i$  is one observed value from the distribution  $f(\psi)$ , the expected value of a random variable from that distribution can then be approximated by a best guess, the sample mean, according to:

$$\mu = E[\Psi] \simeq \bar{\psi} = \frac{1}{N} \sum_{i=1}^N \psi_i \quad (2.49)$$

where the notation  $\simeq$  represents that the expected value of  $\Psi$  will tend to  $\bar{\psi}$  as  $N \rightarrow \infty$ . Note that  $\bar{\psi}$  is an unbiased estimation of  $E[\Psi]$ , which is proven by (2.50).

$$E[\bar{\psi}] = E\left[n^{-1} \sum_{i=1}^N \psi_i\right] = \sum_{i=1}^N E[\psi_i]/n = nE[\psi_i]/n = E[\Psi] \quad (2.50)$$

The variance given by (2.46) can be approximated by the sample variance according to (2.51).

$$\begin{aligned} \sigma^2 &= E\left[(\Psi - E[\Psi])^2\right] \\ &\simeq \overline{(\psi - \bar{\psi})^2} = \frac{1}{N-1} \sum_{i=1}^N (\psi_i - \bar{\psi})^2 \end{aligned} \quad (2.51)$$

(2.51) is said to be an unbiased estimator of the variance. That the variance estimator is unbiased means that it does not deviate from the expected variance. This can be proven by following the reasoning of (2.52).

$$\begin{aligned} \psi_i - \bar{\psi} &= (\psi_i - \mu) + (\mu - \bar{\psi}) \Leftrightarrow \\ (\psi_i - \bar{\psi})^2 &= (\psi_i - \mu)^2 + 2(\psi_i - \mu)(\mu - \bar{\psi}) + (\mu - \bar{\psi})^2 \Rightarrow \\ \sum_{i=1}^N (\psi_i - \bar{\psi})^2 &= \sum_{i=1}^N (\psi_i - \mu)^2 + 2 \sum_{i=1}^N (\psi_i - \mu)(\mu - \bar{\psi}) + \sum_{i=1}^N (\mu - \bar{\psi})^2 \Leftrightarrow \\ \sum_{i=1}^N (\psi_i - \bar{\psi})^2 &= \sum_{i=1}^N (\psi_i - \mu)^2 + 2(\mu - \bar{\psi}) \left(\sum_{i=1}^N \psi_i - N\mu\right) + N(\mu - \bar{\psi})^2 \Leftrightarrow \\ \sum_{i=1}^N (\psi_i - \bar{\psi})^2 &= \sum_{i=1}^N (\psi_i - \mu)^2 + 2(\mu - \bar{\psi})(N\bar{\psi} - N\mu) + N(\mu - \bar{\psi})^2 \Leftrightarrow \\ \sum_{i=1}^N (\psi_i - \bar{\psi})^2 &= \sum_{i=1}^N (\psi_i - \mu)^2 - N(\bar{\psi} - \mu)^2 \Leftrightarrow \\ \sum_{i=1}^N (\psi_i - \bar{\psi})^2 &= N\sigma^2 - N(\sigma)^2/N \end{aligned} \quad (2.52)$$

From which (2.51) can be deduced easily. Finally there is the covariance of samples defined as (2.53).

$$\begin{aligned} \text{Cov}(\Psi, \Phi) &= E\left[(\Psi - E[\Psi])(\Phi - E[\Phi])\right] \\ &\simeq \overline{(\psi - \bar{\psi})(\phi - \bar{\phi})} = \frac{1}{N-1} \sum_{i=1}^N (\psi_i - \bar{\psi})(\phi_i - \bar{\phi}) \end{aligned} \quad (2.53)$$

This ends the introduction to statistics, presented thus far are the terms probability density functions, accumulative probability functions and their properties, Bayes' theorem, the keywords prior and posterior, likelihood, expected value, variance and covariance. It was also then proven that the expected value, the variance and the covariance can be estimated from samples.

This information will be used in the upcoming sections when state estimation by models and data assimilation is discussed.

## 2.2.2 Analysis schemes for combined model prediction and data assimilation

The previous section introduced the notion of an available measurement  $d$  of the state. This section will bring together the measurement with the model prediction in an effort to increase the performance of the state estimate, or rather minimize the difference between the true state and the state estimate produced by the combined model prediction and measurement.

The outcome of this section should be awareness of what is commonly referred to as the analysis scheme [12]. An analysis scheme is a way of combining model predictions with field measurements where both the model state and the measurement usually are dependent of time and space although this section doesn't include any time dependency.

### The analysis scheme for a scalar state-space

Assume that the state at one specific location in space at one specific time is  $\psi^t$  where the superscript  $t$  denotes true state. Now assume that two estimations of  $\psi^t$  are available. The first estimation is expressed by (2.54).

$$\psi^f = \psi^t + p^f \quad (2.54)$$

where the superscript  $f$  denotes forecast, or the prior estimate, or the first guess, of the true state and  $p^f$  is the forecast error,  $\psi^f$  is then the state forecast. The estimate of the true state expressed by (2.54) is a model estimate unlike the estimate expressed by (2.55) which is an measurement  $d$  which contains a measurement error  $\epsilon$ .

$$d = \psi^t + \epsilon \quad (2.55)$$

The task at hand is to find a new estimate  $\psi^a$ , the analysed estimate, which is a better estimate of  $\psi^t$  than both  $\psi^f$  and  $d$ . The issue is that, if new conditions are imposed upon (2.54) in the form of the measurements  $d$ , the system will be overdetermined [13]. The task at hand can then be reformulated to find a new estimate  $\psi^a$  that minimizes the error terms.

Assume that the error terms  $p^f$  and  $\epsilon$  are random variables with zero mean such that:

$$\begin{aligned} \overline{p^f} &= 0 \\ \overline{\epsilon} &= 0 \\ \overline{(p^f)^2} &= C_{\psi\psi}^f \\ \overline{\epsilon^2} &= C_{\epsilon\epsilon} \end{aligned} \quad (2.56)$$

where  $C$  denotes the covariance of the subscripted random variable pair. Assume a linear estimator such that:

$$\psi^a = \psi^t + p^a = \alpha_1 \psi^f + \alpha_2 d \quad (2.57)$$

I.e.  $\psi^a$  is a linear combination of the forecast and the measurement. Given (2.54) and (2.55) stating the model state forecast  $\psi^f$  and the measurement  $d$  (2.57) can be expressed as:

$$\begin{aligned} \psi^t + p^a &= \alpha_1 \psi^f + \alpha_2 d \\ &= \alpha_1(\psi^t + p^f) + \alpha_2(\psi^t + \epsilon) \\ &= (\alpha_1 + \alpha_2)\psi^t + \alpha_1 p^f + \alpha_2 \epsilon \end{aligned} \quad (2.58)$$

for which the expected value is used to find an expression for  $\alpha_1$  and  $\alpha_2$ .

$$\begin{aligned} E[\psi^t + p^a] &= E[(\alpha_1 + \alpha_2)\psi^t + \alpha_1 p^f + \alpha_2 \epsilon] \Leftrightarrow \\ E[\psi^t] + E[p^a] &= (\alpha_1 + \alpha_2)E[\psi^t] + \alpha_1 E[p^f] + \alpha_2 E[\epsilon] \Leftrightarrow \\ E[\psi^t] + 0 &= (\alpha_1 + \alpha_2)E[\psi^t] + \alpha_1 0 + \alpha_2 0 \Leftrightarrow \\ E[\psi^t] &= (\alpha_1 + \alpha_2)E[\psi^t] \Leftrightarrow \\ 1 &= (\alpha_1 + \alpha_2) \end{aligned} \quad (2.59)$$

With this result in mind (2.57) will be reduced according to (2.60):

$$\begin{aligned} \psi^a &= \alpha_1 \psi^f + \alpha_2 d \\ &= (1 - \alpha_2)\psi^f + \alpha_2 d \\ &= \psi^f + \alpha_2(d - \psi^f) \end{aligned} \quad (2.60)$$

It is also possible to deduce the error  $p^a$  of the analyzed state  $\psi^a$  according to:

$$\begin{aligned} \psi^a &= \psi^f + \alpha_2(d - \psi^f) \Leftrightarrow \\ \psi^t + p^a &= \psi^t + p^f + \alpha_2(\psi^t + \epsilon - \psi^t - p^f) \Leftrightarrow \\ p^a &= (1 - \alpha_2)p^f + \alpha_2 \epsilon \end{aligned} \quad (2.61)$$

In which case the error variance of  $\psi^a$ , namely  $\overline{(p^a)^2}$ , is expressed by (2.62). Note the assumption that the model and measurement errors are uncorrelated.

$$\begin{aligned} \overline{(p^a)^2} &= C_{\psi\psi}^a = \overline{((1 - \alpha_2)p^f + \alpha_2 \epsilon)^2} \\ &= (1 - \alpha_2)^2 \overline{(p^f)^2} + 2\alpha_2(1 - \alpha_2)\overline{p^f \epsilon} + \alpha_2^2 \overline{\epsilon^2} \\ &= \left/ \overline{p^f \epsilon} = 0 \right/ = \\ &= (1 - \alpha_2)^2 C_{\psi\psi}^f + \alpha_2^2 C_{\epsilon\epsilon} \end{aligned} \quad (2.62)$$

Since  $C_{\psi\psi}^f$  and  $C_{\epsilon\epsilon}$  are both positive and constant parameters the minimum variance can be found by minimizing  $C_{\psi\psi}^a$  with respect to  $\alpha_2$ :

$$\begin{aligned} \frac{dC_{\psi\psi}^a}{d\alpha_2} &= 0 \Leftrightarrow \\ 0 &= -2(1 - \alpha_2)C_{\psi\psi}^f + 2\alpha_2 C_{\epsilon\epsilon} \Leftrightarrow \\ \alpha_2 &= \frac{C_{\psi\psi}^f}{C_{\psi\psi}^f + C_{\epsilon\epsilon}} \end{aligned} \quad (2.63)$$

The analyzed error variance is minimized for this choice of  $\alpha_2$  and (2.60) becomes

$$\psi^a = \psi^f + \frac{C_{\psi\psi}^f}{C_{\psi\psi}^f + C_{\epsilon\epsilon}}(d - \psi^f) \quad (2.64)$$

And the error variance expressed by (2.62) becomes

$$\begin{aligned} C_{\psi\psi}^a &= C_{\psi\psi}^f \left(1 - \frac{C_{\psi\psi}^f}{C_{\psi\psi}^f + C_{\epsilon\epsilon}}\right)^2 + C_{\epsilon\epsilon} \left(\frac{C_{\psi\psi}^f}{C_{\psi\psi}^f + C_{\epsilon\epsilon}}\right)^2 \Leftrightarrow \\ C_{\psi\psi}^a &= C_{\psi\psi}^f \left(1 - \frac{C_{\psi\psi}^f}{C_{\psi\psi}^f + C_{\epsilon\epsilon}}\right) \end{aligned} \quad (2.65)$$

(2.64) defines an unbiased and optimal estimation of a scalar state variable with a measurement and is an improved estimate of the state that includes both of the estimations given by (2.54) and (2.55).

### The Bayesian analysis scheme for a scalar state-space

Returning to Bayesian statistics and the notion of prior and posterior estimates, assume an initial guess  $\psi^f$  for which there is a probability density function  $f(\psi)$ . The likelihood of getting a measurement  $d$  given  $\psi$  is  $f(d|\psi)$  and it was shown previously that Bayes' theorem gives  $f(\psi|d) \propto f(\psi)f(d|\psi)$ . The notation  $\propto$  means that the probability density function of  $\psi$  given  $d$  is proportional to the right hand side of that expression.

Assume that two estimates of the state similar to (2.54) and (2.55) are available and that all distributions are Gaussian (see (2.36)). The density of the prior estimate of the state is given by

$$f(\psi) \propto \exp\left(-\frac{1}{2}(\psi - \psi^f)(C_{\psi\psi}^f)^{-1}(\psi - \psi^f)\right) \quad (2.66)$$

while the likelihood of  $d$  given  $\psi$ , or simply the likelihood, is given by

$$f(d|\psi) \propto \exp\left(-\frac{1}{2}(\psi - d)C_{\epsilon\epsilon}^{-1}(\psi - d)\right) \quad (2.67)$$

Considering (2.66)-(2.67) the posterior density can be expressed in the same way as (2.43), which gives:

$$f(\psi|d) \propto \exp\left(-\frac{1}{2}\mathcal{J}[\psi]\right) \quad (2.68)$$

Where the operator  $\mathcal{J}$  follows the notation of Evensen and shortens the expression in (2.68) while being defined as equation (2.69) [12].

$$\mathcal{J}[\psi] = (\psi - \psi^f)(C_{\psi\psi}^f)^{-1}(\psi - \psi^f) + (\psi - d)C_{\epsilon\epsilon}^{-1}(\psi - d) \quad (2.69)$$

Since  $\exp(-\beta) = 1/(e^\beta)$ , a minimization of  $\mathcal{J}$  with respect to  $\psi$  will give a

maximization of the posterior density in (2.68).

$$\begin{aligned}
\mathcal{J}[\psi] &= (\psi - \psi^f)(C_{\psi\psi}^f)^{-1}(\psi - \psi^f) + (\psi - d)C_{\epsilon\epsilon}^{-1}(\psi - d) \Rightarrow \\
\frac{d\mathcal{J}}{d\psi} &= 2(\psi - \psi^f)(C_{\psi\psi}^f)^{-1} + 2(\psi - d)C_{\epsilon\epsilon}^{-1} = 0 \Leftrightarrow \\
\psi &= \frac{\psi^f C_{\epsilon\epsilon} + d C_{\psi\psi}^f}{C_{\psi\psi}^f + C_{\epsilon\epsilon}} \Leftrightarrow \\
\psi &= d \frac{C_{\psi\psi}^f}{C_{\psi\psi}^f + C_{\epsilon\epsilon}} + \frac{\psi^f C_{\epsilon\epsilon} + \psi^f C_{\psi\psi}^f - \psi^f C_{\psi\psi}^f}{C_{\psi\psi}^f + C_{\epsilon\epsilon}} \Leftrightarrow \\
\psi &= \psi^f + \frac{C_{\psi\psi}^f}{C_{\psi\psi}^f + C_{\epsilon\epsilon}}(d - \psi^f)
\end{aligned} \tag{2.70}$$

Where the last row is noted to be equivalent to the result of (2.64), so the likelihood maximization given by a variance minimization in the case with Gaussian distributions for all error terms. It is always true that a minimum-variance estimate yields a maximum-likelihood estimate when the distributions are Gaussian. Note that this equivalence is not true for non-Gaussian distributions but that (2.69) still can be used to find the estimator that minimizes the variance [13].

### The analysis scheme for spatial state-spaces

The previous section outlined the problem of combining modelled states with measurements of the same in a scalar case (a single location), this section will expand this problem to a spatial domain (the whole length of a road for example). The task at hand is still to find a way of combining state estimates given by a prediction model with those given by field measurements but now in a spatial domain, say  $\mathbf{x} = (x, y, z)$  or a discrete spatial domain such as a cell representation of a road. The outcome is an analysis scheme. It is used later when the Kalman filter is presented.

Consider a variable  $\psi^f(\mathbf{x})$  expressing the first guess of the state at a location  $\mathbf{x}$ . Let  $\mathbf{d} \in \mathfrak{R}^M$  be a measurement vector with  $M$  measurements. Say that these measurements can be mapped to the true state with  $\mathcal{M} \in \mathfrak{R}^M$  which is called a measurement functional. The basic spatial formulation of the first guess of the state expressed by (2.54) can then be expressed as

$$\psi^f(\mathbf{x}) = \psi^t(\mathbf{x}) + p^f(\mathbf{x}) \tag{2.71}$$

while the measurement vector, that can be compared to (2.55), is

$$\mathbf{d} = \mathcal{M}[\psi^t(\mathbf{x})] + \epsilon \tag{2.72}$$

Both of (2.71) and (2.72) contains an error term with unknown values (otherwise would the true state be known). Note that  $\epsilon \in \mathfrak{R}^M$ . Regarding the measurement functional can such an operator be

$$\mathcal{M}_i[\psi(\mathbf{x})] = \int_{\mathcal{D}} \psi(\mathbf{x}) \delta(\mathbf{x} - \mathbf{x}_i) d\mathbf{x} \tag{2.73}$$

where  $\delta$  is the delta function that is equal to zero for every  $\mathbf{x} \in \mathcal{D}$  except for  $\mathbf{x} = \mathbf{x}_i$ . The subscript  $i$  is a measurement location index.

There is no way to do any further analysis about the state estimate in the spatial domain without making some assumptions about the error statistics, since no information about the errors  $p^f(\mathbf{x})$  or  $\epsilon$  is known [12]. Let the information in (2.56) be true in the spatial case. This means that both of the errors have a mean value equal to zero, and there is information present about the variance of the measurement errors and there is knowledge about the covariance between two spatial points  $\mathbf{x}_i$  and  $\mathbf{x}_j$  in the forecast step.

$$\begin{aligned}\overline{p^f(\mathbf{x})} &= 0 \\ \overline{\epsilon} &= \mathbf{0} \\ \overline{p^f(\mathbf{x}_i)p^f(\mathbf{x}_j)} &= C_{\psi\psi}^f(\mathbf{x}_i, \mathbf{x}_j) \\ \overline{\epsilon\epsilon^T} &= C_{\epsilon\epsilon}\end{aligned}\tag{2.74}$$

where  $C_{\epsilon\epsilon} \in \mathfrak{R}^{M \times M}$  is a covariance matrix which is not to be confused with (2.53) since it is not deduced from sample averages but from the true state.

Given the statistical null hypothesis (2.74) the next step of the discussion is to define a function  $\mathcal{J}$  that can be used to punish large deviations between the state  $\psi$  and the model forecast  $\psi^f$  and the measurements  $\mathbf{d}$  respectively. Evensen [12] stated (2.75) as an example of such a function. (2.75) contains two terms. The first of those two terms states a weighted measurement between the forecast  $\psi^f(\mathbf{x})$  and the state estimate  $\psi(\mathbf{x})$ . The second term measures (in a weighted fashion) the distance between the field measurement  $\mathbf{d}$  and the state estimate.

$$\begin{aligned}\mathcal{J}[\psi] &= \iint_{\mathcal{D}} (\psi^f(\mathbf{x}_1) - \psi(\mathbf{x}_1)) W_{\psi\psi}^f(\mathbf{x}_1, \mathbf{x}_2) (\psi^f(\mathbf{x}_2) - \psi(\mathbf{x}_2)) d\mathbf{x}_1 d\mathbf{x}_2 \\ &+ (\mathbf{d} - \mathcal{M}_{(3)}[\psi_3])^T \mathbf{W}_{\epsilon\epsilon} (\mathbf{d} - \mathcal{M}_{(4)}[\psi_4])\end{aligned}\tag{2.75}$$

$W$  and  $\mathbf{W}$  are the (functional) inverses of  $C_{\psi\psi}^f(\mathbf{x}_1, \mathbf{x}_2)$  and  $C_{\epsilon\epsilon}$  respectively, the subscripts of the expression  $\mathcal{M}_{(i)}[\psi_i]$  indicates dummy variables that are used later during the derivation of the analysis scheme [12]. The inverses of the covariances are the weights mentioned earlier.

The function  $\mathcal{J}[\psi]$  given by (2.75) will grow if the state estimate deviates from the forecast state and the field measurements. It is therefore desirable to find the optimal solution  $\psi(\mathbf{x}) = \psi^a(\mathbf{x})$  that minimizes that expression. If the weights are inverses of error covariances following Gaussian distributions the variance minima equals the likelihood maxima [12].

It can be proven that (2.76) is an optimal solution to (2.75) by deriving the Euler-Lagrange equation from  $\delta\mathcal{J} = \mathcal{J}[\psi + \delta\psi] - \mathcal{J}[\psi] = \mathcal{O}(\delta\psi^2)$  [12]. The idea is to take the variational derivative of the variational functional  $\mathcal{J}$  and assume that  $\mathcal{J} \rightarrow 0$  as  $\delta\psi(\mathbf{x}) \rightarrow 0$ .

$$\psi^a(\mathbf{x}) - \psi^f(\mathbf{x}) = \mathcal{M}_{(3)}^T[C_{\psi\psi}^f(\mathbf{x}, \mathbf{x}_3)] \mathbf{W}_{\epsilon\epsilon} (\mathbf{d} - \mathcal{M}_{(4)}[\psi_4^a])\tag{2.76}$$

(2.76) is not a satisfactory solution to (2.75) since  $\psi^a$  is included in both the right hand side and the left hand side of the expression. It is more desirable to search for an expression for  $\psi_a$  such as the analysis

$$\psi^a(\mathbf{x}) = \psi^f(\mathbf{x}) + \mathbf{b}^T \mathbf{r}(\mathbf{x})\tag{2.77}$$

where

$$\mathbf{b} = \mathbf{W}_{\epsilon\epsilon}(\mathbf{d} - \mathcal{M}_{(4)}[\psi_4^a]) \quad (2.78)$$

is a vector in  $\Re^M$ . (2.77) can be proven to be a variance minimizing (optimal) which is both unique and linear [12].  $\mathbf{r}(\mathbf{x})$  are called representers. If (2.77) is inserted in the optimal solution given by (2.76) can it be shown that

$$\mathbf{r}(\mathbf{x}) = \mathcal{M}_{(3)}[C_{\psi\psi}^f(\mathbf{x}, \mathbf{x}_3)] \quad (2.79)$$

A different expression for the vector  $\mathbf{b}$  can be deduced from (2.77)-(2.79) and is given by (2.80) [12].

$$\left( \mathcal{M}_{(3)}\mathcal{M}_{(4)}^T[C_{\psi\psi}^f(\mathbf{x}_3, \mathbf{x}_4)] + \mathbf{C}_{\epsilon\epsilon} \right) \mathbf{b} = \mathbf{d} - \mathcal{M}_{(4)}[\psi^f(\mathbf{x}_4)] \quad (2.80)$$

The missing piece is to derive an error estimate for (2.77) similar to the one done for the scalar analysis scheme that was expressed by (2.65). This is done with greater effort than for the scalar case and by taking the error covariance given by (2.74) which is

$$\mathbf{C}_{\psi\psi}^a(\mathbf{x}_1, \mathbf{x}_2) = \overline{(\psi^t(\mathbf{x}_1) - \psi^a(\mathbf{x}_1))(\psi^t(\mathbf{x}_2) - \psi^a(\mathbf{x}_2))} \quad (2.81)$$

and substituting this expression in (2.77) which leads to (2.82) if the procedure presented in Evensen [12] is followed.

$$\begin{aligned} \mathbf{C}_{\psi\psi}^a(\mathbf{x}_1, \mathbf{x}_2) &= \mathbf{C}_{\psi\psi}^f(\mathbf{x}_1, \mathbf{x}_2) \\ &\quad - \mathbf{r}^T(\mathbf{x}_1) \left( \mathcal{M}_{(3)}\mathcal{M}_{(4)}^T[C_{\psi\psi}^f(\mathbf{x}_3, \mathbf{x}_4)] + \mathbf{C}_{\epsilon\epsilon} \right)^{-1} \mathbf{r}(\mathbf{x}_2) \end{aligned} \quad (2.82)$$

### Formulation of the analysis scheme for discrete domains

The previous section outlined the analysis scheme in a spatial domain e.g.  $\mathbf{x} = (x, y, z)$  which is continuous. It is usually desirable to make a vector or matrix representation of the spatial domain that is modelled, so as to represent the domain with a discrete numerical grid  $\psi$ . In that case (2.71) and (2.72) would be

$$\begin{aligned} \psi^f &= \psi^t + \mathbf{p}^f \\ \mathbf{d} &= \mathbf{M}\psi^t + \boldsymbol{\epsilon} \end{aligned} \quad (2.83)$$

where  $\psi$  is a vector (numerical grid) representation of the spatial domain,  $\mathbf{p}^f$  and  $\boldsymbol{\epsilon}$  are error terms in vector form,  $\mathbf{d}$  is a measurement vector and  $\mathbf{M}$  is a discrete version of the measurement functional  $\mathcal{M}$ , meaning that  $\mathbf{M}$  is a matrix that maps between the measurement vector and the state vector.

Similar to the procedure made in the previous sections it can be proven that (2.84) is an optimal solution to the minimized variational functional problem [12].

$$\psi^a = \psi^f + \mathbf{r}^T \mathbf{b} \quad (2.84)$$

where

$$\mathbf{r} = \mathbf{M}\mathbf{C}_{\psi\psi}^f \quad (2.85)$$

is a matrix that holds the elements of the error covariance matrix at the spots of the measurements.

The vector  $\mathbf{b}$  in (2.84) can be derived from the discrete formulation of (2.80) which is given by (2.86).

$$\left(\mathbf{M}\mathbf{C}_{\psi\psi}^f\mathbf{M}^T + \mathbf{C}_{\epsilon\epsilon}\right)\mathbf{b} = \mathbf{d} - \mathbf{M}\psi^f \quad (2.86)$$

Finally there is the discrete formulation of the time evolution of the error statistics in the discrete case. The discrete formulation of (2.82) is

$$\mathbf{C}_{\psi\psi}^a = \mathbf{C}_{\psi\psi}^f - \mathbf{r}^T \left(\mathbf{M}\mathbf{C}_{\psi\psi}^f\mathbf{M}^T + \mathbf{C}_{\epsilon\epsilon}\right)^{-1} \mathbf{r} \quad (2.87)$$

Looking back at this section it can be seen that the analysis scheme gives an estimate  $\psi^a$  of the true state  $\psi^t$  and that this estimate is composed by an inverse of the estimate  $\psi^f$  as well as a linear combination  $\mathbf{r}^T\mathbf{b}$  which is just a collection of influence functions, the amount of which is equal to the number of measurements [12]. Given the expression of  $\mathbf{b}$  is it clear that  $\mathbf{b}$  will be small if the forecast state  $\psi^f$  is close to the measurements  $\mathbf{d}$  [12].

This concludes the introduction to analysis schemes, the introduction is used in the coming section where the Kalman filter is presented and, through the Kalman filter, is it also used when the extended Kalman filter and the ensemble Kalman filter are presented.

It is important to note that the analysis scheme so far has been presented as being independent of time, the concept of a time dynamical scheme is introduced with the Kalman filter. This presentation of the analysis scheme in scalar and spatial cases only regarded what was early on presented as a prior, or a forecast,  $\psi^f$  of the state and that a posterior, or the optimal conditional estimate,  $\psi^a$  was computed once extra knowledge in the form of measurements  $\mathbf{d}$  was brought into the picture.

### 2.2.3 Kalman filter

The previous section presented analysis schemes for time independent scalar and spatial cases. The discussion was oriented around the notions of a prior, or a forecast, and measurements of the state and how these two components could be combined to yield the optimal (the variance minimizing, or likelihood maximizing) state estimate  $\psi^a$ .

These analysis schemes can be used to perform sequential data assimilation and model state updates. As previously shown an improved state estimate in vector form  $\psi^a$  can be computed from the forecast  $\psi^f$ , the forecast error covariance matrix  $\mathbf{C}_{\psi\psi}^f$ , the measurement vector  $\mathbf{d}$  and the measurement error covariance matrix  $\mathbf{C}_{\epsilon\epsilon}$ . These components can also be used to compute  $\mathbf{C}_{\psi\psi}^a$ , the analyzed error covariance matrix.

This can be done in the time dynamic and discrete case, this section will outline this process for linear time dynamical cases. It is referred to as the Kalman filter.

#### The scalar case

Let  $G$  be a linear model operator that evolves the state  $\psi_k^t = \psi^t(t_k)$  in time, where  $t_k$  is the  $k$ th time value. This time evolution process is expressed by the

time dynamical system

$$\psi_k^t = G\psi_{k-1}^t + q_{k-1} \quad (2.88)$$

with

$$\psi_0^t = \Psi_0 + a \quad (2.89)$$

where  $q_k$  is a time dependent error term,  $\Psi_0$  is the initial state and  $\psi_0^t$  is that same initial state but including an error term  $a$ .

The true state estimation model cannot be expressed directly as equations (2.88) and (2.89), since the error  $q_k$ , that could represent error sources such as model limitations, is unknown. Instead, the actual model is expressed as (2.90) and (2.91).

$$\psi_k^f = G\psi_{k-1}^a \quad (2.90)$$

with

$$\psi_0^a = \Psi_0 \quad (2.91)$$

Equation (2.90) states the following: A forecast estimate of the state at time  $t_k$  is given by integrating the best guess  $\psi_{k-1}^a$  forward in time according to the model operator  $G$ .

Since (2.88) is the true state and (2.90) is the best guess that can be made, the forecast will have a state error  $(\psi_k^t - \psi_k^f)$ , which can be used to estimate the unknown error statistics:

$$\psi_k^t - \psi_k^f = G(\psi_{k-1}^t - \psi_{k-1}^a) + q_{k-1} \quad (2.92)$$

In which case

$$\begin{aligned} C_{\psi\psi}^f(t_k) &= \overline{(\psi_k^t - \psi_k^f)^2} \\ &= \overline{(G(\psi_{k-1}^t - \psi_{k-1}^a) + q_{k-1})^2} \\ &= \overline{G^2(\psi_{k-1}^t - \psi_{k-1}^a)^2 + 2G(\psi_{k-1}^t - \psi_{k-1}^a)q_{k-1} + q_{k-1}^2} \end{aligned} \quad (2.93)$$

Assuming that the model error  $q_k$  and the state error  $(\psi_k^t - \psi_k^f)$  are uncorrelated is  $\overline{(\psi_{k-1}^t - \psi_{k-1}^a)q_{k-1}} = 0$ , and (2.93) is reduced to

$$\begin{aligned} C_{\psi\psi}^f(t_k) &= \overline{G^2(\psi_{k-1}^t - \psi_{k-1}^a)^2} + 0 + \overline{q_{k-1}^2} \\ &= G^2 C_{\psi\psi}^a(t_{k-1}) + C_{qq}(t_{k-1}) \end{aligned} \quad (2.94)$$

The model evolves in time as given by the dynamical equation (2.90) with starting conditions given by (2.91). The model state error covariance  $C_{\psi\psi}^a(t_{k-1})$ , the model error covariance  $C_{qq}(t_{k-1})$  and the initial error covariance  $C_{\psi\psi}(t_0)$  evolves according to

$$C_{\psi\psi}^a(t_{k-1}) = \overline{(\psi_{k-1}^t - \psi_{k-1}^a)^2} \quad (2.95)$$

$$C_{qq}(t_{k-1}) = \overline{q_{k-1}^2} \quad (2.96)$$

$$C_{\psi\psi}(t_0) = C_{aa} = \overline{a^2} \quad (2.97)$$

Which then concludes the model evolution and the error covariance evolution as a consistent time dynamical system.

When a measurement  $d$  is available the analysis scheme given by (2.64) will yield an analysed state estimate.

### The vector case

The state in a spatial domain can be mapped to a discrete grid represented by the elements in a state vector  $\boldsymbol{\psi}_k^t$ . With a linear model operator in matrix form the true state vector representation will evolve according to

$$\boldsymbol{\psi}_k^t = \mathbf{G}\boldsymbol{\psi}_{k-1}^t + \mathbf{q}_{k-1} \quad (2.98)$$

where the variables follows the description of those in (2.88). The estimation of (2.98) is

$$\boldsymbol{\psi}_k^f = \mathbf{G}\boldsymbol{\psi}_{k-1}^a \quad (2.99)$$

The error covariance matrix  $\mathbf{C}_{\psi\psi}^f(t_k)$  can be found by a procedure similar to the one used to find  $\mathbf{C}_{\psi\psi}^f(t_k)$  in (2.94).

$$\mathbf{C}_{\psi\psi}^f(t_k) = \mathbf{G}\mathbf{C}_{\psi\psi}^a(t_{k-1})\mathbf{G}^T + \mathbf{C}_{qq}(t_{k-1}) \quad (2.100)$$

The time evolution equations (2.99) and (2.100) together with the analysis equations (2.101)-(2.103) is the Kalman filter [12, 14].

$$\boldsymbol{\psi}_k^a = \boldsymbol{\psi}_k^f + \mathbf{K}_k(\mathbf{d}_k - \mathbf{M}_k\boldsymbol{\psi}_k^f) \quad (2.101)$$

where  $\mathbf{d}$  is the measurement vector and  $\mathbf{M}$  is a matrix that maps the state vector to the measurements, it is also known as a measurement matrix. The covariance matrix  $\mathbf{C}_{\psi\psi}^a$  time evolution is given by

$$(\mathbf{C}_{\psi\psi})_k^a = (\mathbf{I} - \mathbf{K}_k\mathbf{M}_k)(\mathbf{C}_{\psi\psi})_k^f \quad (2.102)$$

Both of the equations (2.101) and (2.102) contains a matrix  $\mathbf{K}$ , this matrix is called the Kalman gain and is given by

$$\mathbf{K}_k = (\mathbf{C}_{\psi\psi})_k^f \mathbf{M}_k^T (\mathbf{M}_k (\mathbf{C}_{\psi\psi})_k^f \mathbf{M}_k^T + (\mathbf{C}_{\epsilon\epsilon})_k)^{-1} \quad (2.103)$$

(2.101)-(2.103) are the more common way to post the analysis (or update) equations of the Kalman filter than (2.84)-(2.87) and one of these equation sets can be deduced directly from the other [12, 14].

### 2.2.4 Extended Kalman filter

The previous section presented the Kalman filter, which produces the optimal state estimate for linear functions  $G$ . In the case of non-linear functions  $G$  however, the Kalman filter cannot be used directly. But if  $G$  is differentiable can it be approximated by a Taylor expansion.

This section will briefly outline the extended Kalman filter for sequential data assimilation in which an estimate of the error covariance time evolution is used and for linear measurement operators.

#### The scalar case

Let (2.104) represent a time dynamical non-linear model of the state of a single scalar value  $\psi_k^t$ .

$$\psi_k^t = G(\psi_{k-1}^t) + q_{k-1} \quad (2.104)$$

Just as in the previous section is it noted that the error  $q$  depends on e.g. model inaccuracies and that  $q$  is unknown. Assume a numerical model of the state that is approximated by:

$$\psi_k^f = G(\psi_{k-1}^a) \quad (2.105)$$

In which case the error between the models given as equations (2.104) and (2.105) is the difference

$$\psi_k^t - \psi_k^f = G(\psi_{k-1}^t) - G(\psi_{k-1}^a) + q_{k-1} \quad (2.106)$$

Assume that  $G$  is infinitely differentiable, the next step includes a Taylor expansion of  $G(\psi_{k-1}^t)$  in the close proximity of  $\psi_{k-1}^a$ :

$$\begin{aligned} G(\psi_{k-1}^t) &= G(\psi_{k-1}^a) + G'(\psi_{k-1}^a)(\psi_{k-1}^t - \psi_{k-1}^a) \\ &\quad + \frac{1}{2}G''(\psi_{k-1}^a)(\psi_{k-1}^t - \psi_{k-1}^a)^2 + \dots \end{aligned} \quad (2.107)$$

By substituting this expansion of  $G(\psi_{k-1}^t)$  in equation (2.106) the error distance becomes

$$\begin{aligned} \psi_k^t - \psi_k^f &= G(\psi_{k-1}^a) + G'(\psi_{k-1}^a)(\psi_{k-1}^t - \psi_{k-1}^a) \\ &\quad + \frac{1}{2}G''(\psi_{k-1}^a)(\psi_{k-1}^t - \psi_{k-1}^a)^2 + \dots - G(\psi_{k-1}^a) + q_{k-1} \Leftrightarrow \\ \psi_k^t - \psi_k^f &= G'(\psi_{k-1}^a)(\psi_{k-1}^t - \psi_{k-1}^a) \\ &\quad + \frac{1}{2}G''(\psi_{k-1}^a)(\psi_{k-1}^t - \psi_{k-1}^a)^2 + \dots + q_{k-1} \end{aligned} \quad (2.108)$$

According to equation (2.51) the forecast error covariance time evolution,  $C_{\psi\psi}^f(t_k)$ , can be expressed as the expected value of the square of equation (2.108):

$$\begin{aligned} C_{\psi\psi}^f(t_k) &= \overline{(\psi_k^t - \psi_k^f)^2} \\ &= (G'(\psi_{k-1}^a))^2 \overline{(\psi_{k-1}^t - \psi_{k-1}^a)^2} \\ &\quad + G'(\psi_{k-1}^a)G''(\psi_{k-1}^a) \overline{(\psi_{k-1}^t - \psi_{k-1}^a)^3} \\ &\quad + \left(\frac{1}{2}\right)^2 (G''(\psi_{k-1}^a))^2 \overline{(\psi_{k-1}^t - \psi_{k-1}^a)^4} \\ &\quad + \dots C_{qq}(t_{k-1}) \end{aligned} \quad (2.109)$$

Note that equation (2.109) contains an infinite number of terms, or moments, which can be interpreted as infinite knowledge about the error statistics. Assuming that there is no need to have more information about the error statistics than the first and second order moments will equation (2.109) be simplified to equation (2.110) [12].

$$C_{\psi\psi}^f(t_k) \approx C_{\psi\psi}^a(t_{k-1})(G'(\psi_{k-1}^a))^2 + C_{qq}(t_{k-1}) \quad (2.110)$$

Observe that equations (2.105) and (2.110) states the extended Kalman filter for a scalar state variable when they are combined with equations (2.64) and (2.65) [12]. (2.64) and (2.65) states the analyzed estimate and the error variance of the analyzed estimate. Note that this follows an assumption of linear measurement operators [12].

### The matrix case

Following the same form of calculations as for the scalar case can the extended Kalman filter be derived for the non-scalar case (when a spatial domain is represented by vector notation) [12]. In that case will equation (2.111) represent the true state vector  $\boldsymbol{\psi}_k^t$  at time  $t_k$

$$\boldsymbol{\psi}_k^t = \mathbf{G}(\boldsymbol{\psi}_{k-1}^t) + \mathbf{q}_{k-1} \quad (2.111)$$

while the forecast is stated as

$$\boldsymbol{\psi}_k^f = \mathbf{G}(\boldsymbol{\psi}_{k-1}^a) \quad (2.112)$$

where, in both (2.111) and (2.112),  $\mathbf{G}$  is a non-linear matrix operator.

In line with (2.109)-(2.110) the error statistics (the error covariance) will develop in time according to (2.113) [12, 14].

$$\begin{aligned} \mathbf{C}_{\boldsymbol{\psi}\boldsymbol{\psi}}^f(t_k) &= \mathbf{G}'_{k-1} \mathbf{C}_{\boldsymbol{\psi}\boldsymbol{\psi}}^a(t_{k-1}) \mathbf{G}'_{k-1}{}^T + \mathbf{C}_{qq}(t_{k-1}) + \dots \\ &\simeq \mathbf{G}'_{k-1} \mathbf{C}_{\boldsymbol{\psi}\boldsymbol{\psi}}^a(t_{k-1}) \mathbf{G}'_{k-1}{}^T + \mathbf{C}_{qq}(t_{k-1}) \end{aligned} \quad (2.113)$$

with  $\mathbf{C}_{qq}(t_{k-1})$  as the model error covariance matrix and  $\mathbf{G}'_{k-1}$  being the Jacobian given by

$$\mathbf{G}'_{k-1} = \left. \frac{\partial \mathbf{G}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}} \right|_{\boldsymbol{\psi}_{k-1}} \quad (2.114)$$

Again, note that (2.113) contains an infinite number of terms that are disregarded in the summation. This simplification still equals the assumption that higher order statistical moments are unnecessary since the contribution to the error statistics is assumed to be negligible and that the information can be left out [14].

The extended Kalman filter in matrix form is finally given by (2.112) and (2.113) together with (2.101)-(2.103).

### 2.2.5 Ensemble Kalman filter

The issue with the extended Kalman filter is that it is operated under the assumption that the error covariance  $\mathbf{C}_{\boldsymbol{\psi}\boldsymbol{\psi}}^f$  progress can be approximated by equation (2.109) in the scalar case or (2.113) when lie within the spatial domain and are represented by a numerical grid (a vector).

