Traffic state estimation and prediction in freeways and urban networks

Andrés Ladino Lopez

UNIVERSITÉ GRENOBLE ALPES

THÈSE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spècialité : **Automatique**

Arrêtè ministériel : 25 mai 2016

Prèsentée par
Andrés Alberto LADINO LÓPEZ

Thèse dirigée par **Carlos CANUDAS DE WIT** et
codirigée par **Alain KIBANGOU** et **Hassen FOURATI**

préparée au sein du
laboratoire (INRIA/GIPSA-lab)
dans **l'École doctorale Electronique Electrotechnique Automatique & Traitement du signal (EEATS)**

Traffic state estimation and prediction in freeways and urban networks

Thèse soutenue publiquement le **08 Mars 2018**,
devant le jury composé de:

Christophe BERENGUER
Professeur, Grenoble-INP, Président du Jury
Habib HAJ-SALEM
Directeur de recherche, IFSTTAR, Rapporteur
Lyudmila MIHAYLOVA
Full professor, Sheffield University, Rapporteur
Karl H. JOHANSSON
Chair professor, Automatic Control KTH, Examinateur
Hassen FOURATI
Maître de conférences, Université Grenoble Alpes, Co-encadrant
Alain KIBANGOU
Maître de conférences, Université Grenoble Alpes, Co-encadrant

UNIVERSITÉ GRENOBLE ALPES

ÉCOLE DOCTORALE EEATS

Electronique, Electrotechnique, Automatique & Traitement de signal

T H È S E

pour obtenir le titre de

Docteur en sciences

de l'Université Grenoble Alpes

Mention : AUTOMATIQUE

Présentée et soutenue par

Andrés Alberto LADINO LÓPEZ

Traffic state estimation and prediction in freeways and urban networks

Thèse dirigée par Carlos CANUDAS DE WIT

préparée au laboratoire INRIA/GIPSA-lab

soutenue le 08 Mars 2018

Jury :

Rapporteurs :	Habib HAJ-SALEM	-	IFFSTAR (Noisy-le-Grand, France)
	Lyudmila MIHAYLOVA	-	Sheffield University (Sheffield, England)
Encadrants :	Hassen FOURATI	-	GIPSA-lab (Grenoble, France)
	Alain KIBANGOU	-	GIPSA-lab (Grenoble, France)
Examinateurs:	Karl H. JOHANSSON	-	Royal Institute of Technology KTH (Stockholm, Sweden)
President du jury:	Christophe BERENGUER	-	GIPSA-lab (Grenoble, France)

Acknowledgement

My special thanks to all my colleagues and friends at within the NeCS team, thank you for your help with the adaptation to this country and for being mentally open to make good conversations. My special thanks to my advisors Carlos, Alain and Hassen for their guide during this process of research.

I also gratefully acknowledge the financial support from the Centre Nationale de la Recherche Scientifique (CNRS) through grant 619435 funded by the EU FP7 - SPEEDD project.

Finally, I would like to thank those who were not really concerned with my work in technical terms but whom made their contribution with support and encouragement. They deserve my deepest gratitude all the same. You know who you are.

Contents

List of Figures

List of Tables

List of Symbols

$v(x,t)$: Continuous velocity space/time field.

$v(x_i, s_j)$: Discrete velocity space/time field.

$DTT(x_0, x_n, t_0)$: Dynamic travel time (DTT) from x_0 to x_n at t_0.

$ITT(x_0, x_n, t_0)$: Instantaneous travel time (ITT) from x_0 to x_n at t_0.

$\tilde{v}_l(x_i, t_\tau)$: Imputed velocity from temporal coherence.

$\tilde{v}_s(x_i, t_\tau)$: Imputed velocity from spatial coherence.

$\tilde{v}_d(x_i, t_\tau)$: Imputed velocity from historical coherence.

$\mathcal{K}$: Number of clusters.

$D_{\mathcal{K}}$: Total distortion for a clustered set with $\mathcal{K}$ clusters.

$\mu_q(k)$: Cluster's centroid q at sample time k.

$\Delta\mu_q(k)$: Cluster derivative's centroid q at sample time k.

$y_h(k)$: DTT for day h at sample time k.

$w_{d,q}(k)$: DTT noise for day d with respect to cluster q

$R_q(k)$: Sample based variance associated to cluster q at sample time k.

$v_{d,q}(k)$: DTT noise derivative for day d with respect to cluster q

$V_q(k)$: Sample based variance associated to derivative of cluster q at sample time k.

$\bar{y}_{d,q}(k)$: Level measurement of the DTT of day d with respect to cluster q at sample time k.

$P_{d,q}(k)$: Variance of DTT model of day d with respect to cluster q at sample time k.

$K_q(k)$: Kalman gain for the q-th filter.

$\hat{y}_{d,q}(k)$: Prediction for DTT of day d with respect to cluster q at sample time k.

$S_{d,q}(k)$: Temporal similarity measurement between day d and cluster q.

$U(k)$: Cluster weight similarity matrix .

λ_q: Weight assigned to the q-th cluster.

$\hat{y}_d(k)$: Prediction for DTT of day d at time t_k.

$AE(x_i, t_\tau)$: Absolute error in location x_i at time t_τ.

$APE(k_0 + h, d)$: Absolute percentage error for the prediction at departure time $k_0 + h$ for day d.

$\rho_i(k)$: Density of road i at sample time k.

$\varphi_i^{\text{out}}(k)$: Outgoing flow from road i at sample time k.

$\varphi_i^{\text{in}}(k)$: Ingoing flow from road i at sample time k.

$\hat{\rho}_i(k)$: Estimated density for road i at sample time k.

$\hat{\varphi}_i^{\text{out}}(k)$: Estimated outgoing flow for road i at sample time k.

$\bar{\rho}_i(k)$: Density pseudo observation for road i at sample time k.

$\bar{\varphi}_i^{\text{out}}(k)$: Outgoing flow observation for road i at sample time k.

$\mathcal{I}_g$: Set of roads for junction g.

$\mathcal{I}_g^+$: Set of outgoing roads from junction g.

$\mathcal{I}_g^-$: Set of incoming roads to junction g.

$\mathcal{I}$: Set of roads for the Road Traffic Network (RTN).

$\mathcal{R}_g$: Set of splitting ratios relating all outgoing-ingoing flows from all roads $\mathcal{I}_g^-$ to all roads $\mathcal{I}_g^+$.

$\mathcal{R}$: Set of all splitting ratios for the Road Traffic Network (RTN).

$\mathcal{H}$: Weighted directed graph representing the Road Traffic Network (RTN).

$\mathcal{G}$: Set of all junctions in the Road Traffic Network (RTN)

$D(\rho_i)$: Demand function for road i.

$S(\rho_i)$: Supply function for road i.

$\bar{\mathcal{D}}_g$: Free flow constraints set for junction g.

$\bar{\mathcal{S}}_g$: Congestion constraints set for junction g.

$\mathcal{S}_g^{\max}$: Supply boundary constraints set for junction g.

$\mathcal{D}_g^{\max}$: Demand boundary constraints set for junction g.

$\mathcal{D}_g$: Demand constraints set for junction g.

$\mathcal{S}_g$: Supply constraints set for junction g.

$\mathcal{P}_g^{\varphi}$: Set of constraints for outgoing flow φ^{out} at junction g.

$\mathcal{P}_g^{\rho}$: Set of constraints for density at junction g.

$\mathcal{P}^{\varphi}$: Set of constraints for outgoing flow φ^{out} for Road Traffic Network (RTN).

$\mathcal{P}^{\rho}$: Set of constraints for density ρ for the Road Traffic Network (RTN).

$\Psi(\rho_i)$: Speed-density fundamental diagram function for road i

$\Phi(\rho_i)$: Flow-density fundamental diagram function for road i

List of Acronyms

ACF	Auto Correlation Function
AE	Absolute Error
AP	Access Point / Station Point
APE	Absolute Percentage Error
AID	Aggregated Indicator Data
ARZ	Aw-Rascle-Zheng
ARD	Aggregated Raw Data
ARIMA	Autoregressive Integrated Moving Average
AS	Average Speed
BLUE	Best Linear Unbiased Estimator/Estimation
CL	Congestion Length
CKF	Clustered Kalman filter
CTM	Cell Transmission Model
DIR-CE	Direction Interdépartementale des Routes Centre-Est
DPS	Distributed Parameter System
DTT	Dynamic Travel Time
ECFM	Error Covariance Fusion Method
FCD	Floating Car Data
FI	Fluidity Index
FC	Fuel Consumption
GTL	Grenoble Traffic Lab
GPRS	General Packet Radio Service
GPS	Global Positioning System
INRIA	Institut National de la Recherche en Informatique et Automatique
INSEE	National Institute of Statistics and Economic Studies

IAID	Instantaneous Aggregated Indicator Data
IRD	Individual Raw Data
ITT	Instantaneous Travel Time
ITS	Intelligent Transportation System
KF	Kalman Filter
LP	Linear Programming
LWR	Lighthill-Whitham-Richards
MLD	Magnetic Loop Detector
NeCS	Networked Control Systems
OD	Origin Destination
PACF	Partial Auto Correlation Function
PAID	Predicted Aggregated Indicator Data
PDE	Partial Differential Equation
PLS	Percentage of Lost Samples
PRS	Percentage of Recovered Samples
PW	Payne-Whitham
RTT	Real Travel Time
RTN	Road Traffic Network
SMM	Switching Mode Model
SARIMA	Seasonal Autoregressive Integrated Moving Average
SCP	Station / Collection Point
PSFM	Past Similarity Fusion Method
SPEEDD	Scalable Proactive Event-Driven Decision-making
TT	Travel Time
TTT	Total Travel Time
TTD	Total Travel Distance
TSC	Time Spent in Congestion
TSE	Traffic State Estimation

TSP	Traffic State Prediction
TSEP	Traffic State Estimation & Prediction
V2V	Vehicle to Vehicle
V2I	Vehicle to Infrastructure

Abstract

Workforce centralization, population and economic growth alongside continued urbanization are the main causes of congestion. As cities strive to update or expand aging infrastructure, the management of big data, new models and analytics to better understand and help to reduce traffic congestion are crucial to the health and development of our smart cities of XXI century. Traffic support tools specifically designed to detect, forecast and alert these conditions are highly requested nowadays. Short-term forecasting as well as traffic state estimation are important tasks among these tools for traffic management and they have received a lot of attention during the last decade. In the actual era of data deluge, measurements collected by multiple sensors are important sources of information that require analysis, classification and processing in order to detect patterns and behaviors that can be exploited in traffic estimation and prediction.

This PhD work is dedicated to study techniques to estimate and forecast conditions about a traffic network. First, the problem of Dynamic Travel Time (DTT) short-term forecasting based on data driven methods is considered. Two fusion approaches are proposed to compute short-term forecasts from clustered time series. The first approach uses the error covariance matrix to fuse individual forecasts based on best linear unbiased estimation principles. The second technique exploits similarity measurements between the signal to be predicted and clusters detected in historical data and it performs a fusion as a weighted average of individual forecasts. Tests over real data were implemented in the study case of the Grenoble South Ring, it comprises a highway of 10.5Km monitored through the Grenoble Traffic Lab (GTL) a real time application implemented to provide traffic information for the general public and traffic state agencies.

Based on the previous study, the problem of simultaneous density/flow reconstruction is considered in urban networks based on heterogeneous sources of information. The traffic network is modeled within the framework of macroscopic traffic models, where the Lighthill-Whitham-Richards (LWR) conservation equation is adopted followings a piecewise linear fundamental diagram. The estimation problem considers two key principles. First, the error minimization between the measured and reconstructed flows and densities, and second the equilibrium state of the network which establishes flow propagation within the network. Both principles are integrated together with the traffic model constraints established by the supply/demand paradigm. Finally the problem is casted as a constrained quadratic optimization with equality constraints in order to shrink the feasible region of estimated variables. Some simulation scenarios based on synthetic data for a manhattan grid network are provided in order to validate the performance of the proposed algorithm.

Chapter 1

Introduction

Real life traffic commuting is affected by multiple factors, among them, personal decisions such as origin-destination OD, route choice or way of commuting. In a world where cities tend to grow, congestions become more and more important due to the increasing urbanization which overpasses the rate of development of transport infrastructures. Even though, these aspects may be planned and anticipated other bearings add uncertainty to the traffic commuting problem. Driver behaviors such as reaction time, driving style and traffic condition adaptation introduce new variables to the problem. The impact of these factors is regularly seen in acceleration increments which are sources of pollution and moreover, as a general consequence, reduce the throughput of the traffic network affecting individual commuting times. Specific irrational behaviors are difficult to predict, nevertheless new developments in traffic sensor technologies have increased the amount of traffic information. The development of state estimation and short-term forecasting[1] techniques for traffic networks based on real measured conditions of the network are part of the strategies to create decision support tools for regular commuters.

In this dissertation we consider in a first phase, the problem of Dynamic Travel Time (DTT) short-term forecasting, this indicator provides useful information for multiple users and it is common in actual traffic simulators and applications such as [CDW+15], [Pap+10], [Lin04]. Generally, forecast algorithms in the literature are designed to satisfy a set of constraints given by a traffic model. Since, some factors in driving behaviors are difficult to model and regularly more difficult to predict, this dissertation considers initially an approach conducted by data driven models that infer general relationships between observable traffic variables and historical patterns observed in the network. For data driven approaches, the collection of information can be classified in order to facilitate the data analysis. The separation of the information into multiple groups can be achieved through unsupervised learning techniques such as K-means, where each cluster is created in an automatic way from traffic data patterns.

Since there exists historical patterns, it means the existence of same historical trends during the current day, are likely to happen. In order to construct a consistent forecast, the historical pattern should be smoothed with the current traffic condition. By applying clustering approach, it is possible to characterize different clusters through centroids containing the mean of the data and a given dispersion around it. The evolution of the centroid can be used as future observation that can feed a Kalman Filter (KF) [Oje14]. Therefore the

[1] We precise that the term *prediction* is equivalent to the task of estimating information one step ahead in the future starting from current information. *Short-term forecast* refers to the task of estimating information multiple steps ahead based on current information.

prediction problem can be viewed as a filtering one. Nevertheless the assignment of the observation during the current day to a specific cluster remains an open issue since we don't know its future. The way these forecasts can be selected or mixed, in order to determine the output forecast given to the user, is one of the subjects of this research.

The exploration of data-driven models relies on patterns found on the data but does not take physical constraints of the model into account. Moreover for complex networks the amount of required data starts to be demanding. Despite of existent systems, the problem of how to fuse multiple sources of information to estimate the state of the network remains a challenging problem [ACH14], [ZL11]. Multiple systems appear today as sources of traffic information. The deployment of Magnetic Loop Detectors (MLDs) and stationary collectors have been the traditional way to measure traffic, but they are expensive and highly cost to maintain. Technologies such as Floating Car Data (FCD) appear as a less expensive alternative which requires less maintenance effort with respect to fixed sensors since it exploits existing communication architectures. We have explored in a second phase the problem of assimilation of heterogeneous sources of information for Traffic State Estimation (TSE).

In this introductory chapter we outline the context and background of the traffic problem from the point of view of estimation and forecast, then we motivate why nowadays it is an important problem. We introduce also the Grenoble Traffic Lab (GTL), an experimental platform used as study case to validate experimental results produced during this research. We provide some highlights on the problem of heterogeneous sources of information and how we make use of this information to solve the TSE problem which can be casted as a joint flow-density estimation problem. Next, we present the main objectives of this thesis and narrow down the scope of the developed research.

1.1 Traffic challenges

Traffic congestion has become one of the major issues in most of the metropolises. It is believed that strategies dedicated to identification of congestion characteristics and its origins are the first steps towards a better understanding of the problem. Congestion, both in terms of perception as well as measured condition, impacts the utilization of Road Traffic Networks (RTNs) [MRRR12]. Effects of traffic congestion are directly observed in environmental impact and productivity costs on the society.

Two main principal categories are causes of congestions. *Micro* level factors related to local traffic conditions and *macro* level factors related to the overall use of the traffic infrastructure. In that sense, congestion starts from the micro level (e.g. in roads or intersections) and scales to a macro level where multiple roads are affected with certain severity of congestion. While micro level causes arise on accidents, poor traffic light synchronization, events or weather conditions, macro level factors are more justified in land-use patterns, employment, income levels [WVT03]. Traffic flow theory tries to capture most of these situations and model them as part of mathematical equations [GP06], [TK13]. Nevertheless, in most of these situations a good sensor network is required to measure the state or the mathematical models to describe

particular behaviors grow in complexity.

The traffic problem is solved by regulating the state of the RTN which can be correctly managed through control strategies [Pap+08]. Efficient road traffic management policies strongly depend on the vehicle density information. In road maintenance and traffic monitoring, it is essential to inform the state of the network and to perform preventive maintenance; in designing traffic light control policies, its evolution is essential to construct efficient feedback laws, see [GGCDW16]. Strategies like ALINEA [PHSB90], METALINE [Zha+01], or predictive approaches [Bra+17] constitute alternative ways to approach the problem. In most of the cases control strategies assume full state knowledge on the RTN which is in reality impossible due to the costs of installation and maintenance of these sensors in the network. Furthermore, some of these techniques consider control policies based on forecasted of information obtained from the actual condition of the network [CSS09], [FSS15]. This need is one of the main motivations for the current work. This dissertation focuses its attention on the Traffic State Estimation (TSE) and Traffic State Prediction (TSP) problems. The formal introduction of the problems is detailed in Chapter 2.

1.2 Case-study: The city of Grenoble

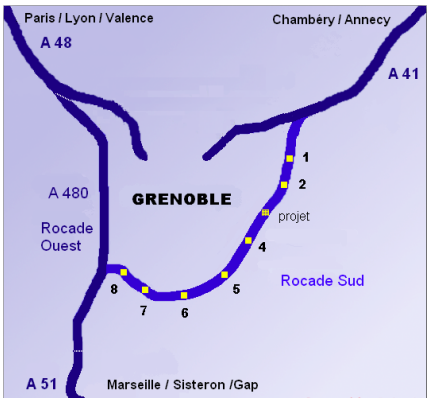
Grenoble covers an area of 18.13km^2 with a total population of 159,953[2]. The city is surrounded by mountains and it constitutes a confluence of rivers Isère and Drac in the north west side from Lyon as well as highways coming from Annecy, Valence going towards Marseille (See Fig. 1.1a, 1.1b). The natural merge existing in the south of the city creates congestions specially during peak hours. The south ring of Grenoble (a.k.a. Route Nationale 87) is a highway enclosing the southern part of the city from A41 to A480 (see Fig. 1.1b). It consists of two carriageways with two lanes. The GTL is a project initiated in 2011 which collects traffic data for this highway (See [CDW+15] and Chapter 7 for further details). The project gathers data from sensors in the east-west direction and it is intended to provide real-time traffic information on the highway.

After the publication of the results in [CDW+15], [Oje14] some functionalities where added to the platform in order to improve the robustness of the online application. Configuration in the sensor network, power supply outages, preventive and corrective maintenances have caused data loss within the network. Since 2014 a series of procedures and tools have been implemented to improve the diagnosis of quality of the data and to better assess in real time the status of the network. More details regarding this aspect are provided in chapter 7.

[2] as of January 2014, according to the National Institute of Statistics and Economic Studies (INSEE)

(a) Grenoble city location in France [a]

[a] http://balanceyourworldbookkeeping.com/where-is-grenoble-france-on-a-map.html



(b) Rocade Sud Grenoble *RN*87. (Source Wikipedia [a])

[a] https://fr.wikipedia.org/wiki/Rocade_sud_(Grenoble)

(c) Pollution peak policy. (Source Le dauphiné - Published. 28/05/2011. [a])

[a] http://www.ledauphine.com/isere-sud/2011/05/27/on-va-rouler-a-70-km-h-sur-la-rocade-sud

Figure 1.1: South ring / Rocade Sud Grenoble

1.3 Heterogeneous data sources in traffic

The last decade has witnessed the creation of multiple decision support tools for traffic management. The development of robust network of sensors, the improvement on technologies for measuring traffic conditions and the increase on data transmission rates have been a key factor for this development. As a result an era of big data in traffic has been impulsed during the last years. Applications for best path selection are on the market and flagship products of millionaire entreprises [Waz], [Her] are available for common use. These products rely on measurements collected and aggregated by each application. European cities along the last decade have invested in deployment of MLD in order to track the traffic conditions. The existence of these infrastructures in addition with information collected from mobile applications constitute a more robust source for traffic estimation. Nevertheless, the nature of the measurements is different. While GPS traces are collected individually, only aggregated versions of the speed profiles are accessible (a.k.a FCD) at specific spots of the network (See Fig. 1.2b) due to penetration rates and privacy issues. On the other hand, MLD are normally installed just in main arterials (See Fig. 1.2a), due to maintenance costs and the required communication infrastructure, leaving secondary roads without measured information important for the TSE task.

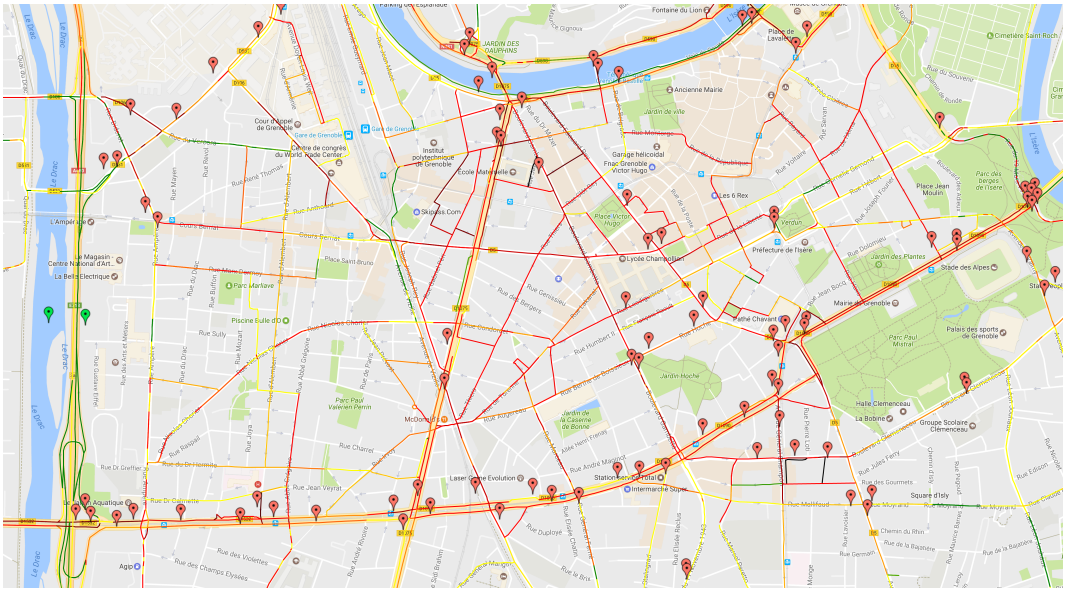
This evolution has been particularly quick in the last few years, due to the prolific growth of GPS sensors deployed and the datasets they generate. This work considers the problem of data fusion seen as an integration of multiple sources. In particular we consider the integration of a spatially dense source of FCD with MLD sensors which are spatially sparse. We face the problem of fusing these types of data to reconstruct the state of the RTN with a particular focus in urban traffic networks. This approach is presented and detailed in Chapter 6.

1.4 Objectives of the research

This dissertation has been elaborated as part of the European project Scalable Proactive Event-Driven Decision-making (SPEEDD) (EU-FP7 619435)[3]. The goal of the proactive event-driven traffic management use case within this project is to make decisions in order to attenuate congestions. Indeed, eliminating completely a congestion in traffic network is neither an affordable, nor feasible goal. Nevertheless, decision support tools can be designed in order to reduce its occurrence and its impact on traffic network users. The idea of proactive traffic management is to employ modern software and hardware to mitigate the congestions' effect by taking decisions in advance to forecasted events.

The primary objective of this dissertation is to develop strategies for short-term traffic forecasting that are efficient by means of their computation cost and reproducible at scale of a real application. The goal of these prediction strategies is to create decision support tools that conduct to improvement of the system's behavior in terms of regular indices for drivers such as Travel Times (TTs) or more complete general indicators such as Total Travel Time (TTT) and

[3] `http://speedd-project.eu`



(a) Magnetic Loop Detector (MLD) measurements in downtown area

(b) Floating Car Data (FCD) measurements [a]

[a] Source: TomTom ®

Figure 1.2: Data traffic measurements examples in Grenoble

Total Travel Distance (TTD). Such indexes are regular criteria used to determine efficiency of traffic infrastructures or even used to design control policies.

A second objective is an extension of these set of methods into a real time application. The full dense network of sensors installed in the South Ring represents nowadays a rich source of information containing events such as regular congestions, stop and go waves, unexpected congestions and seasonal traffic regimes. The infrastructure of sensors initially deployed with the project MOCoPo and then extended with project HYCON2 has been used to collect data for more than 4 years [CDW+15]. This rich source of data encourages the development of online forecasting techniques compatible with dynamic adaptation to the current state of the highway. The objective is then to provide real-time techniques suitable to deal with big datasets and computationally efficient in an online environment.

Finally we aim to explore the TSE problem of simultaneous state density and flow reconstruction based on heterogeneous sources of information. Traditional technologies like MLD are standard measurement devices for traffic systems. New sources of information such as FCD are common nowadays, they are less expensive and require less maintenance effort with respect to fixed sensors since they exploit existing communication architectures. The combination of both sources of information is then used in order to reconstruct the RTN state.

1.5 Contributions

The contributions of this work to the traffic problem are related to the framework of DTT short term forecasting and the problem of traffic density/flow estimation based on multiple sources of data.

First we conduct a detailed explanation on the data workflow and we provide details on the computation of the DTT. The nature of this indicator is based on the dynamic of the speed variable in space and time. In general in order to compute TT at current time it is required a forecasted value of the speed. We propose hence forecast strategy based on current day data and historical data. Historical data is organized into several clusters also denominated partitions. For each cluster, a predictor is designed based on a Kalman filtering strategy. A first approach considers fusion of predictions in the Biased Linear Unbiased Estimation BLUE sense in order to get the optimal prediction. These results are published in [Lad+16b].

A second approach considers the case where DTT is dynamically clustered using a K-means algorithm along the day. Information on the level and the trend of the centroid of each cluster is used to devise a mono cluster based predictors computationally simple to be implemented. To take into account the lack of information in the cluster assignment for the new predicted values, a weighted average fusion based on a similarity measurement is proposed to combine the predictions of each model. For this last case, the algorithm is deployed in a real time application and the performance is evaluated using real traffic data from the South Ring of the Grenoble city in France. This part of our research has been published in [Lad+17] .

As a complement to the previous research, an online real time application was deployed and upgraded from the previous version developed in [CDW+15]. In this version, a whole new system to process events was implemented and new functionalities have been included. Among them the platform is now capable of assessing quality of the data in a precise way, which allows better dataset selection for upcoming research studies, it computes supplementary traffic indicators oriented for traffic operators and regular users for better understanding and use of the highway. Furthermore, support decision tools for traffic operators in order to place speed panels were added. Finally, an official version of this platform was released to the public last December 2016.

Finally we dealt with the problem of flow/density reconstruction in RTNs based on heterogeneous information sources. Following a standard LWR macroscopic modeling approach, the network is partitioned in cells, whose vehicle densities change dynamically in time according to first order conservation laws. We focus our attention particularly on the problem of simultaneous flow and density estimation using as sources of information, MLDs, precise but expensive and regularly with low spatial coverage, and FCD, less precise due to low penetration rates, but already available on most of main roads. A data fusion algorithm is proposed to merge the two sources of information to estimate the network state. These results have been published in [Lad+18].

1.6 Structure of the manuscript

The organization of the rest of the manuscript is as follows:

1. Chapter 2 presents the framework of the TSE and TSP problems as well as the state of the art. It considers the explanation of existing measurement systems and techniques already proposed in order to perform density reconstruction and short-term TT forecast.

2. In chapter 3 we motivate the problem of DTT estimation and possible ways to perform this computation. The chapter introduces first the relationships between microscopic and macroscopic traffic models and from there it derives the computation of DTT. Experiments are validated based on some real data scenarios and a data imputation technique is required as part of the data processing workflow before the final computation.

3. Chapter 4 introduces the problem of time series clustering as a way to differentiate multiple traffic regimes. The chapter is particularly focused on the clustering technique and serves as a preliminary stage for the prediction method. We present and discuss parameters to perform the clustering for a set of a time series and we also explore tests to validate the cluster assignment.

4. In chapter 5 two short term forecast techniques are presented. We propose a simple single predictor data-driven model that is able to produce forecasts for a particular regime of traffic. Then we motivate fusion strategies of multiple forecasts to reduce the

error of forecasting. We present two fusion mechanisms to obtain a prediction and test validations.

5. In chapter 6 we introduce the problem of fusion of multiple sources of data for flow and density reconstruction. The approach for reconstruction is based on the principle of a LWR model for the system and makes use of FCD combined with MLD in order to estimate the state of the system.

6. Chapter 7 consolidates the work from chapters 3, 4 and 5 in a real-time application developed to perform computation of multiple traffic indicators. The application retrieves data from a real network of sensors in the South Ring of Grenoble and performs aggregation and imputation of the measured variables in time and space. A process of data quality assessment is also deployed to verify the quality of the received information. Multiple indicators are commonly used to detect and forecast information of a traffic network.

7. Chapter 8 summarizes the results of the work. The first part will provide analysis of the obtained results and the forecasting/estimation methodologies developed in this thesis. The second part will summarize the main contributions. Finally, the last part will present some possible future perspectives.

Every chapter also contains numerical experiments based on real data or specific designed scenarios, as well as comments and illustrations of the obtained numerical results.

Chapter 2

Models and methods for traffic state estimation and prediction

Contents

2.1 Overview

Europe is the most urbanized continent in the world: over 80% of its population lives in towns and cities [Wal07]. The run towards the urbanization has as direct consequence development and expansion of cities, which in return, leads to increased vehicular traffic. The underlying motivation that brings people to congregate in large urban areas also causes often unbearable levels of traffic congestion in urban roads. Year by year the traffic monitoring company INRIX ® presents statistics of indicators on congestion in urban networks [INR16], revealing the increments of time spent in congestion in big urban poles.

However, along with increased congestions, recent years have also brought some promising improvements, particularly in the areas where the Intelligent Transportation System (ITS) have been put into practice. When dealing with urban traffic, these systems bring innovation introducing new sensing mechanisms intended to improve connectivity towards a full measurement of the network. The existence of multiple sources of measurement and heterogenity on sensor data has promoted the development of traffic monitoring systems and

more recently Vehicle to Vehicle (V2V) and Vehicle to Infrastructure (V2I) technologies are promising sources of new information. This explains why the overall motivations to tackle the problem of improving traffic management are quite strong for traffic researchers and engineers.

In this chapter, we introduce and summarize some of the existing approaches to the TSE and TSP problems. The chapter presents an initial framework of the general modeling or forward problem and its counter part the estimation or inverse problem. The focus is particularly oriented towards the set of techniques used during the last years to solve the TSE problem. The TSE problem literature is deeply reviewed in the highway case by Seo in [Seo+17]. On the other hand, for the case of TSP problem an initial review was presented in 2004 [VGK04], nonetheless the work has been recently updated in [VKG14]. The following sections retrieve the most relevant works related to the problems tackled in this dissertation and position this work with respect to the state of the art.

2.2 The modeling and estimation problems

A Distributed Parameter System (DPS) belongs to certain class of systems that may describe traffic behavior. For this class of systems two main problems can be envisaged, the modeling problem (a.k.a forward problem) and the inverse problem (a.k.a estimation problem). A graphical representation is provided in Fig. 2.1.

2.2.1 Modeling problem

The forward problem consists on determining an abstraction of the physical world typically contained in mathematical equations. This abstraction defines the initial condition of the system, its boundary conditions and parameters required to describe the system. The solution on time of the proposed abstraction illustrates the behavior of the system. This problem is motivated and introduced in the case of traffic systems in [TK13], where physical behavior is explained in highway infrastructures and [GP06] where a good variety of conservation laws models are summarized for traffic. The extension to vehicular traffic networks, in particular urban cases, is detailed in [GHP16] where in addition to conservation laws the problem of dynamic traffic assignment and control and traffic junctions is developed.

