

UNIVERSITY OF TWENTE.

Faculty of Engineering
Civil Engineering & Management

dat  mobility

A model-based approach for state estimation for networks

Orestis Giamarelos
M.Sc. Thesis
January 2018

M.Sc. Thesis Committee:

prof. dr. ir. E. C. van Berkum
dr. ir. L. J. J. Wismans
ir. O. A. L. Eikenbroek
ir. L. C. W. Suijs

Title: **A model-based approach for traffic state estimation for networks**
Master Thesis

Author: **Orestis Giamarelos**
University of Twente
Master's programme: Civil Engineering & Management
Specialization: Transport Engineering & Management
Student number: s1615904
E-mail: orestis@giamarelos.de

Colloquium date: January 31st, 2018

Summary

The aim of this thesis is to develop a model that is able to estimate the traffic state of a network (focusing on urban networks) in real time, by taking into account and fusing traffic data from various sources (e.g. VLOG and Floating Car Data) as they arrive. Most approaches in traffic state estimation focus on freeways and mainly use one source of traffic data. In addition, methodologies designed for use in urban networks usually incorporate a simple node model instead of full modelling of junctions, although the influence of junctions on the traffic state in urban networks is significant. The model developed in this thesis uses Streamline by DAT.Mobility as the process model, which includes detailed modelling of junctions. Data fusion of flow (VLOG) and speed (FCD) data is achieved using Extended Kalman Filtering (EKF).

The model is validated through a series of tests using artificial data (measurements and “ground truth”). Initially, only one uncertainty parameter is enabled in each test, followed by combinations of parameters and finally a last test with all uncertainty parameters enabled. The model could satisfactorily estimate the state variables (densities and speeds) of all links of the network, as the average error values of all tests were at an acceptable level. Estimation improves over time, as the model “learns” through the fusion of measurements. An increasing uncertainty level of the system leads to a reduced rate of improvement. Estimation of the speeds proved to be more accurate in most validation tests.

The tests also indicated the uncertainty parameters that mostly determine the model’s performance. Inaccuracy of the measurements (uncertainty and error on the measurements) is the parameter with the highest impact on the results, followed by the inaccuracy of the fundamental diagram parameters and the inaccuracy of the OD matrix. Combinations of these parameters in most (but not all) cases led to an increased combined impact on the state estimation.

Analysis of the most problematic links in each test showed that an inaccurate OD matrix proved to have a higher impact on the estimation of the densities of links connected to a centroid (destination/origin). In addition, in tests with higher uncertainty it was observed that state estimation of links situated on parts of the network where no measurements were available did not improve over time, underlining the importance of the number of measurement points, as well as their position on the network. Among the fundamental diagram parameters, the free flow speed has the most significant impact on the estimation of the speeds and consequently the densities.

The ability of the developed model to calibrate the fundamental diagram parameters has also been tested. The model succeeds in improving the estimate of the free flow speed over time but does not perform equally well in the estimation of the other fundamental diagram parameters (speed at capacity and capacity per lane). The improvement in the estimation of the free flow speed proved to be independent of the initial values set, meaning that even the values starting further off the ground truth value improve over time and eventually converge to the ground truth value. The standard deviations which control the variation allowed per time step for each fundamental diagram parameter could be increased in order to achieve faster convergence of the free flow speed. However, increasing the standard deviation value for a parameter would lead to more nervous behavior of the estimated value

of the free flow speed, possibly leading to a higher average error over time. Finally, the inclusion of a warm-up phase, to allow the parameters to approach their actual values, is recommended when possible.

Regarding implementation of the method, it was observed that changing the order of fusing the measurements had practically no impact to the state estimation. In addition, setting reasonable thresholds for the state variables proved to be particularly important, as it can prevent the estimation from diverging, especially in cases with high uncertainty.

Finally, future research suggestions include researching the problem of handling latency and using a different definition for the junction modelling factors, in order to improve fusion of the flow measurements. Moreover, application on a more congested network is recommended, in order to determine the model's ability to estimate the speed at capacity and capacity per lane, as well as adding other parameters to the state vector, such as the turning fractions, to be corrected by the Kalman filter.

In memory of Georgios-Marios

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisors:

- *Professor Eric van Berkum* and *Associate Professor Luc Wismans*, whose guidance to concretizing the research topic, giving ideas for deeper analysis and finding workarounds whenever my research had reached a dead end was invaluable.
- *Oskar Eikenbroek*, who devoted a significant amount of his precious time for several months to help me out and often stayed until late at the University working with me to troubleshoot and find bugs in my code. I am grateful for your constant support!
- *Leon Suijs*, who was always available at the company to answer my questions and help me out. During our scheduled weekly meetings, his valuable feedback always helped my research go further.

Thank you all for your understanding, as well as your insightful comments and suggestions all these months!

A big thank you to *DAT.Mobility/Goudappel Coffeng* (and *Luc Wismans* again) for offering me this graduation internship, allowing me to conduct research on a topic of my interest in a great, modern workplace. My thanks also extend to my *colleagues* at DAT.Mobility. A very special thanks to *Jeroen van Oorspronk* for devoting so much of his time answering my Streamline and programming questions and creating for me custom Streamline versions with various features whenever needed.

My thanks also extend to the *CTS group* at the University of Twente for their friendship and support all these months. It was great working and hanging out with all of you!

Last but not least, I can't thank enough my *parents* and my loving *wife*. Nothing would have been possible without their love and support.

Table of contents

1. Introduction	1
2. Problem Description	3
3. Research objective and question	6
4. Method	8
4.1. Introduction	8
4.2. Process model	8
4.3. Full state vector	11
4.4. Full transition model	13
4.5. Junction modeling	14
4.6. Measurement functions	17
4.7. Data fusion using Extended Kalman Filtering	19
4.8. Description of the program flow	21
4.9. Numerical examples	22
5. Validation	31
5.1. Verification	31
5.2. Validation	32
5.3. Test network description	33
5.4. Individual tests	35
5.5. Combinations	46
5.6. Discussion	54
6. Conclusion and future research	63
6.1. Conclusion	63
6.2. Suggestions for future research	65
7. References	67
Appendix A: Simulation settings and parameters (test network)	71
1. Standard OD Matrix	71
2. Modified OD Matrix	72
3. Initial fundamental diagram parameters set	73
4. Traffic light plans	74

1. Introduction

In the latest years, mobility and logistics have become more important than ever before, especially for economic growth. Subsequently, the significance of road transport has elevated globally. However, this ever-growing demand can often not be efficiently served, as it is in many cases much higher than supply. This imbalance between demand and supply leads to overuse of the existing road network and to widespread congestion, high emissions and noise levels (Calvert et al., 2014). In addition, congestion has proven to be extremely costly for the national economies. According to INRIX (2013), traffic congestion is expected to cumulatively cost for the years 2013-2030 \$469 billion to the French national economy, \$480 billion to the British, \$691 billion to the German and \$2.8 trillion to the U.S. economy.

In the past, the most common solution to this problem was to heighten supply levels, by building new roads or by increasing the capacity of the existing network, e.g. by adding extra lanes to existing roads (Wismans et al., 2014). Adding road capacity is no longer a simple solution, as it comes at a high cost and suitable available space is limited. Therefore, effective traffic management becomes increasingly important to maximize the efficiency of the existing road infrastructure.

To achieve the goal of effective traffic management, it is important that adequate information regarding the prevailing traffic conditions is available. Nowadays, a vast and ever-increasing amount of data sources is available, consisting of roadside sensors (e.g. induction loop detectors), in-car sensors (e.g. floating car data) or co-operative roadside and in-car sensors (e.g. automatic vehicle identification via Bluetooth). These data sources can provide information useful to traffic managers, compared to the past when significant effort had to be made to collect measurements and observations from the network. Collection of data has now become easier than ever, but the data coming from many different heterogeneous sources needs to be properly validated (e.g. identifying and removing erroneous measurements and possible systematic errors in the measuring devices) and fused to successfully assess the prevailing traffic conditions and their spatial and temporal dynamics.

The availability of relevant data contributes to and enables the application of real-time traffic management and Intelligent Transport Systems (ITS). A major requirement for these applications is the availability of complete and consistent traffic state estimation. The traffic state includes all elements that describe the prevailing traffic conditions, including parameters needed for prediction. The elements included are set according to the aim of the traffic manager and the assumptions made in the traffic flow model (e.g. fixed or variable fundamental diagram parameters). Example elements that the traffic state can include are the following: speeds, densities and flows of all links of the network, densities of external links interacting with the network (e.g. links receiving from or sending flows to the network), turning fractions, fundamental diagram parameters (parameters that define the shape of the fundamental diagram, e.g. free flow speed, speed at capacity, capacity per lane and

maximum density for the Van Aerde fundamental diagram or similar parameters for other fundamental diagram types) etc.

The current state estimates form the basis for (short-term) prediction of the future traffic state of the network. This prediction leads to informed decisions on traffic control measures to be implemented with a typical goal to prevent or minimize congestion (Yuan et al., 2014). The importance of an accurate traffic state estimation is, thus, highlighted by the fact that an inaccurate estimation will propagate through the decision support model, leading to an incorrect prediction of the (near-)future traffic state and to wrong conclusions regarding the appropriate traffic control measures to be used, as the short-term prediction of the future traffic state is a component of major importance for traffic control and ITS (Van Lint & Van Hinsbergen, 2012).

The accurate estimation of the traffic state is not an easy task. Required data is only partially available, in terms of location and time, through sporadically installed sensors (e.g. induction loop detectors) in the road network or through content service providers (e.g. floating car data). The inaccuracy of sensor data available is an additional issue. Significant research has been conducted on the subject and various methodologies have been suggested to improve traffic state estimation. It must be noted, however, that the majority of related work deals with relatively simple freeway networks. In urban networks, complexity of the problem rises significantly due to various factors. For example, the intersections (regulated by traffic signals or unregulated) and the delay they impose to traffic, as well as the calculation of the fraction of traffic flow turning to each direction, are major factors that further increase the complexity of the problem. Other factors that further complicate the problem in urban networks are the following: the higher number of routing options available from origin to destination for each trip, the variation of road users and travel motives, the higher variation in the speeds of the vehicles, the availability of on-street parking and parking garages within the network, as well as the relatively low density of sensors installed in urban networks.

The goal of this research is to develop a methodology for accurate traffic state estimation for networks, focusing on urban networks. In chapter 2 the problem is described in detail and the research objective and question is presented in chapter 3. Chapter 4 contains an extensive and detailed description of the developed method and chapter 5 the validation of the method using artificial data. In chapter 6, the conclusions of this research, as well as suggestions for future research, are provided.

2. Problem Description

As mentioned in the introduction, the term *traffic state estimation* refers to the estimation of all traffic flow variables necessary to reproduce the traffic conditions on a link or a network, based on available traffic data. Estimation methods include data-driven methods, which use basic statistics, historical data and interpolation based on the sensor data, as well as model-based methods, which also take traffic flow dynamics into account (Yuan, 2013). While the data-driven methods are less complex and computationally expensive than model-based methods, the latter are more accurate and offer the opportunity for full traffic state estimation (e.g. including the fundamental diagram parameters, which are impossible to physically measure), as well as better cope with non-regular traffic conditions. These significant advantages lead to the decision to follow a model-based approach in this research.

A model-based traffic state estimation method consists of three elements (Yuan et al., 2014):

1. **A process model**, used to predict the state variables (e.g. density, speed, flow). Most widely used models are the first-order Lighthill–Whitham–Richards (LWR) (Lighthill & Whitham, 1955 and Richards, 1956) and the Cell Transmission Model (CTM) (Daganzo, 1994 & 1995), which is an operationalization of the LWR model. Also used, especially in freeway traffic state estimation, are higher order Payne-type (Payne, 1971) and METANET (Papageorgiou, 1990 and Wang & Papageorgiou, 2005) models.
2. **An observation model (measurement functions)**, used to compute the expected measurement values to be received from the sensors, considering the uncertainties of the process (e.g. density-flow and density-speed relationships) and the measurements (e.g. due to inefficiencies of the sensors) (Nantes, 2016).
3. **A data assimilation/fusion technique**, used to estimate the most probable traffic state by combining the process model predictions, the measurements received from the sensors and the expected measurement values computed by the observation model. There are various data assimilation techniques, employing either simple or more sophisticated algorithms. An example of a simple algorithm is the Newtonian relaxation (nudging) method (Anthes, 1974), which relaxes system models towards observations, thus omitting the necessity for an observation model to perform data assimilation. A popular more sophisticated data assimilation algorithm is the Kalman filtering (KF) method (Kalman, 1960 and Kalman & Bucy, 1961) in various forms, such as the Unscented Kalman Filter (UKF), Ensemble Kalman Filter (EnKF) and especially the more widely used form, which is the Extended Kalman Filter (EKF) (see e.g. Tampère & Immers, 2007, Van Lint et al., 2008, Wang & Papageorgiou, 2005 and Wang et al. 2008).

In Figure 1, a schematic procedure describing traffic state estimation with a recursive data assimilation method (such as Kalman Filtering) is provided. The method consists of prediction and correction steps, based on the measurements received from the sensors in

each time step. The process model describes the evolution of the system state (e.g. density, speed, fundamental diagram parameters). The observation model relates the system state to the observations. The aim of the data assimilation method is to make an optimal estimation of the system state using all new measurements (observations) that were made available between the previous and the current time step.

For each discrete time step, first a prediction of the system state (z_t^-) is made, based on the process model and the last available estimate. Based on this prediction of the state, the observation model predicts the values of the measurements expected to be received from the sensors. In the correction step that follows, the predicted system state is corrected with an optimal weighting factor, proportional to the error (E_t) between the predicted measurement values and the actual measurement values received from the sensors. The optimal weighting factor is determined in terms of minimizing state estimation errors. The corrected state estimate (z_t^+) is the “*belief*” of the actual traffic state, as it is the result of the combination of the process model and the actual measurements. This procedure iteratively provides state estimates at each time step and it is presented in detail in chapter 4.

Most approaches in traffic state estimation use one source of traffic data/observations (Nantes et al., 2016). A possible reason is the complexity of fusing data from different types of sensors into one model: Each data source can possibly have different spatio-temporal resolutions (e.g. Eulerian/Lagrangian coordinates). Additionally, it could be difficult to integrate various types of sensor data (e.g. speed, flow, travel time, etc.) into the flow model due to nonlinearities in traffic flow (Wang & Work, 2013). However, as Van Lint & Hoogendoorn (2009) testify, fusing data from multiple sources, when properly performed, leads to a more accurate and robust traffic state estimation.

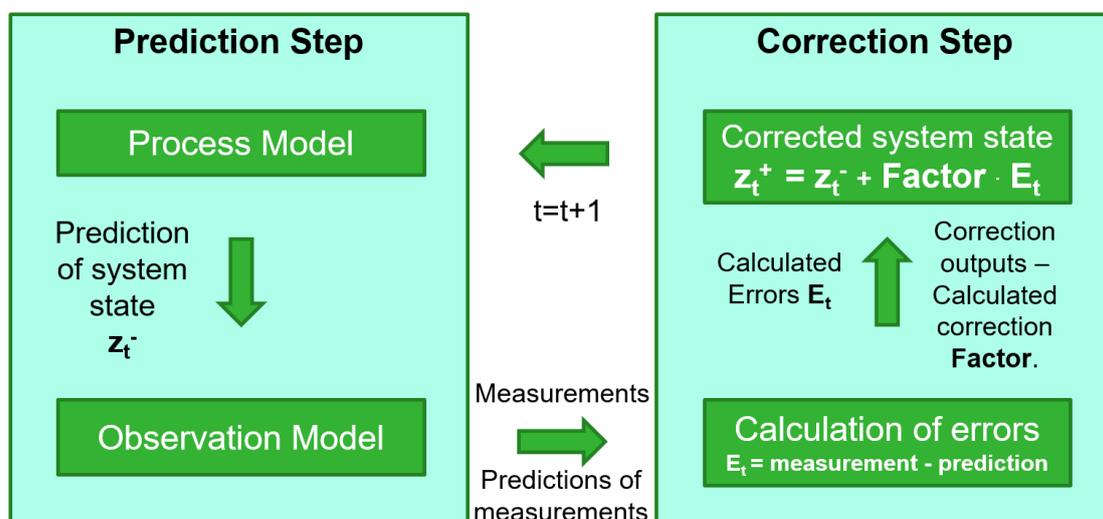


Figure 1. Schematic representation of the prediction-correction data assimilation method. (Based on Yuan, 2013)

There are various data sources providing traffic data measurements. The most common are the induction loop detectors, which are roadside sensors that provide flow data. However, there are not enough induction loop detectors installed, especially in urban networks, to rely upon and deliver a full traffic state estimation. The main reason is the high installation and maintenance cost for an adequately dense sensor network. The data they provide is also not entirely reliable, as they are also prone to measurement errors (e.g. Briedis & Samuels, 2010 and Martins, 2008). Errors may occur either due to malfunctioning (e.g. in Herrera et al. (2010) it is mentioned that in California 30% of the 25,000 installed induction loop detectors does not work properly) or due to the nature of the sensor. For example, in a dual induction loop detector setup, a vehicle approaches and changes lane, passing over only one of the two loop detectors of that lane, resulting in an erroneous measurement. Additionally, a single induction loop detector setup requires additional assumptions to be made, e.g. for the average vehicle length. Therefore, uncertainty for the speed measurement values provided by such a setup is higher.

Very common is also the use of Floating Car Data (FCD), which provides instantaneous speed data through GPS-equipped vehicles that transmit their position and speed. The GPS (in)accuracy of 6m (Owens, 1996) or 7.8m with a 95% confidence level (U.S. Department of Defense, 2008) is a drawback of FCD. This accuracy level can be improved using various methods, such as map matching. Other methods offering improved accuracy exist, e.g. Differential GPS and RTK-GPS (Real Time Kinematics) which can offer a typical accuracy of 1.5m and 2cm respectively (van de Pijpekamp, 2015) or an accuracy range of 1-5m and 1-10cm respectively (Jiménez et al, 2016). However, these methods would require additional equipment to be installed both in-vehicle (e.g. a special antenna) and in the network (reference stations every 100km and 10km respectively). The cost of acquiring floating car data is another drawback, as it is sold by traffic data providers.

Another data source is Bluetooth vehicle identification, which can provide travel times between two specific points in the network. It has the downsides of low penetration rates, high cost of equipment, the uncertain shape and length of the scanning radius, which is based on the surrounding environment (Bhaskar & Chung, 2013) and affects the detection of the vehicles and the measurements (Nantes, 2016).

A less common data source is video image processing (cameras) and automatic number-plate recognition (ANPR). Both sources come at a high cost because of the equipment cost and the fact that they are computationally heavy processes.

Finally, historical traffic data is also being used in practice in various forms. Such examples are the use of initial fundamental diagram, historical link data for every time step (flow, speed, occupancy), historical OD patterns/matrices (Wismans et al., 2014 and van der Vlist et al., 2016).

3. Research objective and question

The **main objective** of the proposed research is to develop a model that will be able to estimate the traffic state of a network (focusing on urban networks) in real time, by taking into account and fusing data arriving from various heterogeneous sources (e.g. VLOG and FCD data) as they arrive. The focus is decided to be on urban networks, as most developed approaches on traffic state estimation focus on freeways (e.g. van Lint & Hoogendoorn, 2009, Wang & Papageorgiou, 2005 and Treiber & Kesting, 2009). Few methodologies, such as the methodology by Nantes et al. (2016), have dealt with urban networks, but still with limitations, such as the lack of a junction model.

An **additional objective** is to observe if the model is able to cope with situations affecting supply of infrastructure, e.g. a reduced free flow speed due to fog covering a part of a large network or a reduced capacity on a link due to an incident. Therefore, the additional objective is to observe if the *fundamental diagram parameters*, the parameters that determine the shape of the fundamental diagram, can also be accurately estimated by the model.

The **main research question**, deduced from the main objective, is the following:

How can a model be developed to provide online estimation of the traffic state of an urban network, taking into account measurements from sensors?

An **additional research question**, covering the additional objective, is the following:

How accurately can this developed model estimate the fundamental diagram parameters of each link of the network?

The **key terms** of the research questions are the following:

The *traffic state* is defined as the densities and speeds of each link, which lead to the estimation of the flows as well, assuming homogeneity of traffic. As the additional research question also requires the fundamental diagram parameters to be estimated, the traffic state has to include these parameters as well.

The term *sensors* refers to measurements coming from sensors installed either roadside (e.g. directional flows derived from VLOG data) or onboard the vehicles (e.g. link speeds derived from FCD). As sensors are not perfect, sensor reliability is also a factor that must be considered in the development of the method.

The term *online* refers to the fact that the measurement data used in the model should not be available beforehand for pre-processing, but should be read from the sensors in real time. In addition, the model calculations should be fast enough to allow it to work in real-time. The handling of possible latency in the measurements is not part of this research. It is

assumed that there is no latency in the measurement values received. This means that a measurement of e.g. the average directional flow for a whole minute is made available immediately, exactly at the end of the 60th second of that minute.

The fact that the methodology must be suitable for *urban* networks increases the complexity of the research problem. Urban links are more challenging to model than freeway links, due to, e.g., the more complex traffic dynamics at intersections and the existence of unregulated intersections, which add a significant amount of uncertainty to the traffic state estimation. The junctions play a very determining role in urban networks, so the developed methodology should be able to cope with junctions. Therefore, junction modelling, as part of the model prediction is essential.

Thus, in order to address this problem, the developed model should be accurate, fast enough, dependent only on the observations available until the current time step (no pre-processing of measurements required), able to accommodate all junction types, estimate supply (fundamental diagram) parameters online and handle the uncertainties of model inaccuracies and sensor reliability.

The proposed method is described in detail in the next chapter.

4. Method

4.1. Introduction

To address the solution of this problem, referring to the elements of a model-based approach for traffic state estimation, mentioned in chapter 2, and taking into account the analysis of the research questions in the previous chapter, the following decisions were made:

- Streamline, a dynamic traffic propagation model based on METANET and implemented in the transport simulation software OmniTRANS by DAT.Mobility, was selected as the **process model**. It is a validated model that has been used in the industry for years and it is more complex and accurate than, for example, implementing a simple first-order LWR model. It additionally includes the powerful junction modeling module (XSTREAM), which provides additional opportunities to incorporate high levels of detail when modeling junctions, in order to simulate the actual situation of an urban network as accurately as possible.
- The measurement functions for the **observation model** will be set up according to the equations used in the process model and the available data sources.
- Finally, the Extended Kalman Filter (EKF), coded in Matlab, is selected as the **data assimilation technique**, as it is suitable for working with non-linear equations. In addition, it is more computationally efficient than other forms, which makes it a good choice for use in large networks (Yuan, 2013), as well as online applications because of the lower computational time it requires. More specifically, a slight modification of the standard EKF algorithm is selected, the incremental EKF proposed by Nantes et al. (2016), which allows incorporating of the various heterogeneous measurements incrementally, whenever they become available, enabling the use of varying sampling rates per data source. Therefore, this methodology offers additional flexibility in relation to the setup of the sensors.

The developed solution is depicted in a flowchart (Figure 2) and the methodology is described in detail in the subchapters that follow.

4.2. Process model

A target road is defined as a set of connected links. The flow of vehicles has a pre-defined direction and it is conserved, in the sense that traffic can enter and leave a link only at the upstream and downstream nodes respectively and not in-between. When dealing with urban traffic, a node can be an intersection, either signalized or unsignalized.