As stated in the previous section, both of these equations can be considered simplifications, since they really should be expressed as an indefinite continuation of the error statistics according to the same equations.

A continuation of the error statistics, in the vector case, means a summation of the higher statistical moments and derivatives of the non-linear model operator  $\mathbf{G}$ .

The ensemble Kalman filter, which is used in the Mobile Millennium system and will be presented here, was developed to overcome the issues associated with the extended Kalman filter [2, 10]. The ensemble Kalman filter is more suitable for handling large state spaces than the original Kalman filter. It is better suited for capturing the time evolution of error statistics, it relieves the user from the

burden of storing and forwarding the error covariance matrix of the model state and no error terms are neglected due to a simplification of the linearization of the model operator. The ensemble Kalman filter has also been noted as being easier on the computational side than the extended Kalman filter [10, 12, 14, 15].

The ensemble Kalman filter was derived from the viewpoint of an ocean model [10]. While such models are uninteresting for the subject of this thesis there are some noteworthy similarities. The oceanic model had, among others, the following features:

1. The model operator was non-linear.
2. The state was integrated forward in time based on an initial state.
3. The initial state was a guess or an estimate from measurements.
4. The initial state guess will have errors originating from e.g. interpolated data (since measurements are sparse in the physical domain) and errors in the actual measurements.

This section outlines the ensemble Kalman filter as it was presented in [10] but in accordance with the corrections made to the filter in [16] and with the notation used in [12]. The first step is an introduction to the time evolution of error statistics represented by an ensemble of model states.

### Evolution and predictability of error statistics

Assume that the state of the (unspecified) model subject at time  $t \in \mathbb{R}^+$  is represented by a vector of state variables  $\boldsymbol{\psi}(t) \in \mathbb{R}^n$  and that this vector holds all dependants.

Let this state vector  $\boldsymbol{\psi}(t)$  be a point in the  $n$ -dimensional space  $\mathcal{D}$ . This point will continuously move around in that space along some trajectory as the model progresses. As noted earlier, the true initial state is unknown and the model imperfect. Assume that this uncertainty is represented by an ensemble, or a cloud, of points of infinite size in the space  $\mathcal{D}$ . A probability density distribution function can then be defined as

$$\phi(\boldsymbol{\psi}) = \frac{dN}{N} \quad (2.115)$$

where  $\boldsymbol{\psi}$  is the initial state guess  $\boldsymbol{\psi}(t = 0)$ ,  $N$  is the number of ensemble members ( $n$ -dimensional points) and  $dN$  is a the number of points per each small increment in volume or simply the point density.

The probability density distribution function can vary depending on the location in  $\mathcal{D}$ , introduce

$$\phi(\boldsymbol{\psi})d\boldsymbol{\psi} \quad (2.116)$$

as the probability that the system is in a state at some location in the volume  $d\boldsymbol{\psi}$  with  $d\boldsymbol{\psi}$  being in the close proximity of  $\boldsymbol{\psi}$ . If  $\phi \geq 0$  for any point  $\boldsymbol{\psi}$  and time  $t$  is

$$\int_{\mathcal{D}} \phi(\boldsymbol{\psi})d\boldsymbol{\psi} = 1 \quad (2.117)$$

Meaning that the probability of finding the system in some state in the entire space  $\mathcal{D}$  is equal to one.

Taking a quantity  $h(\boldsymbol{\psi})$  and applying equation (2.44) with equation (2.116) gives the expected value of  $h(\boldsymbol{\psi})$  as

$$\begin{aligned} E[h(\boldsymbol{\psi})] &= \int h(\boldsymbol{\psi})\phi(\boldsymbol{\psi})d\boldsymbol{\psi} \\ &\simeq \overline{h(\boldsymbol{\psi})} \end{aligned} \quad (2.118)$$

in which case the expected value of the element  $i$  of the state-space vector  $\boldsymbol{\psi}$  is

$$\mu_i \simeq \overline{\psi_i} \quad (2.119)$$

with  $i \in \{1, \dots, n\}$ . According to equations (2.51) and (2.53) will then the elements of the variance matrix  $\mathbf{C}_{ii}$  and the covariance matrix  $\mathbf{C}_{ij}$  become

$$\begin{aligned} \mathbf{C}_{ii} &= E[(\psi_i - \mu_i)^2] \\ &\simeq \overline{(\psi_i - \mu_i)(\psi_i - \mu_i)} \end{aligned} \quad (2.120)$$

and

$$\begin{aligned} \mathbf{C}_{ij} &= E[(\psi_i - \mu_i)(\psi_j - \mu_j)] \\ &\simeq \overline{(\psi_i - \mu_i)(\psi_j - \mu_j)} \end{aligned} \quad (2.121)$$

where  $i \neq j$  in equation (2.121). The results for all elements  $i$  and  $j$  in these two equations are stored at the locations  $(i, j)$  in the covariance matrix  $\mathbf{C}_{ij}$  (with the variances along the diagonal for  $i = j$ ).

The expected value and the covariance matrix substitute the first and second order statistical moments. In the same fashion can the third and the fourth order moments (the skewness and the kurtis) be defined [10]:

$$\begin{aligned} \Theta_{ijk} &\simeq \overline{(\psi_i - \mu_i)(\psi_j - \mu_j)(\psi_k - \mu_k)} \\ \Gamma_{ijkl} &\simeq \overline{(\psi_i - \mu_i)(\dots)(\dots)(\psi_l - \mu_l)} \end{aligned} \quad (2.122)$$

Even higher order moments can be defined in this way.

Assume that the initial state guess follows a normal distribution, i.e. the probability density function  $\phi(\boldsymbol{\psi}, t = 0)$  is a Gaussian distribution with the mean  $\boldsymbol{\mu}(t = 0)$  coinciding with  $\boldsymbol{\psi}(t = 0)$  and with a covariance  $\mathbf{C}_{ij}(t = 0)$  that describes the uncertainty of the initial state guess [10].

Since the distribution is Gaussian, ensemble members far from the initial guess will be less probable than members close to the initial estimate  $\boldsymbol{\mu}(t = 0)$  [10].

Each of the possible state vectors, or ensemble members will evolve forward in time by whatever non-linear model  $\mathbf{G}$  they are subject to. This evolution is a motion in the space  $\mathcal{D}$  which is deterministic with an outcome depending only on the location of the initial ensemble. The evolution given by the model is also called the forecast [10].

As previously stated, every initial ensemble member is generated around the initial guess, which, at the time, was the mean value of the distribution. During the time evolution however, the member that represent the mean might drift away from what is the most likely state among the members of the forecast [10].

It is also possible that the ensemble forecast might grow or shrink in  $\mathcal{D}$ , according to e.g. equation (2.120) the variance will grow as well. A result of this is that the prediction error statistics changes with the change of variance of the forecast. If the ensemble changes with respect to spread in  $\mathcal{D}$  the associated probability density function that describes the ensemble will change as well. The conclusion is that the ensemble density probability can be predicted by the same equations that are given by the model (since the density probability is derived from the ensemble) [10].

Consider the members in a fraction of the state space,  $d\boldsymbol{\psi}$ . Those members are subject to the same time evolution model, which is deterministic. No ensemble members can be destroyed during this evolution, neither can they be created, thus will the probability be conserved as stated by (2.123) [10].

$$\frac{\partial\phi}{\partial t} + \sum_{i=1}^n \frac{\partial G_i \phi}{\partial \psi_i} = 0 \quad (2.123)$$

where  $G_i$  is the  $i$ :th component of the model operator  $\mathbf{G}$ . (2.123) can be compared to the conservation of vehicles with respect to traffic flow; the inflow and outflow of probability over  $d\boldsymbol{\psi}$  is balanced by the change in probability density over time. The chain rule  $(f \cdot g)' = f' \cdot g + f \cdot g'$  is applied in (2.123):

$$\begin{aligned} \frac{\partial\phi}{\partial t} + \sum_{i=1}^n \frac{\partial G_i \phi}{\partial \psi_i} &= 0 \Leftrightarrow \\ \frac{\partial\phi}{\partial t} + \sum_{i=1}^n \left( \phi \frac{\partial G_i}{\partial \psi_i} + G_i \frac{\partial\phi}{\partial \psi_i} \right) &= 0 \Leftrightarrow \\ \frac{\partial\phi}{\partial t} + \sum_{i=1}^n G_i \frac{\partial\phi}{\partial \psi_i} &= -\phi \sum_{i=1}^n \frac{\partial G_i}{\partial \psi_i} \end{aligned} \quad (2.124)$$

(2.124) is useful since the right hand side of the final equality states that the probability density will increase or decrease with the model integrations. An alternative viewpoint is that it describes a motion of the point cloud in a small portion of the state space  $\mathcal{D}$ . Therefore, that cloud is contracting or expanding [10].

Following a single ensemble member (point) will  $d\phi/dt$  describe if the ensemble is contracting (when  $d\phi/dt > 0$ ) or expanding (when  $d\phi/dt < 0$ ). These two situations would then equal cases where the error is becoming smaller (during a cloud contraction since the deviation around the mean is smaller) and larger respectively [10].

Assume a stochastic differential equation with a non-linear model operator for the coming steps:

$$d\boldsymbol{\psi} = \mathbf{G}(\boldsymbol{\psi})dt + d\mathbf{q} \quad (2.125)$$

where  $d\boldsymbol{\psi}$  is the state vector,  $\mathbf{G}$  is a non-linear model operator and  $d\mathbf{q} \in \mathbb{R}^n$  are the white noise model error terms in vector representation. This error follow a Gaussian distribution with zero mean and covariance  $\mathbf{C}_{qq}$ . (2.125) states that a small increment in time,  $dt$ , equals a state location change in  $\mathcal{D}$  denoted as  $d\boldsymbol{\psi}$ .

Considering the ensemble of points in the state space the probability density function of the model state  $\phi(\boldsymbol{\psi})$  will follow (2.126) as the points are forwarded

in time according to  $\mathbf{G}$  [10, 12]:

$$\frac{\partial \phi}{\partial t} + \sum_i \frac{\partial (g_i \phi)}{\partial \psi_i} = \sum_{i,j} \frac{\mathbf{C}_{qq}}{2} \frac{\partial^2 \phi}{\partial \psi_i \partial \psi_j} \quad (2.126)$$

(2.126) is comparable to (2.123), the difference is the model error that were introduced in (2.125) which have been assumed throughout this chapter. Evensen [12] notes that (2.126) is not the result of any simplifications of importance and that it is a fundamental expression for how the error statistics are evolved in time. If (2.87) could be solved, it would be possible to deduce the error statistics at an arbitrary time but it becomes overpowering for state spaces other than the scalar case [10, 12].

The question that remains is just how the error statistics represented by the probability density function  $\phi$  should be determined and forwarded in time during the analysis step. It was already concluded that sometimes it is not desirable to do the linearization of the non-linear model operator as it was done for the extended Kalman filter. It is not even possible in some cases (considering e.g. a certain update algorithm that is the CTM-v model). Neither is it possible to solve (2.126) in an efficient way, if it even is possible [10].

The next assumption made in this discussion is that all initial conditions for the process (2.125) are assumed to follow a Gaussian distribution, just as the error terms  $d\mathbf{q}$ . In accordance with (2.36) is such a distribution characterized by the mean  $\mu$  and the variance  $\sigma^2$ . If the model operator  $\mathbf{G}$  in (2.125) was linear, these two statistical moments would be completely sufficient to perfectly describe the probability density [12].

In the original paper Evensen [10] presented Monte Carlo methods as an alternative for prediction of the error statistics. Recall the idea of a large collection of points in the state space, a form of a cloud that is called an ensemble. At some time instant, each of these points represent a different state and each of the points are forwarded in time according to the same non-linear model operator. The mean value of these states and the covariance matrix may be insufficient to describe the error statistics completely, but they will describe the average path through time in the state space (by the mean, or the approximated expected value) as well as how the points spread around that average path (the dispersion is given by the covariance matrix). The good thing is, that if this ensemble contains enough points, the approximation of these two statistical moments will be equivalent to solving (2.126) [10, 12].

Consider a large collection  $\{\psi_1, \dots, \psi_N\}$  of points in the state space  $\mathcal{D}$ . The expected value of these points is given by an approximation from a sample and is denoted  $\bar{\psi}$ . According to (2.120) and (2.121) the error covariance matrices for the forecast estimate (the prior) and the analysed estimate (the posterior) then becomes

$$\begin{aligned} \mathbf{C}_{\psi\psi}^f &= \overline{(\psi^f - \bar{\psi}^f)(\psi^f - \bar{\psi}^f)^T} \\ \mathbf{C}_{\psi\psi}^a &= \overline{(\psi^a - \bar{\psi}^a)(\psi^a - \bar{\psi}^a)^T} \end{aligned} \quad (2.127)$$

The major assumption made here is that the ensemble mean, the mean value among all states represented by the points in the clouds is the best estimate of the true state  $\psi^t$  which is completely unknown. (2.127) then states the best guess of the covariance as the spread around this mean value given by the prior and posterior guesses of the true state.

### Analysis scheme for the ensemble Kalman filter

In the original report, Evensen [10] made a slight error by not treating the measurements  $\mathbf{d}$  as an ensemble of random variables in the derived analysis scheme that was presented [16]. If the measurements are not treated as an ensemble of random variables with a mean equal to the first observation value, one might risk an update with too low variance [16]. An ensemble of observations is therefore introduced:

$$\mathbf{d}_j = \mathbf{d} + \boldsymbol{\epsilon}_j \quad (2.128)$$

where  $j = 1, \dots, N$  denotes the measurement member index,  $\mathbf{d}$  is the original observation,  $\mathbf{d}_j$  is the  $j$ :th random observation drawn and  $\boldsymbol{\epsilon}_j$  is a random term with zero mean. The covariance matrix of the measurement ensemble is then approximated by

$$\mathbf{C}_{\epsilon\epsilon}^e = \overline{\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T} \quad (2.129)$$

where the superscript  $e$  denotes that  $\mathbf{C}_{\epsilon\epsilon}^e$  is an ensemble representation of the true measurement errors covariance matrix  $\mathbf{C}_{\epsilon\epsilon}$  that was applied in the Kalman filter. Note that  $\mathbf{C}_{\epsilon\epsilon} \simeq \mathbf{C}_{\epsilon\epsilon}^e$  as  $N \rightarrow \infty$ .  $N$  will be limited for practical implementations and the error statistics of the measurements will therefore be an approximation (another source of error). Evensen [12] notes that this approximation can be justified since the observation noise seldom is known anyway and that a sufficient amount of ensemble members can be enough to make a fair representation.

Just as for the Kalman filter and the extended Kalman filter is the analysis  $\boldsymbol{\psi}_j^a$  made at each time when measurements are available and for every single ensemble member  $j$  according to (2.101). The sample mean can thus be computed as

$$\bar{\boldsymbol{\psi}}^a = \bar{\boldsymbol{\psi}}^f + \mathbf{K}_e(\bar{\mathbf{d}} - \mathbf{M}\bar{\boldsymbol{\psi}}^f) \quad (2.130)$$

where  $\mathbf{K}_e$  is the Kalman gain given by (2.103) but with the approximation  $\mathbf{C}_{\psi\psi}^f = (\mathbf{C}_{\psi\psi}^e)^f$  and  $\bar{\mathbf{d}}$  is given by the average of all  $\mathbf{d}_j$  given by (2.128).

Given the definition of  $\mathbf{C}_{\psi\psi}^e$  in (2.127) it is also possible to derive an expression for the time evolution of the error statistics [12]:

$$(\mathbf{C}_{\psi\psi}^e)^a = (\mathbf{I} - \mathbf{K}_e\mathbf{M})(\mathbf{C}_{\psi\psi}^e)^f \quad (2.131)$$

which is the error covariance minimization also used in the standard Kalman filter.

This marks the end of the presentation of the ensemble Kalman filter. The discussion was started off with a presentation of statistical definitions and Bayesian statistics where the notion of a prior and a posterior estimate was given. The analysis scheme was then presented in general as a way of assimilating observed data with a prior guess, or a forecast of the state for time independent cases in both scalar and spatial domains. The next step was the presentation of the original Kalman filter which is the optimal estimate of the state for linear systems. The extended Kalman filter was presented as an alternative for nonlinear systems and it could be noted that both of these two schemes may be impossible to use in some applications. Implementation difficulties might be

due to a non-linear model operator (demanded by the Kalman filter), a non-differentiable model operator or underestimation of the error statistics as well as computational difficulties in the linearisation of the model operator (for the extended Kalman filter). Finally the ensemble Kalman filter was presented as an alternative and this is the analysis scheme used in the Mobile Millennium system.

## 2.3 Combining CTM-v and EnKF

Presented in the thesis so far is a network update algorithm  $\mathcal{M}$  given by (2.32) that adopts a velocity based cell transmission model (the CTM-v).

The adoption of the CTM-v was motivated as a way of simplifying the velocity estimation problem. This is especially efficient when considering probes<sup>6</sup>. The extension of the CTM-v from a single edge model to arbitrary networks makes  $\mathcal{M}$  non-linear and non-differentiable and, as a result, neither the Kalman filter or the extended Kalman filter cannot be used for velocity data assimilation from probe data with  $\mathcal{M}$  [2]. The (so far general) ensemble Kalman filter (EnKF) for data assimilation that was outlined in Section 2.2.5) is a suitable alternative for velocity data assimilation because the subject, the general model operator  $\mathbf{G}$  was neither differentiable nor linear. We are also relieved by the problems (unbounded error growth) that can arise from the simplifications in the linearisation in the extended Kalman filter.

The combination of the network update algorithm  $\mathcal{M}$  and the EnKF data assimilation technique is summarized by the following EnKF-algorithm [2]. Relating to Section 2.2.5 we leave the arbitrary model operator  $\mathbf{G}$  and state space  $\psi$  by assuming that  $\mathbf{G} = \mathcal{M}$  and  $\psi = \mathbf{v}$ .

- Step 1. Make an initial state guess  $\bar{\mathbf{v}}_a$ , this can be an observed state. Repeat a process where  $N$  ensemble members indexed by  $j$  are created by adding perturbations from a Gaussian distribution with zero mean and the covariance  $\mathbf{C}^a$  to the guess  $\bar{\mathbf{v}}_a$ .
- Step 2. This is the forecast step. Update each of the  $N$  ensemble members  $\mathbf{v}_j$  by the network update algorithm  $\mathcal{M}$  and add perturbations  $\boldsymbol{\eta}_j$  to each of those time evolutions, i.e:

$$\mathbf{v}_j^f = \mathcal{M}[\mathbf{v}_j^a] + \boldsymbol{\eta}$$

where  $\boldsymbol{\eta} \sim (0, \mathbf{Q})$  is the Gaussian zero-mean white noise with covariance  $\mathbf{Q}$  that captures inaccuracies in the model  $\mathcal{M}$ . Then update the error statistics  $(\mathbf{C}_{vv}^e)^f$  according to (2.127).

- Step 3. This is the analysis step. Once measurements  $\mathbf{d}$  are available; compute the Kalman gain  $\mathbf{K}_e$ . Then draw an ensemble of observations  $\mathbf{d}_j$  according to (2.128) and then update each ensemble member  $\mathbf{v}_j^g$  according to (2.101).
- Step 4. Return to step 2, which is repeated until measurements are available.

---

<sup>6</sup>A probe is a mobile sensor that traverses the highway and that is capable of transmitting velocity data. It can e.g. be a cell phone equipped with GPS.

Note that the EnKF-algorithm above is time dynamical. Step 1 gives the state and error statistics at the time origin. The forecast is then made in order to get the state at the net time step, i.e. the time step is incremented in step 2 and the analysis carried from step 3 to step 2 then lags one step behind the forecast.

### 2.3.1 Implementation of the CTM-v and EnKF combination

Reference [2] noted that the above mentioned EnKF-algorithm posed some practical issues for their real-time and large scale implementation of the same algorithm. First, they noted that correlations between distant parts of the highway network might arise from the covariance representation even though networks edges far apart are unlikely to have correlated states. Second, the forecast error covariance  $((C)_{vv}^e)^f$  for a large cell transmission model might reach the extent where the computer's working memory is heavily loaded.

Such issues were tackled by an adoption of a covariance localization method in [2]. The authors assumed that such correlations were not network wide but that only neighbouring links could have correlated states. The update equations then, were solved in a distributed fashion with smaller covariance matrices. The alternative EnKF-algorithm was adopted for each individual edge, the forecast error covariance  $((C)_{vv}^e)^f$  was calculated for each of the edges and measurements used in the analysis for each edge was only obtained from the current edge and for its neighbouring links, see [2] for details.

## Chapter 3

# Literature Review Part II: Calibration

Being the second part of our literature review, the aim for this chapter is sort out the what and the hows of model calibration in general. This chapter forms the foundation upon which the calibration framework later presented rests.

The first part of this chapter, section 3.1, sorts out the what of model calibration. The terms validation and verification are often mentioned together with model calibration. This literature review makes no exception from this rule but calibration remains the primary focus, especially the methods that defines such a process.

The second part of this chapter, section 3.2 presents an automatic empirical calibration method, tailored for estimation of Daganzo-Newells fundamental diagram, see 2.1.4 for more details about the fundamental diagram. The calibration method is then described via an example when applying the method to a real data set.

The final part of the this chapter, section 3.3 is section 3.3 where a review of the non-gradient search algorithm called the Complex method. The behaviour of the method is demonstrated by an application to a simple optimization problem with two unknown and a differentiable objective function and for a more relevant problem were a single edge cell transmission model for velocity has its parameters estimated by the Complex method.

These two later parts then form the initial how of model calibration, the start of point to the upcoming calibration method implementation presented in a later chapter, i.e. the framework.

### 3.1 Calibration and validation of traffic models

The centrepiece of this thesis is calibrating a macroscopic traffic model. Traffic models have many useful applications that are not mentioned in this report. As models in general however, they must first be proven to be valid replications of the infrastructure and the traffic dynamics that they capture before the applications can be made [17]. It must be proven that the model is a good enough representation of reality before the user can make any analysis. An important step towards having a valid model is model calibration.

To present traffic model calibration is important since it has been noted that calibration of macroscopic traffic models lack the amount of literature the researcher would like to have available [17]. It might therefore be considered to be an unclosed topic, or a subject of debate and therefore it is important to define what the calibration process is for the researcher that applies it. The following quotes are examples of this:

*"The processes of model verification, validation and calibration are critical to the credibility and reliability of the model results. However, in current traffic engineering practice, there appears to be little uniformity in the definition and conduct of these process elements."*

– Hellinga [17].

*"So far relatively few calibration results for general macroscopic traffic flow models have been reported."*

– Ngoduy & Maher [18].

*"[...] there appears to be a lack of consensus among traffic model developers as to the terminology to be used in the traffic model development and testing exercise."*

– Rakha, et al. [19].

*"[...] it was indicated that simulation model calibration and validation often were discussed and informally practiced among researchers, but seldom have been formally proposed as a procedure or guideline."*

– Park & Qi [20].

Summarizing the literature, it would seem that traffic model calibration is a non-trivial topic and that the subject must be presented more precisely so that there is no question about the meaning of the word in the context of this thesis.

This section will sort out what model calibration and validation is. Model verification will be presented as well as these three terms often are mentioned together.

Calibration of traffic models is here described in such a way that it can be applied in the development of both microscopic<sup>1</sup> and macroscopic traffic models. Much of the literature reviewed regards microscopic traffic models in particular but without being specific to a greater extent than placing model calibration in the model development process and defining what calibration is.

### 3.1.1 Literature survey

In order to remove any of the doubt that was captured in the quotations earlier, a literature survey of twelve different papers was conducted. Some of the reviewed papers covers efforts of how to define calibration and validation in the context of both macroscopic and microscopic models [17–25]. Some of those papers also

---

<sup>1</sup>Microscopic simulation models capture the movement of individual vehicles in highly detailed infrastructure representations. Such model can include signal plans, public transportation schemes, pedestrian movements, individual driver behaviour. Such models are of a lower scale compared to macroscopic simulation models where e.g. the velocity of the entire traffic stream is modelled.

covers examples of implementation of calibration and validation. The other papers make no effort to define these terms, but they do apply calibration techniques [26–28].

The literature survey is summarized by Tables 3.1 (macroscopic models) and 3.1 (microscopic models and one unknown). It must be pointed out that the placement of the authors definitions has been done to the best ability and that if the exact words match it will determine the placement in the table.

### **Verification of traffic models**

Although it is not of primary interest of this thesis it can be noted that the term verification has not passed unnoticed in the cited papers [17, 19, 25, 27, 28].

For the papers regarding microscopic simulation software, can it be noted that model verification leans towards being a process of code reviewing which is carried out in the software development stage [17, 19, 25].

Among the two macroscopic simulation papers that mentioned verification is this term a part of a greater process called quantitative model validation in which parameter values are determined and sanity checked by evaluating model output [27, 28].

### **Calibration of traffic models**

From the literature survey summarized by Tables 3.1 and 3.2 can it be noted that calibration of traffic models leans towards being an iterative process where the model parameters are adjusted in such a way that the model output matches field observations [17, 18, 20, 23–25, 27, 28]. If possible it would be preferable if this iterative change in model parameters was managed by an automated algorithm [18, 20, 23, 27, 28].

### **Validation of traffic models**

The definition of model validation differs between the reviewed papers (see Tables 3.1 and 3.2). However most of those papers, for both macro and microscopic models, states that validation in this context is a process where the output from the calibrated model is compared to field observations [18, 20, 23, 24, 26]. These papers also states that this validation should be performed for datasets spanning different time periods than those used during the calibration process.

Cremer [23] took model validation one step further than this as they included a sensitivity analysis of the model parameters with respect to model output deviation against field observations.

**Table 3.1:** Summary of the literature review. Every row gives the identified definition of the respective term. The table states the literature survey for sources where macroscopic models were considered.

<b>Verification is</b>	<b>Calibration is</b>	<b>Validation is</b>
Not included [18].	An automated iterative process where model parameters are adjusted by an algorithm in such a way that it increases model performance [18].	To compare model output against a dataset for another time period than the calibration time period [18]. To make sure that congestion is captured and propagates correctly in time and space.
Not included [21].	To chose model parameters in such a way that they match field observations [21].	Not included as such, although model output was compared to field observations for a separate day [21].
Not included [26].	To identify model parameters and chose the value of those parameters in such a way that the model output match field observations [26].	To compare calibrated model against a dataset for another time period than the calibration's [26].
Called stage 2 of quantitative model calibration [27]. Application of a model for which parameters have been estimated as to compare the model result against a dataset for another time period than the calibration.	Called stage 1 of quantitative model validation [27]. Iterative process. It aims at parameter value estimation with an algorithm for local parts of the study area.	Mentioned but not presented [27].
Called stage 2 of quantitative model calibration [28]. Application of a model for which parameters have been estimated so as to compare the model result against a dataset for another time period than the calibration's.	Called stage 1 of quantitative model validation [28]. Iterative process. It aims at parameter value estimation with an algorithm for local parts of the study area.	Called qualitative validation and aims to adopt parameter values apprehended during the quantitative validation for the entire study area and manually tune parameters so as to make the model result fit field observations [28].
Not included [22].	A method that includes certain steps such as the definition of geometric properties and split ratio estimations [22].	Not included as such, although they did compare model output to field observations for a separate day [22].
Not included [23].	Called the parameter estimation problem [23]. Iterative process. It aims at parameter value estimation with an algorithm.	To compare the calibrated model against a dataset for another time period than the calibration time period [23]. To perform sensitivity analysis of each individual parameter for this new data set in order to ensure transferability.

**Table 3.2:** Summary of the literature review. Every row gives the identified definition of the respective term. The table states the literature survey for sources where microscopic models were considered, except for the top row where the scale is undetermined.

Verification is	Calibration is	Validation is
The process of evaluating if the model logic is a fair representation of reality [17].	The iterative process of defining and supplying model parameter values to the model depending on the properties of the model subject [17].	The process of comparing the proposed model logic to the actual computer code [17].
Not included [20].	An iterative process where model parameters are adjusted by an algorithm in such a way that it increases model performance [20].	To compare calibrated model against a dataset for another time period than the calibration time period [20]. Make sure that congestion is captured and propagates correctly in time and space.
The process of debugging, unit testing and validate code logic [19]. Also includes sensitivity analysis of the output that corresponds to a common input range.	Ideally the process of selecting input parameter values based on field observations [19].	To determine if the model rules will yield an outcome that is consistent observations and the accepted theory [19].
Not included [24].	An iterative process where model parameters are adjusted by an algorithm in such a way that increases model performance [24].	To compare calibrated model against a dataset for another time period than the calibration time period [24].
A process where the software developer works together with researchers in order to assure that the software implementation agrees with the current traffic theory [25].	An iterative process [25]. The adjustment of parameters in such a way that the model is better suited to reproduce field observations.	Process where the model is evaluated prior to use by testing its ability to replicate field observations [25]. Done for a dataset other than the one used in the calibration.

This concludes the more quantitative part of the calibration literature review. The purpose was to answer what traffic model calibration is. The opening citations certainly left the question open and hopefully it has been answered through the quantitative literature review.

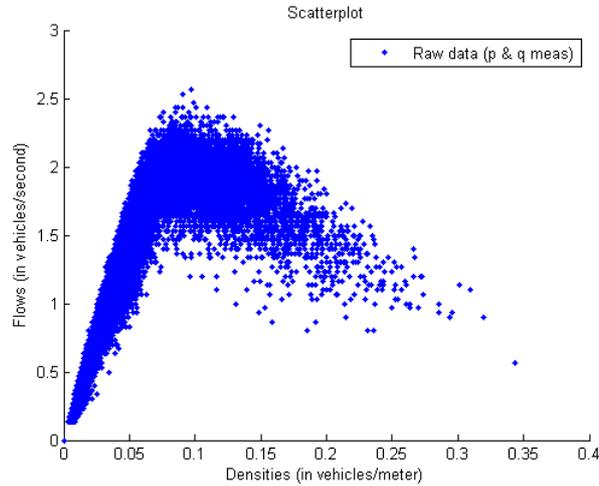
The next section then presents one of two methods presented in this thesis as the answer to how.

### 3.2 Automatic empirical calibration method

In 2009 Dervisoglu et. al [21] presented a calibration method for calibrating parameters directly connected to the fundamental diagram, using data from PeMS. PeMS stands for the Caltrans Performance Measurement System, a monitor of

traffic data which is collected in real-time<sup>2</sup>. The data is collected from the Interstate 880 in the San Francisco Bay area. The procedure was developed for cell transmission models which uses parameters directly connected to the fundamental diagram, for more information on fundamental diagrams, see 2.1.4.

This calibration method requires a triangular representation of the fundamental diagram (i.e. the Daganzo-Newell case) [21]. Another prerequisite is the fundamental relationship  $Q = v * \rho$  and that data from stationary sensors are used, e.g. data from PeMS stations. The calibration algorithm is ordered into three steps. In the first step is the free flow parameter  $v_{max}$  estimated. The next step is to estimate the flow capacity. This gives all parameters for the uncongested state of the diagram. The last step is for estimating the shock wave speed  $w_f$  and the density capacity  $\rho_{max}$ . The algorithm is presented below using an example with real data. Note that the definition of the maximum (capacity) flow is adopted from [29].



**Figure 3.1:** This figure displays the scatter plot of the data (flow and density measurements) from the PeMS station with ID 400309. The data was collected during the dates 2012-01-01 to 2012-03-01.

The PeMS station supplying the data which was chosen as input to the example, around which the algorithm will be explained, is a station on I-880 with PeMS station ID 400309. The station covers five lanes. The scatter plot in Figure 3.1 shows the data collected from 2012-01-01 to 2012-03-01. Note that all data points are represented by tuple of flow and density and that they have been filtered.

### 3.2.1 Regression analysis: Equality constrained least squares

This section will present the regression analysis used for data fitting in steps I in section 3.2.2 and step III in section 3.2.4 of the calibration method. The regression analysis is derived from the constrained least squares method presented

<sup>2</sup>We simply refer to these detectors as PeMS, PeMS detectors, PeMS stations or alike in this thesis. Visit the [www-page pems.dot.ca.gov](http://www.pems.dot.ca.gov) for more information.

in [30]. Equality Constrained Least Squares is a common data fitting tool. It is an optimization algorithm that minimizes the sum of squared residuals. The difference between an observed value and a fitted value provided by a model is the definition of a residual. The result from this algorithm provides a slope and the intersections between the line and the x-axis.

The formulation of the equality constrained least squares problem for solving the regression analysis in step I and step III of the calibration procedure, is formulated by (3.1). The optimization method for the equality constrained least squares is described in [30].

$$\begin{aligned} \text{minimize: } f(x) &= \frac{1}{2}x^T Px + q^T x + r \\ \text{subject to: } Ax &= b \end{aligned} \quad (3.1)$$

where  $P \in S_+^n$  and  $A \in \mathbb{R}^{p \times n}$ . According to [30] the optimal condition for 3.1 can be written as (3.2).

$$Ax^* = b, \quad Px^* + q + A^T v^* = 0 \quad (3.2)$$

(3.2) rewritten is (3.3).

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ v^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix} \quad (3.3)$$

where  $x^*$  and  $v^*$  are variables that help to form  $n + p$  sets of linear equations with  $n + p$  variables. These in turn form the Karush-Khun-Tucker system (KKT system) for this problem. The formulation (3.3) is the KKT system for the constrained equality least squares problem (3.1) and the coefficient matrix is the KKT matrix. The KKT matrix can be of two types, either singular or non-singular.

For non-singular KKT matrices, there is a unique optimal primal-dual pair  $(x^*, v^*)$ . To solve non-singular KKT-matrix, invert the left hand side and the optimal solution gives the optimal  $(x^*, v^*)$  and the Langrange multipliers. The inverse is made via a Cholesky decomposition.

If the KKT matrix is singular but solvable, do a QR decomposition<sup>3</sup>. By doing a QR decomposition it is possible to find the solution to 3.1.

If the problem is infeasible, do a block elimination of the KKT system, using the following system: First reform the problem 3.3 to 3.4.

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} = - \begin{bmatrix} g \\ h \end{bmatrix} \quad (3.4)$$

The elimination algorithm presented by [30], for solving the KKT system is block elimination algorithm. The algorithm consists of four steps.

1. Form  $H^{-1}A^T$  and  $H^{-1}g$ . This is made using a Cholesky factorization of  $H$ .
2. Form the Schur complement  $S = -AH^{-1}A^T$ .
3. Determine  $w$  by solving  $Sw = AH^{-1}g - h$ . This is made using a Cholesky factorization of  $-S$ .

---

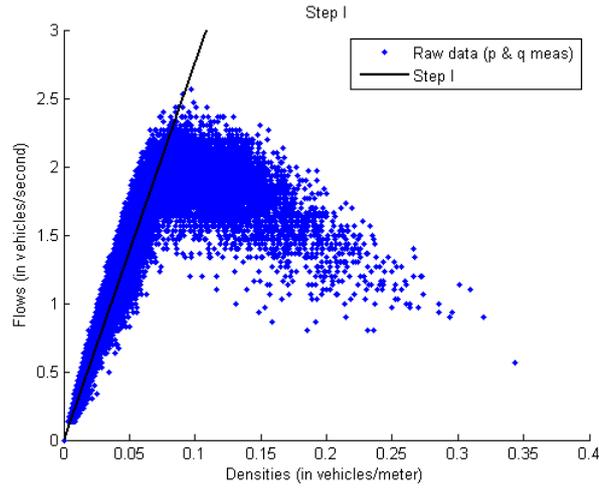
<sup>3</sup>A QR composition is a decomposition of a matrix. Take the matrix  $A$ , let the orthogonal matrix be defined as  $Q$  and the  $R$  be the upper triangle matrix. Then the decomposition of  $A$  is defined as  $A = QR$

4. Determine  $v$  by solving  $Hv = -A^T w - g$ .

Note that further explanation of the linear regression analysis is outside of the scope of this thesis.

### 3.2.2 Step I

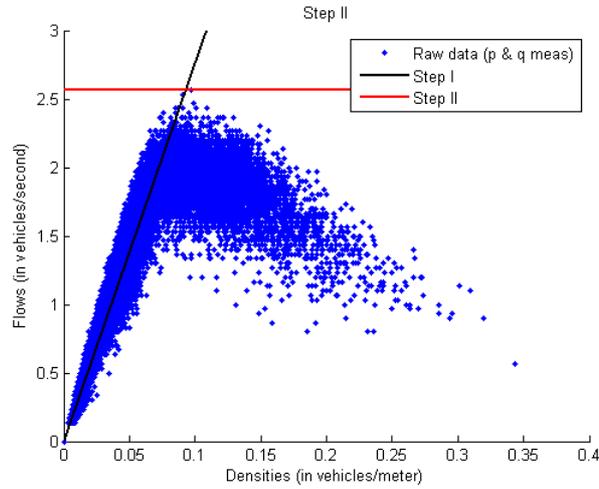
As mentioned earlier, step I is about estimating the free flow velocity. First sort the data points into two sets. One data sets that represents the uncongested regime and one data sets that represents the congested regime. All data points with higher velocity than a certain limit are considered to be in the uncongested regime, in [21] this limit is set to  $> 55$  miles/hour. An equality constrained least squares fit is used on all data points within the uncongested regime. This gives the regression line. The result is an estimated free flow velocity, this value corresponds to  $v_{\max}$ . When using the Daganzo-Newell fundamental diagram the relationship  $v_{\max} = v_{cr}$  is valid. See calibration result from step I in figure 3.2.



**Figure 3.2:** Shows the regression line for step I, where the aim is to estimate the free flow velocity.

### 3.2.3 Step II

To estimate the maximum measured capacity, all measured data are considered. Find the data point representing the highest flow among all data. Let this flow represent  $q_{\max}$ . Solve the trivial equation  $q_{\max} = \rho * v_{\max}$ . This equation gives  $\rho_{cr}$ , thus the critical density is estimated.  $q_{\max} = \rho_{cr} * v_{\max}$  represents the capacity of the highway, where the station is located. It is also necessary to mention that this point is representing the equality constraint when doing the constrained regression analysis on the data in step III.

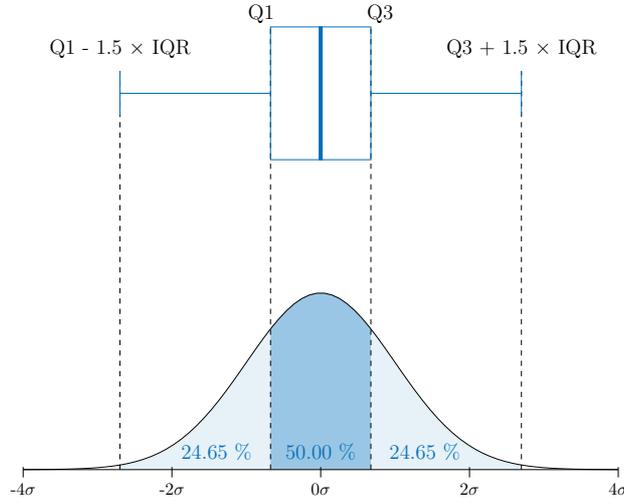


**Figure 3.3:** This figure displays the line for the maximum flow. It also display the intersection between the regression line for free flow velocity and the maximum flow.

### 3.2.4 Step III

Use all data points from the congested regime, the second set of data. Only data from the second set is considered during step III. The data from the second set is partitioned into bins of non overlapping data along the horizontal axis with a certain amount of data in each bin. Each bin contains a defined amount of data points, in [21] they choose a bin size of ten data points, for a graphical explanation see Figure 3.5.

To determine how to estimate if a data point is an outlier or not, the authors of [21] chooses to make use of quartiles in statistical theory. See figure 3.4 for a graphical representation. Figure 3.4 also shows how the box plot relates to a Gaussian distribution. The theory for finding data points that lie within the box plot is used in the binning procedure. If the data point lie outside a specified interval e.g. a value  $x > Q_3$  then it is considered as an outlier. Note that the use of quartiles does not require a Gaussian distribution [31] and that a Gaussian distribution is not assumed later in the automatic empirical calibration procedure.