2.2.1.1 Traffic flow models

Traffic flow model theory is the one in charge of describing behavior of traffic systems. Along history traffic theory has determined three levels of abstraction of traffic dynamics. We present them in order of aggregation:

- *Microscopic*: The basic models are the car-following ones or models based on Newton's law. The advancement of sensing technology, data collected through GPS, mobile sen-

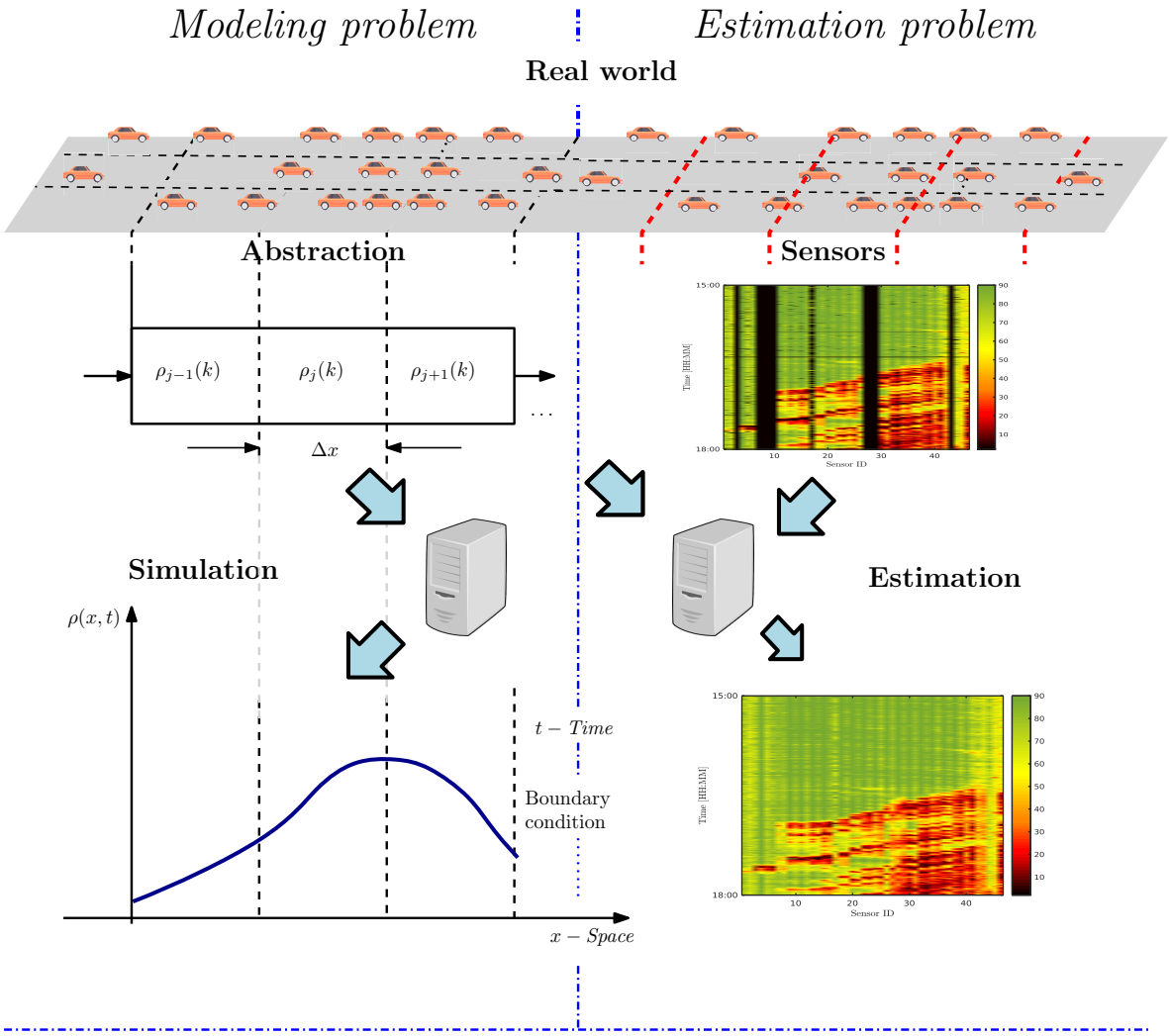


Figure 2.1: Forward and Inverse problem representation. Left side of the figure represents the modeling problem in which a mathematical abstraction represents the behavior of the system. The right side represents the complementary problem of estimation. Data retrieved from sensors and a model are used to compute the state of the system

sors have motivated the development of these models. In this formulation the state of the system considers information of individual vehicles within the network. The main assumption of the car-following models is that an individual car's motion only depends on the car ahead information, for more details see [BM99], [New93], [CHM58], [Edi61]. Given the influence of the driver in traffic, car-following models have been lately oriented towards a better comprehension of the human reactions. [TYG16] studies the influence of aggressive behaviors, and [MS16] analyzes the effect of these behaviors in the throughput of the network. Emotions can be included as part of the model, as it is the case of [HA14a], [Ro+17] where emotional behavior has been added to the dynamics of single vehicles. These and other behaviors such as lane exchange can be added at the cost of complexity when increasing the size of the network under study.

- *Mesoscopic*: These models are aimed to combine microscopic and macroscopic approaches in a hybrid model by integrating heterogeneous vehicle characteristics and moving bottlenecks. They do not describe detailed behaviors of vehicles as particles, instead they provide a complete description to distinguish changes of the traffic dynamics within a low computational cost while keeping an individual tracking of vehicles characterized in an event based form [LB11]. A second class within this category are Kinetic models. They use Boltzmann-like equations where main traffic description are given by density distributions. A function $f(x, v, t)$ expresses the probability of having a vehicle at time t at position x runninng at velocity v. Details about these models are explained in [MP69], [PA60], [ID16].

- *Macroscopic*: The next level of aggregation corresponds to macroscopic traffic models. Developed independently by [LW55] and [Ric56], they describe traffic equations from the point of view of fluid dynamics. In this case, the use of a Partial Differential Equation (PDE) represents a DPS in which the state describes aggregation of particle states which leads to a more global description. In a single road, this nonlinear model is based on the conservation of cars described by the scalar hyperbolic conservation law

$$\partial_t \rho + \partial_x \Phi(\rho) = 0, \tag{2.1}$$

  where $\rho = \rho(t, x) \in [0, \rho^{\max}]$ represents the density of cars and $\rho^{\max} > 0$ is the maximum density of the road. The function $\Phi(\rho)$ is the flow of cars which is expressed in terms of the density and the local speed v, written $\Psi(\rho)$ as a function of the density. Solutions to fluid-dynamic models for traffic flow seem to be the most appropriate to detect macroscopic phenomena as shocks formation and propagation of waves backwards along roads. However, they can develop discontinuities in a finite time even starting from smooth initial boundary conditions.

  Second order models have been also proposed in traffic systems. The first one was proposed in the 1970s independently in [WF75] and [Pay71]. It was the first model to couple velocity dynamics as a second equation, and it is referred to as the Payne-Whitham (PW) model. In the LWR scalar equation model, a particular form of v was assumed where velocity is a function of density, but in high order models, v and ρ are assumed to be independent and a second equation is formed to link them. The equations

of the PW model in nonconservative form is then given by

$$\begin{aligned} \partial_t \rho + \partial_x(\rho v) &= 0 \\ \partial_t v + v\partial_x v &= \frac{\Psi(\rho) - v}{\tau} - \frac{C_0^2}{\rho}\partial\rho_x, \end{aligned} \tag{2.2}$$

where τ is the relaxation parameter and C_0^2 is parameter related to driver anticipation. The Aw-Rascle-Zheng (ARZ) model proposed in [AR00] and improved in [Ras02] is argued to be an improvement on the PW model. Authors claim that the drawback in the PW model (letting information travel faster than the flow [Dag95a]) can be corrected by adding information on how the average driver behaves within the supplementary speed equation (microscopic information). In this case, the ARZ model is expressed as

$$\begin{aligned} \partial_t \rho + \partial_x(\rho v) &= 0 \\ \partial_t(v - \Psi(\rho)) + v\partial_x(v - \Psi(\rho)) &= -\frac{\Psi(\rho) - v}{\tau}. \end{aligned} \tag{2.3}$$

The ARZ model converges to the LWR model at $\tau \to 0$; therefore, it can be regarded as a generalization of the LWR model [Seo+17]. We refer to the corresponding references for more details on these models. In the approach delivered in this thesis we consider the conservation equation which establishes the foundations of traffic flow theory. Speed dynamics required to exploit the capabilities of second order models are achieved only with GPS at high penetration rates and high sampling rates. Most of the dynamics are smoothed when FCD is examined due to the aggregation time and penetration rate of data providers.

Simplifications on numerical schemes of these family of models have been found in [GR96] and [LR02]. Among all models the most celebrated one is the PDE based Lighthill-Whitham and Richards (LWR) model [LW55] based on fluid kinematics, is able to reproduce crucial phenomena such as traffic shock waves. Discretization of the LWR-PDE is not straightforward but stable numerical schemes have been proposed, the most well known being the Cell Transmission Model (CTM) [Dag94]; [Dag95b]. The CTM relies on linear piecewise fundamental diagrams shown to be consistent with the Godunov discretization scheme for scalar conservations laws. In its former formulation the problem was presented only on single roads, extensions to more complex networks are presented in [GHP16]. Numerical approximations of the junction problem are presented in [BNP06] while formal solutions of the problem are found [CGP05].

Other variables apart from density have been also used in macroscopic models. [Wor+10] proposed a macroscopic model where the state is represented by the speed, one of the requirements is the invertibility of the speed-density map which is not the case for the linear case. We try to tackle this issue by separating congested/free-flow zones via FCD. This dissertation takes advantages of the LWR model formulation at a network level to propose an estimation approach which based on the available measurements and the RTN model. In Chapter 6 we introduce details on this formulation of the RTN model used.

Model	Representation
White noise	ARIMA$(0,0,0)$
Random walk	ARIMA$(0,1,0)$
Autoregressive	ARIMA$(p,0,0)$
Moving average	ARIMA$(0,0,q)$

Table 2.1: Representation of multiple models through ARIMA models

2.2.1.2 Data driven models

The use of historical information takes advantage of repetitive patterns in traffic. Patterns found in data allow exploiting particular situations in which traffic models require more development [VKG14]. However, the exclusive use of this information does not allow the algorithms to react to current conditions. The use of same day data, on the other hand, overcome this problem, but at the cost of being limited for shorter evolution conditions. When a mixture of the information is used, the algorithms can capture current dynamics as well as enlarge their forecasting horizons of evolution of models.

- *Autoregressive Integrated Moving Average (ARIMA)*: In an autoregression model, we forecast the variable of interest using a linear combination of past values of the variable. In a moving average model rather than using past values of the forecast variable in a regression, only past forecast errors are used in a regression-like model. In this model case the objective is to combine both approaches. The description is given by

$$\begin{array}{cccccc} \phi_p(B) & (1-B)^d & y(k)=c & \theta_q(B) & e(k) & \\ \uparrow & \uparrow & & \uparrow & & \\ AR(p) & I(d) & & MA(q), & & \end{array} \tag{2.4}$$

  where $B^j y(k) = y(k-j)$, $\phi_p(B) = (1-\phi_1 B - \cdots - \phi_p B^p)$ denotes the autoregressive polynomial of order p, $\theta_q(B) = (1+\theta_1 B + \cdots + \theta_q B^q)$ refers to the moving average polynomial of order q, and $(1-B)^d$ denotes the order of differentiation involved. $y(k)$ correspond to the value of the variable at the current time instant and the term $e(k)$ intends to model the error. The model can express general families of models ARIMA(p,d,q). Many of the models we discuss in this dissertation are special cases of the ARIMA model, as shown in Table 2.1.

  The generic representation through ARIMA allows a complete representation of the main characteristics of y, the models can be estimated using Box-Jenskins method [Box+15]. The good degree of freedom in p, q, d provides flexibility to fit data parameters. High order models may overfit and add complexity, at same time a rich data set is required for calibration. Nevertheless, when dealing with good calibrated models the approach can accurately predict the behavior with relative small errors [HAMS95].

- *Seasonal Autoregressive Integrated Moving Average (SARIMA).* In this case a seasonality parameter is added to the ARIMA model in order to capture regularity patterns reflected

in cases of traffic in a daily/weekly basis. In the traffic case, the notion that seasonality could be achieved by a weekly seasonal difference was first recognized by [WH03]. This model requires strong repetitive pattern in order in order to have a good performance [KV15]. Its application is particularly oriented in the TSP problem in particular for relative long term forecasts (days). The description is given by

$$\Psi_P(B^S)\phi_p(B)(1-B)^d(1-B^S)^D y(k) = c\theta_q(B)\Theta_Q(B^S)e(k), \tag{2.5}$$

where $\Psi_P(B^S) = 1 - \Psi_1 B^S - \cdots - \Psi_P B^S$ is the seasonal autoregressive polynomial of order P, $\Theta_Q(B^S) = 1 + \Theta_1 B^S + \cdots + \Theta_Q B^{QS}$ is the seasonal moving average polynomial of order Q, $(1-B^S)^D$ the polynomial of seasonal differentiation. The terms $\phi_p(B) =$ represents the non-seasonal auto regressive polynomial of order p. The terms $\phi_p(B)$,$\theta_q(B)$ correspond to the non-seasonal autoregressive polynomial and non-seasonal moving average polynomial from ARIMA model.

We motivate the use of data driven models for the TSP problem due to:

- Adaptation to the real time traffic conditions in terms of same day information. This advantage makes the mechanisms more robust, therefore allowing a correct evolution even when non regular conditions during the day are present in the measurements.
- Model structure and its components include common behaviors in time series (trend, seasonal variation, cycle, etc.). The effects of observed variables and explanatory variables are modeled separately before being integrated into the final model.
- Low computational complexity suitable for real-time implementation. Given the linearity of the model, this feature is desirable in order to perform computation of multiple traffic variables in relatively short times.
- It is straightforward to predict future values into the future, or estimate values in current time given the causality of the model.

We provide more details on the model used in the Chapter 5 in which we make use of these family of models in order to devise predictors for Travel Time.

2.2.2 Estimation problem

The modeling problem presents two fundamental limitations. First, the abstraction is a representation of the real world which is regularly simplified given the existent tools. These simplifications conduct to inherent errors which are not necessary measurable and at same time they are hard to determine. Second, the information contained inside the abstraction such as initial condition, boundary condition are regularly known by approximation therefore adding uncertainty to the model.

In order to fill the gap and address this limitations a related problem can be studied. The estimation problem is a complementary problem to the former one, in this case additional

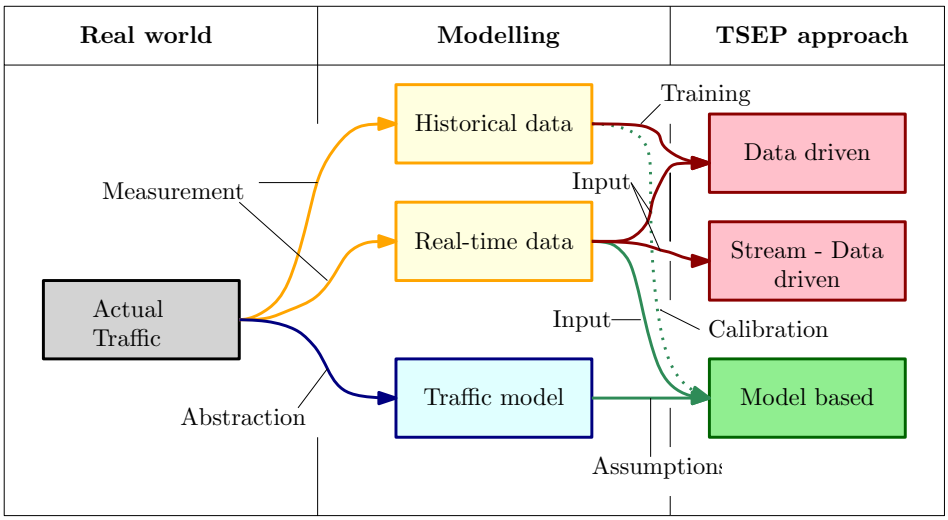


Figure 2.2: Framework for TSE & TSP problems[Seo+17]

information is added from the physical world. This information is taken from the real world represented as observations in form of data measured by sensors. The process in which these observations are integrated with an abstraction is known as estimation. The objective of the estimation is to reconstruct the state of the system [Can16].

The framework of TSE and TSP can be classified into three main components (See Fig 2.2 [Seo+17]). The first component corresponds to the reality in which two main sources of information are available. Abstractions as explained in section 2.2.1.1 and measurements which can feed data-driven mathematical representations of type 2.2.1.2. A second component is a classification of family of models. These can be based on historical information or realtime information for the data-driven cases. Finally a third component categorizes the families of algorithms or approaches based on this information.

The *model-driven* approach relies on information from a physical model describing the traffic dynamic, i.e. traffic flow model. They estimate traffic states in unobserved areas by means of the model and real-time data is used as supplementary information to estimate actual state of the traffic network. Model parameters can be calibrated in an offline process by using historical data or included within the problem as a parameter estimation problem. The data-driven approach relies on historical-data inferred knowledge about the system. These learning process is achieved through statistical techniques or machine learning methods. The estimation task can then be achieved integrating real-time data within the algorithms. This means that it does not require a priori knowledge of traffic being modeled explicitly as in physical traffic flow models. The approach generally requires large amount of historical-data. Finally, the *streaming-data-driven* uses real-time (i.e., streaming) data only and is not strongly characterized by empirical relation which can be found in common traffic flow models (e.g., the fundamental diagram). It means that the approach less relies on a priori knowledge on traffic network [Che+03].

In this dissertation we make use of the three families of models. First, streaming-data-

driven models are used in the imputation task required to guarantee data completion as it is detailed in chapter 3. Second, data-driven models are used for DTT short-term forecasting problem explained in chapter 5. Finally, we make use of model-driven algorithms in order to reconstruct flow and density in urban type traffic networks as detailed in chapter 6.

Before summarizing the main contributions to the TSE, TSP problems in the recent years, we briefly describe the sensor technologies available to capture measurements of traffic systems.

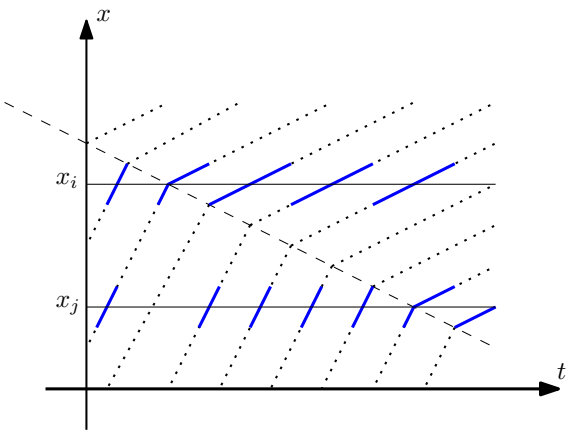
2.3 Measurement systems

Measurement systems in traffic systems have been developed since the beginning of traffic flow theory (See. Fig. 2.4). Along with technology development the detail on the measurement and the quality of variables collected has been improved. Data types are grouped into two categories according to the measurement methodology: stationary and mobile (See Fig. 2.3). Stationary data (See. Fig. 2.3a, 2.3b) is collected by fixed sensors, such as inductive loops or other magnetic based technologies [Baj+11], and can be considered as conventional technology. This stationary data makes uses of the network infrastructure, it is regularly maintained by state agencies and the cost of expansion of this technology is very high.

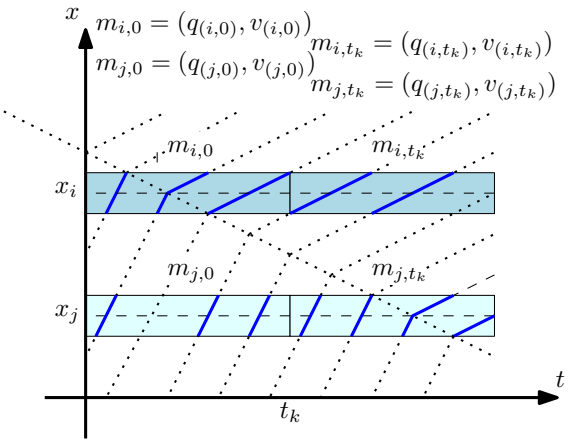
In the recent years, the development of mobile applications using GPS trace information has become a new way to collect data (See. Fig. 2.3c, 2.3d). The emergence of Vehicle to Vehicle (V2V), Vehicle to Infrastructure (V2I) technologies, connected and automated vehicles makes of this methodology a promising source for the development of ITS. Some of the difficulties on collection of this data are privacy issues on the driver side and low penetration rate which is regularly low due to the technology adoption and the marketshare of multiple companies. As a consequence, FCD introduce another category of this data in which spatio-temporal aggregation is the main difference. In aggregated data as is the case of FCD, information from multiple vehicles is aggregated (e.g., 1 min time interval) and stored. The benefit is a solid and more spatially dense source of information which can be integrated to TSE problems.

Fig. 2.3 displays a classification of different technologies. Figs. 2.3a and 2.3b summarize two ways to capture stationary data, meaning data retrieved at fixed positions (a.k.a Eulerian data). In Fig. 2.3a loops are placed in positions x_i and x_j. Each time a vehicle crosses over a sensor (blue lines) an event is registered. In practice, all events during a certain amount of time are collected and then transmitted to the network in charge of data collection. This measurement mechanism produces Aggregated Indicator Data (AID) which is one important source of data in order to calibrate mesoscopic models [LB11] [DY17]. The amount of events registered varies depending on the capacity of the RTN. Hence, storage and good performance of the data network infrastructure is required particularly during peak periods of the day.

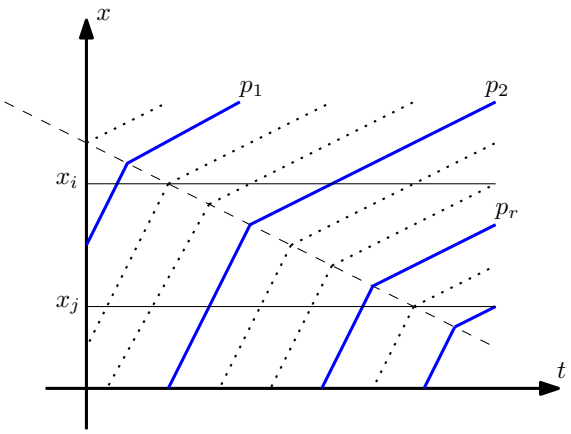
Loop event information suffers from problems like false detection, aliasing, unregistered events, or lost packages due to high number of events during peak times. An alternative procedure is to aggregate event information from stationary detectors in time windows as



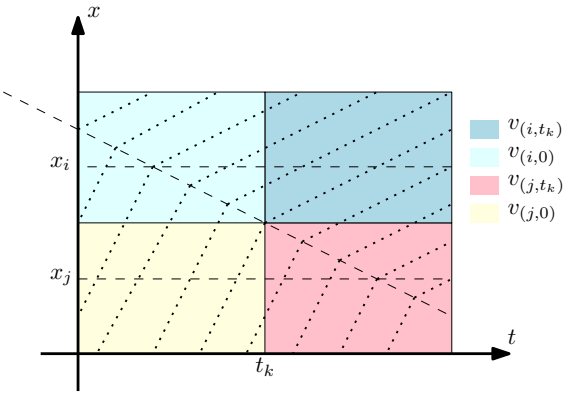
(a) Stationary: Individual MLD data



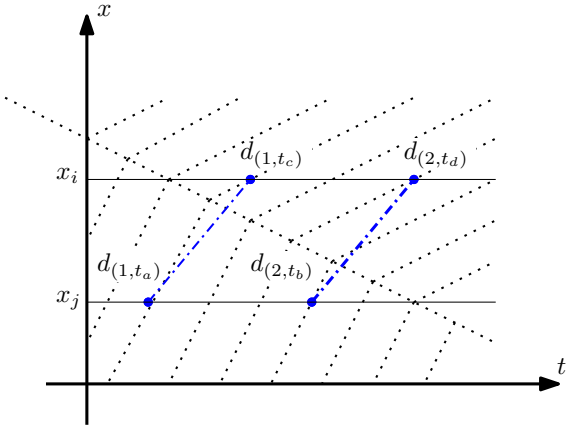
(b) Stationary: Aggregated MLD data



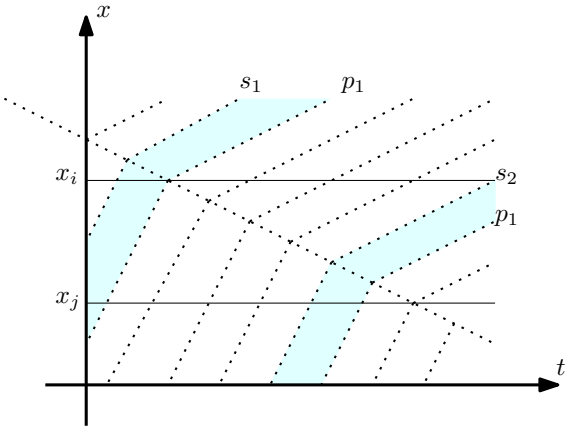
(c) Mobile: Individual GPS data



(d) Mobile: Aggregated GPS data : FCD



(e) Hybrid: Individual Travel Time (TT) data



(f) Mobile: Individual Spacing data

Figure 2.3: Measurement technologies in traffic systems

depicted in Fig 2.3b. Here m_{i,t_k} defines the aggregated information of the sensor aggregated at certain periods of time (See details in Section 3.3.1). Variables can be aggregated according to the Edie's principles [Edi63]. (See details in chapter 3). Loop data is typically aggregated in 1 to 5 min time slots. Variables collected in this sensors are aggregated average speeds v, vehicle counts q and occupancies in some cases [CDW+15].

An alternative to stationary sources of data are mobile sources of data (a.k.a Lagrangian) as depicted in Figs. 2.3c and 2.3d. In this case, individual traces collecting the position of some vehicles (also denominated probe vehicles) are shown in Fig. 2.3c. This kind of measurement regularly collects position, time and speed corresponding to the probe vehicle under study. In order to perform estimation based on these measurements, the amount of probe vehicles equipped with GPS should be sufficiently high in order to capture the spatio-temporal characteristic of the system [Her+10]. Regularly GPS is a source of data nowadays in high adoption, it is integrated in smartphones or directly incorporated in vehicles. Actual technology allows precisions in the order of less of 10m, nevertheless penetration rate is not regularly uniform in the network leading to very few data collection in some spaces in the network. In addition privacy issues due to laws constraints regulate access to this information. Aggregation (See. Fig. 2.3d) in time and space is also the one of the alternatives to protect drivers' identity and to provide at same time a good representation of the traffic state. The aggregation process is explained later in this dissertation.

Fig. 2.3e displays the principle of bluetooth detection. It is a growing technology during the last years [SKA15]. The travel time is determined by re-identifying bluetooth devices in the probe vehicles such as cell phones between multiple sites. A single bluetooth reader mounted on the side of the road can be used to determine the travel time of a vehicle. In the case of the Fig. 2.3e the measurement $d(1, t_a)$ represents a vehicle label detected at time t_a at position x_i, then after some time the same label is detected at time t_b indicating the DTT $t_a - t_b$ of the vehicle. Finally, the arrival of connected vehicles/autonomous vehicles has augmented to the capability of measurement leading to the appearance of new technologies (See Fig. 2.3f). One example is headspace defined as the inter-vehicle distance. This information can be represented either in time or referred to a particular speed. These kind of measurements are regularly used in probe vehicles to monitor statistical conditions on infrastructures (e.g. [HB10]) but they will acquire more importance with the development of autonomous and connected vehicles.

Supplementary systems to monitor traffic like video cameras make also part of measurement systems. The extracted information represents a direct measurement of the state for macroscopic traffic models [Urr+09]. A video-processing stage is always required before the obtention of traffic variables which makes this kind of systems difficult to implement for real-time applications.

(a) Measuring scheme (b) Measuring scheme

Figure 2.4: Characterization of traffic systems [Gre35]

2.4 Traffic state estimation overview

Fig. 2.4b illustrates the experimental setup used in [Gre35] to characterize traffic flow. First measurements in traffic systems where registered based on stationary observations. This was the first methodology to solve the TSE problem. Several issues may lead to an estimation task from a traffic application, one may be interested in reconstructing partial information of missing samples namely data imputation problem. This problem is regularly a middle stage in the TSE problem. Other possibility is to do state reconstruction for macroscopic models. This is equivalent to perform density reconstruction, a quantity not directly measured with most of current technologies apart from video. Another measurement of interest is the Travel Time (TT) which can be considered as a state of the traffic system [HHB12].

2.4.1 Data imputation

Literature about this topic can be classified among several techniques. Tensor methods offer extreme good performance for offline applications since the fitting problem is adapted to a multi dimensional array of data which can be arranged with a particular predefined structure. However, for large dimensions its computational cost augments due to decomposition techniques often used to implement the imputation algorithm [Tan+13],[Ran+15], [GKF17] . Distributed strategies have been proposed to cope with large scale decomposition and distribution is possible at the cost of an agreement among smaller subproblems [DAK14].

Bayesian methods are good for online applications, the principle assumes a high correlation among close spatio - temporal neighbors which model the *a-posteri* distributions. As a consequence a first stage requires estimation of statistics of the a priori usually via parameter estimation which is required in order to fit distributions on the *a-priori* distribution [LLL13]. This method offer slight better performance than regular averaging methods, nevertheless the

implementation is based on a recursive algorithms which do not offer explicit solutions and moreover, statistics about *a-priori* information should be computed in advance.

Interpolation methods exploit also the correlation between spatial neighbors, even when the performance of these methods is highly dependent on the amount of missing data, the results given by [Che+03], [Boy11] show promising results in terms of efficacy and computational cost. In [Che+03] a stream-data-driven technique is devised in order to produce estimates of speeds based on neighborhood data, the main objective is the problem is primary speed reconstruction. In a model-driven approach [HSL02] an LWR equation is used to reconstruct missing information.

2.4.2 Density reconstruction

Density is a quantity that can be measured only via camera systems. Regularly, it is more desirable to estimate this quantity based on a traffic model. The problem of traffic estimation begins in the 1970's with early work from [SG72] when a Kalman filter strategy was applied to reconstruct density in a highway. At the beginning of this century, [Muñ+03] made use of a Switching Mode Model (SMM) and embedded it into Kalman Filter (KF) mixture filter to reconstruct density based on loop data. In an alternative approach [Coi03] produced density estimation based on vehicle re-identification in double-loops at different locations. [Mor+14] proposed a robust mode selector and an observer for the constrained-graph SMM developed in [WOK12].

More recent [SW17] presented a scalable filtering algorithm for multi-agent traffic estimation. The task is achieved by modeling the traffic network in the framework of a SMM and performing spatially distributed KF with overlapping regions. A consensus algorithm is used in the boundaries to guarantee the agreement between the spatially distributed filters. The aforementioned approaches are based on boundary measurements in order to perform the reconstruction.

In [SKA15] headspace between vehicles is used in probe vehicles to reconstruct flow and density. The approach considers a stochastic model for the head space in order to deal with the penetration rate issue. In [DLZ13] multiple sources of data are integrated into the three detector model. In more recent works, [LCK15], [LCK16], authors proposed to combine both FCD and MLD for flow and density estimation. [WH16] combines GPS probe data and MLD data within a framework of non-linear filtering, with a modified CTM to formulate the state estimation problem. The study in [BLRP16] considers sources of connected vehicles and not connected ones for the density reconstruction. A different approach [CC16] formulated the problem for lagrangian and loop data integration into a model from variational calculus perspective. [HB10] incorporates GPS information in the conservation equation in order to correct errors introduced by the aggregation effect.

[Viv+15] contributes to the topic proposing a distributed switched observer design method, based on Luenberger-like and consensus ideas. To this aim, a switching mode model SMM is employed, representing the freeway system switching among different sets of linear differ-

ence equations. In the proposed approach, each local observer reconstructs the densities of the overall system with very limited information exchange requirements. [WFW16] proposes efficient multiple model particle filter (EMMPF) to solve the problems of traffic state estimation and incident detection, which requires significantly less computation time compared to existing multiple model nonlinear filters. To incorporate the on ramps and off ramps on the highway, junction solvers for a traffic flow model with incident dynamics are developed. The effectiveness of the proposed EMMPF is assessed using a benchmark hybrid state estimation problem, and using synthetic traffic data generated by a micro-simulation software.

Estimation schemes have been also proposed for second order models, works in [WPM06], [Wan+09], [Mih+12] are some examples where Luenberger observers, Kalman Filter (KF) and non linear filtering techniques are applied to reconstruct densities and speeds. In particular, [Mih+12] uses parallelized particle filters and complex probability density functions. In this approach the high-dimensional traffic state can be decomposed into functions with simpler forms and the whole estimation problem is solved in an efficient way. [Bla+12] explores Unscented-KF and Ensemble-KF applied to second order models, the results although require high amount of computations for single roads estimation in order to be efficient.

2.4.3 Travel time estimation

Travel time can be defined as the period of time to go through a route between any two points of interest. It is a fundamental measurement in transportation. Travel time is also one of the most readily understood and communicated measured indices by a wide variety of users, including transportation engineers, planners and consumers, yet it is rigorous enough for technical analysis.

Travel time can be directly measured via probe vehicles, it basically consists of a vehicle that is specifically dispatched to drive within the traffic stream for the express purpose of data collection. The technology to collect data can be GPS, or license plate matching. When using license plate matching video-recognition techniques can be set up to accurately detect vehicles [Sum+12]. Direct measurements of TT are regularly precise but as a counterpart they can be only applied to very low amount of vehicles within the traffic network.

As an alternative to direct measurements of TT, indirect methods to estimate TT can be used. Dual-loop detectors is one of the more precise techniques. These are placed on a freeway a fixed distance apart, approximately 5 meters, producing more accurate speed measurements between the loop detectors. A widely used method to estimate travel time from dual loop detector measurements is using piece-wise constant speeds between detectors. [LZ03] develops a method to improve this estimation on the basis of a piecewise-linear of function speed which corrects errors introduced by a piecewise-constant approximation. [Che04] improved the method by interpolating the average speed as a polynomial functions of both distance and time while iteratively calculating the actual link travel times.

Double-loop detectors are not regularly found in traffic infrastructures. Instead, single loop detectors are more common. For these cases, only flow rate and occupancy over fixed

time intervals are measured. Two main approaches can be considered to perform travel time estimation. TT can be estimated directly by inferring information from flows and occupancies directly as in [Ath65] or by inferring speed and performing then the computation.

More recently integration of GPS data into traffic equations was used to estimate travel time [Woo+17]. At a network level [Yua+14] proposed KF filter technique for a variational approach of the LWR model in order to reconstruct TT. Other approaches exploit capabilities of machine learning methods like [HHB12]. [Hun+13] considers regularization techniques to compute travel times along paths in urban networks.

The first shortcoming is that most of the methods for travel time estimation are developed based on data generated by dual-loop detectors with large time aggregations (e.g. 5min) which contain averaged information about speeds and smooth stop and go waves perceived by systems. Second, the majority of existing studies focus on travel time estimation instead of travel time prediction when using traffic measurements from single loop detectors. In this dissertation we encourage computation of DTT in short-time intervals and in a predictive way.

2.5 Traffic state prediction overview

The short-term forecast problem is deeply studied in [VKG14]. We summarize some of the main challenges presented in that review and we highlight important references for the current research. The following are the main subset of challenges taken from [VGK04] and worked in this dissertation.

- *Responsive algorithms*: Transportation agencies require forecasts that are robust to short and longer term changes in traffic conditions. In cases where these changes are unexpected accidents or adverse weather conditions which can be somehow measured, traffic management systems should optimize management and advisory strategies.
- *Short-term predictions from volume to travel time*: Over the past 10 years, travel time prediction has attracted interest because of its importance as a network performance measurement and its straightforward use to inform road users about traffic conditions.
- *Data resolution, aggregation and quality*: The selection of the suitable forecasting interval (step) is critical and relates to the type of ITS application to which the algorithms are integrated into.

In this dissertation we focus on models that rely on historical travel time information. As shown in [Chr+04], typical real patterns can be found in large data sets of historical data. [Lin04] make uses of neural networks which describes data dependancies in order to forecast travel-time. Later, [LZT08] studies skew distributions in TT day distribution and it argues that it must be considered an important contributing factor to travel time unreliability. For example [OKC13b] proposes a forecasting of travel time based on an Adaptive Kalman Filter

(AKF) strategy in which observations are built from historical data sets of speed and flow. [OKC13a] applies a similar strategy in flow prediction and presents the problem of multi-step ahead forecasting based on clustered time series by applying several predictors such as Gaussian maximum likelihood (GML) and AKF. [WSH14] presents an approach for short term flow forecasting using multiple ARMAX based predictors obtained from clustered data. The ARMAX model is adapted independently to different groups of flow time series and a single prediction is selected based on one criteria that considers minimum error estimation for the predicted signals. [WGH14] uses Link Node Cell Transmission Model calibrated via Monte Carlo methods in order to generate a prediction using the expectation maximization algorithm. Although all these methods present strategies for the selection of the predicted sequence, regimes described by clustered data are not totally separable and studies like [ZL11] have shown the improvement of performance with combined forecasts.