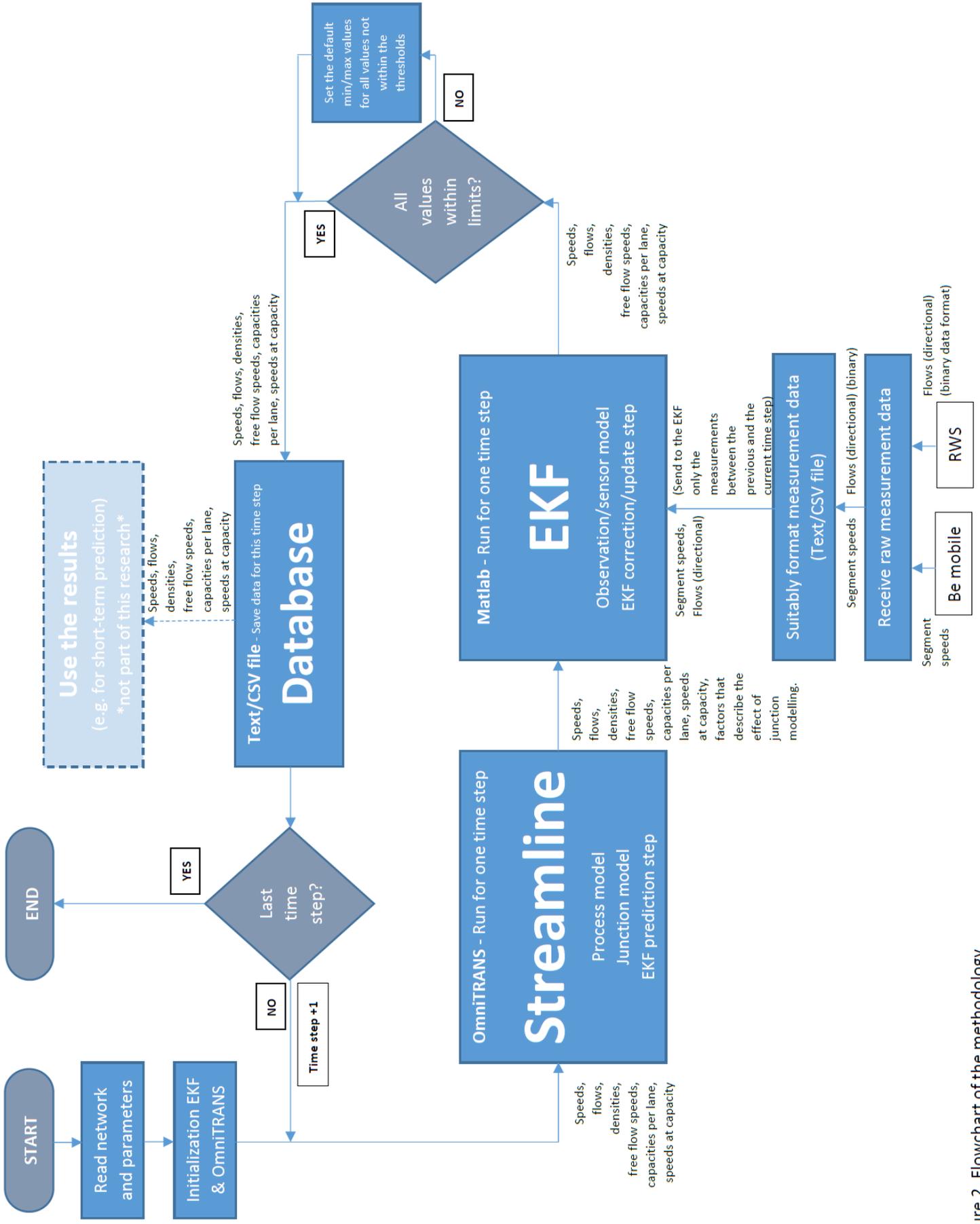


Figure 2. Flowchart of the methodology

Each link i is designed to have a distance L_i and the simulation time is divided in time steps of time T , so as to satisfy the Courant-Friedrichs-Lewy condition (Sod, 1985). This condition ensures that the distance covered within a time step, which is equal to the free flow speed (v^{free}) multiplied by the duration of the time step T , cannot exceed the length of the link ($v_i^{free} \cdot T \leq L_i$).

The equations used in the process model are the METANET (Technical University of Crete & Messmer, 2012) equations for the outflow, density propagation and speed propagation. These are presented below:

The outflow of each link i at time k is given by the equation:

$$q_i(k) = \rho_i(k) \cdot v_i(k) \cdot \lambda_i, \quad (1)$$

where

$q_i(k)$ denotes the total outflow of link i at time k ,

$\rho_i(k)$ denotes the density per lane of link i at time k ,

$v_i(k)$ denotes the speed of link i at time k and

λ_i denotes the number of lanes of link i .

The density of each link i at time $k + 1$ is given by the equation:

$$\rho_i(k + 1) = \rho_i(k) + \frac{T}{L_i \lambda_i} (q_{i-1}(k) - q_i(k)), \quad (2)$$

whose intuitive, physical meaning is that the density of the link at the next time step is the sum of the density at the current time step and the difference between the inflow to link i (outflow of the upstream link $i - 1$) and the outflow from link i .

The speed of each link i at time $k + 1$ is given by its speed at time k , plus a *relaxation term* that includes a fundamental diagram calculation $V(\rho)$, a *convection term* that expresses the change in speed caused by the inflow of vehicles and an *anticipation term* that expresses the speed decrease caused by a density increase downstream. The relevant equation is the following:

$$v_i(k + 1) = v_i(k) + \frac{T}{\tau} (V(\rho_i(k)) - v_i(k)) + \frac{T}{L_i} v_i(k) (v_{i-1}(k) - v_i(k)) - \frac{vT}{\tau L_i} \frac{\rho_{i+1}(k) - \rho_i(k)}{\rho_i(k) + \kappa} \quad (3)$$

where τ , v and κ are model parameters.

The fundamental diagram calculation $V(\rho)$, which is included in (3), expresses the speed-density relationship. Streamline uses by default the Van Aerde fundamental diagram, but it is possible to use other fundamental diagram types as well. Because the use of the Van Aerde fundamental diagram led to the forming of very long and complex equations and even more complex Jacobians, it was decided to switch to the simpler METANET fundamental diagram, by using the appropriate setting in Streamline. The form of this fundamental diagram is shown in Figure 3.

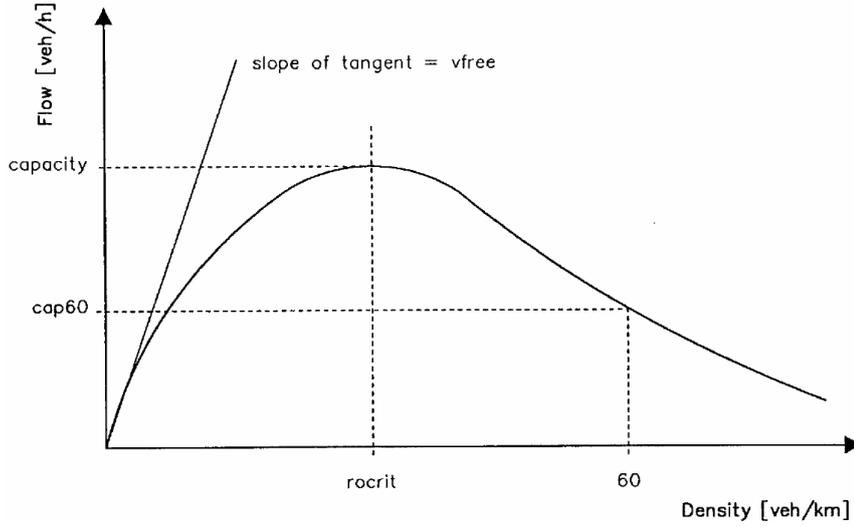


Figure 3. The METANET fundamental diagram.

The equation of the METANET fundamental diagram is the following:

$$V(\rho_i(k)) = v^{free} \cdot \exp \left[-\frac{1}{\alpha_m} \cdot \left(\frac{\rho_i(k)}{\rho_{crit,i}(k)} \right)^{\alpha_m} \right], \quad (4)$$

where ρ_{crit} signifies the critical density, which can be expressed as:

$$\rho_{crit} = \frac{f^{cap}}{v^{cap}}, \quad (5)$$

where f^{cap} signifies the capacity per lane and v^{cap} signifies the speed at capacity.

The term α_m is defined as:

$$\alpha_m = -\frac{1}{\ln \left(\frac{f^{cap}}{v^{free} \cdot \rho_{crit}} \right)} \quad (6)$$

and using (5):

$$\alpha_m = -\frac{1}{\ln \left(\frac{v^{cap}}{v^{free}} \right)} \quad (7)$$

Therefore, the fundamental diagram equation (4), using (5) and (7) becomes:

$$V(\rho_i(k)) = v^{free} \cdot \exp \left[\ln \left(\frac{v^{cap}}{v^{free}} \right) \cdot \left(\frac{\rho_i(k) \cdot v^{cap}}{f^{cap}} \right)^{-\frac{1}{\ln \left(\frac{v^{cap}}{v^{free}} \right)}} \right] \quad (8)$$

4.3. Full state vector

In order to form the full state vector, the approach and the annotations used in Wang & Papageorgiou (2005) will be followed.

The full state vector x has the following general form:

$$x = (l, d, p), \quad (9)$$

where l signifies the vector of link variables, d the vector of boundary variables and p the vector of fundamental diagram parameters.

The flow $q_i(k)$ for every link i at time k can be calculated from (1) by replacing the values of $\rho_i(k)$ and $v_i(k)$ from equations (2) and (3). Therefore, the independent link variables for each link i are ρ_i and v_i .

Consequently, the vector of the link variables for a network consisting of N links has $2N$ elements in total and the following form:

$$l = (\rho_1, v_1, \rho_2, v_2, \dots, \rho_N, v_N) \quad (10)$$

For the calculation of $\rho_i(k+1)$ from equation (2) and the calculation of $v_i(k+1)$ from equation (3), the variables $q_{i-1}(k)$, $v_{i-1}(k)$ and $\rho_{i+1}(k)$ are also required to be available for all links. These values can be calculated for all links, except the links at the edges of the network, which have a centroid either as an origin or a destination of the link. Therefore, for all centroids, through which traffic enters and exits the network, the variables $q_{i-1}(k)$, $v_{i-1}(k)$ and $\rho_{i+1}(k)$ need to be provided as well, in order to enable the estimation of the link variables for all links of the network. The vector of the boundary variables for C centroids has the following form:

$$d = (q_1^{origin}, v_1^{origin}, \rho_1^{destination}, \dots, q_C^{origin}, v_C^{origin}, \rho_C^{destination}) \quad (11)$$

In Streamline, the convection and anticipation terms of the speed equation (3) containing the variables $v_{i-1}(k)$ and $\rho_{i+1}(k)$ respectively, are omitted from the equation when the upstream link is an origin and the downstream link is a destination respectively. Therefore, the variables that give the upstream speed of an origin and the downstream density of the destination, are omitted from (11).

The flow entering the network from the origins is modeled with a simple queue model. It mainly depends on the demand, which is read from the OD matrix/matrices Streamline uses to route traffic in the network for the duration of the simulation. The OD matrices are estimated from a variety of data sources such as home and roadside interviews, historical traffic data and observed link volumes. As matrix estimation is beyond the scope of this thesis, it is assumed that the provided OD matrices for the test network are accurate. Therefore, the origin flows can be considered as input to the system, provided by Streamline every time step, and not a state variable.

Thus, no boundary variable is part of the state vector in the developed model, as the upstream speed of the origins and the downstream density of the destinations are omitted in Streamline, while the flow of the origins is input to the system. Therefore, $d = \emptyset$.

Finally, the fundamental diagram calculation in (8) requires the values of additional parameters, which determine the shape of the fundamental diagram. The three parameters required for the fundamental diagram calculation are the free flow speed (v^{free}), the capacity per lane (f^{cap}) and the speed at capacity (v^{cap}). In a large network, the values of the fundamental diagram parameters may greatly vary between links. For example, the network could consist of major urban links (speed limit 50 km/h) and minor urban links (speed limit 30 or 20 km/h). For such a mixed network, the values of the three fundamental diagram parameters between all links of the network would greatly differ. In addition, the

fundamental diagram parameters could be affected by various factors, such as adverse weather conditions, which could affect differently the values of the fundamental diagram parameters of each link. For example, fog could be covering only a part of a large network, causing a significant decrease of the free flow speed to the links of that area, while in other areas of the network there could be no or very little decrease of the free flow speed due to the fog. Therefore, it is considered important to set different fundamental diagram parameters for each link and include them in the state vector. The vector of fundamental diagram parameters for a network consisting of N links has $3N$ elements in total and the following form:

$$p = (v_1^{free}, \dots, v_N^{free}, f_1^{cap}, \dots, f_N^{cap}, v_1^{cap}, \dots, v_N^{cap}) \quad (12)$$

Based on the analysis above, the full state vector x for the model consists of $5N$ elements and it is presented in (13):

$$x = (\rho_1, v_1, \dots, \rho_N, v_N, u_1^{free}, \dots, u_N^{free}, f_1^{cap}, \dots, f_N^{cap}, v_1^{cap}, \dots, v_N^{cap}) \quad (13)$$

4.4. Full transition model

By replacing equation (1) into equation (2), we receive the following equation for the density:

$$\rho_i(k+1) = \rho_i(k) + \frac{T}{L_i \lambda_i} (\rho_{i-1}(k) \cdot v_{i-1}(k) \cdot \lambda_{i-1} - \rho_i(k) \cdot v_i(k) \cdot \lambda_i) + \xi^q(k) \quad (14)$$

The density equation is modelled exact, as it describes the conservation of vehicles. However, the term $\xi^q(k)$ is added after incorporating the flow equation (1), to reflect the modelling inaccuracy of the approximate flow equation.

Similarly, the term $\xi^v(k)$ is added to the empirical speed equation (3) to reflect the modelling inaccuracy of this equation. The resulting equation is the following:

$$v_i(k+1) = v_i(k) + \frac{T}{\tau} (V(\rho_i(k)) - v_i(k)) + \frac{T}{L_i} v_i(k) (v_{i-1}(k) - v_i(k)) - \frac{vT}{\tau L_i} \frac{\rho_{i+1}(k) - \rho_i(k)}{\rho_i(k) + \kappa} + \xi^v(k) \quad (15)$$

Therefore, the transition model for the link variables can be written in a compact form as follows:

$$l(k+1) = g(x(k)) + \xi(k) \quad (16)$$

Or using vectors:

$$\begin{bmatrix} \rho_i(k+1) \\ v_i(k+1) \end{bmatrix} = \begin{bmatrix} \rho_i(k) \\ v_i(k) \end{bmatrix} + \begin{bmatrix} \frac{T}{L_i \lambda_i} (\rho_{i-1}(k) \cdot v_{i-1}(k) \cdot \lambda_{i-1} - \rho_i(k) \cdot v_i(k) \cdot \lambda_i) \\ \frac{T}{\tau} (V(\rho_i(k)) - v_i(k)) + \frac{T}{L_i} v_i(k) (v_{i-1}(k) - v_i(k)) - \frac{\nu T}{\tau L_i} \frac{\rho_{i+1}(k) - \rho_i(k)}{\rho_i(k) + \kappa} \end{bmatrix} + \begin{bmatrix} \xi^q(k) \\ \xi^v(k) \end{bmatrix} \quad (17)$$

The fundamental diagram parameters are modeled with a random walk strategy, meaning that the current value of a variable is composed of the past value plus an error term defined as zero-mean Gaussian white noise:

$$p(k+1) = p(k) + \xi(k) \quad (18)$$

Or using vectors:

$$\begin{bmatrix} v^{free}(k+1) \\ f^{cap}(k+1) \\ v^{cap}(k+1) \end{bmatrix} = \begin{bmatrix} v^{free}(k) \\ f^{cap}(k) \\ v^{cap}(k) \end{bmatrix} + \begin{bmatrix} \xi^{vfree}(k) \\ \xi^{fcap}(k) \\ \xi^{vcap}(k) \end{bmatrix} \quad (19)$$

All ξ terms in equations (14)-(19) denote zero-mean Gaussian white noise processes. They are defined as:

$$\xi^* = \mathcal{N}(0, \sigma^{*2}), \quad (20)$$

where σ^* denotes the standard deviation of link variable/fundamental diagram parameter $*$. The standard deviation for each variable must be set according to the typical time variations expected to be observed in the respective variables/model parameters to be tracked. A higher noise standard deviation for an estimated model parameter indicates a parameter that is more time variant. Thus, a lower standard deviation value would lead to slower convergence of the parameter estimates, while a higher standard deviation value would lead to more nervous behavior (larger fluctuations) of the parameter estimates.

4.5. Junction modeling

Signalized and unsignalized junctions are integral elements of urban networks. Therefore, for a model that can be applied to urban networks, handling junctions is essential. OmniTRANS incorporates XSTREAM, a powerful junction modeling module, which can be used to model signalized junctions and roundabouts, as well as uncontrolled or sign-controlled junctions. It calculates the average delay for each turning movement based on the junction layout, turning flows and signal settings (for signalized junctions).

XSTREAM adds extra turning links to the network with properties that affect traffic propagation accordingly. The data received as output is practically limited to the outflow, speed and density of these extra turning links.

The complex calculations involved and the fact that it is hard to obtain calculation data from the XSTREAM module lead to the need to follow another approach to capture the effects of junction modelling on the state variables without implementing all the calculations in the model.

The workaround to this problem is to sum up the effect of junction modeling using appropriate factors, by exploiting the information obtained from Streamline for the state of the next time step (second). Two points are of interest for the state estimation problem:

1. The possible reaching of an outflow limit on the link upstream the junction, as a result of the traffic light ahead and possibly congestion spillback from the link downstream the turn.
2. The inflow that enters the link downstream the turn, which is affected by the turn delay imposed by junction modeling, as well as the turning fractions.

Therefore, if it is possible to get the information regarding these two points for every time step, the actual details of the calculations made within the junction modeling module are not necessary.

The process used for receiving these factors that sum up the effects of junction modeling is presented below:

1. A factor that summarizes the effect of the first point (using the term “outflow limit factor”) is received by examining the Streamline state of the next time step. If the flow is not equal to the product of density, speed and the number of lanes, as in equation (1), then an outflow limit because of either the traffic light or congestion downstream has been reached. So, the outflow limit factor will be equal to 1, unless an outflow limit has been reached, in which case the factor will be less than 1 and calculated as:

$$\alpha_i(k) = \frac{q_i(k)}{\rho_i(k) \cdot v_i(k) \cdot \lambda_i'} \quad (21)$$

where α_i signifies the outflow limit factor.

The flow equation (1) is modified to include the outflow limit factor, to reflect the limited outflow when an outflow limit has been reached, taking the following form:

$$q_i(k) = \rho_i(k) \cdot v_i(k) \cdot \lambda_i \cdot \alpha_i \quad (22)$$

2. The second point affects the inflow of the links downstream the turn. The inflow of this link is the sum of the outflows of the extra turning links that enter this link. The outflows of the extra turning links can be expressed using factors that relate them to the outflow of their upstream link. Consider the junction in the example below:

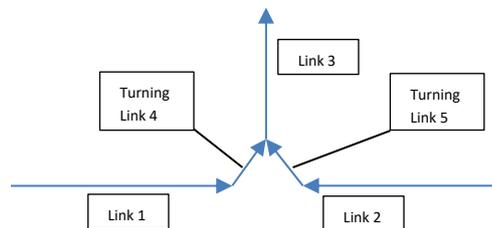


Figure 4. Example junction.

In this example, the inflow of link 3 will be equal to the sum of the outflows of turning links 4 and 5. However, these can be expressed using a factor that relates them to the outflows of links 1 and 2 respectively. With the Streamline state of the next time step

known, we can calculate these factors and end up with an equation that only contains standard network links (no turning links) and incorporates the effects of junction modeling.

Therefore, for the example junction of Figure 4:

$q_3^{in} = q_4^{out} + q_5^{out} = \beta_4 \cdot q_1^{out} + \beta_5 \cdot q_2^{out}$, where the factors β are calculated as $\beta = \frac{q_{turning\ link}^{out}}{q_{network\ link}^{out}}$. So, in this example, $\beta_4 = \frac{q_4^{out}}{q_1^{out}}$ and $\beta_5 = \frac{q_5^{out}}{q_2^{out}}$. The values of the factors are calculated using the Streamline state of the next second (simulation time step), which is known, and their values are replaced in the equation which will be used for this time step.

Thus, the outflows of the turning link j , which is located downstream network link i , can be expressed as:

$$q_j^{out}(k) = \rho_i(k) \cdot v_i(k) \cdot \lambda_i \cdot \alpha_i \cdot \beta_j \quad (23)$$

These modified flow equations (22, 23) are also used in all equations that include flows, for example in the density equation (2), as well as in the flow measurement functions that will be presented in the next subchapter.

The advantages of using this approach are the following:

- Full XSTREAM settings and features can be enabled and any combination of options for turning delays and traffic light timings may be used.
- The dependencies between the previous and next links of the network are preserved.
- Significantly easier implementation than attempting to implement the full XSTREAM module in the developed model.

The main disadvantage is that the factors need to be recalculated every time step and their values must be updated in all equations and their jacobians. This costs some additional running time, but this process is still faster overall, as with this solution the equations used in the model remain simpler and are therefore faster to calculate.

It has to be noted that the method described calculates the factors using data of the last two consecutive seconds, which leads to capturing the effect of junction modelling on the exact second the EKF is applied. However, for technical reasons it was decided to use average minute data for the calculation of the junction modelling factors. This approach solved a technical problem but led to an inaccuracy in the calculations, which is in most cases negligible, except in cases of sharp increases/decreases of the density.

Especially on links directly downstream an intersection, where more than one links would send flow to a leaving link from the junction, it was observed that on extremely congested situations this link would not accept any flow, so the outflow of the upstream (turning) links would be zero. The reason for this is that the density of the leaving link has already

exceeded the maximum density value, so this behavior is reasonable, as a precautionary measure to allow the density to fall back to acceptable levels (less than the maximum density). However, the initial use of per second data to calculate the factors would in that case possibly be using an outflow of zero, which would lead to the calculation of a factor with a value of zero and eventually zeroing out the whole flow equation and producing wrong estimates of the state variables. To prevent this problem from occurring, it was decided to use average data of the whole minute for the calculation of the junction modelling factors. While the effect of this averaging of data over a minute does not lead to an observable difference in uncongested conditions, it does cause a slight temporary difference, especially to the estimation of the density when congestion is forming and dissolving rapidly, due to e.g. a sharp increase/decrease in demand. When an increase in density takes place, the “outflow limit factor” is relatively underestimated (depending on how sharp the increase is), because the average density of 1 minute is a value, which is lower than the density of the last second that should have been used instead. Therefore, an underestimated factor for the same flow will lead to an overestimated density, in order for the flow equation to be valid. The opposite effect is observed, but to a significantly lower extent, in case of an equally sharp dissolving of congestion.

4.6. Measurement functions

With the measurement functions h , the expected value of the measurements from the sensors are expressed, based on the state of the system and the system input. The measurement functions generally consist of the predicted values, adding the uncertainties of the process and the measurement.

The **speed measurements** are given from Floating Car Data. The speed measurement function for link i is given by the speed plus measurement noise:

$$h_{v_i}(k) = v_i(k) + \gamma_i^v(k), \quad (24)$$

where γ_i^v denotes speed measurement noise which is modeled as zero-mean Gaussian white noise, similarly to the ξ values in (20).

The **flow measurements** are available from VLOG data at the stop line of the regulated intersections. The VLOG data gives the outflow of the turning links, so the relevant flow measurement function is derived from (23):

$$h_{q_j^{out}}(k) = \rho_i(k) \cdot v_i(k) \cdot \lambda_i \cdot \alpha_i \cdot \beta_j + \xi^q(k) + \gamma_j^q(k), \quad (25)$$

where $\xi^q(k)$ reflects the modeling inaccuracy of the flow equation, as previously mentioned, and $\gamma_j^q(k)$ denotes the flow measurement noise which is modeled as zero-mean Gaussian white noise.

Due to the way the factors are defined, an overall outflow limit factor is calculated for the outflow of each network link. This method could be problematic for the entering links to a

junction, where the flow could be limited e.g. only to one direction, due to congestion downstream that direction. In this case, the use of this common factor to each directional flow separately would be wrong, as the actual outflow limit factor to the congested direction should have an even lower value, while the flow limit factors to the other directions could even be 1 (totally unrestricted), depending on the traffic conditions downstream and the demand toward these directions.