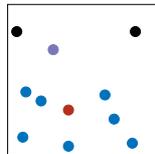
**Figure 3.4:** This figure represents how the different quartiles relates to the mass of probability.

Each bin is then summarized by BinDensity which is equal to the mean of the density values in the bin. After each horizontal bin value (BinDensity) has been determined, then each bins vertical value (flux) is determined via (3.5) and (3.6). The result from (3.6) gives the largest non outlier with the highest flow and density value.

$$\text{Bin} = \{f_1, f_2, \dots, f_i\} \quad (3.5)$$

$$\text{BinFlow} = \max_{f_i} (f_j | f_j \in \text{Bin}, f_i < Q3 + 1.5\text{IQR}) \quad (3.6)$$

where  $i$  is the index of data point  $i$  in each BinDensity,  $j$  is the  $j$  : th bin computed by (3.6) and  $Q3$  is the 75th percentile of the data points in the bin and IQR is the interquartile range.



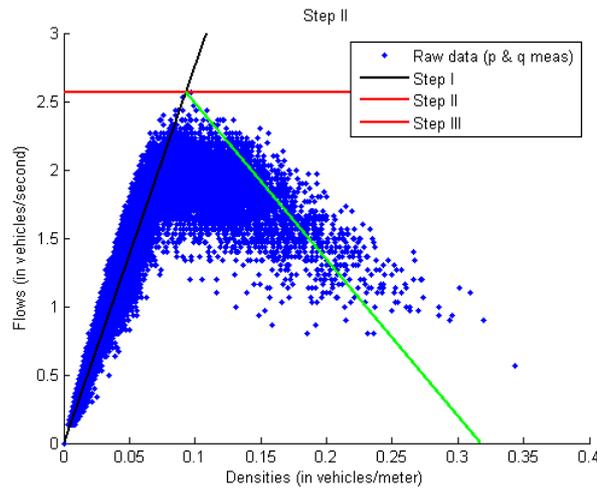
**Figure 3.5:** This figure displays an example bin. The bin size is ten data points. The red point represents the mean density (it does not belong to the actual set and can therefore be placed anywhere in the bin). Black points are outliers, sorted out using (3.6). The purple point refers to the chosen flow. The final value for the bin is the flow from the purple point and the density from the red point.

The result from calculating (3.6) creates a BinDensity - BinFlow pair (Bin-Pair). When all bins have been calculated, a constrained least-square regression method is then applied to the BinPairs, thus obtaining the last relationship of

the fundamental diagram representation. The constraint is given by the free flow capacity. The result from the calibration is summarized in Table 3.3 and in figure 3.6.

**Table 3.3:** This table presents the calibration results from the example. Note that the parameters are estimated using aggregated data from PeMS station ID 400309 from five lanes.

Parameter	Value	Unit
Free-flow velocity	27.65	meter/second
Backwave speed	11.44	meter/second
Maximum density	0.3172	vehicles/meter
Flow capacity	2.567	vehicles/second
Density capacity	0.0928	vehicles/meter



**Figure 3.6:** Shows the resulting fundamental diagram from the calibration process.

### 3.2.5 Remarks

Note that this calibration procedure assumes that the Daganzo-Newell fundamental diagram is used. Due to this, the calibration method needs to be adapted to a hyperbolic-linear fundamental diagram instead. The main reasoning behind this is, when using the Daganzo-Newell fundamental diagram it is assumed that  $v_{cr} = v_{max}$ , which is not the case when using a hyperbolic triangular diagram. How the algorithm is adopted to fit this kind of diagram will be explained in 5.1.

## 3.3 Complex method

The Complex method, introduced by M.J. Box in 1965, is a non-gradient search algorithm for finding the optimum value for non-linear functions depending on

multiple decision variables [32]. The Complex method, or simply the algorithm considering its nature, assumes that the decision variables are enclosed within an admissible region or, in other words, that the target function is the subject to at least one explicit constraint. All constraints are assumed to be in-equality constraints. The fact that the Complex method is a non-derivative search algorithm speaks in favour of many applications since the objective function derivative would never need to be calculated. This would allow the user to choose any objective function during the evaluation of the model performance without any changes to the implementation of the Complex method.

Several authors of different traffic modelling papers have used the Complex algorithm for solving their parameter estimation problems in an orderly fashion, see for instance [23, 27, 28].

This section gives an introduction to the Complex method as it is described in [32–34]. The initial presentation is general and later explained with an example of how to solve trivial optimization problem with the Complex method. The later part of this section concerns the application of the Complex method for parameter estimation in traffic models.

A disadvantage of the algorithm that should be noted is that a solution cannot be proven to be a global optimum (the search effort might get stuck at local minimum or maximum) [33, 34]. This issue is usually handled by running the algorithm several times with different initial conditions within the admissible region. It should also be noted that the algorithm demands that the target function is evaluated many times, which could be costly from a computational perspective. The user should therefore be careful when choosing the dataset to calibrate against, there is probably no need to run a calibration process over a long time period that only encompasses a single traffic state for example.

### 3.3.1 Formal presentation

For explanation of the Complex method, consider the following problem:

$$\begin{aligned}
 \min \quad & f(x_1, \dots, x_N) \\
 \text{subject to} \quad & x_i^L \leq x_i \leq x_i^U, \quad i = 1, \dots, N \\
 & x_i^L \leq x_i \leq x_i^U, \quad i = (N + 1), \dots, (N + M)
 \end{aligned}$$

where the  $N$  constraints are explicit, while the  $M$  constraints are implicit,  $x_i^L$  is the lower limit and  $x_i^U$  is the upper limit for the variable  $x_i$ .

The difference between explicit and implicit constraints is that explicit constraints define a range for the decision variables  $x_i$ ,  $i = 1, \dots, N$  while implicit constraints are either linear or non-linear functions of the decision variables  $x_i$ ,  $i = 1, \dots, N$  [33]. In other words, the variables  $x_i$ ,  $i = (N + 1), \dots, (N + M)$  are definitely functions of  $x_i$ ,  $i = 1, \dots, N$  while the lower bounds  $x_i^L$  and the upper bounds  $x_i^U$  for  $i = (N + 1), \dots, (N + M)$  are either functions of  $x_i$ ,  $i = 1, \dots, N$  or constants.

During the initialization of the Complex method is a so called complex generated from the admissible region that is expressed by the explicit constraints while any violation of implicit constraints are handled retroactively [33]. The complex is a collection of  $N$ -dimensional points and it is assumed that complex

has at least  $K > (N + 1)$ ,  $N$ -dimensional feasible points once it has been generated. The point of choosing a complex with  $K > N + 1$  points is to ensure that the complex keeps its dimensionality when the points are moved around in the  $N$ -dimensional space during the iterations of the Complex method. If the complex has too few members, it could collapse, or rather, flatten itself against the first constraint that is reached during the movement in the  $N$ -dimensional space [32].

The initial complex  $\{x_{ij}\}$ ,  $i = 1, \dots, N$ ,  $j = 1, \dots, K$ , where the subscript  $i$  indicates the variable and the subscript  $j$  indicates the complex member, is generated as follows. Assume that one feasible point is known (one that does not violate any of the implicit and explicit constraints). All other  $K - 1$  points  $j$  are generated randomly as

$$x_{ij} = x_i^L + r_{ij}(x_i^U - x_i^L), \quad i = 1, \dots, N, \quad j = 2, \dots, K \quad (3.7)$$

where  $r_{ij}$  is a uniformly distributed random variable expressed as  $r_{ij} \sim U(0, 1)$ . The reason for why a feasible point must be given is that (3.7) leaves no guarantee of non-violated implicit constraints and a feasible point is needed in order for the Complex method to know in which direction an infeasible point should be moved. Note however that there is no need for an initial point if no implicit constraints are present, the entire initial complex can be generated with (3.7) in that case.

If an implicit constraint is unsatisfied by a generated point  $k$  during the initialization of the complex method, that point is moved half the distance towards the centroid of the points generated thus far. This reveals the importance of having an initial feasible point since if the first randomly generated point violates an implicit constraint, it must be moved towards that point until it is feasible. The centroid  $x_{iC}$  along dimension  $i = 1, \dots, N$  is defined as

$$x_{iC} = \frac{1}{k-1} \sum_{j=1}^k (x_{ij}), \quad j \neq k \quad (3.8)$$

where  $k$  is the index of the last point that was generated and thus the size of the generated complex  $\{x_{ij}\}$  thus far along dimension  $j$ , including the point that violated the constraint. A new complex point is generated according to

$$x_{ik}^N = \frac{x_{iC} - x_{ik}^O}{2}, \quad i = 1, \dots, N \quad (3.9)$$

where the superscript  $N$  denotes a new point,  $O$  denotes the old point that violated any constraint and  $x_{ik}^N$  is the value of the  $i$ th coordinate of the new complex point. This procedure leads to a feasible point, if the feasible region is assumed to be convex, since it can be repeated until all constraints are satisfied.

The points of the complex are randomly scattered over the feasible region once the complex of feasible points is generated. The value of the target function is then evaluated for every single point and it is noted which of the complex members that have the best and worst value. Those two points are used to checked the imposed convergence criteria

$$f_{max} - f_{min} < \beta \quad (3.10)$$

Reference [33] notes that  $(10^{-3} - 10^{-4})f_{max}$  is a common value of  $\beta$ .

The worst point generated, called  $x^W$ , is rejected if (3.10) is a false statement. If the problem is a minimization problem, the worst point is of course the one with the greatest function value. In this case, rejecting the point means that it is moved towards the centroid of the remaining complex members, creating a new point  $x^{New}$  according to (3.11).

$$x_{ik}^{New} = \alpha(x_{iC} - x_{ik}^W) + x_{iC}, \quad i = 1, \dots, N \quad (3.11)$$

where the expression for the centroid  $x_{iC}$  is similar to that expressed by (3.8). As for the value of the reflection parameter, or step length,  $\alpha$  in (3.11), [32] found that an  $\alpha = 1.3$  should be a suitable choice.

The new point  $x_{ik}^{New}$  is evaluated to see if it satisfies the explicit constraints. If it is not, it is moved a small distance  $\delta$  within the limit of the constraint that is violated (and then rechecked again). Reference [32] states that  $10^{-6}$  has proven to be a functional value of  $\delta$  and also that the chance of the corrected point to be within the subspace of the other points should be fairly small.

If any implicit constraint is unfulfilled after the reflection of the worst point, the trial point should be moved halfway towards the centroid of the better points just as during the generation of the initial complex and according to (3.9), but then with  $K - 1$  in the denominator.

In the original formulation of the complex method the progress stated by (3.10) - (3.11) is repeated until the entire complex (the set of points) has moved to the centroid. At that time the convergence criteria expressed by (3.10) should be fulfilled. In other words, the algorithm stops when the complex (the points) has collapsed to a singular location in the parameter space, in which case the target values are nearly the same for each individual point of the complex.

It has been noted that the Complex method can get stuck around a local optimum if that local optimum happened to be the centroid [34]. By repeating the movement for the worst member<sup>4</sup> is a way to overcome this issue. A trial point can be expressed by a gradual movement of the point towards the minimum value as long as the trial point continues to be the worst one of the complex. It is a risk though that this might result in that a trial member is too near another point which could risk a collapse of the complex. This risk of a collapse can be handled by the introduction of a new random value that is added to the trial point. Also, this random search increases the search effort in relation to the simple reflection. This modification (or extension since this new movement is introduced only if a trial point is a repeater) of the Complex method can be expressed by (3.12).

$$x_{ik}^{N(new)} = \frac{x_{ik}^{N(old)} + \varepsilon x_{iC} + (1 - \varepsilon)x_{ik}^B}{2} + (x_{iC} - x_{ik}^B)(1 - \varepsilon)(2R - 1) \quad (3.12)$$

where  $x_{ik}^B$  is the best point of the complex,  $R$  is a uniformly distributed random number,  $R \sim U(0, 1)$  and  $\varepsilon$  is defined by (3.13).

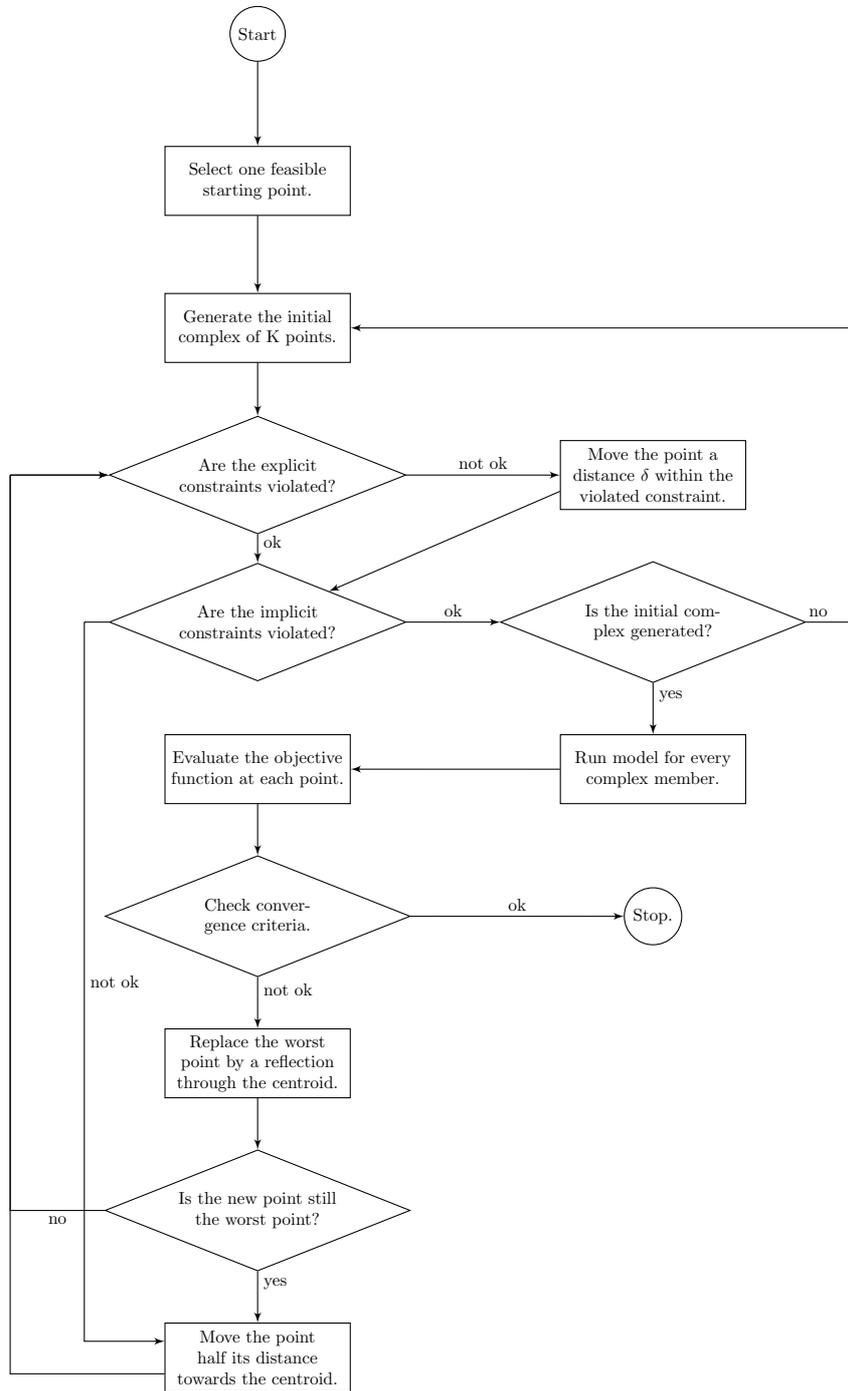
$$\varepsilon = \left( \frac{n_r}{n_r + k_r - 1} \right)^{\left( \frac{n_r + k_r - 1}{n_r} \right)} \quad (3.13)$$

where  $k_r$  is the number of times that  $x_{ik}^N$  in (3.12) has been the worst point (a repeater),  $n_r$  is a constant. A  $n_r = 4$  could be an adequate choice for that

<sup>4</sup>The repeated point is called a trial point prior to that, if it is not longer the worst one.

constant. Note that the previous worst point  $x_{ik}^{N(ol d)}$  in (3.12) is always updated to be the previous trial point.

A flow chart for Complex method can be found in [33, p. 119]. That model was followed when Figure 3.7 was created, this figure shows the overall process of the Complex method if one chooses to utilize the half distance movement towards the centroid and if there are implicit constraints present.



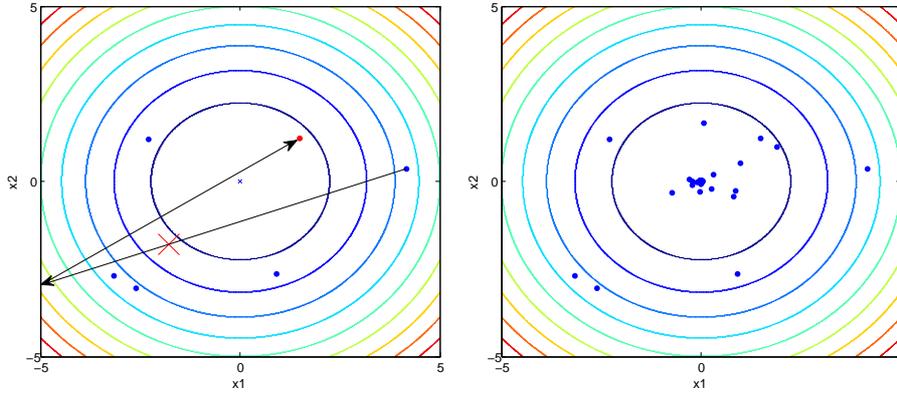
**Figure 3.7:** Complex method (algorithm) presented as a flow-chart for solving in-equality constrained optimization problems.

### 3.3.2 Example: Visualization of the complex movement

In this section, an example of the Complex algorithm application is presented. The goal is to find optimum solution to the following minimization problem:

$$\begin{aligned} \min \quad & f(x_1, x_2) = x_1^2 + x_2^2 \\ \text{subject to} \quad & -5 \leq x_i \leq 5, \quad i = 1, 2 \end{aligned}$$

The target function  $f(x_1, x_2)$  is one with a global minima within the admissible domain at  $(x_1, x_2) = (0, 0)$  for which  $f(x_1, x_2) = 0$ . In the figure below it is possible to follow the movement of the complex members as they end up in the complex centroid near the optima.



**Figure 3.8:** An example of how the complex moves towards the optima with the first movement on the left and the complete movement on the right. The red cross marks the current centroid. The leftmost figure depicts a movement where the complex method tries to move a point outside the boundaries.

### 3.3.3 Example: Calibration of a single edge highway model

This section should be seen as a bridge between a highway model and the Complex model. The target of highway model calibration is to decrease the difference between model output, predicted state and field observations. This difference is expressed by a function that often returns a single scalar value. In this section the traffic model will be described as dynamic state vector equations so that the parameter estimation problem is well posed.

Recall the cell transmission model for velocity (CTM-v). The CTM-v expressed that the velocity field on the highway evolves in time according to certain rules expressed by the transformed Godunov velocity flux functions. With the ambition to calibrate by comparing model output define three vectors:

$$\mathbf{v}^T = [v_0, v_1, \dots, v_{i_{max}}] \quad (3.14)$$

$$\mathbf{u}^T = [v_{-1}, v_{i_{max}+1}] \quad (3.15)$$

$$\mathbf{y}^T = [\mathbf{v}, \mathbf{x}] \quad (3.16)$$

Where  $\mathbf{v}$  state vector holding the velocity for each cell  $i \in \{0 \dots i_{max}\}$ ,  $\mathbf{u}$  is an input vector holding the conditions to be imposed at the boundaries and  $\mathbf{y}$  is an output vector holding the state at the measurement point as well as the (current) parameter set  $\mathbf{x}$  that led to that output.

The state vector  $\mathbf{v}$  represents the velocity field over a single edge (for which one fundamental diagram is used). The edge consist of  $i_{max} + 1$  cells.

The input vector  $\mathbf{u}$  holds measurements from the boundaries that will substitute references to what was referred to as ghost points earlier, i.e. they do not lie in the physical domain but provide the boundary conditions to be imposed on the modelled highway.

The output vector  $\mathbf{y}$  can in general be any traffic state variable. It is important to note which kind of data source the model output should be compared with. If the user wants to use stationary speed detectors then the velocity field from the model output may have to be transformed from space mean speed to time mean speed (since the detector only measures speed at a single point in space).

The model is characterized by a vector of unknown parameters  $\mathbf{x} = (\omega, v_{max}, \rho_{max})$  (e.g. the free flow speed). This vector is the estimation subject and is treated as a variable during the parameter estimation procedure.

Assuming that the velocity field evolves according to the CTM-v, which now is label with  $F$  (and not the standard update algorithm label  $\mathcal{M}$  since only a single edge is considered in this explanation), can a non-linear dynamic state vector equation be expressed according to

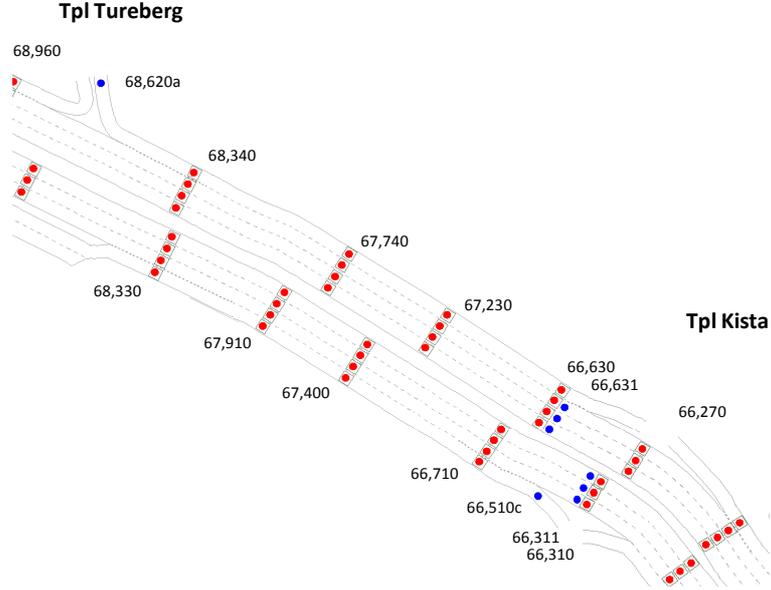
$$\mathbf{v}(n+1) = F[\mathbf{v}(n), \mathbf{u}(n), \mathbf{x}] \quad (3.17)$$

The output vector is defined by

$$\mathbf{y}(n) = T[\mathbf{v}(n), \mathbf{x}] \quad (3.18)$$

where  $T$  is an operator that transforms the values in the output vector  $\mathbf{v}$  to the domain in which the field measurements lie in (e.g.  $T$  represents a process where a single cell is extrapolated). The objective is to find the parameter set  $\mathbf{x}$  that minimizes the deviation between model output and the field observations  $\hat{\mathbf{y}}$ . As a short proof of concept the Complex method is now to be used for parameter estimation of a small CTM-v model.

Consider Figure 3.9, it depicts sensor arrays (detectors) on the highway E4 near Kista, Sweden. The test site stretches from detector 68,330 to detector 66,710 and traffic travels in that direction. In between these two detectors lies detector 67,400. All of these detectors captures the velocity on the highway over time. This stretch of highway spans over 1.7 km and is divided into 17 cells  $i \in \{0 \dots 16\}$ , each being 100 meters long ( $\Delta x = 100m$ ).



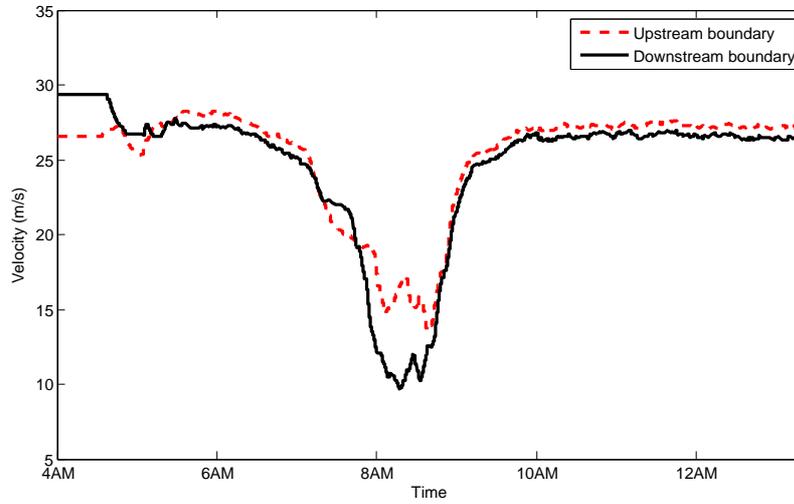
**Figure 3.9:** Map over the example area. The dots show sensor arrays placed along the highway. Tpl is a short form of "trafikplats", the Swedish word for interchange.

Assume that detector 68,330 provides velocity measurements that can be imposed as boundary conditions for cell  $i = 0$ , i.e. these measurements are the ghost cell references where  $i = -1$ . Similarly, the measurements of detector 66,710 are imposed as boundary conditions for cell  $i_{max} = 16$ . The boundary conditions are given in Figure 3.10.

Now assume that detector 67,400 provides measurements that are considered to be the target of the model output, i.e. this detector provides measurements that are considered to be the best possible result that the model can produce. The Complex method can be integrated with the CTM-v model that replicates the highway stretch. The parameters of the CTM-v model ( $\omega, v_{max}, \rho_{max}$ ) can then be estimated by following the process of Figure 3.7 (without the burden of any implicit constraints). The only thing missing is an objective function that expresses the agreement between the measurements given by detector 67,400 and the model output. Introduce the mean absolute percentage error (MAPE) for this purpose:

$$MAPE = \frac{100\%}{n_{max}} \sum_{n=0}^{n_{max}} \left| \frac{\hat{y}^n - y^n}{\hat{y}^n} \right| \quad (3.19)$$

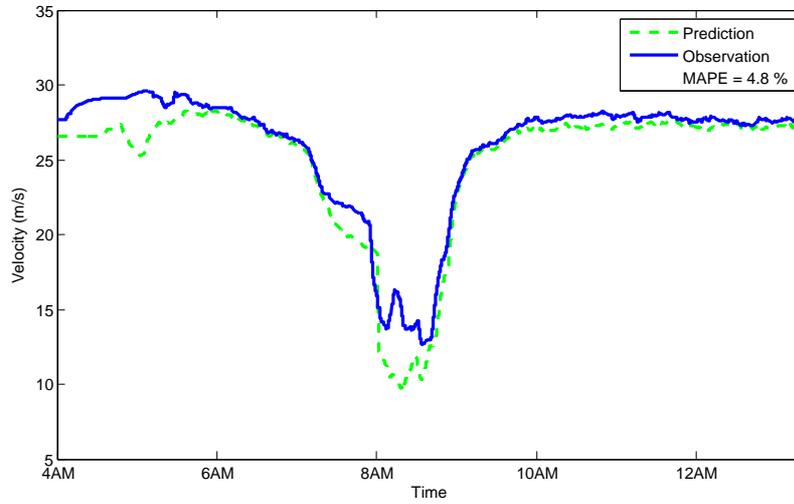
The MAPE measures the mean absolute deviation and is therefore the accuracy of the model output  $y^n$  at one cell  $i \in 0 \dots i_{max} = 16$  compared to the observations  $\hat{y}^n$  at that same location.



**Figure 3.10:** The boundary conditions that were imposed on the example model. The graph shows velocity vs. time for the upstream and downstream detectors 68,300 and 66,710 respectively.

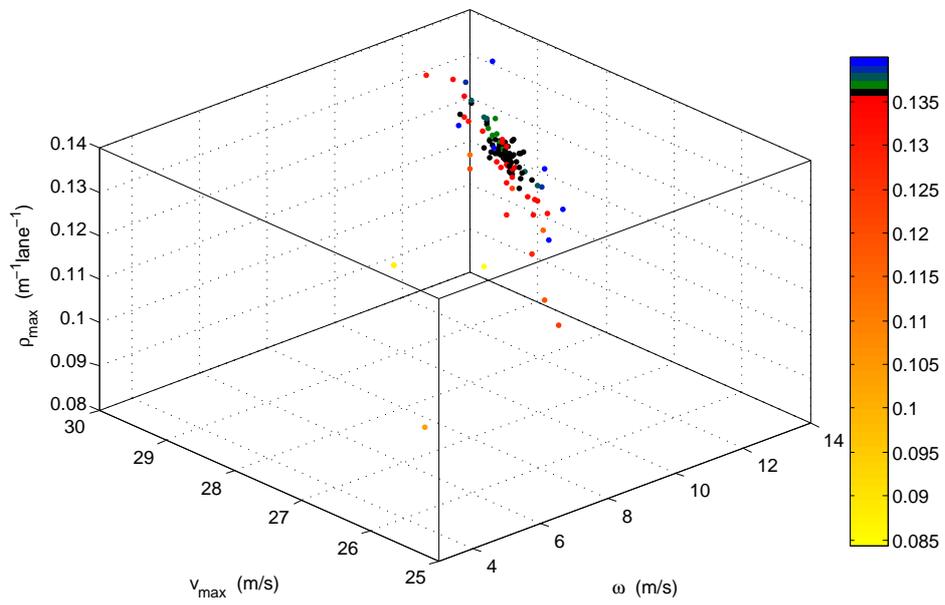
Figure 3.11 gives the final result from the calibration effort using the Complex method. The time series labelled prediction is  $\mathbf{y} = \mathbf{v}_9$  i.e. the model output from the 10th cell. The other time series in Figure 3.11 is the actual observation  $\hat{\mathbf{y}}$  done by detector 67,400. This shows that the Complex method can be used to determine unknown parameter sets for the CTM-v model.

Since only three parameters were variable during the calibration it is possible to plot the movement of the complex as it converges to the local minima. This movement is depicted in Figure 3.12. The colour scale maps the values of the maximum density  $\rho_{max}$ , specifically, the colour black maps to near optimal values of  $\rho_{max}$ . The purpose of Figure 3.8 is to show the behaviour of the Complex method.



**Figure 3.11:** Model output after calibration in dashed green versus observed velocities. The mean absolute percent error was 4.8 % after calibration.

This concludes the presentation of the Complex method for parameter estimation. The Complex method, a non-gradient search algorithm was presented. That it is of the non-gradient type speaks in favour of its use since  $\mathcal{M}$ , the update algorithm of the CTM-v presented earlier, is not differentiable. On top of that the ensemble Kalman filter is wrapped around the  $\mathcal{M}$  for the purpose of data assimilation. A black-box algorithm, such as the Complex method, is therefore attractive. The actual implementation of the Complex method is presented in a later chapter.



**Figure 3.12:** The movement (all iterations) of the complex during the calibration of the example model. Each complex point is given in the coordinates  $(\omega, v_{max}, \rho_{max})$ .

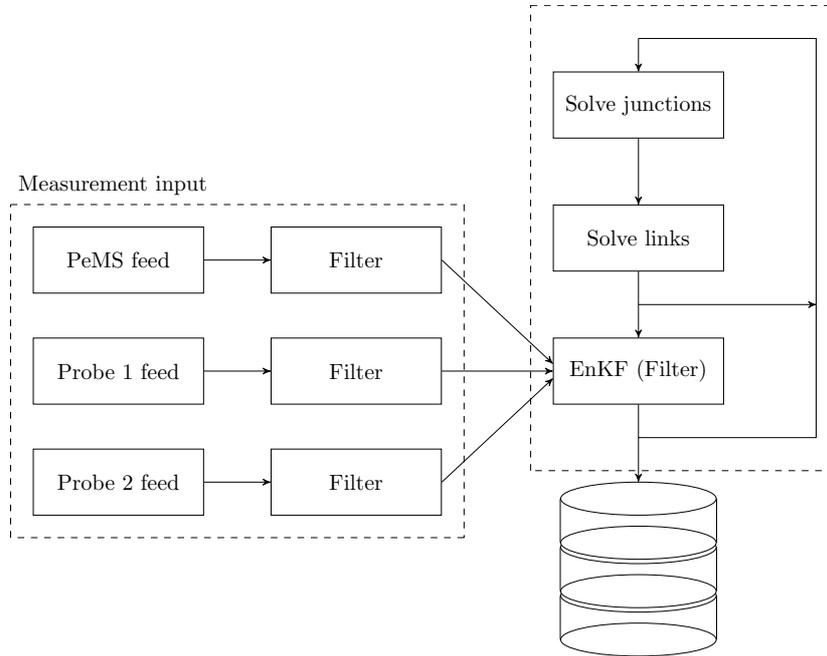
## Chapter 4

# System Description

This chapter is mainly an outcome from the system analysis that were conducted and the purpose of introducing this chapter, is to provide an overview of the relevant parts of the Mobile Millennium system for the reader. It also serves the purpose of mapping the different parts of the system that are used in the calibration framework that this report is presenting. It will also present a list of all parameters that the framework will try to estimate and calibrate.

### 4.1 Overview

The system mainly consists of three parts; measurement loader, highway model and the ensemble Kalman filter. All of these three parts have hard coded, ad-hoc calibrated parameters. Figure 4.1 shows how the data flows throughout the system. Data is loaded from a database where data sources can be of type probe or stationary data. All data that is loaded is pre filtered. The user can choose to use data from no data source, a specific data source, some specific data sources or all data sources. The highway model utilize the CTM-v model together with the velocity from the current time step to predict the velocity field for the next time step. The default time step is six seconds. The third part is the ensemble Kalman filter; it fuses the predicted data with the measured data. All parameters found in the system when conducting the system analysis are listed in the Appendix A. As mentioned earlier in this section, all parameters listed in this section are all static and hard coded. This makes it impossible to calibrate them in an automatic matter. Therefore system modifications were necessary. The modifications made will be explained later in the thesis.



**Figure 4.1:** This figure shows the data flow for the Mobile Millennium system. The measurement loader loads data from the database with stored data and feeds it to the highway model predicts and fuses the data using the ensemble Kalman filter.

## 4.2 Measurement loader

The main purpose of the measurement loader is to enable the user to choose and load any data from any source available in the database. There are hard coded parameters with significance to the system in the measurement loader. More about why it is significant is mentioned in after Table 4.1. Therefore this part of the highway model is relevant for this thesis. The relevant parameters for measurement loader is summarized in Table 4.1.

**Table 4.1:** This table displays all parameters connected to measurement loader that is relevant for the thesis. Note that the parameters in this table are for one lane.

Parameter	Current Value	Unit
Free-flow velocity	27.0	meter/second
Critical velocity	25.0	meter/second
Backwave speed	5.36	meter/second
Maximum density	0.124	vehicles/meter
Flow capacity	0.667	vehicles/second
Density capacity	0.0247	vehicles/meter

The data where  $v > v_{cr}$ <sup>1</sup> is too noisy for the ensemble Kalman filter when

<sup>1</sup>These measurements are considered to exist in the uncongested regime

using stationary sensors. Therefore, a fit of the data in the uncongested region is therefore made using the fundamental diagram. The flow gets mapped according to the fundamental diagram and transformed to velocity<sup>2</sup>. Therefore the fundamental diagram parameters for each link is connected to the measurement loader.

### 4.3 Highway model

The highway model predicts the traffic state for the next time step. The input into highway model is the current state for the current time step. Parameters connected to highway model is displayed in Table 4.2.

**Table 4.2:** This table displays all parameters connected to the highway model that is relevant for the thesis. All parameters connected to the fundamental diagram is based on lane.

Parameter	Current Value	Unit
Split ratio	defined for each junction	percentage
Sinks	5*max capacity	vehicles
Sources	0.6*max capacity	meter/second
Free-flow velocity	27.0	meter/second
Backwave speed	5.36	meter/second
Maximum density	0.124	vehicles/meter
Flow capacity	0.667	vehicles/hour
Density capacity	0.0247	vehicles/meter

Note that measurement loader and the highway model have common parameters. This is due to certain properties in the system. The system is made for using the ensemble Kalman filter, which is not stable when having to noisy measurements. This means when measurements are considered to be in uncongested regime<sup>3</sup> from stationary sensors is tampered with, since the ensemble Kalman filter cannot cope with noisy data.

### 4.4 Ensemble Kalman filter revisited

The ensemble Kalman filter is responsible for fusing the predicted state with the measured state. In Table 4.3 all relevant parameters for the ensemble Kalman filter are summarized.

<sup>2</sup>This is only true if the measurement loaded lies within the density domain

<sup>3</sup>The measurements in uncongested regime are generally too noisy for the ensemble Kalman filter.

**Table 4.3:** This table displays all parameters connected to the ensemble Kalman Filter that is relevant for the thesis.

<b>Parameter</b>	<b>Current Value</b>	<b>Unit</b>
Model mean error	multiple values	meter/second
Model covariance error	multiple values	meter/second
Observation mean error probes	multiple values	meter/second
Observation covariance error probes	multiple values	meter/second
Observation mean error stationary detector	multiple values	meter/second
Observation covariance error stationary detector	multiple values	meter/second

## Chapter 5

# Implementation of Calibration Procedures

This chapter will present how the authors implemented the calibration procedures and the framework. The first section, section 5.1 will focus on the implementation of the automatic empirical calibration method. The next section, section 5.2, will focus on how to assign parameters connected to the fundamental diagram to links. The last section of this chapter, section 5.3, presents how the complex method have been adapted, implemented and integrated into the system.

### 5.1 Implementation of the automatic empirical calibration method

As mentioned in 4.3, due to the fundamental diagram that is used in the system, see (2.20), it is not possible to apply the calibration method without modifications. This section will present three adapted versions of the modified calibration method. Note that later on, the unmodified calibration method is referred to as the original, the modified calibration methods are referred to as versions.

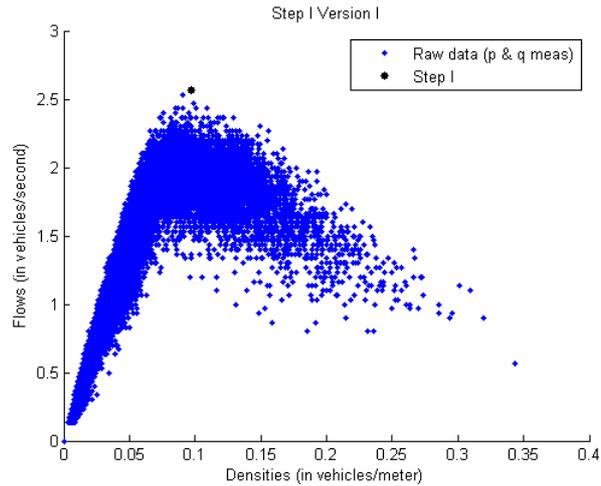
The method is going to calibrate the parameters;  $\rho_{\max}$ ,  $\rho_{cr}$ ,  $v_{cr}$ ,  $v_{\max}$ ,  $w_f$ , and  $q_{\max}$ , the parameters for the Hyperbolic-Linear diagram for each stationary sensor in the system. The example from the section 3.2, is used as an example in this section as well.

#### 5.1.1 Version I

This version utilize somewhat the same approach as the original method. However, the order in which the parameters are estimated as well as the methods for estimating them are changed. Instead of finding  $v_{\max}$  first, this version focus on finding the capacity first, which is expressed by  $v_{cr} * \rho_{cr} = q_{\max}$ . After that, the second step is to find the back wave propagating speed  $w_f$  and the maximum density  $\rho_{\max}$ . The last step is to estimate the free flow velocity  $v_{\max}$  and to do this, the continuity constraint (2.21) is used. A step by step description of the method is as follows.