Different models can be inferred from different clusters of data. A series of new methods have been emerging to combine information from these models. For instance [Zha11] proposes a forecast based on interactive multiple models by combining different individual forecasting methods. [DPK12] proposes an adaptive fusion method combining historical information and current day data. Other hybrid forecasts includes support vector machine methods like in [Hon+11] and model based approaches as in [Cha+12].

Generally, forecast algorithms in the literature are designed to satisfy a set of constraints given by the forecast problem. Most of the algorithms take into account availability of a full set of measurements for all possible locations and time instants within the traffic network and they overcome the problems of missing data, low penetration ratio or unbalanced spatial coverage by introducing additional steps such as imputation algorithms [HC16]. Moreover, the great majority of literature provides fixed forecasting scenarios in which predefined OD schemes are considered [MJW13],[Chr+04]. Some recent approaches have emerged considering flexible OD, moreover the possibility to exploit internal relations of the network to reconstruct information [DPK12]. In this dissertation we formulate the problem by considering prediction over single OD pair of the highway and replicating the same methodology for other OD combinations, this due to the properties of DTT. In addition, since measurements for all the OD pairs cannot be available all the time, we face a missing data problem. To overcome this issue, we resort to a data imputation based on interpolation.

2.6 Final comments on the chapter

In this chapter we have presented a TSE & TSP problems review. We have introduced a general framework to study the modeling and the estimation problem. Then we have presented the main classification of techniques and measurement technologies existing nowadays to measure ITS. Finally we have made a review of the most relevant works used in this dissertation by classifying them into estimation and prediction frameworks.

The remaining chapters will develop the main content of this dissertation. Data retrieval and processing are presented in chapters 3 and 4. The problem of short-term forecast is

presented and deployed in a realtime application explained in chapters 5 and 7. We study in addition the TSE problem, it is formulated in the urban traffic network case in chapter 6.

Chapter 3

Traffic data collection and travel time computation

Contents

3.1 Overview

During the last years some considerable increments in traffic volumes have been observed and measured thanks to new emerging technologies in the area of ITS. Population and economic growth alongside continued urbanization are the root causes of congestion. While cities develop strategies to increase their sizes and development of services these do not match by a comparable extension to the development in road infrastructures. As main consequence, important corridors have been steered to states of congestion on a daily basis period and increased travel times are perceived by regular drivers within the traffic network.

While most of traffic service providers have been interested in qualifying traffic indexes within cities, travel time is one of the most readily and understood traffic index by a wide variety of users among them transportation engineers, planners and regular drivers. Recent

press releases such as [INR16] are an evidence of this fact. As an example in France, drivers waste up to 70 h/year within congestions and Grenoble in particular occupies the 8^{th} position with an average of 33 h/year.

The accuracy on the computation of Travel Time (TT) is an important task and research during last years has been focused on more precise estimates [LZ03], [YG13], [Oje14], [DG16]. The importance of this indicator has led to the development of TT measurement systems based on multiple sources of information. Despite the existence of multiple traffic monitoring tools, studies formalizing the impact of these estimates tend to put aside the impact of the quality of the data and intermediate computations.

This chapter discusses computation of TT in particular the basic principles for DTT estimation and explains how this computation is achieved from MLD data which is the main input source of data. In particular, we provide details in the computation of DTT which is the same index used in the GTL explained in details in Chapter 7. We highlight the differences between the DTT and the Instantaneous Travel Time (ITT) and errors incurred from an experimental point of view.

First, we describe the existing relationships among macroscopic and microscopic traffic variables. Second, we consider the process of data collection and data aggregation which makes part of the solution implemented in GTL, then we deal with cleaning strategies for the measured variables used in the reconstruction of DTT. Finally, we provide illustrative comparisons between the two approaches of TT.

3.2 Macroscopic-microscopic relationships

In practice two type of measurements can be distinguished among the classical systems for traffic measurements. The first comes out from measurements directly taken from sensors onboard within a vehicle. The key idea behind using probe vehicles for travel time estimation is that a probe vehicle traveling in traffic should be a reasonable representation of the characteristics of the traffic. A sufficiently large number of probe vehicles should be representative of the traffic conditions experienced.

Let consider a *link*[1] defined as a segment of road without any internal merge or diverge allowing traffic on a particular direction. At *microscopic* scale vehicles are represented by trajectories $x_i(t)$, defining the position x with respect to the origin of the section for the i^{th} car at a particular time t. Similarly we denote by $v_i(x,t)$ the speed of the i-th vehicle at time t and location x. For the sake of simplicity lateral dynamics are not considered. At a *macroscopic* level, traffic over the link is represented by *flow* φ, *density* ρ and *average speed* v [TK13]. Positions at the microscopic level can be tied to macroscopic behavior and the relationships were given by [Edi63].

First let us consider the graphical representation given by Fig. 3.1. Within the figure

[1] a.k.a section

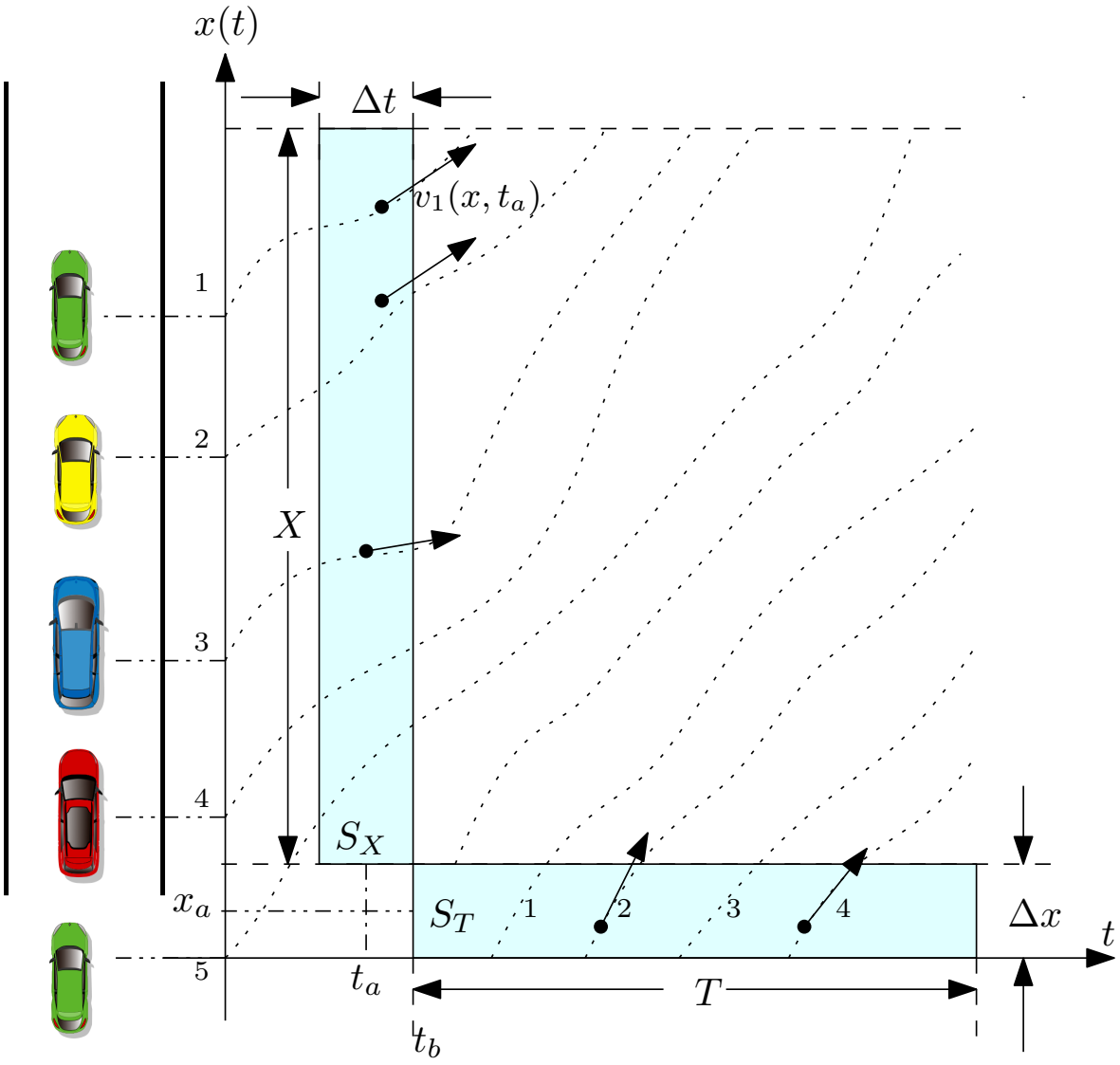


Figure 3.1: Relationship between microscopic and microscopic variables

two traffic intervals are depicted, they represent measurements in two surfaces. The first one correspond to a space interval X characterized by measurements taken within the surface S_X, the second one corresponds to a time interval T which corresponds to measurements captured within the surface S_T. The data collection in the space interval covers the road segment of length X and an infinitesimal time interval Δt. During that time interval we assume n_X (e.g. $n_X = 5$ in Fig. 3.1) the total number of vehicles in the space interval. The data collection in the time interval consists on measurements collected during a specific time T in a small piece of road Δx, for this case we consider n_T as the number of vehicles crossing a small space region at a fixed location e.g double loop detectors separated Δx during the period of time T. We define then the macroscopic variables.

3.2.1 Density

Density ρ is defined as the number of vehicles per distance unit at a particular time. t_a (See. Fig. 3.1). For the measurement interval S_X density over the road segment can be calculated as

$$\rho_{S_X}(t_a) = \frac{n_X(t_a)}{X}. \tag{3.1}$$

This expression can be alternatively represented as

$$\rho_{S_X}(t_a) = \frac{n_X(t_a)}{X} = \frac{n_X(t_a)\Delta t}{X\Delta t} = \frac{TTT_{n_X}(t_a)}{S_X},$$

where the numerator corresponds to the Total Travel Time spent by all vehicles $n_X(t_a)$ divided by the area of interest S_X . A similar definition can be extended to the measurement interval S_T. Considering the definition of the individual Travel Time $tt_i(t) = \Delta x / v_i(x_a, t)$, $t \in [t_b, t_b + \Delta T]$ over the surface S_T

$$\rho_{S_T}(x_a) = \frac{\sum_{i=1}^{n_T(x_a)} \frac{\Delta x}{v_i(x_a,t)}}{T\Delta x} = \frac{\sum_{i=1}^{n_T(x_a)} \frac{1}{v_i(x_a,t)}}{T}. \tag{3.2}$$

3.2.2 Flow

We proceed in analogous way for the flow. Traffic flow is defined as the number of vehicles crossing a particular section per time unit. For a time interval T at a specific location x_a where n_T vehicles cross a section of length Δx, the flow can be defined as

$$\varphi_{S_T}(x_a) = \frac{n_T(x_a)}{T}. \tag{3.3}$$

Which can be expressed as

$$\varphi_{S_T}(x_a) = \frac{n_T(x_a)\Delta x}{T\Delta x} = \frac{TTD_{n_T}(x_a)}{S_T},$$

where the numerator represents in this case the Total Travel Distance travelled by all vehicles in S_T divided by the measurement area. The previous expression may also be expressed in

the location measurement interval S_X. By doing an analogous treatment as in 3.2 the flow can be written as

$$\varphi_{S_X}(t_a) = \frac{\sum_{i=1}^{n_X(t_a)} v_i(x,t_a)\Delta t}{X\Delta t} = \frac{\sum_{i=1}^{n_X(t_a)} v_i(x,t_a)}{X}. \tag{3.4}$$

3.2.3 Speed

Speed, namely v is the macroscopic variable that describes the fundamental relation of traffic flow [LW55]; [Ric56]

$$\varphi = v\rho, \tag{3.5}$$

thus, the speed can be defined as the quotient between the flow and the density. In the case of measurement interval S_X

$$v_{S_X}(t_a) = \frac{\varphi_{S_X}(t_a)}{\rho_{S_X}(t_a)} = \frac{\sum_{i=1}^{n_X(t_a)} v_i(x,t_a)}{n_X(t_a)}, \tag{3.6}$$

and in the case of S_T

$$v_{S_T}(x_a) = \frac{\varphi_{S_T}(x_a)}{\rho_{S_T}(x_a)} = \frac{n_T}{\sum_{i=1}^{n_T(x_a)} 1/v_i(x_a,t)}. \tag{3.7}$$

Equation (3.6) computes the mean average value of multiple measurements while (3.7) computes the harmonic average of the speeds (a.k.a. space mean speed). Having in consideration these definitions we proceed to explain the interconnection with data collected from the highway.

3.3 Data recovery

In this section, we describe first the whole data aggregation process from each individual sensor measurement to the aggregated velocity used for the travel time computation. Later, these concepts will define *instantaneous* and *dynamic* travel times for a given origin-destination path. Associated computations of the corresponding times are established using the aggregated velocity.

3.3.1 Data aggregation process

Fig. 3.2 shows the sensor layout used in the GTL, an experimental platform for real time collection of traffic data coming from a dense wireless sensor network installed in the south ring of Grenoble, see [CDW+15]. Sensor technology provided by SENSYS [Baj+11] consists in magnetometers, which are passive devices detecting the perturbation over a magnetic field when a vehicle crosses over the sensor. Sensors are regularly deployed by pairs separated about 5 m at each lane. They provide aggregated data every 15 s: velocities, flows, occupancies, but

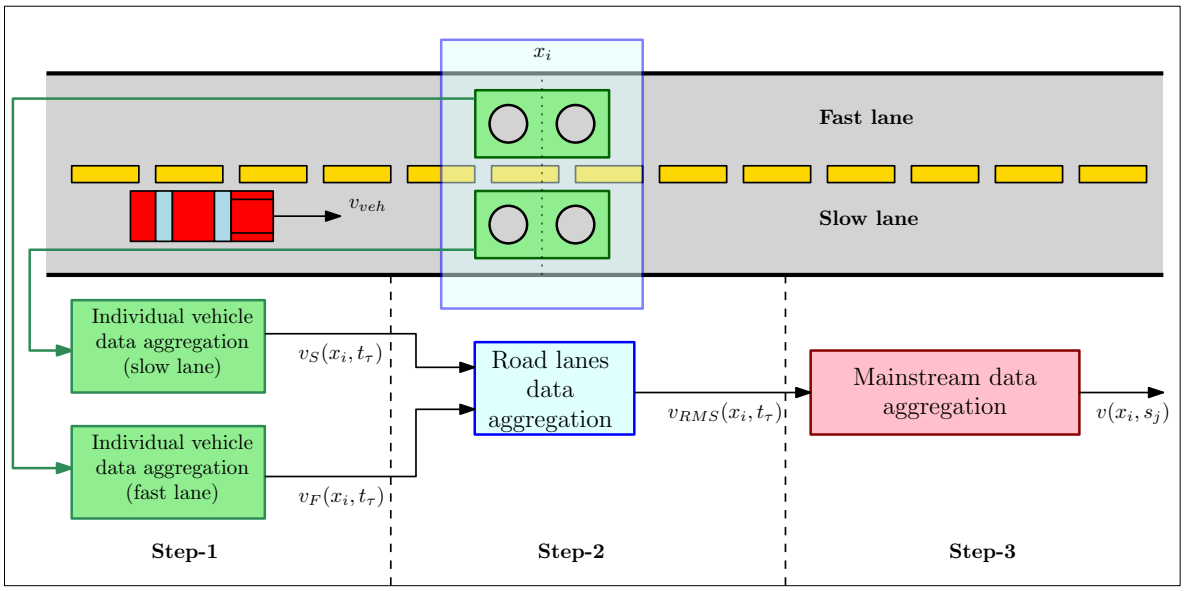


Figure 3.2: Schematic representation of the 3-levels data aggregation from the sensor samples to the final velocity used for travel time prediction.

also histograms of vehicle lengths and speeds for each lane. The whole data aggregation consists in three main steps:

- *Step 1. Individual vehicle data aggregation.* Each pair of sensors samples individual vehicle measurements per lane for: flow, speed and time of detection every time-interval of $\delta = 15\,\text{s}$ at each location x_i. For a vehicle labelled i, its velocity at location x_i, sensed at time t, is denoted by $v_i(x_i, t)$. This data is aggregated at the level of the collection points every time $t_\tau = t_{\tau-1} + \delta$, with $\tau \in \mathbb{Z}^+$, and then sent to a server where information is stored in a database. The aggregated lane speed, $v_{\text{lane}}(x_i, t_\tau)$, results by taking the harmonic average all individual vehicle velocities $v_i(x_i, t)$ as in (3.7), collected at each lane during the time-interval $[t_\tau - \delta, t_\tau]$, that is

$$v_{\text{lane}}(x_i, t_\tau) = \frac{N_{\text{lane}}(x_i, t_\tau)}{\sum_{t \in I_\tau} \sum_i \frac{1}{v_i(x_i,t)}}, \quad I_\tau = [t_\tau - \delta, t_\tau], \tag{3.8}$$

  where $N_{\text{lane}}(x_i, t_\tau)$ is the total number of vehicles during the interval I_τ per lane at location x_i. When no vehicle is detected, the system prevents for such an event by sending a specific code and then v_{lane} is replaced in priority order by: *i*) an average from its neighbourhood lanes values, *ii*) a moving time-horizon average of the actual velocity lane, or *iii*) the free flow velocity (Section 3.4).

- *Step 2. Road lanes data aggregation.* The lane velocities v_{lane} resulting from *Step 1* are now aggregated at the level of a *road measurement section* v_{RMS}. For this, we use a weighted average to better represent the speed center of gravity as a function of the

number of vehicles per lane N_{lane}.

$$v_{\mathrm{RMS}}(x_i, t_\tau) = \frac{\sum\limits_{\mathrm{lane}} v_{\mathrm{lane}}(x_i, t_\tau) N_{\mathrm{lane}}(x_i, t_\tau)}{\sum\limits_{\mathrm{lane}} N_{\mathrm{lane}}(x_i, t_\tau)} \tag{3.9}$$

- *Step 3. Main stream data aggregation.* Finally the lane velocities v_{lane} which are obtained every time-interval δ are now averaged over the database time-interval $\delta_s = N_m\delta$, $N_m \in \mathbb{Z}^+$ (in our case $N_m = 4, \delta_s = 1\mathrm{min}$) which will be used to store velocity information in the database named here the velocity-heat map describing the velocity quantification in the (x_i, s_j)-map, i.e.

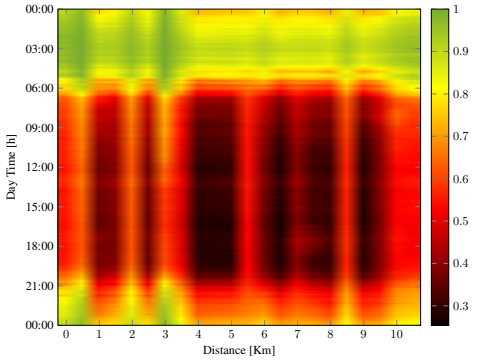
$$\begin{aligned} v(x_i, s_j) &= \frac{1}{N_m} \sum_{t_\tau = s_{j-1}}^{s_j} v_{RMS}(x_i, t_\tau), \\ s_j &= s_{j-1} + N_m(t_\tau - t_{\tau-1}), j \in \mathbb{Z}^+ \end{aligned} \tag{3.10}$$

3.4 Imputation of missing values in traffic speed data

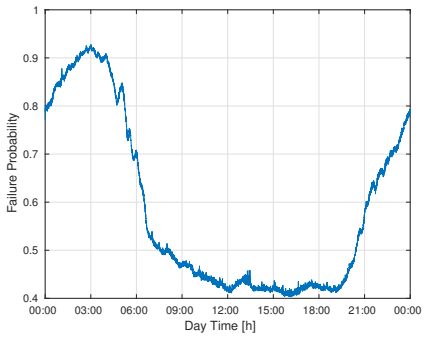
Data loss is any process or event, in information systems, that results in data being destroyed by failures in acquisition, transmission, and storage. Usually, in a traffic sensor network, we are faced to the following failures:

- *Acquisition errors*: When a vehicle, within the time interval $[t_\tau - \delta, t_\tau]$, crosses just one of the two sensors in a specific location, the pair of sensors cannot measure the time window this vehicle took to cross the space between them (See Fig. 3.2), therefore it is not possible to measure the speed of this vehicle. In this case, an arbitrary value is set as speed measurement; the value (-1) is used in this case for Sensys network (See table 7.5)[Baj+11].
- *Transmission errors*: When data is sent by the sensor but the communication link is broken at some point due to networking infrastructure problems, the samples are not received and as a consequence holes appear in the raw data table. Once this case is identified, another arbitrary value is set to the data; the value (-2) is used in this case for Sensys network.
- *Technical problems*: Sensor deployment due to maintenance in the road pavement, power supply malfunctioning, bad configuration constitute other sources of missing samples, battery replacements among others.

Data loss can therefore have a particular occurrence. Let us examine the probability function of missing samples related to each one of the sensors in the network [CDW+15]. We conducted a deep analysis on data recorded from September 1$^{\mathrm{st}}$, 2015 to March 31$^{\mathrm{st}}$, 2016. For each location, we first computed the relative frequency of missing samples at each time and location



(a) Space-time distribution of missing samples - Relative frequency of missing samples with respect to the full historical data set



(b) Network missing samples - Relative frequency of missing samples with respect to the full space data expected at each sample time

Figure 3.3: Missing data evaluation of the GTL from September 1^{st}, 2015 to March 31^{st}, 2016. The figure on the left panel illustrates the distribution of missing samples analyzed from an historical view point. At each sample time and specific location the amount of missing samples is counted within the historical data set. The right panel displays the distribution of missing samples along the day, as a remark the peaks of missing data are concentrated out of the peak hours which benefits the algorithm in this study.

for the velocity examining the amount of missing samples among the full historical dataset. Then we evaluate the network-wide performance. The results are depicted in Fig. 3.3 where the left panel describes the space-time distribution of relative frequency for missing data while the right panel describes the time evolution of data missing at a network level. We observe in Fig. 3.3b the relative frequency of missing samples in time for a historical data set of speeds recorded between September 1^{st}, 2015 and March 31^{st}, 2016. From these results, we can note high values of missing data relative frequency in non peak hours $12:00\text{AM}$-$06:00\text{AM}$ and $09:00\text{PM}$-$12:00\text{AM}$ which reflects the fact that few vehicles are detected by the sensors and no speeds are captured during this intervals of time (acquisition errors). During peak hours, the probability of missing samples varies between 40% and 60%. It is therefore necessary to resort to data imputation in order to replace missing data by their approximations which must be good enough to not impair the travel time prediction.

3.4.1 Mechanism of imputation

A crucial step for the traffic data reconstruction is the imputation of missing or erroneous values. In this case we promote fast computational methods given the needs of real time requirements for the experimental part (See Chapter. 7).

As stated above, missing velocity samples are assigned to some negative values, typically $v(x_i, t_\tau) = -1$ or $v(x_i, t_\tau) = -2$. In the sequel, we introduce the following indicator of valid

data:

$$\mathcal{I}(x_i, t_\tau) = \begin{cases} 0 & \text{if} \quad v(x_i, t_\tau) = -1 \quad \text{or} \quad v(x_i, t_\tau) = -2 \\ 1 & \text{elsewhere.} \end{cases} \tag{3.11}$$

Every sample $v(x_i, t_\tau)$ considered as non valid is assigned as `NaN` [2]. An efficient strategy is to set up an order to replace it by the best approximation according to some type of consistency. Three types of consistency mechanisms can be envisioned: *temporal*, *spatial*, and *historical*.

1. *Temporal consistency*: Based on the assumption that $v(x_i, t_\tau) \approx v(x_i, t_\tau - \delta)$ for some value of $\delta > 0$ the aim is to substitute the missing value at time t_τ by the average value within the time window $[t_{\tau-r}, t_\tau]$

$$\tilde{v}_l(x_i, t_\tau) = \begin{cases} \dfrac{\sum_{l=1}^{r} v(x_i, t_{\tau-l})\mathcal{I}(x_i, t_{\tau-l})}{n(t_{\tau-r}, t_\tau)}, & n(t_{\tau-r}, t_\tau) \neq 0 \\ \texttt{NaN} \quad elsewhere \end{cases}, \tag{3.12}$$

   where $\mathcal{I}$ filters only the valid samples. $n(t_{\tau-r}, t_\tau) = \sum_{l=1}^{r} \mathcal{I}(x_i, t_{\tau-l})$ is a counter of the number of valid samples in the interval $[t_{\tau-r}, t_\tau]$. In the average model described by (3.12) it is interesting to examine the behavior of the absolute error explained in (3.15). The behavior of the absolute error may vary in terms of several factors. First, the amount of missing samples which is represented by the Percentage of Lost Samples (PLS) with respect to the total received information in the data set. Second, the length of the window r taken for the model described in (3.12).

2. *Spatial consistency*: Considering the assumption $v(x_i, t_\tau) \approx v(x_j, t_\tau)$ for x_j within a neighborhood $\mathcal{N}_i$ the value is replaced as

$$\tilde{v}_s(x_i, t_\tau) = \begin{cases} \dfrac{\sum_{x_i \in \mathcal{N}_i} v(x_i, t_\tau)\mathcal{I}(x_i, t_\tau)}{|\mathcal{N}_i|}, & |\mathcal{N}_i| \neq 0 \\ \texttt{NaN} \quad elsewhere. \end{cases} \tag{3.13}$$

   In the context of a highway $\mathcal{N}_i = \{x_{i-1}, x_{i+1}\}$ represents the sensor ahead in the following location and the sensor behind in the previous location in the space axis. The spatial imputation described in (3.13) seeks to recover samples from neighbors. The strategy is useful when a particular sensor fails leading to increase the amount of missing samples at specific collection points. Even though this strategy is a good alternative when one single sensor is failing, it does not offer the same performance when a set of consecutive sensors fail. In this case the case the method is not sufficently robust, one alternative is considering neighbors at larger distances or create interpolation algorithms like in [GKS15]. We rely on historical trends to provide additional information which in fact contributes to better recovery of the information.

[2] `NaN` corresponds to a non-numerical value which avoids erroneous identification and operation of missing data.

3. *Historical consistency*: A third step in the process is applied when neither the local imputation nor the spatial imputation work. We denote $v^{(h)}(x_i, t_\tau)$ as the speed in space-time plane for some day h in the set of past days and $\mathcal{I}^{(h)}(x_i, t_\tau)$ its corresponding valid data indicator. Under the assumption of historical similarities $v^{(h)}(x_i, t_\tau) \approx v^{(d)}(x_i, t_\tau)$ for two different days h and d, the value is imputed as an average of historical values of preselected days, hence

$$\tilde{v}_d(x_i, t_\tau) = \begin{cases} \dfrac{\sum_{h \in \mathcal{H}_d} v^{(h)}(x_i, t_\tau)\mathcal{I}^{(h)}(x_i, t_\tau)}{|\mathcal{H}_d|}, & |\mathcal{H}_d| \neq 0 \\ \texttt{NaN} \quad elsewhere \end{cases} \tag{3.14}$$

3.4.1.1 Evaluation of the imputation schemes

In order to assess the strategies we examine a day (October 16^{th}, 2015) with no missing data from 07 : 00AM to 07 : 00PM. To emulate data loss, missing samples are generated according to a specific pattern, then each imputation mechanism is evaluated according to the following absolute error AE:

$$AE(x_i, t_\tau) = 100\Big|\frac{\tilde{v}(x_i, t_\tau) - v(x_i, t_\tau)}{v(x_i, t_\tau)}\Big|(1 - \mathcal{I}(x_i, \tau_t)). \tag{3.15}$$

Figure 3.4 displays the behavior of the mean absolute error in terms of the order r and the PLS. The analysis starts at $7AM$ and ends at $7PM$, during this time interval a random set of samples is generated according to the probability sample distribution of each one of the sensors (See Fig. 3.3a). The samples are retired from the original data set and the algorithm is applied for several values of r. At the end, the absolute error described in (3.15) where $\tilde{v}(x_i, t_\tau) = \tilde{v}_l(x_i, t_\tau)$ is considered and the mean value is computed (See. Left-panel in 3.4). On the other hand the amount of recovered samples is analyzed in the right panel of Fig. 3.4. In this case the index Percentage of Recovered Samples (PRS) denotes the ratio between the number of recovered samples with respect to the total number of missing samples and it is computed for different values of r. The result shows that in general this method performs better for lower time aggregations but in it is not sufficiently robust when dealing with high amount of missing samples. It is worth to notice, when dealing with low flows the selection of a low value for r is not convenient for the performance of the method, for his reason we select $r = 4$ which corresponds to small variations during the last 1 min of history. The low robustness of this methods lead to the proposition of supplementary sources of information as the spatial source.

We also evaluate the quality of imputation by examining PRS in a real case scenario of missing data. In this case, a pattern of missing samples as ocurred in the day October 7^{th}, 2015 is emulated over the day October 16^{th} in order to evaluate the method in more realistic conditions. The absence of data can be seen in Fig. 3.5 and the results for the PRS and AE are detailed in Table 3.1. As a main result it can be seen that the method with better performance is the historical due to the amount of information available for the recovery,

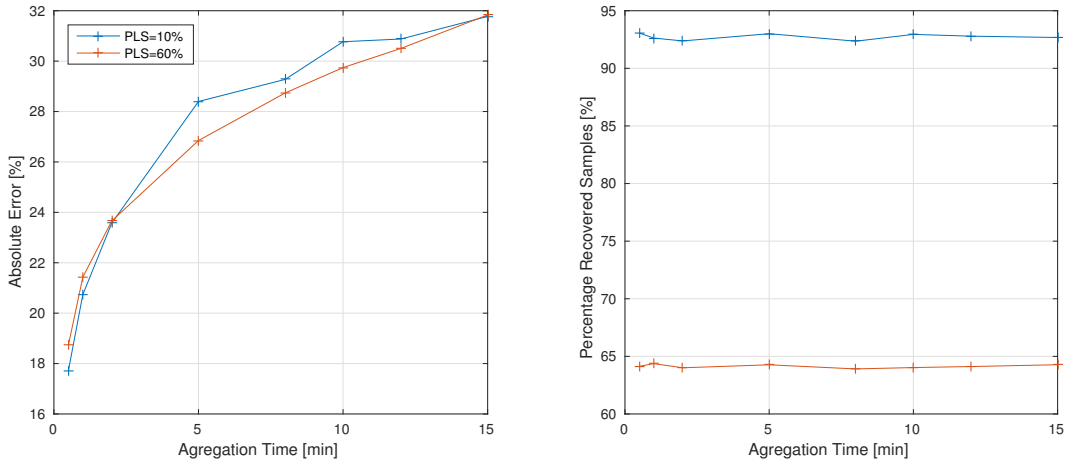


Figure 3.4: Temporal imputation performance 16$^{\text{th}}$ October, 2015. Left panel - The mean value of the absolute error in terms of variation of r. It's worth to remark $4 < r < 60$ which corresponds to variation in the aggregation time between 1 min and 15 min. Colors in the curves denote different PLS Right panel - PRS. In this case ratio of between the number of recovered samples with respect to the number of lost samples is presented for two cases of PLS.

Method	PRS (%)	AE (%)
Temporal	10.6	29.4
Spatial	83.1	25.7
Historical	88.2	22.0

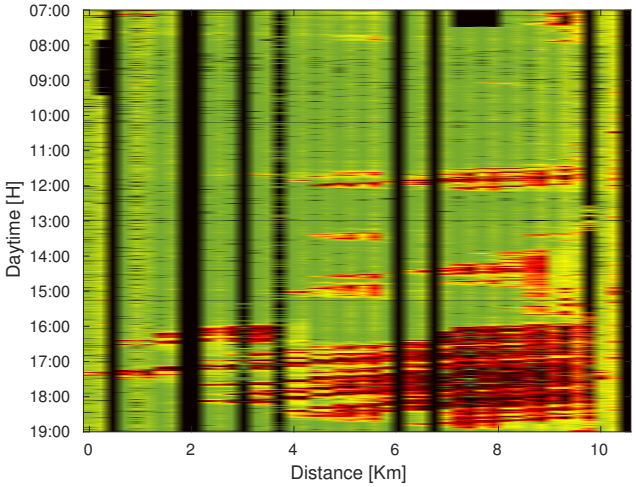
Table 3.1: Percentage of Recovered Samples (PRS) and Absolute Error (AE) for each method applied in the day October 16th, 2015. $r = 4$ is the parameter *temporal* strategy

.

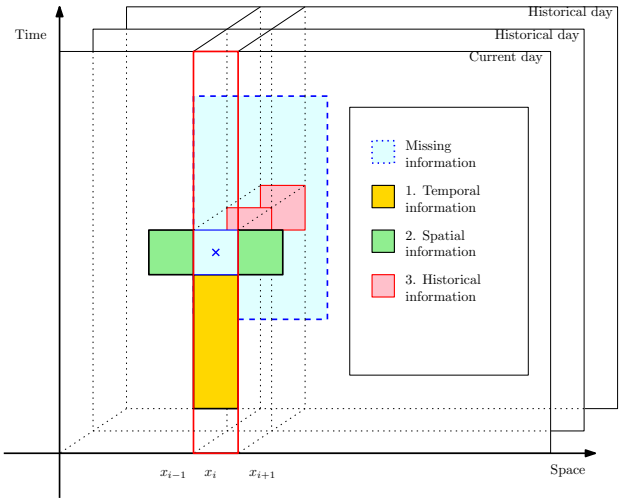
however it is computationally expensive since each replaced sample should be obtained from the average of historical sets, the quality of the recovery may also vary on the amount of historical days.