To take into account these cases as well, it was decided to use the sum of all directional flows of the entering links to a junction (summing up the measurement values and the respective measurement functions), instead of fusing each individual directional flow measurement separately.

Therefore, for the example junction presented in Figure 5, there are measurements from VLOG data available for the outflows of turning links 4 and 5. The outflow limit factor of link 1 is calculated on the total outflow of link 1, which is divided to the turning links 4 and 5. However, as link 2 is congested and link 3 is not, and depending on the demand to each direction, it is possible that the outflow is limited to only one of these directions or that it is unevenly restricted. The solution to this problem is to sum the measurement values of the outflows of the turning links 4 and 5 and also combine the respective measurement functions, in order to form one combined measurement to fuse using the Extended Kalman Filter.

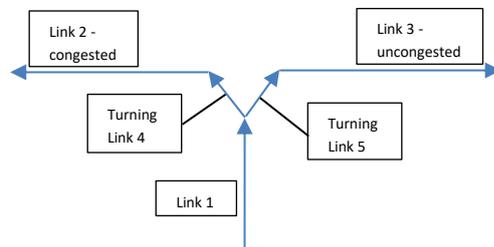


Figure 5. Example junction.

The individual measurement functions for links 4 and 5 would have been (based on equation 25) the following:

- $h_{q_4^{out}}(k) = \rho_1(k) \cdot v_1(k) \cdot \lambda_1 \cdot \alpha_1 \cdot \beta_4 + \xi^q(k) + \gamma_4^q(k)$ for link 4
- $h_{q_5^{out}}(k) = \rho_1(k) \cdot v_1(k) \cdot \lambda_1 \cdot \alpha_1 \cdot \beta_5 + \xi^q(k) + \gamma_5^q(k)$ for link 5.

The combined measurement function for the total outflow of the entering link 1 to the junction is the following:

- $h_{q_1^{out}}(k) = \rho_1(k) \cdot v_1(k) \cdot \lambda_1 \cdot \alpha_1 \cdot (\beta_4 + \beta_5) + \xi^q(k) + \gamma_4^q(k) + \gamma_5^q(k)$

Notice in the combined measurement function the participation of the process noise once (which refers to the modelling inaccuracy of the flow equation) and the measurement noise of all involved sensors (in this example, $\gamma_4^q(k)$ and $\gamma_5^q(k)$).

A drawback of this approach is that the number of flow measurements to be assimilated in every time step is now reduced. Therefore, the information the Kalman filter obtains from VLOG data is slightly deteriorated, as it comes in the form of a more complex function with more uncertainty parameters included (e.g. all measurement noise elements in one function). Therefore, capturing the errors and error covariances by the Kalman filter is with this approach more difficult, compared to an approach which would use separate flow measurement functions for each measurement.

Finally, it must be noted that the standard deviation values set for γ_j^q and γ_j^v should reflect the reliability level of the corresponding measurements and they depend on the reliability of the sensors. Their values can optionally be different per link, reflecting different accuracy levels of different sensor types that could possibly be installed throughout the network.

4.7. Data fusion using Extended Kalman Filtering

The Extended Kalman Filter has been selected as a computationally efficient data assimilation method. More specifically, the incremental Extended Kalman Filter algorithm has been selected. This slightly modified EKF algorithm, proposed by Nantes et al. (2016), is implemented to handle measurements with different sampling frequencies (e.g. loop detector data aggregated every minute while instantaneous speed data from the floating cars can be available every 15 seconds). The filter adds measurements as they are received without a state transition in-between every measurement. A precondition for this incremental addition of information from multiple measurements is the assumption that all measurements z are independent, given the current state and input vectors.

Based on the previously described formulation of the model, the equations are applied to the EKF algorithm (Nantes, 2016).

Definitions:

The state-space model consists of the following equations in compact form:

$$x(k) = f(x(k-1)) + \xi \quad (26)$$

$$z(k) = h(k, x(k)) + \gamma \quad (27)$$

Equation (26) is formed by combining (17) and (19), while equation (27) is formed by combining (24) and (25).

The process noise covariance matrix is defined as:

$$T = \text{diag}(\xi) \quad (28)$$

i.e. a matrix whose diagonal terms are the relevant variances of the state elements.

The Jacobian of the state transition function f , with respect to the full state vector x and computed at $x(k-1)$ is defined as:

$$F_{k|k-1} = \left. \frac{\partial f(k,x)}{\partial x} \right|_{x=x(k-1)} \quad (29)$$

The Jacobian of each measurement function h , with respect to the full state vector x and computed at the “a priori” estimate of the state $\hat{x}(k)$ is defined as:

$$H_k = \left. \frac{\partial h}{\partial x} \right|_{x=\hat{x}(k)} \quad (30)$$

Prediction step:

1. State estimate propagation:

$$\hat{x}(k) = f(x(k-1)) \quad (31)$$

Equation (31) gives the “a priori” estimate for the state, based on the model equations and the state of the previous simulation time step. In this setup, the state estimate is obtained directly from the Streamline simulation.

2. Error covariance propagation:

$$\Sigma(k) = F_{k|k-1} \cdot \Sigma(k-1) \cdot F_{k|k-1}^T + T, \quad (32)$$

based on the previous error covariance matrix, the Jacobians of the process model and the process noise covariance matrix.

Correction step:

The correction step is repeated for all measurements received in the current time step:

1. Measurement prediction

$$\hat{z}(k) = h(\hat{x}(k)), \quad (33)$$

where z refers to a speed or flow measurement and h refers to the relevant speed or flow measurement function, calculated at the “a priori” state estimate.

2. Kalman gain

$$K(k) = \Sigma(k) \cdot H^T(k) \cdot [H(k) \cdot \Sigma(k) \cdot H^T(k) + \xi + \gamma]^{-1} \quad (34)$$

The Kalman gain is a vector that works as a regulator between the “a priori” estimate of the state and the received measurement. Based on the “knowledge” it has accumulated in the error covariance matrix and the uncertainties of the process and measurement, it decides on which of the two values to give more weight to: the estimate or the measurement.

3. State estimate update

$$\hat{x}(k) = \hat{x}(k) + K(k) \cdot (z(k) - \hat{z}(k)) \quad (35)$$

The difference between the measurement value and the predicted measurement multiplied by the Kalman gain vector is added to the “a priori” state estimate, in order to correct the state with the results of the fusion of this specific measurement. This updated state estimate is constantly updated after each measurement of the current time step is assimilated.

4. Error covariance update

$$\Sigma(k) = (I - K(k) \cdot H(k)) \cdot \Sigma(k) \quad (36)$$

The error covariance matrix is similarly updated after each measurement is fused.

The last resulting state estimate (called the “a posteriori” state estimate) represents the updated state after fusing all measurements available in the current time step. This corrected “a posteriori” state estimate will be used as the initial state of the next time step. The same applies to the error covariance matrix as well.

Thresholds:

After all observations have been assimilated, the resulting state variables must be restricted within reasonable margins, in order to prevent the estimation from significantly diverging from the actual solution. As an example, the values of the densities need to be non-negative and lower than the maximum density value ($0 \leq \rho_i \leq k_j$). Therefore, if a density value is calculated to be, for example, 185 vehicles/km and the maximum density value is set to be 180 vehicles/km, then the calculated density value will be lowered to 180 vehicles/km.

At the end of this process, the resulting values of the state elements are retained as the state of the system for the current time step. The process continues with the next time step, using these values as the initial state for the next time step.

4.8. Description of the program flow

The method described in this chapter is programmed in Matlab and OmniTRANS (in the form of Ruby scripts). OmniTRANS runs the simulation per second for one minute, a Matlab program fuses the available measurements and feeds the updated state values back to OmniTRANS for the simulation of the next time step. The process is automated using a control character scheme, where a special character is written in a text file, which signifies to these programs when they should pause and resume running.

The process begins by running a “start job” in OmniTRANS, which sets the initial fundamental diagram parameters in the network links and then proceeds with running the first minute of the simulation. When the simulation is over, OmniTRANS creates a file containing a database dump of the state values and other parameters used in the equations (e.g. flows from the origins, downstream densities etc.) for every second of simulation and a save state file containing the state of the last and previous second of simulation. It then starts Matlab and writes a specific control character in the control text file, which signifies that OmniTRANS has finished running.

After the Matlab program is initialized and the network and the relevant equations are loaded, Matlab reads the file containing the database dump of the simulation and from this data it forms the initial state vector for the EKF and the parameter vectors required for the calculations, which include the downstream densities, the origin flows, the factors required

for junction modelling (outflow limit factors and inflow reduce factors). Using these values, the EKF algorithm runs, fusing the measurements that belong to the current time step, and the resulting updated values are checked if they are within reasonable thresholds (e.g. no value can be negative, all densities have to be lower than the maximum density etc.). If not, they are set within the defined limits.

The updated values of the densities and speeds, as well as the calculated values of the flows, using the updated values of the densities and speeds and a possible outflow limit factor, replace the relevant values for the speeds, densities and flows in the OmniTRANS save state file. The updated fundamental diagram parameters are written in a separate CSV file. Finally, a control character enabling OmniTRANS to proceed is written in the control text file.

OmniTRANS then proceeds with reading the file containing the updated fundamental diagram parameters and updates the relevant values of the links in the network. It then reads the updated save state file and proceeds with the simulation of the next minute, using the data from the save state file as the initial state.

The process continues for the desired number of time steps. After the last time step, a special control character is written in the control text file, which instructs the OmniTRANS job to stop running and the Matlab program to proceed with calculating performance indicators and designing graphs.

4.9. Numerical examples

In order to help the reader understand the developed methodology, some numerical examples are provided, where the formulation of the equations and the state vector is presented. In addition, the numerical calculations for the first time step are made, in order to illustrate the calculation process.

4.9.1. Network consisting of one link and fundamental diagram parameters estimation

For the first numerical example, we consider a network that consists of only one link, as shown in Figure 6. The link consists of one lane and its length is 1 km. For this link, a fixed inflow is considered, with a value of $q^{in}=1000$ veh/h. The time step of the simulation is set to 1 second.

The full state vector in this example contains the density (ρ_m) and speed (v_m) of the link, as well as the fundamental diagram parameters (v^{free} , f^{cap} and v^{cap}). Therefore, the full state vector is the following:

$$x = (\rho_m, v_m, v^{free}, f^{cap}, v^{cap}) \quad (37)$$

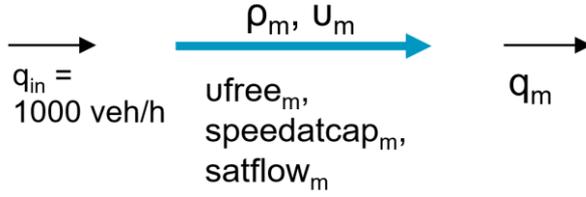


Figure 6. Network consisting of one link.

The density propagation equation is based on (14), after replacing the values for the duration of the time step T , number of lanes λ and length of the link L , as well as the term $\rho_{i-1}(k) \cdot v_{i-1}(k) \cdot \lambda_{i-1}$ with 1000, which is the value of q^{in} :

$$\rho_m(k+1) = \rho_m(k) + \frac{1}{3600} \cdot (1000 - \rho_m(k) \cdot v_m(k)) + \xi^q(k) \quad (38)$$

In order to avoid complex equations in this example, a much simpler speed equation is used, which has been randomly setup, simply making sure that it contains all fundamental diagram parameters in it:

$$v_m(k+1) = v_m(k) + \frac{f^{cap}}{v^{free} \cdot v^{cap}} + \xi^v(k) \quad (39)$$

Therefore, the full state transition model can be described as:

$$\begin{bmatrix} \rho_m(k+1) \\ v_m(k+1) \\ v^{free}(k+1) \\ f^{cap}(k+1) \\ v^{cap}(k+1) \end{bmatrix} = \begin{bmatrix} \rho_m(k) + \frac{1}{3600} \cdot (1000 - \rho_m(k) \cdot v_m(k)) \\ v_m(k) + \frac{f^{cap}(k)}{v^{free}(k) \cdot v^{cap}(k)} \\ v^{free}(k) \\ f^{cap}(k) \\ v^{cap}(k) \end{bmatrix} + \begin{bmatrix} \xi^q(k) \\ \xi^v(k) \\ \xi^{v^{free}}(k) \\ \xi^{f^{cap}}(k) \\ \xi^{v^{cap}}(k) \end{bmatrix} \quad (40)$$

The Jacobian F with respect to the state x of the state transition function f is a 5x5 matrix:

$F =$

$1 - \frac{v_m}{3600}$	$-\frac{\rho_m}{3600}$	0	0	0
0	1	$-\frac{f^{cap}}{(v^{free})^2 \cdot v^{cap}}$	$\frac{1}{v^{free} \cdot v^{cap}}$	$-\frac{f^{cap}}{(v^{cap})^2 \cdot v^{free}}$
0	0	1	0	0
0	0	0	1	0
0	0	0	0	1

(41)

The relevant measurement functions for the speed and flow are the following:

$$h_{v_m}(k) = v_m(k) + \gamma_m^v(k) \quad (42)$$

$$h_{q_m}(k) = \rho_m(k) \cdot v_m(k) + \xi^q(k) + \gamma_m^q(k) \quad (43)$$

The Jacobians of the measurement functions are the following:

$$H_v = (0, 1, 0, 0, 0) \quad (44)$$

$$H_q = (v_m, \rho_m, 0, 0, 0) \quad (45)$$

The standard deviations for the uncertainties of the process and the measurements are set to $\xi^q = \xi^v = \gamma_m^v = \gamma_m^q = 0.25$ and for the uncertainties of the fundamental diagram parameters to $\xi^{vfree} = \xi^{fcap} = \xi^{vcap} = 1$.

The process noise covariance matrix $T = \text{diag}(\xi^q, \xi^v, \xi^{vfree}, \xi^{fcap}, \xi^{vcap}) = \text{diag}(0.25, 0.25, 1, 1, 1)$.

The EKF is initialized with the following settings:

Initial state: $x(0) = (19.9, 50, 49, 1610, 34)^\top$

Initial covariance matrix $\Sigma(0) = \text{diag}(1)$

The example measurements that will be used for the first time step are the following:

$z_q(1) = 1020$ and $z_v(1) = 48$.

Prediction step

The state estimate propagation for the first time step $\hat{x}(1)$ is given by (40), by replacing the values of the initial state $x(0)$. The resulting state is:

$\hat{x}(1) = (19.9014, 50.9664, 49, 1610, 34)^\top$.

The Jacobian F calculated at $\hat{x}(1)$ is calculated from (41):

$F =$

0.9858	-0.0055	0	0	0
0	1	-0.0197	0.0006	-0.0284
0	0	1	0	0
0	0	0	1	0
0	0	0	0	1

The error covariance propagation (Σ) is given by (32):

$\Sigma =$

1.2219	-0.0055	0	0	0
-0.0055	1	-0.0197	0.0006	-0.0284
0	-0.0197	2	0	0
0	0.0006	0	2	0
0	-0.0284	0	0	2

Correction step

The application of the correction step must be repeated twice, as there are measurements of both the speed and the flow of the link.

Starting with the flow measurement, the process begins with the prediction of the measurement from the measurement functions:

$$\hat{z}q(1) = \rho_m(1) \cdot v_m(1) = 19.9014 \cdot 50.9664 = 1014.3 \text{ veh/h}$$

The Jacobian of the flow measurement function from (45) is:

$$H_q = (v_m, \rho_m, 0, 0, 0) = (50.9664, 19.9014, 0, 0, 0)$$

The Kalman gain is calculated from (34):

$$K = [0.0170, 0.0067, -0.0001, 0.000003, -0.000155]^T$$

The state estimate update is calculated from (35):

$$\hat{x}(1) = [19.9982, 51.0047, 48.9994, 1610.00002, 33.9991]^T$$

The error covariance matrix Σ is calculated from (36):

$$\Sigma =$$

0.1657	-0.4238	0.0067	-0.000203	0.0096
-0.4238	1.0855	-0.0171	0.00052	-0.0246
0.0067	-0.0171	2	0.0000013	-0.0000607
-0.000203	0.00052	0.0000013	2	0.00000185
0.0096	-0.0284	-0.0000607	0.00000185	1.9999

The same process is repeated for the fusion of the speed measurements, using the above new values for the $\hat{x}(1)$ and Σ after fusion of the flow measurement:

$$\hat{z}v(1) = v_m(1) = 51.0047 \text{ km/h}$$

$$H_v = (0, 1, 0, 0, 0)$$

$$K = [-0.2673, 0.68465, -0.010773, 0.000328, -0.015526]^T$$

$$\hat{x}(1) = [20.8014, 48.9475, 49.032, 1609.99903, 34.046]^T$$

$\Sigma =$

0.0524	-0.1337	0.0021	-0.000064	0.0030
-0.1337	0.3423	-0.0054	0.000164	-0.0078
0.0021	-0.0054	1.9998	0.000007	-0.000326
-0.000064	0.000164	0.000007	2	0.00001
0.0030	-0.0078	-0.000326	0.00001	1.9995

After fusing all measurements, the resulting value of \hat{x} is the “a posteriori” estimate of the traffic state, the corrected state for the first time step. This value of the state and the resulting value of Σ are used as the respective initial values for the next time step.

Therefore, the values $x(1) = \hat{x}(1) = [20.8014, 48.9475, 49.032, 1609.99903, 34.046]^\top$ and $\Sigma(1) = \Sigma$ are the corrected state for the first time step. The process continues with the next time step, where these values are used for the “a priori” state estimate of the second time step.

4.9.2. Network consisting of two links, no fundamental diagram parameters estimation

For the second numerical example, we consider a network consisting of two links, as shown in Figure 7. The links consist of one lane and their length is 1 km each. For the first link, a fixed inflow is considered, with a value of $q^{in}=1000$ veh/h. The time step of the simulation is set to 1 second. In this example we omit the fundamental diagram parameters and use a simpler speed equation for simplicity reasons.

The full state vector in this example contains the densities (ρ_1 and ρ_2) and speeds (v_1 and v_2) of the links. Therefore, the full state vector is the following:

$$x = (\rho_1, v_1, \rho_2, v_2) \quad (46)$$

The density propagation equations are based on (14), after replacing the values for the duration of the time step T , number of lanes λ and length of the link L , as well as the term $\rho_{i-1}(k) \cdot v_{i-1}(k) \cdot \lambda_{i-1}$ with 1000, which is the value of q^{in} for the first link:

$$\rho_1(k+1) = \rho_1(k) + \frac{1}{3600} \cdot (1000 - \rho_1(k) \cdot v_1(k)) + \xi^q(k) \quad (47)$$

$$\rho_2(k+1) = \rho_2(k) + \frac{1}{3600} \cdot (\rho_1(k) \cdot v_1(k) - \rho_2(k) \cdot v_2(k)) + \xi^q(k) \quad (48)$$

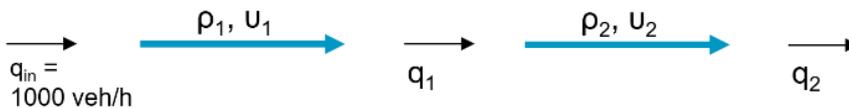


Figure 7. Network consisting of two links.

In order to avoid complex equations in this example, too, a simpler speed equation is used, keeping only the convection term. For the first link, a constant upstream speed of 50 km/h is assumed:

$$v_1(k+1) = v_1(k) + \frac{1}{3600} \cdot v_1(k) \cdot (50 - v_1(k)) + \xi^v(k) \quad (49)$$

$$v_2(k+1) = v_2(k) + \frac{1}{3600} \cdot v_2(k) \cdot (v_1(k) - v_2(k)) + \xi^v(k) \quad (50)$$

Therefore, the full state transition model can be described as:

$$\begin{bmatrix} \rho_1(k+1) \\ v_1(k+1) \\ \rho_2(k+1) \\ v_2(k+1) \end{bmatrix} = \begin{bmatrix} \rho_1(k) + \frac{1}{3600} \cdot (1000 - \rho_1(k) \cdot v_1(k)) \\ v_1(k) + \frac{1}{3600} \cdot v_1(k) \cdot (50 - v_1(k)) \\ \rho_2(k) + \frac{1}{3600} \cdot (\rho_1(k) \cdot v_1(k) - \rho_2(k) \cdot v_2(k)) \\ \rho_2(k) + \frac{1}{3600} \cdot (\rho_1(k) \cdot v_1(k) - \rho_2(k) \cdot v_2(k)) \end{bmatrix} + \begin{bmatrix} \xi^q(k) \\ \xi^v(k) \\ \xi^q(k) \\ \xi^v(k) \end{bmatrix} \quad (51)$$

The Jacobian F with respect to the state x of the state transition function f is a 4x4 matrix:

$F =$

$1 - \frac{v_1}{3600}$	$-\frac{\rho_1}{3600}$	0	0
0	$1.0139 - \frac{v_1}{3600}$	0	0
$\frac{v_1}{3600}$	$\frac{\rho_1}{3600}$	$1 - \frac{v_2}{3600}$	$-\frac{\rho_2}{3600}$
0	$\frac{v_2}{3600}$	0	$\frac{v_1}{3600} - \frac{v_2}{1800} + 1$

(52)

The relevant measurement functions for the speed and flow are the following:

$$h_{v_1}(k) = v_1(k) + \gamma^v(k) \quad (53)$$

$$h_{v_2}(k) = v_2(k) + \gamma^v(k) \quad (54)$$

$$h_{q_1}(k) = \rho_1(k) \cdot v_1(k) + \xi^q(k) + \gamma^q(k) \quad (55)$$

$$h_{q_2}(k) = \rho_2(k) \cdot v_2(k) + \xi^q(k) + \gamma^q(k) \quad (56)$$

The Jacobians of the measurement functions are the following:

$$H_{v_1} = (0, 1, 0, 0) \quad (57)$$

$$H_{v_2} = (0, 0, 0, 1) \quad (58)$$

$$H_{q_1} = (v_1, \rho_1, 0, 0) \quad (59)$$

$$H_{q_2} = (0, 0, v_2, \rho_2) \quad (60)$$

The standard deviations for the uncertainties of the process are set to $\xi^q = \xi^v = 0.25$ and for the uncertainties of the measurements to $\gamma^v = \gamma^q = 0.5$.

The process noise covariance matrix $T = \text{diag}(\xi^q, \xi^v, \xi^q, \xi^v) = \text{diag}(0.25, 0.25, 0.25, 0.25)$.

The EKF is initialized with the following settings:

Initial state: $x(0) = (19.9, 49.9, 25, 49)^\top$

Initial covariance matrix $\Sigma(0) = \text{diag}(1)$

The example measurements that will be used for the first time step are the following:

$z_{q1}(1) = 1020$, $z_{v1}(1) = 45$ for link 1 and $z_{q2}(1) = 1050$, $z_{v2}(1) = 48$ for link 2.

Prediction step

The state estimate propagation for the first time step $\hat{x}(1)$ is given by (51), by replacing the values of the initial state $x(0)$. The resulting state is:

$\hat{x}(1) = (19.9019, 49.9014, 24.9356, 49.0123)^\top$.

The Jacobian F calculated at $\hat{x}(1)$ is calculated from (52):

$F =$

0.9861	-0.0055	0	0
0	0.9862	0	0
0.0139	0.0055	0.9864	-0.0069
0	0.0136	0	0.9866

The error covariance propagation (Σ) is given by (32):

$\Sigma =$

1.2225	-0.0055	0.0136	-0.00007527
-0.0055	1.2225	0.0055	0.0134
0.0136	0.0055	1.2232	-0.0068
-0.00007527	0.0134	-0.0068	1.2236

Correction step

The application of the correction step must be repeated four times, as there are measurements of the speeds and flows of both links.

Starting with the speed measurement of link 1:

The measurement value received from the sensor is $z_{v1}(1) = 45$ km/h. The prediction of the measurement from the measurement function is $\hat{z}_{v1}(1) = v_1(1) = 49.9014$ km/h.