### Step I

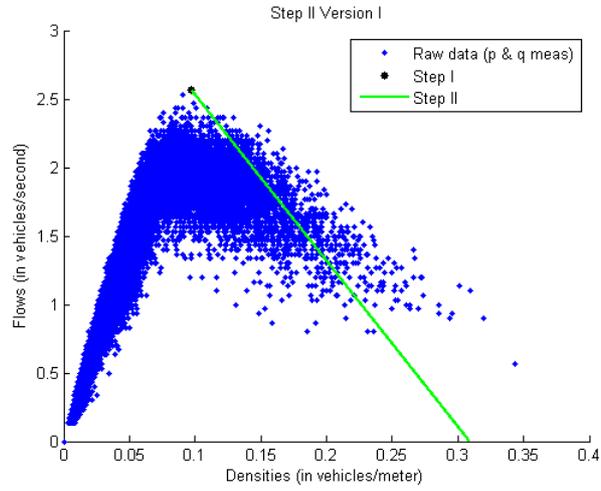
Find the maximum capacity by searching through all data points. Search through the measured data points, find the data point with highest flow measured. Let that data point with the highest measured flow and density represent the capacity. Alas  $v_{cr}$ ,  $\rho_{cr}$  and  $\rho_{max}$  estimated. The graphical result from step I is presented in 5.1.



**Figure 5.1:** Shows the maximum capacity found.

### Step II

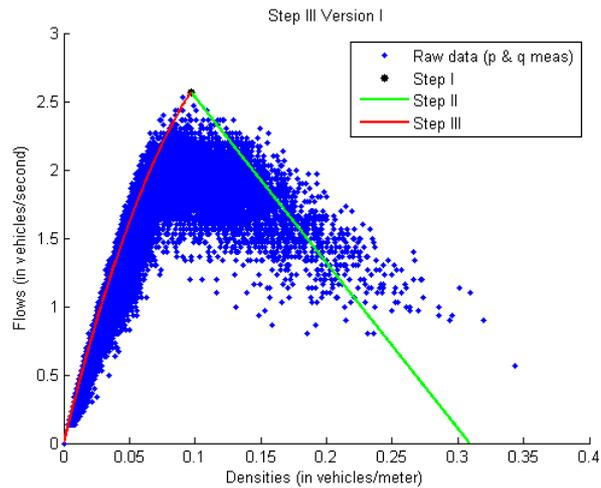
The back wave propagating speed  $w_f$  as well as the maximum density  $\rho_{max}$  is estimated through the same binning method as in the original, for the exact method see 3.2.4. The graphical result from step II is presented in 5.2.



**Figure 5.2:** Shows the result from the binning process and regression analysis in step II.

### Step III

This step involves finding the last parameter  $v_{\max}$  using the continuity constraint. Since  $\rho_c$ ,  $\rho_{\max}$ ,  $w_f$  already have been estimated, the continuity constraint (2.21) from 2.1.4 can be used to estimate  $v_{\max}$ . The graphical result can be viewed in figure 5.3. All calibrated parameters for version I can be found in 5.1.



**Figure 5.3:** This figure shows the calibration result from step III version I.

**Table 5.1:** Calibration results for the example in section 3.2. Note that the parameters are estimated using aggregated data from five lanes.

Parameter	Value	Unit
Free-flow velocity	38.62	meter/second
Backwave speed	12.11	meter/second
Maximum density	0.3087	vehicles/meter
Flow capacity	2.567	vehicles/second
Density capacity	0.0968	vehicles/meter

### 5.1.2 Version II

This version is using the same approach as version I, the main difference is how the capacity is estimated. Instead of using the intersection between the free flow velocity and the data point with the highest flow measured, this version instead applies the binning concept to the data points with the highest flow measured. The result from this binning is used to find the capacity of the road. To estimate the back wave propagating speed and density max, the same approach as the original is used. The last step is to estimate the free flow velocity and to do this, the continuity constraint is used in this version as well.

#### Step I

Just as version I, this step involves finding the maximum capacity. Instead of using the data point with the highest flow, bin the data points with the highest flows instead. The characteristics of data, where the flow is near its maximum, makes it reasonable to bin the tip in the following way. Sort all data point by flow. Choose a bin size  $N$ . Binning of the tip is motivated by the fact that the data point, even though filtered, are not trust worthy. Note that the filtering is more strict in this version of the binning method. The binning is made using (5.1)-(5.3).

$$\text{Bin} = \{f_1, f_2, \dots, f_n\} \quad (5.1)$$

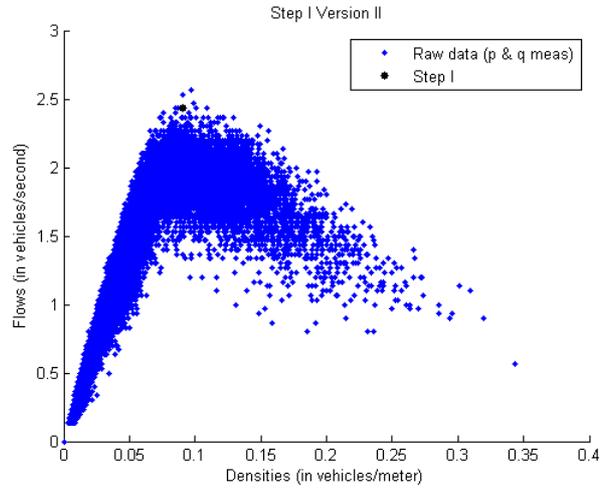
Start to estimate the density of the bin, by using (??) and (5.1).

$$\text{BinDensity} = \max_{\mathbf{f}_i} (f_j | f_j \in \text{Bin}, f_i < Q3) \quad (5.2)$$

Start to estimate the density of the bin, by using the bin from (5.1) and (5.3).

$$\text{BinFlow} = \min_{\mathbf{f}_i} (f_j | f_j \in \text{Bin}, f_i > Q1) \quad (5.3)$$

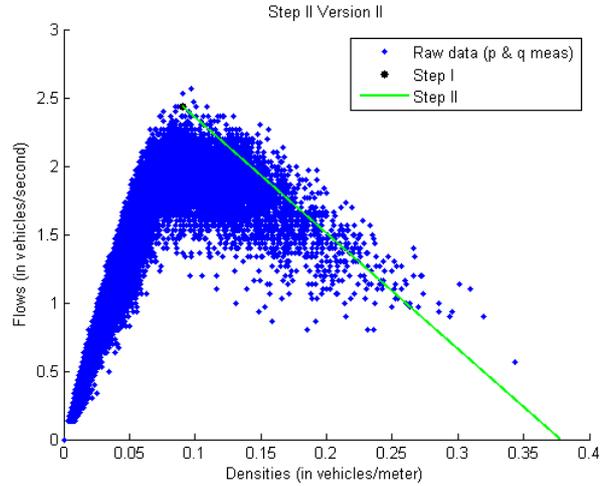
The result from the binning method of the tip is an estimate of the maximum capacity of the highway. Alas the  $\rho_c$ ,  $v_c$  and  $q_{\max}$ .



**Figure 5.4:** This figure depicts the results from step I, when trying to calibrate the parameters for the fundamental diagram using version II.

### Step II

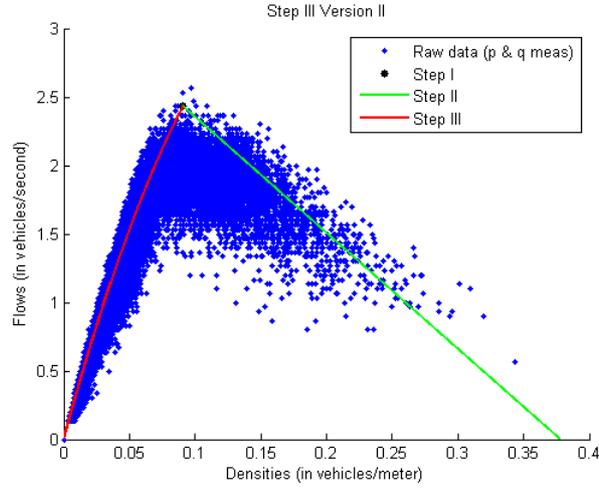
The back wave propagating speed  $w_f$  as well as the maximum density  $\rho_{max}$  is estimated through the same binning method as in the original, see 3.2.4.



**Figure 5.5:** This figure depicts the results from step II, when trying to calibrate the parameters for the fundamental diagram using version II.

### Step III

Use the continuity constraint to estimate  $v_{max}$  exactly in the same way as described in Version I step III. The graphical solution is visualized in figure 5.6



**Figure 5.6:** This figure shows the final results from using version II.

**Table 5.2:** This table presents the calibration results for the example in section 3.2.

Parameter	Value	Unit
Free-flow velocity	35.33	meter/second
Backwave speed	8.469	meter/second
Maximum density	0.3778	vehicles/meter
Flow capacity	2.433	vehicles/second
Density capacity	0.09057	vehicles/meter

### 5.1.3 Version III

Step I, step II and step III is identical to the original automatic empirical calibration method except that it have been extended with one step. The last step in this version , the same method for estimating  $v_{\max}$  as version I and version II.

#### Step I

Just as the original method, see 3.2.2, find an initial value of  $v_{\max}$ , using a linear regression on all data points below a certain density to find  $v_{\max}$ . For graphical representation see figure 3.2. For more detailed information see section 3.2.2.

#### Step II

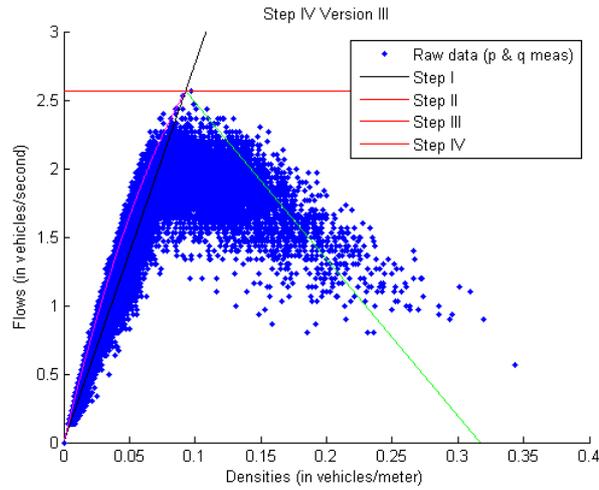
To find the maximum capacity ( $v_c$ ,  $\rho_c$  and  $q_{\max}$ ), find the largest flow. Just as step I, step II is a direct copy of step II from the original method, see 3.2.3. Use the intersection between this line and the regression line from  $v_{\max}$ . For graphical representation see figure 3.3.

### Step III

The back wave propagating speed  $w_f$  as well as the maximum density  $\rho_{\max}$ , is estimated through the same binning method and constrained least square regression analysis as in the original method, see 3.2.4.

### Step IV

This step involves finding the last parameter  $v_{\max}$  using the continuity constraint, see 5.1.1. The graphical result is presented in 5.7 and the estimated parameters is summarized in Table 5.3.



**Figure 5.7:** This figure depicts the results from step IV, when trying to calibrate the parameters for the fundamental diagram using version IV

**Table 5.3:** This table presents the calibration results for the example in section 3.2 using version III.

Parameter	Value	Unit
Free-flow velocity	39.08	meter/second
Backwave speed	11.44	meter/second
Maximum density	0.3172	vehicles/meter
Flow capacity	2.567	vehicles/second
Density capacity	0.0928	vehicles/meter

#### 5.1.4 Implementation remarks

The fundamental diagram calibration procedures was implemented into the system as an independent module. In general this means that the parameters can be calibrated without any interaction with the highway model. The calibration procedure loads filtered data from a PeMS with a certain time aggregation for a defined time interval, the default setting for the aggregation period is 30 second long. The outcome of the measurement during the aggregation is the mean

traffic state for the aggregated time period on the point on the highway where the PeMS station placed. The measured data point consists of three measurements, the velocity, the flow and the density, where the velocity measured is the mean velocity over the 30 second period; the flow is equal to the mean flow; the density is the mean density. After all data have been loaded by project, the fundamental diagram calibration procedure starts. When all parameters have been estimated, they are inserted into the database and keyed by the sensors ID. This enables the system to allocate a fundamental diagram to each link specifically. Since the calibration method is automated and the data from the PeMS stations are not trustworthy, even though the data is filtered, a quality metric was introduced. This quality metric estimates the ratio of data points between the congested regime and the uncongested regime, data availability and if the parameters are within reasonable bounds.

## 5.2 Link assignment

It is a requirement that each link should have an individual fundamental diagram. To assign the fundamental diagrams to specific links, two different methods were developed. This section will present the two methods for assigning the fundamental diagrams. The two methods are similar to each other, the main difference between them, is how the fundamental diagram assignment is made relative to which direction the traffic is streaming. The methods consists of three steps each. Figure 5.8 shows the three steps graphically. Step I and step II is identical for both of the methods. A link is defined as  $k_{n,j}$  where  $n$  is the order,  $n = 1, \dots, N$  and  $j$  is the assigned fundamental diagram. Note that the notations for links are specific for this section only.

### Step I

Define a list of links that is going to be assigned a flux function. The list needs to be sorted according to position and traffic direction,  $\{k_{1,j}, \dots, k_{n,j}\} = K$ .

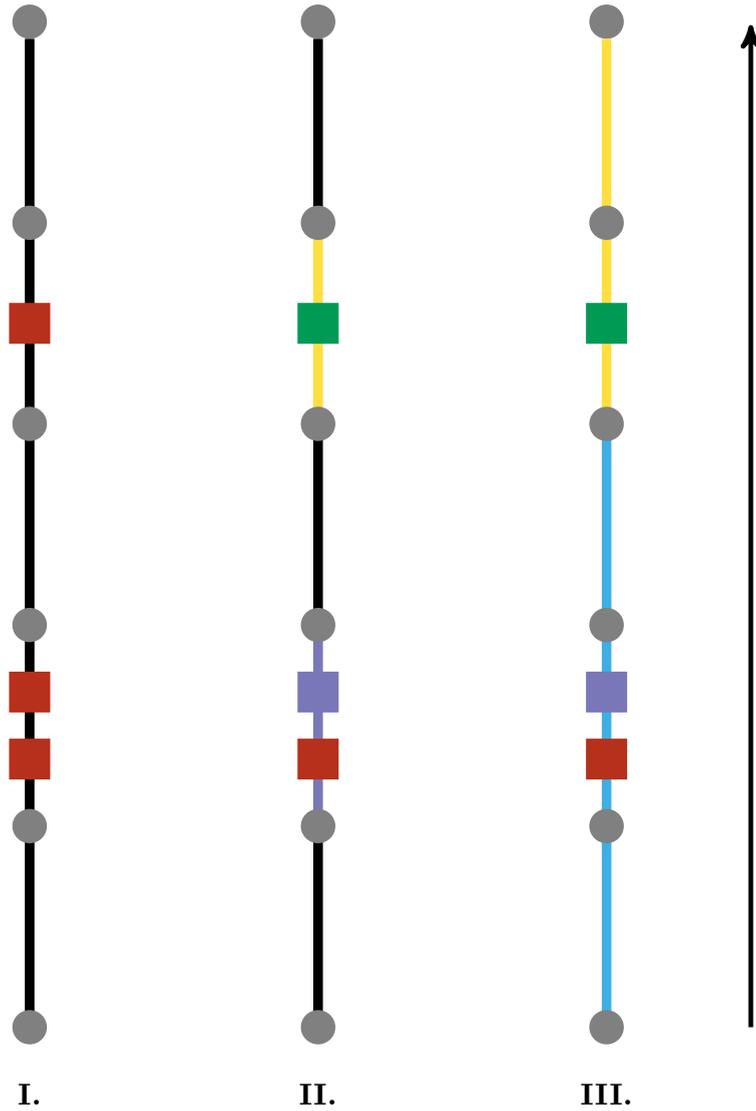
Example: Consider a small stretch of road represented by five links  $k_{1,j}$ ,  $k_{2,j}$ ,  $k_{3,j}$ ,  $k_{4,j}$  and  $k_{5,j}$ . The links are positioned geographically in the following order, first  $k_{1,j}$ , second  $k_{2,j}$  and third  $k_{3,j}$ . The traffic travels from link  $k_{1,j}$  to  $k_{5,j}$ . They can be considered sorted according to position and direction only and only if,  $\{k_{1,j}, k_{2,j}, k_{3,j}, k_{4,j}, k_{5,j}\} = K$  where  $K$  is the list of the ordered set of links. See figure 5.8 I for graphical representation for step I.

### Step II

Go through each link in the network. For each  $k_{n,j}$  do: if there are no sensor connected to the current link, do nothing. If there are one sensor connected to the current link, assign the fundamental diagram  $j$  related to the sensor to the link  $k_{n,j}$ . Example: Go through each link, if there exist a sensor on the link, assign the fundamental diagram from that sensor to the link. If there are more than one sensor connected to the current link. Assign the fundamental diagram with the highest capacity to the link. Since the placement of the sensor on the link relative to the on and off ramps is not known, this is a way for avoiding ghost bottlenecks introduced by low capacity. The result from step

II is  $\{k_{1,1}, k_{2,1}, k_{3,2}, k_{4,2}, k_{5,2}\} = K$ . See the graphical result for step II in figure 5.8 II.

The method requires that there is at least one sensor connected to a link in the network. It is also required that there is an existing fundamental diagram related to the sensor.



**Figure 5.8:** This figure displays a graphical example for each step for the link assignment. Circles is junctions, red squares is calibrated PeMS stations yet to be assigned to a link, black lines are links. The arrow displays the traffic direction. Links and PeMS stations with other colours are assigned a fundamental diagram.

### 5.2.1 Forward link assignment

The forward link assignment assumes that the capacity and other road properties propagates in the same direction as the traffic stream.

#### Step III

Find the first link  $k_{m,1}$  with an assigned fundamental diagram in the network, set  $j = j_{current}$ . Assign the fundamental diagram from  $k_{m,1}$  where links with  $n < m$ , this means to all upstream links. Then go through each link with  $n > m$  (all downstream links from  $n$ ). If there is a sensor on link  $n$  with a fundamental diagram assigned then set  $j_{current} = j + 1$ . If there is no sensor on the link, assign current fundamental diagram  $j$  to that link.

Example I: Since step I and step II already have been made, we have  $\{k_{1,j}, k_{2,1}, k_{3,j}, k_{4,2}, k_{5,j}\} = K$ . Find the first assigned link, that is  $k_{2,1}$ , set  $j = j_{current}$ . Then assign all upstream links. The result becomes  $k_{1,1}$ . After that, assign all downstream links. No sensor found on  $k_{3,j}$ , alas  $k_{3,j} = k_{3,1}$ . Go to next link. A fundamental diagram have been assigned to  $k_{4,2}$ , update  $j$ ,  $j = 1 + 1$ . Go to next link. No fundamental diagram assigned to this link, alas the last link is assigned according to  $k_{5,2}$ . For graphical representation of step III see 5.8 III.

### 5.2.2 Backward link assignment

The backward link assignment method assumes that the fundamental diagram can be inherited upstream instead of downstream.

#### Step III

The same approach is used for backward link assignment method as for the forward link assignment method. First reverse the sorted set of links, so that the sorted list becomes the following  $\{k_{n,j}, \dots, k_{1,j}\} = K$ . Since the methods propagates upstream, update  $j$  according to  $j = j - 1$ . Then use the exact same procedure as step III to assign the links.

Example: Step I and step II are finished, we have that  $\{k_{1,j}, k_{2,1}, k_{3,j}, k_{4,2}, k_{5,j}\} = K$ . Reverse the order  $\{k_{5,j}, k_{4,2}, k_{3,j}, k_{2,1}, k_{1,j}\} = K$ , then find the first link with an assigned fundamental diagram, set  $j = j_{current}$ , in this example  $j = 2$ . After that, assign the fundamental diagram to all links downstream, the result becomes  $k_{5,2}$  and  $k_{5,1}$ . After that begin assign fundamental diagrams to links upstream. The next link  $k_{3,j}$  will be  $k_{3,2}$ . Go to next link  $k_{4,1}$ , which have a fundamental diagram, therefore update according to  $j = j - 1$ . Go through rest of the links. The final result from step III is  $\{k_{1,1}, k_{2,1}, k_{3,2}, k_{4,2}, k_{5,2}\} = K$ . For the graphical representation see figure 5.8.

## 5.3 Complex method implementation

This part of the thesis describes the implementation of the complex method. The purpose with the upcoming sections is to visualize exactly what is happening in the calibration process, but with figures and words rather than formulas.

The first section is a presentation of what is referred to as ground truth followed by a section on the conceptual thought behind the implementation in order to increase the reader's awareness of the calibration process. Then the major data flow is presented to show how the different system parts interact. This system description is similar to the one initially made to the Mobile Millennium system, see figure 4.1. The major difference is the introduction of a feedback loop and an increased system boundary where an extra data source is introduced, which is used as the best possible estimate during the calibration. This section also presents simplifications made to the complex method prior to its implementation.

### 5.3.1 Ground truth

It has already been argued that a traffic model needs to be calibrated against some sets of field data, a ground truth that is treated as something like an ideal model result. The ground truth can be observations at single points in space, such as traffic flow or time mean speeds, they can also be vehicle trajectories given by traffic probes that are constantly recording their location over time or space mean speeds computed from travel time estimates. For this thesis Bluetooth readings from the highway were available. From such readings travel times on the highway as well as space mean speeds can be computed. This section touches on the subject of space mean speed estimation from Bluetooth readings.

In the first quarter of 2012 a set of Bluetooth detectors was placed along two different experimental sites (the highways that were modelled) during approximately two weeks for each site. These sites are commonly referred to as I-880 and I-15 Ontario. Bluetooth detectors of this kind have the ability to scan the Bluetooth frequency band for any transmissions and capture the MAC-address<sup>1</sup> of the unit that is transmitting as well as the time for that transmission.

Simplified, the data collection procedure can be said to work as follows. A Bluetooth detector is located at the beginning of a highway segment. It detects a passing transmitting Bluetooth unit at time  $t_1 = 0$  by identifying and storing<sup>2</sup> that unit's MAC-address, see Figure 5.9. If the same MAC-address is detected later on by another detector downstream of the first detector, and in a reasonable amount of time, say at  $t_2 = 10$  s the travel time between those detectors is  $t_2 - t_1 = 10$  s. Since the exact locations are known for both detectors the space mean speed of the Bluetooth device can be computed.

This detection process is repeated between each Bluetooth detector pair along the highway and for many different unique transmitters (vehicles) that are travelling on the highway. The travel time can then be computed as an average over a time period since many different transmitters are detected. The uncertainty in travel time can then be expressed as a variance computed around the sample mean. For this deployment the Bluetooth readings were aggregated over 1, 5 and 15-minute intervals.

The final outcome from this procedure (collecting, filtering, aggregating, transforming) is the space mean speed for different time intervals along the

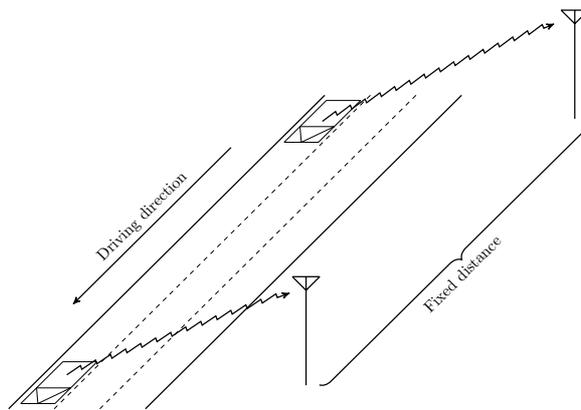
---

<sup>1</sup>A Media Access Control address (or MAC-address) is an unique label that identifies the network interface of the unit in question.

<sup>2</sup>All Bluetooth readings are uploaded to a database for filtering and travel time computations.

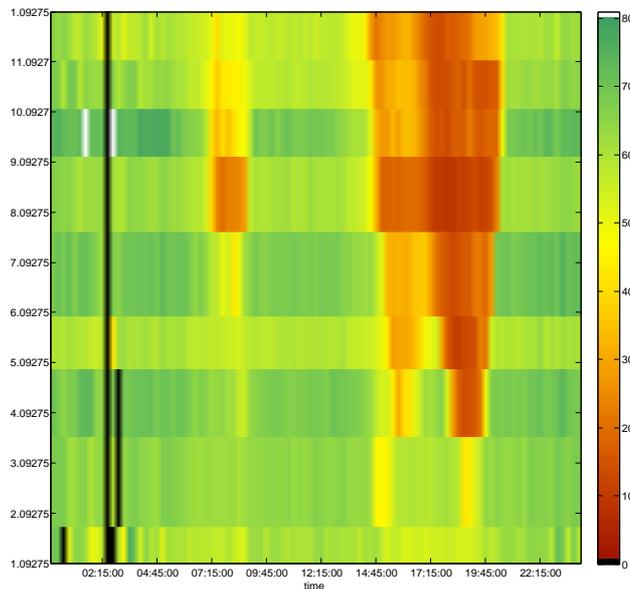
path between each Bluetooth detector pair, a path named Bluetooth route. The transformation mentioned is to compute the space mean speeds from the travel time estimations given by the sample means. This can be done given that the fixed distance between the Bluetooth readers is known.

Figure 5.10 is an example of the outcome from a complete procedure (collecting, filtering, aggregating, transforming). The figure shows the velocity field in a space-time diagram where a higher speed corresponds to a darker shade of green and a lower speed corresponds to a darker shade of red. Speeds greater than 80 mph have been mapped to white while a speed equal to zero has been mapped to black. This velocity field was estimated from Bluetooth detections.



**Figure 5.9:** Example of the ground truth data collection deployment where Bluetooth readers are located along the highway. The Bluetooth readers detects and stores MAC-addresses of Bluetooth units that passes by along with the time of that reading. The travel time and space mean speed between two readers can be computed since the location of the Bluetooth detectors are known.

Note the vertical band at 2:30AM in Figure 5.10. It is a common feature of the current deployment and is correlated with the upload of Bluetooth data to the database that stores those detections. This is bad from a calibration viewpoint but can be avoided by limiting the time frame of the calibration.



**Figure 5.10:** Example of how the ground truth is visualized in the space-time domain. The space-time diagram depicts the aggregated velocity (in mph) for I-880 northbound with aggregation period of 15 minutes. The time range for this estimation is between 2012-03-14 00:00 and 2012-03-15 00:00.

This concludes the presentation of what is regarded as ground truth. The ground truth is named as if it was an absolute, which is not the case. The exact location of the transmitter is not known upon detection of the Bluetooth broadcast, data is filtered and then aggregated as sample means. To investigate the quality of the ground truth is outside the scope of this report, it is simply argued that the filtered data can be trusted to a degree that is high enough and that any uncertainty is reflected in the variance that is computed during the aggregation of the Bluetooth travel times.

### 5.3.2 Complex method: Implementation concept

This section presents how a single iteration of the calibration loop (process) functions. The purpose of this section is also to visualize the iterative calibration process in the same way that the system user would be able to view the different steps of that process. This is done from a data processing perspective, from filtered field observations to the comparison between the model result and the ground truth.

Consider Figure 5.11, it depicts the increased level of data processing made during a single iteration of the calibration loop. The single iteration is given in the space-time perspective which is the most frequent view a user has of the different steps. Starting from the top, the different layers in Figure 5.11 depicts:

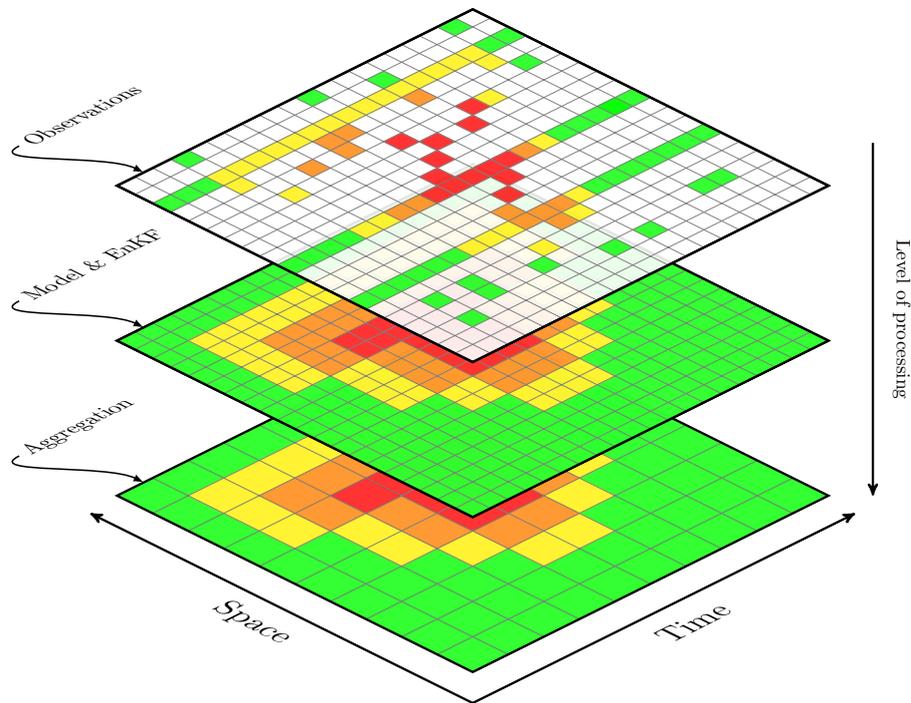
1. Filtered observations given by probes and stationary detectors,
2. the model output as a space-time diagram in relatively high resolution (e.g.  $\Delta T = 30$  s and cell lengths of a couple of hundred meters) and

3. an aggregation of the velocity field (e.g.  $\Delta T = 15$  min and longer routes) so as to bring it into the same domain as the ground truth.

Moving from measurements in the top layer to the velocity estimate given by the highway model is done by the basic formulation of the Mobile Millennium system. This is the domain where parameters have influence which makes these layers (specifically the step from the top layer to the middle one) the calibration subject. Note however that the aggregated layer of Figure 5.11 is the test subject when the model output is evaluated against the ground truth.

There is more to the aggregation step than just averaging data. During the field deployment the Bluetooth detectors were stationary with intermediate distances that far surpassed those of any cells or links of the highway model. In other words, the resolution of the data from Bluetooth detectors was lower than that of the highway model. The aggregation procedure (the step between the bottom two layers of Figure 5.11) was then to:

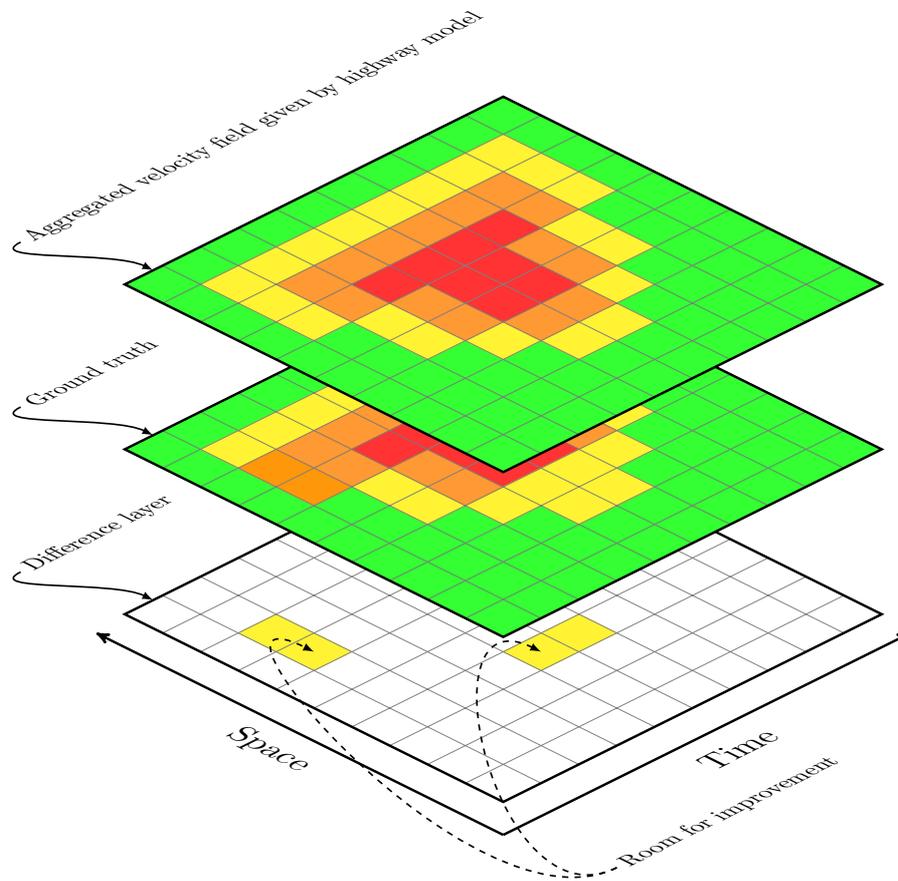
1. Estimate travel times for each link from the velocity field given by the highway model.
2. Aggregate those travel times to encompass the routes between each Bluetooth detector pair
3. Aggregate the travel times in time.



**Figure 5.11:** Depiction of the increased level of processing made during one calibration step. Top layer: Observations are available at different space-time coordinates. Middle layer: Observations are combined with model results through the ensemble Kalman filter (EnKF) to predict the velocity field. Bottom layer: The resulting velocity field is then aggregated through a travel time estimation process from which the space mean speed is determined.

The increased level of processing depicted in Figure 5.11 is done for a single parameter set suggested by the complex method. The next step of the complex method is then to evaluate the performance of that parameter set. This is made by comparing the ground truth and with the output. This is the reason for the aggregation and transformation made from the second to the third layer of Figure 5.11.

The evaluation of the current parameter set is depicted in Figure 5.12 where the bottom layer of Figure 5.11 now is the top layer (the aggregated model result for a certain parameter set). Available is also the ground truth which has the same resolution and domain as the model result, the ground truth is given in the middle layer of Figure 5.11. The difference between the two layers is then given by the bottom layer of Figure 5.12 where some room for improvement is noted (the layers did not match completely but should be better than for the previously evaluated parameter set). A performance value can be computed from this difference.



**Figure 5.12:** The final step of a single iteration of the calibration loop where the current parameter set is evaluated by comparing the aggregated model velocity field (top layer) to the ground truth (middle layer). The bottom layer symbolises the difference between the two top most layers. In this case there was some room for improvement. A performance metric can be computed that expresses this difference.

This concludes the presentation of the implementation concept. The conceptual thought of the calibration procedure was presented visually for a single calibration iteration where the model output is aggregated and brought into the same grid as the ground truth. The two different fields are then compared and from this comparison a performance metric can be computed. The less the difference between the two fields that there is, the better the parameter set and this should be reflected in the performance metric.

The evaluation of the model result against the ground truth does not have to be visualized with space-time diagrams. In the final implementation of the complex method it was time series for each Bluetooth route rather than space-time diagrams that were compared and then averaged, but these are simply different perspectives of the same data.

### 5.3.3 Complex method: Implementation data flow

This section about the complex method implementation shows the data flow between the different system parts. Recall the initial system presentation represented by figure 4.1. It described the system as it functions when the highway model is executed. Field observations are combined with a model result through an ensemble Kalman filter in order to increase model performance, assimilate data and fill in the blanks that are not given by observations. The outcome can be displayed as space-time diagrams showing the velocity field over the studied highway section.

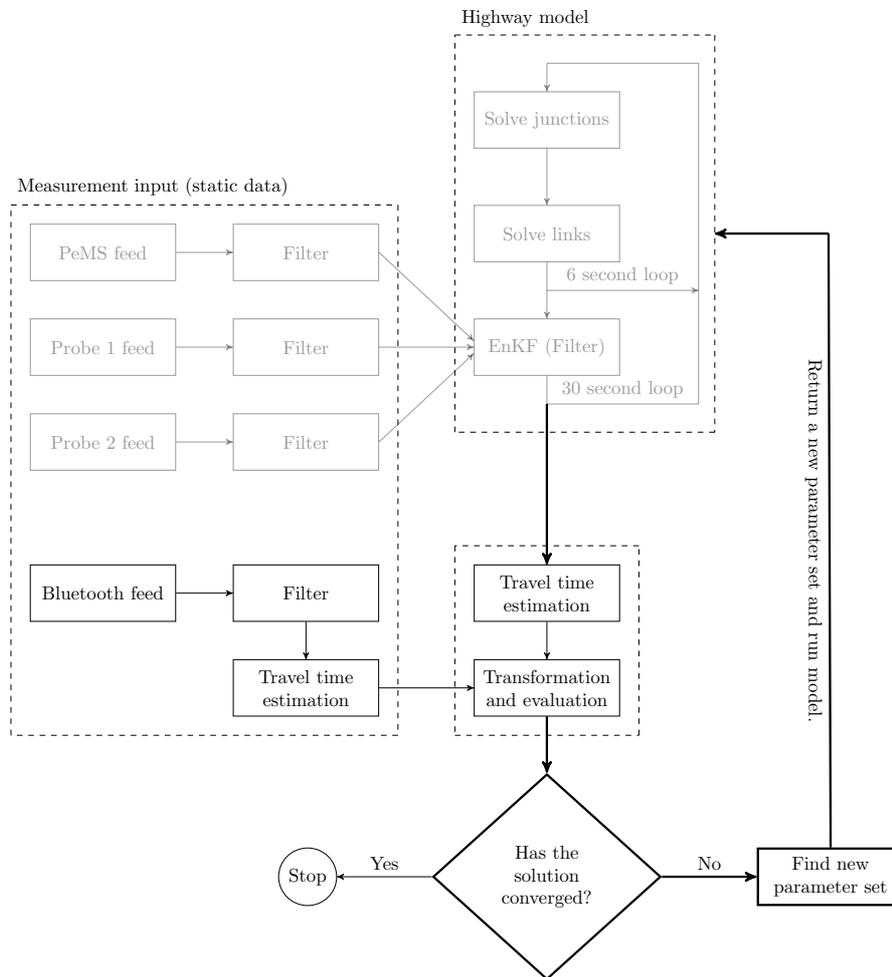
The outcome of the process depends on model parameter values, such as free-flow speeds, measurement error covariances and so forth. All such parameters were originally static, hard-coded parameters in the system code. One of the changes made in order to make both the complex method and the automatic empirical calibration procedure to function properly was then to enable the highway model to take user specified parameters.

As stated in the previous section the ground truth and the velocity field did not have the same resolution ( $\Delta T$  and  $\Delta x$ ). Another change that then had to be made was to take velocity fields given by the highway model and transform these into the same domain as ground truth.

The complex method finally had to be integrated with the new more dynamic highway model. The complex method was implemented to be a separated from any specific problem (in the current form it can solve any non-linear functions that are subject to explicit constraints). This was done by adding a calibration manager class to the calibration process, a sort of heartbeat and coordinator that calls all the other programs (highway model, travel time estimator, aggregation procedures, the complex method etc.) in the right order.

The data flow of the calibration process is shown in Figure 5.13, the primary differences between this system and the one given for the basic highway model is the feedback-loop (which is high-lighted by the thicker black borders and arrows), the travel time estimate and velocity aggregation, the comparison to ground truth and the existence of a new data source (Bluetooth). The original system parts are de-emphasized to a lighter shade of grey in the figure.

The data flow in Figure 5.13 can be summarized as follows: (1) the highway model is the centre of the system. A parameter set is fed to the model through the feedback-loop for which the model is executed, this is a trigger signal for the loop. The six-second loop is iterated five times at which point the ensemble Kalman filter brings in field measurements. After this data assimilation is the regular six-second loop invoked again. (2) when the highway model is finished travel times are estimated from the velocity field. (3) the ground truth given by Bluetooth readings are brought into the system and both of the travel time estimates are transformed to time mean speeds and compared by the calculation of a performance value. (4) a convergence criteria is evaluated, the loop is stopped if the convergence criteria (or the maximum number of iterations) is reached. (5) a new parameter set is found if the convergence criteria is not met, this new parameter set is then fed into the highway model etc.



**Figure 5.13:** The data flow of the calibration procedure. The figure includes the original system in a de-emphasised shade of gray as well as the new parts in black. The new parts include the Bluetooth feed and travel time estimation. The travel time estimation made from highway model results, a block that transforms the model results to have the same resolution as the ground truth and the complex method which is used to find new parameter values. The feedback-loop is highlighted by thicker arrows and borders.