Temporal or spatial consistency is cheap from a computational point of view since only the current day data is used for imputation in contrast to the historical one. On the other hand, the amount of error in the case of spatial and historical remain quite close. Finally, we adopt the following imputation sequence (detailed in Algorithm 1 in Fig. 3.6) which seeks to complete gradually the day by first recovering information from smaller components in the network (sensors) and then up scaling to bigger aggregations (neighbours of sensors, full network):

1. Spatial consistency (since it produces a low AE and it is cheaper than historical, also it could potentially cover a PRS similar to the historical), else
2. Temporal consistency (since it provides the local information and the cost is very cheap



(a) Missing sample distribution- Speed spatio-temporal distribution, in black it is showed the spatio-temporal distribution of missing samples according to a specific day of the week. The distribution of missing samples is particularly found on October 7^{th}, 2015



(b) Sources of information for the imputation algorithm

Figure 3.5: Real traffic scenario of missing samples. The figure describes a typical missing data distribution in which information at particular locations is lost during the full day. The sources of imputation to replace the missing data are shown in 3.5b

Data: Raw: $v(x_i, t_\tau)$
Result: Imputed value $v(x_i, t_\tau)$
Define parameters: $r = 4\text{min}$
$\mathcal{N}_i = \{x_{i-1}, x_{i+1}\}$
$\mathcal{H}_d = \{h_d^{(1)}, \ldots, h_d^{(H)}\}$
Compute $\tilde{v}_l, \tilde{v}_s, \tilde{v}_h$, as (3.12), (3.13) and (3.14)
while $v(x_i, t_\tau) < 0$ **do**
  $v(x_i, t_\tau) \leftarrow$ `NaN`
  $v(x_i, t_\tau) \leftarrow \tilde{v}_s(x_i, t_\tau)$
  if $v(x_i, t_\tau) =$ *NaN* **then**
    $v(x_i, t_\tau) \leftarrow \tilde{v}_l(x_i, t_\tau)$
  end
  if $v(x_i, t_\tau) =$ *NaN* **then**
    $v(x_i, t_\tau) \leftarrow \tilde{v}_h(x_i, t_\tau)$
  end
end

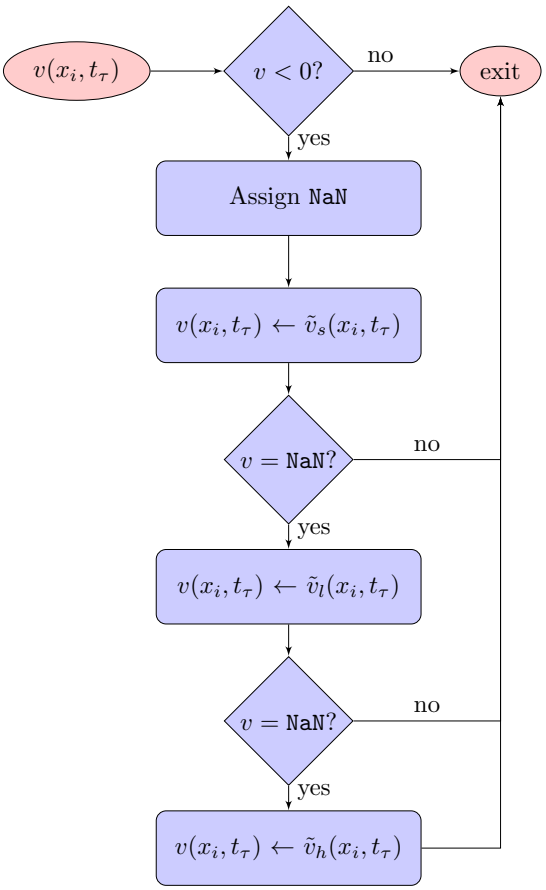


Figure 3.6: Algorithm applied for imputation. Left-hand side: Algorithm sequence / Right-hand side: Flow diagram to illustrate data work flow

with respect to the historical one), else

3. Historical consistency in which the objective is to recover a reduced set of missing samples.

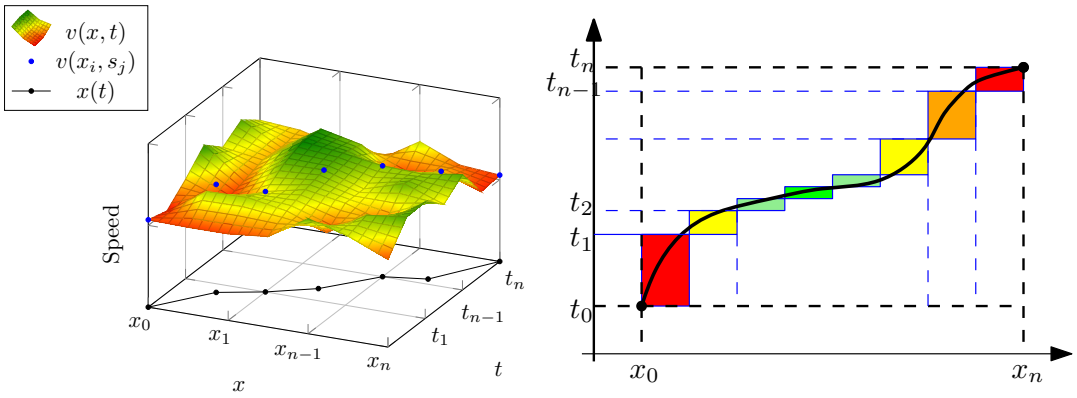
3.5 Travel time: definition and computation

We have presented already techniques of speed collection. At this point, we focus our attention on travel time estimation based on indirect methods based from loop data. This section aims to establish concepts and properties about Travel Time. In addition we introduce notions on how the travel time is forecasted. For this, let us imagine a map of a vehicle trajectory traversing through a route limited by the space interval $[x_0, x_n]$. Consider the case of single origin-destination path, and let $v(x, t)$ be the continuous space/time distribution of the velocity field along the considered single road segment as seen in Fig. 3.7. Here, we are be interested

in obtaining a mathematical expression which describes the Travel Time between the given positions. This formula will be particularly helpful in the forecasted travel time formulation. If the velocity distribution $v(x,t)$ in $[x_0, x_n]$ is known, the vehicle travel time at infinitesimal space intervals is given by $dt \triangleq dx/v(x,t)$ which yields the following integral equation, that gives, in continuous time, the exact value of travel time

$$t_n - t_0 = \int_{x_0}^{x_n} \frac{dx}{v(x,t)},$$

The case of perfect knowledge of $v(x,t)$ is far from reality. In general, a discretized grid of speeds is feasible based on installed sensors within roads. These points allow an accurate approximation of $v(x,t)$ as seen in Fig. 3.7a



(a) Space/time speed surface

(b) Speed updates for DTT computation

Figure 3.7: Space/time speed evolution for DTT computation

We define two approximations for the travel time given by:

Definition 3.1 (Dynamic Travel Time)
The Dynamic Travel Time (DTT) from position x_0 to x_n starting at time $t = t_0$ and ending at time $t = t_n$ is defined as:

$$DTT(x_0, x_n, t_0) = t_n - t_0 = \sum_{i=0}^{n-1} \frac{x_{i+1} - x_i}{v\big(x_i, DTT(x_0, x_i, t_0)\big)}, \tag{3.16}$$

where $x_i, i = 1, \ldots, n-1$ stand for intermediate locations between x_0 and x_n.

Definition 3.1 considers the effect of traffic progression along the road since speeds dynamically change in time and space affecting the *DTT* term inside the summation. It states that the vehicle's arrival time at the downstream position x_n as a function of the time spent up to the location $i-1$. Literature also proposes the denominated ITT.

Definition 3.2 (Instantaneous Travel Time)
The Instantaneous Travel Time (ITT) from position x_0 to x_n starting at time $t = t_0$ and

ending at time $t = t_n$ is defined as:

$$ITT(x_0, x_n, t_0) = t_n - t_0 = \sum_{i=0}^{n-1} \frac{x_{i+1} - x_i}{v(x_i, t_0)}. \tag{3.17}$$

Clearly, it is seen then that while DTT is consistent with the traffic conditions experienced by a driver along the road, ITT is based on a very strong and not always realistic assumption (i.e. speed does not change in time), which becomes more critical as the length of space interval increases. Therefore, *DTT* can be considered as the "true" travel time of a vehicle along the velocity flow whereas the *ITT* is a crude approximation of the true one. Definition 3.16 implies a recursive computation of the travel time and the computation select points as seen in Fig. 3.7a from discrete velocity field $v(x_i, s_j)$ as stored in the database.

Computing (3.16) implies an approximation for the integral term in consistence with the discrete velocity field $v(x_i, s_j)$ as stored in the data base. Several approximation are possible. For instance, the DTT from two neighbor positions $(x_i,\ x_{i+1})$, at t_k can be approximated using the Euler integration formula, where $v(x, t)$ is considered constant by holding his value in time and the space; $v(x, t) \approx v(x_i, t_k)$, in $I_i = [x_i, x_{i+1}] \times [t_k, t_{k+1}]$, or by averaging over two neighbor locations at the initial time; $v(x, t) \approx (v(x_i, t_k) + v(x_{i+1}, t_k))/2$ and holding this value over I_i. The DTT according to (3.16) for two consecutive locations is given by

$$DTT(x_i, x_{i+1}, t_k) = t_{k+1} - t_k \approx \frac{x_{i+1} - x_i}{v(x_i, t_k)}, \quad t_k := DTT(x_0, x_i, t_0).$$

Note however that the time argument, t_k, in $v(x_i, t_k)$ is a continuous variable (k is consider here as an event), and that velocity data is stored in a discrete-time base $v(x_i, s_j)$. Therefore, index j need to be defined from t_k to match $s_j = \lfloor t_k \rfloor$, that is $j(t_k) = \frac{\lfloor t_k \rfloor}{\delta_s}$. Under this consideration, it is illustrative to see that

$$\begin{aligned}
t_1 &= t_0 + (x_1 - x_0) v(x_0, s_0)^{-1} \\
t_2 &= t_1 + (x_2 - x_1) v(x_1, s_{j(t_1)})^{-1} \\
&\vdots \\
t_{k+1} &= t_k + (x_{k+1} - x_k) v(x_k, s_{j(t_k)})^{-1} \\
&\vdots \\
t_n &= t_{n-1} + (x_n - x_{n-1}) v(x_{n-1}, s_{j(t_{n-1})})^{-1}
\end{aligned}$$

Then, DTT can be written in a compact form as

$$DTT(x_0, x_n, t_0) = \sum_{k=0}^{n-1} (x_{k+1} - x_k) v(x_k, s_{j(t_k)})^{-1}. \tag{3.18}$$

Similarly, the following formula is obtained for ITT

$$ITT(x_0, x_n, t_0) = \sum_{k=0}^{n-1} (x_{k+1} - x_k) v(x_k, s_{j(t_0)})^{-1}. \tag{3.19}$$

3.5.1 DTT and ITT main properties

It is worth to underline the main properties of two traveling time definition introduced earlier in terms of: *precision*, *causality*, and *composability*.

3.5.1.1 Precision.

Let for simplicity assume homogenous sensor distributions $(x_{j+1} - x_j) = \delta_x$. The error introduced in the ITT computation (we assume that the DTT is the true travel time) crossing these sections of length δ_x can be estimated as:

$$ITT_{error} = DTT - ITT = \delta_x \sum_{k=1}^{n-1} (v(x_k, s_{j(t_k)})^{-1} - v(x_k, s_{j(t_0)})^{-1}). \tag{3.20}$$

This error will grow as the traffic conditions experience important changes in the near future. In presence of traffic shock waves the term $(v(x_k, s_{j(t_k)})^{-1} - v(x_k, s_{j(t_0)})^{-1})$ will induce large differences. This case is illustrated in Fig. 3.8. Except for free flow situations, when the DTT and the ITT become close, important errors in the ITT are also expected when considering large and complex networks with heterogeneous traffic conditions along trajectory paths. For this particular case the ITT under estimates the DTT and the main differences start to appear in the middle of the trajectory when the congestion wave reaches the vehicle trajectory, i.e at $x = 5$Km, $t = t_0 + 7$min. The free flow travel time is here 7min. Error bounds incurred by the ITT computation from May to December 2015. Maximum error reaches up to 3 times the free-flow travel time.

Furthermore, ITT can be computed by the sum of the individual ITT values of each road section (region between two sensor points) while the DTT is computed by the accumulation of several terms, this effect adds computational complexity for multiple origin destinations to the algorithm and prediction strategies as the one developed in [Lad+16b] cannot be applied directly.

3.5.1.2 Composability

Consider tree arbitrarily locations (x_0, x_1, x_2) ordered as $x_2 > x_1 > x_0$, then it is easy to see that the ITT and the DTT have the following properties:

i) $ITT(x_0, x_2, t_0) = ITT(x_0, x_1, t_0) + ITT(x_1, x_2, t_0)$

i) $DTT(x_0, x_2, t_0) \neq DTT(x_0, x_1, t_0) + DTT(x_1, x_2, t_0)$

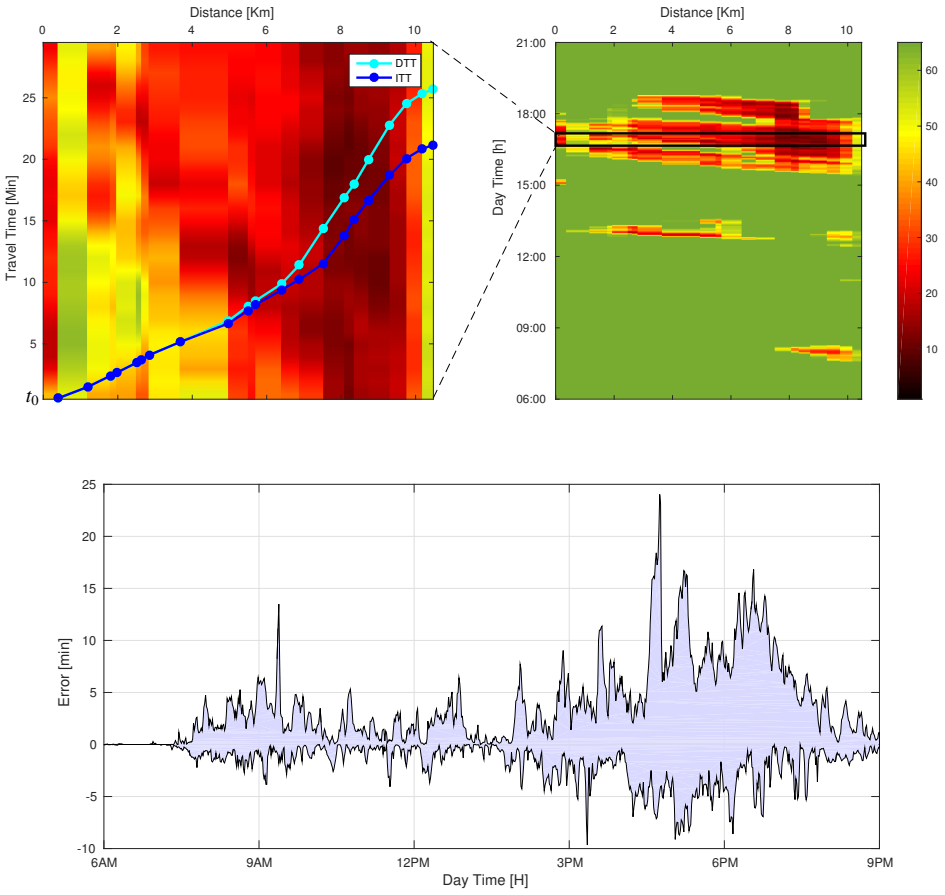


Figure 3.8: Top-Figure. Illustration of the time/space evolution of the DTT vs. the ITT computed with data from the GTL [CDW+15] for September 1^{st}, 2015. The heat map in the figure displays the measured speed profile in Km/h for the day.

The ITT can thus be computed by the sum of the individual ITT values of each road section (region between two sensor points) while the DTT is computed by the continuation of the flow and cannot be determined by the individual values of each section.

3.5.1.3 Causality

Note that the ITT can be straightforwardly computed using information at the current time t_0. However, to compute the DTT it is necessary to build predictions ahead in time for $v(x_k, t_0)$, namely to device prediction algorithms for $\hat{v}(x_k, s_j), \forall s_j \in [t_0, t_0 + \Delta_f]$, where Δ_f is the time-horizon necessary for such a prediction. The size of Δ_f should be selected based on the maximum Travel Time spent in the road. The predicted DTT is then computed using such a predictions as

$$\widehat{DTT}(x_0, x_n, t_0) = \sum_{k=0}^{n-1} (x_{k+1} - x_k) \hat{v}(x_k, s_{j(t_k)})^{-1}, \tag{3.21}$$

with $\hat{v}(x_k, s_0) = v(x_k, t_0)$, and $j(t_k) = \frac{\lfloor t_k \rfloor}{\delta_s}$.

3.5.2 Imputation effects on DTT

Data imputation is a required step in order to have a complete full matrix of speed information that allows to compute the travel time. Without this step, some errors are generated in the computation of the travel time or even inconsistencies due to the nature of measurements. In this case we check the performance of Algorithm 1 in terms of reconstruction error on DTT.

After applying Algorithm 1, we obtained the results depicted in Fig. 3.9. In this case the DTT is computed with the real speed profile and compared with respect to the values obtained from imputed one. Increasing the proportion of missing samples causes the error to rise by a small amount. In 90% of the cases the error in the DTT does not overpasses 5%. From the panels on scattered data the concentration of error occurs more often in congested time when travel time is high, however the error incurred can still be considered satisfactory for the estimation of the DTT.

3.6 Final comments on the chapter

The current chapter has dealt with definitions and introduction of the TT definition used in this dissertation. It has been shown experimentally the errors incurred when the ITT is used with respect to the DTT and the reasons for selecting DTT as a variable to be forecasted. Furthermore, the chapter explores the full process for data collection and consequent imputation of missing values. An imputation method under analysis based on interpolation techniques

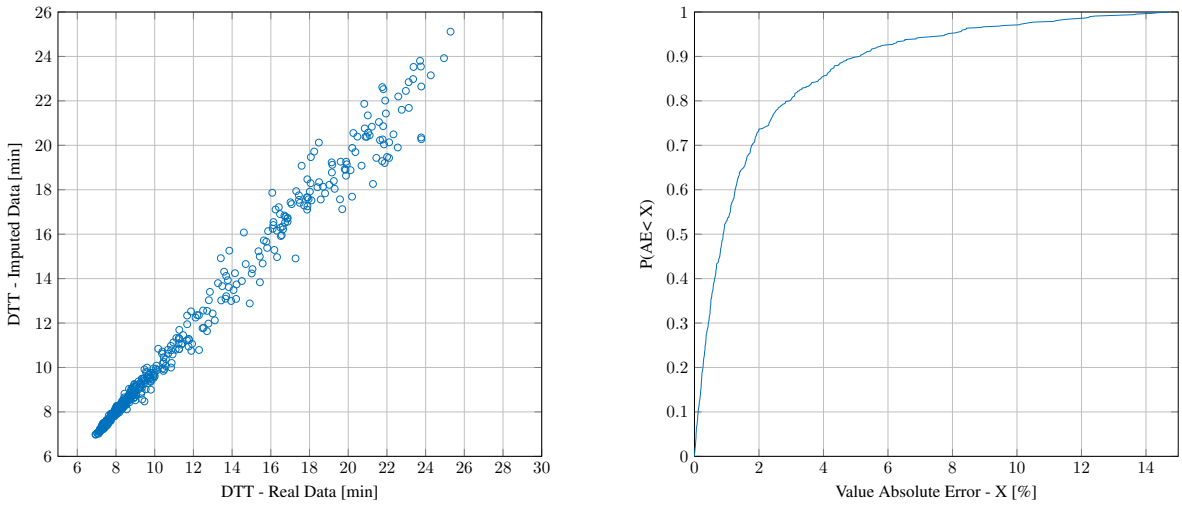


Figure 3.9: Left Panel. Scatter plot of actual measured DTT vs estimated DTT after imputation. The test in this case shows the scenario 2 when a regular pattern of missing data is removed. The imputation algorithm is applied over speeds (See Algorithm. 1), and DTT is then computed as (3.18). Right Panel. Absolute Error Cumulative Distribution Function

from different sources is fully described and characterized in an experimental way. Results have established that the impact of this method has low impact when computing the DTT which will be then used as a main source of information for the clustering and forecasting techniques described in the subsequent chapters.

The remaining of the manuscript will be dedicated to the mechanism used to forecast . The DTT obtained with the imputed speed will be used later to constitute the historical database from which the prediction algorithm will be devised. In the next chapters, we describe how to exploit such historical knowledge through clustering and a more detailed description on the prediction mechanism of travel time $\widehat{DTT}(x_0, x_n, s_j), \forall s_j \in [t_0, t_0 + \Delta_f]$.

Chapter 4

Data clustering

Contents

4.1 Overview

The previous chapters have provided an introduction to the problem of TSE and had presented the ways to compute TT based on MLD. The data reconstruction problem afforded via imputation algorithms solves an estimation problem in which correlation among variables is either instantaneous or just related to past data either in the history or earlier during the day. In particular, for the case of DTT, computation is performed along the variation of the speed profile in the space/time plane according to (3.18). This intrinsic dynamic dependance highlights the natural difficulty of the problem in TSP. Pattern identification is a tool that provides accuracy to the DTT computation and hence the forecasting problem. Given the conditions of big streams of data stored on the GTL platform, we aim to exploit properties contained in the data for forecasting purposes.

This chapter exploits features in the data by means of clustering of historical data. The main goal when studying historical traffic information is to identify repetitive patterns in the underlying data. In other words, detect whether the traffic patterns on any given day have similarities with traffic patterns in other days. In order to find these patterns we carry out a separation by using suitable clustering techniques. These techniques organize data into homogeneous groups where the within-group-object similarity is minimized and the between-group-object dissimilarity is remains maximal as possible. This section focuses on the development of a methodology for data clustering, where the similarities between the profiles within limited time periods are indeed difficult to extract.

We introduce and advocate clustering as a tool to separate traffic regimes for DTT and speeds. Observations in real traffic data taken from the GTL illustrate some examples in which classification of data can lead to identification of partitions also corresponds to separation of traffic regimes (see Fig. 4.1). This empirical evidence is used as a main support to find dynamic partitions of data along the day. The main objective of the partitions are to characterize a particular traffic regime (free flow, light congestion, strongly congested) and to compress historical information by representing with very few features each one of the partitions. This features will then be integrated into an observation model used to devise short-term time forecasts detailed in Chapter 5.

The mechanism proposed to create the multiple clusters is the K-means algorithm. The setup of parameters for the implementation of this technique is also provided. First, a specific initialization chosen for this algorithm is presented in order to avoid empty clusters and generate repetition. We make an evaluation of the technique proposed by [PDN05] within the traffic case to determine the number of partitions required for the clustering stage. The results show that the number of partitions vary according to the moment of the day and therefore a dynamic clustering is more desired for better accuracy at the forecast stage.

4.2 A motivation for data traffic clustering

The evolution of ITS technologies and data collection technologies has created historical databases of TT where clustering has been a key tool to analyze traffic. Approaches like [YRG15] have used this technique with the main idea of separating characteristics on traffic. One main reason to use clustering techniques is the possibility to compress information, extract an underlying structure, and also as a way to perform natural classification sufficiently significant to characterize the behavior of the system [Jai10]. In common highway applications TT is obtained through indirect measurements (see Chapters 2 and 3) starting from speeds collected by MLD. For this purpose we first analyze and visualize patterns existing in speed profiles.

Fig. 4.1 is a prominent example of this behavior. We depict the spatio-temporal speed distribution in the Rocade Sud during November 2015. 30 days of data are displayed in the x-axis and for each day the axis contains spatial information of the speeds while the y-axis contains temporal data during the day. Blue boxes show the boundary of the space data in the x-axis. Boxes denote the limits of data for a single day. As it can be seen similarities along the days of the week emerge, a first intuition would be to identify *a-priori* patterns given a full set by classifying them into weekdays or week-ends [WB05]. Although this approach is naive, it separates out two different types of usage of the traffic network. To observe more similarities among data, let us consider the spatio-temporal matrix V containing the speeds for a single day. Let us define $v = vec(V)$ a vectorized version of the variable by piling in a single column all network measurements. Fig. 4.2 illustrates via a radar diagram the average cross correlation of v between days of the week during the year December 2015 - December 2016. The behavior of correlation shows that in general days are in all cases well highly correlated

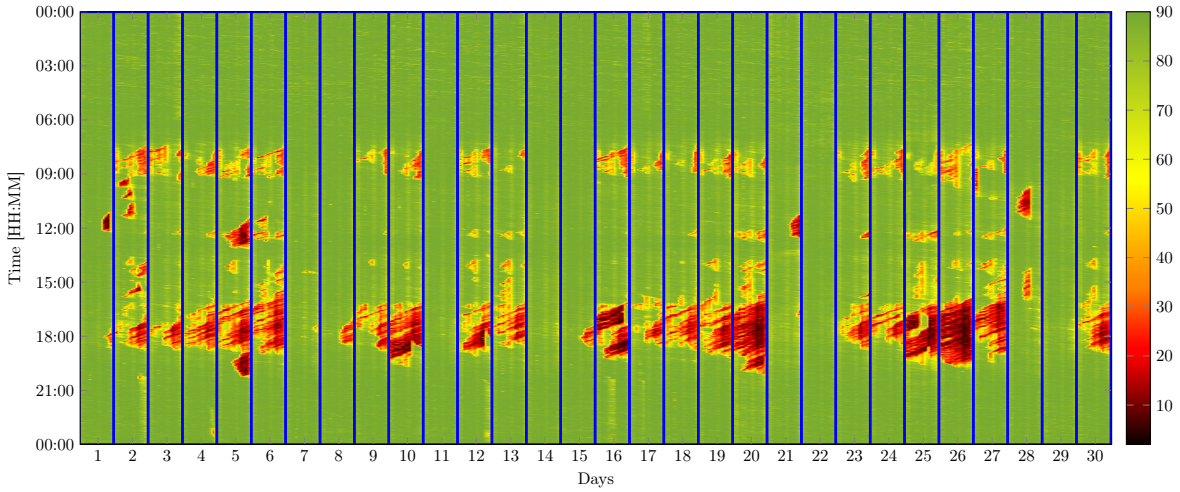


Figure 4.1: Spatio-temporal speed patterns in $\mathrm{km\,h^{-1}}$ during November 2015. Weekday patterns are clearly distinguishable from weekend patterns. Among weekdays some days are similar among them sharing some particular speed profiles

with the corresponding same day in the history of the year. This behavior suggests that the highway may also follow weekly patterns.

Although the same weekday correlations are strong, in between correlations (meaning correlations of weekday i with weekday j) remain comparable and sufficiently important not to be ignored, so further pattern detection algorithms are required in this case. The trip starting and ending time, travel distance, travel frequency, activity duration, and some analogous features are the typical of vehicle travel behaviors. All these aspects have a significant effect on the traffic condition in a direct or indirect way. For example, the distribution of the trip starting and ending time of all vehicles will decide the peak-hour time. Better understanding of these characteristics will be helpful to analyze the travel pattern and travel mode of vehicle.

Based on the DTT computations established in Chapter 3 we analyze the clustering of DTT considering the problem from a time series point of view.

4.3 Clustering of DTT data

The aim of clustering is to divide a set of elements, time series of DTT in our case, into subsets or partitions denominated clusters. Elements belonging to the same cluster share natural properties explained by a similarity measurement. The tool has been widely used in applications for pattern recognition, image segmentation and also in traffic [MGF11] even in DTT computation as in [ECR14]. In this case, we consider data clustering as a tool for identification of multiple regimes in traffic.

To illustrate the interest in clustering, let us consider a set of DTT data computed between September 2015 and May 2016, from 07 : 00AM to 07 : 00PM everyday. In order to highlight

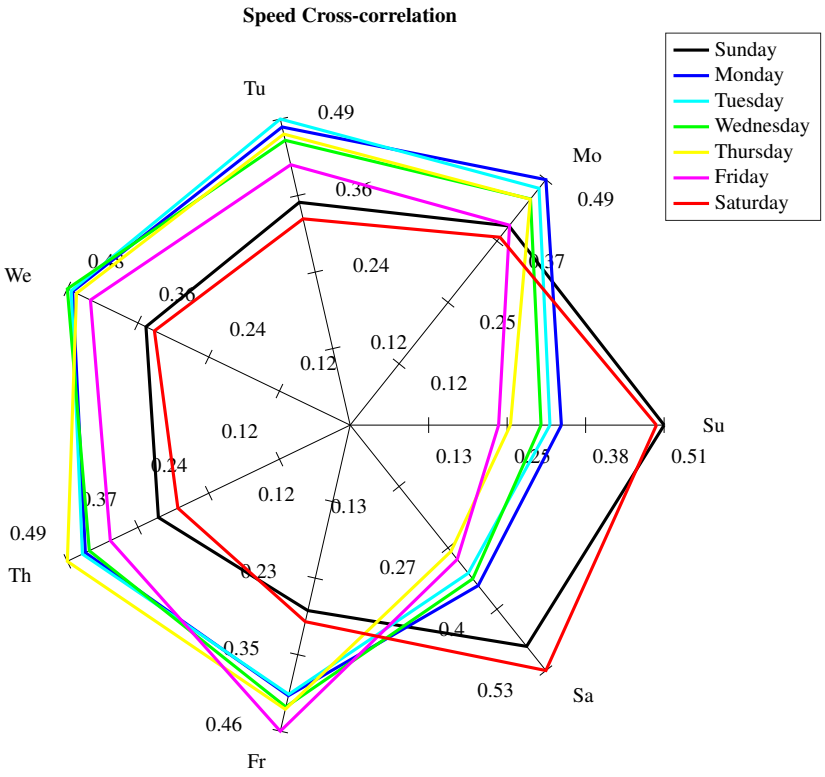
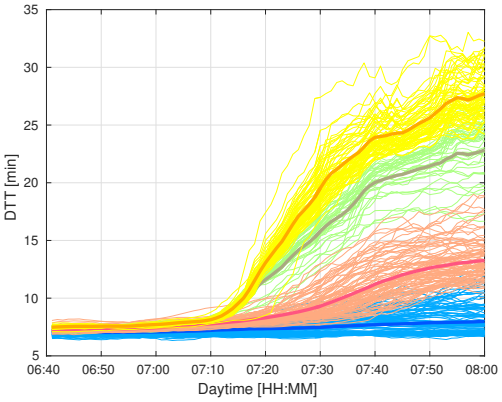


Figure 4.2: Speed crosscorrelation between days of the week

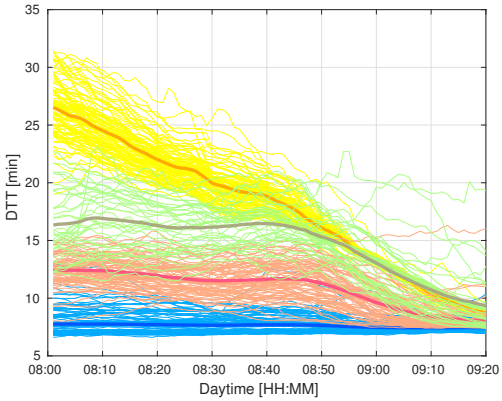
similarities between these data, we show the full set in short intervals of time along several times in the day (See Fig. 4.3). In addition we plot in bold color the effect of 4 identified centroids for this data. As it is seen the data might be very variable during the day, nevertheless it contains historical similarities, that, in this particular case, distinguish a congested and free flow traffic regimes. In Fig. 4.3, color represents elements whose euclidean distance is the smallest to the corresponding centroid. We illustrate several moments of the day and its corresponding traffic behaviors. In order to identify these partitions, clustering is indeed a suitable tool [Kha+16]. Several aspects must be taken into account for the clustering algorithm, such as: time series to be clustered, type of algorithm, number of partitions. The identification and assessment of these aspects are presented in the following subsections.

4.3.1 Dynamic clustering

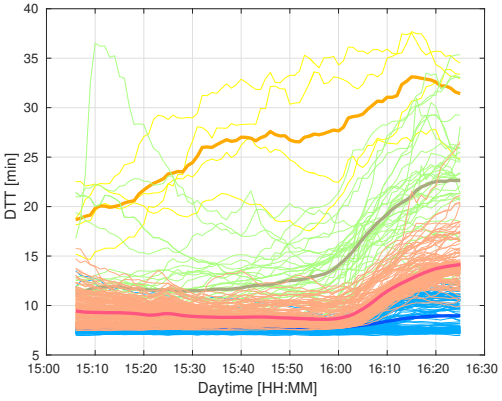
Traffic data exhibits regular patterns day after day, and the time window in which these patterns are identified constitute an important aspect within the clustering technique. There is a clear distinction between working days and weekends for instance. One can think clustering per day is enough. Before providing details on the clustering algorithm used in our case, let us consider Fig. 4.4 where full days time series have been used. A similarity index is given by the smallest Euclidean distance between the day in solid line and each one of the days in dashed lines. In Fig. 4.4, it establishes high similarity (minimum distance) between the day in solid



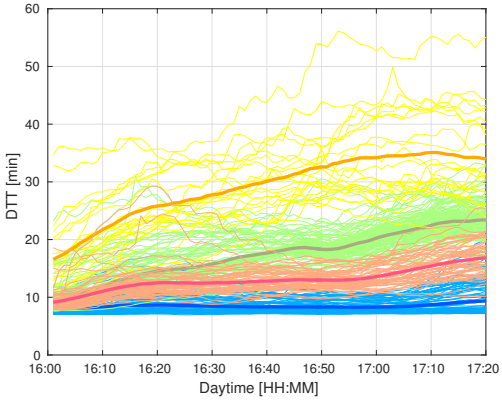
(a) Morning appearing congestion



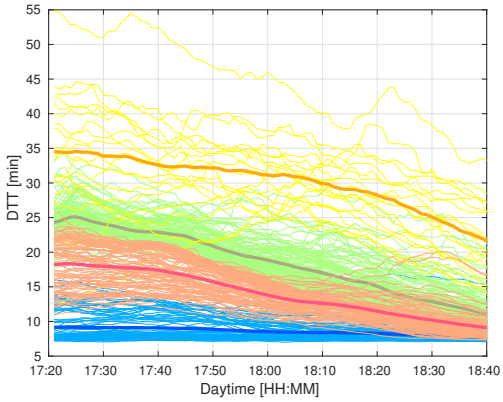
(b) Morning vanishing congestion



(c) Mid-afternoon appearing congestion



(d) Afternoon congestion



(e) Evening vanishing congestion

Figure 4.3: DTT time series for an historical dataset. In color, time signals representing centroids for particular groups. Along the day 4 fixed clusters are identified describing regimes of the traffic behavior such as free-flow conditions, middle level congestions and strong congestions

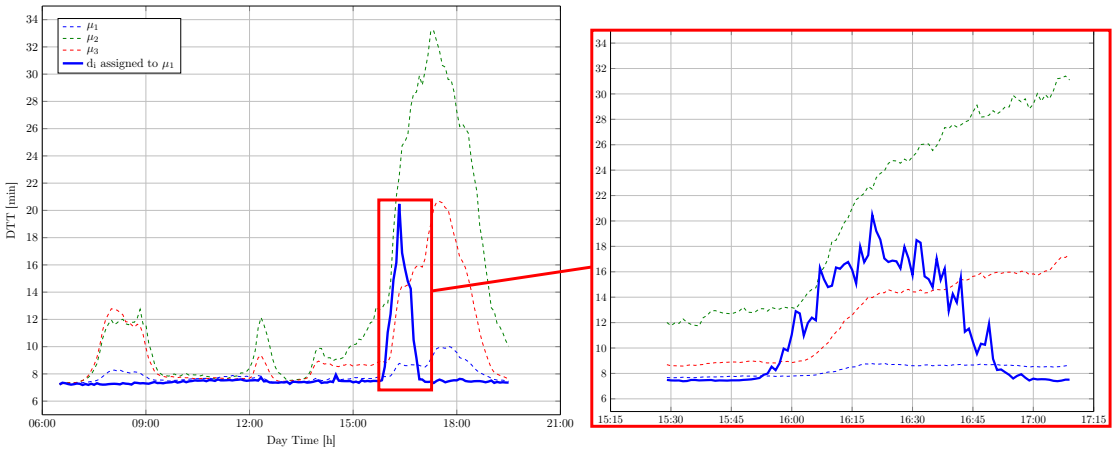


Figure 4.4: Full days clustering performance vs dynamic clustering. Bad assignment may occur when long windows of time are considered to establish similarity.

blue and the day in dashed line marked as μ_1. We can indeed note a strong similarity almost all the day except for the time window 03 : 45PM to 05 : 00PM. By reducing the analysis window to this time period, the day of interest is rather assigned to the dashed day marked as μ_3. Since the knowledge gained from the clusters will be used for prediction purpose, we promote dynamic clustering where clustering is achieved using a moving time window.