The Jacobian of the speed measurement function from (57) is:

$$H_{v1} = (0, 1, 0, 0, 0)$$

The Kalman gain is calculated from (34):

$$K = [-0.0028, 0.6198, 0.0028, 0.0068]^T$$

The state estimate update is calculated from (35):

$$\hat{x}(1) = [19.9155, 46.8636, 24.9220, 48.9789]^T$$

The error covariance matrix Σ is calculated from (36):

$$\Sigma =$$

1.2225	-0.0021	0.0137	-0.000038
-0.0021	0.4648	0.0021	0.0051
0.0137	0.0021	1.2232	-0.0068
-0.000038	0.0051	-0.0068	1.2235

The same process is repeated for the fusion of the speed measurement of link 2, using the above new values for the $\hat{x}(1)$ and Σ after fusion of the speed measurement of link 1:

$$z_{v2}(1) = 48 \text{ km/h}$$

$$\hat{z}v2(1) = v_2(1) = 48.9789 \text{ km/h}$$

$$H_{v2} = (0, 0, 0, 1)$$

$$K = [-0.000019, 0.0026, -0.0034, 0.6200]^T$$

$$\hat{x}(1) = [19.9155, 46.8611, 24.9254, 48.3720]^T$$

$$\Sigma =$$

1.2225	-0.0021	0.0137	-0.000015
-0.0021	0.4648	0.0021	0.0019
0.0137	0.0021	1.2232	-0.0026
-0.000015	0.0019	-0.0026	0.4650

Fusion of flow measurement of link 1, using the above new values for the $\hat{x}(1)$ and Σ :

$$z_{q1}(1) = 1020 \text{ veh/h}$$

$$\hat{z}q1(1) = \rho_1 v_1(1) = 933.2624 \text{ veh/h}$$

$$H_{q1} = (v_1, \rho_1, 0, 0) = (46.8611, 19.9155, 0, 0)$$

$$K = [0.0200, 0.0032, 0.00024, 0.00001325]^T$$

$$\hat{x}(1) = [21.6481, 47.1383, 24.9460, 48.3732]^T$$

$$\Sigma =$$

1.2225	-0.0021	0.0137	-0.000015
-0.0021	0.4648	0.0021	0.0019
0.0137	0.0021	1.2232	-0.0026
-0.000015	0.0019	-0.0026	0.4650

Fusion of flow measurement of link 2, using the above new values for the $\hat{x}(1)$ and Σ :

$$z_{q2}(1) = 1050 \text{ veh/h}$$

$$\hat{z}_{q2}(1) = \rho_2 v_2(1) = 1206.72 \text{ veh/h}$$

$$H_{q2}(1) = (0, 0, v_2, \rho_2) = (0, 0, 48.3732, 24.9460)$$

$$K = [-0.0000055, 0.000013, 0.0188, 0.0036]^T$$

$$\hat{x}(1) = [21.6490, 47.1363, 22.0018, 47.8015]^T$$

$$\Sigma =$$

0.0790	-0.1850	0.000366	-0.00071
-0.1850	0.4355	-0.00086	0.0017
0.000366	-0.00086	0.1128	-0.2181
-0.00071	0.0017	-0.2181	0.4231

After fusing all measurements, the resulting value of \hat{x} is the “a posteriori” estimate of the traffic state, the corrected state for the first time step. This value of the state and the resulting value of Σ are used as the respective initial values for the next time step.

Therefore, the vector $x(1) = \hat{x}(1) = [21.6490, 47.1363, 22.0018, 47.8015]^T$ is the corrected state for the first time step and $\Sigma(1) = \Sigma$ is the covariance matrix after the first time step. The process continues with the next time step, where these values are used for the “a priori” state estimate of the second time step and the error covariance propagation of the prediction step ($\hat{x}(2)$ and $\Sigma(2)$).

5. Validation

In this chapter, the performance of the developed model is assessed. The first step in this process is the *verification* of the model, by confirming that all equations and computations are properly implemented into the model. The second step is the *validation* of the model, which examines if the model accurately represents the real system. It is achieved through a series of tests using artificial measurements. The results of the tests are thoroughly analyzed and discussed at the end of the chapter.

5.1. Verification

The first step to be taken before proceeding with the validation of the model is the *Verification* step. In this step, the extent to which the concept described in the model has been correctly transferred to the program is examined. It includes the verification of the individual equations, as well as their correct implementation in the code. However, it must be mentioned that verification only ensures that there are no unintentional computational/implementation errors in the programmed model. It does not evaluate the accuracy of the structure or results of the model, which is assessed in the validation step that follows.

The developed code has been extensively documented and commented, describing what each part of the code intends to do and how. This practice facilitates checking of the code by the programmer him-/herself, as well as by other people and/or parties, for robustness and accuracy and makes identification of errors easier.

In the case of the developed model, which relies on both Streamline and a custom-coded Matlab program, it is additionally important to verify that the equations used in both Streamline and Matlab are the same. A possible difference between the two would add an additional error to the model results. By running the model without the use of measurements and comparing the values received from Streamline and the values calculated by the equations used in Matlab, some very minor differences in the results are observed, which are attributed to roundings taking place in the calculations. The maximum differences observed are smaller than half a thousandth of a unit (<0.0005 , e.g. if a speed calculated by Streamline is 49.9920 km/h, the relevant result received by Matlab will be between 49.9915 and 49.9925 km/h). Since the resulting differences are very small, they have practically no effect on the results and they may safely be ignored.

The main causes for these differences between the Streamline and Matlab calculations are considered to be the following:

- The major cause is the fact that the junction modelling part of Streamline is simulated using appropriate calculated factors, as described in the previous chapter. This approach inserts additional rounding to the calculations both when the factors

are calculated and when the values of e.g. speed and density are calculated using these rounded factors.

- An additional minor cause is the use of the save state functionality of OmniTRANS to exchange data with Matlab, which provides rounded numbers to the 6th decimal digit. Combined with the relatively long speed equations leads to slight, but negligible inaccuracies.

The advantage of preserving junction modelling when modelling an urban network significantly outweighs the addition of roundings in the calculations. Therefore, these slight differences are only mentioned as results of the verification process, but are practically negligible, as the resulting differences are minor.

5.2. Validation

Finding an adequate validation scheme for a model that uses real measurements is not an easy task. The reason is that there is no “ground truth” to compare the results of the model with, in order to evaluate its performance. The smaller the difference between the calculated values and the relevant “ground truth” values, the more successful the model is at representing the actual situation. The vast majority of other work in the field relies on artificial data/measurements to produce “ground truth” data, obtained e.g. from a microscopic simulation (e.g. Duret et al., 2017, Fountoulakis et al., 2017, Nantes et al., 2016, Sunderrajan et al., 2016 etc.). A microscopic simulation can provide all kinds of data, measurements and comparisons to rely upon, in order to validate the results of the model. On the other hand, when using real-life data there is no other information at the researcher’s disposal, apart from the actual measurement values, as received from the installed sensors.

The validity of the developed model can be supported theoretically, based on the fact that the model builds on already validated theories and methodologies that are widely accepted and adopted by many researchers in the field, as has already been mentioned in chapter 2. Streamline is based on the METANET model, which together with the EKF algorithm implemented in the developed Matlab program, are acknowledged and used by professionals and researchers in the field for decades. More specifically, the *incremental Extended Kalman Filtering* method implemented in the model, presented by Nantes et al. (2016), is in effect a slight modification of the EKF method. This method and the additional assumptions made for it (e.g. the independence of each measurement) are validated by the authors using a microscopic simulation to produce ground truth and sensor data to compare with, yielding satisfactory results. Therefore, it can be claimed that in terms of *face validity* (Eddy et al., 2012), the developed model is valid because it is based on already validated models and methodologies. What remains to be assessed is how this particular setup performs in practice.

The validation scheme selected for this specific model will be based on artificial measurements. The reason is that the use of artificial measurements provides the necessary controlled environment and the opportunity to examine the impact each uncertainty parameter has on the accuracy of the estimation. Therefore, the use of artificial measurements provides more options and flexibility for the validation of the method. The validation scheme, results and discussion is presented in the rest of this chapter.

The performance evaluation is achieved by calculating the calculation of the root mean square error (RMSE) and/or mean absolute percentage error (MAPE) between estimated and “ground truth” values. The root mean square error is generally calculated as follows, with y_i the “ground truth” value, \hat{y}_i the respective estimated value through the model and n the number of estimated values:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (61)$$

The MAPE is calculated as follows:

$$MAPE = \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{y_i} \quad (62)$$

An overall RMSE and/or MAPE can be calculated as the average RMSE or MAPE for all links of the network (Wismans et al., 2014).

The MAPE value gives an indication of a percentage error, which can give a first idea on its significance. The RMSE value gives an indication of the actual error value, allowing to draw a conclusion if the error is significant from a traffic engineer’s point of view. A combination of both values is required for a proper analysis. For example, a 50% MAPE for a traffic flow value on a link operating near capacity is a very unsuccessful result. However, this is not the case for the same MAPE value observed in a link with hardly any flow (Van Lint & Van Hinsbergen, 2012). A difference, e.g., of an estimate of a flow of 15 veh/h from a measurement of 10 veh/h leads to a MAPE value of 0.5 or 50%, which is very high. However, the actual impact of this error from a traffic engineer’s point of view is negligible, as the RMSE value of 5 veh/h is negligible.

From each one of the validation tests, the links displaying the highest MAPE and RMSE error values are determined and possible reasons for their behavior is presented in the discussion section at the end of the chapter.

5.3. Test network description

For the validation using artificial measurements, a test network with known conditions was set up in OmniTRANS. It was designed so as to include all possible sequences of links (link-

link, node-link, link-node, centroid-node, node-centroid), in order to be able to identify sequences for which state estimation could be particularly problematic.

The test network consists of 30 links in total (15 bi-directional links, one lane each, 50m length), 7 centroids (entries to/exits from the network), two 4-way regulated intersections and one 3-way unregulated intersection (give way) and is depicted in Figure 8. The asterisks indicate the centroids and they are numbered with the blue-colored numbers preceded by C. The intersections are depicted using green-colored numbers, preceded by U for unregulated intersections and R for regulated intersections. The unregulated intersection U1 is a “Give-way” intersection, with the priority set to the horizontal direction (1 → 3 and -3 → -1). The black-colored numbers on the right of each link direction display the link number. e.g. link 1 has a direction from left to right (from C1 to U1) and link -6 has a direction from top to bottom (from R1 to C3).

An origin-destination (OD) matrix, which contains information on the trips going from each origin to each destination, is also needed for the simulation. In other words, in the OD matrix the demand to and from each centroid is set. Instead of using a fixed OD matrix for the whole simulation period, it was decided to use two different OD matrices, one having a higher demand and the other one having a lower demand, in order to introduce variation. Simulation for minutes 1-20 and 41-60 is run using the higher-demand OD matrix, while simulation for minutes 21-40 uses the lower-demand OD matrix.

For the two regulated intersections, fixed traffic light plans were used, even though the developed methodology allows for any type of traffic light timings, either static or dynamic. The reason is that the traffic light timings together with the OD matrices are set accordingly, in order to ensure that there is congestion on specific sections of the network, while the rest of the network operates in free flow conditions. The OD matrices and traffic light plans set for the test network are presented in Appendix A1 and A4 respectively.

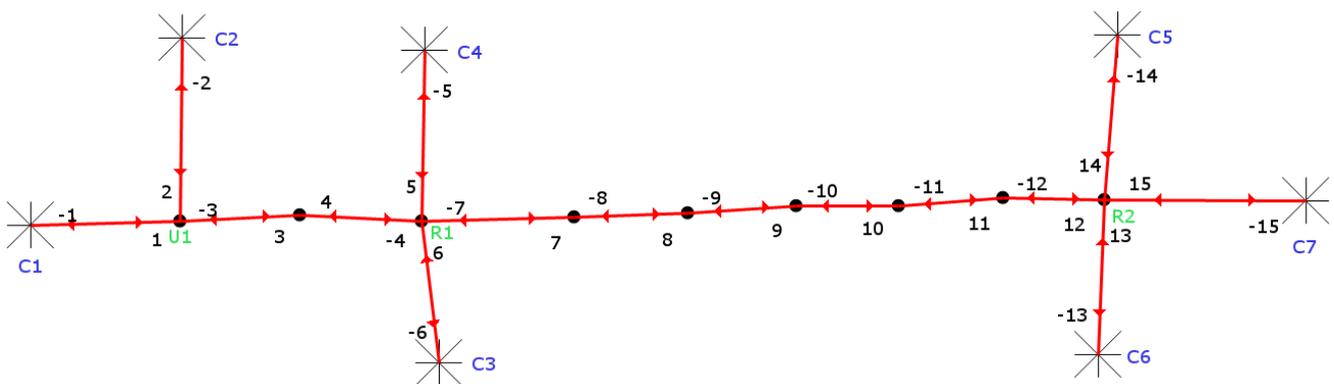


Figure 8. The test network.

The fundamental diagram parameters are set accordingly for an urban network: Free flow speed is set to 50 km/h, speed at capacity to 35 km/h and capacity per lane to 1650 veh/h. All links of the network share the same fundamental diagram parameters.

Finally, for the system to work, a minimal uncertainty value for the uncertainties for the process and measurements ($\xi^q, \xi^v, \gamma^q, \gamma^v$) is required to be set. Therefore, the default values for the standard deviations for $\xi^q, \xi^v, \gamma^q, \gamma^v$ are set to a low value (0.1). This value was set as an adequately low value that did not cause calculation problems in any of the tests. Previous tests with zero or lower values (e.g. 0.01) led to problems in all or some tests, depending on the setup.

5.4. Individual tests

The validation process begins by testing the effect each one *uncertainty parameter* has on the estimation. The uncertainty parameters in the developed methodology are the following:

- Base case with all uncertainty parameters set equal to the ground truth
- Correctness of measurement values (uncertainty about the actual values received, error on the measurements due to inefficiencies of the sensors)
- Availability of measurements
- Correctness of the OD matrix used
- Fundamental diagram parameters (correct or not, changing or not)

In each of these tests, only one uncertainty parameter applies, as all other uncertainty parameters are kept identical to the ground truth, in order to capture the effect of the examined uncertainty parameter to the state estimation.

The ground truth is created by running a one-hour simulation on the test network with the appropriate parameters, as required by the setup of each test. At the end of the simulation, the average speeds and densities per minute are collected.

The artificial measurements are created following a similar approach: After the end of the simulation, matrices containing the average flows per minute for all turning links (flow measurements) and the average speeds per minute for all network links (speed measurements) are created. Depending on the setup of each test, a smaller sub-matrix containing the values of only the parameters required by the test is constructed. When required, random noise within a given threshold ($\pm 10\%$) is added to the measurement values.

5.4.1. Test 1: Base case scenario – no uncertainty parameters enabled

This test forms the basis for comparison with the results of the tests that follow. In this test, no uncertainty parameters are enabled. Therefore, the OD matrix, fundamental diagram parameters and measurements used in the simulation are equal to the ground truth. Uncertainties for the process and measurements are set at a minimal value (0.1). Finally, speed measurements from all links of the network and flow measurements from all intersections are fused.

Ideally the MAPE and RMSE for this test would be zero. However, this is not the case because of the minimal uncertainty that is required to be set for the system to run, as well as the amount of number roundings that take place in the process. Therefore, the expected results of this test are very low MAPE and RMSE values, that can be attributed to these causes.

The results of the simulation are practically identical to the ground truth in most cases. As can be seen from the graphs in Figure 9 showing the results of the simulation for the density and speed of a random link of the network, all lines coincide (ground truth, simulation before correction and simulation after correction), as their differences are minor and practically not observable.

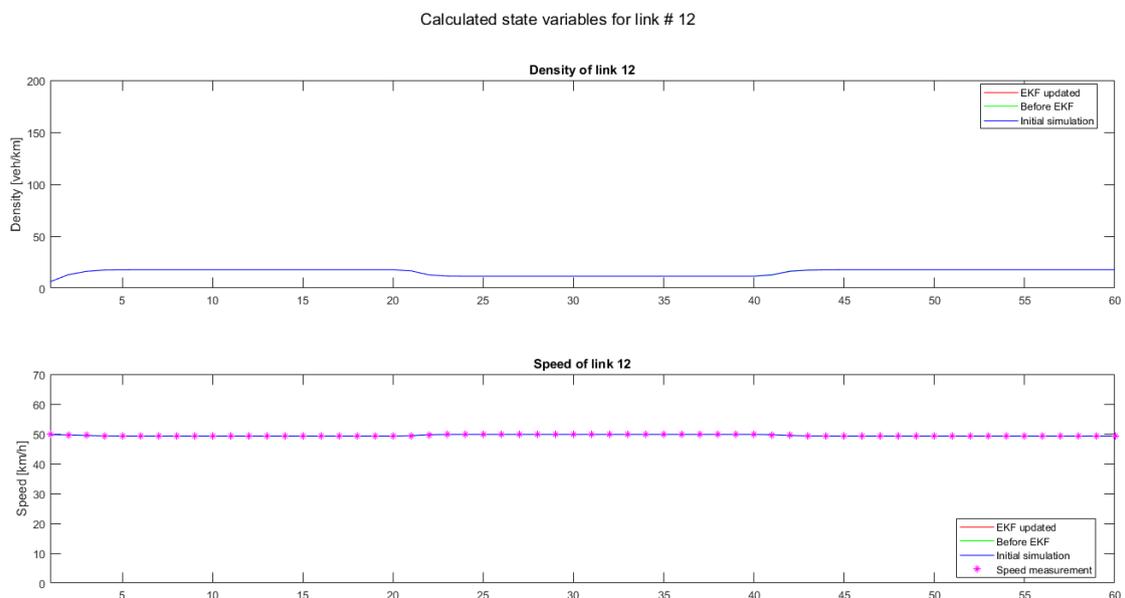


Figure 9. Calculated state variables for link #12.

An exception is link 13 shown in Figure 10, on which we have set up a very high demand to create congestion. This very abrupt increase in density occurring in this link cannot be entirely accurately followed by the model, as shown in the graph by the different red and blue lines, indicating the corrected and ground truth values respectively. The same behavior is also observed after time step 40, when, after an interval of 20 time steps with a lower-demand OD matrix, the previous higher-demand OD matrix is used again and congestion forms again. This behavior is generally observed in sharp increases or decreases particularly of the density and is thoroughly discussed in the discussion part of this chapter.

The network MAPE and RMSE values are satisfactorily low, $5.5 \cdot 10^{-4}$ and 0.0381 respectively. By calculating separate network MAPE and RMSE values for the densities and speeds, the results are 0.0011 and 0.0539 for the densities, $7.9 \cdot 10^{-6}$ and $1.5 \cdot 10^{-4}$ for the speeds.

The average RMSE values show that link 13 contributes to this RMSE more than all other links: the average RMSE value of link 13 over the whole simulation period is 0.9753, while the next highest RMSE value is 0.0176 (link -11). The same applies to the average MAPE value, where the relevant value for link 13 is 0.0076 and the next highest average MAPE is 0.0009 (links -11 and -10).

By calculating separate average MAPE and RMSE values per link, link 13 again shows the highest MAPE and RMSE values for both the densities and speeds. The average MAPE for the densities is 0.0149 (approximately 9 times higher than the second highest) and the average MAPE for the speeds is 0.0002 (approximately 30 times higher than the second highest). Similarly, the average RMSE for the densities is 1.3792 and the average RMSE for the speeds is 0.0026 (both values approximately 50 times higher than the second highest).

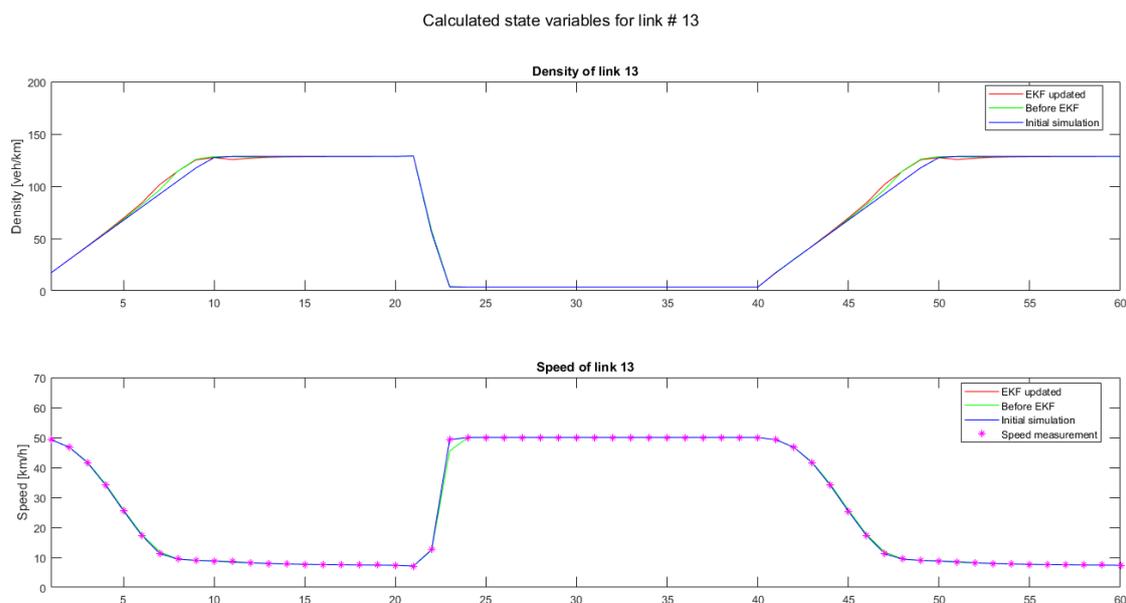


Figure 10. Calculated state variables for link #13.

The results are presented in Table 1 that follows:

Table 1: Results of test 1.

TEST 1	Network	Network (speeds)	Network (densities)	Max Average per link	Max Average per link (ν)	Max Average per link (ρ)
MAPE	$5.5 \cdot 10^{-4}$	$7.9 \cdot 10^{-6}$	0.0011	0.0076 [13]	0.0002 [13]	0.0149 [13]
RMSE	0.0381	$1.5 \cdot 10^{-4}$	0.0539	0.9753 [13]	0.0026 [13]	1.3792 [13]

5.4.2. Test 2: Error on the measurements and uncertainty values for the process and measurements

The goal of this test is to identify the impact of the uncertainty of the measurements. As the sensors are not perfect, each measurement received possibly contains an error. In all other individual tests run, artificial measurements without error were used. However, it is considered important to examine how the model works when the measurements are not perfectly accurate, because this is what is always expected in a real-case scenario.

The measurements were created with a random deviation from the ground truth of $\pm 10\%$. Uncertainty standard deviation values for the process (flow and speed) and flow measurements were set to 0.5, while the uncertainty for the speed measurements was set to 1. This selection of setting a higher uncertainty is justified by the experience working with the real data available for the application of the methodology, which showed that the available Floating Car Data was less reliable.

The results of the simulation are satisfactory, as they follow the ground truth well in most cases. The average situation, relatively common for all links of the network, is shown in Figure 11, showing the results of the simulation for the density and speed of a random link of the network. The red line, which expresses the corrected values follows well the blue line (ground truth). In the speeds section of the graph, it can also be observed that the corrected values are in the vast majority of time steps between the measurement and simulated/ground truth values. From the graphs, the differences in the values seem acceptably low.

The exception of link 13 is in this test made more clear, as the errors on the measurements are causing a higher impact in this link, as shown in Figure 12. This is discussed in more detail in the discussion subchapter at the end of this chapter.

The calculated error values are presented in Table 2.

Table 2: Results of test 2.