No system architecture has been mentioned in depth thus far. It can be said that the highway model functions as in the initial version of the Mobile Millennium system. Whenever the 30-second loop is executed data is written to a database. Measurements are static and are read into the highway model from a database in the same 30-second loop. Once the highway model is finished the travel times are estimated with a process that loads the velocity field from the database, estimate travel times and then save the results to the same database. That data is then loaded and transformed to another domain and a performance metric is computed. The performance metric is stored in a database together

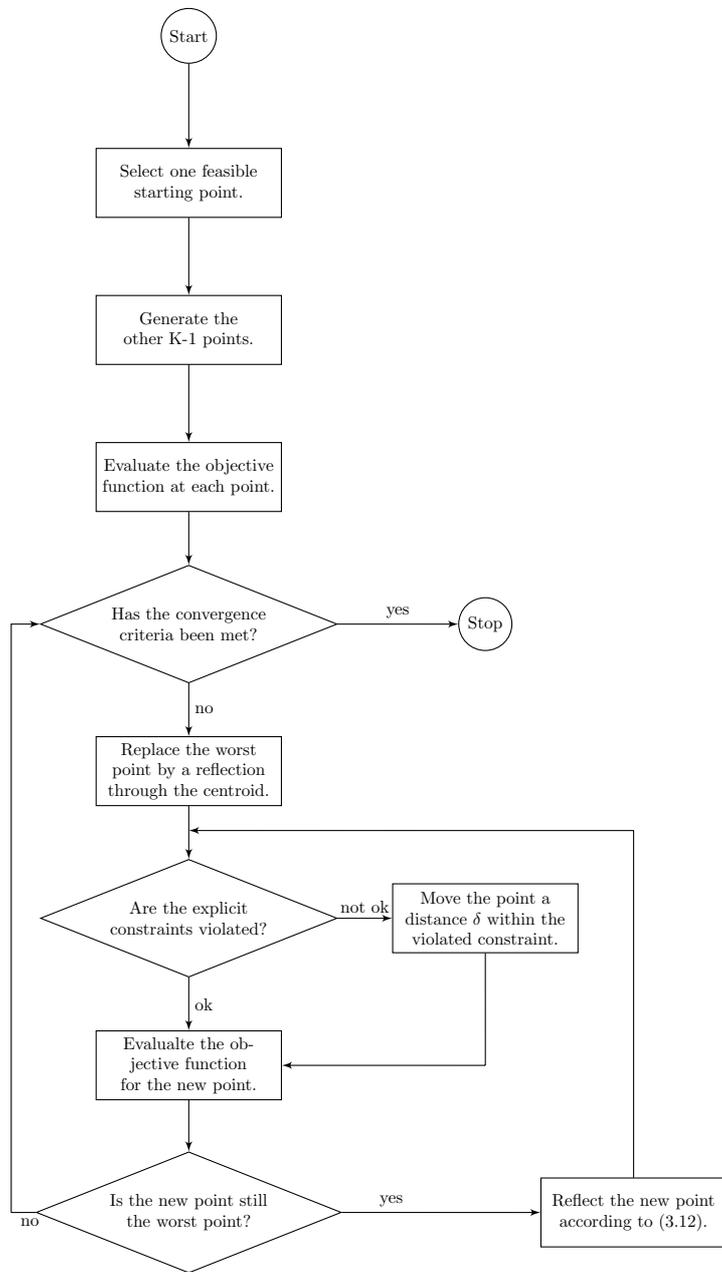
with the current parameter set. A new parameter set is chosen and written to the database, that parameter set is brought into the highway model and the process is repeated. A database is then always the communication media between the different processes depicted in Figure 5.13.

### 5.3.4 Complex method: Simplification for problems with explicit constraints

This section presents a flowchart of a modified complex method that was ultimately implemented with the Mobile Millennium system. For the this thesis, there was no need to implement a complete algorithm as it is depicted in Figure 3.7. The reason was circumstances associated with the models, which implied that no implicit in-equality constraints were present during the calibration or that such could be handled in the interface between the complex method and CTM-v model.<sup>3</sup> Also, (3.12) was chosen for expressing the movement of a repeating worst complex member. These two points change the nature of complex algorithm as it was expressed in Figure 3.7. The alternative formulation, which can be called the complex method for explicitly constrained optimization, is expressed by Figure 5.14.

---

<sup>3</sup>No parameters of the CTM-v except for the turning proportions at intersections could be the subject to implicit constraints. The special circumstances for the models used was that no intersection had more than one off-ramp. I.e. there was only need to calibrate the proportion of the vehicle stream that stayed on the highway,  $0 \leq \lambda_1 \leq 1$  from which the proportion of vehicles turning off the road,  $\lambda_2$  could be computed as  $\lambda_2 = 1 - \lambda_1$ .



**Figure 5.14:** The modified complex method, or the complex method for explicitly constrained optimization appearing as a flowchart representing how it was implemented in the Mobile Millennium system code.

## Chapter 6

# Experiments

This chapter shows the reader how the experiments are made. The first section 6.1 in this chapter introduce the test sites on which the experiments are made. Section 6.2 sheds light on the subject on how the authors evaluates which developed fundamental calibration procedure that performs the best. The next section, section 6.3 explains how the test sites are being calibrated by different frameworks. Why different frameworks are being evaluated, is due to lack of knowledge on how they perform. So to be able to present which of the framework that performs the best, all relevant ones are tested.

### 6.1 Test site introduction

This section will introduce the two different test sites that will be the test cases for the calibration frame work. The different test sites are chosen because of data availability. The date ranges for data availability is different for each test case as well. This depends on that Bluetooth data are available only during certain time periods. The differences between the test sites will be presented in detail in the sections 6.1.1 and 6.1.2.

#### 6.1.1 Test site I: Interstate 880, CA

This site is on I-880, the part of the highway that runs from Hayward down to Fremont in the Bay Area, near San Francisco. The data availability on this highway are from the sources PeMS stations, probes and Bluetooth units. Figure 6.1 shows the stretch of the highway that is modelled. A more detailed map can be found in Appendix B. The date range for this network and test site is from 2012-03-03 to 2012-03-15. This site has been a subject for research on many occasions regarding the Mobile Millennium in the past [2].

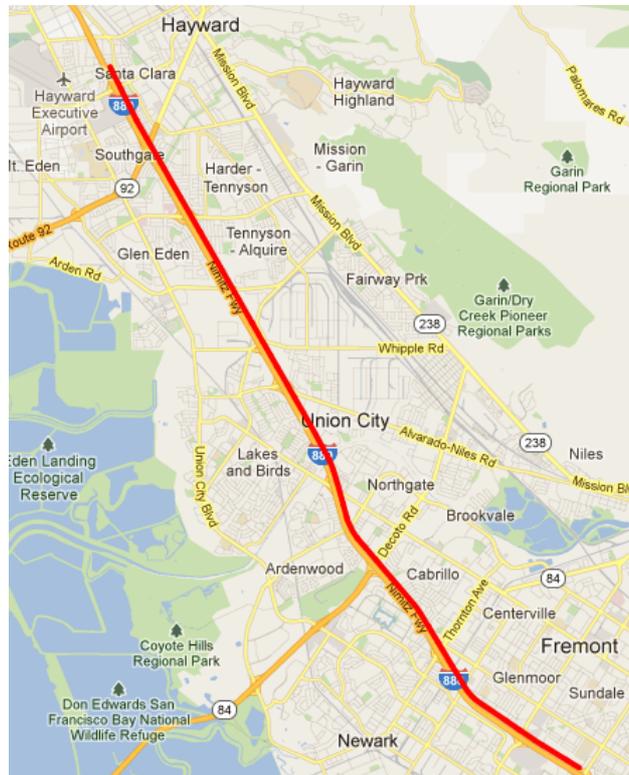


Figure 6.1: Shows the network representation of the test site I-880

### 6.1.2 Test site II: Interstate 15 Ontario

This site is placed on I-15 Ontario, north of Los Angeles, California. The data availability for this stretch of road are from the sources PeMS station, Bluetooth detectors and probes. Figure 6.1.2 shows the geographical area of the site on I-15 Ontario. The relevant date ranges for this test site is from 2012-03-31 to 2012-04-12. For a more detailed map see Appendix B. I-15 Ontario is a relatively new test site in the Mobile Millennium project, which makes it an interesting site with respect to the thesis and testing the framework.

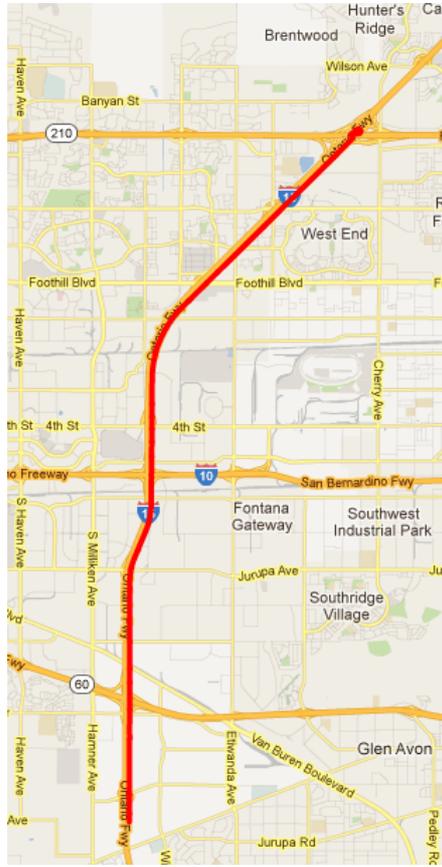


Figure 6.2: Shows the network representation of the test site I-15 Ontario

## 6.2 Experiment layout: Fundamental diagram estimation

This section explains which settings, see Table 6.1, and how the experiments for estimating which fundamental diagram and link assignment procedure that will be used for later experiments.

In order to evaluate which fundamental diagram calibration as well as link assignment procedure that performs better, the experiments were conducted with the settings presented in Table 6.1. The configure ID is the identifier for the experiment setting and applies to both I-880 and I-15 Ontario. Note that results from fundamental calibration version I, due to bad performance during the development phase, is not included in the final experiments in the thesis.

**Table 6.1:** Presents the global settings for the system when running the different experiments with parameters set from the fundamental diagram calibration procedure.

Configure ID	Link assignment procedure	Fundamental Diagram Calibration Version
1	-	Original
1001	Forward	III
1005	Backward	III
1009	Backward	II
1013	Forward	II

The experiments was conducted in the following order for all settings<sup>1</sup>, for both test sites:

1. Choose a setting from Table 6.1.
2. Calibrate according to the version in Table 6.1. Assign a unique identifier to keep track of the version and which sensor.
3. Assign the fundamental diagrams according to the link procedure for the setting in Table 6.1. Assign a unique identifier for the setting on each fundamental diagram assigned.
4. Run the system with the unique setting identifier.

To estimate the performance of a specific setting, a comparison was made between the output from the system with the setting and the ground truth. This was made using root mean squares error (RMSE) and normalized root mean squares error (NRMSE). The Root Mean Square Error NRMSE describes the difference between two different samples. RMSE is calculated using (6.1).

$$\begin{aligned}
 \text{RMSE}(\theta_1, \theta_2) &= \sqrt{\text{MSE}(\theta_1, \theta_2)} \\
 &= \sqrt{\text{E}((\theta_1 - \theta_2)^2)} \\
 &= \sqrt{\frac{\sum_{i=1}^n (x_{1,i} - x_{2,i})^2}{n}}
 \end{aligned} \tag{6.1}$$

where  $\theta_1$  is a vector of samples and  $\theta_2$  is the comparison vector (in this case the ground truth). The Normalized Root Mean Square Error (NRMSE) instead of describing the distance, it describes the percentage of error. The NRMSE is calculated using (6.2).

$$\text{NRMSE} = \frac{\text{RMSE}}{\text{mean}(\theta_2)} \tag{6.2}$$

For results see 7.1.

### 6.3 Experiment layout: Automated calibration

This section presents three different method layouts that were followed during the calibration of the different test subjects, or frameworks for calibration of

<sup>1</sup>The configure ID is the unique identifier for each setting.

traffic models. Common for all three of these methods are that calibration is followed by validation, where the estimated parameters are evaluated for another time period than the calibration time period.

### 6.3.1 Local and source alternating calibration framework using standard fundamental diagrams

The first method, or framework, that was proposed is one where the model's output was observed (validated) when default settings were used and where parameters associated with non-valid areas were adjusted with an automated algorithm, i.e. the Complex method (calibrated).

An initial setting was also that only stationary detectors were used as input to the model for the initial validation and calibration. The motivation for this approach, was that since stationary detectors have a fixed coverage<sup>2</sup>, we wish to find a reasonable operating point (parameter set) that enables the model to capture the conditions using data sources are better guaranteed over time and space. Note that all model parameters are static over time, which means that the flow at on-ramps are consistent over time and that the proportion of vehicles always leaves the highway at node  $x$  are consistent and so on. Static inflows and turning proportions would imply that the model would be static if it would be invoked without any sources at all. More specifically would this imply a completely uncongested scenario with default settings. The supply flow of off-ramps were left to have a big capacity. The hypothesis was that the adjustment of the turning proportions could compensate for lack of capacity<sup>3</sup>.

Adding stationary detectors, such as the PeMS inductive loop system, as a data source introduces congestion in the model's output if low speeds are detected. But sparse stationary detectors might not be able to induce a system-wide congestion if the model is biased towards free-flow conditions. The initial calibration effort was therefore made so that it enabled the possibility to see if the model's parameters could be tuned in a way that induced congestion were there no stationary detectors were present.

The final steps of this calibration framework included the addition of probes as a data source in order to make the final calibration of observation noise parameters and validate the calibrated model during the best of circumstances.

This framework was followed when an I-880 model was calibrated and validated. The calibration was a special case given that some hands-on parameter estimation had been done already [2], i.e. the first step of this process was validation of the current parameter set. The outcome of the validation was deemed unsatisfactory on some local parts. The default parameter set was set as the initial point in the parameter space once the calibration commenced. The calibration and validation of this model was made for two different but very congested days, the 9th and 14th of March, 2012, between 12PM and 9PM on both days.

To be more specific on the changes made with the default settings during the calibration: In the default setting and using only stationary detectors the congestion on Bluetooth route 3 was deemed to be too high, which was indicated

---

<sup>2</sup>In space, but not always in time, this due to unforeseen downtime.

<sup>3</sup>I.e. if the capacity on the off-ramp really is low then a low proportion should leave the highway and this should only be noticeable during the hours of high demand.

that the inflow was too high at the end of this route. The parameter stating the static inflow from Decoto Road was set as variable during the calibration effort, see figure B.2 in Appendix B. Focusing further on route 3, it was possible that the overestimated congestion was due to an excess of vehicles driving on the highway. The split ratio at the intersection between I-880 and Decoto Road was therefore set as variable during the new calibration effort.

Too little congestion on route 8 indicated that too little inflow in the vicinity of that route. The static parameter stating the inflow parameter from Jackson Street was set as variable during the calibration effort. Too little congestion on route 9 might indicate that the supply on the network border was too great. The parameter indicating the capacity of the end of the mainline was not set to be variable during the calibration, this to avoid a constant bottleneck at that location. Physically speaking there is an on-ramp leading onto the modelled highway just after the endpoint of the model. The inflow from this on-ramp is not modelled since it lies outside the physical domain of the model.

The validation of the calibrated model revealed a persistent lack of congestion for routes 7-9 and to much of it on route 3, even though some positive changes were observable. Since the model did not pass the validation it could be rolled back to calibration but since probe measurements are available for this highway stretch were such measurements introduced instead so as to see if they had any positive influence on the model's result. The introduction of probes invoked a second calibration phase where EnKF-related parameters were set as variable<sup>4</sup>. To summarize, this method, or framework consisted of the following steps:

1. Run the highway model on the validation date using the default parameters using only stationary detectors.
2. Validate the results, check if the model is able to capture congestion and if it does so to an acceptable degree.
3. Calibrate chosen parameters, more specifically such parameters that might increase (decrease) (un)wanted congestion.
4. Validate the results again, check if the model is able to capture congestion and if it does so to an acceptable degree.
5. Be able to state that the model is incapable of capturing all of the congestion without probes. Include probes as a data source and calibrate observation noise parameters.
6. Validate the calibrated model by running it for another congested day and by comparing the aggregation field and travel time estimates on a route-by-route basis against the ground truth.

The initial validation results, the initial calibration and validation results and the final validation results are given in the next chapter.

---

<sup>4</sup>Not the observation noise mean or the model noise mean though, those were all set to be equal to 0 while the standard deviations were subject to calibration.

### 6.3.2 Global and type-differentiated calibration framework using estimated fundamental diagrams

The second framework that was tested was done so together with estimated fundamental diagrams. The parameter types were differentiated by type during the calibration effort, which was done step-wise. I.e. source parameters were calibrated before sink parameters. In addition, for this framework it was suggested that source parameters were calibrated during free flow conditions only so as to find a reasonable operating point. Just as in the previous framework proposed, only stationary detectors were used in the initial calibration effort.

The motivation for this type of method was, that for sites where congestion is common in time as well as space and where stationary detectors and probes are abundant, should such observers provide the input needed for the model to capture the congestion. The underlying model can then be tuned so as to function well in the non-congested states and provide a good operating point from which the state moves into the congested one.

The introduction of new fundamental diagrams changes the model's output. Not only can a new fundamental diagram impose changes in capacity from link to link, which could indicate new bottlenecks or reduce the effect of others, it also means changed free flow speeds, shock wave speeds and changed maximum densities.

This framework was tested for another run for the I-880 network where estimated fundamental diagrams were used. The calibration and validation dates were the same as for the other I-880 model. All source parameters<sup>5</sup> were calibrated during a short free flow period so as to get a reasonable estimate of the travel time during these conditions. The standard deviations of the model error noise and stationary detector observation noise were set as variable as well. Only stationary detectors were used during this calibration step. All split ratios took the values of the default settings<sup>6</sup>.

When the calibration of the sources was done, split ratios were added to the calibration list. Using only stationary detectors, the split ratios were calibrated for the 14th of March, 2012 between 12PM and 9PM. Then probes were added as a data source and the observation noise for these probes was set to variable during the calibration. To summarize, this method, or framework consisted of the following steps:

1. Estimate fundamental diagrams from stationary detector data and assign such values for each link.
2. Choose a short free-flow time period and calibrate source parameters and EnKF-noise parameters for this time period.
3. Add split ratios to the calibration list and leave the now calibrated source parameters and EnKF-noise parameters as static. Extend the calibration time period to encompass a congested state.
4. Introduce probes as a data source and add that observation noise standard deviation as variable during the calibration. Leave the now calibrated split

---

<sup>5</sup>The inflows at the start of the modelled highway and at on-ramps are regarded as source parameters.

<sup>6</sup>Meaning that 100 % of all vehicles on the highway were assumed to stay on the highway when they reached an intersection.

ratios as static.

5. Validate the calibrated model by running it for another congested day and by comparing the aggregation field and travel time estimates on a route-by-route basis against the ground truth.

All in all this is a divide and conquer concept. Having a subset of parameters variable during the calibration should make the effort more controllable and the solution should converge quicker. Calibrating over a short time period also decreases the calibration time since the time of a model run often decides the calibration time. Stationary detectors were used first of the same reasons as mention in the previous framework.

### 6.3.3 Global calibration framework using estimated fundamental diagrams

In some cases the model users might like to deploy a highway model similar to the Mobile Millennium model for areas where stationary detectors are rare, have low coverage, or low reliability and so on. For such circumstances a framework proposed where all model parameters are calibrated in one go and for a time period that includes the transition from free-flow to the congested regime and then back again. If data of congestion is available for many days separate from the calibration date could the developer choose to validate the calibrated model for several days worth of data.

The motivation is that a lack of stationary detector coverage leaves the model operator with no choice than to base the model's validity on the presence of probes that can sense the state of the road. I.e. there is no reason to leave these probes out of the initial calibration effort. This proposed framework is also motivated by automation: Would one be able to produce a valid model by simply setting all model parameters as variable during the calibration effort.

The calibration effort of I-15 near Ontario, CA followed this framework. The modelled highway stretches around 11 miles (almost 18 km), it intersects with three other major roads (M. Baker Memorial Fwy, San Bernardino Fwy, Foothill Fwy) and several other roads. The ground truth was unfortunately only available from seven (and not nine) Bluetooth routes, the comparison between the model output and the ground truth was then done for a subset of the model's edges.

Every turning proportion parameter were set as variable during the calibration if an off-ramp was connected to that node. Similarly, every source (on-ramp flow) parameters at the intersections were also set as variable.

The calibration of the I-15 model was done for April 4, 2012 between 10AM and 8PM since congestion was noted during this time period. Given all sources and split ratios as well as observation noise parameters that were subject to the calibration, were a total of 38 parameters subjects to this effort. This approach is different to the ones that were done for I-880 and demanded a larger number of model iterations.

For validation of these results the model was run for another time period than the one used during the calibration. The 3rd and 5th of April showed signs of congestion and were chosen for validation purposes.

This calibration framework can be summarized to the following steps:

1. Estimate fundamental diagrams from stationary detector measurements and assign the estimated values to the model's links.
2. Introduce both probes and stationary detectors as model inputs, set all noise parameters as variable during the calibration and add all source and split ratio parameters to this calibration list. Calibrate the model for a day of significant congestion.
3. Check results to see if any parameters were pushed against their boundaries (a valid point for all calibration efforts).
4. Recalibrate such parameters that were deemed close to their boundaries.
5. Validate the calibrated model by running it for two other congested days and comparing the aggregation field and travel time estimates on a route-by-route basis against the ground truth.

The map outlining the Bluetooth detector locations for I-15 in Appendix B will indicate the presence of ten detectors and therefore nine Bluetooth routes. Unfortunately were only seven Bluetooth routes available during the calibration of the I-15 model, meaning that the model's performance was not evaluated over the first and last Bluetooth route.

# Chapter 7

## Results

This chapter presents the results from all relevant experiments for this thesis. The outline for this chapter is first the results from the fundamental diagram calibration and link assignment procedures. In the next section, the results from running the system with the calibration frameworks are presented.

### 7.1 Fundamental Diagram Calibration Results

This section presents all the results from the fundamental calibration and link assignment procedure evaluation. These results are the foundation on which fundamental calibration version that will be chosen for the next step in the calibration framework.

**Table 7.1:** This table shows the RMSE for I-880 when comparing the modelled output with the ground truth. The unit in the table is in meter/second.

Day	CID 1	CID 1001	CID 1005	CID 1009	CID 1013
2012-03-03 07:00-23:00	2.01	3.06	3.16	2.79	3.79
2012-03-04 07:00-23:00	2.29	3.43	3.48	3.07	3.99
2012-03-05 07:00-23:00	2.80	3.32	3.31	3.17	4.30
2012-03-06 07:00-23:00	3.29	3.51	3.50	3.66	4.41
2012-03-07 07:00-23:00	3.61	4.08	3.86	3.82	4.71
2012-03-08 07:00-23:00	3.48	3.69	3.63	3.54	4.67
2012-03-09 07:00-23:00	3.90	4.20	4.09	4.07	4.96
2012-03-10 07:00-23:00	2.39	2.67	2.96	3.07	3.93
2012-03-11 07:00-23:00	2.68	2.94	3.40	3.33	4.11
2012-03-12 07:00-23:00	2.92	3.28	3.51	3.44	4.50
2012-03-13 07:00-23:00	3.59	4.11	3.89	3.81	4.88
2012-03-14 07:00-23:00	3.84	6.75	4.24	4.27	5.21
2012-03-15 07:00-23:00	3.15	3.36	3.49	3.45	4.43
MAX	3.90	6.75	4.24	4.27	5.21
MIN	2.01	2.67	2.96	2.79	3.79
MEAN 3.07	3.72	3.57	3.50	4.45	
STDEV	0.61	1.02	0.36	0.43	0.43

**Table 7.2:** This table shows the NRMSE for I-880 when comparing the modelled output with the ground truth.

<b>Day</b>	<b>CID 1</b>	<b>CID 1001</b>	<b>CID 1005</b>	<b>CID 1009</b>	<b>CID 1013</b>
2012-03-03 07:00-23:00	0.07	0.11	0.11	0.10	0.13
2012-03-04 07:00-23:00	0.08	0.12	0.12	0.10	0.14
2012-03-05 07:00-23:00	0.11	0.13	0.13	0.12	0.16
2012-03-06 07:00-23:00	0.13	0.14	0.14	0.15	0.18
2012-03-07 07:00-23:00	0.15	0.17	0.16	0.16	0.20
2012-03-08 07:00-23:00	0.14	0.14	0.14	0.14	0.18
2012-03-09 07:00-23:00	0.17	0.18	0.18	0.18	0.22
2012-03-10 07:00-23:00	0.08	0.09	0.10	0.11	0.14
2012-03-11 07:00-23:00	0.09	0.10	0.11	0.11	0.14
2012-03-12 07:00-23:00	0.11	0.12	0.13	0.13	0.17
2012-03-13 07:00-23:00	0.15	0.17	0.16	0.16	0.20
2012-03-14 07:00-23:00	0.17	0.30	0.19	0.19	0.23
2012-03-15 07:00-23:00	0.12	0.13	0.13	0.13	0.17
MAX	0.17	0.30	0.19	0.19	0.23
MIN	0.07	0.09	0.10	0.10	0.13
MEAN	0.12	0.15	0.14	0.13	0.17
STDEV	0.03	0.05	0.03	0.03	0.03

**Table 7.3:** This table shows the RMSE for I-15 Ontario when comparing the modelled output with the ground truth. The unit in the table is in meter/second.

<b>Day</b>	<b>CID 1</b>	<b>CID 1001</b>	<b>CID 1005</b>	<b>CID 1009</b>	<b>CID 1013</b>
2012-03-31 07:00-23:00	3.14	2.52	2.81	4.04	4.75
2012-04-01 07:00-23:00	3.65	2.90	3.14	4.67	5.37
2012-04-02 07:00-23:00	4.66	4.77	4.69	4.89	4.87
2012-04-03 07:00-23:00	5.00	5.32	5.11	5.03	5.10
2012-04-04 07:00-23:00	5.72	5.96	5.92	5.57	5.45
2012-04-05 07:00-23:00	5.30	5.68	5.29	5.32	5.33
2012-04-06 07:00-23:00	5.68	6.03	5.67	5.66	5.53
2012-04-07 07:00-23:00	3.47	2.74	3.21	4.17	5.02
2012-04-08 07:00-23:00	3.47	2.15	3.03	4.39	5.37
2012-04-09 07:00-23:00	4.92	5.09	5.09	4.74	4.98
2012-04-10 07:00-23:00	4.66	4.85	4.75	4.61	4.85
2012-04-11 07:00-23:00	3.58	3.85	3.72	3.97	4.35
2012-04-12 07:00-23:00	4.11	4.13	4.19	4.11	4.64
MAX	5.72	6.03	5.92	5.66	5.53
MIN	3.14	2.15	2.81	3.97	4.35
MEAN	4.41	4.32	4.36	4.71	5.05
STDEV	0.89	1.36	1.07	0.57	0.35

**Table 7.4:** This table shows the RMSE for I-15 Ontario when comparing the modelled output with the ground truth.

<b>Day</b>	<b>CID 1</b>	<b>CID 1001</b>	<b>CID 1005</b>	<b>CID 1009</b>	<b>CID 1013</b>
2012-03-31 07:00-23:00	0.10	0.08	0.09	0.13	0.16
2012-04-01 07:00-23:00	0.12	0.09	0.10	0.15	0.17
2012-04-02 07:00-23:00	0.17	0.17	0.17	0.18	0.17
2012-04-03 07:00-23:00	0.18	0.19	0.19	0.18	0.19
2012-04-04 07:00-23:00	0.22	0.22	0.22	0.21	0.21
2012-04-05 07:00-23:00	0.20	0.21	0.20	0.20	0.20
2012-04-06 07:00-23:00	0.21	0.23	0.21	0.21	0.21
2012-04-07 07:00-23:00	0.11	0.09	0.11	0.14	0.17
2012-04-08 07:00-23:00	0.11	0.07	0.10	0.14	0.17
2012-04-09 07:00-23:00	0.18	0.19	0.19	0.17	0.18
2012-04-10 07:00-23:00	0.17	0.17	0.17	0.17	0.18
2012-04-11 07:00-23:00	0.13	0.13	0.13	0.14	0.15
2012-04-12 07:00-23:00	0.15	0.15	0.15	0.15	0.17
MAX	0.22	0.23	0.22	0.21	0.21
MIN	0.10	0.07	0.09	0.13	0.15
MEAN	0.16	0.15	0.15	0.17	0.18
STDEV	0.04	0.06	0.05	0.03	0.02

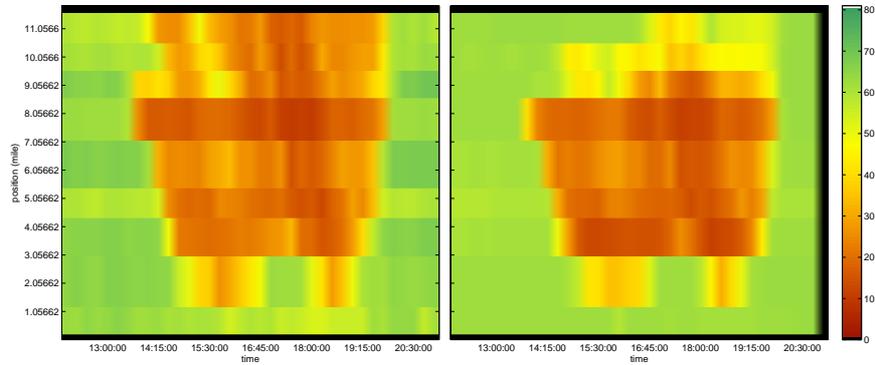
## 7.2 Complex method calibration results

This section presents the calibration and validation results for the three framework implementations. The first section presents the results for the local and source alternating calibration framework for the I-880 model, where standard fundamental diagrams were used. This section is followed by the results of the I-880 calibration, where the global and type-differentiated calibration framework with estimated fundamental diagrams were used. Finally are the results for the I-15 model presented.

### 7.2.1 Local and source alternating calibration framework using standard fundamental diagrams: Interstate 880, CA

This section provides the calibration results for the I-880 northbound model, a model capturing an 11 mile (approx. 17.7 km) highway stretch near Union City, CA, with eleven highway intersections.

Figure 7.1 shows the initial assessment of the model's performance, i.e. this perspective is part of the initial validation. The left hand side of Figure 7.1 is the ground truth space mean speed deduced from Bluetooth travel time measurements while the right hand side is the aggregated model output. Again, only stationary detectors were used as input to the model in this initial step. The space mean speeds that are depicted in the figure were given by dividing route lengths by the measured/estimated travel time.



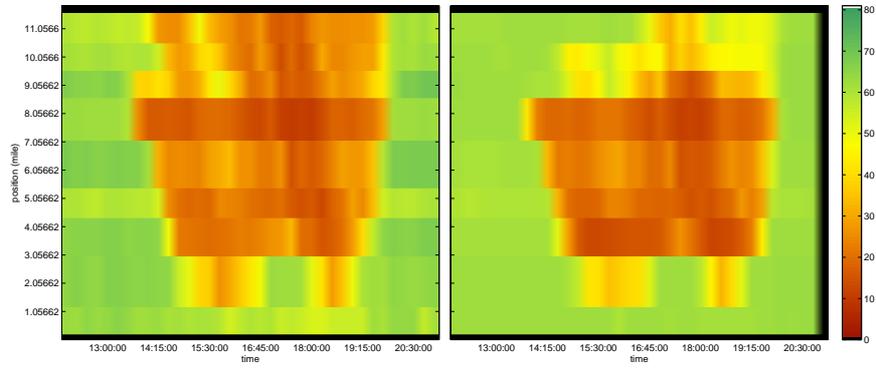
**Figure 7.1:** Ground truth (left) and the initial model prediction (right) as aggregated velocity fields in miles per hour (as a reference, 80 mph  $\approx$  129 km/h, 40 mph  $\approx$  64 km/h) for I-880 northbound on March 9, 2012. Only stationary detectors were used during the model run. The Bluetooth routes are clearly visible as horizontal stripes in both of the plots.

Nine horizontal bands are apparent in Figure 7.1 which stretches a fixed distance. These are the Bluetooth routes, and are colour-coded so as to give the average space mean speed between each Bluetooth detector during 15-minute periods, either deduced from travel time measurements (ground truth) or from the model output (the right hand figure).

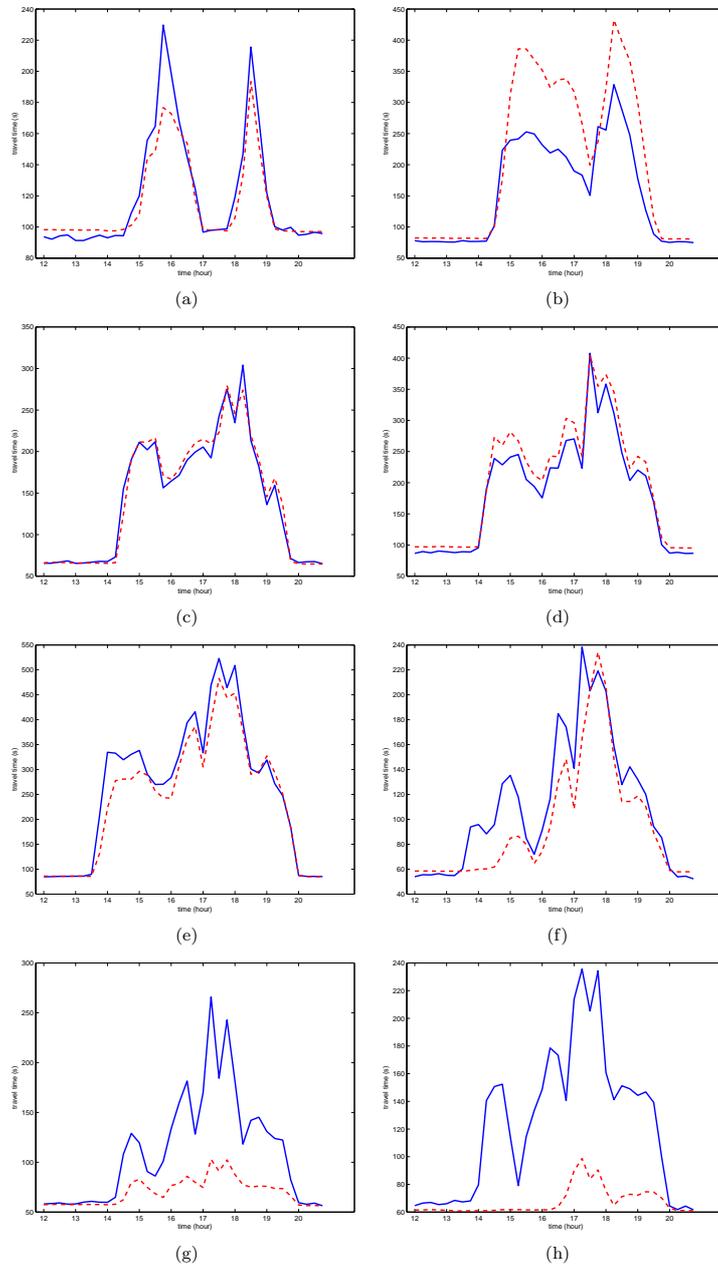
The data in Figure 7.1 can be viewed in another perspective. Consider Figure 7.3, it depicts the travel times for I-800 NB routes 2-9<sup>1</sup>. The blue lines indicate the ground truth travel time and the dashed red line is the model prediction of the travel time for that route. The x-axis (time of day) has the same scale in all figures, note however that the scale of the y-axis (travel time in seconds) alternates between the figures.

As was indicated in section 6.3.1 there were three parameters in total in the calibration effort: The source flow from Jackson Street and Decoto Road as well as the proportion of vehicles leaving for Decoto Road. The complex had not converged after 30 iterations. The calibrated model was then validated for another time period than the calibration time period. Again, March 9 was used for this purpose. The aggregated model output is depicted by Figure 7.2. The travel time estimates are given on a route by route basis in Figure 7.4.

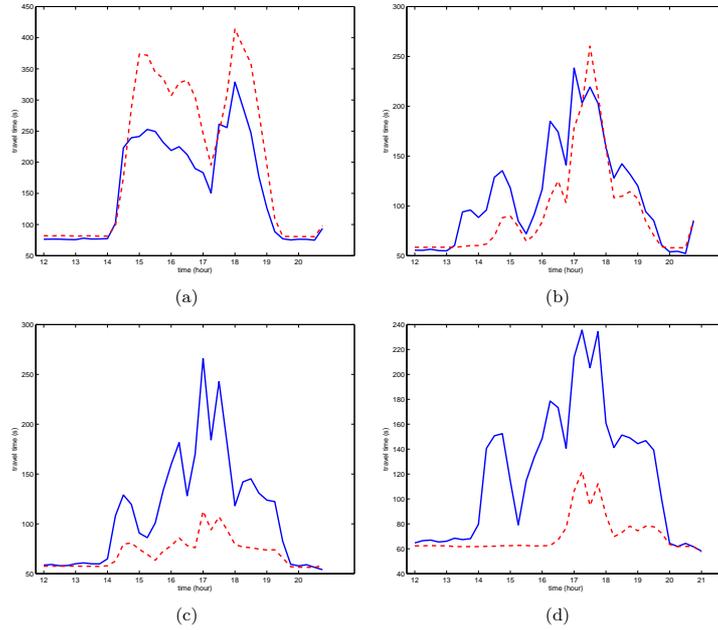
<sup>1</sup>Route 1 never showed any sign of major congestion and it was therefore effortless to predict the travel time for that route.



**Figure 7.2:** Validation of calibration effort. Ground truth (left) is compared to the aggregated model output (right) for I-880 northbound on March 9, 2012. Only stationary detectors were used during the model run. The model output does not agree with the ground truth on some routes.



**Figure 7.3:** Model predictions of the travel time in dashed red and ground truth (Bluetooth) in solid blue for each I-880 route 2 to 9 (a-h). These predictions were done for I-880 on March 9, 2012 between 12PM and 9PM. Only stationary detectors were used during the model run. No calibrated parameters were used, only standard values.

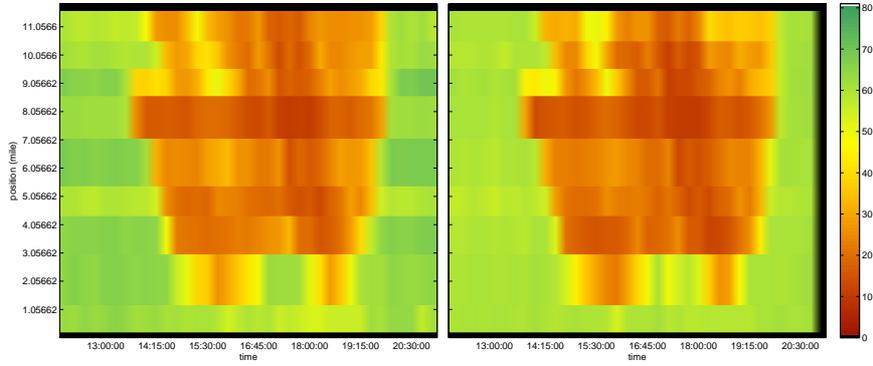


**Figure 7.4:** Validation of model predictions of the travel time in dashed red and ground truth (Bluetooth) in solid blue for each I-880 route 3 (a) and routes 7-9 (b-d). These predictions were done for I-880 on March 9, 2012 between 12PM and 9PM. Only stationary detectors were used during the model run. Only a subset of the modelled routes are shown in the figure.

As a final step, probes were included as a data source and the observation noise standard deviation of this data source was set as variable during the final calibration which was made during March 14, 2012 between 12PM and 9PM.

Figure 7.5 depicts the aggregated velocity field for March 9, 2012, i.e. the validation results for the calibrated model. Table 7.5 gives a quantitative measure<sup>2</sup> of the calibration result on a route basis and for the entire model for both the validation date and calibration date. Figure 7.6 depicts the travel times for I-800 NB routes 2-9 (a-h) for March 9, 2012.