One of the parameters to be chosen is the size of the sliding window that is being clustered. The Auto Correlation Function (ACF) is a tool that allows the correlation among samples at time k with samples at time $k-h$ where h denote the lag. The ACF is given by

$$\rho_y(h) = \frac{\sum_{k=h+1}^{T}(y(k)-\bar{y})(y(k-h)-\bar{y})}{\sum_{k=h+1}^{T}(y(k)-\bar{y})^2}, \tag{4.1}$$

where $\bar{y}$ represents the mean value along the journey. We measure the ACF in a historical dataset and we illustrate the mean behavior and its variance in Fig. 4.5. From the results it can be seen that along the historic dataset a lag of 60 min is good enough to capture information that may be associated to the same cluster. This happens when the smallest possible value in the distribution reaches $\rho(h) \approx 0.2$ typically considered as a low correlation.

The dynamic clustering selected for this phase is fixed as follows. For a particular time of predicition k_0 the window to be considered for the clustering is a symmetrical window of size $[k_0 - \Delta_f/2, k_0 + \Delta_f/2]$ where $\Delta_f = 90$ min. Now we introduce the separation mechanism under study.

4.3.2 Clustering techniques

Several techniques have been developed in order to extract features from data, in general all of them identify the underlying structure in data, perform natural classification and in addition

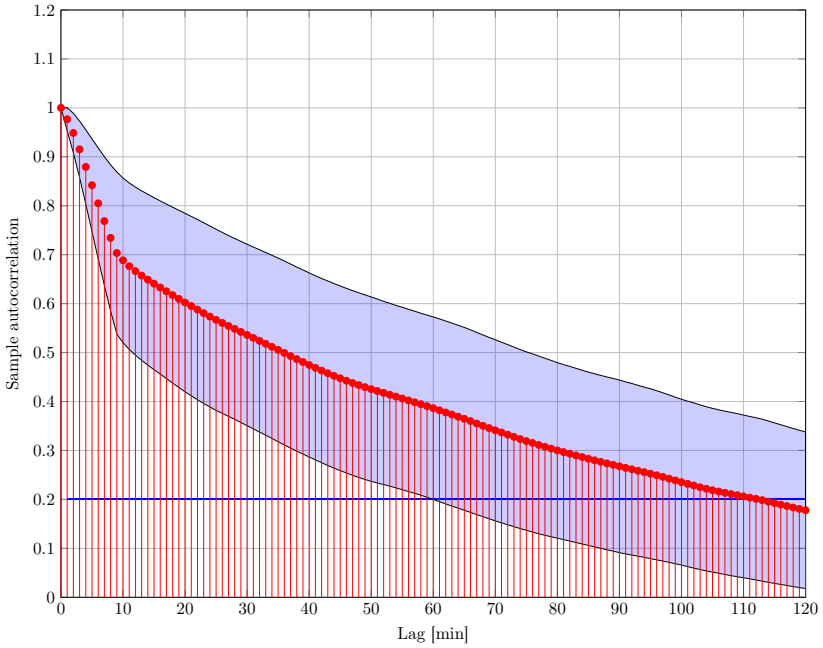


Figure 4.5: DTT autocorrelation function for 244 days Sept 2015 - May 2016. The red points correspond to the mean of the ACF for the historical dataset. The blue value represents the standard deviation. The optimal value is reached when the lag is around 60 min.

serve as a tool to compress information [Jai10]. In our case we aim to identify determined traffic regimes from time series by compressing multiple historical realizations of variable in partitions. Clustering time series seeks similarity between fluctuations of a variable in time, 4 main categories are commonly applied in time series classification: hard partitioning, connectivity based, distribution based and density model based [XT15].

Partition based methods are oriented to identify disjoint partitions of the elements. $K-$means [Llo82] and $K-$medioids [PJ09] are the most representative methods. Connectivity methods seek to establish relationships of each member with a subset of elements, i.e $K-$ nearest neighbor finds homogeneous partitions of size K for a given dataset [FH89]. Hierarchical clustering is another technique in this group that creates agglomerations with hierarchy of elements that share some similarities [BLG14].

Among the multiples alternatives we seek specific features for the clustering. Since the aim is to identify traffic regimes we are particularly interested in clustering algorithms that are able to provide strict partitioning, it means an element must belong just to a single cluster, this requirement makes obsolete the use of fuzzy or overlapping clustering techniques. It is also desired to support independence between the partitions since later on we aim to extract statistical properties of each group. Strategies like hierarchical clustering may perform better in terms of separation but the existence of parental relationships between sub-groups generates dependencies we try to avoid.

Besides, due to the nature of big data collected historically it is also commendable to

represent with minimum information the available partitions if possible a certain unique characteristic. Finally, we aim to apply the clustering to time series by preserving the identity of each time series, since it is desirable to assign a single element to a specific partition. The aforementioned conditions lead to a reduced group of techniques among them: Gaussian Mixture Expectation Maximization (GMEM), $K-$ means/medioids or more sophisticated techniques as explained in [XT15]. In this work we focus on the K-means algorithm since it is a technique with low time complexity while being highly scalable with large datasets [Jai10]; [AV07].

4.3.3 K-means algorithm

K-means algorithm is a partition type clustering algorithm that creates $\mathcal{K}$ clusters, $\mathcal{K}$ being a positive input parameter. The method introduced by [Llo82] has as a unique goal a space quantization process starting from a density of points defined in $\mathbb{R}^n$. The partitions are obtained through a sequential algorithm that creates centroids which minimize the following criterion,

$$\underset{j}{argmin} \sum_{j=1}^{\mathcal{K}} \sum_{x_i \in j} \|x_i - \mu_j\|_2^2. \tag{4.2}$$

In general the problem of clustering is an *NP-hard* problem and the algorithm may produce local minimum solutions [MNV09]. However, there are efficient heuristic algorithms that are commonly employed and converge quickly to a local optimum. Given a initial set of $\mathcal{K}$ mean points $\mu_1^{(1)}, \ldots, \mu_{\mathcal{K}}^{(1)}$, the algorithm finds the solution by executing two main steps.

1. *Assignment*: Assign each observation to the cluster whose the similarity is closest mean point. In this case, similarity corresponds to the shortest euclidean distance to the mean point. Alternative methods as $K-$medioids consider the shortest median, or another similarity methods as the correlation, cosine distance or ℓ_1 norm.

$$\mathcal{C}_i^{(t+1)} = \{x_p : \|x_p - \mu_i^{(t)}\|^2 \leq \|x_p - \mu_j^{(t)}\|^2 \ \forall \ j, 1 \leq j \leq \mathcal{K}\} \tag{4.3}$$

2. *Update*: Compute new mean points to be the new centroids of the observations in the new clusters.

$$\mu_i^{(t+1)} = \frac{1}{|\mathcal{C}_i^{(t+1)}|} \sum_{x_p \in \mathcal{C}_i^{(t+1)}} x_p \tag{4.4}$$

The convergence of the algorithm is finished either by a finite number of iterations or when the centroid achieves a minimization for (4.2). A typical stop criteria for the algorithm is when $\mu_i^{(t)} - \mu_i^{(t+1)} \approx 0$ meaning no substantial updates on the centroid are introduced. The generic clustering process is illustrated in Fig. 4.6. The assignment step is achieved in steps $2), 4)$ for two separate iterations while the update step is illustrated in 3). In the context of traffic application multiple approaches have been proposed to apply clustering techniques. [Liu+14] identifies regimes of traffic by considering the elements x_p as data points in a time series along a full day. [Oje14] follows a similar approach but in this case the elements x_p are divided

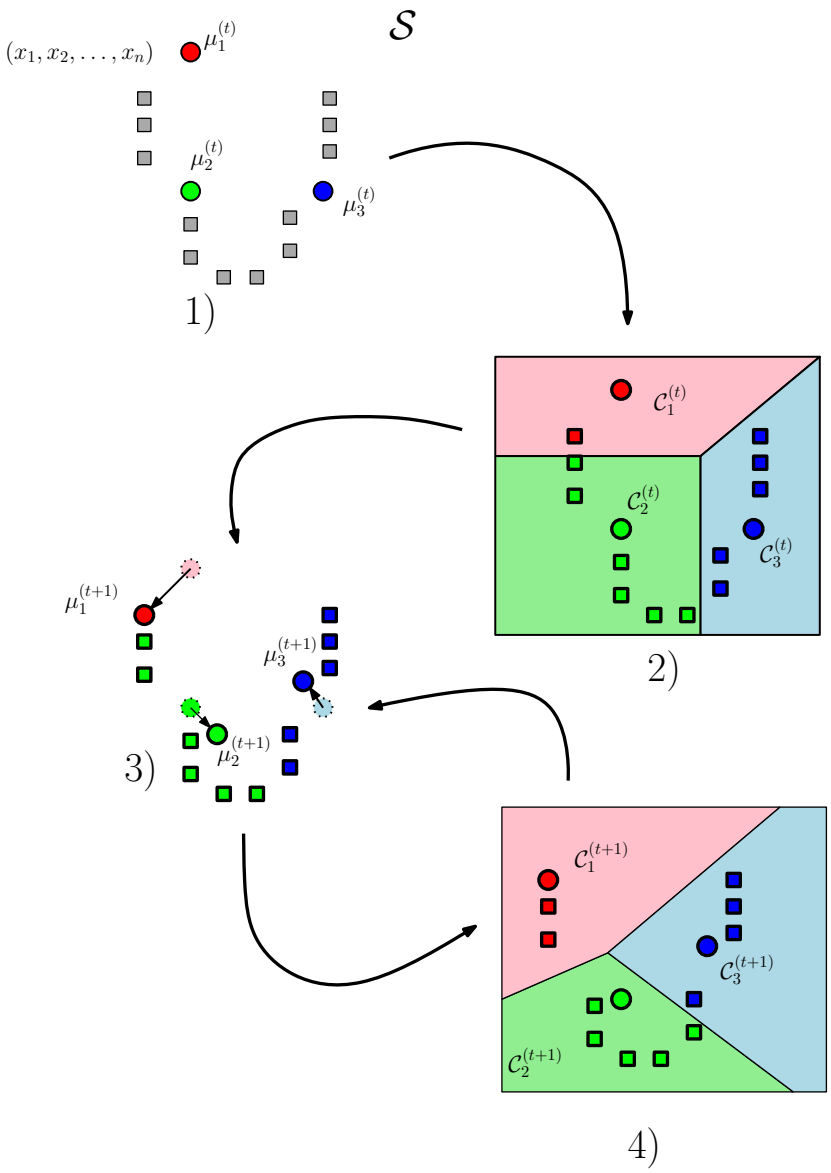


Figure 4.6: General stages in K-mean clustering: Step 1) first step corresponds to the initialization. Step 2) corresponds to the elements assignment to a particular cluster $\mathcal{C}_i$ according to a shortest euclidean distance criteria. Steps 3) and 4) correspond to the update of the partitions and the reassignment task based on updated information.

into fixed partitions during the day each one setting an offline independent clustering process. More recently, [LHC16] proposes a hierarchical approach in which assignments are done based on data points at a single instant of time, thus letting apart the dynamic component of the signal. In this case we consider each element of the set as a time series on the dimension $\mathbb{R}^{\Delta}$. For the forecasting application the sliding window is centered with respect to the prediction time instant (See. Fig. 4.7). The goal is to capture past behaviors and future behaviors in the same sliding window. Past information will be used as known a priori information in the forecast strategy to better select the future. (See Chapter 5 for further details).

In our case, given a set of time series with TT given by $\mathcal{S} = \{DTT_1, \ldots, DTT_n\}$ for n different days, the objective is to assign each DTT_i to a cluster $\mathcal{C}_j$ such that $\mathcal{C}_i \cap \mathcal{C}_j = \emptyset$, $\mathcal{C}_i \neq \emptyset$ and $\cup_{j=1}^{\mathcal{K}} \mathcal{C}_j = \mathcal{S}$. The notion of distance between elements of the set $\mathcal{S}$ is characterized in this case with the euclidean norm. Here μ_j represents the centroid of each one of the clusters that are being updated sequentially, this process is illustrated in Fig. 4.7.

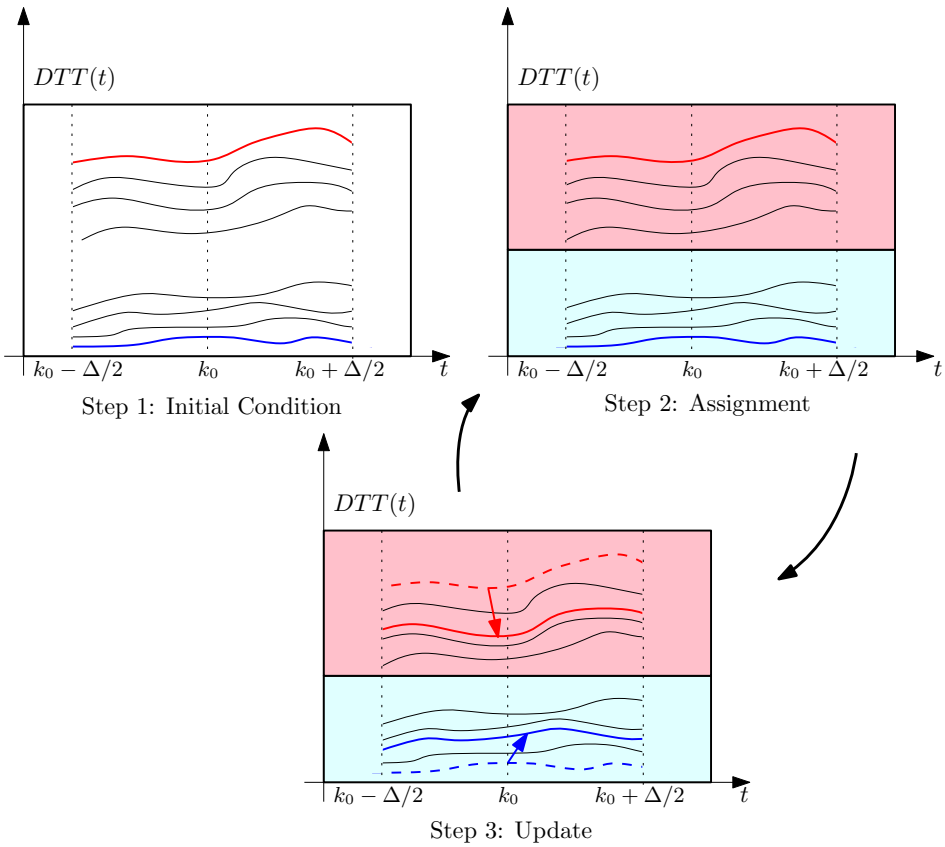


Figure 4.7: Diferent Separation stages in K-means in time series: Step 1 - Given a number of clusters, the corresponding centroids are selected randomly among the elements of $\mathcal{S}$. Step 2 - then the remaining elements are assigned to the clusters according to their distance to the centroids. Step 3 - the process is iterated until a stable partitioning is obtained.

4.3.4 Parameter setup

When performing K-means, a two main input components are required to perform the task. First is a well defined dataset which was defined in the previous section and second the number of partitions required to perform the task. Due to the sub-optimal nature of the solution of (4.2) it is important to run diagnostic checks for determining the number of clusters in the dataset and a mechanism to establish reproducible results in some extent.

The *initial condition* as well as the *number of clusters* represent important parameters, so we explain the process of selection of each one:

1. *Definition of number of clusters*: There exists multiple techniques to select $\mathcal{K}$. We consider the one based on the work developed in [PDN05]. In this case the optimal number of clusters for a data set is given by:

$$\mathcal{K}^* = \underset{\mathcal{K}\in\{2,3,\ldots,\mathcal{K}_{max}\}}{argmin} \quad f(\mathcal{K}). \tag{4.5}$$

The aim of the problem is to minimize a determined function f which will depend on the *total distortion* for a fixed number of partitions. The idea behind is that if the measured distortion decreases for $\mathcal{K}$ clusters the optimal value is reached. For a fixed value $\mathcal{K}$ the *total distortion* is defined as

$$D_{\mathcal{K}} = \sum_{j=1}^{\mathcal{K}} \sum_{DTT_i\in\mathcal{C}_j} \|DTT_i - \mu_j\|_2. \tag{4.6}$$

(4.6) represents the dispersion within the cluster by adding errors between the elements and the centroids. In this case, the euclidean norm is applied since distortion is aimed to be explained with the same similarity measurement as in the clustering algorithm. The role of $f(\mathcal{K})$ is to reveal trends in the data distribution and therefore it is important to keep it independent of the number of objects. In addition $f(\mathcal{K})$ is an indicator of the total distortion $D_{\mathcal{K}}$ of different values of $\mathcal{K}$. In this sense, from [PDN05]:

$$f(\mathcal{K}) = \begin{cases} 1 & \text{if} \quad \mathcal{K} = 1 \\ \frac{D_{\mathcal{K}}}{\alpha_{\mathcal{K}} D_{\mathcal{K}-1}} & \text{if} \quad D_{\mathcal{K}-1} \neq 0 \quad \forall \quad \mathcal{K} > 1 \\ 1 & \text{if} \quad D_{\mathcal{K}-1} = 0 \quad \forall \quad \mathcal{K} > 1 \end{cases} \tag{4.7a}$$

$$\text{where} \qquad \alpha_{\mathcal{K}} = \begin{cases} 1 - \frac{3}{4N_d} & \text{if} \quad \mathcal{K} = 2, N_d > 1 \\ \alpha_{\mathcal{K}-1} + (1 - \alpha_{\mathcal{K}-1})/6 & \text{if} \quad \mathcal{K} > 2, N_d > 1. \end{cases} \tag{4.7b}$$

In (4.7a) the term $\alpha_{\mathcal{K}} D_{\mathcal{K}-1} = \hat{D}_{\mathcal{K}}$ is an estimation of $D_{\mathcal{K}}$, so once $D_{\mathcal{K}}$ decreases with respect to an estimated reference value it means the cluster reached less total distortion. The term $\alpha_{\mathcal{K}}$ is derived under the basis that the set of time series of length N_d is uniformly distributed (see [PDN04]). The principle lying behind this idea is: when data distribution is uniform, there is not any irregularity. Therefore, data sets with uniform distribution could be used to calibrate and verify the clustering result.

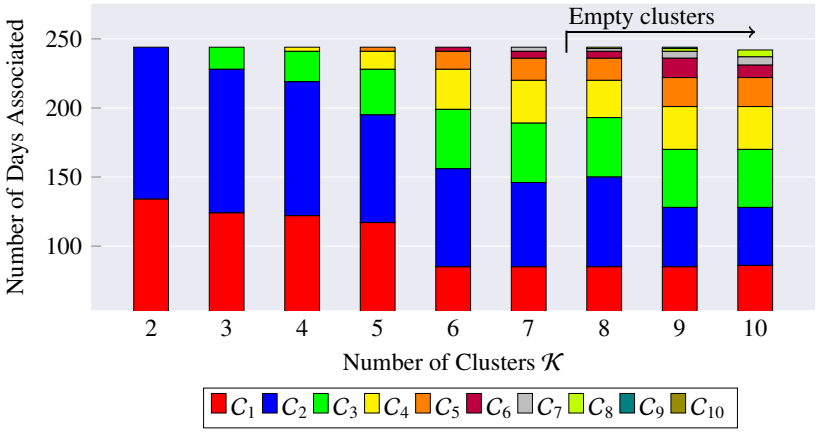


Figure 4.8: Number of days assigned to cluster $\mathcal{C}_i$ at 17h00.

Now let us find the optimal number of partitions by executing a test of clustering for a given historical dataset. For this purpose, we consider a set of 244 days of data, containing the DTT from Meylan to Rondeau [CDW+15]. A sliding window from 07 : 00AM to 07 : 00PM selects the interval of time in historical data to be clustered, the length of the sliding window is 90min. We proceed in steps of 1min. Each time data is clustered up to 10 clusters as depicted in Fig. 4.8. We can note that empty clusters start to appear from $\mathcal{K} > 7$. Therefore to find the optimal number $\mathcal{K}^*$ we solve (4.5) by setting $\mathcal{K}_{max} = 7$. Now, let check the correlation between the optimal number of clusters and the traffic regime given by the DTT. Fig. 4.9 depicts the distribution (median value and dispersion) of optimal number of clusters $\mathcal{K}$ with respect to the average DTT. We can highlight the increasing number of partitions for congested periods, however no predominant values can be preselected without incurring in errors. For the rest of the analysis, $\mathcal{K}^*$ varies according to the solution of (4.5) for a specific time of the day. It is important to remark that in the real time tool [CDW+15] the clustering is also performed in a dynamic way each time a prediction is launched (See section 4.3.1).

2. *Initialization of centroids*: The nature of the $\mathcal{K}$-means algorithm needs a random initial condition. In order to avoid empty clusters and to achieve repetition on the clustering process, we follow the approach given in [AV07]. The initial seed for the first cluster is chosen randomly according to a uniform distribution, the following ones are chosen with a probability proportional to the distance between the new point and the previous selected centroid. This modified algorithm of $\mathcal{K}$-means is shown to have faster convergence. In order to achieve similarity in the results, the clustering process is replicated a fixed amount of times. The iteration taken into account is the one that minimizes the total distortion (4.6). Although the nature of the algorithm provides local mínima, experimentally it was found that 10 replicates of the clustering are good enough to find same cluster distributions.

In order to visualize one example of this dynamic behavior Fig. 4.10 depicts the cluster-

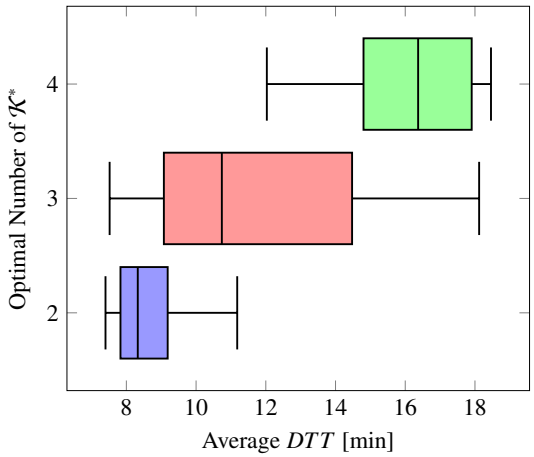


Figure 4.9: Box plot between the optimal number of clusters $\mathcal{K}^*$ and the historical average DTT (Meylan - Rondeau). The distribution of average DTT shows a direct correlation between the DTT and $\mathcal{K}^*$. However, it is important to remark a high variance of the average DTT in relation with a particular value of $\mathcal{K}^*$. Due to this effect, the value $\mathcal{K}^*$ is recovered from the solution of (4.5).

ing obtained for the time window 6:15AM to 7:45AM. The three centroids are depicted in color lines and the area around them corresponds to the subspace occupied by the elements associated to the corresponding cluster.

Here, the objective of capturing and putting together patterns with similar structure is reached: we can notice that blue cluster characterized by μ_1 recovers elements in free flow conditions where patterns are grouped together. Patterns that show build-up congestions are also separated in two different partitions. The main difference between them are the level of congestion as well as congestion peak times.

4.4 Final comments on the chapter

The current chapter has presented the problem of data clustering applied in traffic for historical data classification. It is shown experimentally that the K-means technique can lead to identification of traffic regimes in historical patterns which are a key tool in order to perform short-term forecasts.

Given the autocorrelation function and possible miss assignments of signals to a particular cluster we promote a dynamic clustering along the day. The cost of this dynamic clustering augments computational cost for the benefit of better assignment and as a consequence accuracy in the forecasting process. Nonetheless, the initialization of the clustering process still benefits the convergence of the K-means algorithm and makes the tool suitable for an real time data process.

The selected technique requires input parameters in order to be executed. A methodology

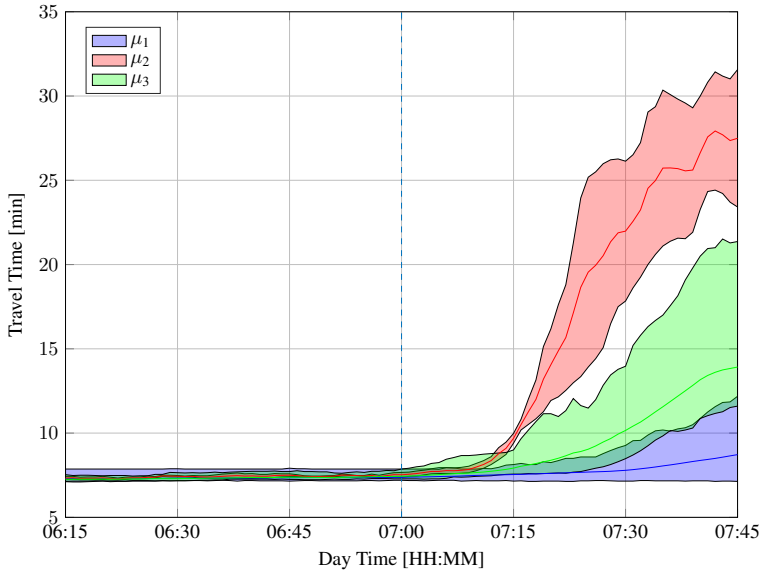


Figure 4.10: Example of clustered historical DTT data for a time window of 90min centered around $k_0 = 07:00\text{AM}$.

of an optimal number of clusters $\mathcal{K}$ based on previous works [PDN05] is evaluated in the traffic case for this purpose. The strategy conducts to the usage of more partitions during congested zones. This result conducts to more shrinked clusters by diminishing the total distortion.

The following chapter will integrate the discussed tool into short-term forecast strategies. The clustered information will be used as an indicator for construct candidate profiles of the future of the signal that will be merged with the current measured information.

Chapter 5

Short-term travel time forecasting

Contents

5.1 Overview

Traffic forecasting is one of the most desired tools for traffic management requested by operators and commuters. Nowadays, the multiple sources contributing data to traffic information systems rely on different sensor networks. In the era of data deluge in which we are, these measurements collected by different kind of sensors such as MLD or GPS build up unique sources of information that require analysis, classification, and processing in order to detect patterns and behaviors that can be exploited for traffic prediction [LCK15]; [VKG14]. Several indicators like travel time, queue length, density, delay are used as performance indexes to determine the status of a traffic network [WB05].

We concentrate our efforts on DTT and we consider it as a state of the network. This work is motivated by the forecasting problem originally proposed in [OKC13a]; [Oje14] based on the existence of traffic regimes identified from a clustering process. The main idea to reconstruct the forecast relies on the existence of DTT patterns identified along the history as described in [WB05] and [Chr+04] when large historical datasets are observed. Precisely, the chapter focuses on the development of two short-term forecasting strategies based on data driven models. These strategies are suitable to perform multi-step ahead forecasts. We then

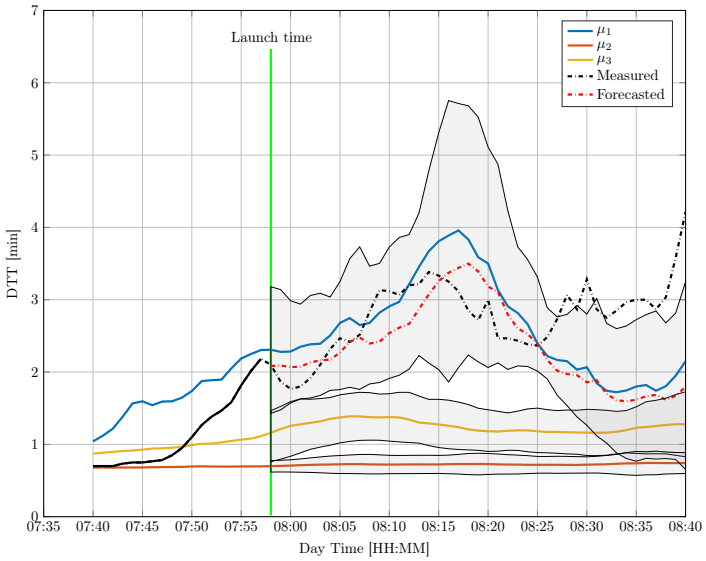


Figure 5.1: Short term forecasting problem. The actual measured signal is depicted in dashed black lines. The forecast is depicted in red dashed lines

evaluate these strategies based on a dataset of TT taken from the GTL. The methods consider evaluations on DTT model as described in Chapter 3.

This chapter is organized as follows. First, we describe the forecasting problem, the inputs of this problem and the objective of the short-term forecasting algorithm. We introduce then the historical data structure and we link the clustering process explained in Chapter 4 as an input for the current algorithm. Furthermore, details on assumptions and statistical measurements taken on the datasets for the forecasting problem are provided. Second, we derive a complete structure for a forecasting algorithm and we detail two fusion mechanisms that provide two alternative algorithms. Finally, we address an experimental validation and we assess performance of the proposed methods using traffic data from the South Ring of the Grenoble city in France.

5.2 Forecasting problem

The forecasting problem consists in the process of making predictions of the future based on past and present data and most commonly by analysis of trends. The problem can be classified in the family of estimation problems where future information needs to be constructed based on a mathematical model or information read in data [HA14b]. Let us consider for example Fig. 5.1, the picture depicts an example of a forecasted signal. The black solid line represents the current measure of travel time. The green vertical line is an indicator of the moment in which the forecast is launched, as a result of this process the red dashed signal is produced, later this estimation can be evaluated with respect to the original value once the real travel time depicted in black-dashed is measured. The rest of the information in this plot corresponds

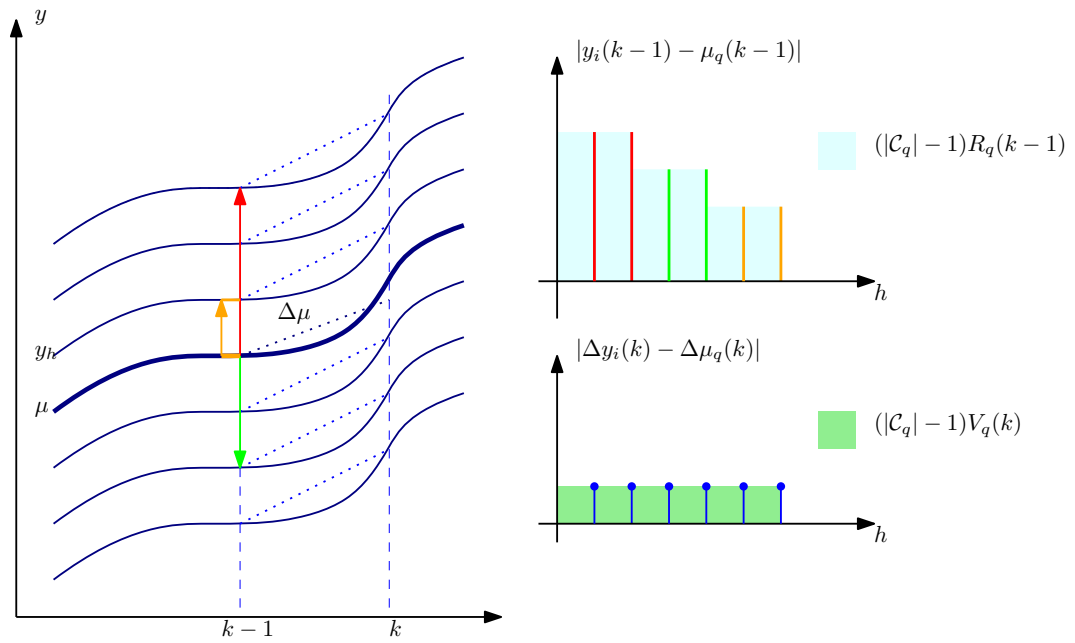


Figure 5.2: Sample based variance estimator in pseudo observations. The information related to the variance of the cluster is illustrated in the left side. At time k the difference between each day i and μ_q is computed. The area lying behind the curve is a non normalized representation of R_q. An analog process is achieved with V_q

to supplementary information we use to construct the pseudo observation model detailed in section 5.3. The solid lines in blue, yellow and red represent the cluster centroids while the gray zones around them correspond to the variance of these centroids.

The forecast problem can be stated as follows. Let $y_h(k)$ be the variable of interest to be forecasted at time k and day h. We aim to predict $\hat{y}_d(k)$ for the current day d and $k > k_0$, k_0 the current time, using the available measurements given $y_d(k_0)$ and $y_h(k)$, $h \neq d$, $k \in I = [k_0 - \Delta_p + 1, k_0 + \Delta_f]$. Here, Δ_p and Δ_f denote the past and the future horizons, respectively and $y_d(k_0)$ is the last available value of y based on the current measurements.

In order to perform this task we make use of clustered data as described in Chapter 4 that describes information in the future time horizon of interest. This information should be combined with the current day information, we rely on the Kalman Filter (KF) theory for this mixture process. This approach was selected for the following reasons: it is very useful in real time applications and more important, it can be adapted to fuse measurements and it provides an alternative to combine same day and historical data.