TEST 2	Network	Network (speeds)	Network (densities)	Max Average per link	Max Average per link (ν)	Max Average per link (ρ)
MAPE	0.0277	0.0179	0.0375	0.0428 [13]	0.0274 [13]	0.0582 [13]
RMSE	0.8257	0.8679	0.5019	2.8943 [13]	1.3490 [5]	3.7714 [13]

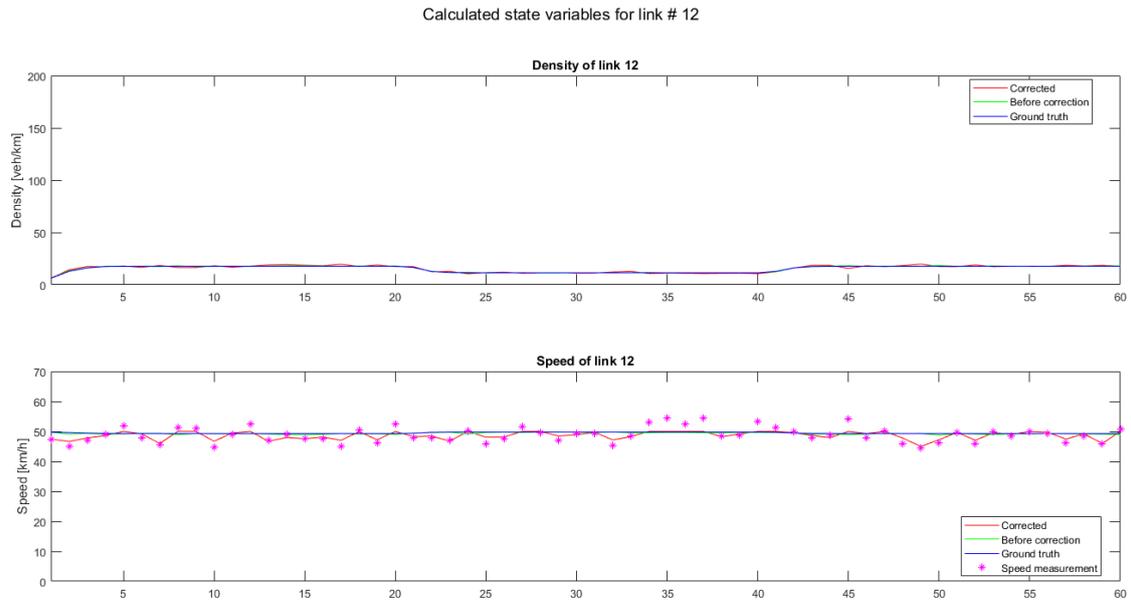


Figure 11. Calculated state variables for link #12.

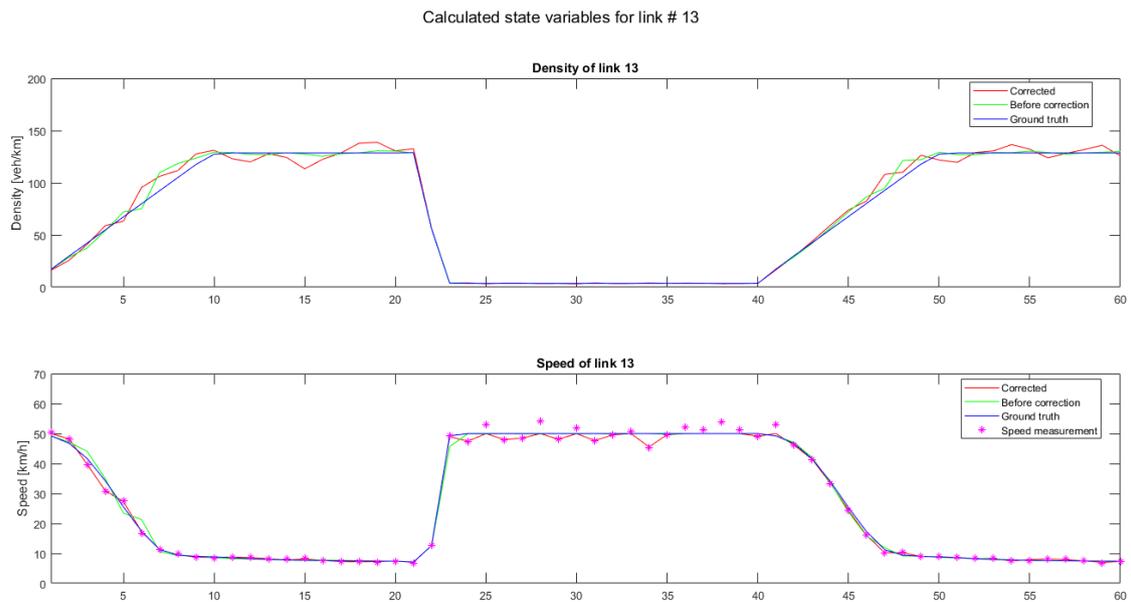


Figure 12. Calculated state variables for link #13.

The maximum values are again almost all observed on link 13. The difference is that in this second test, the MAPE values of other links as well have values that approach those of link 13. However, the average RMSE for the speeds of link 13 had one of the lowest values of all links in the network. This was expected, as link 13 is significantly more congested than all other links of the network and, thus, lower speeds are expected, as can be observed from the two graphs for links 12 and 13. Therefore, the RMSE as an indicator of the error as an absolute value, and not as a percentage, is expected to be lower.

5.4.3. Test 3: Less measurements available

The goal of such a test is to identify the impact to the estimation of the speeds and densities due to a limited availability of measurements. In the previous tests, speed measurements from all links of the network and directional flow measurements from all three intersections were used. As so many measurement points are unlikely to happen in a real-case scenario, it is important to examine if the estimation significantly deteriorates due to a lack of measurements.

In this specific setup, this test cannot offer much insight to that direction, because the measurements are derived from the ground truth, without any error on the measurements added, and they are produced using the same model used for the actual simulation. With each measurement that is fused, the minor uncertainty parameters required for the running of the model ($\xi^q = \xi^v = \gamma^q = \gamma^v = 0.1$) and the roundings of the values that take place lead to a minor error being added. Therefore, in this test we are testing this hypothesis and expect to observe a slightly lower error compared to test 1, as we are fusing less measurements and all other uncertainty parameters are not enabled. The impact of the availability of less measurements will be made apparent when combining uncertainty parameters in the next subchapter, as the availability of less measurements will offer fewer opportunities to improve estimation and “correct” the errors introduced due to other enabled uncertainty parameters.

As in the previous tests, the ground truth was created using the standard OD matrix previously mentioned and the standard fundamental diagram parameters ($v^{free}=50$ km/h, $f^{cap}=1650$ veh/h and $v^{cap}=35$ km/h). The measurements were created from these ground truth values without adding any error and only a subset of all measurements was kept. Finally, the uncertainty standard deviation values (ξ and γ) are set to their minimal values (0.1).

This test could be performed in various ways. For example, a suggestion could be to randomly remove scattered measurements or to remove speed measurements from specific links or flow measurements from whole intersections. It was decided to run this test in consistence with data that could be available in a real case. Therefore, the measurements available are speed measurements of only one of the two directions of traffic and directional flow measurements from the regulated intersections of the network (R1 and R2).

As expected, the results regarding the error values are similar to those of the first test. In the resulting graphs, no difference between the ground truth and the corrected values can practically be observed. The results for link 13, which has always displayed the greatest variation, are also significantly more satisfactory, as can be seen in Figure 13.

The calculated error values are presented in the table below:

Table 3: Results of test 3.

TEST 3	Network	Network (speeds)	Network (densities)	Max Average per link	Max Average per link (ν)	Max Average per link (ρ)
MAPE	0.0006	0.0006	0.0006	0.0090 [13]	0.0146 [13]	0.0034 [13]
RMSE	0.0149	0.0158	0.0089	0.2199 [13]	0.2742 [13]	0.1032 [13]

The maximum values are all observed on link 13. Similarly to test 1, the error values of link 13 are significantly higher than those of the other links of the network. The difference compared to the error values of the other links in the network is high, but lower than the difference observed in test 1.

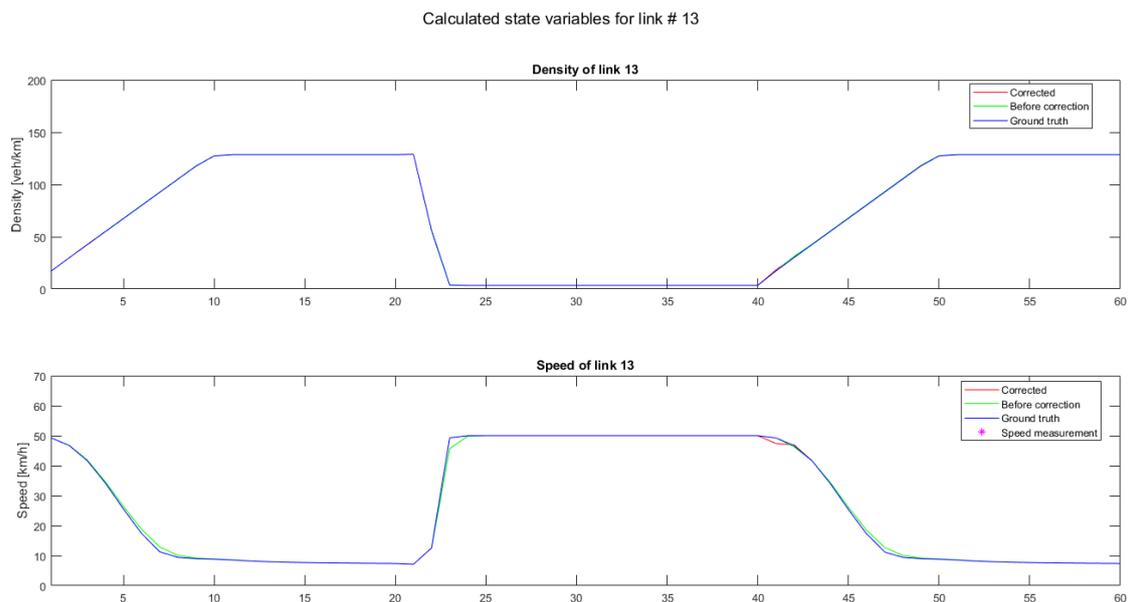


Figure 13. Calculated state variables for link #13.

5.4.4. Test 4: Different OD matrix

The goal of this test is to identify the impact a not-perfectly-calibrated OD matrix can have to the estimation of speed and density. This is also a scenario that is common in practice, as the OD matrix is set up using data from various sources and can reach high accuracy levels but it is unlikely that it can be considered an absolutely accurate representation of the ground truth. Therefore, a slight or even major deviation from the “ground truth” is always expected when working with a real network.

As in the previous test, the ground truth was created using the standard fundamental diagram parameters ($v^{free}=50$ km/h, $f^{cap}=1650$ veh/h and $v^{cap}=35$ km/h), which are kept constant for the whole simulation. The measurements were created from these ground truth values without adding any error to the relevant values. However, the values of the OD matrix that is used to produce the ground truth and measurements have been randomly altered within limits ($\pm 10\%$) (Appendix A2). The actual simulation is run using the OD matrix presented in chapter 5.3 (Appendix A1). Finally, the uncertainty standard deviation values (ξ and γ) are set to their minimal values (0.1).

The calculated error values are presented in the table below:

Table 4: Results of test 4.

TEST 4	Network	Network (speeds)	Network (densities)	Max Average per link	Max Average per link (ν)	Max Average per link (ρ)
MAPE	0.0045	$4.0 \cdot 10^{-6}$	0.0089	0.0119 [-5]	$7.4 \cdot 10^{-5}$ [13]	0.0238 [-5]
RMSE	0.1011	$9.9 \cdot 10^{-5}$	0.1430	0.9242 [13]	0.0008 [13]	1.3070 [13]

Most maximum values are observed on link 13. Differences between the ground truth and the corrected values are already evident from the graph (Figure 14).

Links -5 and -6 both show similarly high MAPE values, but as they correspond to low values of densities, they are not accompanied by high RMSE values as well. Therefore, actual differences are not high enough to warrant a significant inaccuracy that would make a difference from a traffic engineer’s point of view. The insignificance of the errors in link -5 can be seen in the relevant graph (Figure 15), where the lines of the corrected values and the ground truth practically coincide.

Calculated state variables for link # 13

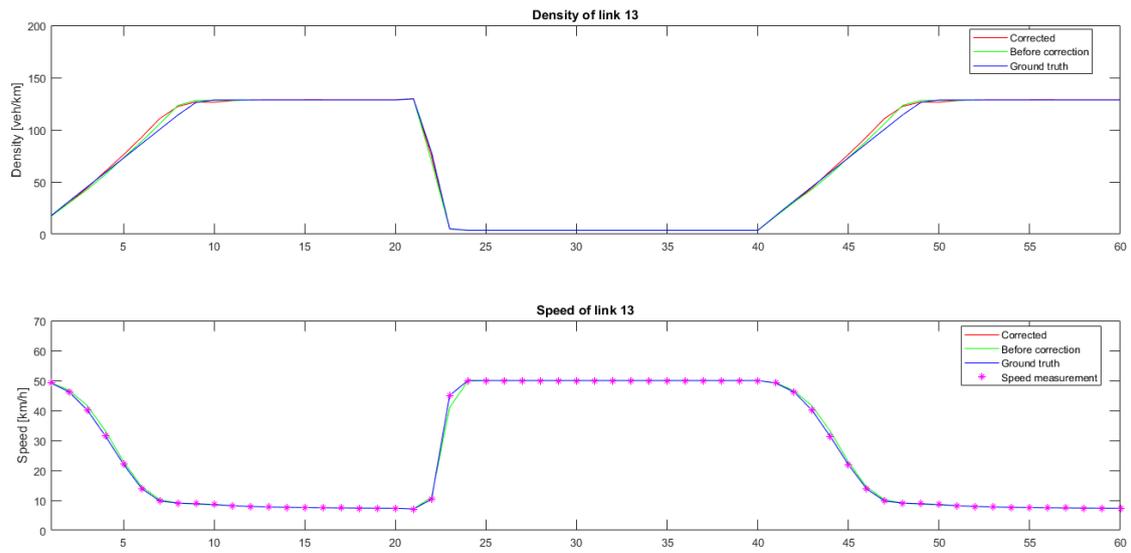


Figure 14. Calculated state variables for link #13.

Calculated state variables for link #-5

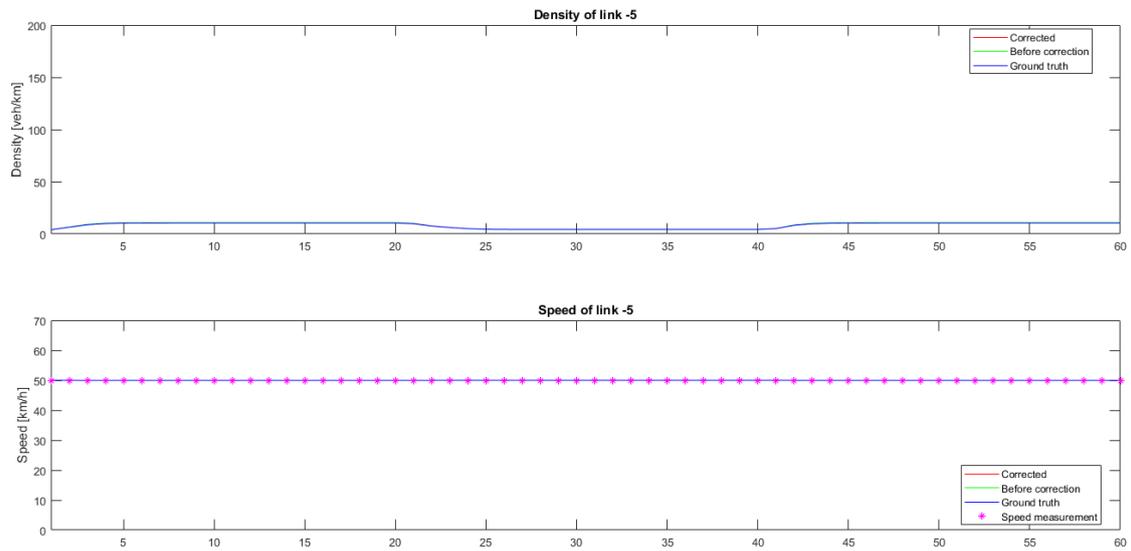


Figure 15. Calculated state variables for link #-5.

5.4.5. Test 5: Inaccurate fundamental diagram parameters

The goal of this test is to check the impact of using slightly inaccurate fundamental diagram parameters to the estimation of the densities and speeds. As the fundamental diagram parameters are not directly measurable, this scenario of inaccurate fundamental diagram parameters is expected to occur in a real case. A second goal is to check if the fundamental diagram parameters are corrected over time. The results of the fundamental diagram parameter estimation are presented in chapter 5.7.

The ground truth was created using the standard OD matrix (Appendix A1) and the standard fundamental diagram parameters of our test network ($v^{free}=50$ km/h, $f^{cap}=1650$ veh/h and $v^{cap}=35$ km/h). The measurements were created from these ground truth values without adding any error to those values and with only the minimum required uncertainty for the system to run ($\xi^q = \xi^v = \gamma^q = \gamma^v = 0.1$).

The original fundamental diagram parameters on all links were altered within limits ($\pm 10\%$). Free flow speeds were set between 45 and 55 km/h, speeds at capacity between 32 and 38 km/h and capacities per lane were set between 1500 and 1790 veh/h. The individual values set for each link can be found in Appendix A3. As in this test we want to enable changing of the fundamental diagram parameters for our system, we set the relevant standard deviation values of the fundamental diagram parameters to appropriate values $\xi^{vfree} = \xi^{vcap} = 2$ and $\xi^{fcap} = 500$. The selection of especially this very high value for the standard deviation of the capacity per lane is discussed in chapter 5.6.2, in which the performance of the system in estimating the fundamental diagram parameters is presented. All other uncertainty parameters are kept stable, with the same values as in the ground truth.

The results of the estimation of the density and speed show satisfactory results. While initially larger differences are observed, especially on the links whose fundamental diagram parameters are further off the ground truth values, soon the estimation of the densities and speeds improves, as the fundamental diagram parameters (especially the free flow speed) are corrected over time. More on the evolution of the estimation of the fundamental diagram parameters is presented in chapter 5.6.2.

The calculated error values for the calculation of the speeds and densities are presented in the table below:

Table 5: Results of test 5.

TEST 5	Network	Network (speeds)	Network (densities)	Max Average per link	Max Average per link (v)	Max Average per link (ρ)
MAPE	0.0062	0.0072	0.0052	0.0209 [-10]	0.0312 [-10]	0.0159 [13]
RMSE	0.3020	0.3588	0.0916	1.1089 [-10]	1.5488 [-10]	1.1511 [13]

Link -10 shows the highest maximum average MAPE and RMSE values. At the same time it is one of the links that had its initial fundamental diagram parameters (especially the free flow speed) furthest away from the ground truth values. As the free flow speed over time approached the ground truth value of 50 km/h, the estimation of speed and density improved as well. The same behavior can be observed on all links and it is even more evident on the most congested link 13. Comparing the density and speed evolution in time steps 1-10 with those of time steps 41-50, when the high-demand OD matrix is used, a much lower estimation error is observed in the second case, when the free flow speed is close to the ground truth value. The relevant graphs for the speed and density estimation of links -10 and 13 are presented in Figures 16 and 17.

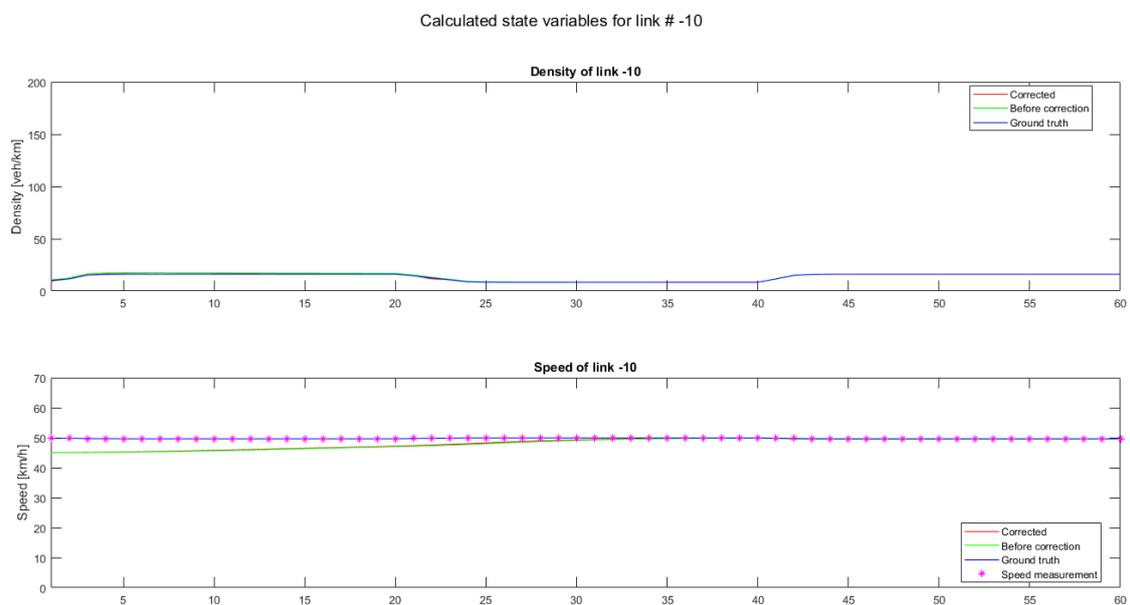


Figure 16. Calculated state variables for link #-10.

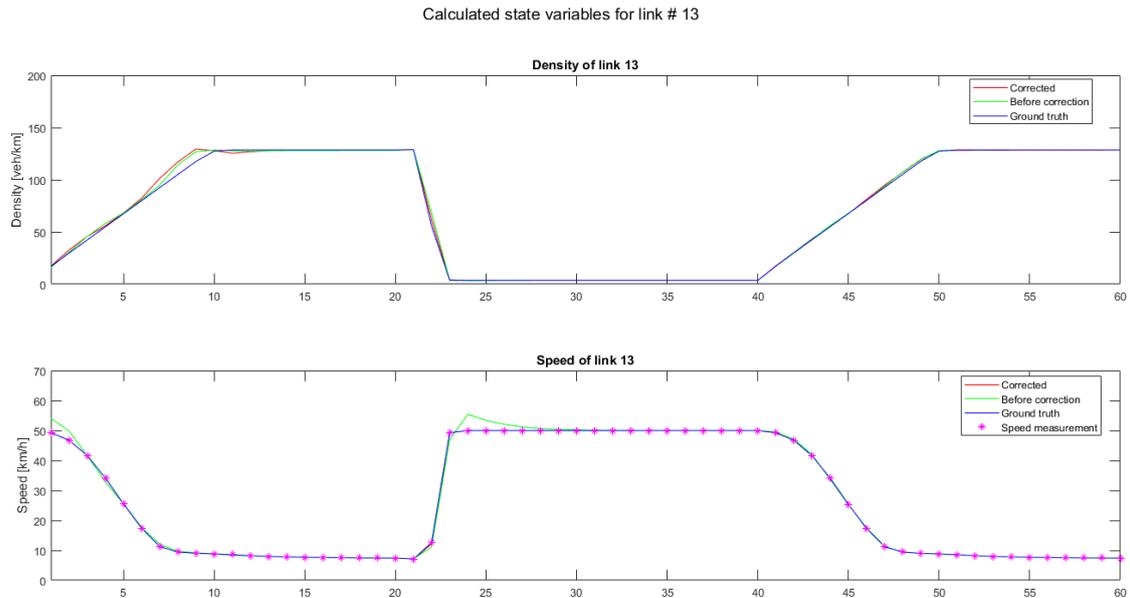


Figure 17. Calculated state variables for link #13.

5.5. Combinations

Apart from the individual tests performed in the previous subchapter, three combinations between the uncertainty parameters were tested, in order to provide more insight on the combined effect of the parameters. The parameters to combine in these tests were selected based on situations most likely to encounter in a real-case scenario, as well as combinations where it was logically expected that the combination would significantly increase their individual impact on the estimation. Finally, a test with all uncertainty parameters enabled was run. The results for these tests are presented in the rest of this subchapter.

5.5.1. Test 6: Inaccurate OD matrix and lower availability of measurements

The combination of an inaccurate OD matrix with the lower availability of measurements is a combination that was expected to significantly increase the impact of the inaccurate OD matrix. This happens because, by fusing less measurements, there are less opportunities to correct and alleviate the effect of the OD matrix. In the individual test presented in subchapter 5.4.4, speed measurements of all links, as well as directional flow measurements from all equations were available in all time steps. In this test, speed measurements to only one direction (link #1, 3, -4, 7, 8, 9, 10, 11, 12 and -15) are fused in each time step. In addition, directional flow data from only the regulated intersections (R1 and R2) is available.