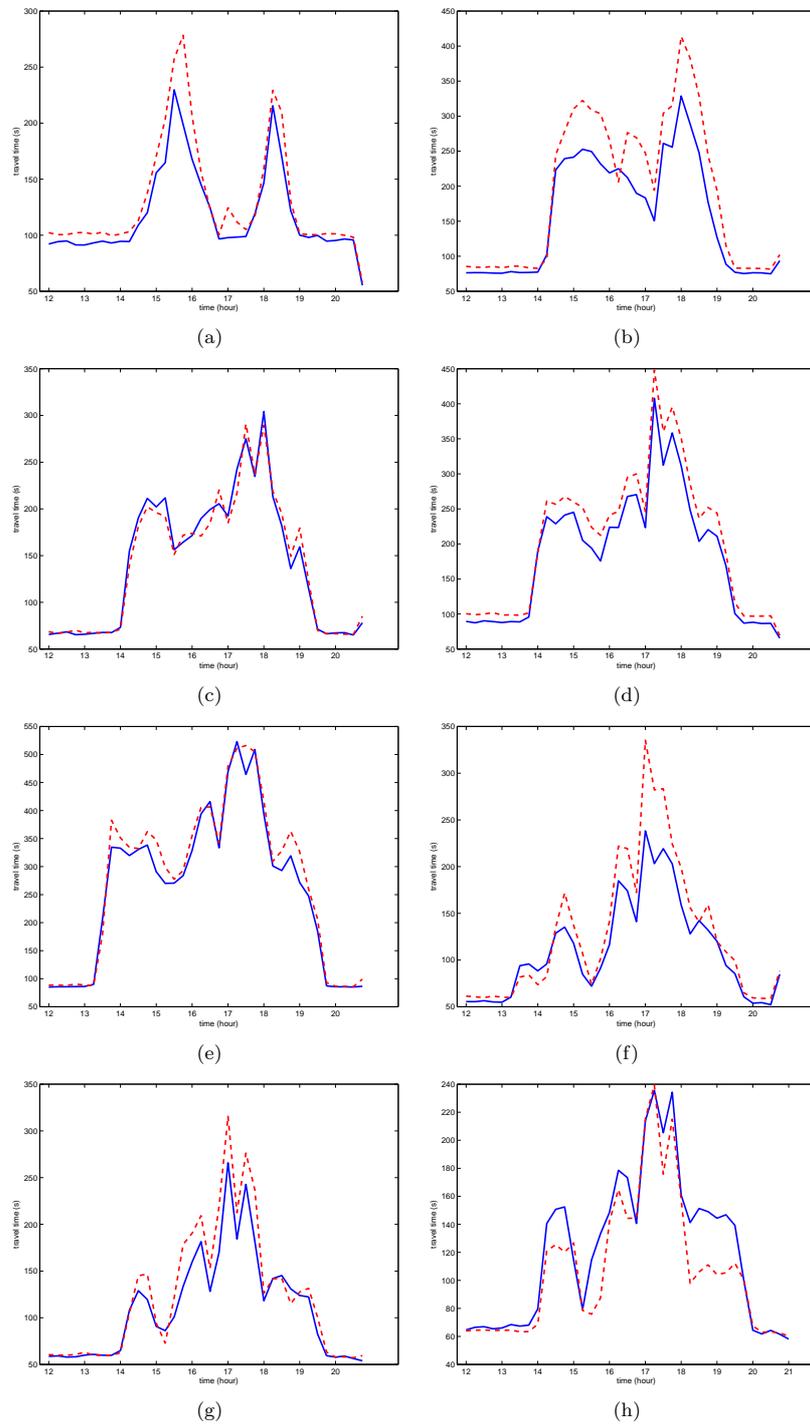
<sup>2</sup>The mean absolute percent error (MAPE) that was introduced earlier.



**Figure 7.5:** Validation of the final calibration effort. Ground truth (left) is compared to the aggregated model output (right) for I-880 northbound on March 9, 2012. Stationary detectors and probes were used during this model run.

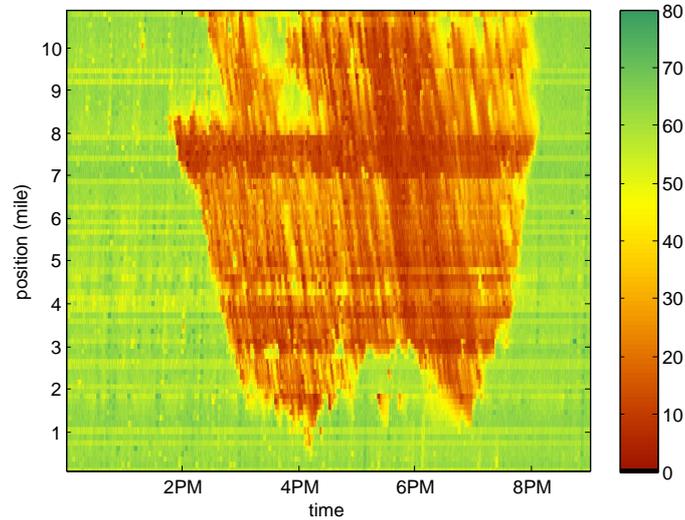
**Table 7.5:** The calibrated model’s performance in relation to Bluetooth ground truth on a route-by-route basis for the validation date March 9, 2012 and the calibration date March 14, 2012. Both stationary detectors and probes of type B were used during these model runs. A value within parenthesis gives the performance prior to the calibration effort.

Route	MAPE (%) 9th of March	MAPE (%) 14th of March
1 (-)	4.1	5.3
2 (a)	9.7	8.6
3 (b)	20.0	13.0
4 (c)	4.7	6.5
5 (d)	12.0	10.0
6 (e)	6.3	8.3
7 (f)	15.0	17.0
8 (g)	10.0	11.0
9 (h)	11.0	14.0
<b>Model</b>	<b>9.5 (9.7)</b>	<b>10.1 (11.9)</b>



**Figure 7.6:** Validation the final model's travel time predictions in dashed red and ground truth (Bluetooth) in solid blue for each I-880 route 2-9 (a-h). These predictions were done for I-880 on March 9, 2012 between 12PM and 9PM. Both stationary detectors and probes were used during the model run.

As a final validation perspective of the model is the velocity field (given in miles per hour) in Figure 7.7 considered. Note that this is the actual model output for the 9th of March prior to aggregation.



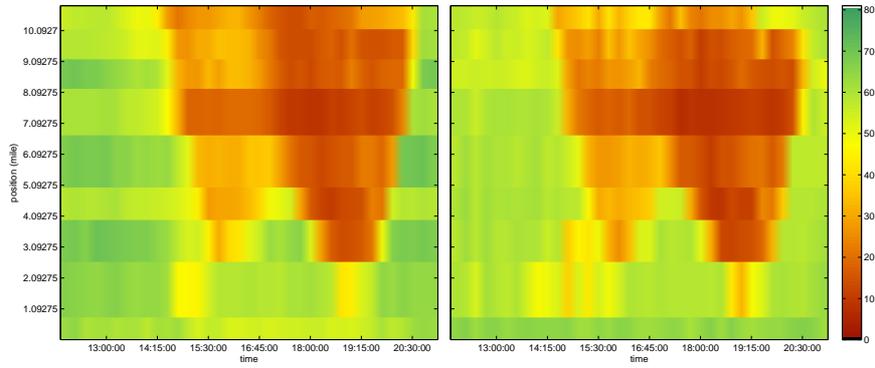
**Figure 7.7:** Qualitative validation of the actual (non-aggregated) I-880 model output, a velocity field given in miles per hour on March 9, 2012 between 12PM and 9PM. Both stationary detectors and probes were used during the model run. Standard fundamental diagrams were used with the calibrated parameters.

## 7.2.2 Global and type-differentiated calibration framework using estimated fundamental diagrams: Interstate 880, CA

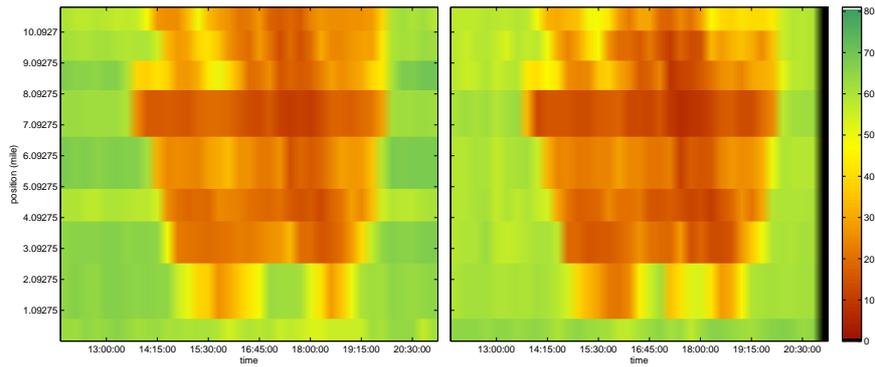
The final calibration result for the actual calibration date is depicted in Figure 7.8 as the aggregated model output on the right hand side versus the ground truth on the left hand side.

Figure 7.9 shows the validation results, i.e. the aggregated model output for March 9, 2012 using the calibrated parameter set. This is the outcome of the process that was proposed in Section 6.3.2. Travel time graphs are given in Figure 7.10 for the same date, time and aggregation period. Table 7.6 shows the mean absolute percentage errors for each route on both of the calibration and validation dates. The values within parenthesis in the same table are the MAPE-metrics for the I-880 model that was calibrated using standard fundamental diagrams and that was presented in the previous section.

Finally there is the actual model output for the validation date, i.e. the velocity field for I-880 northbound as predicted by the calibrated model using estimated fundamental diagrams and calibrated parameters. This actual model output, the velocity field is given in Figure 7.11.



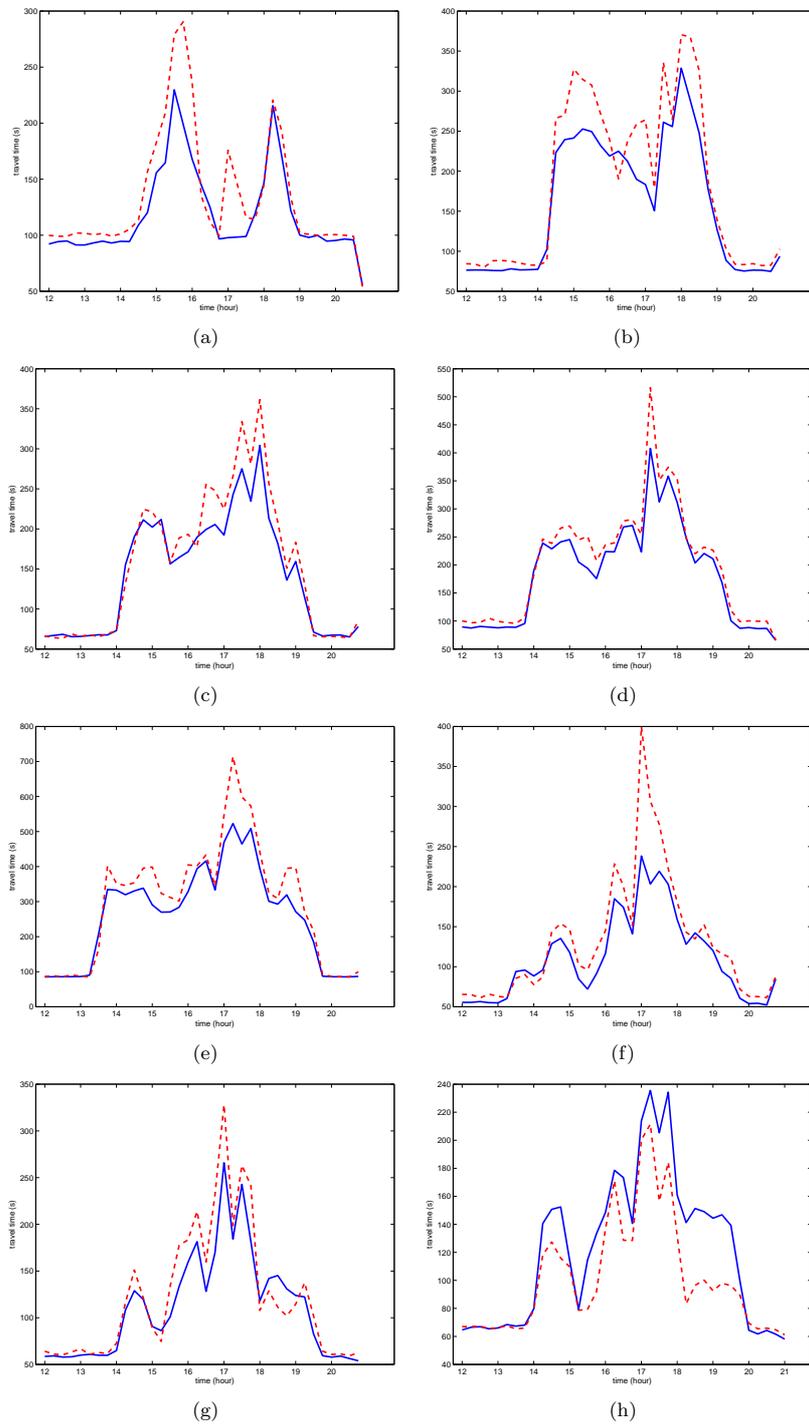
**Figure 7.8:** Outcome of the calibration effort. Ground truth (left) is compared to the aggregated model output (right) for I-880 northbound on March 14, 2012. Stationary detectors and probes were used during this calibration run together with estimated fundamental diagrams.



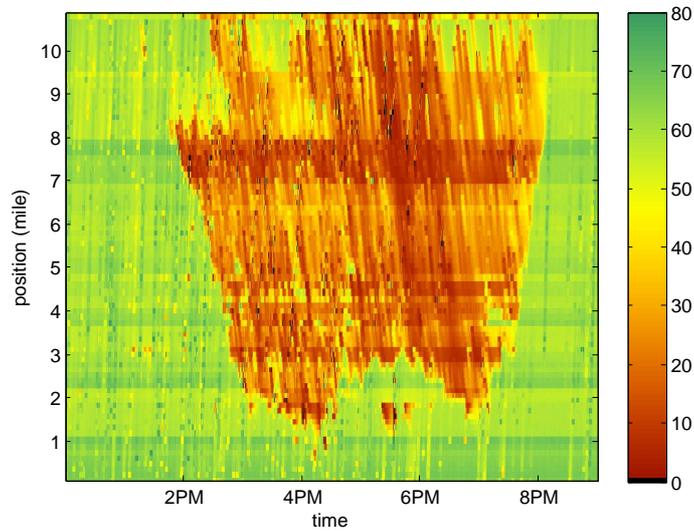
**Figure 7.9:** Validation results for the I-880 model. Ground truth (left) is compared to the aggregated model output (right) for I-880 northbound on March 9, 2012. Stationary detectors and probes were used during this calibration run together with estimated fundamental diagrams.

**Table 7.6:** Calibrated model performance in relation to Bluetooth ground truth on a route-by-route basis for the validation date March 9, 2012 and the calibration date March 14, 2012. Both stationary detectors and probes of type B were used during these model runs. The corresponding metrics for the previous I-880 model, where standard fundamental diagrams were used are included in parenthesis.

<b>Route</b>	<b>MAPE (%) March 9</b>	<b>MAPE (%) March 14</b>
1 (-)	5.3 (4.1)	10.0 (5.3)
2 (a)	14.0 (9.7)	11.0 (8.6)
3 (b)	16.0 (20.0)	15.0 (13.0)
4 (c)	9.4 (4.7)	7.4 (6.5)
5 (d)	11.0 (12.0)	11.0 (10.0)
6 (e)	12.0 (6.3)	15.0 (8.3)
7 (f)	18.0 (15.0)	19.0 (17.0)
8 (g)	13.0 (10.0)	12.0 (11.0)
9 (h)	14.0 (11.0)	21.0 (14.0)
<b>Model</b>	<b>11.6 (9.5)</b>	<b>13.4 (10.1)</b>



**Figure 7.10:** Validation result, the model's travel time prediction in dashed red and ground truth (Bluetooth) in solid blue for each I-880 route 2-9 (a-h). These predictions were done for I-880 on March 9, 2012 between 12PM and 9PM. Both stationary detectors and probes were used during the model run. Estimated fundamental diagrams were used.



**Figure 7.11:** Qualitative validation of the actual (non-aggregated) I-880 model output, a velocity field given in miles per hour on March 9, 2012 between 12PM and 9PM. Both stationary detectors and probes were used during the model run. Estimated fundamental diagrams were used with the calibrated parameters.

### 7.2.3 Global calibration framework using estimated fundamental diagrams: Interstate 15, Ontario, CA

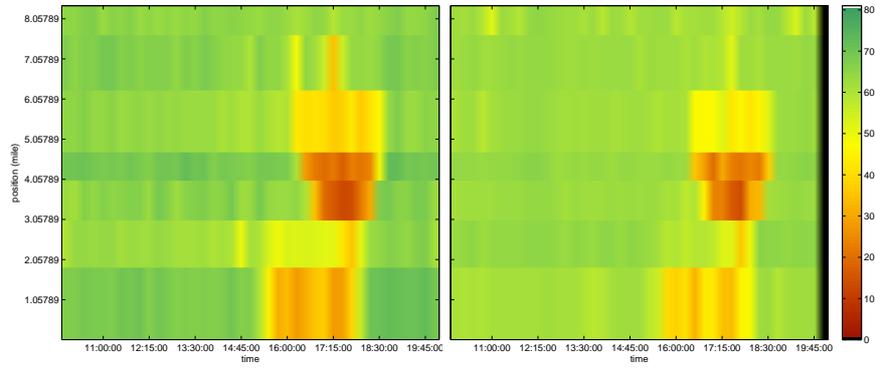
This section presents the results that were acquired following the framework presented in . The complex had not converged after 500 + 30 iterations<sup>3</sup>, the best parameter set had a MAPE equal to 7.4 % on the calibration date range 10AM-8PM on April 4, 2012.

The aggregated outputs for the validation dates April 3 and 5 are shown in figures 7.13 and 7.15 together with the ground truth for each of these two days. For the sake of comparison is the model output prior to calibration included in figures 7.12 and 7.14.

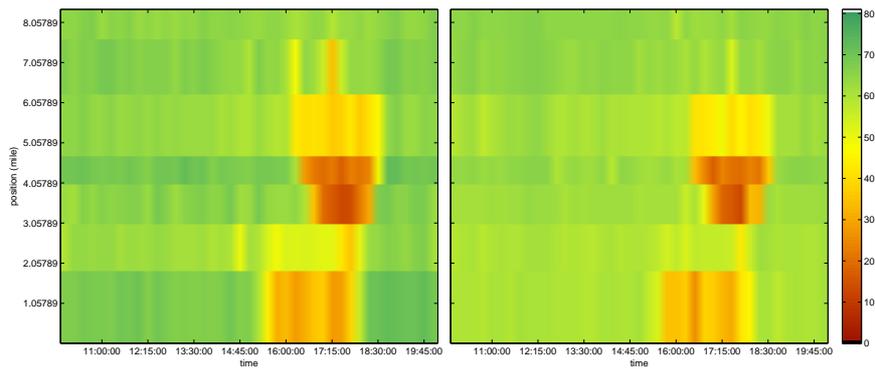
The performance metric is given on a route-by-route basis in Table 7.7 and for the model as a whole. The validation dates (April 3 and 5) as well as the calibration date (April 4). Within parenthesis in Table 7.7 is the model's performance for each of these three days prior to calibration.

The travel times for each route and validation day are given in Figures 7.16 and 7.17. These are the results of the highway model after calibration. Note that the seventh route was omitted from these figures since that section seldom shows any sign of congestion. All values are given in seconds.

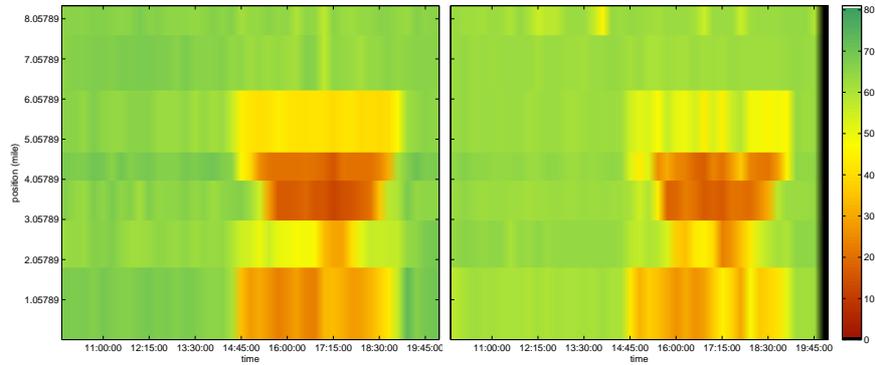
<sup>3</sup>Some parameters had approach the bounds and were therefore recalibrated, thus the extra 30 iterations.



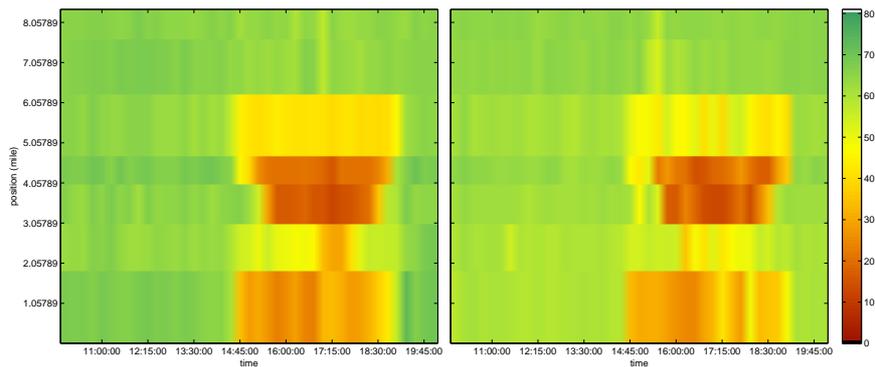
**Figure 7.12:** Ground truth (left) and the aggregated model output prior to calibration (right) for I-15 northbound on April 3, 2012 between 10AM and 8PM. Induction loop detectors and probes type B were used during the model run. The seven Bluetooth routes are clearly visible as horizontal stripes in both of the plots.



**Figure 7.13:** Ground truth (left) and the aggregated model output after calibration (right) for I-15 northbound on April 3, 2012 between 10AM and 8PM. Induction loop detectors and probes type B were used during the model run. The seven Bluetooth routes are clearly visible as horizontal stripes in both of the plots.



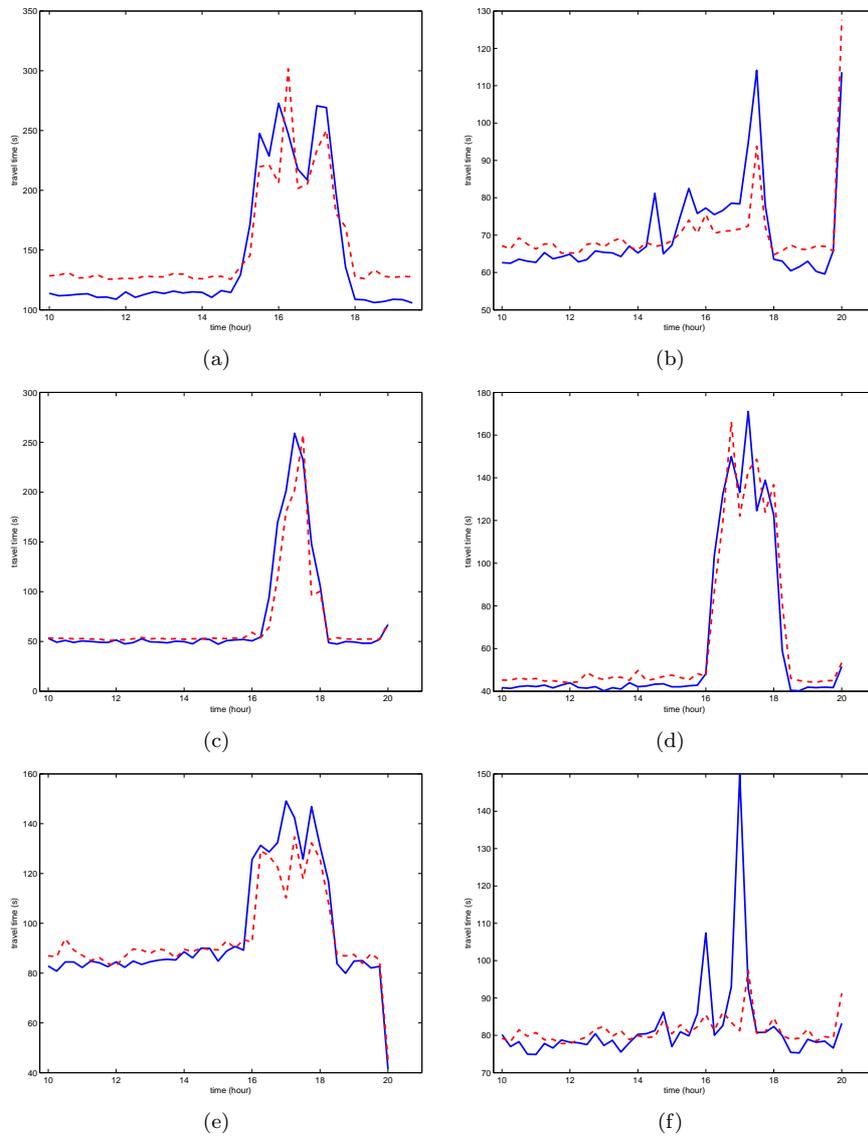
**Figure 7.14:** Ground truth (left) and the aggregated model output prior to calibration (right) for I-15 northbound on April 5, 2012 between 10AM and 8PM. Induction loop detectors and probes type B were used during the model run. The seven Bluetooth routes are clearly visible as horizontal stripes in both of the plots.



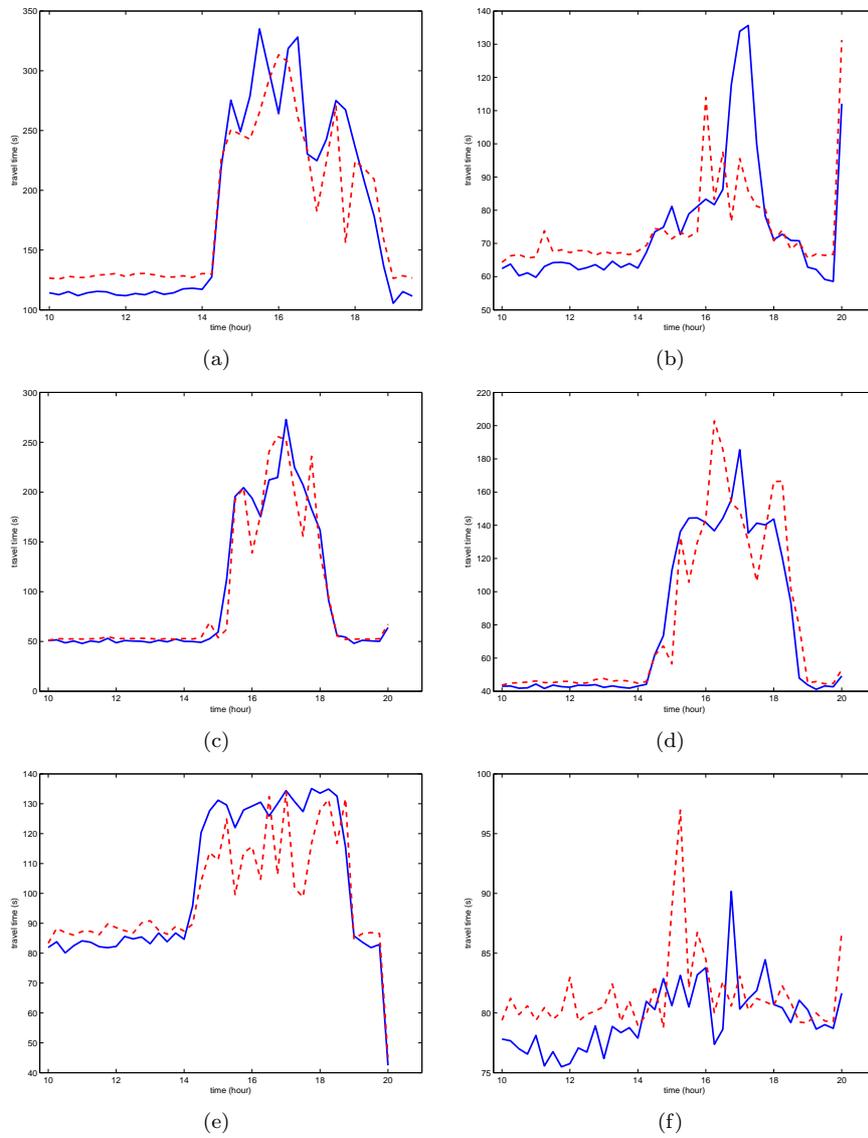
**Figure 7.15:** Ground truth (left) and the aggregated model output after calibration (right) for I-15 northbound on April 5, 2012 between 10AM and 8PM. Induction loop detectors and probes type B were used during the model run. The seven Bluetooth routes are clearly visible as horizontal stripes in both of the plots.

**Table 7.7:** Calibrated model performance in relation to Bluetooth ground truth on a route-by-route basis for the validation dates April 3 and 5, 2012 and the calibration date April 4, 2012. A value within parenthesis gives the performance prior to the calibration effort.

<b>Route</b>	<b>MAPE (%) April 3</b>	<b>MAPE (%) April 5</b>	<b>MAPE (%) April 4</b>
1 (a)	14.0	12.0	12.0
2 (b)	6.7	9.6	5.8
3 (c)	8.4	9.0	7.0
4 (d)	10.0	12.0	10.0
5 (e)	5.5	7.9	8.7
6 (f)	4.5	3.8	3.4
7 (-)	2.0	2.4	2.4
<b>Model</b>	<b>7.4 (8.2)</b>	<b>8.5 (9.7)</b>	<b>7.4 (9.3)</b>

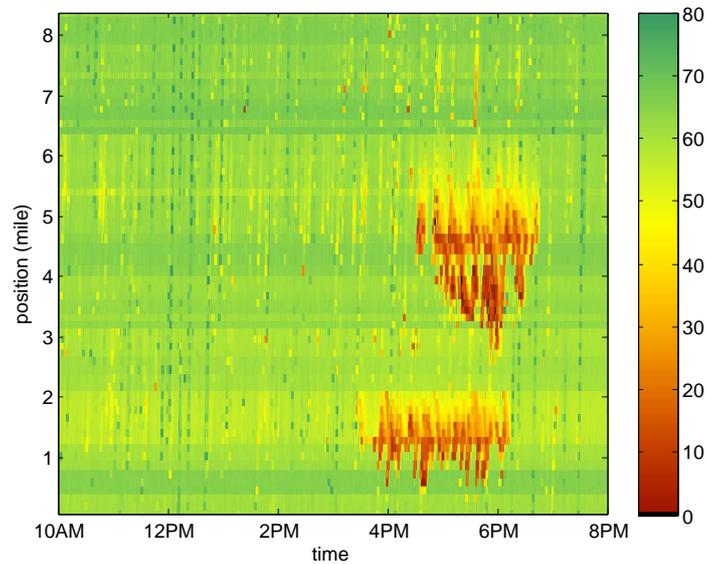


**Figure 7.16:** Validation of model predictions of the travel time in red and ground truth (Bluetooth) in blue for each I-15 route 1-6 (a-f). These predictions were done for April 3, 2012 between 10AM and 8PM. Both stationary detectors and probes type B were used during the model run.

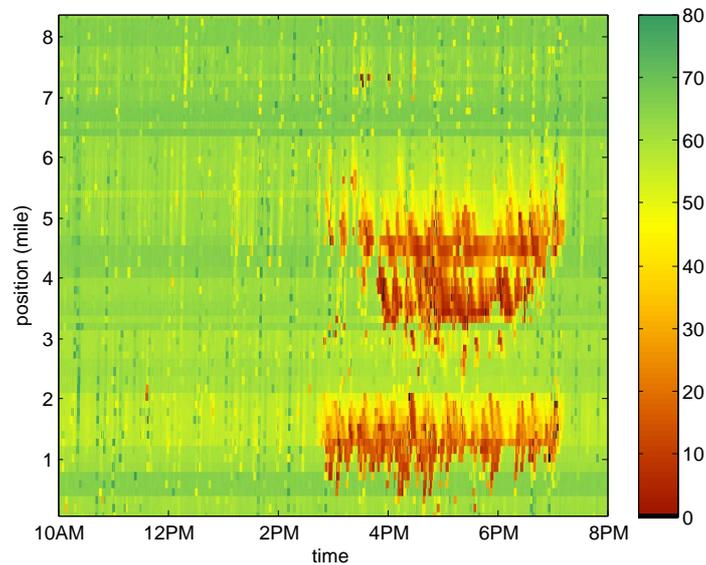


**Figure 7.17:** Validation of model predictions of the travel time in red and ground truth (Bluetooth) in blue for each I-15 route 1-6 (a-f). These predictions were done for April 5, 2012 between 10AM and 8PM. Both stationary detectors and probes type B were used during the model run.

Finally, and as means to enable a more qualitative validation of the model's output after calibration is the highway model's actual output included in figures 7.18 and 7.19. These are the velocity fields where the speed is given in miles per hour in space and time.



**Figure 7.18:** Qualitative validation of the actual (non-aggregated) I-15 model output, a velocity field given in miles per hour on April 3, 2012 between 10AM and 8PM. Both stationary detectors and probes were used during the model run. Estimated fundamental diagrams were used with the calibrated parameters.



**Figure 7.19:** Qualitative validation of the actual (non-aggregated) I-15 model output, a velocity field given in miles per hour on April 5, 2012 between 10AM and 8PM. Both stationary detectors and probes were used during the model run. Estimated fundamental diagrams were used with the calibrated parameters.

## Chapter 8

# Analysis and Discussion

The purpose of this chapter is to present an analysis made from all of the work in the thesis. The main focus will lie on the performance of the calibration frameworks, this includes all the procedures as well. The main focus also lie on the results from using them. In this thesis there is no separate sections for discussion or analysis, they are instead interwoven with each other throughout the chapter.

### 8.1 Fundamental diagram estimation and link assignment procedure

If a comparison of the results from tables 7.1 and 7.2 is made for I-880, it is possible to proclaim that the performance of the standard fundamental diagram is not bad. At least when relating to the result from the system using the calibrated fundamental diagrams. The best performance when discussing the results from I-880 is using the standard fundamental diagram, but the difference is not that great when using calibrated fundamental diagram<sup>1</sup> with backward link assignment. This is somewhat vexing, since the expectation was that the system would perform better in general after calibrating the fundamental diagram<sup>2</sup>. If instead the comparison is made visually from the graphical outputs for I-880 and I-15 Ontario in Appendix C, the calibration procedures looks more promising than just using the standard fundamental diagram.

When considering the results from I-15 Ontario as well it is also possible to conclude that the standard fundamental diagram performs good relative to the calibrated fundamental diagrams overall. However, for I-15 Ontario using the standard fundamental diagrams, the system does not perform as good as for CID 1009.

Another interesting observation is that the performance of the system really differs when using different link assignment procedures. This points out that the link assignment procedure is vital for the system performance. Since only two different approaches of fundamental diagram link assignments are discussed in this thesis, one cannot rule out that there is another link assignment procedure

---

<sup>1</sup>This is true for both fundamental calibration procedure version II and version III.

<sup>2</sup>We did not consider redefining a global fundamental diagram by calibration, since literature states that road capacity varies over time and space [35].

that make the system perform better for I-880 when using either version II or version III of the fundamental diagram calibration procedure.

### 8.1.1 Calibration framework

The model MAPE for I-880 with the first framework<sup>3</sup> and with the standard fundamental diagrams, for the calibration date 9:th of March 2012 is 9.5 % respectively for the validation date 14:th of March is 10.1 %. Whilst with the second framework<sup>4</sup> the model MAPE is for the calibration date 11.6 % respectively 13.4 % for the validation date.

Why the model run with the second framework<sup>5</sup> performs worse for I-880, is probably due to the fact that the system in general performs worse with estimated fundamental diagrams, see the results from Table 7.1 and 7.2. It can also depend on where the stationary sensors are located. The sensors act as a good boundary condition, since the state at these locations almost always is known. It would therefore probably be better for the system performance to halt the estimation where the last sensor is placed. We have no suggestions for networks without stationary sensors though.

The model MAPE for I-15 Ontario is 8.5 % for the calibration date 4:th of April, 7.4 % for the validation dates 3:rd and 5:th of April. Why the model performance is better than for I-880 is probably due to less congestion, see figure C.17(a), C.18(a) and C.19(a). Less congestion practically means that the system does not need to estimate the changes in the traffic state. So therefore, as long as the free flow conditions are calibrated right, the model performance will be more accurate for a network where the system have less congestion in general, at least when comparing to a more congested network.

It is also very noteworthy, that the complex method does not converge when calibrating I-15 Ontario. This means that the model result is probably going to improve, when run over more iterations. Since the time for calibrating the model was limited and system limitations this was the best that we could do. But, it would have been very interesting to see how well the model performance would be when the complex method converges. Why the complex method does not converge could depend on several different things. One is that the convergence criteria can be set to be too exact. Another can be that the amount of iterations is too few.

Even though the second framework performs less well than the first, estimating the fundamental diagrams has its advantages. The advantage is in the ability to adapt the capacity with respect to actual measurements, rather than trust that the standard fundamental diagram is accurate. When calibrating an unfamiliar road and the conditions regarding that unfamiliar road (test site I-15 Ontario); assuming that the same conditions can be applied from those on I-880 can be a false statement and a problem. Therefore adapting the fundamental diagrams with respect to local road conditions is a good approach for coping with this.

---

<sup>3</sup>Local and source alternating calibration framework.

<sup>4</sup>Global calibration and type-differentiated calibration framework using estimated fundamental diagrams

<sup>5</sup>Global and type-differentiated calibration framework using estimated fundamental diagrams.

## Chapter 9

# Conclusion and Future Work

This chapter will present the conclusion of all the work done in the thesis. In this chapter we will provide suggestions and recommendations of how to calibrate a traffic state space model. Suggested future work will also be presented in this chapter.

It is possible to conclude that this thesis is able to present a framework for calibrating a state space model. The performance of the calibration framework and the results are not as exhaustive as we would have preferred. But all of the results for the framework points in one direction and that is, with calibration we get a higher model accuracy than with the default parameters. Since the complex method did not converge, we cannot state that the model is calibrated, but it is possible to state that improvement of the accuracy of the model have increased.

The preferable and suggested framework for further use, according to the authors, is the global and type-differentiated calibration framework using estimated fundamental diagram. The main reason for this is that the approach when calibrating parameters is more general and adaptive than with the other framework. Even considering it does not perform as good when compared to the other framework, but the differences between the output is not substantially large. There is still an improvement when using this framework when compared to the default parameters in the system.

It is also possible to conclude that there is still much work left to do, especially regarding testing and further development of the framework. Not only testing more calibration procedures, but also testing the framework with other similar state space models as well e.g. the density based cell transmission model. The suggested main area for further research is certainly testing the framework. Another relevant subjects of research would be to test other procedures for assigning fundamental diagrams.

Another interesting area for further research, is to explore the impact of different modifications to the Mobile Millennium. One of this would be testing other version than the Hyperbolic Linear fundamental diagram.