5.3 Observation model

In order for the KF to provide updated state estimates at every time step along the considered time horizon, three key pieces of information are needed: noisy observations of the real system,

statistics of the observations noise, and statistics of the process model noise. As stated in this dissertation, historical data are organized in clusters. Each cluster $\mathcal{C}_q$ is defined by its centroid $\mu_q(k)$ and the corresponding forward finite difference $\Delta\mu_q(k) = \mu_q(k+1) - \mu_q(k)$. In other words, a cluster is assumed to exhibit a similarity on both the level and the trend. Therefore, claiming that y for the current day d belongs to cluster $\mathcal{C}_q$ yields to:

$$y_d(k) = \mu_q(k) + w_{d,q}(k) \tag{5.1}$$
$$\Delta y_d(k) = \Delta\mu_q(k) + v_{d,q}(k), \tag{5.2}$$

for any $k \in I$, where the evolution of $y_d(k)$ is considered to be an stochastic process in which

$$E[w_{d,q}(k)] = 0 \text{ and } E[w_{d,q}^2(k)] = R_q(k)$$
$$E[v_{d,q}(k)] = 0 \text{ and } E[v_{d,q}^2(k)] = V_q(k).$$

Moreover, this noise will be assumed to be given by the dispersion present in clustered historical data, as illustrated in Figs. 5.1, 5.2. Here, at each time instant k, the variance of the realizations $y_h(k)$ and $\Delta y_h(k)$, historical values and historical increments respectively are computed based on data. The variances $V_q(k)$ and $R_q(k)$ can be computed from the cluster as follows:

$$R_q(k) = \frac{1}{|\mathcal{C}_q| - 1} \sum_{h \in \mathcal{C}_q} (y_h(k) - \mu_q(k))^2,$$

$$V_q(k) = \frac{1}{|\mathcal{C}_q| - 1} \sum_{h \in \mathcal{C}_q} (\Delta y_h(k) - \Delta\mu_q(k))^2,$$

where $|\mathcal{C}_q|$ stands for the cardinality of the cluster.

5.4 Prediction algorithm

Two situations may be considered in the prediction scenario. The first one is, given the advantage of the current information and potential future information obtained from each cluster, an individual prediction can be obtained based on single cluster information like the trend of the centroid and the value of y for this particular cluster in the future. The effect of this strategy has been previously studied in [OKC13b]. However, the hard selection of one single prediction is a risky choice due to the limited information of y up to the current time and the limited knowledge of the future. A second situation emerges when considering the fusion of multiple individual cluster based predictions, this possibility opens the idea to reduce the error for predictions where the cluster assignment in the interval $I_{\Delta_p} = [k_0 - \Delta_p, k_0]$ provides close similarity to multiple clusters. Therefore it is needed to define a proper similarity in I_{Δ_p} and to use this information for mixing mono-cluster based predictions. We explore in the following subsections the derivation of individual predictions and then the strategy used to obtain the final prediction.

5.4.1 Mono-cluster based prediction

In a similar approach based on data driven models [OKC13a] compared two models, the one proposed in [OS84] and a random walk model, concluding that the latter presented better results in terms of flow forecasting accuracy, for their specific scenario. Motivated by these findings and its straightforward structure, the dynamic evolution is modeled through a random walk model ARIMA$(0, 1, 0)$.

From similarity on the trend in (5.2), given a prediction $\hat{y}_d(k|k)$ with an error variance $\hat{P}_d(k|k)$, the one step-ahead prediction is given by

$$y_{d,q}(k+1|k) = \hat{y}_d(k|k) + \Delta\mu_q(k). \tag{5.3}$$

with the error variance:

$$P_{d,q}(k+1|k) = \hat{P}_d(k|k) + V_q(k). \tag{5.4}$$

On the other hand, similarity the level gives the following predictor:

$$\bar{y}_{d,q}(k+1) = \mu_q(k+1) \tag{5.5}$$

with error variance $R_q(k+1)$.

The trend-based predictor is a mixture of historical data, by means of the cluster centroid's trend, and current data (initial value of the predictor) while the level-based predictor is only based on the historical level. The objective is then to find a linear combination of these two predictors in order to minimize the variance of the prediction error:

$$\begin{aligned} &\min_{K_q(k+1)} E[(y_d(k+1) - \hat{y}_{d,q}(k+1|k+1))^2] \\ \text{s.t.} \quad &\hat{y}_{d,q}(k+1|k+1) = (1 - K_q(k+1))y_{d,q}(k+1|k) + K_q(k+1)\bar{y}_{d,q}(k+1) \end{aligned} \tag{5.6}$$

The solution of this optimization problem is given by [Kal60]:

$$K_q(k+1) = \frac{1}{1 + \frac{R_q(k+1)}{\hat{P}_{d,q}(k|k)+V_q(k)}} \tag{5.7}$$

One can note that if the dispersion around the centroid is small enough ($R_q(k+1)$ very small) the level-based predictor is preponderant. The reverse is true if the trend coherence is stronger (smaller value of $V_q(k)$). In addition, the predictor

$$\hat{y}_{d,q}(k+1|k+1) = (1 - K_q(k+1))y_{d,q}(k+1|k) + K_q(k+1)\bar{y}_{d,q}(k+1) \tag{5.8}$$

with a gain given by (5.7) has error variance

$$\hat{P}_{d,q}(k+1|k+1) = \frac{R_q(k+1)P_{d,q}(k+1|k)}{R_q(k+1) + P_{d,q}(k+1|k)}. \tag{5.9}$$

Equations (5.3), (5.4), (5.5), (5.7), (5.8) and (5.9)) constitute the so-called Clustered Kalman filter (CKF).

5.4.2 Multi-cluster based prediction

Sometimes the performance of the clustering algorithms leads to partitions that may not characterize a forecast in the future (See Fig. 4.10) where overlapping of clusters can be noticed before 07 : 15AM). For this reason, we present two fusion algorithms to tackle this situation. The first algorithm exploits the existing information in the error covariance matrix and performs a fusion inspired on best linear unbiased principles. The second algorithm relies on a similarity measurement established within the recent past of the variable to be forecast to perform the fusion of multiple forecasts emitted by the CKF.

5.4.2.1 Error Covariance Fusion Method (ECFM)

The main objective by introducing an Error Covariance Fusion Method (ECFM) is justified by the fact that the combined forecast might improve its performance as it is discussed in [ZL11], particularly during peak times. In order to make the fusion of the forecasts provided by the clusters, we take into account the fact that $\tilde{y}_{d,q}(k) = \hat{y}_{d,q}(k|k) - y_d(k)$, the estimation error for each one of the mono-cluster based strategies, has variance $P_{d,q}(k|k)$. We assume that each forecast is a noisy version of the actual value of $y_d(k)$. A block diagram of the algorithm is presented in Fig. 5.3.

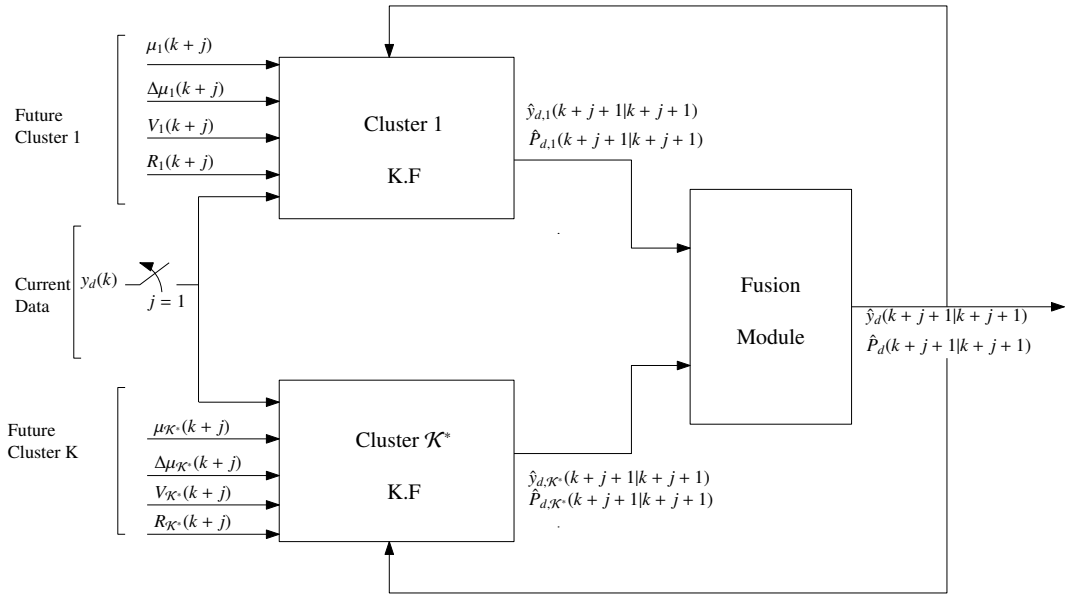


Figure 5.3: Error Covariance Fusion Method (ECFM) Block diagram. At each time instant, individual predictions are generated on the CKF. The fusion is result of an optimization problem inspired on BLUE principles

In what follows, we aim to fuse the forecasts provided by each cluster with a weighting term according to their error covariance matrix. The weighting information will provide weight to those clusters with more confidence for the future horizon. For this purpose, let concatenate all the forecasts in a vector $\theta(k) \in \mathbb{R}^{\mathcal{K}^*}$, $\mathcal{K}^*$ being the optimal number of clusters. We aim to find

$$\hat{y}_d(k) = \underset{y(k)}{argmin} \|\theta(k) - Gy(k)\|^2_{U(k)}, \tag{5.10}$$

with $G = \mathbb{1} \in \mathbb{R}^{\mathcal{K}^*}$ is the column one vector, and $U(k)$ the forecasts error covariance matrix. The main objective of the problem (5.10) is to perform a weighted data fitting problem based on measurements provided by the individual cluster forecasts and weighted by $U(k)$ which is conformed by $\mathcal{K}^* \times \mathcal{K}^*$ matrix denominated $u_{[ij]}(k)$ denoting the error covariance between the estimates from two different clusters i and j. In this case $u_{[ij]}(k)$ is given by the following proposition:

Proposition 5.1
Consider the individual estimates $\theta(k)$ from the mono cluster based predictors and the optimal fusion given by (5.10)*. Then the error covariance matrix of the fusion is given by:*

$$\begin{aligned} u_{[ij]}(k+1) =& \Upsilon_i u_{[ij]}(k) \Upsilon_j + \\ & \delta(i-j)\Big(\Upsilon_i V_j(k) \Upsilon_j + K_i(k+1) R_j(k+1) K_j(k+1)\Big), \end{aligned} \tag{5.11}$$

where $\delta(i-j)$ is the Kronecker delta function and $\Upsilon_i = 1 - K_j(k+1)$.

Proof. Consider the prediction equation (5.8) that computes the update in the filter. Consider also the replacement of the terms $\hat{y}_{d,q}(k+1|k)$ and $\bar{y}_{d,q}(k+1)$ given by (5.3), (5.1) and (5.5)

$$\begin{aligned} \hat{y}_j(k+1|k) =& A_j(k+1)\Big(\hat{y}_d(k|k) + \Delta\mu_j(k)\Big) + \\ & K_j(k+1)\Big(y_d(k+1) + w_{d,j}(k+1)\Big), \end{aligned} \tag{5.12}$$

where $A_j(k+1) = 1 - K_j(k+1)$. Substracting the term $y_{d,j}(k+1)$ and using (5.2) on the rightside part of the equation we obtain.

$$\tilde{y}_{d,j}(k) = A_j(k+1)(\tilde{y}_{d,j}(k) + v_{d,j}(k) + \Delta\mu_j(k)) + K_j(k+1) w_{d,j}(k+1). \tag{5.13}$$

where $\tilde{y}_{d,j}(k) = \hat{y}_{d,j}(k) - y_{d,j}(k)$. We find the covariance between two clusters l, j. Furthermore, we group the vector $X_j = [\tilde{y}_j(k) \;\; v_j(k) \;\; w_j(k+1)]^T$ and consider the covariance between the affine term $C_j X_j + D_j$ where $C_j = \begin{bmatrix} A_j(k+1) & A_j(k+1) & K_j(k+1) \end{bmatrix}^T$, $D_j = A_j(k+1)\Delta\mu_j(k)$ is given by:

$$E[\tilde{y}_{d,l}(k)\tilde{y}_{d,j}(k)] = C_l E[X_l X_j^T] C_j^T. \tag{5.14}$$

In equation (5.14), by previous hypothesis of the KF over independency between observation noise and process noise. (See. (5.3)). The expression leads to:

$$\begin{aligned} u_{[lj]}(k+1) =& A_l(k+1) u_{[lj]}(k) A_j(k+1) + \\ & \delta(l-j) A_l(k+1) E[v_{d,l}(k) v_{d,j}(k)] A_j(k+1) + \\ & \delta(l-j) K_l(k+1) E[w_{d,l}(k) w_{d,j}(k)] K_j(k+1), \end{aligned} \tag{5.15}$$

which can be written as (5.11). □

For details in the case where the state is a vector instead of a scalar, a similar procedure is followed in [Lad+16b]. The optimal solution in this case is given by the best linear unbiased estimator as:

$$\hat{y}^*(k) = \Lambda(k)\theta(k), \tag{5.16}$$

where $\Lambda(k) = (\mathbb{1}^T U(k)\mathbb{1})^{-1}\mathbb{1}^T U(k)$. The solution is then computed iteratevely each sample time with updates in the weighting matrix and the individual forecasts contained in $\theta(k)$. The variance of the estimator given in (5.16) can be computed in this case as:

$$\hat{\Gamma}(k) = \Lambda(k)\mathcal{D}(\hat{P}_d(k|k))\Lambda^T(k), \tag{5.17}$$

where $\mathcal{D}(\hat{P}_d(k|k))$ is the block-diagonal matrix with $\hat{P}_{d,q}(k|k)$ as elements of the diagonal.

5.4.2.2 Past Similarity Fusion Method (PSFM)

For this case we consider the *temporal similarity measurement (TSM)* between the cluster q and the current day d as,

$$S_{d,q}(k) = \sum_{j=k-\Delta_p+1}^{k} \big(y_d(j) - \mu_q(j)\big)^2 e^{-\lambda(k-j)} + \gamma\big(\Delta y_d(j) - \Delta\mu_q(j)\big)^2 e^{-\lambda(k-j)} \tag{5.18}$$

where the exponential term introduces a forgetting factor λ that gives more importance to recent observed measurements. The objective of (5.18) is to provide a measurement of closeness between the current day d and the particular cluster. Furthermore, we aim to find weights at the prediction launch time k_0 that correlate in a proper way the current day data and the clustered historical data. For that we introduce, the *cluster weight similarity* given by

$$\Gamma(k) = \begin{bmatrix} e^{-\zeta S_1(k)} & \dots & 0 \\ 0 & e^{-\zeta S_2(k)} & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & e^{-\zeta S_q(k)} \end{bmatrix}. \tag{5.19}$$

Information given by $\Gamma(k)$ allow us to fuse data based on observed measurements. For this purpose, let us define by $\hat{y}_d(k|k)$ the combined prediction from $\hat{y}_{d,q}(k|k)$, $q = 1, 2, \cdots, \mathcal{K}^*$. Our aim is to compute a weighted linear combination of the predictions provided by each cluster. In this case the full mechanism of fusion is depicted in Fig. 5.4. For the computation of the combined prediction we get:

$$\hat{y}_d(k|k) = \sum_{q=1}^{\mathcal{K}^*} \lambda_q \hat{y}_{d,q}(k|k), \quad \lambda_q = \frac{e^{-\zeta S_q(k_0)}}{\sum_{q=1}^{\mathcal{K}} e^{-\zeta S_q(k_0)}}. \tag{5.20}$$

The overall process is depicted in Fig. 5.4. It is a two-steps process:

- All of the cluster-based predictors are fed locally at time $k + j$ in order to perform cluster-based predictions according to the subsection 5.4.1.
- The cluster-based predictions are fused according to (5.20).

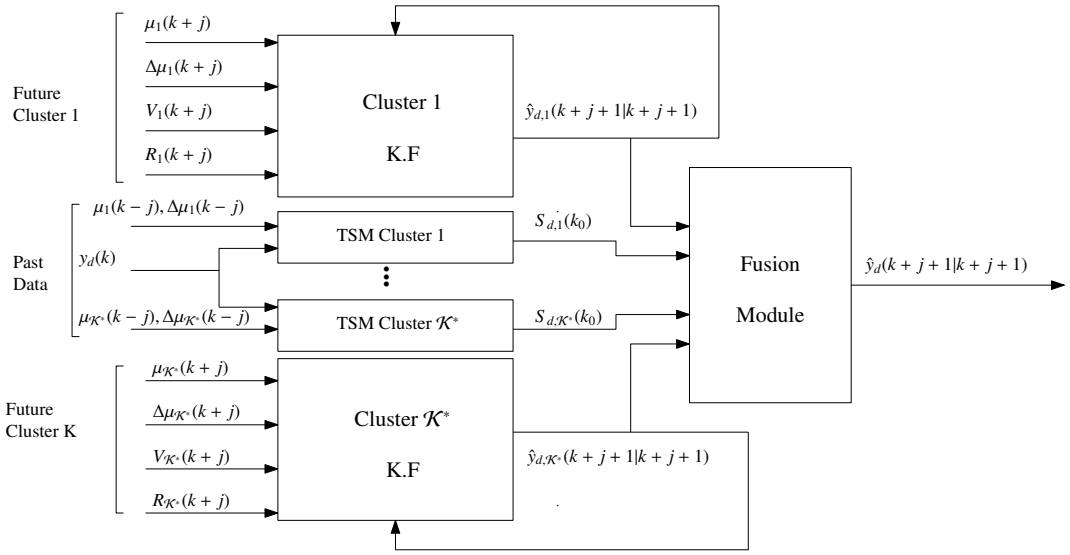


Figure 5.4: Past Similarity Fusion Method (PSFM) Block diagram. In this case, at launch prediction a similarity TSM is computed and used to weight each one of the indivividual forecast obtained from CKFs

5.5 Experimental evaluation

In the previous chapters, we have described the data acquisition and processing procedures allowing to constitute a reliable database that will be used for prediction purpose. In this section, we evaluate the prediction scheme proposed in the previous section.

The experimental data used for the test is obtained from the GTL (see chapter 7). The network consists of 135 magnetometer sensors distributed along 10.5 Km connecting the highway A31 (north-west) to A480 (south) in Grenoble France. The network offers 68 possible locations for collecting speeds in fast and slow lanes. In this experiment, we consider the scenario of a car entering the highway at Meylan and leaving at Rondeau [1]. Predictions are performed between 07 : 00AM and 07 : 00PM time in which it is found peaks of congestion. The system collects information from sensors in real time each time step k, after proceeding with the data imputation and aggregation steps. The aim is to construct the predicted travel time as defined in (3.18). Based on the real measured speeds at each collection point in the interval, the DTT is constructed during the past days and for the current day according to the available measurements. DTT varies from 7min in free flow conditions to 40min in congested cases.

[1]Both places correspond to the beginning and the end of the 10.5 Km of installed sensors

For the tests, we set the parameters λ , γ and ζ to some fixed values. The parameter λ is the amount of past information to be considered for the launched prediction, we fix $\lambda = 0.5$ in order to correlate at most the last 15 minutes of traffic regime which have shown to be highly correlated with current information [OKC13a]. γ is chosen such that the following relation is preserved for each cluster q:

$$\frac{\sum_{j=k-\Delta_p+1}^{t} \left(y_d(j) - \mu_q(j)\right)^2}{\sum_{j=k-\Delta_p+1}^{t} y_d(j)^2} = \gamma \frac{\sum_{j=k-\Delta_p+1}^{t} \left(\Delta y_d(j) - \Delta \mu_q(j)\right)^2}{\sum_{j=k-\Delta_p+1}^{t} \Delta y_d(j)^2}$$

The idea with the term is to normalize the amount of energy of the error between the day and the cluster with the same amount of energy given by its corresponding derivatives, this aims to consider with same importance the information given by the trend and the one given by the level. Finally we set the value of $\zeta = 0.5$. This value explains the degree of selection of the term $S_{d,q}(k_0)$, we measured the variance of multiple clusters and by examining the dispersion we fix the value such that the term $e^{-\zeta S_q(k_0)} \approx 0$ for variances larger than 4min around the centroid of the corresponding cluster.

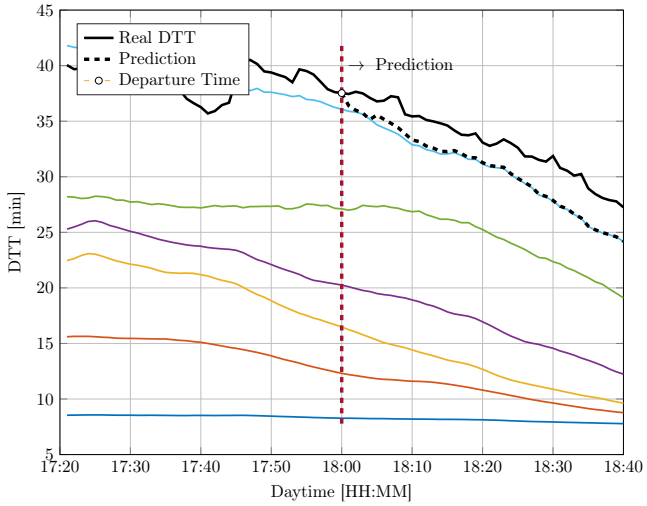
5.5.1 Individual Predictions Assessment

5.5.1.1 ECFM Approach

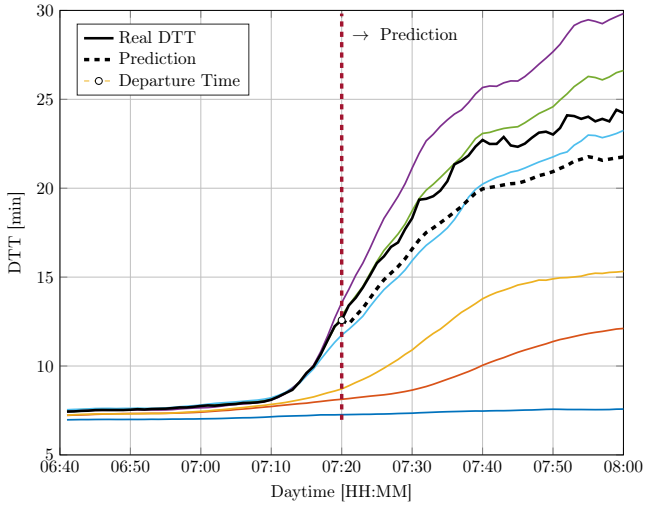
We assess individual predictions for a fixed amount of clusters $\mathcal{K}^*$ as shown in Fig. 5.5. Here we aim to show behaviors at specific moments of the day to verify the performance of the ECFM prediction algorithm. The performance of the method is verified at several times and the results are shown in 5.5a,5.5b and 5.5c. Each one of the scenarios depict one of the results of increasing appearing congestion, disappearing congestion and cases where multiple regimes appear in the historical data. It can be seen in particular in 5.5a that the prediction tends fast align with the cluster. This results illustrates that the method tends to incorporate pseudo observations into the estimation as part of the future prediction. 5.5b, shows an appearing congestion in which the forecast error tends to be small in the short term and increasing in long term, this due to the error propagation through the time. Finally 5.5c illustrates the case in which multiple regimes may exist. Although the behavior presents small errors in short-term, the influence of other clusters influence the error also in long term predictions. This fact suggest that dynamic adaptation of the forecast time should be handle carefully since it may introduce bias in the estimation due to trend in the predicted value. The evaluation of the PSFM technique where only past information in ΔI_p is considered to establish the weights helps to clear out this issue.

5.5.1.2 PSFM Approach

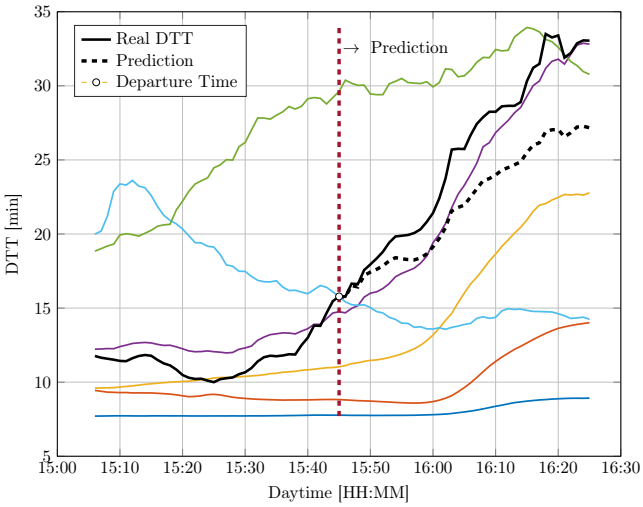
In this case we check in detail the behavior for the PSFM approach in Fig. 5.6. Each one of the figures 5.6a,5.6c,5.6e depicts the prediction given by PSFM approach, the data collected during the corresponding day and the centroids of the clusters considered for each case. The



(a) Prediction Day - 22nd April,2016



(b) Prediction Day - 17th February,2016



(c) Prediction Day - 10th October,2015

Figure 5.5: ECFM Individual predictions in several days at multiple times of the day. In color we visualize the centroids of the cluster.

performance of the method is visually verified in Fig. 5.6a where there is a clear distinction between the different regimes and hence the prediction provides good results, a proof of this fact is also checked in the weight assigned in Fig. 5.6b. As a second case, Fig. 5.6c shows a particular initial morning congestion that regularly starts around 07 : 20AM. As seen in Fig. 5.6c the increasing behavior of the congestion is captured by the TSM and hence by the forecast algorithm. The weights λ_q in this case select the three closest centroids indexed as $(2, 4, 6)$ as seen in Fig. 5.6d. Finally 5.6e shows a particular mid afternoon congestion. It is important to remark that in this case the similarity term that provides the good selection of the cluster is strongly supported by the trend of the cluster.

Even though the amount of error is high in this last case the behavior of this prediction can be improved along the day. Consider for instance the same prediction as in Fig. 5.6e launched at different times (See Figures 5.7a to 5.7f). In this case, as the prediction is computed at different hours it is noticeable the reduction of error due to a higher similarity with respect to the purple cluster as time advances during the day. A more precise comparison is given in terms of the Absolute Percentage Error (5.21) in Fig. 5.7f.

5.5.2 Crossvalidation test

For a quantitative test, we constrain the set of historical data to 244 days collected from September 2015 to May 2016 [PL16], [Lad+16a]. We will evaluate the prediction scheme for different departure times in two periods: morning from 07 : 00AM to 10 : 00AM and afternoon from 04 : 00PM to 07 : 00PM. In what follows, we consider k_0 as the current time and $k_0+\Delta T$ as the desired departure time in future with $\Delta T \in \{5, 10, 15, 20, 25\}$.

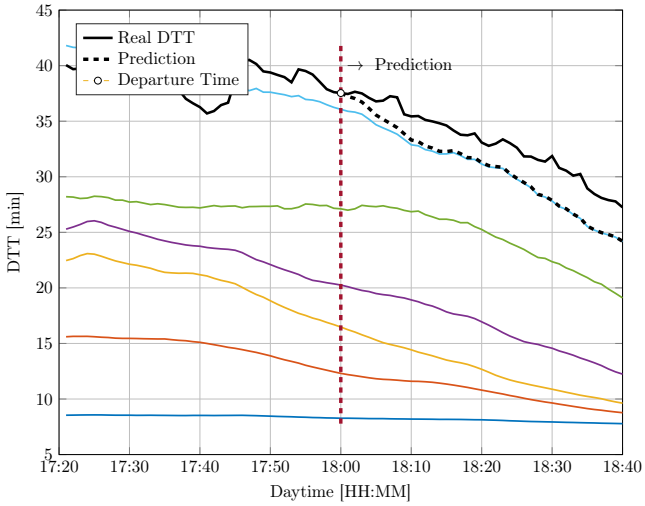
According to the results of Section 4.3, for all the tests, the number of cluster is adapted according to (4.5). In order to assess the performance of the proposed method, we resort to the APE as performance index. For a particular departure at some desired future horizon $1 \leq \Delta_f$ the APE is given by:

$$APE(k_0 + h, d) = 100\frac{|\widehat{y}_d(k_0 + h) - y_d(k_0 + h)|}{y_d(k_0 + h)}, \tag{5.21}$$

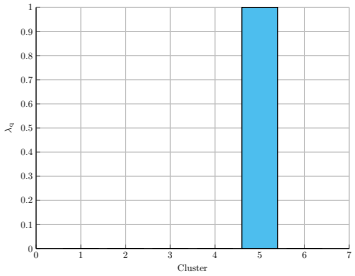
In order to study the behaviour of the strategy, the fixed amount of historical data will be selected, then a leave one out cross validation test is performed over the data. The element out of the set is the candidate day d for performing the prediction. We perform the prediction at some specific time k_0 of the day, the APE index in (5.21) is computed for each one of the forecasts, then the process is repeated along the day by advancing in steps of 1min. The results given hereafter are averaged over k_0 and d.

5.5.2.1 Evaluation of the mono-cluster prediction scheme

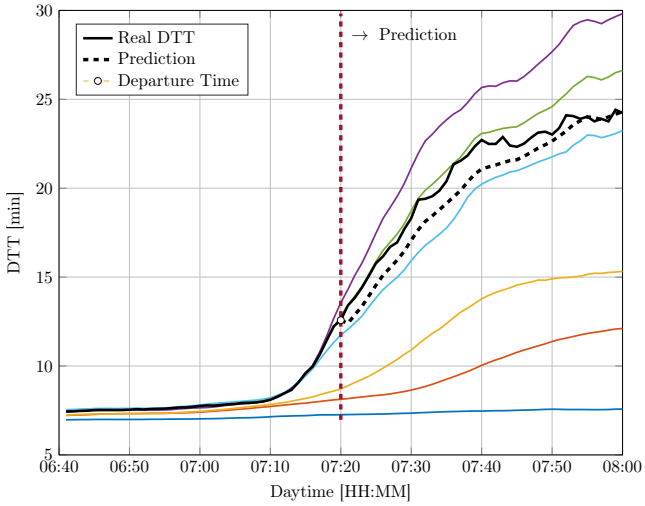
In this subsection, we first evaluate the prediction scheme derived when the cluster to which the time series of interest belongs is well known.