The fundamental diagram parameters remain fixed at their ground truth values and no error or uncertainty on the measurements is applied (the minor required values are used again, $\xi^q = \xi^v = \gamma^q = \gamma^v = 0.1$).

The calculated error values are presented in the table below:

Table 6: Results of test 6.

TEST 6	Network	Network (speeds)	Network (densities)	Max Average per link	Max Average per link (ν)	Max Average per link (ρ)
MAPE	0.0082	0.0027	0.0137	0.0563 [13]	0.0728 [13]	0.0397 [13]
RMSE	0.1842	0.0495	0.2449	2.6359 [13]	1.0134 [13]	3.4414 [13]

The highest error values are all observed on link 13. The relevant graph is presented in Figure 18.

Contrary to the individual test presented in subchapter 5.4.4, no speed measurements for link 13 are available, leading to an increased error in the speed and, therefore in the density estimation as well. The fused speed measurements for link 13 in the individual test had helped alleviate the effect of the inaccurate OD matrix. In this combined test, the values are mostly corrected by the fusion of flow measurements of the regulated intersection R2, as well as from the correlations to other links as well.

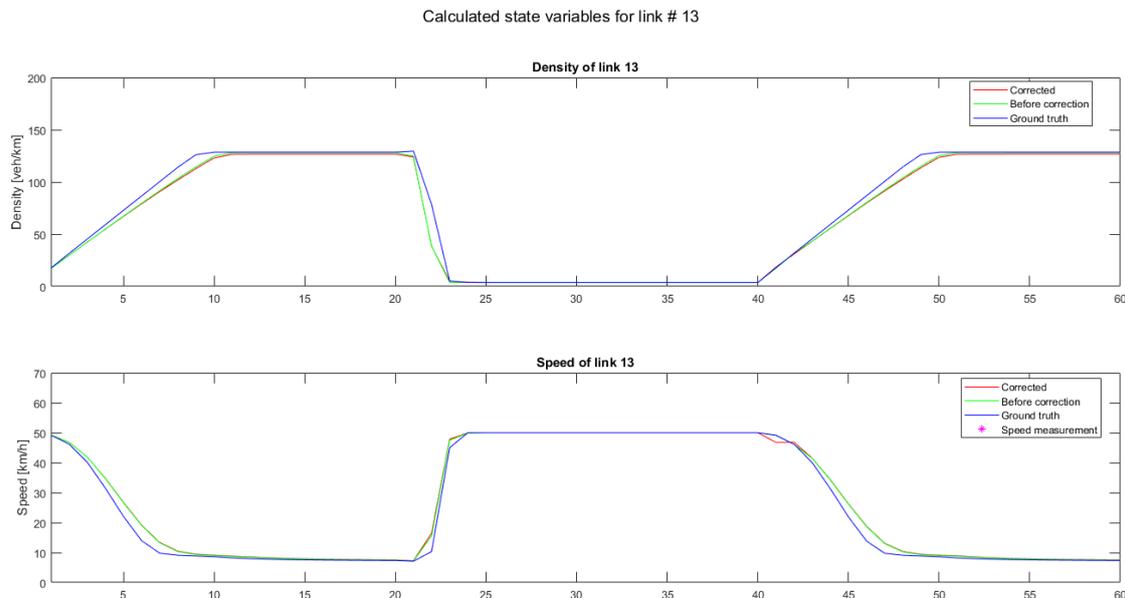


Figure 18. Calculated state variables for link #13.

5.5.2. Test 7: Inaccurate fundamental diagram parameters and error/uncertainty on the measurements

This pair of uncertainty parameters was selected because they proved to have the highest impact on the estimation of the densities and speeds, based on the network MAPE and RMSE values calculated in the relevant individual tests. At the same time, as the fundamental diagram parameters are not directly measurable and the sensors providing the measurements are not perfect, it is expected in a real-case scenario to have slightly inaccurate fundamental diagram parameters and errors on the measurements. Therefore, this is another scenario that can be encountered in reality, while at the same time having great interest because of the high impact on the estimation already observed for these two parameters.

The ground truth is created using the standard OD matrix (Appendix A1). The standard fundamental diagram parameters are also used ($v^{free}=50$ km/h, $f^{cap}=1650$ veh/h and $v^{cap}=35$ km/h), and they are kept constant for the creation of the ground truth. The measurements are created adding a random error of $\pm 10\%$. Speed measurements from all links and directional flow measurements from all intersections are created to be used in the simulation.

In the actual simulation, uncertainty standard deviation values for the process (flow and speed) and flow measurements are set to 0.5, while the uncertainty for the speed measurements was set to 1. The initial fundamental diagram parameters are set as in test 5 (detailed table with all the fundamental diagram parameter values per link is provided in Appendix A3) and are allowed to change during the simulation according to the standard deviation values of $\xi^{vfree} = \xi^{vcap} = 2$ and $\xi^{fcap} = 500$. The OD matrix in the simulation is kept the same as in the ground truth.

The calculated error values are presented in the table below:

Table 7: Results of test 7.

TEST 7	Network	Network (speeds)	Network (densities)	Max Average per link	Max Average per link (v)	Max Average per link (ρ)
MAPE	0.0374	0.0336	0.0411	0.0578 [-10]	0.0751 [-10]	0.0647 [2]
RMSE	1.3320	1.6488	3.1849	3.1849 [13]	3.7378 [-10]	3.9400 [13]

The links that show the highest error values are links -10 and 13, the same links that displayed the highest errors in the individual test of the fundamental diagram parameters (in 5.4.5). An improvement over time is observed in this test, as well. However, the improvement is much slower, as is the convergence of the free flow speed. Therefore, the introduction of error and uncertainty for the measurements greatly reduces convergence speed. In the individual fundamental diagram parameter test, the free flow speed of almost all links had already reached the ground truth free flow speed value by the 30th time step. In

this test, very few links have managed to achieve converge at the 30th time step and by examining the speed of convergence, it appears that it would take 3 times the simulation horizon, in order for the free flow speed of all links to converge to their ground truth values.

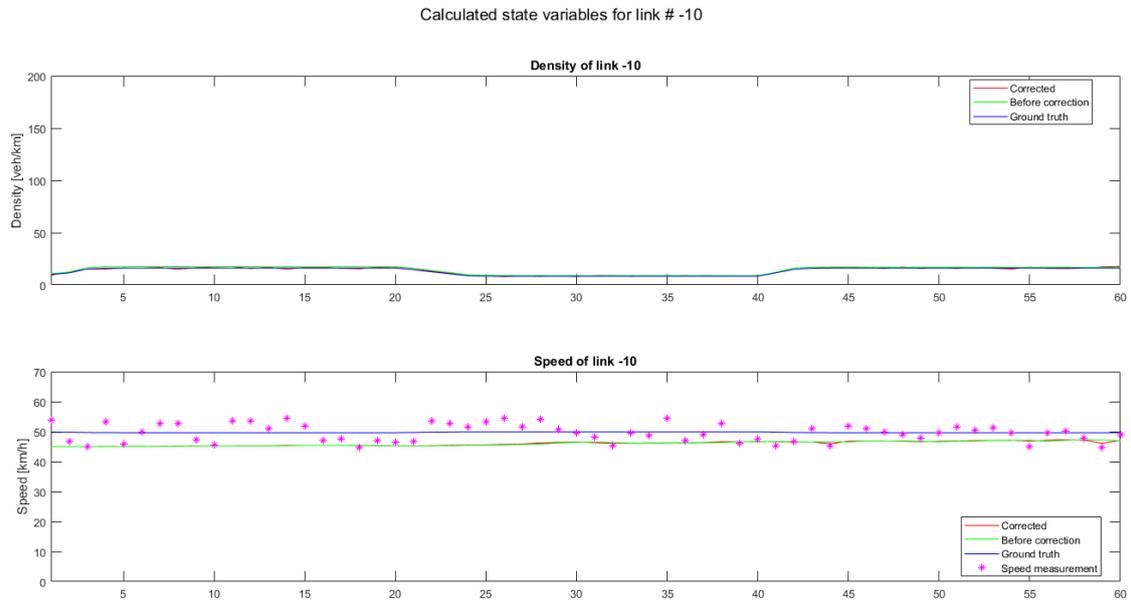


Figure 19. Calculated state variables for link #-10.

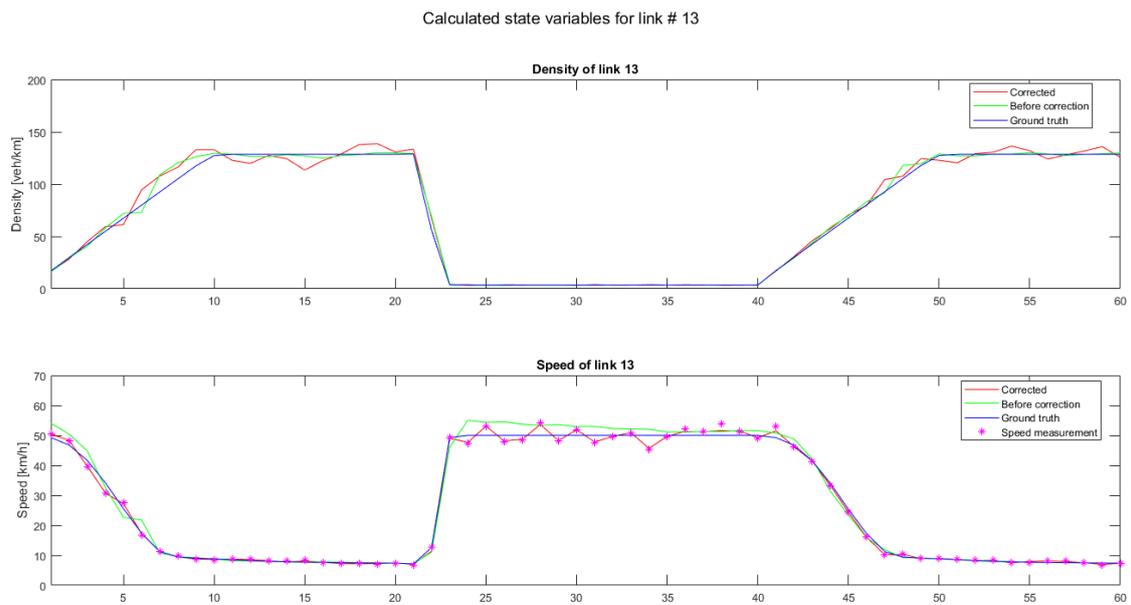


Figure 20. Calculated state variables for link #13.

Therefore, the relevant graphs for the speeds and densities for these links (Figures 19 and 20) show the gradual improvement over time in the estimation and a comparison with the same graphs of the individual test (Figures 16 and 17) shows the different in the improvement achieved over time, comparing the errors in speed and density estimation of time steps 1-10 with those of time steps 41-50.

5.5.3. Test 8: Inaccurate fundamental diagram parameters and OD matrix

The third parameter pair also represents a pair that can be found in a real-case scenario, as there is always uncertainty regarding the fundamental diagram parameters, which are not directly measurable, and the OD matrix, which is compiled using data from various sources and is not likely to be entirely accurate.

The ground truth is created using the altered OD matrix (Appendix A2). The standard fundamental diagram parameters are also used ($v^{free}=50$ km/h, $f^{cap}=1650$ veh/h and $v^{cap}=35$ km/h), and they are kept constant for the creation of the ground truth. The measurements are created without adding any error. Speed measurements from all links and directional flow measurements from all intersections are created to be used in the simulation.

In the actual simulation, the minor uncertainty standard deviation value of 0.1 is set to the ξ and γ values. The OD matrix used is the standard OD matrix (Appendix A1). The initial fundamental diagram parameters are set as in test 5 (Appendix A3) and are allowed to change during the simulation according to the standard deviation values of $\xi^{vfree} = \xi^{vcap} = 2$ and $\xi^{fcap} = 500$.

The calculated error values are presented in the table below:

Table 8: Results of test 8.

TEST 8	Network	Network (speeds)	Network (densities)	Max Average per link	Max Average per link (v)	Max Average per link (ρ)
MAPE	0.0087	0.0072	0.0101	0.0193 [-10]	0.0312 [-10]	0.0304 [-6]
RMSE	0.3409	0.3575	0.1592	1.1204 [-10]	1.5497 [-10]	1.4644 [13]

The highest error values are mainly observed in link -10, similarly to the results of the individual fundamental diagram parameter test (in 5.4.5). The relevant graphs for this link are provided in Figure 21. It shows a gradual improvement over time, as can be seen more easily in the speed graph, which occurs as the free flow speed gradually converges to the ground truth value. Link -10 is the slowest to converge compared to other links in the network whose initial free flow speed values are equally far from the ground truth value. Such an example is link 3, whose free flow speed value appears to converge after

approximately 10-15 time steps, significantly faster than the 30-35 time steps required for link -10. This observation is further discussed in chapter 5.6. The relevant graph for link 3 is provided for comparison in Figure 22.

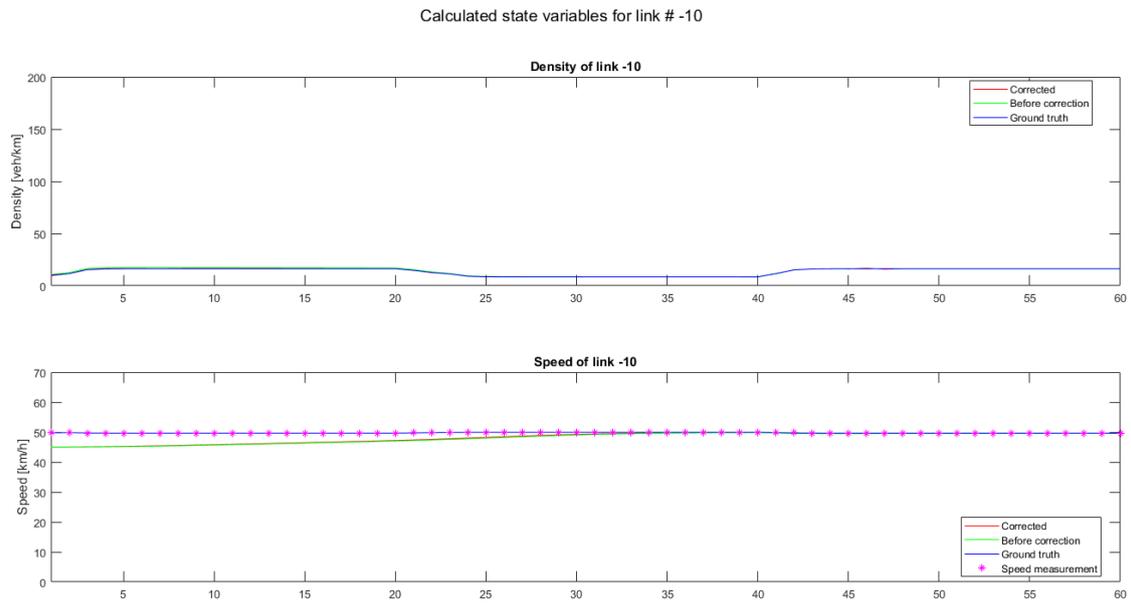


Figure 21. Calculated state variables for link #-10.

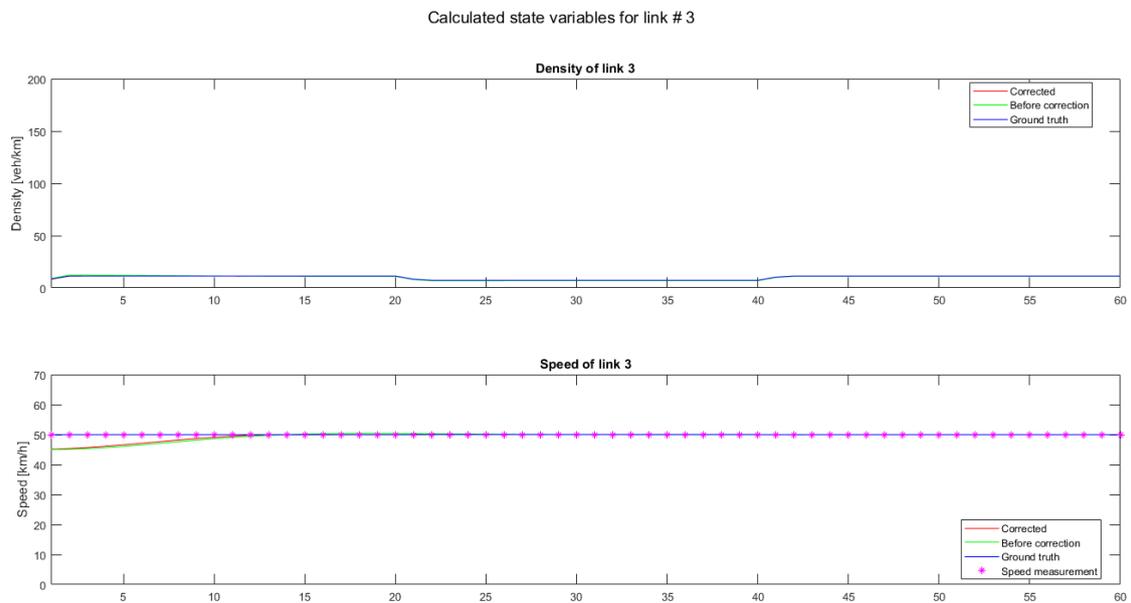


Figure 22. Calculated state variables for link #3.

5.5.4. Test 9: All uncertainty parameters combined

The final test incorporates all uncertainty parameters. All these uncertainty parameters will be evident in the real case. Therefore, a test incorporating all parameters is an opportunity to test the performance of the developed model in a scenario that resembles real life but in a controlled environment.

The ground truth is created using the altered OD matrix (Appendix A2). The standard fundamental diagram parameters are also used ($v^{free}=50$ km/h, $f^{cap}=1650$ veh/h and $v^{cap}=35$ km/h), and they are kept constant for the creation of the ground truth. The measurements are created adding a random error of $\pm 10\%$. Speed measurements to only one direction (links #1, 3, -4, 7, 8, 9, 10, 11, 12 and -15) are fused in each time step. In addition, directional flow data from only the regulated intersections (R1 and R2) is available.

In the actual simulation, uncertainty standard deviation values for the process (flow and speed) and flow measurements are set to 0.5, while the uncertainty for the speed measurements is set to 1. The initial fundamental diagram parameters are set as in test 5 (Appendix A3) and are allowed to change during the simulation according to the standard deviation values of $\xi^{vfree} = \xi^{vcap} = 2$ and $\xi^{fcap} = 500$. The OD matrix in the simulation is the standard OD matrix (Appendix A1), different to the one used in the ground truth by $\pm 10\%$.

The calculated error values are presented in the table below:

Table 9: Results of test 9.

TEST 9	Network	Network (speeds)	Network (densities)	Max Average per link	Max Average per link (v)	Max Average per link (ρ)
MAPE	0.0614	0.0595	0.0632	0.1763 [2]	0.1664 [2]	0.1862 [2]
RMSE	2.2092	2.8653	0.7849	5.8914 [2]	8.3179 [2]	5.3015 [13]

The highest error values for this test are mainly observed in link 2. In this link, the estimation of the free flow speed has failed (diverges from the ground truth value) leading to an increasing error. In general, the fundamental diagram parameter estimation seems to be failing in most links, as few links show a convergence of the free flow speed toward the ground truth value and still at a lower pace compared to the previous tests. The inability to estimate the free flow speed has a significant impact to the estimation, as most links of the network are in free flow conditions. The relevant graphs for the two links showing the highest error values (links #2 and #13) are presented in Figures 23 and 24.

The network MAPE value of 0.0614 is the highest of all tests, as expected, but still low enough to support the claim that the developed model works sufficiently well for the estimation of the speeds and densities. The main source of problems appears to be the

inability to estimate the free flow speed, leading to an inaccurate speed and consequently a wrong density value.

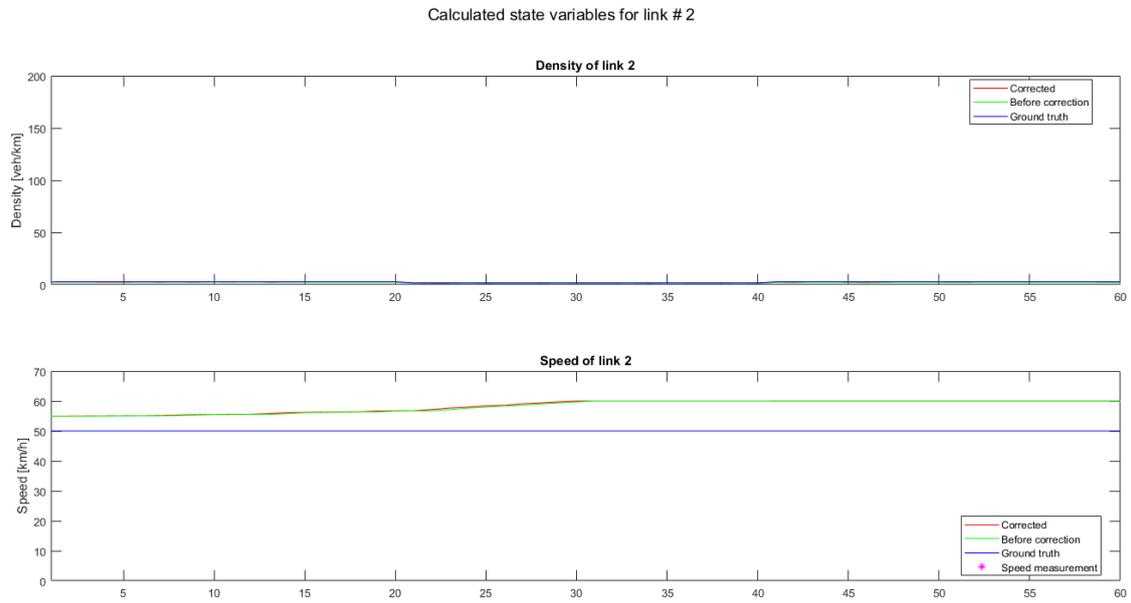


Figure 23. Calculated state variables for link #2.

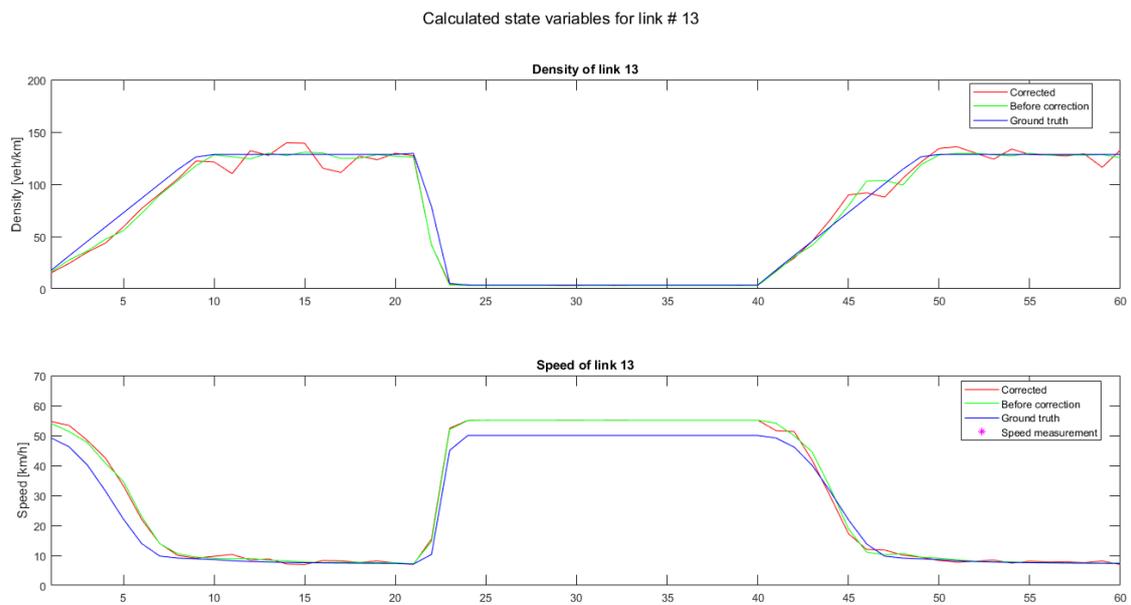


Figure 24. Calculated state variables for link #13.