# Bibliography

- [1] A. Bayen, J. Butler, and A. D. Patire, *Mobile Millennium Final Report*. California Center for Innovative Transportation Institute of Transportation Studies University of California, Berkeley, 2011.
- [2] D. B. Work, S. Blandin, O. Tossavainen, B. Piccoli, and A. M. Bayen, “A traffic model for velocity data assimilation,” *Applied Mathematics Research eXpress*, vol. 2010, pp. 1–35, April 2010.
- [3] M. J. Lighthill and G. B. Whitham, “On kinetic wave ii: a theory of traffic flow on crowded roads,” *Proceedings of the Royal Society of London*, vol. Series A, pp. 317–345, 1955.
- [4] P. Richards, “Shockwaves on the highway,” *Operations Research*, vol. 4, pp. 42–51, 1956.
- [5] B. D. Greenshield, “A study of traffic capacity,” *Highway Research Board*, vol. 14, pp. 448–477, 1935.
- [6] C. F. Daganzo, “The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory.,” *Transportation Research Part B* 28, no. vol. 28, pp. 269–287, 1994.
- [7] C. J. Herrera and A. M. Bayen, “Incorporation of lagrangian measurements in freeway traffic state estimation,” *Transportation Research Part B*, vol. 44, pp. 460–481, 2012.
- [8] C. Bardos, A. Le Roux, and J. Nedelec, “First order quasilinear equations with boundary conditions,” *Communication in Partial Differential Equations*, vol. 4, pp. 1017–1034, 1979.
- [9] C. F. Daganzo, “The cell transmission model. part i: A simple dynamic representation of highway traffic,” *California Path Program Institute of Transportation Studies, University of California, Berkeley*, 1993.
- [10] G. Evensen, “Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics,” *Journal of Geophysical Research*, vol. 99, pp. 10,143–10,162, May 1994.
- [11] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *Transactions of the ASME–Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.

- [12] G. Evensen, *Data Assimilation: The Ensemble Kalman Filter*. Springer-Verlag Berlin Heidelberg, 2 ed., 2009.
- [13] P. J. van Leeuwen and G. Evensen, “Assimilation data and inverse methods in terms of a probabilistic formulation,” *Monthly Weather Review*, vol. 124, pp. 2898–2913, 1996.
- [14] G. Evensen, “The ensemble kalman filter for combined state and parameter estimation,” *IEEE Control Systems Magazine*, vol. 29, pp. 83–104, 2009.
- [15] G. Evensen, “The ensemble kalman filter: Theoretical formulation and practical implementation,” *Ocean Dynamics*, vol. 53, pp. 343–367, 2003.
- [16] G. Burgers, P. J. van Leeuwen, and G. Evensen, “Analysis scheme in the ensemble kalman filter,” *American Meteorological Society*, vol. 126, pp. 1719–1724, 1998.
- [17] B. Hellinga, “Requirement for the calibration of traffic simulation models.” Unpublished, <http://www.civil.uwaterloo.ca/bhellinga/publications/Publications/CSCE-1998-Calibration.PDF>, accessed 2012-07-31.
- [18] D. Ngoduy and M. J. Maher, “Calibration of second order traffic models using continuous cross entropy method,” *Transportation Research Part C*, vol. 24, pp. 102–121, 2012.
- [19] H. Rakha, B. Hellinga, M. Van Aerde, and W. Perez, “Systematic verification, validation and calibration of traffic simulation models.” Paper presented at the Annual Meeting, TRB, Washington, DC, 1996.
- [20] B. Park and H. Qi, “Development and evaluation of a calibration and validation procedure for microscopic simulation models,” tech. rep., Virginia Transportation Research Council, 2004.
- [21] G. Dervisoglu, G. Gomes, J. Kwon, R. Horowitz, and P. Varaiya, “Automatic calibration of the fundamental diagram and empirical observations on capacity,” 2009. Presented at 88th Annual Meeting of the Transportation Research Board, Washington, D.C., 2009.
- [22] L. Munoz, X. Sun, D. Sun, G. Gomes, and R. Horowitz, “Methodological calibration of the cell transmission model,” in *Proceeding of the 2004 American Control Conference*, vol. 1, pp. 798–803, 2004.
- [23] M. Cremer and M. Papageorgiou, “Parameter identification for a traffic flow model,” *Automatica*, vol. 17, no. 6, pp. 837–843, 1981.
- [24] Z. Li, H. Liu, and J. Li, “A calibration and validation procedure for microscopic simulation model,” in *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference*, pp. 563–568, International IEEE, September 2010.
- [25] R. Dowling, A. Skabardonis, and V. Alexiadis, “Traffic analysis toolbox volume iii: Guidelines for applying traffic microsimulation software,” tech. rep., Dowling Associates, Inc., 2004.

- [26] C. Braban-Ledoux, “Metacor - a macroscopic modelling tool for corridor application to the stockholm test site,” tech. rep., The Royal Institute of Technology, Center For Traffic Engineering & Traffic Simulation, 2000.
- [27] A. Kotsialos, Y. Pavlis, M. Papageorgiou, and G. Vardaka, “Daccord - development and application of co-ordinated control of corridors, annex b,” tech. rep., Hague Consulting Group, 1997.
- [28] A. Kotsialos, M. Papageorgiou, C. Diakaki, Y. Pavlis, and F. Middelham, “Traffic flow modeling of large-scale motorway networks using the macroscopic modeling tool metanet,” *IEEE Transactions on Acoustics Speech and Signal Processing on Intelligent Transportation Systems*, vol. 3, pp. 282–292, December 2002.
- [29] *Highway Capacity Manual*. Transportation Research Board, 2000.
- [30] P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [31] G. Blom, *Sannolikhetsteori och statistikteori med tillampningar*. Studentlitteratur AB, 2005.
- [32] M. J. Box, “A new method of constrained optimization and a comparison with other methods,” *The Computer Journal*, vol. 8, pp. 42–52, 1965.
- [33] J. Farkas and K. Jarmai, *Analysis and Optimum Design of Metal Structures*. A.A. Balkema Publishers, 1997.
- [34] J. Andersson, *Multiobjective Optimization in Engineering Design - Applications to Fluid Power Systems*. PhD thesis, Linkopings universitet, 2001.
- [35] C. Antoniou, B. R., J. Barcelo, M. Ben-Akiva, H. Hanabusa, R. Hoiguchi, N. Koutsopoulos, H., D. Krajzewicz, M. Kuwahara, R. Liu, P. Vortisch, Y. Wang, and Q. Yang, *Fundamentals of Traffic Simulation*. Springer Science, 2010.

# Appendix A

## Identified Parameters

This appendix present the parameters found during the system analysis. All parameters are divided sections depending on which class they belong to in the system structure. Thereafter the parameters are sorted into classes. All parameters found are presented below. The parameters which are not directly correlated (or have not found the correlating interpretation) to the literature are denoted as  $\varphi_i$  where  $i$  denotes the  $i^{th}$  parameter found.

### Measurement loader

#### Timers

$\varphi_1$  = swedish lag.

This timer makes sure that there is time enough for the system to fill the database with filtered data. Current value is set to 300f (seconds).

*Code:*

```
if (Monitor.get_db_env().equalsIgnoreCase("kth01")) {  
    float swedishLag = 300f;//FIXME: hard coded value  
    ...  
}
```

$\varphi_2$  = Seconds Since Last Fixed Measure.

This timer sets how long time it is required for resetting the confidences. Current value set to 180f (seconds).

*Code:*

```
if (secondsSinceLastFixedMeas > 180f) { //FIXME: hard coded value  
    resetConfidences = true;  
}
```

$\varphi_3$  = interval + 240f

The timer is used when getting Probe A speeds. The purpose of timer is probably used because of the same reasons as the Swedish lag timer, which is implemented due to some system lag.

*Code:*

```
if (loadProbes) {
```

```

...
    Probe A Speed = DataType.Probes A Feed.getSpeeds(net, time, interval +
    240f); //pre filter for probe B speeds and adding the std
...
}

```

### Confidence parameters

$\varphi_4$  = Link Confidence

This parameter is for setting the default confidence on a link when creating the Hashmap. Default value is set to 0.025.

*Code:*

```

//Create Hashmap for link confidences
HashMap< Integer, Float> linkConfidenceMap = new HashMap< Integer,
Float> ();
for (Integer linkId : this.id) {
    linkConfidenceMap.put(linkId, 0.025f); //FIXME: hard coded value
}

```

$\varphi_5$  = PeMS Confidence.

If there is a PeMS on the link, then set the confidence in the Hashmap to  $\varphi_5$ . Current value of 0.5.

*Code:*

```

//change confidence to 0.5 if pems data is in the link
for (Datum.Speed pemsSpeedMeas : pemsSpeed) {
    linkConfidenceMap.put(pemsSpeedMeas.spot.link.id, 0.5f);
    //FIXME: hard coded value
}

```

$\varphi_6$  = Radar Confidence.

Sets the confidence if radar data is available on the link. The current default value is 0.5.

*Code:*

```

//change confidence to 0.5 if radar data is in the link
for (Datum.Speed radarSpeedMeas : radarSpeed) {
    linkConfidenceMap.put(radarSpeedMeas.spot.link.id, 0.5f);
    //FIXME: hard coded value
}

```

$\varphi_7$  = info24radar (*Note: The same type as radar confidence*).

This parameter is mainly for the Swedish system. Set the confidence on the link if not a radar is present. Current set value is 0.0.

*Code:*

```

//change confidence to 0.5 if radar data is in the link
for (Datum.Speed info24SpeedMeas : info24Speed) {
    linkConfidenceMap.put(info24SpeedMeas.spot.link.id, 0f);
    //FIXME: hard coded value
}

```

$\varphi_8$  = Probe B Speed Confidence (is it deviation?).

Sets the confidence if a Probe B is on a link at a current spot with a current speed during a certain time. Current value 4.0.

*Code:*

```
if (ds.speed > 1f) {  
    dsArrayList.add(new Datum.Speed(ds.time, ds.spot, ds.speed, 4F));  
    //FIXME: hard coded value  
...  
}
```

$\varphi_9$  = confidence checker for B Probes.

Checks whether the confidence is below a certain level. If it is then increase the confidence with a certain amount. Current value 0.05.

*Code:*

```
if (prevConfidence < (1f - 0.05f)) { //FIXME: hard coded value  
    linkConfidenceMap.put(linkId, prevConfidence + 0.05f);  
    //FIXME: hard coded value  
}
```

$\varphi_{10}$  = Add confidence.

Adds to current confidence if probe B is on the link. Current value is 0.05.

*Code:*

```
if (prevConfidence < (1f - 0.05f)) { //FIXME: hard coded value  
    linkConfidenceMap.put(linkId, prevConfidence + 0.05f);  
    //FIXME: hard coded value  
}
```

$\varphi_{11}$  = Probe A Speed Confidence (standard deviation?).

Sets the confidence if a Probe A is on a link at a current spot with a current speed during a certain time. Current value 4.0.

*Code:*

```
if (ds.speed > 1f) { //FIXME: hard coded value  
    dsArrayList.add(new Datum.Speed(ds.time, ds.spot, ds.speed, 4F));  
    //FIXME: hard coded value  
...  
}
```

$\varphi_{12}$  = confidence checker (Probe A).

Checks whether the confidence is below a certain level. If it is then increase the confidence with a certain step. Current value 0.05.

*Code:*

```
if (prevConfidence < (1f - 0.05f)) { //FIXME: hard coded value  
    linkConfidenceMap.put(linkId, prevConfidence + 0.05f);  
    //FIXME: hard coded value  
}
```

$\varphi_{13}$  = add confidence (Probe A).

adds to the confidence a certain amount. Current value 0.05.

*Code:*

```

if (prevConfidence < (1f - 0.05f)) { //FIXME: hard coded value
    linkConfidenceMap.put(linkId, prevConfidence + 0.05f);
    //FIXME: hard coded value
}

```

$\varphi_{14}$  = confidence reset value.

If no data can be found for a certain time interval, then reset according to  $\varphi_{14}$ . Current value is set to 0.0.

*Code:*

```

//if no fixed sensor data for the last 3 minutes:
if (resetConfidences) {
    for (Integer linkId : this.id) {
        linkConfidenceMap.put(linkId, 0f);
    }
}

```

$\varphi_{15}$  = measurementNoiseCov[ii][ii] = stdev != null ? Math.pow(stdev, 2) : 16.0; Set the diagonal in the covariance matrix. Either Math.pow(stdev, 2) or to 16.0.

*Code:*

```

for (int ii = 0; ii < numberOfSpeedMeasurements; ii++) {
    measurements[ii] = speedData.get(ii).speed;
    Float stdev = speedData.get(ii).stdDev;
    measurementNoiseCov[ii][ii] = stdev != null ? Math.pow(stdev, 2) : 16.0;
    //FIXME: hard coded value
}

```

### Other parameters

$\varphi_{16}$  = Localization distance.

Sets how far the localization operator should check for neighbouring edges. Current value 100.0.

*Code:*

```

this.localizationDistance = 100f; //FIXME: hard coded value

```

## Estimate Manager

### Timers

$\varphi_{17}$  = Get data interval timer.

This timer is for setting the get data interval timer to a default value. This value should probably depend on how much lag the system needs to filter and store measurements. Current value set to 180.0.

*Code:*

```

if (tempgetdatainterval != null) {
    this.get_data_interval = tempgetdatainterval;
}
else {
    this.get_data_interval = 180.0f;
    //FIXME: hard coded value
}

```

```
...  
}
```

### Statistic Based Parameters

$\varphi_{18}$  = Model Error Mean.

If not the model error mean can be found in the data base then set it to a default value. Current default value is 0.0.

*Code:*

```
try {  
    modelErrorMean = this.network.attributes.getDouble(  
        "highway_modelErrorMean");  
    if (null == modelErrorMean) throw new NetconfigException(null, null);  
}  
catch (NetconfigException ex) {  
    modelErrorMean = 0d;  
}
```

$\varphi_{19}$  = Model Error Covariance.

If not the model error covariance can be found in the data base, then set it to a default value. Current is 0.1.

*Code:*

```
try {  
    modelErrorCov = this.network.attributes.getDouble(  
        "highway_modelErrorCov");  
    if (null == modelErrorCov) throw new NetconfigException(null, null);  
}  
catch (NetconfigException ex) {  
    modelErrorCov = 0.1d;  
}
```

## Velocity Flux Function Experimental

### CTM-v model parameters

$\varphi_{20}$  = Speed tolerance.

The purpose of this speed tolerance parameter is to set the difference between the free flow speed and the critical speed  $v_c$ . Current value is set to 2 m/s.

*Code:*

```
double vtolerance = 2f; //FIXME: hard coded value
```

$\varphi_{21}$  = Critical speed.

This parameters purpose is to indicate at which speed the critical flow and density occurs. Current default value is 26 m/s.

*Code:*

```
double vcritical = 26f; //FIXME: hard coded value
```

## Gudnov Link

### CTM-v model parameters

$\varphi_{22}$  = Parameter for increasing the free flow speed.

In the Gudnov Link class there is a procedure called `getFreeFlowSpeed`. This procedure fetches the free flow speed and adds speed for some apparent reason. Current value added =  $(3 / 2.23693629)$ .

*Code:*

```
public double getFreeFlowSpeed() {
    return this.freeFlowSpeed + (3 / 2.23693629);
    //FIXME: hard coded value
}
```

## Flow Model Network

### CTM-v model parameters

$\varphi_{23}$  = Shock wave speed.

The speed for how fast the shock wave travels are set in the flow model network class. Current value is -5.36448 m/s.

*Code:*

```
public static final double shockWaveSpeed = -5.36448;
//FIXME: hard coded value.
```

$\varphi_{24}$  = Maximum density.

Sets the maximum density per lane. Current value is 0.12430080 cars/m.

*Code:*

```
public static final double rhoMax = 0.124300808; //0.108763207;
//FIXME: hard coded value.
```

## Flow Model Runner

### CTM-v model parameters

$\varphi_{25}$  = `initialVelocitySigma`.

This parameter provides an interface of propagating the velocity field forward.

*Code:*

```
private final double initialVelocitySigma = 1;
//FIXME: hard coded value.
```

$\varphi_{26}$  = Free Flow Speed Correction.

This parameter changes the Free Flow Speed for the initial states. Current value is set to - 0.1 m/s.

*Code:*

```
initState[ij] =entry.getValue().getFreeFlowSpeed() - 0.1
```

## Parameter Store

The split ratios, number of lanes and free flow speed can be hard coded in this class.

## Network (Database)

### EnKF Parameters

$\varphi_{27}$  = Highway model error mean.

Sets the default model error mean. Current value 0.5.

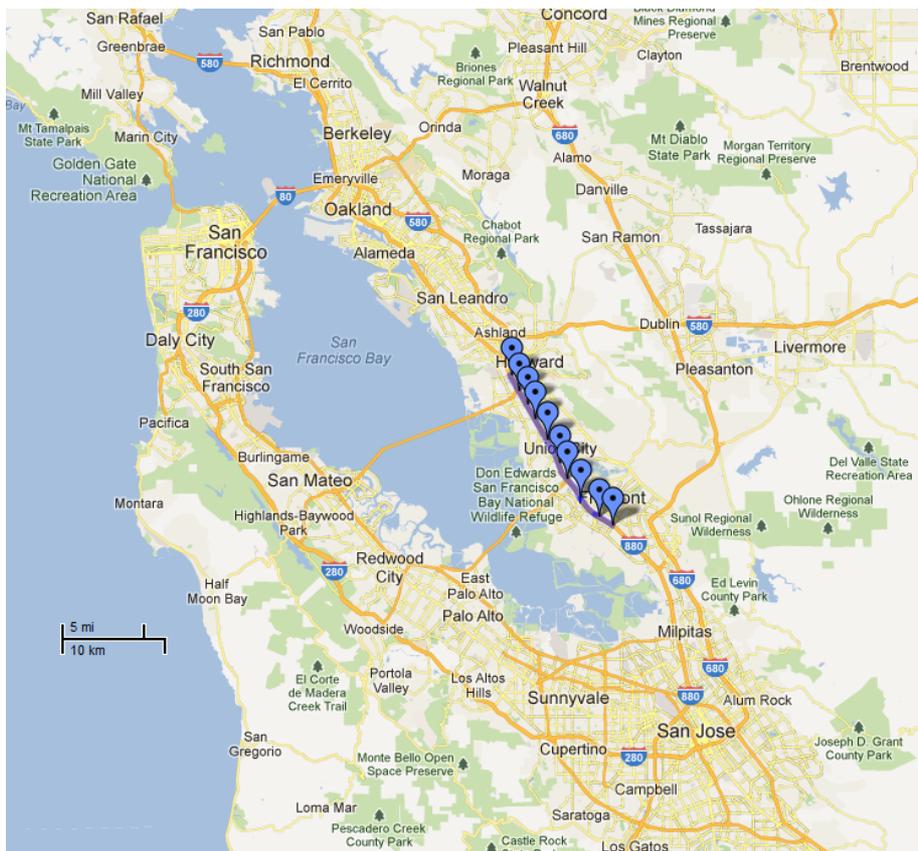
$\varphi_{28}$  = Highway error covariance.

Sets the default model error covariance. Current value is set to 2.

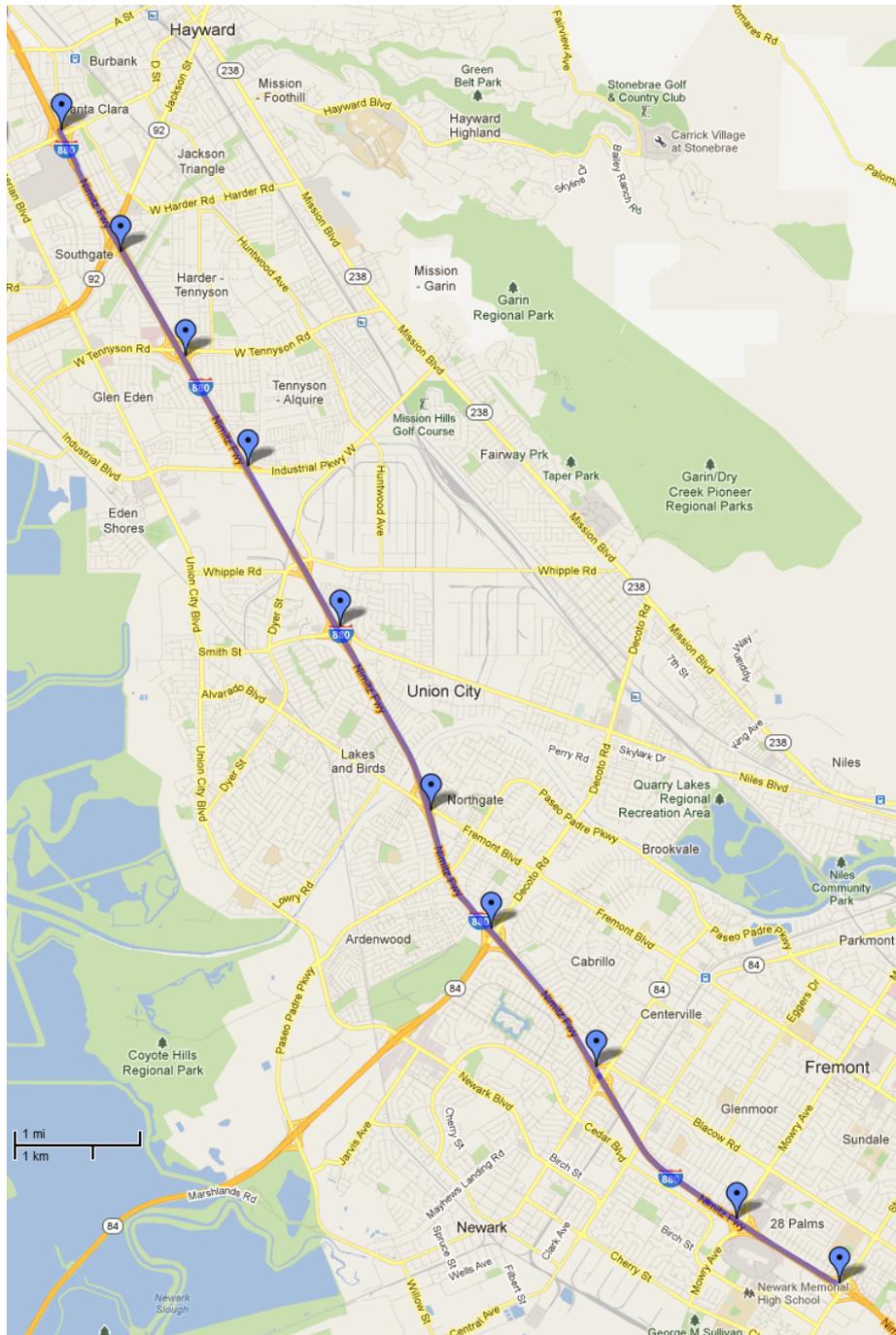
## Appendix B

# Bluetooth Deployment

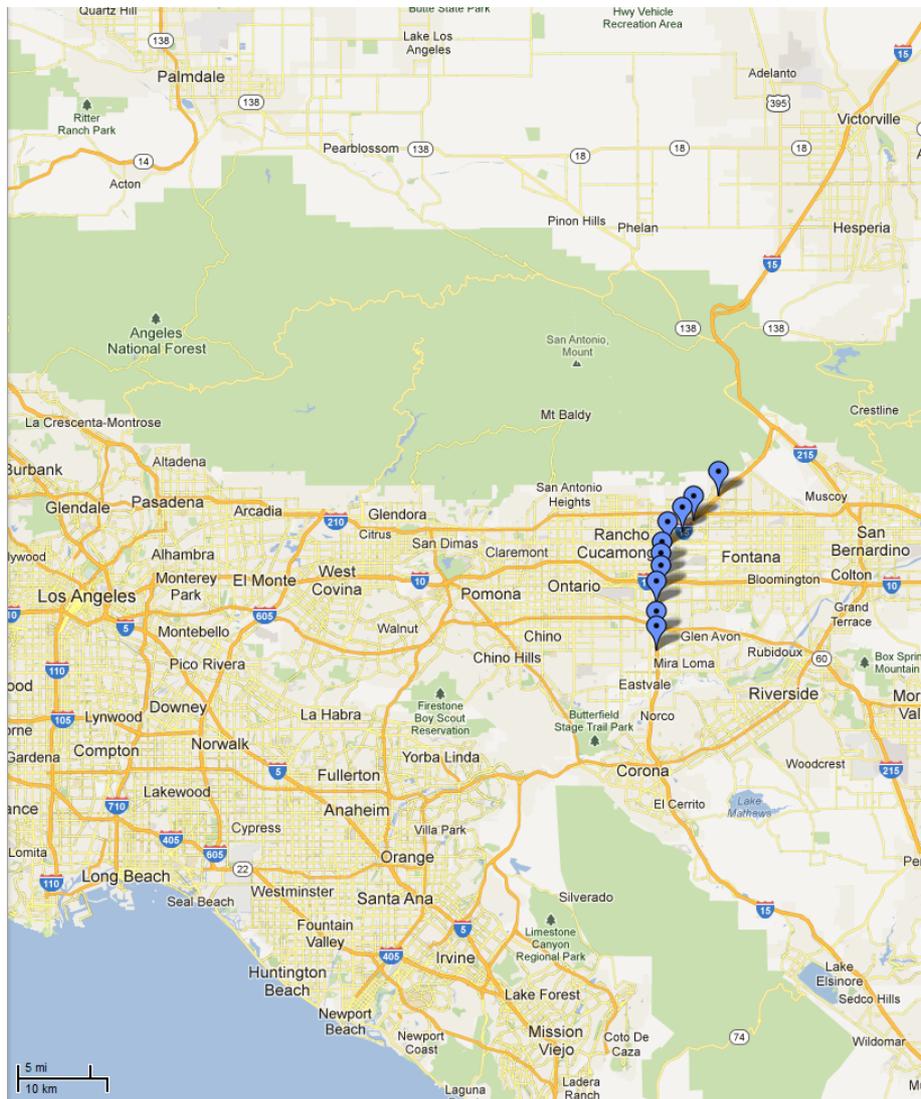
This appendix shows detailed maps as well as maps with overview over the Bluetooth deployments.



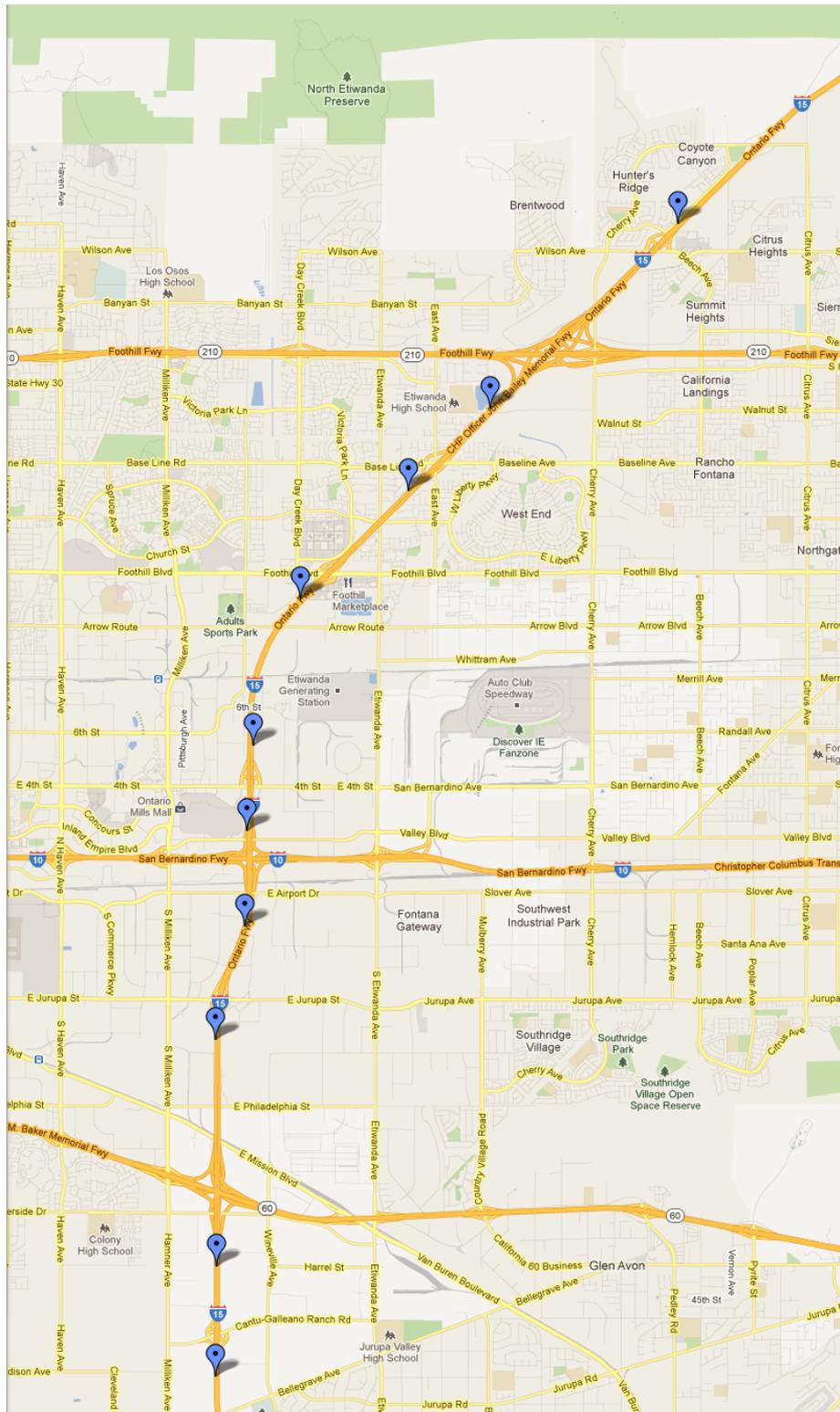
**Figure B.1:** A map presenting an overview of the Bluetooth deployment on Interstate 880, CA and the surroundings areas. Retrieved on September 22, 2012 from the webpage [maps.google.com](http://maps.google.com)



**Figure B.2:** A more detailed map of the Bluetooth deployment on Interstate 880, CA. Retrieved on September 22, 2012 from the webpage [maps.google.com](http://maps.google.com)



**Figure B.3:** The map show an overview of the Bluetooth deployment on I-15 Ontario, CA and the surroundings areas. Retrieved on September 22, 2012 from the web page maps.google.com



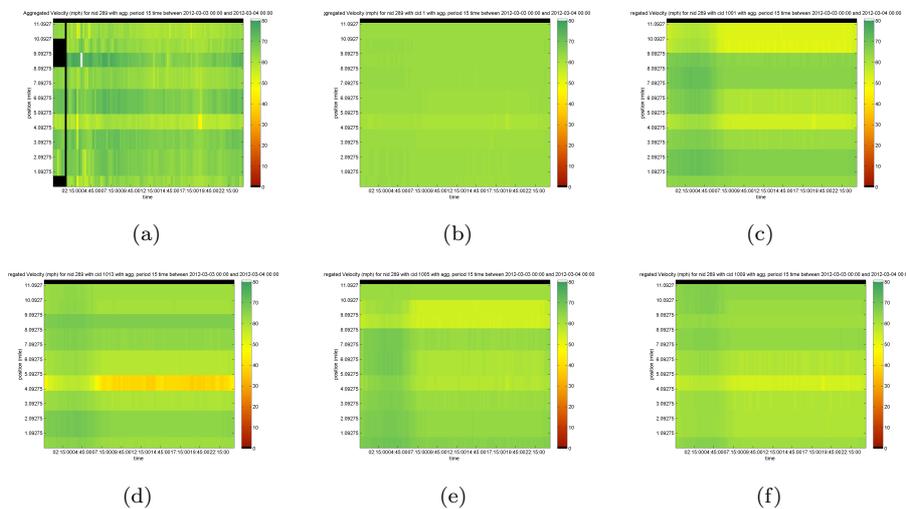
**Figure B.4:** A detailed map over the Bluetooth deployment on I-15 Ontario, CA. Retrieved on September 22, 2012 from the web page [maps.google.com](http://maps.google.com)

# Appendix C

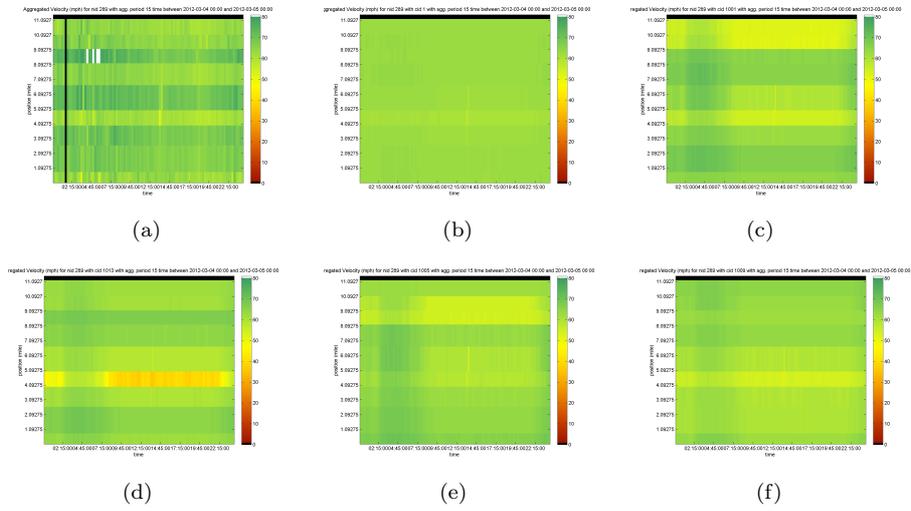
## Fundamental Diagram Calibration Results

This appendix contains a all of the space mean speed diagrams from running the systems with all of the configurations for the fundamental calibration procedure. This as well as all of the reference space mean speed diagrams from the Bluetooth and running the highway model with default parameters.

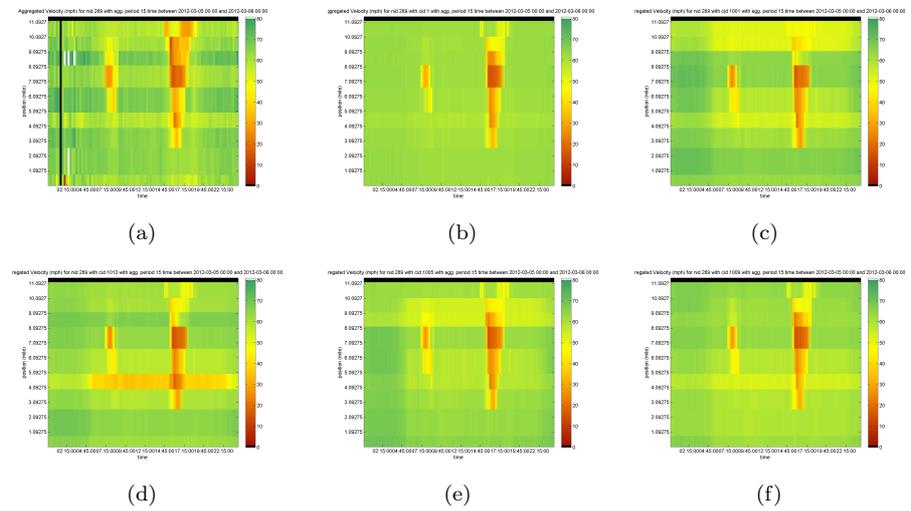
### I-880 Northbound



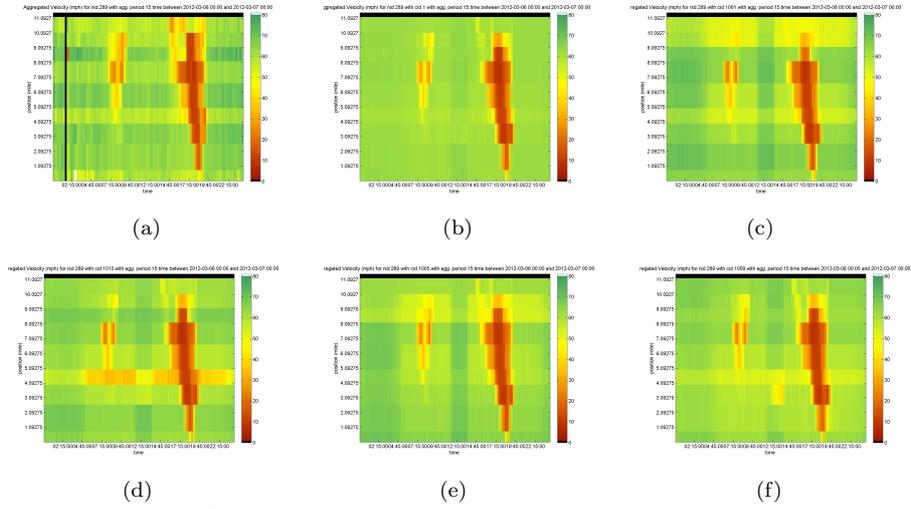
**Figure C.1:** Shows the space mean speed for the calibration methods and the ground truth from running the highway model on test site I-880. The aggregation period is 15 minutes over the time period from 2012-03-03 00:00-24:00. C.1(a) shows the ground truth. C.1(b) use the ad-hoc calibration. C.1(c) and C.1(e) use the fundamental diagrams from version III with the forward respectively backward link assignment. C.1(d) and C.1(f) use the fundamental diagrams from version II with the forward respectively backward link assignment.



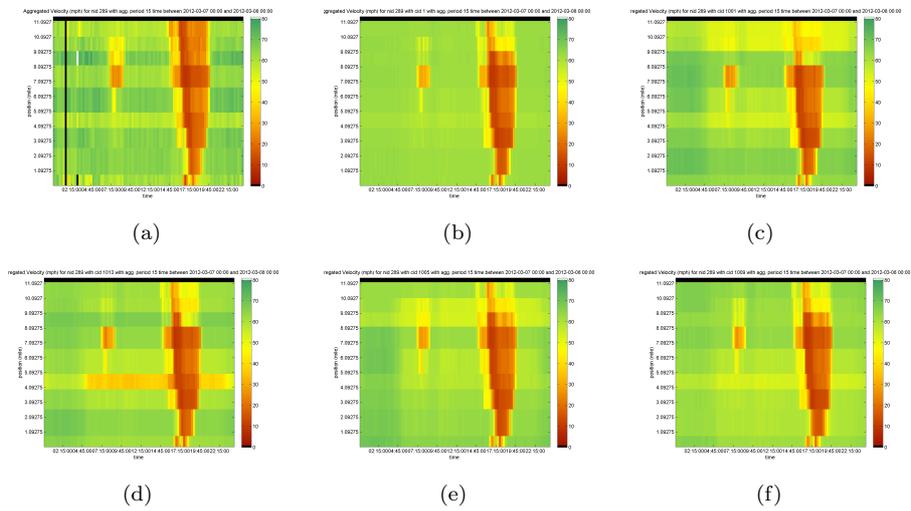
**Figure C.2:** Shows the space mean speed for the calibration methods and the ground truth from running the highway model on test site I-880. The aggregation period is 15 minutes over the time period from 2012-03-04 00:00-24:00. C.2(a) shows the ground truth. C.2(b) use the ad-hoc calibration. C.2(c) and C.2(e) use the fundamental diagrams from version III with the forward respectively backward link assignment. C.2(d) and C.2(f) use the fundamental diagrams from version II with the forward respectively backward link assignment.



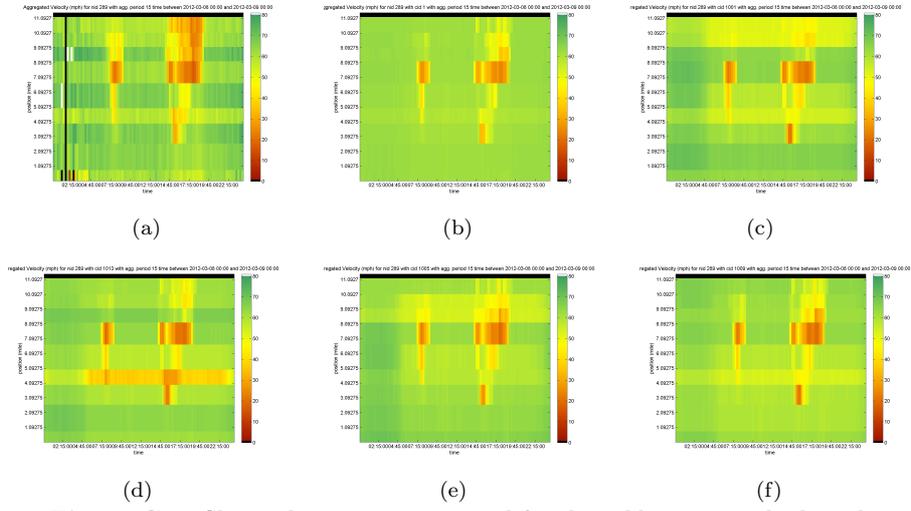
**Figure C.3:** Shows the space mean speed for the calibration methods and the ground truth from running the highway model on test site I-880. The aggregation period is 15 minutes over the time period from 2012-03-05 00:00-24:00. C.3(a) shows the ground truth. C.3(b) use the ad-hoc calibration. C.3(c) and C.3(e) use the fundamental diagrams from version III with the forward respectively backward link assignment. C.3(d) and C.3(f) use the fundamental diagrams from version II with the forward respectively backward link assignment.