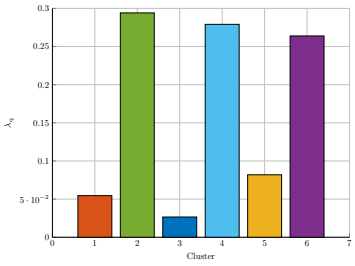
(a) Prediction Day - 22nd April,2016



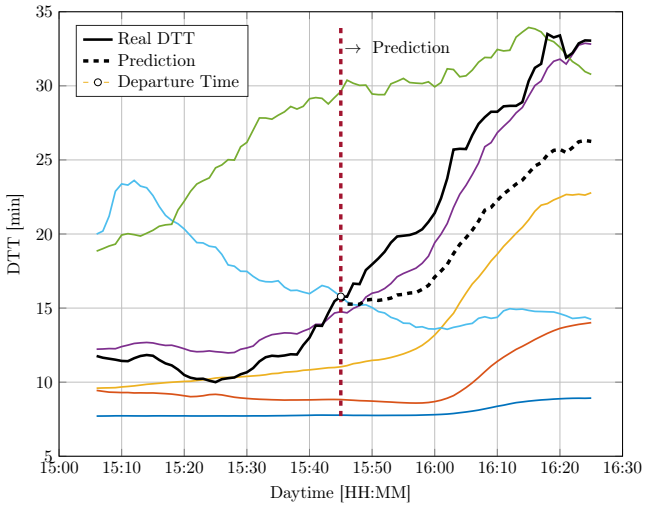
(b) λ_q - 22nd April,2016



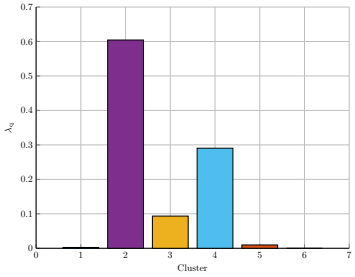
(c) Prediction Day - 17th February,2016



(d) λ_q - 17th February,2016

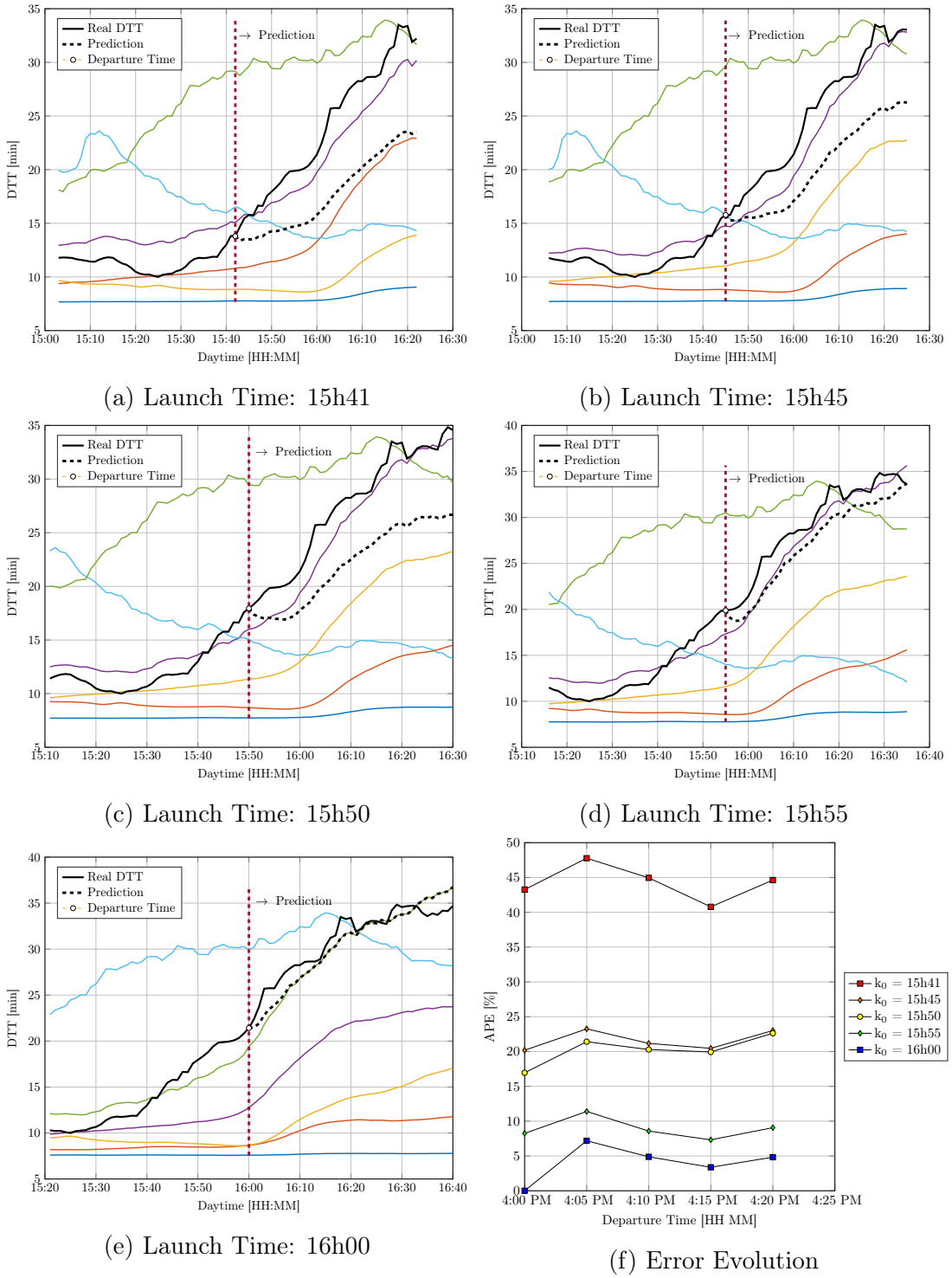


(e) Prediction Day - 10th October,2015



(f) λ_q - 10th October,2015

Figure 5.6: Individual predictions on the PSFM in several days at multiple times of the day. In color we visualize the centroids of the cluster. The assigned weights corresponds in color to the cluster color depicted in the left side figures



(a) Launch Time: 15h41

(b) Launch Time: 15h45

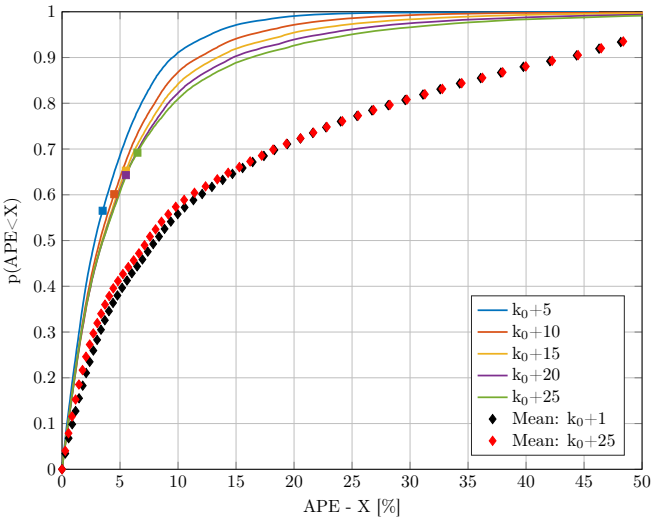
(c) Launch Time: 15h50

(d) Launch Time: 15h55

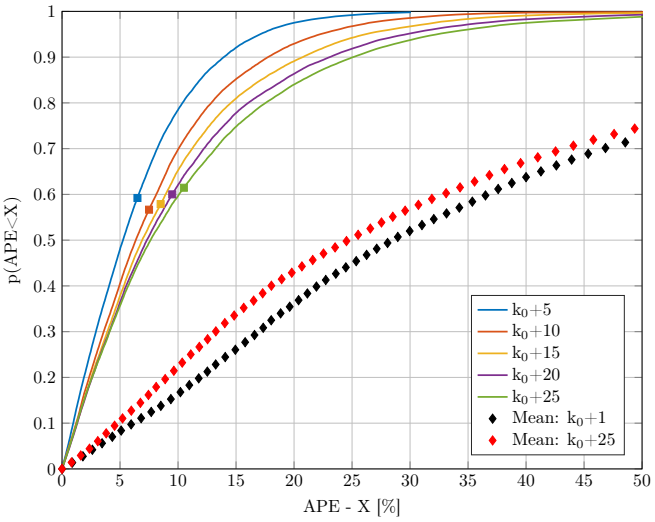
(e) Launch Time: 16h00

(f) Error Evolution

Figure 5.7: Analysis of same day predictions at different times during the day



(a) APE examined between 7AM and 10AM



(b) APE examined between 4PM and 7PM

Figure 5.8: ECDF plot of the prediction performance for the mono-cluster case, while assuming that the cluster is well known

	$p < X$	$\Delta = 5$	$\Delta = 10$	$\Delta = 15$	$\Delta = 20$	$\Delta = 25$
Morning	90%	9.74%	11.73%	13.45%	14.74%	16.19%
	80%	6.84%	7.96%	8.62%	9.159%	9.62%
Afternoon	90%	13.84%	17.70%	20.54%	22.94%	24.98%
	80%	10.24%	12.86%	14.54%	15.98%	17.48%

Table 5.1: APE guaranteed error based for multiple horizons. Mono-cluster case

	$p < X$	$\Delta = 5$	$\Delta = 10$	$\Delta = 15$	$\Delta = 20$	$\Delta = 25$
Morning	90%	7.44%	11.20%	17.80%	27.06%	34.38%
	80%	5.55%	7.23%	9.21%	11.87%	14.42%
Afternoon	90%	11.02%	19.80%	26.24%	31.19%	35.75%
	80%	7.60%	13.83%	19.22%	23.55%	26.88%

Table 5.2: APE guaranteed error based for multiple horizons. ECFM case

In Fig. 5.8, the empirical cumulative distribution function (ECDF) of the APE is depicted for different horizons and different periods of time during the day. We compare the performance with that of methods considering historical mean as prediction. For historical mean, whatever the horizons the ECDF is similar. For the proposed method, we can note that the performance decreases when the horizon of prediction increases. The gap of performance observed between horizon 5min and horizon 25min is slightly bigger in the afternoon than in the morning. Whatever the horizon, the proposed scheme outperforms the historical mean.

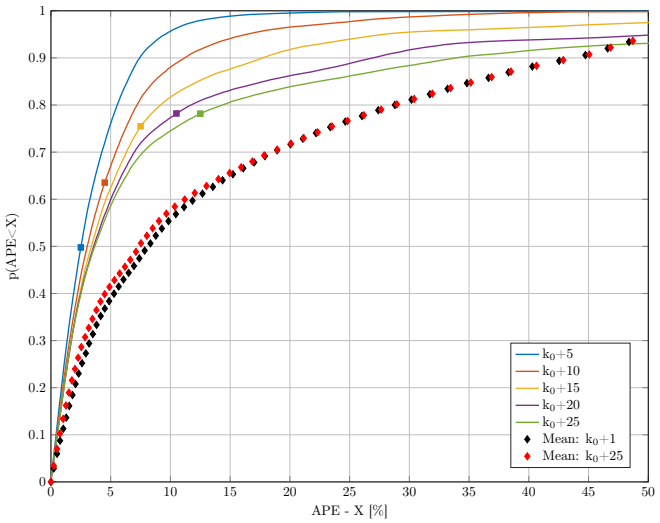
80% of time, the proposed method allows to get predictions error less than 7% for horizon 5min and less than 10% for horizon 25min in the morning case. As for the afternoon case 80% of the times the error is less than 11% for horizon 5min and less than 18% for horizon 25min. (See. table 5.1). In terms of time, this means for horizon 5min the error is guaranteed to be less than 42sec in free flow [2] and less than 4min24sec during congestion. For horizon 25min, we get an error less than 1min15s in free flow and less than 7min in congestion. We can notice than the proposed method, despite its simplicity, is efficient enough. However, we have assumed the cluster is perfectly known. Such assumption is hard to fulfill. So in next section, we evaluate the multi-cluster strategy with fusion policy derived in previous section.

5.5.2.2 ECFM fusion strategy

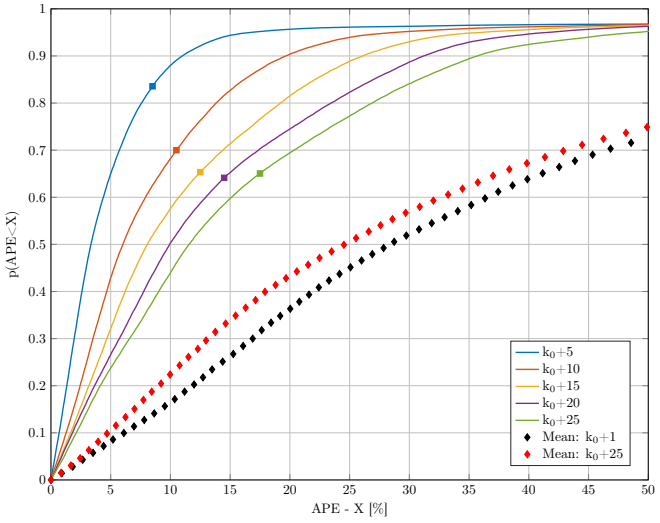
We assess the scenario of the ECDF based on the behavior of the error. Fig 5.9 shows de behavior of the ECDF for the ECFM scenario. In this case we consider no knowledge of the association of the time series and the cluster. In this case it is noticeable, a degradation in the performance with respect to the mono cluster case.

Table 5.2 presents two possible scenarios of evaluation. In the morning for horizons of 5 we can guarantee an error of maximum 7.45% with a probability of 0.9 which for the current

[2] Travel time is considered here as 7min for a free flow condition and 40min for a congested case



(a) APE examined between 7AM and 10AM



(b) APE examined between 4PM and 7PM

Figure 5.9: ECDF plot of the prediction performance for ECFM

	$p < X$	$\Delta = 5$	$\Delta = 10$	$\Delta = 15$	$\Delta = 20$	$\Delta = 25$
Morning	90%	9.04%	11.82%	14.19%	17.26%	19.59%
	80%	6.93%	8.35%	9.57%	10.62%	11.42%
Afternoon	90%	14.86%	18.97%	21.89%	24.35%	26.24%
	80%	10.93%	13.41%	15.27%	16.79%	18.20%

Table 5.3: APE guaranteed error based for multiple horizons. PSFM case

validation scenario represents 33sec in free cases up to 3min in congestion period. For longer horizons the APE augments up to 34.38% which may lead to errors over 10mins. The points over the curves are signalization of the mean errors measured in the distribution. In this case for all the time horizons these errors are bounded up to 13% in the morning case and 18% in the afternoon case. The method in general out performs the historical mean but we observed that these results could be improved with PSFM.

5.5.2.3 PSFM fusion strategy

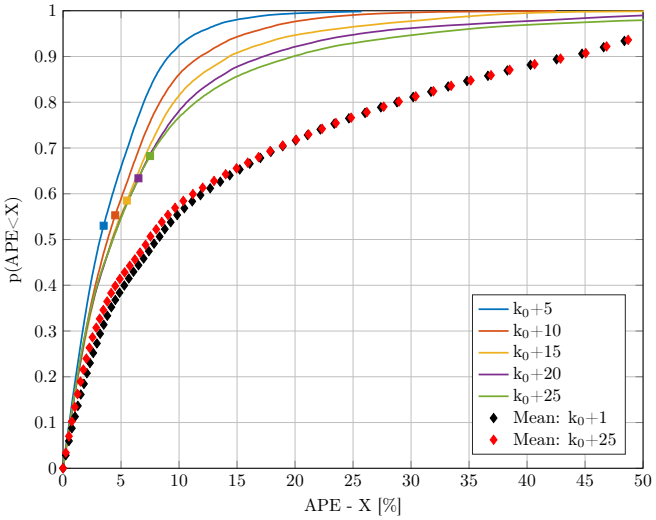
Fig. 5.10 depicts the ECDF for the multi-cluster fusion strategy for the PSFM scenario. In this case, same as the ECFM evaluation we have no *a priori* knowledge on the cluster to which the time series belongs. Therefore we are far from the ideal case. However, we can note that the method still perform better that the historical mean.

In the morning case, for the horizon of 5min the guaranteed error is 6.93% (meaning 29sec in free flow and 2min46sec during congestion) while for horizon of 25min we get 11.42% (meaning 47sec in free flow and 4min34sec during congestion). As for the afternoon case the error augments to 10.93% for horizon of 5min (meaning 46sec in free flow and 4min22sec during congestion) and 18.20% for horizon of 25min. For short term predictions the performance is still good while for longer term predictions there is a significant loss of performance compared to the ideal case of the previous section. (See table 5.3)

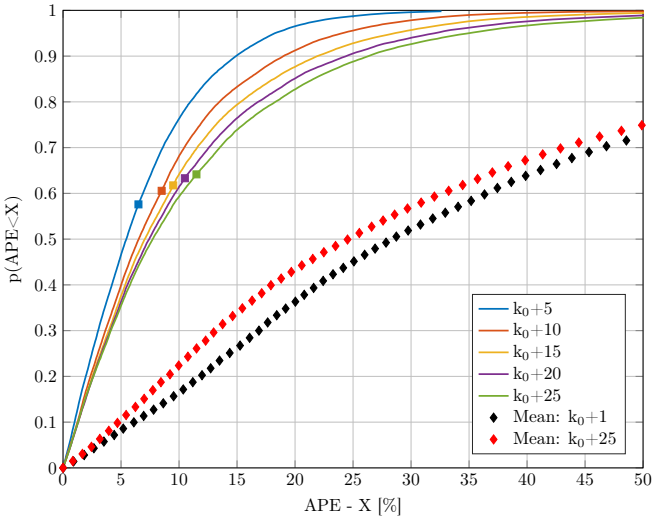
5.6 Discussion about performance of ECFM and PSFM

The two main fusion methods introduced in the previous sections deal with the forecast problem in two different ways. While ECFM incorporates the error at each time instant k when the filter updates, it is noticeable from Tables 5.3, 5.2 that error in the short term behaves better for the ECFM case. This fact is in consistence with the adaptation of the forecast in the future. While in ECFM (Fig 5.3) a permanent feedback from the predicted value is introduced in the mechanism, the approach in PSFM considers a weight which is fixed in the time.

In general in individual tests as well as cross validation tests PSFM is a method that performs slightly better than ECFM. For the PSFM, the information stored in $\Gamma(k)$ given



(a) APE examined between 7AM and 10AM



(b) APE examined between 4PM and 7PM

Figure 5.10: ECDF plot of the prediction performance for PSFM

by (5.19) contains information of the trend and at same time a weighted average of past information in the system. This acts in favor of the prediction due to the close correlation existing between the variable and the clusters during the interval ΔI_p. On the other hand in ECFM, the dynamic adaptation of $U(k)$ based on a permanent feedback from the output introduces error in the estimation due to the innovation term when the variance tends to augment. See (5.7).

5.7 Final comments on the chapter

This chapter has presented short term forecast techniques applied to DTT predictions in transportation networks. The data driven inspired techniques combine individual forecasts performed for clustered time series. Statistical features and models of the clustered data are provided as well as two fusion mechanisms, the first one based on best linear unbiased principles and the second one based on past similarity measurements. Prior to this individual forecasts models are developed under a Clustered Kalman filter approach.

The efficacy of the prediction methodologies is studied in terms of the prediction error along the future step horizon under a cross validation schema. It is found a growing behavior of the error with the prediction horizon. The proposed schemes outperform the historical mean and provide interesting results when a priori knowledge on the cluster to which belongs the time series to be predicted is available.

For the case proposed methodology PSFM improves the results obtained with respect to the ECFM. Particular individual predictions and statistical results are detailed in order to validate these results. Finally the study and design of these techniques is based on the classic Kalman Filter (KF) theory and it is oriented towards a real time application where new measurements of data can be received in short periods of time ([15 s-1 min]) as in the analyzed cases. This real time application will be presented in details in the subsequent chapter, in which the main benefit of this work is to perform low computational workload to obtain online DTT forecasts under real time traffic conditions.

Chapter 6

Simultaneous density and flow reconstruction

Contents

6.1 Overview

The previous chapters have studied the formalization of the TSP problem. The formulation has been stated under the basis of data driven models where historical data provides the main guideline to perform the predictions, in addition, we have studied examples of the approaches in freeway traffic networks. In this chapter, we focus the attention into the fusion of multiple sources of data that may contribute to the solution of the TSE problem, furthermore, we examine urban type networks. Urban networks are the root of majority of congestions due to complexity of the network topology.

Congestion is related to an excess of vehicles on a portion of road, the quantification of this event can be established through the density of the road. Despite existing systems this quantity is not easy to measure in complex traffic networks. The deployment of MLD has been the traditional technology but it is expensive and highly costly to maintain. In the previous chapters we have exploited a data driven model approach in order to estimate and forecast

DTT. The main objective on this approach was to provide a decision support tool for drivers using the infrastructure. The problem of tTT reconstruction as seen in chapter 3 makes use only of available speed measurements taken from MLD data. This problem was also studied in highway networks. In recent years, the development of mobile technologies has enormously increased, bringing along a strong market for relatively abundant FCD. FCD is part of the new group of technologies that enriches traffic information and may facilitate the task of TSE.

The combination of fixed sensor measurements from MLD and mobile data concentred in FCD provides rich information that we employ to address the problem of estimating road density and flow in a traffic network. Density is considered a good representation of the state of the system, providing more information than average speed alone. Appart from the state reconstruction, the following are three main aspects of interest which can be considered of big importance to motivate the problem. Solving the TSE problem serves as

- *Decision support tool*: Providing public authorities with statistical data to monitor the state of the network and predict dangerous scenarios. Simulated and estimated scenarios based on models and retrieved real data contribute to a better understanding of the system and better decisions making.
- *Information of current state*: Informing in real-time drivers about the state of the network through navigation systems. Current and accurate state information of the network provides an empowering tool for decisions about the usage of the traffic network.
- *Design of traffic lights*: Efficient road traffic management policies for networks consider assumptions on availability of state measurements which is not regularly the case [GGCDW16]. The high costs of this implementation makes the solution unfeasible from a practical point of view. An accurate state reconstruction of the network is regularly required as an initial stage to provide more accurate information for a control action.

This chapter is devoted to the problem of joint reconstruction of flow and density in traffic networks using heterogeneous data sources. To the best of our knowledge, this problem has been partially treated in some simple scenarios like highways, but very seldom in urban traffic networks. The estimation process is carried out on the basis of the steady-state equilibrium of traffic network with an explicit model for the road intersections. Each single road section is modeled by the well known macroscopic fluid traffic model LWR, while the intersections are modeled by solving the single junction problem [BNP06] as a set of optimization problems (maximum outflows) which turn out to be Linear Programming (LP) if the piece-wise linear fundamental (triangular) diagram is adopted.

The whole estimation problem is finally described as a minimization of the quadratic error between the measurement and their estimates under linear equality constraints coming from the network model. One key general difficulty in this problem, which is inherent to the nature of the system, is the "non-invertibility" of the velocity-to-density map in free-flow conditions. Namely, several density values are possible when the network operate in free-flow velocity. Then, in this free-flow regime, one velocity point maps to a bounded density set. Although

several regularization schemes are possible, the results will remain very sensitive to noise in this domain, and the errors will be spread out to the whole set of estimated variables. To overcome this difficulty, we make use of the equilibrium state of the system which intends to provide flow propagation within the network and we adopt equality constraints which allows the density recovery by promoting solutions in the fundamental diagram.

The chapter is organized as follows. We first introduce the network traffic model adopted in this paper and formulate the problem under study in sections 6.2 and 6.3 respectively. Then we present the optimization algorithms for both the density and flow reconstructions which are evaluated in a simulation scenario which considers a Manhattan like grid.

6.2 Road traffic network model

A urban network is made of roads and junctions between them. In what follows, we describe the dynamics governing each road and their splitting or merging in junctions.

6.2.1 Macroscopic Traffic model for a single section

We consider one of the most used instance of a continuous macroscopic traffic model, the LWR [LW55], [Ric56] given in (2.1). The flux function $\Phi(\cdot)$ is assumed to be concave and it takes values in the interval $[0, \varphi^{\max}]$. The characteristic curve of $\Phi(\rho)$, widely known as fundamental diagram, may take multiple forms including a triangular ones [New93]. The discretized version of this model is known as the CTM [Dag94]; [Dag95b], which is undoubtedly one of the main and most well known traffic models to date. It is based on a first order Godunov approximation of LWR [Leb96].

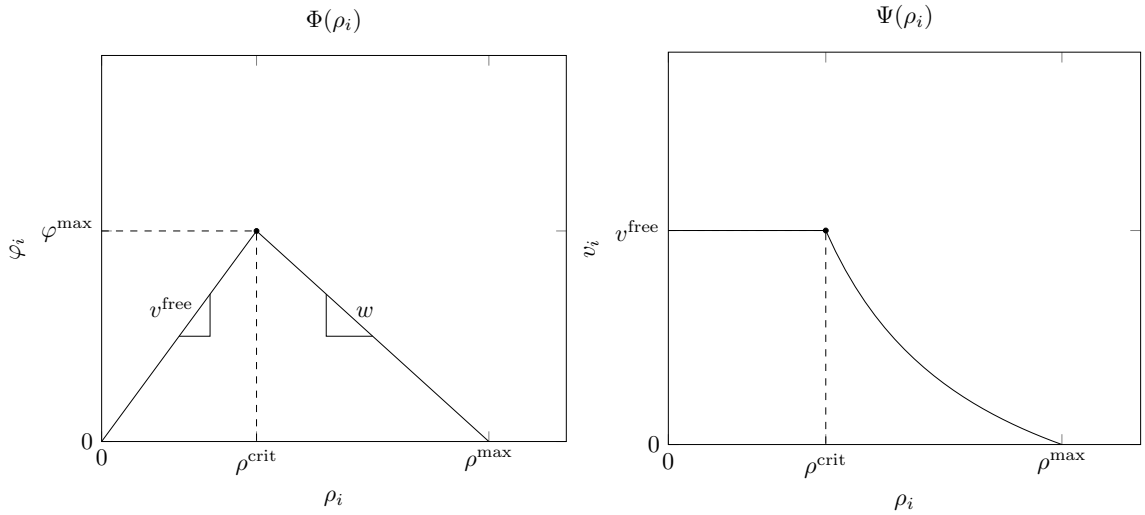
In the CTM, the road segment of interest is first partitioned into a sequence of cells. The propagation of traffic dynamics in each cell is given by the following set of equations:

$$\begin{aligned} \rho_i(k+1) &= \rho_i(k) + \frac{T}{l_i}(\varphi_{i-1}(k) - \varphi_i(k)) \\ \varphi_{i-1}(k) &= \min\left(v^{\text{free}}\rho_{i-1}(k), \varphi^{\max}, w(\rho^{\max} - \rho_i(k))\right) \end{aligned} \tag{6.1}$$

where $\rho_i(k)$ denotes the current vehicles density in the ith cell while $\varphi_i(k)$ stands for the interface flow between the ith and $(i+1)$th cells, l_i being the length of the cell. The followings are the set of parameters associated to each cell[1]:

- $\rho^{\max}$, the maximum density, often referred as *jam* density,
- $\varphi^{\max}$, the maximum capacity flow,
- v^{free}, the maximum velocity of vehicles in the cell, said the *free–flow* velocity,

[1] We omit the i index in the notation of parameters of each cell to simplify the reading.



(a) Flow-density (b) Speed-density

Figure 6.1: Density-flow and density-speed fundamental diagrams. The value of density ρ^{crit} is usually said *critical density*, and it represents the switching point between a free-flow regime and a congested one

- w, the speed of the congestion wave in back propagation.

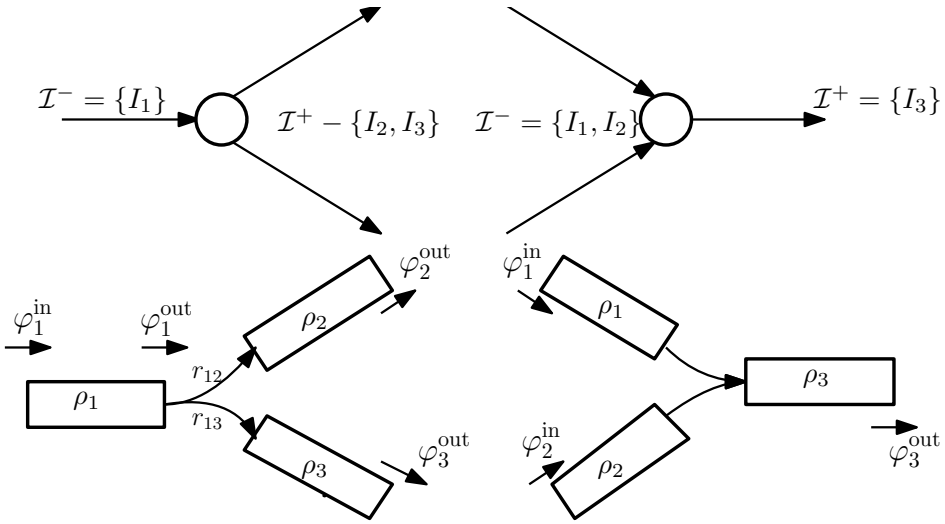
These parameters can be easily found in the fundamental diagram (See Fig. 6.1).

6.2.2 Junction model

Let $\mathcal{G}$ denote the set of junctions within the network. A junction labeled $g \in \mathcal{G}$ represents a physical connection between 2 or more roads and they can be found in shapes of bottlenecks, divergences, or merges. A junction g is represented by the tuple $(\mathcal{I}_g, \mathcal{R}_g)$ where $\mathcal{I}_g = \{I_i : i = 1, \dots, n+m\}$ represents a set of roads and $r_{[ij]} \in \mathcal{R}_g$ are the splitting ratios denoting driving preferences. Each element of $\mathcal{G}$ symbolizes the existing junction between the set of n upstream roads $\mathcal{I}_g^- = \{I_i : i = 1, \dots, n\}$ and the set of m downstream roads $\mathcal{I}_g^+ = \{I_j : j = n+1, \dots, n+m\}$ as depicted in Fig. 6.2. The set of junctions $\mathcal{G}$ along with the corresponding sets $\mathcal{I}_g^-, \mathcal{I}_g^+, \mathcal{R}_g$ constitute the called Road Traffic Network (RTN).

Each cell represents a node in a bigger network and it is characterized by a set of unique variables describing the macroscopic behavior: $\rho_i(k)$, the density, $\varphi_i^{\text{in}}(k)$, the ingoing flow, and $\varphi_i^{\text{out}}(k)$, the outgoing one. We refer to $\rho(k)$, $\varphi^{\text{in}}(k)$, $\varphi^{\text{out}}(k)$ as the vectors collecting all the densities and flows of the network.

Definition 6.1 (Single Junction Problem)
Consider a junction g with n incoming roads $\mathcal{I}_g^-$ and m outgoing ones $\mathcal{I}_g^+$ under the vehicle



a) Diverge junction b) Merge junction

Figure 6.2: Simplified case of junction types

conservation law, i.e.

$$\sum_{i=1}^{n} \varphi_i^{out}(\rho_i(k)) = \sum_{j=n+1}^{n+m} \varphi_j^{in}(\rho_j(k)). \tag{6.2}$$

The single junction problem consists in determining values of inflow and outflow consistent with the solutions of the LWR model (or its discretized counterpart (6.1)*) [GHP16], [CGP05].*

Since solutions of φ^{in}, φ^{out} to the traffic distribution problem in (6.2) are non unique, additional rules to incorporate drivers' behavior can be added to define a particular solution. Typical rules for Manhattan-like networks[2] are:

A) *Drivers follow fixed routes.* Hence there exists traffic routing coefficients $r_{ij} \in (0, 1]$, representing the splitting ratio from road I_i to road I_j. The resulting matrix, called Splitting Ratio matrix, $R_g = [r]_{ij}$ is row stochastic $\sum_{j=n+1}^{n+m} r_{ij} = 1$. This matrix is assumed to be known (at least in average).

B) *Drivers tends to maximize the network throughput.* So drivers behavior is such that incoming flows to the junctions φ^{out} are maximized.

A solver for the junction problem Consider the rules A) and B) together with the triangular fundamental diagram of Fig. 6.1 having the density-flow map,

$$\Phi(\rho_i) = \begin{cases} v^{\text{free}}\rho_i & 0 \le \rho_i \le \rho^{\text{crit}} \\ w(\rho^{\max} - \rho_i) & \rho^{\text{crit}} < \rho_i \le \rho^{\max} \end{cases}, \tag{6.3}$$

[2]For the sake of simplicity we have limited our approach to Manhattan-like networks where there are no merging junctions in which $n > m$. However, it is also possible to generalize for any other type of networks including also merging junctions by adding priority driving rules, see [GHP16],[BNP06]

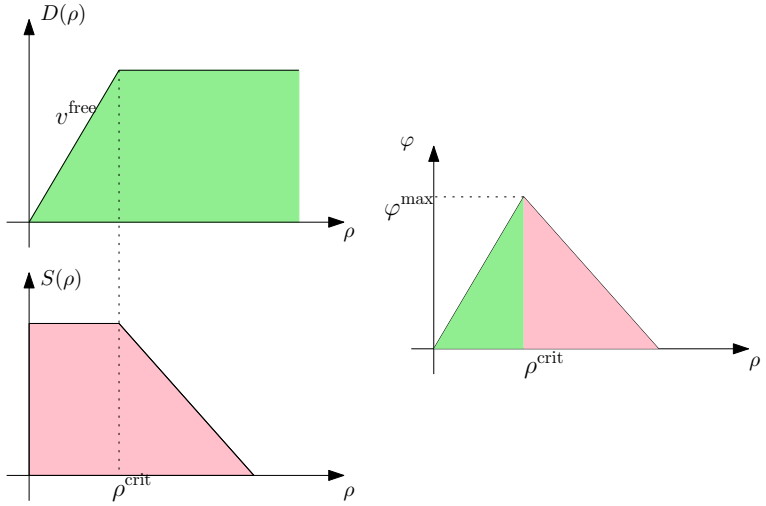


Figure 6.3: Demand and Supply functions. The intersection of both functions characterize the triangular fundamental diagram

and, its corresponding speed-density map,

$$\Psi(\rho_i) = \begin{cases} v^{\text{free}} & 0 \leq \rho_i \leq \rho^{\text{crit}} \\ w\left(\dfrac{\rho^{\max}}{\rho_i} - 1\right) & \rho^{\text{crit}} < \rho_i \leq \rho^{\max} \end{cases} . \tag{6.4}$$

Note that the admissible solutions are those satisfying the Riemann problem [GHP16]. The Riemann's admissible solutions can be rewritten using the Demand-Supply formalism, where the demands $D(\rho_i)$ and supplies $S(\rho_i)$ functions are given as (see also Fig 6.3):

$$D(\rho_i) = \min\left(v^{\text{free}}\rho_i, \varphi^{\max}\right), \tag{6.5}$$
$$S(\rho_i) = \min\left(\varphi^{\max}, w(\rho^{\max} - \rho_i)\right). \tag{6.6}$$

In a simple junction of one inflow and one outflow, the interface flow corresponds to

$$\varphi_i^{\text{out}} = \min\left(D(\rho_i), S(\rho_{i+1})\right), \tag{6.7}$$

and the solutions φ_i^{out} can be expressed as upper bounds of the inequalities

$$\begin{aligned} 0 \leq \varphi_i^{\text{out}} &\leq D(\rho_i) \\ 0 \leq \varphi_i^{\text{out}} &\leq S(\rho_{i+1}). \end{aligned} \tag{6.8}$$

Maximizing the throughput as suggested by rule B), implies maximizing

$$\max_{\varphi^{\text{out}}} \quad \sum_{i=1}^{n} \varphi_i^{\text{out}} \tag{6.9}$$

and introducing rule A), implies that inflows and outflows are linearly related by the relationship $\varphi_j^{\text{in}} = \sum_{i \in \mathcal{I}^-} r_{ij}\varphi_i^{\text{out}}$, transforming the inflows in

$$\varphi^{\text{in}} = R^T\varphi^{\text{out}} + \varphi^{\text{ext}}. \tag{6.10}$$

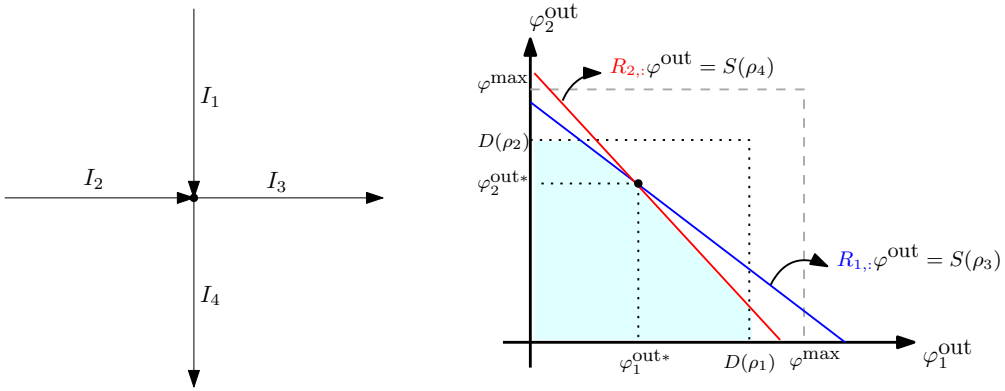


Figure 6.4: Solution junction problem case 2×2. Left hand side illustrates the network topology. Right hand side illustrates the solution of (6.14)

where $\varphi^{\text{ext}} = B\lambda_e$ is a vector containing exogenous external inflows to the network in the corresponding positions of φ^{in} and zeros elsewhere. B is a selection matrix for the incoming boundary flows of the network.

In problem (6.9) the flow φ^{out} should respect the relationships (6.8) for each one of the entering roads to the intersection. Let consider first an organization of the set of constraints for a single junction g as

$$\begin{aligned} \check{\mathcal{D}}_g &: \Big\{ \varphi_i^{\text{out}} \leq v^{\text{free}} \rho_i \quad \forall \quad i \in \mathcal{I}^- \Big\}, \\ \mathcal{D}_g^{\max} &: \Big\{ \varphi_i^{\text{out}} \leq \varphi^{\max} \quad \forall \quad i \in \mathcal{I}^- \Big\}, \end{aligned} \tag{6.11}$$

$$\begin{aligned} \check{\mathcal{S}}_g &: \Big\{ \sum_{i=1}^{n} r_{ij} \varphi_j^{\text{out}} \leq w(\rho^{\max} - \rho_j) \quad \forall \quad j \in \mathcal{I}^+ \Big\}, \\ \mathcal{S}_g^{\max} &: \Big\{ \sum_{i=1}^{n} r_{ij} \varphi_j^{\text{out}} \leq \varphi^{\max} \quad \forall \quad j \in \mathcal{I}^+ \Big\}. \end{aligned} \tag{6.12}$$

The set of constraints $\mathcal{D}_g = \check{\mathcal{D}}_g \cap \mathcal{D}_g^{\max}$ and $\mathcal{S}_g = \check{\mathcal{S}}_g \cap \mathcal{S}_g^{\max}$, complete the formulation required for the junction problem. With $\mathcal{P}_g^{\varphi} : \mathcal{D}_g \cap \mathcal{S}_g$ the solution to the junction problem can then finally stated as:

$$\begin{aligned} \max_{\varphi^{\text{out}}} \quad & \sum_{i=1}^{n} \varphi_i^{\text{out}} \\ \text{s.t.} \quad & \varphi^{\text{out}} \in \mathcal{P}_g^{\varphi}. \end{aligned} \tag{6.13}$$

Solutions of the optimization problem (6.13) typically reach the upper boundaries of the constraints (See Fig. 6.3), and they provide a single unique solution for cases where $m > n$.