For the reader's convenience, the results of the MAPE and RMSE values from all tests are gathered in Table 10.

Table 10: Aggregated results of all tests.

TEST #		Network	Network (speeds)	Network (densities)	Max Average per link	Max Average per link (u)	Max Average per link (ρ)
1	MAPE	$5.5 \cdot 10^{-4}$	$7.9 \cdot 10^{-6}$	0.0011	0.0076 [13]	0.0002 [13]	0.0149 [13]
	RMSE	0.0381	$1.5 \cdot 10^{-4}$	0.0539	0.9753 [13]	0.0026 [13]	1.3792 [13]
2	MAPE	0.0277	0.0179	0.0375	0.0428 [13]	0.0274 [13]	0.0582 [13]
	RMSE	0.8257	0.8679	0.5019	2.8943 [13]	1.3490 [5]	3.7714 [13]
3	MAPE	0.0006	0.0006	0.0006	0.0090 [13]	0.0146 [13]	0.0034 [13]
	RMSE	0.0149	0.0158	0.0089	0.2199 [13]	0.2742 [13]	0.1032 [13]
4	MAPE	0.0045	$4.0 \cdot 10^{-6}$	0.0089	0.0119 [-5]	$7.4 \cdot 10^{-5}$ [13]	0.0238 [-5]
	RMSE	0.1011	$9.9 \cdot 10^{-5}$	0.1430	0.9242 [13]	0.0008 [13]	1.3070 [13]
5	MAPE	0.0062	0.0072	0.0052	0.0209 [-10]	0.0312 [-10]	0.0159 [13]
	RMSE	0.3020	0.3588	0.0916	1.1089 [-10]	1.5488 [-10]	1.1511 [13]
6	MAPE	0.0082	0.0027	0.0137	0.0563 [13]	0.0728 [13]	0.0397 [13]
	RMSE	0.1842	0.0495	0.2449	2.6359 [13]	1.0134 [13]	3.4414 [13]
7	MAPE	0.0374	0.0336	0.0411	0.0578 [-10]	0.0751 [-10]	0.0647 [2]
	RMSE	1.3320	1.6488	3.1849	3.1849 [13]	3.7378 [-10]	3.9400 [13]
8	MAPE	0.0087	0.0072	0.0101	0.0193 [-10]	0.0312 [-10]	0.0304 [-6]
	RMSE	0.3409	0.3575	0.1592	1.1204 [-10]	1.5497 [-10]	1.4644 [13]
9	MAPE	0.0614	0.0595	0.0632	0.1763 [2]	0.1664 [2]	0.1862 [2]
	RMSE	2.2092	2.8653	0.7849	5.8914 [2]	8.3179 [2]	5.3015 [13]

5.6. Discussion

5.6.1. Estimation of the speeds and densities

The aim of the tests that were run in the validation process was to examine the accuracy of the state estimation, as well as to discover the uncertainty parameters that have the highest impact on the estimation of the speeds and densities of all links of the network. From the individual tests of chapter 5.4, the most impacting parameter is the error and uncertainty of the measurements (test 2), which displayed the highest network MAPE and RMSE values. Using the same criterion, the second most impacting parameter is the inaccuracy of the fundamental diagram parameters, followed by the inaccuracy of the OD matrix and the lower availability of measurements.

The combinations tested in chapter 5.5 provided more insight on the combined effect of the parameters, approaching the conditions of a real case where more uncertainty parameters are involved. The combined effect of the inaccurate OD matrix and lower availability of measurements (test 6), as well as that of the inaccurate fundamental diagram parameters combined with an error and uncertainty on the measurements (test 7), appeared to be significantly higher, as the resulting network MAPE values for these tests is significantly higher than the sum of MAPE values of the relevant individual tests. In test 6, the model had

less opportunities to correct the errors imported due to the inaccurate OD matrix, as the amount of available measurements was lower, leading to a higher impact of that imported error, until the model could gradually “learn” to handle it. In test 7, the error and uncertainty on the measurements weakened the ability of the system to estimate the free flow speed and consequently improve the estimation of speeds and densities. The free flow speed was still correctly estimated after some time, but convergence took longer.

On the other hand, in test 8 (inaccurate OD matrix and fundamental diagram parameters) the combined effect of the parameters appeared to be lower than their individual effect. Taking into account that part of each individual effect is also due to the minimum required uncertainty for the model to run, as well as roundings in the calculations and exchange of data between Streamline and Matlab, as mentioned in 5.4.1, the conclusion that we can draw from this test is that the model is still able to handle the inaccuracy of both parameters adequately, at least for this inaccuracy level of $\pm 10\%$ selected for the test.

In test 9, the combined effect of all uncertainty parameters together significantly increased. The model failed to calculate the free flow speeds of links on parts of the network that were not measured. On links with speed measurements and/or close to the regulated intersections for which measurements are available, there was convergence but it was slower than in previous tests because of the additional errors/uncertainty imported to the model in the last test. The free flow speed value has a strong influence on the speed estimation and consequently on the density estimation. Thus, this parameter proved to be the most disrupting to the estimation because the co-existence of the other uncertainty parameters severely weakened the model’s ability to correct it with time.

The overall network MAPE value of test 9 (0.0614) is significantly higher than the values observed in the previous tests. However, it can be considered marginally acceptable, as its practical meaning of an average error of about 6% is in most cases adequate, meaning for example an estimated speed of 47 km/h instead of the correct speed of 50 km/h. However, when dealing with larger values, e.g. a three-lane road with a traffic flow of 5400 veh/h, a 6% error means a difference in the flow of over 320 veh/h, which consists a much more significant difference from a traffic engineer’s point of view.

By calculating separate overall MAPE values for the densities and speeds of the links of the network (columns 4 and 5 of Table 10), it can be derived that the speed estimation is significantly more accurate in cases where the fundamental diagram parameters are correctly estimated. The main reason for the fact that the estimation of the density is less accurate than the estimation of the speed is that while we have direct measurements of the speed, we can only indirectly estimate/correct the density with the help of the flow measurements. Flow measurements, at least in this network, can only be obtained from the intersections and, thus, are not available for all links of the network. Furthermore, the method selected of summing up flow measurements, due to limitations imposed from the junction modelling factors, as analyzed in chapter 4.6, has an additional weakening effect to their ability to correct the state. Therefore, it is expected for the density estimation to be in most cases less accurate than the speed estimation. The parameter that seems to mostly

disrupt speed estimation is the fundamental diagram parameters (free flow speed), as can be understood from the fact that the difference in the MAPE values of the densities and speeds for the network (columns 4 and 5 of Table 10) is smaller in the tests that involve inaccurate fundamental diagram parameters. In test 5, in which the only uncertainty parameter enabled is the inaccuracy of the fundamental diagram parameters, the MAPE value for the speeds even exceeds the MAPE value for the densities.

The columns 6, 7 and 8 of Table 10 provide insight on the links that contribute more to the error values calculated for each test. The link whose error values were the highest per test is written within brackets in the relevant cell on that table. The link showing most often the highest values is link 13, followed by links -10, 2, -5 and -6.

Link 13 is the most congested link of the network. However, it is the selected way of calculating the factors simulating the effect of junction modelling (“flow limit factor” and “inflow reduce factors”) that seems to be playing the most significant role to the problem observed on that link, as described in the previous chapter. This can be derived from the fact that when more uncertainty parameters are involved, the errors on other links exceed the error values of link 13.

Link -10 is situated in the middle of the long stretch between the two regulated intersections with a direction from right to left. Link -10 shows the highest error values in the tests that had inaccurate fundamental diagram parameters. This can be explained by the fact that link -10 is situated the furthest away from the regulated intersections, so it is less affected by the available directional flow measurements on the intersections. Therefore, it relies almost exclusively to its speed measurements, as well as speed measurements of the neighboring links, in order to correct its free flow speed and eventually speed and density. As was observed in these tests, link -10 showed slower convergence of the free flow speed and this led to a slower improvement in the estimation of its speed and density. In addition, another factor is its starting free flow speed value of 45 km/h, which is the furthest away from the actual value of 50 km/h. These two factors contribute to its slower convergence of the free flow speed and its higher error values.

Convergence speed could be increased by using higher values for the relevant standard deviation. However, this method should be used with caution, as it could lead to more nervous behavior of the free flow speeds possibly leading to higher overall error values in the network. Therefore, it is recommended to run several tests with different standard deviation values for the fundamental diagram parameters, in order to check the behavior of the system and achieve a balance between speed of convergence and nervous behavior.

The improvement achieved in the estimation of the free flow speeds over time, which leads to improved estimation of the speeds and densities as well, could justify the introduction of a warm-up phase to allow the free flow speed to converge or at least move closer to the actual values before initiating state estimation. An important advantage of the developed model, as was observed in the results and previously discussed, is that it can converge to the actual fundamental diagram parameter values (at least the free flow speeds), irrespective of

the difference of their initial values from the ground truth values. The disrupting effect of wrong fundamental diagram parameters to the speed and consequently the density estimation was also observed. Therefore, introducing a warm-up phase allows the user to make more relaxed assumptions for the initial fundamental diagram parameter values to set, as they will be for the most part corrected during this warm-up phase, without affecting speed and density estimation which will begin after this phase. The possible addition of a warm-up phase, as well as its duration, would be subject to the exact conditions of the network in question, such as the usual congestion levels and the estimated uncertainty levels of the parameters.

In addition, it has to be noted that the addition of a warm-up phase would be useful not only in cases with constant fundamental diagram parameter values as the ground truth, as were the cases tested in this thesis, but also in cases with varying values over time. The behavior that the estimation of the free flow speed displays, gives confidence to claim that it can also follow and adapt to changing values of the parameter. However, this will need to be verified using relevant tests. In addition, in most cases with variable fundamental parameter values, e.g. due to adverse weather conditions, the change of the parameters occurs gradually, allowing the system to adapt to the new conditions. Therefore, alleviating the uncertainty of the initial fundamental diagram parameter values set is also important in cases with varying values over time and an introduction of a warm-up phase in these cases is recommended as well.

Link 2 is situated upstream the unregulated intersection (U1) and downstream centroid C2. It shows the highest error values in test 9 and also the highest error on the speed estimation in test 7. Its starting free flow speed value is 45 km/h, as in link -10, and there are no speed measurements available for this link. Therefore, it is reasonable that the estimation of link 2 is failing, as the system has practically no information available for that part and direction of the network, while an error due to the different OD matrix is constantly added by the centroid upstream (C2) and the speed of the link is constantly estimated wrongly because of the inaccurate free flow speed. In general, it was observed in test 9 that the links whose free flow speeds move over time towards convergence were those that had available measurements, either speed measurements, or were close to a regulated intersection and were therefore corrected by the fusion of the directional flow measurements.

A possible improvement could be achieved using sets of common fundamental diagram parameters for several links for which it is reasonable to claim that they would have the same fundamental diagram parameter values (e.g. all links between two consecutive intersections). Convergence would then be achieved faster, as the model would receive more information on these parameters from various links of the network. A hypothesis can reasonably be made that such a strategy would prove to be helpful especially for links on parts of the network for which there is not enough information, such as link 2. However, thorough testing of such a hypothesis would be required.

Another point that needs to be stressed is the importance of the thresholds applied at the end of each time step. It is important that these thresholds are set as realistically as possible,

in order to prevent estimation from significantly diverging from the ground truth, especially in cases with high uncertainty. For example, as can be seen from the results of link 2 in test 9 (Figure 23), the introduction of an upper threshold for the free flow speed at 60 km/h (on a link with a speed limit of 50 km/h) also kept the diverging of the speed estimation to reasonable levels, as an additional threshold is applied to the speeds, which prevents them from being much higher than the free flow speed of the link.

Links -5 and -6 are situated downstream the middle regulated intersection and upstream centroids 4 and 3 respectively. These links show higher error values, especially for the density estimation, in tests 4 and 8, which incorporate an inaccurate OD matrix. This result can be explained by the fact that the OD matrix affects routing and the calculated turning fractions. These turning fractions, which in this case are calculated from an inaccurate OD matrix, take part in the flow measurement equations through the “inflow reduce factors”, leading to a similarly inaccurate value downstream the intersection on the links leading to a centroid/exit from the network.

A major topic of discussion is also the existence of bias in this research, which could be affecting the results. An example is the range of $\pm 10\%$ set for the addition of errors in the measurements, fundamental diagram parameters and OD matrix. This value was selected as being low enough to reflect e.g. a relatively well-calibrated OD matrix and high enough to make differences observable. The same value was used in all tests, in order to create a fair base for comparison. However, it is not necessarily true that using the same range for all uncertainty parameters is fair, due to the different characteristics of each parameter. For the uncertainty parameter of lower availability of measurements, as parameters such as the dispersion of the measurements (e.g. remove scattered measurements or concentrated in one part of the network) are also of importance, the criterion selected was the resemblance to a real-case scenario, where e.g. flow data would be available only at the regulated intersections and speed data could be available only to one direction of traffic.

Furthermore, a major source of bias is the fact that the measurements are created using basically the same model used for simulation (Streamline). Using real measurements would certainly lead to higher error values. Therefore, validation with real measurements as well is considered necessary.

Moreover, the form of the test network, as well as the demand set for it is another source of bias. The fact that most links are on (or close to) free flow conditions could possibly be the reason behind the increased significance of the free flow speed to the speed estimation, compared to the capacity per lane and the speed at capacity. In addition, the failure of estimating those two fundamental diagram parameters could possibly be attributed to the fact that there are almost no measurements close to capacity conditions. More tests using more congested conditions would be required to test this hypothesis.

A final point that needs to be mentioned is the order of fusing the measurements. As thorough testing proved, the order of fusing the measurements practically plays no role in the state estimation. Resulting differences in the corrected state and error covariance matrix

are negligible. The effect of the order of fusing the measurements declines over time, as the system “learns” the network. After running many tests where extreme measurements were introduced in the first time step(s) or later in the simulation period, it was discovered that the impact of the extreme measurements on the first time step was significant. However, after a few more time steps this effect was negligible. Therefore, on a practical application of this method, the order of fusing the measurements makes practically no difference and extreme measurements can be handled adequately well by the system.

5.6.2. Estimation of the fundamental diagram parameters

From literature (e.g. Wang et al. 2007 & 2008) it is known that convergence of the fundamental diagram parameters in a freeway network takes a long time, and in this specific work the fundamental diagram parameters used were common for all links of the network, something that could facilitate convergence, together with the smaller complexity of a freeway network compared to an urban network. The slow convergence can be explained from the fact that these parameters cannot be directly measured, so there are no measurements of these specific parameters which could be fused to improve accuracy. They are indirectly changed due to the fusion of the flow and speed measurements, as the fundamental diagram parameters take part in the speed equations, specifically in the element that contains the fundamental diagram calculation. However, it would be expected to see the altered fundamental diagram parameter values gradually changing towards the “ground truth” values and the unchanged values to remain practically unaltered, fluctuating over time around the “ground truth” values.

A first note regarding the estimation of the fundamental diagram parameters concerns the standard deviation values to be used. In a first run with standard deviation values of 2, 5 and 2 respectively for the free flow speed, capacity per lane and speed at capacity, it was noticed that there was almost no change in the values of the capacity per lane parameter. This can possibly be explained by the fact that the capacity per lane parameter participates less in the forming of the speed equation compared to the other two parameters (the free flow speed and speed at capacity elements variables appear three times in the equation, while the capacity per lane only once), at least with the equations as formed using the METANET fundamental diagram, as well as from the fact that the actual numerical value is much higher than the values of the free flow speed and the speed at capacity. Therefore, the use of a significantly higher standard deviation value for the capacity per lane parameter is justified. The test was then repeated with a very high standard deviation value of 500 for the capacity per lane and in this case some change on the capacity per lane values was observed.

The estimation of the fundamental diagram parameters was enabled in tests 5, 7, 8 and 9. The main difference observed in the tests was that as the number of uncertainty parameters involved increased, convergence speed of the free flow speed became lower. In test 9, with all uncertainty parameters enabled, convergence speed was very low and only for links that were backed by measurements. Calculation of the free flow speed in other links practically

failed, with some examples even diverging from the ground truth values. The discussion of the results that follows is based on test 5, where all other uncertainty parameters were eliminated, so the actual fundamental diagram parameter estimation can be examined.

The results on the estimation of the fundamental diagram parameters were mixed. The free flow speed estimation can be considered successful, as in all links the value of the free flow speed changed over time towards the “ground truth” value and when this value was reached fluctuated around this value. The free flow speed values of the links that were already correct at the start of the simulation, fluctuated around that correct value over the whole simulation period, as well. Therefore, in terms of estimation accuracy, the estimation of the free flow speed is considered successful.

As shown in Figure 25, the MAPE and RMSE values decrease over time, indicating the improvement achieved in the calculation of the free flow speed over time. The maximum values of the MAPE and RMSE were found to occur on the first time step (with values of 0.0556 and 2.7809 respectively), followed by a relatively fast decline. A minor increase of the MAPE and RMSE values occurs in time step 41. At time step 41, the OD-matrix used for the simulation is changed again to the higher-demand OD-matrix, causing a slight difference to which the model adapts already from the next time step. A minimum value is reached at time step 54 (with values of 0.0055 and 0.2743 respectively for the MAPE and RMSE value) and is practically kept constant at the remaining time steps.

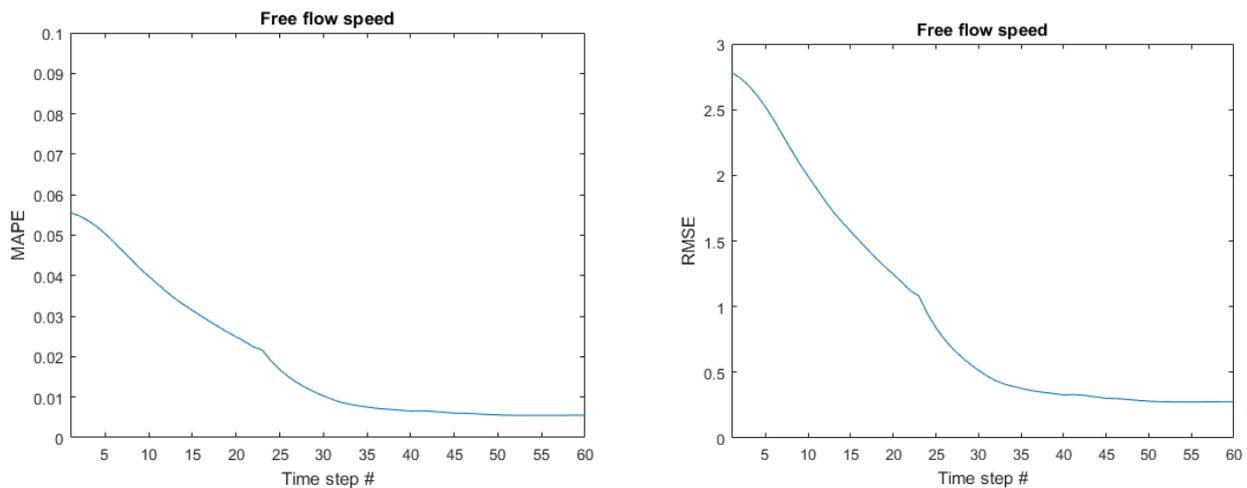


Figure 25. Evolution of the MAPE and RMSE values of the network for the estimation of the free flow speed.

Contrary to the good results observed for the estimation of the free flow speed, the estimation of the speed at capacity and the capacity per lane is unsuccessful. Their values seem to randomly fluctuate over time, often to the wrong direction. Therefore, some values seem to be converging to the correct values, while others are diverging. As shown in Figures 26 and 27, the MAPE and RMSE values for these two parameters remain practically constant over time, showing that there is no improvement achieved in their estimation for the whole network over time, at least for this time period of 60 minutes. What is also observed is a minor change shortly after time steps 20 and 40, when the demand changes. This is practically observed only on the speed at capacity estimation, as in the capacity per lane only very minor fluctuations can be observed, practically imperceptible in the graphs.

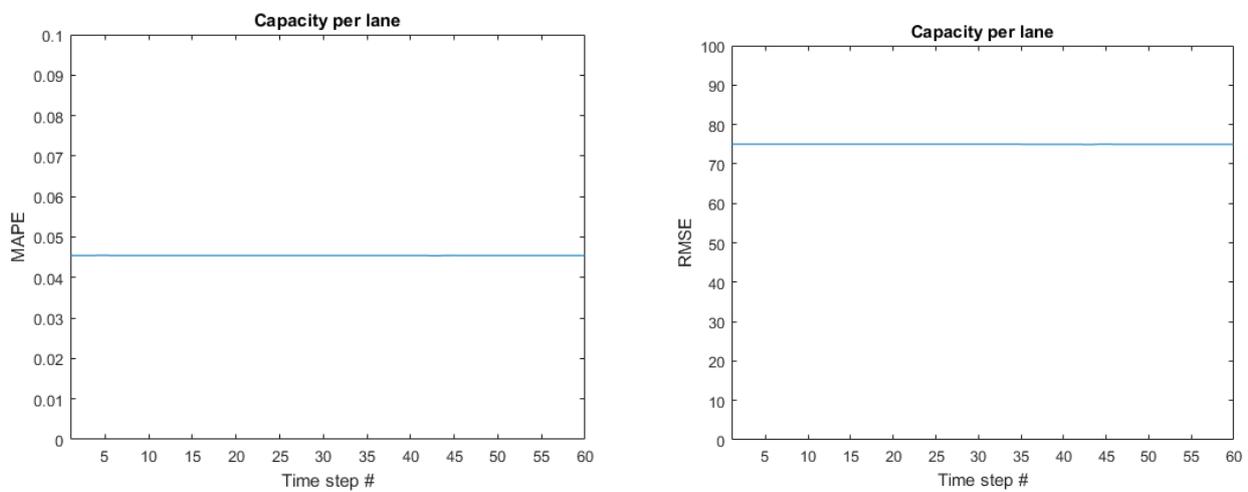


Figure 26. Evolution of the MAPE and RMSE values of the network for the estimation of the capacity per lane.

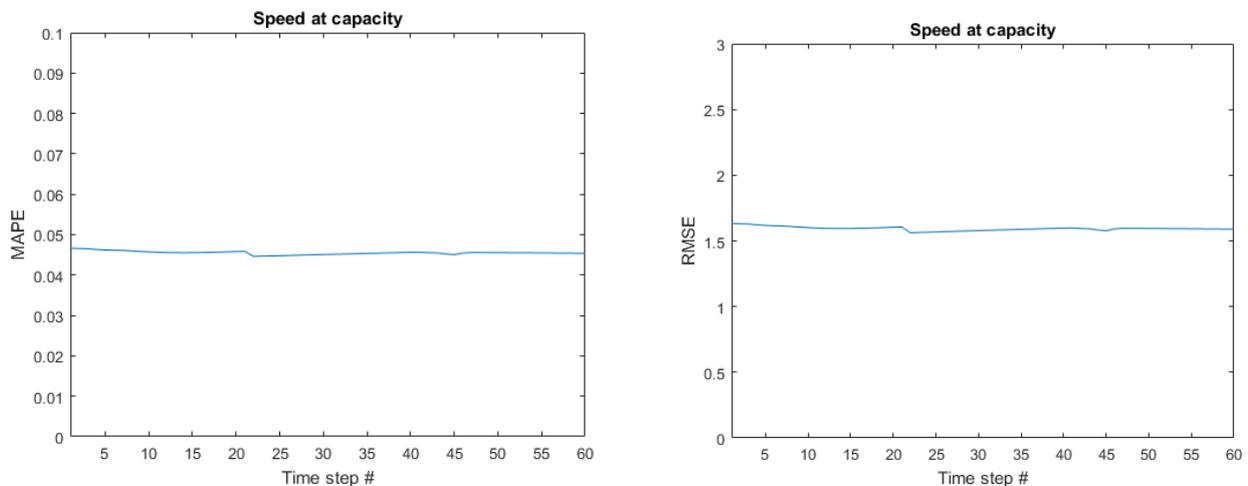


Figure 27. Evolution of the MAPE and RMSE values of the network for the estimation of the speed at capacity.