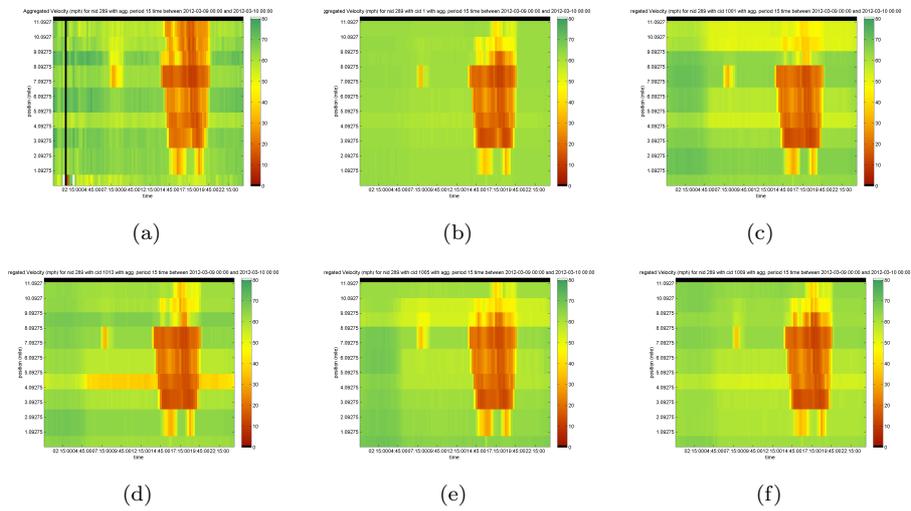
**Figure C.4:** Shows the space mean speed for the calibration methods and the ground truth from running the highway model on test site I-880. The aggregation period is 15 minutes over the time period from 2012-03-06 00:00-24:00. C.4(a) shows the ground truth. C.4(b) use the ad-hoc calibration. C.4(c) and C.4(e) use the fundamental diagrams from version III with the forward respectively backward link assignment. C.4(d) and C.4(f) use the fundamental diagrams from version II with the forward respectively backward link assignment.



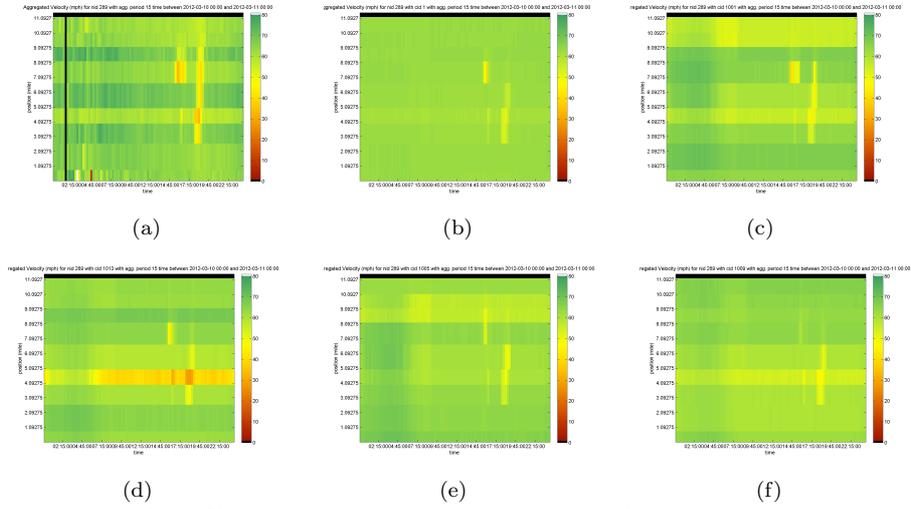
**Figure C.5:** Shows the space mean speed for the calibration methods and the ground truth from running the highway model on test site I-880. The aggregation period is 15 minutes over the time period from 2012-03-07 00:00-24:00. C.5(a) shows the ground truth. C.5(b) use the ad-hoc calibration. C.5(c) and C.5(e) use the fundamental diagrams from version III with the forward respectively backward link assignment. C.5(d) and C.5(f) use the fundamental diagrams from version II with the forward respectively backward link assignment.



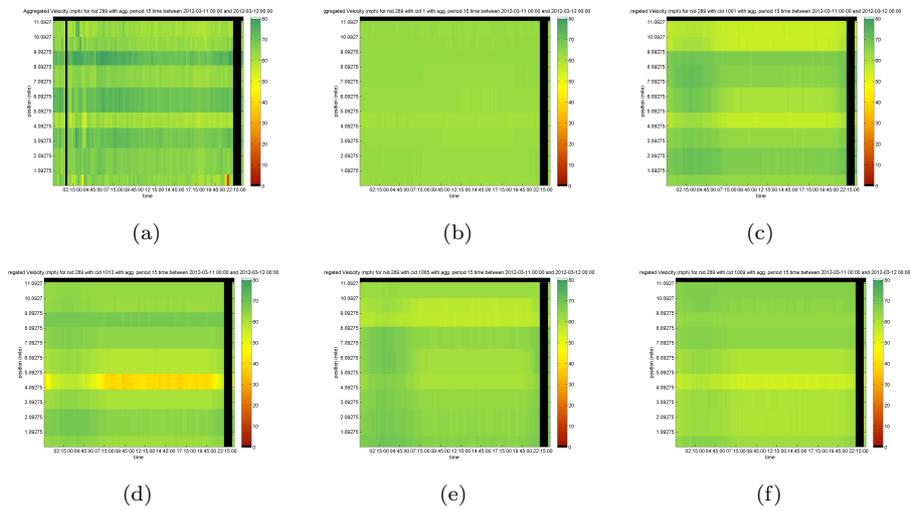
**Figure C.6:** Shows the space mean speed for the calibration methods and the ground truth from running the highway model on test site I-880. The aggregation period is 15 minutes over the time period from 2012-03-08 00:00-24:00. C.6(a) shows the ground truth. C.6(b) use the ad-hoc calibration. C.6(c) and C.6(e) use the fundamental diagrams from version III with the forward respectively backward link assignment. C.6(d) and C.6(f) use the fundamental diagrams from version II with the forward respectively backward link assignment.



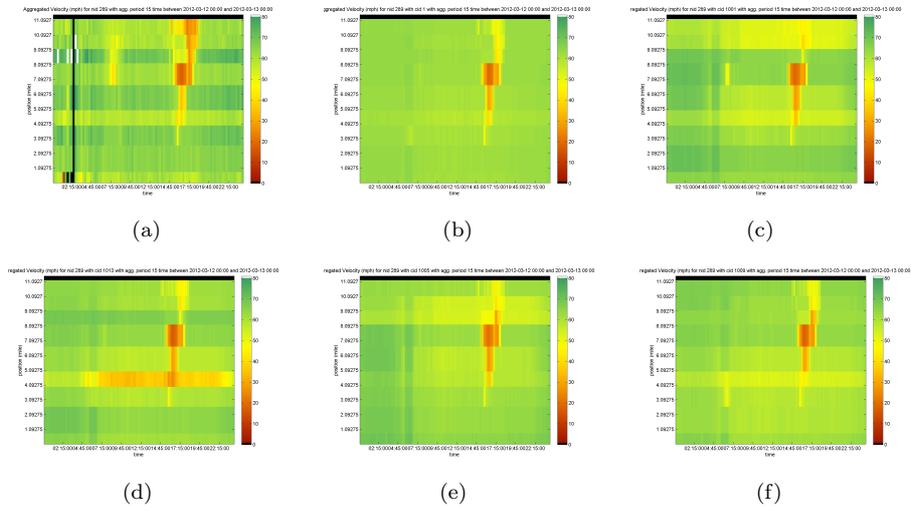
**Figure C.7:** Shows the space mean speed for the calibration methods and the ground truth from running the highway model on test site I-880. The aggregation period is 15 minutes over the time period from 2012-03-09 00:00-24:00. C.7(a) shows the ground truth. C.7(b) use the ad-hoc calibration. C.7(c) and C.7(e) use the fundamental diagrams from version III with the forward respectively backward link assignment. C.7(d) and C.7(f) use the fundamental diagrams from version II with the forward respectively backward link assignment.



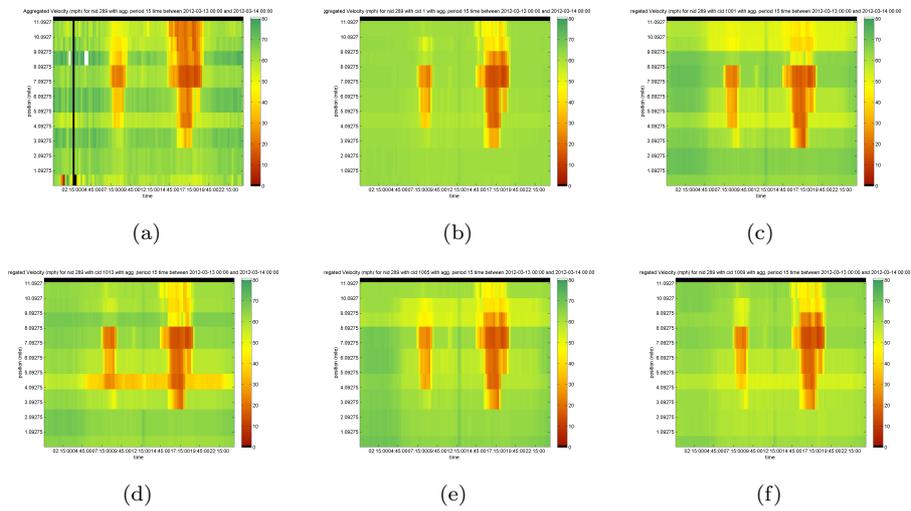
**Figure C.8:** Shows the space mean speed for the calibration methods and the ground truth from running the highway model on test site I-880. The aggregation period is 15 minutes over the time period from 2012-03-10 00:00-24:00. C.8(a) shows the ground truth. C.8(b) use the ad-hoc calibration. C.8(c) and C.8(e) use the fundamental diagrams from version III with the forward respectively backward link assignment. C.8(d) and C.8(f) use the fundamental diagrams from version II with the forward respectively backward link assignment.



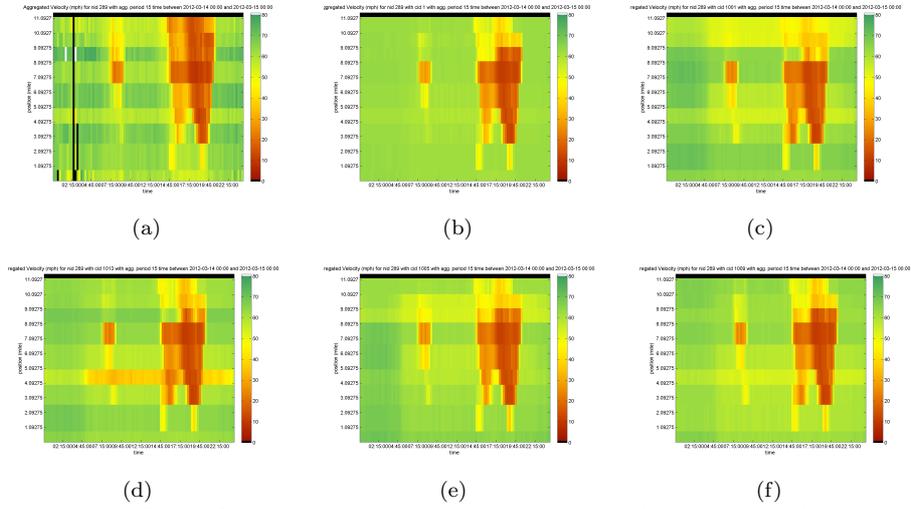
**Figure C.9:** Shows the space mean speed for the calibration methods and the ground truth from running the highway model on test site I-880. The aggregation period is 15 minutes over the time period from 2012-03-11 00:00-24:00. C.9(a) shows the ground truth. C.9(b) use the ad-hoc calibration. C.9(c) and C.9(e) use the fundamental diagrams from version III with the forward respectively backward link assignment. C.9(d) and C.9(f) use the fundamental diagrams from version II with the forward respectively backward link assignment.



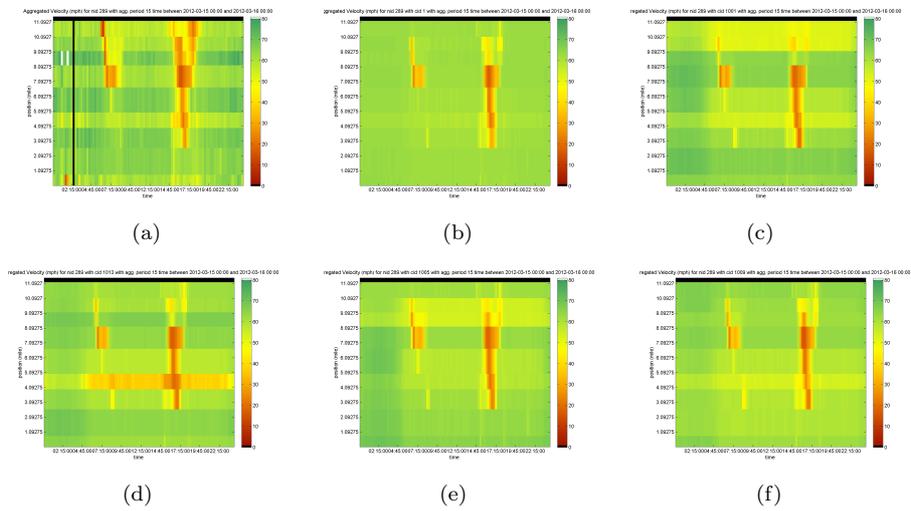
**Figure C.10:** Shows the space mean speed for the calibration methods and the ground truth from running the highway model on test site I-880. The aggregation period is 15 minutes over the time period from 2012-03-12 00:00-24:00. C.10(a) shows the ground truth. C.10(b) use the ad-hoc calibration. C.10(c) and C.10(e) use the fundamental diagrams from version III with the forward respectively backward link assignment. C.10(d) and C.10(f) use the fundamental diagrams from version II with the forward respectively backward link assignment.



**Figure C.11:** Shows the space mean speed for the calibration methods and the ground truth from running the highway model on test site I-880. The aggregation period is 15 minutes over the time period from 2012-03-13 00:00-24:00. C.11(a) shows the ground truth. C.11(b) use the ad-hoc calibration. C.11(c) and C.11(e) use the fundamental diagrams from version III with the forward respectively backward link assignment. C.11(d) and C.11(f) use the fundamental diagrams from version II with the forward respectively backward link assignment.

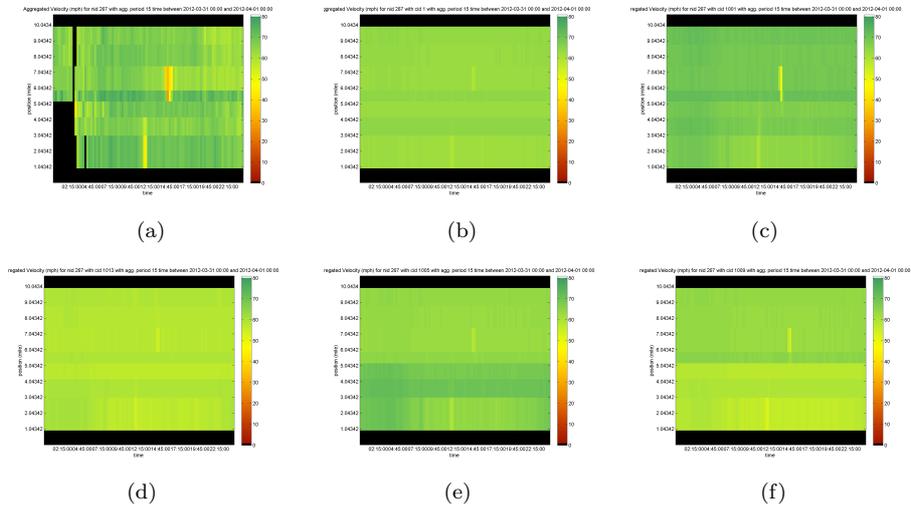


**Figure C.12:** Shows the space mean speed for the calibration methods and the ground truth from running the highway model on test site I-880. The aggregation period is 15 minutes over the time period from 2012-03-14 00:00-24:00. C.12(a) shows the ground truth. C.12(b) use the ad-hoc calibration. C.12(c) and C.12(e) use the fundamental diagrams from version III with the forward respectively backward link assignment. C.12(d) and C.12(f) use the fundamental diagrams from version II with the forward respectively backward link assignment.

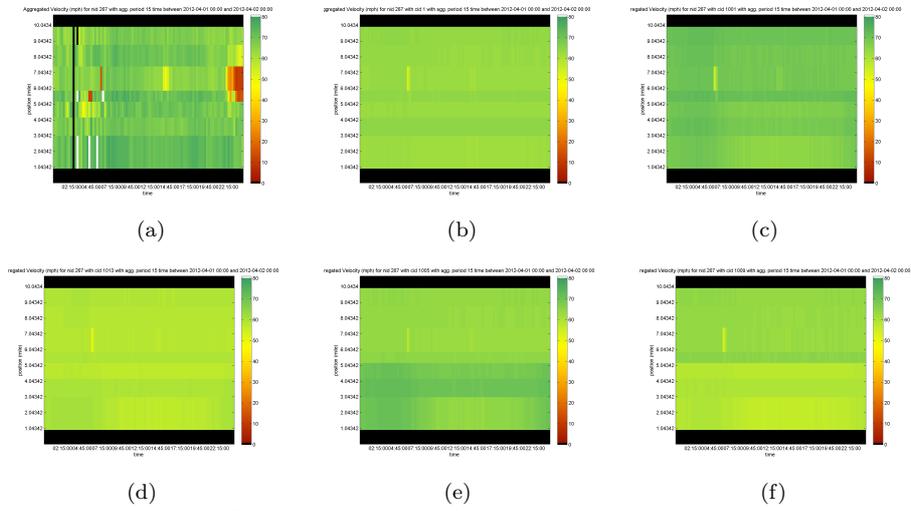


**Figure C.13:** Shows the space mean speed for the calibration methods and the ground truth from running the highway model on test site I-880. The aggregation period is 15 minutes over the time period from 2012-03-15 00:00-24:00. C.13(a) shows the ground truth. C.13(b) use the ad-hoc calibration. C.13(c) and C.13(e) use the fundamental diagrams from version III with the forward respectively backward link assignment. C.13(d) and C.13(f) use the fundamental diagrams from version II with the forward respectively backward link assignment.

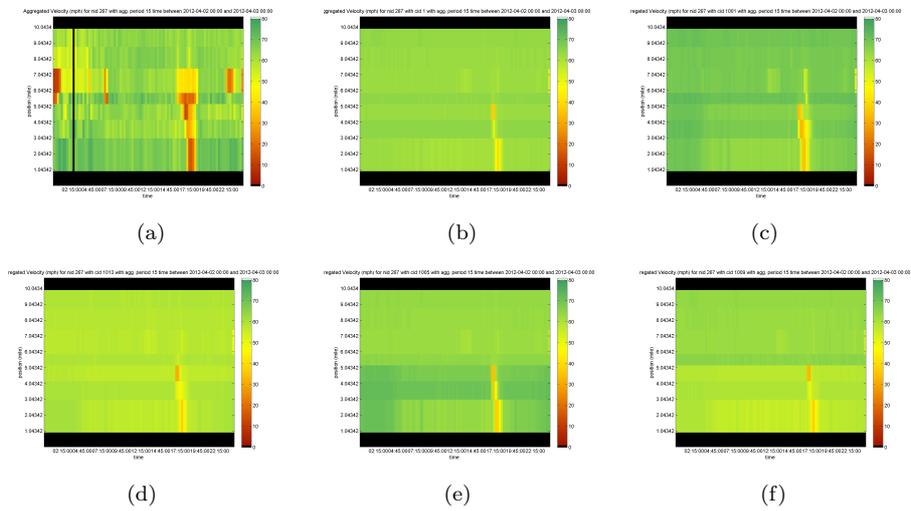
# I-15 Ontario Northbound



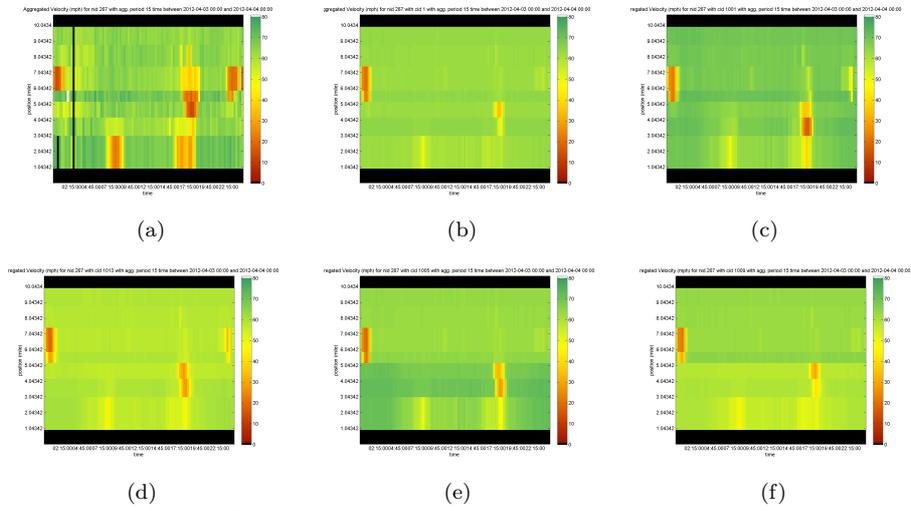
**Figure C.14:** Shows the space mean speed for the calibration methods and the ground truth from running the highway model on test site I-15 Ontario. The aggregation period is 15 minutes over the time period from 2012-03-31 00:00-24:00. C.14(a) shows the ground truth. C.14(b) use the ad-hoc calibration. C.14(c) and C.14(e) use the fundamental diagrams from version III with the forward respectively backward link assignment. C.14(d) and C.14(f) use the fundamental diagrams from version II with the forward respectively backward link assignment.



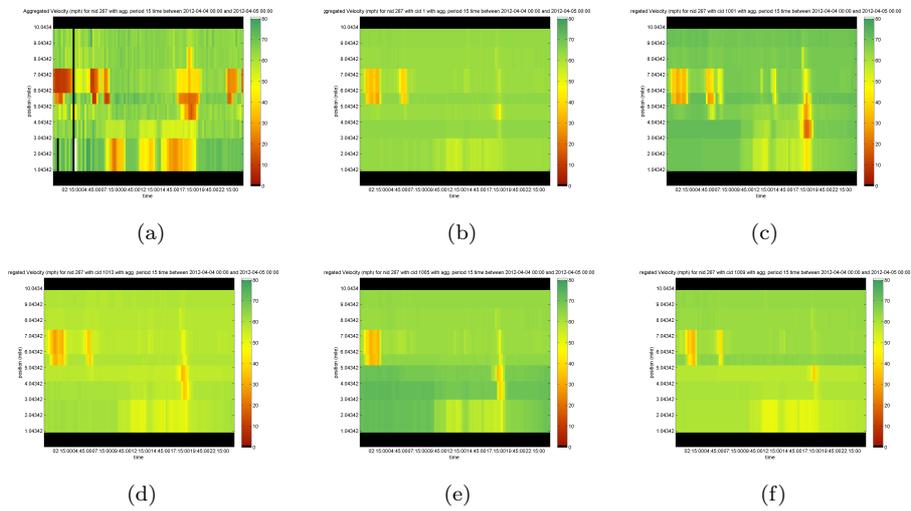
**Figure C.15:** Shows the space mean speed for the calibration methods and the ground truth from running the highway model on test site I-15 Ontario. The aggregation period is 15 minutes over the time period from 2012-04-01 00:00-24:00. C.15(a) shows the ground truth. C.15(b) use the ad-hoc calibration. C.15(c) and C.15(e) use the fundamental diagrams from version III with the forward respectively backward link assignment. C.15(d) and C.15(f) use the fundamental diagrams from version II with the forward respectively backward link assignment.



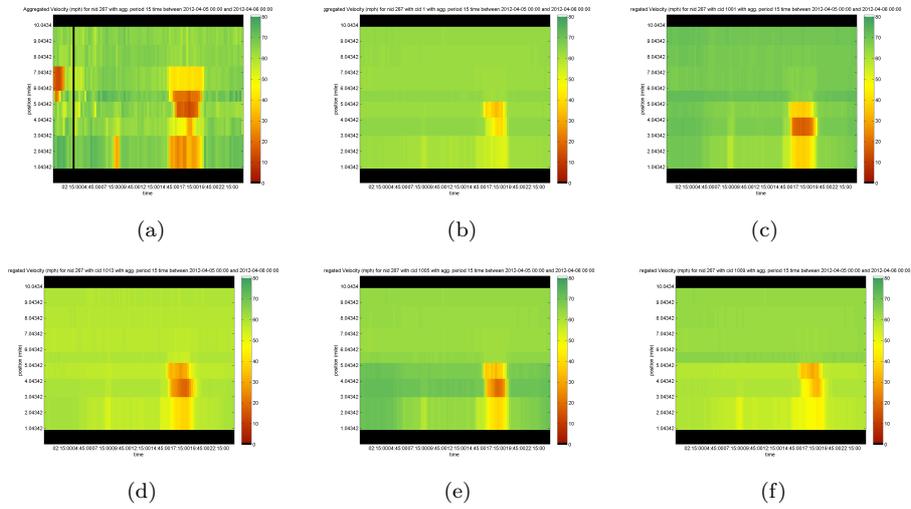
**Figure C.16:** Shows the space mean speed for the calibration methods and the ground truth from running the highway model on test site I-15 Ontario. The aggregation period is 15 minutes over the time period from 2012-04-02 00:00-24:00. C.16(a) shows the ground truth. C.16(b) use the ad-hoc calibration. C.16(c) and C.16(e) use the fundamental diagrams from version III with the forward respectively backward link assignment. C.16(d) and C.16(f) use the fundamental diagrams from version II with the forward respectively backward link assignment.



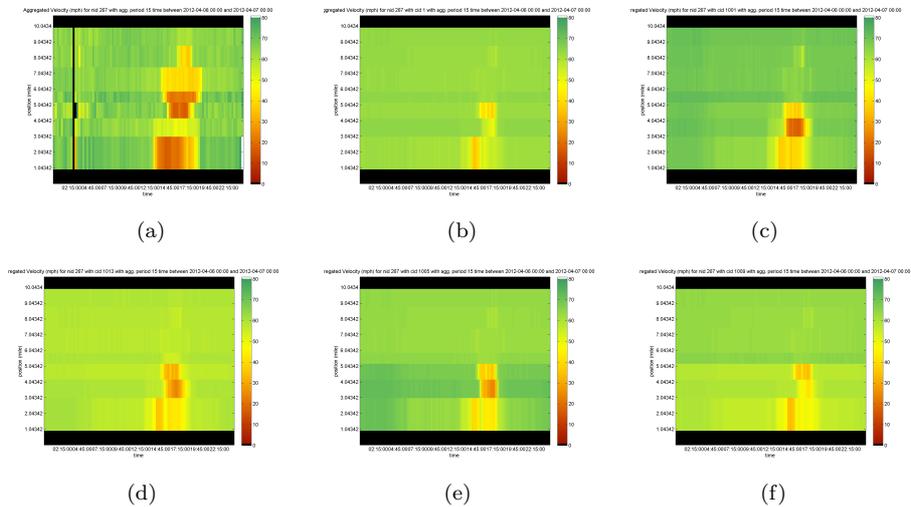
**Figure C.17:** Shows the space mean speed for the calibration methods and the ground truth from running the highway model on test site I-15 Ontario. The aggregation period is 15 minutes over the time period from 2012-04-03 00:00-24:00. C.17(a) shows the ground truth. C.17(b) use the ad-hoc calibration. C.17(c) and C.17(e) use the fundamental diagrams from version III with the forward respectively backward link assignment. C.17(d) and C.17(f) use the fundamental diagrams from version II with the forward respectively backward link assignment.



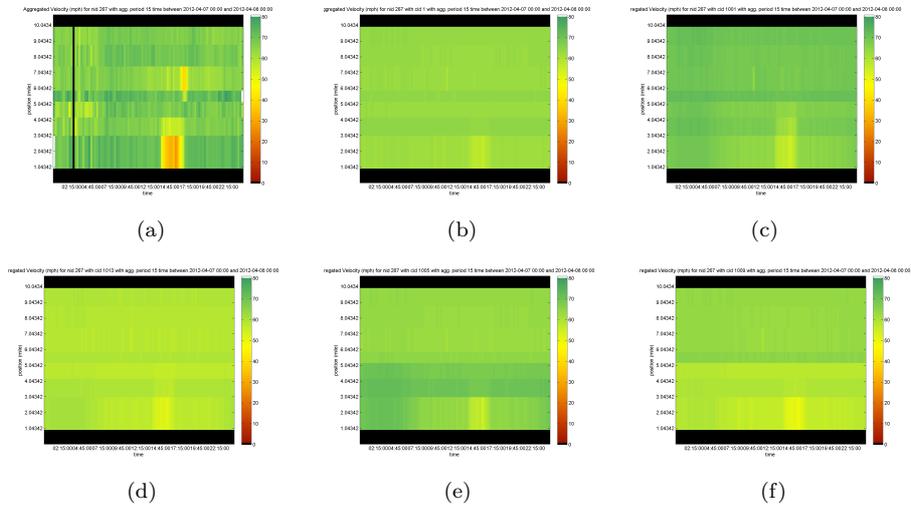
**Figure C.18:** Shows the space mean speed for the calibration methods and the ground truth from running the highway model on test site I-15 Ontario. The aggregation period is 15 minutes over the time period from 2012-04-04 00:00-24:00. C.18(a) shows the ground truth. C.18(b) use the ad-hoc calibration. C.18(c) and C.18(e) use the fundamental diagrams from version III with the forward respectively backward link assignment. C.18(d) and C.18(f) use the fundamental diagrams from version II with the forward respectively backward link assignment.



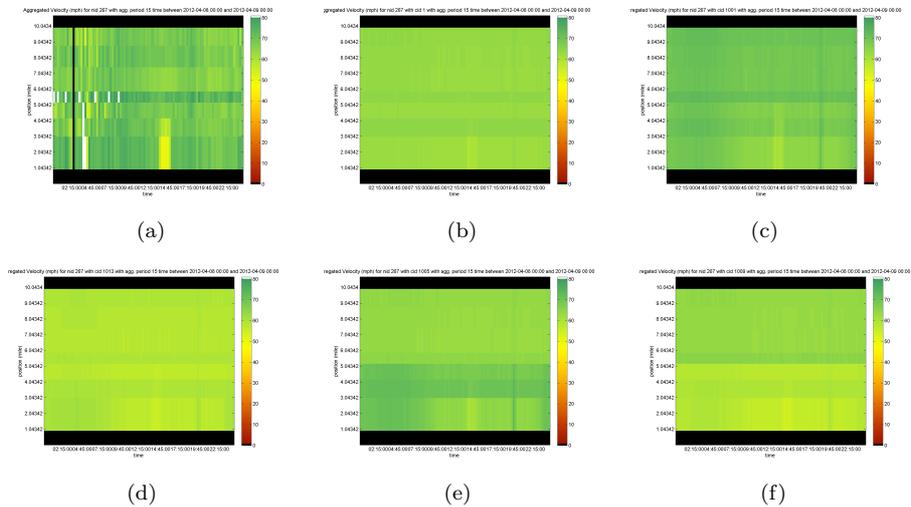
**Figure C.19:** Shows the space mean speed for the calibration methods and the ground truth from running the highway model on test site I-15 Ontario. The aggregation period is 15 minutes over the time period from 2012-04-05 00:00-24:00. C.19(a) shows the ground truth. C.19(b) use the ad-hoc calibration. C.19(c) and C.19(e) use the fundamental diagrams from version III with the forward respectively backward link assignment. C.19(d) and C.19(f) use the fundamental diagrams from version II with the forward respectively backward link assignment.



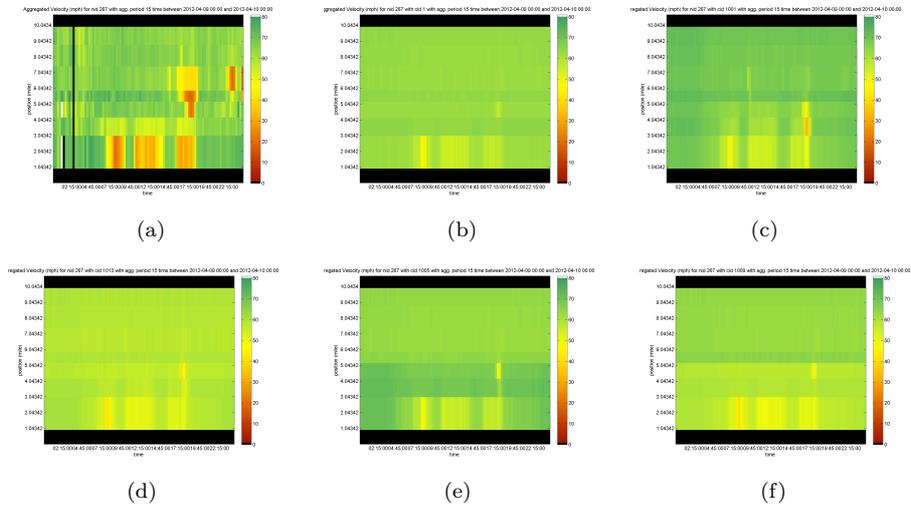
**Figure C.20:** Shows the space mean speed for the calibration methods and the ground truth from running the highway model on test site I-15 Ontario. The aggregation period is 15 minutes over the time period from 2012-04-06 00:00-24:00. C.20(a) shows the ground truth. C.20(b) use the ad-hoc calibration. C.20(c) and C.20(e) use the fundamental diagrams from version III with the forward respectively backward link assignment. C.20(d) and C.20(f) use the fundamental diagrams from version II with the forward respectively backward link assignment.



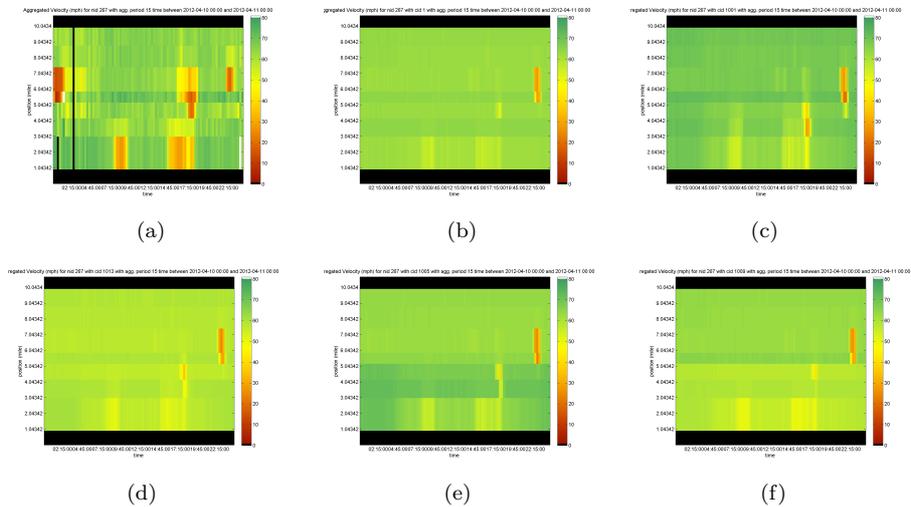
**Figure C.21:** Shows the space mean speed for the calibration methods and the ground truth from running the highway model on test site I-15 Ontario. The aggregation period is 15 minutes over the time period from 2012-04-07 00:00-24:00. C.21(a) shows the ground truth. C.21(b) use the ad-hoc calibration. C.21(c) and C.21(e) use the fundamental diagrams from version III with the forward respectively backward link assignment. C.21(d) and C.21(f) use the fundamental diagrams from version II with the forward respectively backward link assignment.



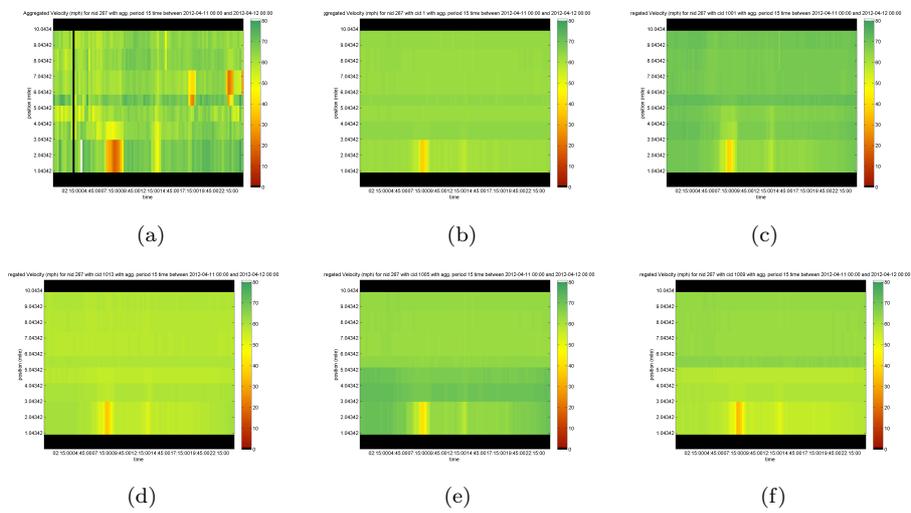
**Figure C.22:** Shows the space mean speed for the calibration methods and the ground truth from running the highway model on test site I-15 Ontario. The aggregation period is 15 minutes over the time period from 2012-04-08 00:00-24:00. C.22(a) shows the ground truth. C.22(b) use the ad-hoc calibration. C.22(c) and C.22(e) use the fundamental diagrams from version III with the forward respectively backward link assignment. C.22(d) and C.22(f) use the fundamental diagrams from version II with the forward respectively backward link assignment.



**Figure C.23:** Shows the space mean speed for the calibration methods and the ground truth from running the highway model on test site I-15 Ontario. The aggregation period is 15 minutes over the time period from 2012-04-09 00:00-24:00. C.23(a) shows the ground truth. C.23(b) use the ad-hoc calibration. C.23(c) and C.23(e) use the fundamental diagrams from version III with the forward respectively backward link assignment. C.23(d) and C.23(f) use the fundamental diagrams from version II with the forward respectively backward link assignment.



**Figure C.24:** Shows the space mean speed for the calibration methods and the ground truth from running the highway model on test site I-15 Ontario. The aggregation period is 15 minutes over the time period from 2012-04-10 00:00-24:00. C.24(a) shows the ground truth. C.24(b) use the ad-hoc calibration. C.24(c) and C.24(e) use the fundamental diagrams from version III with the forward respectively backward link assignment. C.24(d) and C.24(f) use the fundamental diagrams from version II with the forward respectively backward link assignment.



**Figure C.25:** Shows the space mean speed for the calibration methods and the ground truth from running the highway model on test site I-15 Ontario. The aggregation period is 15 minutes over the time period from 2012-04-11 00:00-24:00. C.25(a) shows the ground truth. C.25(b) use the ad-hoc calibration. C.25(c) and C.25(e) use the fundamental diagrams from version III with the forward respectively backward link assignment. C.25(d) and C.25(f) use the fundamental diagrams from version II with the forward respectively backward link assignment.