In general, it is observed again that the changes on the values of the capacity per lane are much smaller per time step, compared to the values of the speed at capacity, despite the big difference in the standard deviation values set for these two parameters (2 for the speed at capacity and 500 for the capacity per lane). As previously mentioned, it can be attributed to the increased “participation” of the speed at capacity parameter to the speed equation, compared to the capacity per lane parameter.

Finally, a factor that could be causing the failure of the speed at capacity and capacity per lane is the fact that in the test network there are almost no measurements close to capacity. Almost all links of the test network are in free flow or close to free flow conditions for the whole simulation period. As a result, the links are almost never measured in congestion or near-capacity conditions. This could prove to be the major cause the Kalman filter cannot improve estimation of the speed at capacity and the capacity per lane parameters, as there are no measurements of the complete fundamental diagram. Therefore, there is not adequate information to adapt the values of these parameters as well. In addition, as the test network is an urban network, it is the junctions that mainly determine the capacities within the network, making the acquisition of measurements of links near capacity even more difficult. At the same time, this dominance of the junctions in determining the capacities could be the reason why speed at capacity and capacity per lane appear to be much less significant for the state estimation in an urban network than the free flow speed. To examine this topic, several tests with a different, more congested network must be carried out. Therefore, this remark is provided as a topic for future research.

6. Conclusion and future research

6.1. Conclusion

As presented in chapter 3, the research questions were the following:

1. *How can a model be developed to provide online estimation of the traffic state of an urban network, taking into account measurements from sensors?*
2. *How accurately can this developed model estimate the fundamental diagram parameters of each link of the network?*

To address the research questions, the developed solution uses Streamline as the process model. It is a higher order LWR model, which is more suitable than a first order model because of the more correlations between links that can be captured and lead to more opportunities for correction on links for which no measurements are available. In addition, Streamline incorporates a junction model (XSTREAM), which is a vital element of a model designed for urban networks, as the role of the junctions is decisive. The observation model was set up accordingly, using the appropriate equations (measurement functions) of the process model. Extended Kalman Filtering was selected as the data assimilation method, as it is suitable for working with non-linear equations and is computationally efficient. As it was required to develop a model suitable for online state estimation, the model does not require any pre-processing of the data and is capable of running faster than real-time, at least for relatively small networks. Finally, the state vector was set up to include the speeds and densities, which are needed to answer the first/main research question, as well as the fundamental diagram parameters, in order to cover the second/additional research question as well. To answer the research questions and assess the effectiveness of the developed approach, the model was validated through a series of tests using artificially created measurements.

Regarding the first research question, it can be concluded that the model could satisfactorily estimate the densities and speeds of all links of the network, as the average error values of all tests were at an acceptable level. Estimation improves over time, as the model “learns” through the fusion of measurements. An increasing uncertainty level of the system leads to a reduced rate of improvement overall for the network. Estimation of the speeds appears in most tests to be more accurate than the estimation of the densities. This can be attributed to the fact that the densities are indirectly corrected through the fusion of flow measurements, while the speeds are directly corrected through the fusion of speed measurements. Another contributing factor is the method selected for fusing the flow measurements in the developed approach, due to technical reasons. By summing up flows from each direction instead of fusing each directional flow measurement, the correction effect of the flow measurements is slightly reduced.

Uncertainty and error on the measurements is the parameter with the highest impact on the results, followed by the inaccuracy of the fundamental diagram parameters and the

inaccuracy of the OD matrix. Combinations of these parameters in most cases led to an increased combined impact on the state estimation. Among the fundamental diagram parameters, the free flow speed has the most significant impact on the estimation of the speeds and consequently the densities.

The results of the last test, which combined all uncertainty parameters showed that, although estimation improved overall in that test as well, the results of certain links showed a deteriorating estimation, as the error values of these links kept increasing over time. This occurred in links on which the uncertainty parameters had the highest impact, e.g. links connected to centroids (and therefore affected by the OD matrix inaccuracy) while at the same time being situated away from any measured link, offering practically no opportunity to correct its state. This result especially underlined the importance of the availability of measurements. It showed, in addition, that very high uncertainty levels (perhaps even higher than those tested) may lead to a failure in estimation. Therefore, it is important to attempt to reduce uncertainty on parameters that can be pre-processed, such as OD matrix calibration or setting the initial fundamental diagram parameter values.

In cases where estimation of the fundamental diagram parameters is not of importance, the model can be modified to exclude them from the state vector. The immediate result would be a significant increase in simulation speed, as the dimensions of the Jacobian matrices would drastically decrease allowing for faster calculations. It can also be argued that estimation of speed and density will also slightly benefit from this change, as the model will have less parameters to calibrate when applying the correction. However, additional tests would be required to validate this hypothesis.

From the analysis of the most problematic links in each test, it was observed that an inaccurate OD matrix proved to have a higher impact on the estimation of the densities of links connected to a centroid (destination/origin). In addition, in tests with higher uncertainty it was observed that state estimation of links situated on parts of the network where no measurements were available did not improve over time, underlining the importance of the number of measurement points on the network, as well as their location.

Regarding the second research question, the model succeeds in improving the estimate of the free flow speed over time but fails to estimate the speed at capacity and capacity per lane. The improvement in the estimation of the free flow speed proved to be independent of the initial values set. The standard deviations which control the variation allowed per time step for each fundamental diagram parameter could be increased in order to achieve faster convergence of the free flow speed. However, a high standard deviation value would lead to more nervous behavior of the estimated value of the free flow speed, possibly leading to a higher average error over time. Therefore, a series of preliminary tests could be run beforehand, in order to decide on their values. Finally, the inclusion of a warm up phase, to allow the parameters to approach their actual values, before actual state estimation begins, is recommended whenever possible.

The most probable cause of the observed failure of the estimation of the speed at capacity and the capacity per lane appears to be the fact that in the tests performed the network was mostly in free flow conditions. As there were almost no measurements near capacity, the system cannot estimate the relevant fundamental diagram parameters. However, this hypothesis needs further validating through relevant tests on networks on near-capacity conditions.

A recommendation would be to use sets of common fundamental diagram parameter values for consecutive/"similar" links in the network, instead of separate values per link. By following this strategy, it is expected that the system will converge to the correct fundamental diagram parameter values and respond to changes in their values faster even in cases with lower availability of measurements, as the EKF will have more sources of information for the same parameters. However, this hypothesis has not been tested yet.

Finally, regarding implementation, it was observed that changing the order of fusing the measurements had a negligible impact to the state estimation. In addition, the importance of setting reasonable thresholds for the state variables needs to be stressed, as it can prevent the estimation from diverging, especially in cases with high uncertainty.

6.2. Suggestions for future research

Future research could focus on the alleviation of the inefficiencies of the developed model, its expansion, as well as topics that were derived from the discussion of the results of this research.

An important element of the developed model that could be improved is the definition selected for the outflow limit factors. Defining and using outflow limit factors per turn direction, as well, would lead to an improvement in the fusion of flow measurements and consequently an improvement in state estimation, especially the estimation of the densities. Introducing sets of common fundamental diagram parameters is another suggestion that could improve estimation of the fundamental diagram parameters, as well as the speeds and densities.

Applying the model on a more congested network would shed light on its ability to estimate the speed at capacity and capacity per lane, too, instead of only the free flow speed. By fusing more measurements near capacity, it is expected that the estimation of those fundamental diagram parameters would improve, as the model would have more information on conditions related to these parameters.

A possibly determining factor that has not been examined at all in this research is latency. It is assumed that the measurements for each minute are immediately available at the end of the 60th second. However, in a real network, this assumption is not realistic, as measurement data may take several minutes to become available, depending on various

factors, such as the data size, transmission speed and the required processing of sensor data. A suggestion offered as a starting point when researching how to handle latency in a real network is the following: The traffic state of each time step (e.g. minute) would be predicted from the process model, based on the traffic state of the previous time step. Kalman Filtering would be applied for a time step several minutes in the past, as soon as measurement data for that time step is made available, resulting to a corrected traffic state for that time step. Using this corrected traffic state, the predictions of the traffic states of all subsequent time steps would then be recalculated. For example, assuming a latency of five minutes, the process would begin with the traffic state being predicted solely by the process model for the first six minutes. At the end of minute 6, the measurements for the first minute would be made available. The traffic state of minute 1 would then be corrected accordingly, by fusing the measurements for minute 1. This updated traffic state of minute 1 would be used to predict again the traffic state of minutes 2-7 using the process model. Similarly, at the end of minute 7, measurements for minute 2 would be available, leading to an updated traffic state for minute 2 and a new prediction for all subsequent minutes etc.

Finally, the use of the Kalman filter provides the opportunity to estimate other parameters as well. For example, routing could be affected through the use of the turning fractions, which could be included in the state vector and changed by the Kalman filter with the information from the measurements. The turning fractions would then replace the relevant junction modelling factor ("inflow reduce factor") in the equations, as the factor practically imitates the turning fractions. OD matrix calibration could also be included as an additional process to be followed at the end of each time step.

7. References

- Anthes, R. A. (1974). *Data assimilation and initialization of hurricane prediction models*. Journal of the Atmospheric Sciences, 31(3), pp. 702–719.
- Bhaskar, A. & Chung, E. (2013). *Fundamental understanding on the use of Bluetooth scanner as a complimentary transport data*. Transportation Research Part C: Emerging Technologies 37, 42-72.
- Briedis, P. & Samuels, S. (2010). *The accuracy of inductive loop detectors*. 24th ARRB Conference – Building on 50 years of road transport and research, Melbourne, Australia, 2010.
- Calvert, S.C., Snelder, M., Bakri, T., Heijligers, B. & Knoop, V.L. (2015). *Real-Time Travel Time Prediction Framework for Departure Time and Route Advice*. Transportation Research Record: Journal of the Transportation Research Board 2015 2490:, 56-64, DOI: 10.3141/2490-07.
- Daganzo, C. F. (1994). *The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory*. Transportation Research Part B: Methodological, 28(4), pp. 269–287.
- Daganzo, C. F. (1995). *The cell transmission model, part ii: Network traffic*. Transportation Research Part B: Methodological, 29(2), pp. 79–93.
- Deng, W., Lei, H. & Zhou, X. (2013). *Traffic state estimation and uncertainty quantification based on heterogeneous data sources: a three-detector approach*. Transportation Research Part B, 57, pp. 132-157.
- Duret, A. & Yuan, Y. (2017). *Traffic state estimation based on Eulerian and Lagrangian observations in a mesoscopic modeling framework*. Transportation Research Part B, 101, pp. 51-71.
- Eddy, D.M., Hollingworth, W., Caro, J.J., Tsevat, J., McDonald, K.M. & Wong, J.B. (2012). *Model Transparency and Validation: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7*. Value in Health 15 (2012), pp. 843-850.
- Fountoulakis, M., Bekiaris-Liberis, N., Roncoli, C., Papamichail, I. & Papageorgiou, M. (2017). *Highway traffic state estimation with mixed connected and conventional vehicles: Microscopic simulation-based testing*. Transportation Research Part C, 78, pp. 13-33.
- Herrera, J.C., Work, D.B., Herring, R., Ban, X., Jacobson, Q. & Bayen, A.M. (2010). *Evaluation of Traffic Data Obtained via GPS-Enabled Mobile Phones: The Mobile Century Field Experiment*. Transportation Research Part C, Vol. 18, No.4, 2010, pp. 568-583.
- INRIX (2013). *Economic and Environmental Impact of Traffic Congestion in Europe & the U.S.* URL: <http://inrix.com/economic-environment-cost-congestion/>

- Jiménez, F., Monzón, S. & Naranjo, J.E. (2016). *Definition of an Enhanced Map-Matching Algorithm for Urban Environments with Poor GNSS Signal Quality*. *Sensors* 2016, 16(2), 193.
- Kalman, R. (1960). *A new approach to linear filtering and prediction problems*. *Journal of Basic Engineering*, 82(1), pp. 35–45.
- Kalman, R.E., Bucy, R.S. (1961). *New results in linear filtering and prediction theory*. *Transactions of the ASME Series D* 83 (1961), pp. 95-108.
- Lebacque, J. P. (1996). *The Godunov scheme and what it means for first order traffic flow models*, in: Lesort, J., ed., *Proceedings of the 13th International Symposium on Transportation and Traffic Theory*, Lyon, France, pp. 647–677.
- Lighthill, M. & Whitham, G. (1955). *On kinematic waves II: A theory of traffic flow on long crowded roads*. *Proceedings of Royal Society*, 229A (1179), pp. 317–345.
- Martin, P.T. (2008). *Detector Technology Evaluation*. DCEE Technical Report, Department of Civil and Environmental Engineering, University of Utah, USA, pp. 66.
- Nantes, A., Ngoduy, D., Bhaskar A., Miska M. & Chung E. (2016). *Real-time traffic state estimation in urban corridors from heterogeneous data*. *Transportation Research Part C* 66 (2016), pp. 99-118.
- Owens, T. (1996). *Estimating the spatial accuracy of coordinates collected using the global positioning system*. Tech. rep., Onalaska, Wis.: National Biological Service, Environmental Management Technical Center.
URL: <http://purl.access.gpo.gov/GPO/LPS27981>
- Papageorgiou, M., Blosseville, J.M. & Hadj-Salem, H. (1990). *Modelling and real-time control of traffic flow on the southern part of boulevard peripherique in paris: Part i: Modelling*, *Transportation Research Part A: General*, 24(5), pp. 345–359.
- Payne, H. (1971). *Models of freeway traffic and control*, *Simulation Councils Proceedings Series: Mathematical Models of Public Systems*, 1(1), pp. 51–61.
- Richards, P. J. (1956). *Shock waves on the highway*. *Operations Research*, 4, pp. 42–51.
- Smits, E.-S., Bliemer, M.C.J., Pel, A.J. & van Arem, B. (2015). *A family of macroscopic node models*. *Transportation Research Part B* 74 (2015), pp. 20-39.
- Smulders, S. (1990). *Control of freeway traffic flow by variable speed signs*. *Transportation Research B* 24, pp. 111-132.
- Sod, G.A. (1985). *Numerical Methods in Fluid Dynamics*. Cambridge University Press, USA.
- Sunderrajan, A., Viswanathan, V., Cai, W. & Knoll, A. (2016). *Traffic state estimation using floating car data*. *Procedia Computer Science* 80 (2016), pp. 2008-2018.

- Tampère, C., Corthout, R., Cattrysse, D. & Immers, L.H. (2011). *A generic class of first order node models for dynamic macroscopic simulation of traffic flows*. Transportation Research Part B 45 (2011), pp. 289-309.
- Tampère, C. & Immers, B. (2007). *An extended kalman filter application for traffic state estimation using CTM with implicit mode switching and dynamic parameters*. In: Proceedings of the 2007 IEEE Intelligent Transportation Systems Conference, IEEE, Seattle, WA, USA, pp. 209–216.
- Technical University of Crete, Dynamic Systems and Simulation Laboratory & Messmer, A. (2012). *METANET – A simulation program for motorway networks – Documentation*. URL: https://www.researchgate.net/profile/Albert_Messmer/publication/282285780_METANET_a_macroscopic_simulation_program_for_motorway_networks/links/568a788e08aebccc4e19fcc8/METANET-a-macroscopic-simulation-program-for-motorway-networks.pdf
- Treiber, M. & Kesting, A. (2009). *Reconstructing the traffic state by fusion of heterogeneous data*. Computer-Aided Civil Infrastructure Engineering, 26, pp. 408-419.
- U.S. Department of Defense (2008). *Global Positioning System Standard Positioning Service Performance Standard*. National Coordination Office for Space-Based Positioning, Navigation and Timing, September 2008. URL: <http://www.gps.gov/technical/ps/2008-SPS-performance-standard.pdf>
- Van de Pijpekamp (2015). *ICT Support*. Lecture Notes, course Intelligent Transport Systems, University of Twente, September 2015.
- Van der Vlist, M., Wismans, L., Van Beek, P. & Suijs, L. (2016). *Virtual Patrolling*. Transportation Research Procedia 14 (2016), pp. 3370-3379.
- Van Lint, J.W.C. & Hoogendoorn, S.P. (2009). *A robust and efficient method for fusing heterogeneous data from traffic sensors on freeways*. Computer-Aided Civil Infrastructure Engineering, 24, pp. 1-17.
- Van Lint, J.W.C., Hoogendoorn, S.P. & Hegyi A. (2008). *Dual EKF state and parameter estimation in multi-class first-order traffic flow models*, in: Preprints of the 17th IFAC World Congress, IFAC, Seoul, Korea, pp. 14078–14083.
- Van Lint, J.W.C. & Van Hinsbergen, C.P.I.J. (2012). *Short-Term Traffic and Travel Time Prediction Models*. TR Circular E-C168 (2012), pp. 22-41.
- Wang, R. & Work, D.B. (2013). *Joint parameter and state estimation algorithms for real-time traffic monitoring*. USDOT Region V Regional University Transportation Center Final Report, NEXTRANS Project No 0971Y04. URL: <https://www.purdue.edu/discoverypark/nextrans/assets/pdfs/Joint%20parameter%20and%20state%20estimation%20algorithms%20for%20real-time%20traffic%20monitoring%20097%20.pdf>

- Wang, Y. & Papageorgiou, M. (2005). *Real-time freeway traffic state estimation based on extended Kalman filter: a general approach*. Transportation Research Part B 39 (2005), pp. 141-167.
- Wang, Y., Coppola, P., Tzimitsi, A., Messmer, A., Papageorgiou, M. & Nuzzolo, A. *Real-Time Freeway Network Traffic Surveillance: Large-Scale Field-Testing Results in Southern Italy*. IEEE Transactions on Intelligent Transportation Systems, Vol. 12, No. 2, June 2011.
- Wang, Y., Papageorgiou, M. & Messmer, A. (2007). *Real-time freeway traffic state estimation based on extended Kalman filter: A case study*. Transportation Science, Vol. 41, No. 2, May 2007, pp. 167–181.
- Wang, Y., Papageorgiou, M. & Messmer, A. (2008). *Real-time freeway traffic state estimation based on extended Kalman filter: Adaptive capabilities and real data testing*. Transportation Research Part A 42 (2008), pp. 1340-1358.
- Wismans, L., de Romph, E., Friso, K. & Zantema, K. (2014). *Real time traffic models, decision support for traffic management*. 12th International Conference on Design and Decision Support Systems in Architecture and Urban Planning, DDSS 2014. Procedia Environmental Sciences 22 (2014), pp. 220-235.
- Yuan, Y. (2013). *Lagrangian multi-class traffic state estimation*. Doctoral dissertation, Delft University of Technology.
- Yuan, Y., Van Lint, H., Van Wageningen-Kessels, F. & Hoogendoorn S. (2014). *Network-Wide Traffic State Estimation Using Loop Detector and Floating Car Data*, Journal of Intelligent Transportation Systems, 18:1, 41-50, DOI: 10.1080/15472450.2013.773225

Appendix A: Simulation settings and parameters (test network)

1. Standard OD Matrix

Time steps 1-20 and 41-60:

Centroids	1	2	3	4	5	6	7
1	0,000	1,500	1,500	1,500	1,500	1,500	1,500
2	0,500	0,000	0,350	0,350	0,350	0,350	0,350
3	1,500	1,500	0,000	1,500	1,500	1,500	1,500
4	1,500	1,500	1,500	0,000	1,500	1,500	1,500
5	0,700	0,700	0,700	0,700	0,000	0,700	0,700
6	1,000	0,000	1,000	3,500	2,700	0,000	3,075
7	2,000	0,800	1,100	1,500	0,800	1,500	0,000

Time steps 21-40:

Centroids	1	2	3	4	5	6	7
1	0,000	0,750	1,000	0,750	1,000	0,750	1,000
2	0,250	0,000	0,250	0,250	0,250	0,250	0,250
3	1,000	1,000	0,000	1,000	1,000	1,000	1,000
4	1,000	1,000	1,000	0,000	1,000	1,000	1,000
5	0,500	0,500	0,500	0,500	0,000	0,500	0,500
6	0,500	0,500	0,500	0,500	0,500	0,000	0,500
7	1,000	1,000	0,500	0,500	0,500	0,500	0,000

2. Modified OD Matrix

Time steps 1-20 and 41-60:

Centroids	1	2	3	4	5	6	7
1	0,000	1,400	1,600	1,360	1,550	1,650	1,550
2	0,550	0,000	0,400	0,320	0,330	0,370	0,360
3	1,350	1,400	0,000	1,550	1,580	1,650	1,400
4	1,400	1,600	1,350	0,000	1,550	1,400	1,600
5	0,770	0,750	0,680	0,650	0,000	0,750	0,720
6	1,080	0,000	1,100	3,350	2,500	0,000	3,300
7	1,900	0,750	1,210	1,570	0,880	1,450	0,000

Time steps 21-40:

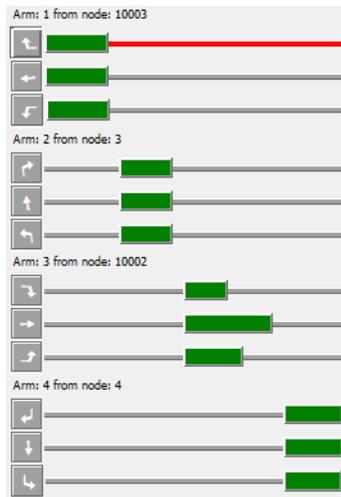
Centroids	1	2	3	4	5	6	7
1	0,000	0,700	1,070	0,790	1,100	0,800	1,060
2	0,230	0,000	0,270	0,230	0,270	0,240	0,270
3	1,050	1,090	0,000	1,100	0,950	0,920	0,970
4	1,100	1,040	0,980	0,000	0,960	1,020	0,940
5	0,480	0,550	0,490	0,450	0,000	0,480	0,550
6	0,460	0,480	0,530	0,550	0,540	0,000	0,510
7	1,100	1,050	0,450	0,480	0,530	0,550	0,000

3. Initial fundamental diagram parameters set

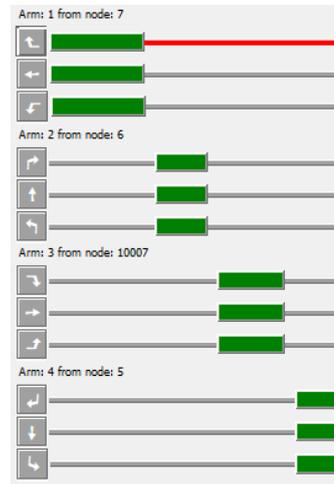
Link ID	v^{free} (km/h)	f^{cap} (veh/h)	v^{cap} (km/h)		Link ID	v^{free} (km/h)	f^{cap} (veh/h)	v^{cap} (km/h)
-15	50	1500	35		1	46	1650	36
-14	49	1510	36		2	45	1660	37
-13	48	1520	37		3	55	1670	38
-12	47	1530	38		4	54	1680	32
-11	46	1540	32		5	53	1690	33
-10	45	1550	33		6	52	1700	34
-9	55	1560	34		7	51	1710	35
-8	54	1570	35		8	50	1720	36
-7	53	1580	36		9	49	1730	37
-6	52	1590	37		10	48	1740	38
-5	51	1600	38		11	47	1750	32
-4	50	1610	32		12	46	1760	33
-3	49	1620	33		13	45	1770	34
-2	48	1630	34		14	55	1780	35
-1	47	1640	35		15	54	1790	36

4. Traffic light plans

Traffic light plans (green times per turn direction) for the R1 and R2 regulated intersections respectively. Cycle time is 60 seconds to coincide with the arrival of each round of measurements.



R1



R2