Example 6.1

Let consider the classical intersection with two incoming roads $\mathcal{I}^- = \{I_1, I_2\}$ and two exiting roads $\mathcal{I}^+ = \{I_3, I_4\}$ (See Fig. 6.4). The single junction problem for the case 2×2 can be

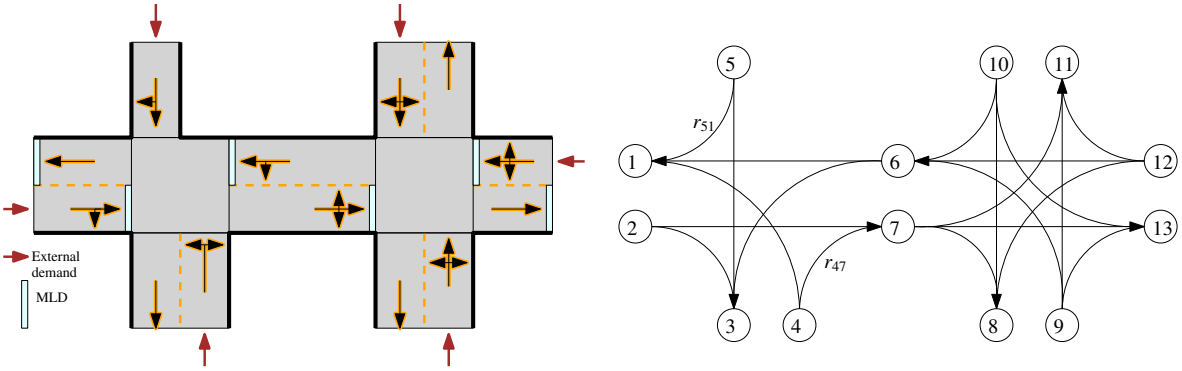


Figure 6.5: RTN and its graph model. Left hand side panel shows the physical infrastructure. Right hand side the graph $\mathcal{H}$

casted as the following optimization problem,

$$\begin{aligned}
\min_{\varphi_1^{out},\varphi_2^{out}} \quad & \varphi_1^{out} + \varphi_2^{out} \\
s.t. \quad & \varphi_1^{out} \leq v^{free}\rho_1, \quad \varphi_1^{out} \leq \varphi^{max} \\
& \varphi_2^{out} \leq v^{free}\rho_2, \quad \varphi_1^{out} \leq \varphi^{max} \\
& \begin{bmatrix} r_{13} & r_{23} \\ r_{14} & r_{24} \end{bmatrix} \begin{bmatrix} \varphi_1^{out} \\ \varphi_2^{out} \end{bmatrix} \leq \begin{bmatrix} w(\rho^{max} - \rho_3) \\ w(\rho^{max} - \rho_4) \end{bmatrix}, \\
& \begin{bmatrix} r_{13} & r_{23} \\ r_{14} & r_{24} \end{bmatrix} \begin{bmatrix} \varphi_1^{out} \\ \varphi_2^{out} \end{bmatrix} \leq \begin{bmatrix} \varphi^{max} \\ \varphi^{max} \end{bmatrix}, \\
& \varphi_1^{out} \geq 0, \quad \varphi_2^{out} \geq 0
\end{aligned} \tag{6.14}$$

The solution of (6.14) *namely* $(\varphi_1^{out*}, \varphi_2^{out*})$ *corresponds to one of the extreme points of the intersection of all constraints [GHP16]. Right hand side of Fig. 6.4 illustrates the solution of the proposed optimization problem.*

6.2.3 Full traffic network model

The full network of roads can be described by a weighted directed graph $\mathcal{H} := (\mathcal{I}, \mathcal{R})$. The graph is represented by the union of all tuples $\left(\mathcal{I} := \bigcup_{g\in\mathcal{G}} \mathcal{I}_g, \mathcal{R} := \cup_{g\in\mathcal{G}}\mathcal{R}_g\right)$ for each intersection resulting in n_H roads[3].

The complete traffic network model combines, dynamic equations of the density evolution of each road (6.1), in its vector form

$$\rho(k+1) = \rho(k) + TL^{-1}(\varphi^{\text{in}}(k) - \varphi^{\text{out}}(k)). \tag{6.15}$$

or equivalently, using rule A),

$$\rho(k+1) = \rho(k) + TL^{-1}\Big((R^T - I)\varphi^{\text{out}}(k) + \varphi^{\text{ext}}(k)\Big), \tag{6.16}$$

[3]We consider, all the roads in the graph $\mathcal{H}$ to be re-labelled with a single index i so that each road keeps a unique identifier.

where T the sampling time and $L = diag\{l_i\}$, with the junction models compactly represented by a graph resulting form the union of multiple single junction problems, as shown in Fig. 6.5.

Definition 6.2 (Network Junction Problem)
The extension of the single junction problem to the network case is represented as the union of all local problems respecting simultaneously all the constraints imposed by the traffic model at each junction.

$$\max_{\varphi^{out}} \quad \mathbb{1}^T \varphi^{out} \quad s.t.\ \varphi^{out} \in \mathcal{P}^{\varphi}. \tag{6.17}$$

The network junction problem is the solution to the optimization problem (6.17)*. The union of all problems is obtained through the maximization of the total throughput* $\sum_{i=1}^{n_H} \varphi_i^{out} = \mathbb{1}^T \varphi^{out}$ *and the constraints* $\mathcal{P}^{\varphi} := \bigcup_{g \in \mathcal{G}} \mathcal{P}_g^{\varphi}$.

6.3 Density and flow reconstruction

In this section we present the joint density and flow reconstruction problem. We first indicate the main data characteristics and their associated observation model. Then, based on such properties we propose a new reconstruction scheme based on the error minimization with respect to the available measurements, the minimization of the error between the estimated variables and the steady state condition of the system and finally, we constraint the space of solutions to be the same as the boundary set obtained from the solution of the network junction problem. The algorithm boils down to solve a quadratic optimization problem with linear equalities.

6.3.1 Observations model and its mathematical properties

Let $\mathcal{I}_{FCD}$ be the set of roads where FCD are collected. FCD measurements can be considered to be available everywhere in the network, i.e. $|\mathcal{I}_{FCD}| = n_H$. They describe the average velocity at each road. On the other hand, let $\mathcal{I}_{MLD}$ be the set where MLD data is collected. They measure outflows at the road ends where sensors are installed. Loop detectors are not available at all roads $|\mathcal{I}_{MLD}| = N_M < n_H$. These measurements have the following observation models,

$$\bar{v}_i(k) = \Psi(\rho_i, k) + \eta_v(\rho_i, k), \quad i \in \mathcal{I}_{FCD} \tag{6.18}$$

$$\bar{\varphi}_l^{\text{out}}(k) = \Phi(\rho_l, k) + \eta_{\varphi^{\text{out}}}(\rho_l, k), \quad l \in \mathcal{I}_{MLD} \tag{6.19}$$

where the terms $\eta_{\varphi^{\text{out}}}, \eta_v$ represent additive noise produced by factors such as aggregation time, penetration rate, measurement noise, etc, which naturally may affect the measurement quality. In appendix B we provide an experimental characterization of this effect. The velocity measurement (6.18) will be used as a basis for density reconstruction as velocity are sensed in all the network road. However, its inverse map (noise apart)

$$\rho_i \quad = \quad \Psi^{-1}(\bar{v}_i + \eta_v) \tag{6.20}$$

is not invertible in the free-flow part (See Fig. 6.1). On the other hand the inverse map of the flow,

$$\rho_l \quad = \quad \Phi^{-1}(\bar{\varphi}_l^{\text{out}} + \eta_{\varphi^{\text{out}}}) \tag{6.21}$$

has the problem that is not uniquely defined meaning that two values of density could lead to the same flow. The non-unicity of (6.21) can be tackled by using the velocity measures allowing to discriminate congested from free-flow regimes, but can only be used in some roads. Therefore, velocity measurements and flow measurements will be used in different way to facilitate the signal reconstructions. The density/flow reconstruction algorithm is designed on the basis of the following assumptions:

Assumption 1 (Boundary flows)
All inflows and outflows at the network boundaries are measurable.

Assumption 2 (Measured FCD speeds)
Speeds captured by FCD are measured everywhere in the network.

Assumption 3 (Density pseudo-observation)
Let consider the fundamental diagram and a measurement of speed taken in the congested regime, then there exists a density observation which can be uniquely recovered from the map (6.4)*,* $\forall \rho_i \in [\rho^{crit}, \rho^{max}], \forall \bar{v}_i \in [0, v^{free})$

$$\bar{\rho}_i = \Psi^{-1}(\bar{v}_i) = \frac{\rho^{max}}{1 + \bar{v}_i/w}, \tag{6.22}$$

Assumption 4 (Network Equilibrium)
The reconstruction is done in a fast enough time scale so that the network can be considered to be at the equilibrium. At the network equilibrium, the flows are then related by the steady-state equation of (6.16)*,*

$$(R^T - I)\varphi^{out} + B\lambda_e = 0. \tag{6.23}$$

Remark: Under full rank conditions for the matrix $R^T - I$ a unique solution $\varphi^{\text{out}} = -(R^T - I)^{-1}B\lambda_e$ can be obtained for the vector of outflows in the traffic network.

6.4 Reconstruction problem

We consider the reconstruction problem expressed as the solution to an optimization problem at a fixed time instant

$$\min_{\hat{\varphi}^{\text{out}}, \hat{\rho}} \qquad J_{\varphi^{\text{out}}} + J_\rho \quad \text{s.t } \mathcal{M}_{\varphi,\rho} \tag{6.24}$$

where $\mathcal{M}_{\varphi,\rho}$ denotes a RTN model. The terms $J_{\varphi^{\text{out}}}$ and J_ρ represent penalty functions for the error between the estimated values $\hat{\varphi}^{\text{out}}, \hat{\rho}$ and the measurements from the observation model in the systems.

The integration of these penalty functions allow the system to consider the reality captured by the observation model (6.20), (6.21). In particular, J_φ integrates into the problem direct information contained in the MLD measurements. This cost function minimizes the error between available measurements $\bar{\varphi}^{\text{out}} = \{\bar{\varphi}_l^{\text{out}}, l \in \mathcal{I}_{MLD}\}$ through the quadratic norm $\sum_{l\in\mathcal{I}_{MLD}} (\hat{\varphi}_l^{\text{out}} - \bar{\varphi}_l^{\text{out}})^2$.

On the other hand, assumption (3) allows a good recovery of information in the congested zones. According to Fig. 6.1 density values on congested zones are contained in the speed measurements. Let us consider

$$S_{\bar{v}_i} = \begin{cases} 1, & \bar{v}_i < v^{\text{free}} \\ 0, \text{elsewhere} \end{cases}$$

For all the network the selection matrix $S_{\bar{v}} = diag\{S_{\bar{v}_i}\}$ is a transformation that contains on its diagonal ones for cells in congested zones and zeros everywhere else. The penalty goal for the density case is achieved by considering $\sum_{\bar{v}_j<v^{\text{free}}} \left(\hat{\rho}_j - \bar{\rho}_j\right)^2$, which has as an objective the minimization of density error in congested measured links. Finally the two terms can be exchanged transforming the problem into

$$\min_{\hat{\varphi}^{\text{out}},\hat{\rho}} \quad \sum_{i=1}^{N_M} (\hat{\varphi}_l^{\text{out}} - \bar{\varphi}_l^{\text{out}})^2 + \sum_{j=1}^{N_H} \left(S_{\bar{v}_j}(\hat{\rho}_j - \bar{\rho}_j)\right)^2 \text{ s.t } \mathcal{M}_{\hat{\varphi},\rho} \tag{6.25}$$

Difficulties and relaxation of the estimation problem Some difficulties lie in the formulation of the problem (6.25). Non-linearities of $\mathcal{M}_{\varphi,\rho}$ may appear in the fundamental diagram and no explicit solution is known to the problem. When introducing non linear constraints to the estimation problem, the solutions are regularly hard to approximate and most algorithms extract local minima. In addition, the computational cost and the complexity to solve the problem increases when the size of the network augments.

We introduce then a relaxed version of this problem which considers two main aspects. First, given the sparse nature of the flow measurements, and considering assumption 4 the steady state of the system constitutes an additional source of information to the flow reconstruction problem. This can be integrated into the cost function through the norm $\left\|(R^T - I)\varphi^{\text{out}} + B\lambda_e\right\|^2$.

Even though this term adds information to (6.25), it does not tackle the nonlinearity. In order to relax the non-linear constraints we make use of piecewise linear fundamental diagrams (6.3), (6.4). The set of constraints is then transformed into linear equalities. They shape boundaries of the space of solutions established by the network junction problem. This aims to push solutions of the estimation within the fundamental diagram. Given assumption 2 the information about speeds provides a way to classify congested from free-flow cell constraints as:

$$\begin{aligned} \bar{\mathcal{D}}_g &: \left\{\varphi_i^{\text{out}} = v^{\text{free}}\rho_i \quad \forall \quad i \in \mathcal{I}^- \wedge \bar{v}_i = v^{\text{free}}\right\}, \\ \bar{\mathcal{S}}_g &: \left\{\sum_{i=1}^{n} r_{ij}\varphi_j^{\text{out}} = w(\rho^{\max} - \rho_j) \quad \forall \quad j \in \mathcal{I}^+ \wedge \bar{v}_j < v^{\text{free}}\right\}, \end{aligned} \tag{6.26}$$

Such constraints can be expressed in terms of the selection matrix $S_{\bar{v}}$ as

$$A(\bar{v})\varphi^{\text{out}} = B(\bar{v})\rho + C(\bar{v}), \tag{6.27}$$

where

$$\begin{aligned} A(\bar{v}) = \begin{bmatrix} S_{\bar{v}} - I \\ R^T S_{\bar{v}} \end{bmatrix}, \quad B(\bar{v}) = \begin{bmatrix} v^{\text{free}}(S_{\bar{v}} - I) \\ -wS_{\bar{v}} \end{bmatrix}, \\ C(\bar{v}) = \begin{bmatrix} \mathbb{0} \\ w\rho^{\max}\mathbb{1}S_{\bar{v}} \end{bmatrix}. \end{aligned} \tag{6.28}$$

The resulting optimization problem is quadratic with linear equalities which provide good properties for the solution. The reconstruction problem can be formulated as follows.

Problem 1 (Joint density / flow reconstruction)
Given a set of measurements of flow $\bar{\varphi}^{out} \in \mathcal{I}_{MLD}$ and a set of speed measurements $\bar{v} \in \mathcal{I}$ the inverse problem defined to recover $\hat{\rho}, \hat{\varphi}^{out}$ is given by the solution of the following optimization problem:

$$\begin{aligned} \min_{\hat{\varphi^{out}}, \hat{\rho}} \quad & \left\| C_M \hat{\varphi}^{out} - \bar{\varphi}^{out} \right\|^2_{\gamma_\varphi} + \left\| S_{\bar{v}} \left(\hat{\rho} - \Psi^{-1}(\bar{v}) \right) \right\|^2_{\gamma_\rho} + \\ & \left\| (R^T - I)\hat{\varphi}^{out} + B\lambda_e \right\|^2_{\gamma} \\ s.t. \quad & A(\bar{v})\hat{\varphi}^{out} = B(\bar{v})\hat{\rho} + C(\bar{v}) \\ & \hat{\rho} \in \mathcal{P}^\rho. \end{aligned} \tag{6.29}$$

In (6.29) $\mathcal{P}^\rho := \bigcup_{g \in \mathcal{G}} \mathcal{P}^\rho_g$ define the boundaries for ρ. In this case this particular boundaries can be determined from the speed measurements $\bar{v}$. C_M selection matrix for the outgoing flows wherever they are available. The terms $\gamma_\varphi, \gamma_\rho, \gamma$ are weighting factors.

6.5 Simulations and scenarios

6.5.1 Scenario description

We have built a manhattan topology network as depicted in Fig. 6.6. In this case, each one of the junctions is specified by 2 ingoing roads and 2 outgoing roads. The following set of equations describe the dynamics for each one of the local intersections.

$$\begin{aligned} \rho_1(k+1) &= \rho_1(k) + \frac{T}{l_1}(\varphi_1^{\text{in}} - \varphi_1^{\text{out}}) \\ \rho_2(k+1) &= \rho_2(k) + \frac{T}{l_2}(\varphi_2^{\text{in}} - \varphi_2^{\text{out}}) \\ \rho_3(k+1) &= \rho_3(k) + \frac{T}{l_3}(r_{13}\varphi_1^{\text{in}} + r_{23}\varphi_2^{\text{in}}) - \varphi_3^{\text{out}} \\ \rho_3(k+1) &= \rho_4(k) + \frac{T}{l_4}(r_{14}\varphi_1^{\text{in}} + r_{24}\varphi_2^{\text{in}}) - \varphi_4^{\text{out}} \end{aligned} \tag{6.30}$$

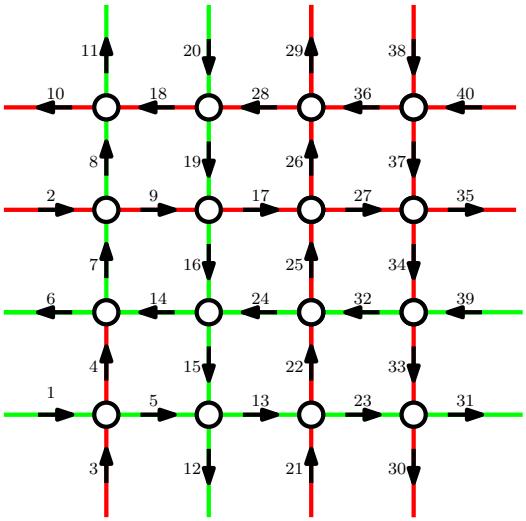


Figure 6.6: Manhattan network topology under study

Each one of the intersections solve the model (6.30) where the solution for the outgoing flows are found via the optimization problem presented in (6.14). The system is excited externally with random values for external inflows and outflows selected from a uniform distribution in $[0, \varphi^{\max}]$, meaning the boundary conditions of the network are known. The selection of the splitting ratios is fixed so that 70% of the flow continues in a straight direction while the remaining turns. We consider all cells are uniform $l_i = 500m$, $v^{\text{free}} = 50Km/h$, $\varphi^{\max} = 2000veh/h$. The corresponding value of w is obtained from (6.3). For the purpose of this simulation $\gamma_\varphi = \gamma_\rho = \gamma = 1$

6.5.2 Performance measurement

In order to assess performance of the method, we consider the Absolute Error (AE) as a reference:

$$AE_\rho(k) = \frac{1}{N}\sum_{i=1}^{N} |\rho_i(k) - \hat{\rho}_i(k)| \tag{6.31a}$$

$$AE_{\varphi^{\text{out}}}(k) = \frac{1}{N}\sum_{i=1}^{N} |\varphi_i^{\text{out}}(k) - \hat{\varphi}_i^{\text{out}}(k)| \tag{6.31b}$$

We illustrate the performance and limitations of the discussed estimation method. Initially, a simulation of the Manhattan grid using the CTM and the solution for the network junction problem was run until it reached equilibrium. These results were taken as ground truth. Using this equilibrium state, the traffic state estimation algorithm was executed by using measurements of the inflows at the boundaries and speed information everywhere. Results for density and flow are shown in Figure 6.7.

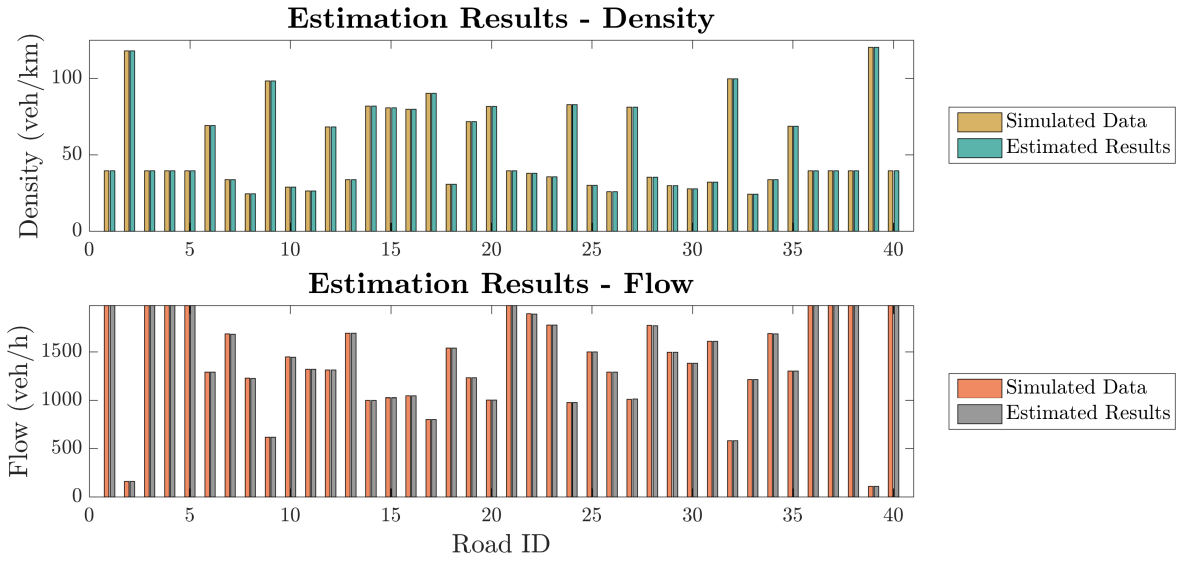


Figure 6.7: Comparison of ground truth and estimated values for density (top) and flow (bottom) at network equilibrium

As it can be observed, the estimation algorithm is able to capture the free-flow and congested locations of the network, with values close to ground truth. For the case of vehicle density, the mean absolute error is 0.025 veh/km with a maximum error of 0.037 veh/km, whereas vehicle flow presented a mean error of 0.45 veh/h and a maximum error of 1.5 veh/h.

Even though the algorithm has been proposed for a static case, to illustrate the performance of the algorithm we consider a dynamical scenario. For this, a second experiment was carried out. For this instance, the traffic network was given a random initial condition of flow and density, and the inflows at the boundaries were initialized with a demand equal to random constant value uniformly distributed in the interval $[0, \varphi^{\max}]$. Each of the iterations of the simulation process represents 15 seconds. At every step, the estimation method was applied with the assumption that the network had reached equilibrium. Results are presented in Figs. 6.8 and 6.9. It is worth to see that the algorithm has good convergence in terms of the APE indicator in steady state as desired. Although this error may augment in fast transients due to discontinuities in the solution, the reconstruction of the density and flow replicate similar dynamics.

In Figs. 6.8 and 6.9, green sections correspond to values of low density (or high flow) whereas red/black sections correspond to values of high density (or low flow). As it can be observed, even by assuming equilibrium at every moment of the simulation, the estimation method is able to capture the general dynamics of the system. It is noted that whilst in the ground-truth case the time transitions are smooth, the estimated results present jumps at some time moments. This is due to the incorrect assumption of equilibrium and the difficulty of the algorithm to recover density at the free-flow regime.

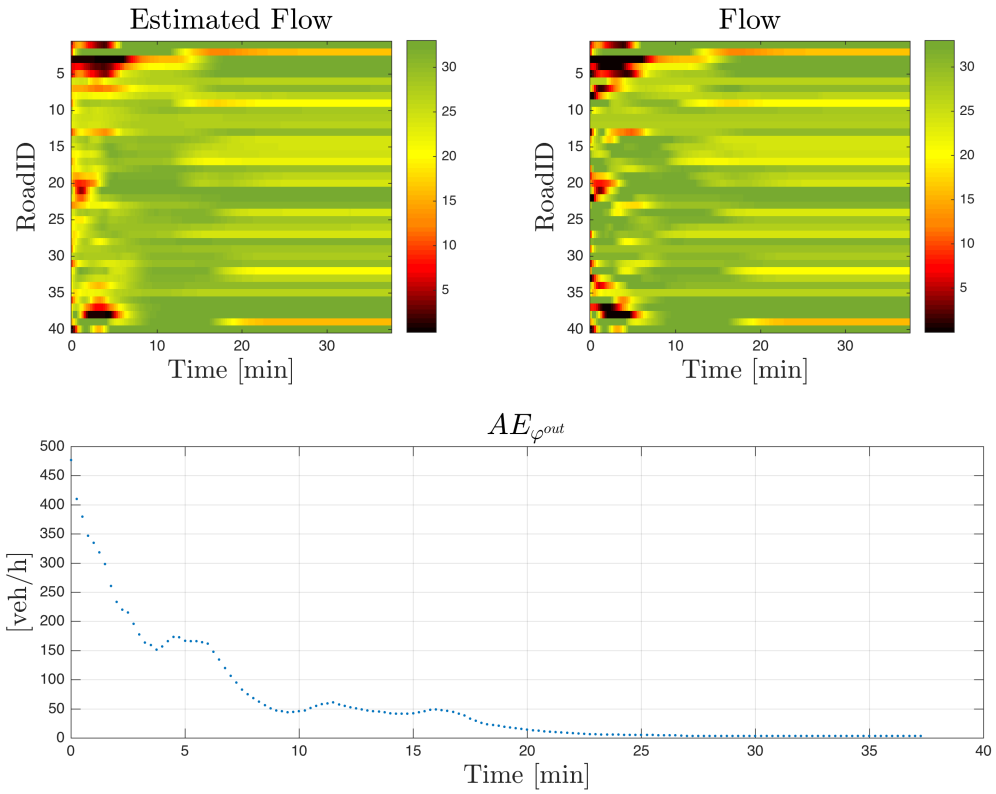


Figure 6.8: Flow dynamic reconstruction: Ground truth (right) and estimated (left) time series in a RTN.[4]

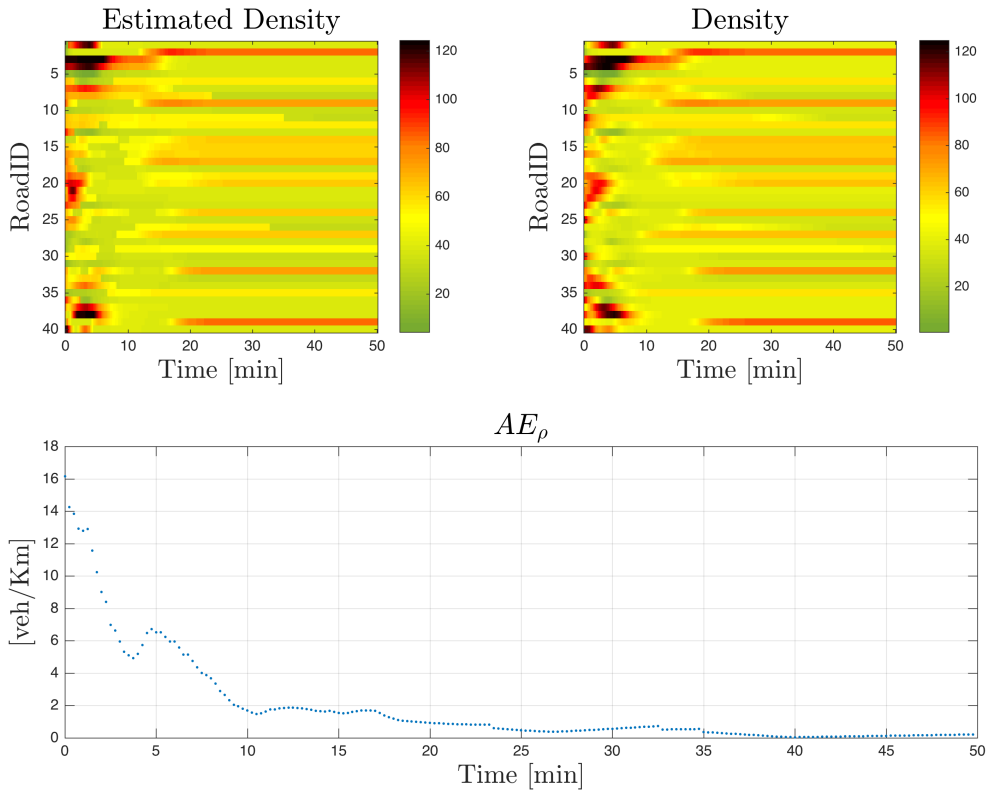


Figure 6.9: Density dynamic reconstruction: Ground truth (right) and estimated (left) time series in a RTN.

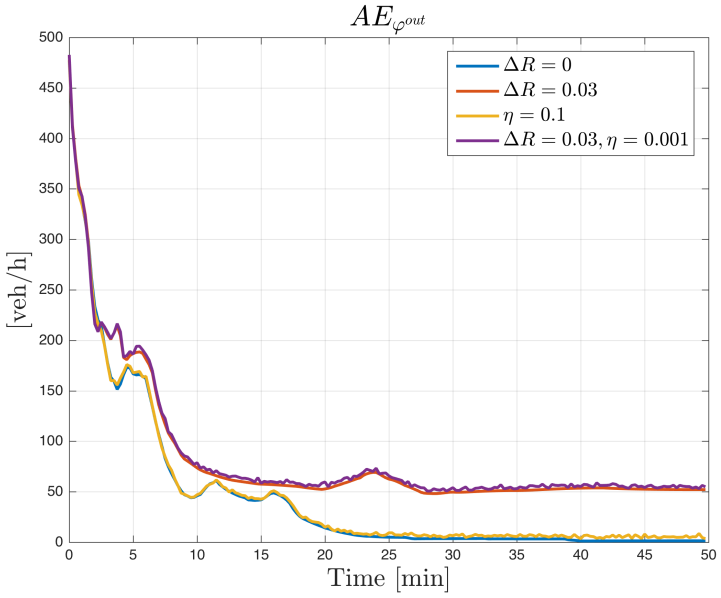


Figure 6.10: $AE_{\varphi^{\text{out}}}(k)$ for a RTN corresponding to several cases of perturbations and measurement noises.

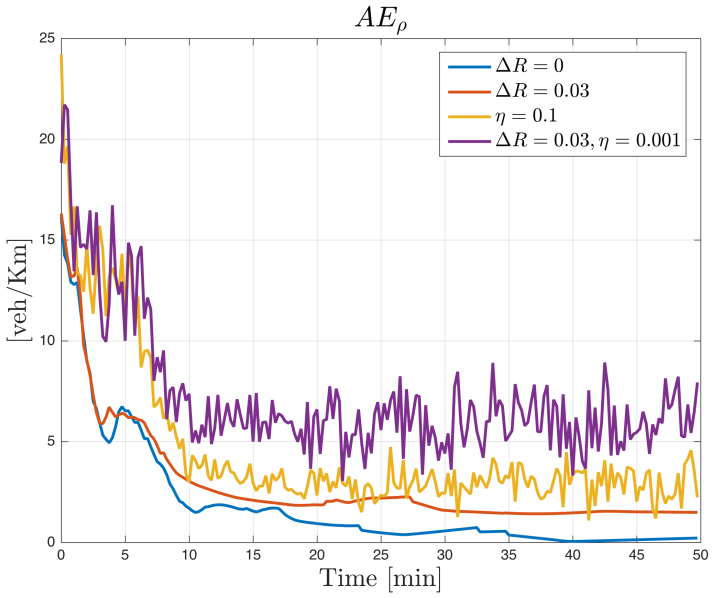


Figure 6.11: $AE_{\rho}(k)$ for a RTN corresponding to several cases of perturbations and measurement noises.

Scenario	$\sum_{k\in\Delta T_s} AE_\rho(k)$ [veh/km][6]	$\sum_{k\in\Delta T_s} AE_{\varphi^{\text{out}}}(k)$ [veh/h]
Case 1- Reference	0.17	1.3797
Case 2- (ΔR)	1.49	52.23
Case 3- (η)	2.81	5.30
Case 4- ($\Delta R + \eta$):	6.23	55.00

Table 6.1: Robustness test results for multiple scenarios. *AE*

Robustenss tests We have also verified the robustness of the proposed technique with respect to variations on different parameters. In fact the model can be affected in parameters such as R, or measurements can be contaminated by noises η_v, $\eta_{\varphi^{\text{out}}}$. In this case we provide the behavior of the *AE* in multiple cases.

Case 1- (*Reference*): We consider the evolution described in Figs. 6.9 and 6.8 as the reference scenario.

Case 2- (ΔR): We introduced a random perturbation on $R = R^{\text{nom}} + \Delta R$ where we denote R^{nom} the nominal value of the parameter and $\Delta R \in [-0.3, 0.3]$ is distributed following a uniform distribution [5]. Based on the knowledge of R^{nom} the estimation is performed and the *AE* indicators are computed according to (6.31a), (6.31b).

Case 3- (η): We introduced a random noise on the measurement given by $\eta_v = \eta_{\varphi^{\text{out}}}$ is a gaussian distribution with 0 mean and variance 0.001. Based on the knowledge of R^{nom} the estimation is performed and the *AE* is computed for ρ and φ^{out}.

Case 4- ($\Delta R + \eta$): We combined scenarios of *Case 2* and *Case 3*.

The results for all cases are depicted in Figs. 6.10, 6.11. We summarize the results in addition in table 6.1. The results show that in general the estimation in density is very sensitive to R as well as $\eta_v, \eta_{\varphi^{\text{out}}}$ but still its relative error is slow. Flow estimations are less likely to be altered by the noise but the deviations of the R parameter introduce new information that is not taken into account by the estimator. At same time, adding deviations to the parameter R introduces a big source of error since the model reference (steady state of the system) is modified. On the other hand introducing noise in the measurements can be handled by the optimization problem since this is a performance measurement for the optimization.

6.6 Final comments on the chapter

This chapter has addressed the problem of joint density/flow reconstruction over urban traffic networks based on the fusion of multiple sources of information. The recovery process was

[5] For a particular intersection g whenever the $r^{\text{nom}}_{[ij]}$ parameter is modified we highlight that the matrix R keeps the property of row stochasticity $R\mathbb{1} = \mathbb{1}$

achieved by introducing an estimation problem in which the cost function considers the integration of measurements and the network traffic model. We have introduced a relaxation to the original reconstruction problem and we have converted it into a quadratic problem under linear constraints which presents nice properties to be solved. The solutions of the problem show a good recovery in a static as well as in a dynamic situation.

The following chapter will lead to a real time implementation application, where we make use of some of the techniques already studied in this thesis.

Chapter 7

Grenoble Traffic Lab (GTL) Toolbox

Contents

7.1 Overview

Some of the results, typically those on TSP were integrated into a software platform. This chapter summarizes the integration of the developed algorithms and other features created for enhancing GTL, which is an experimental platform dedicated to advanced monitoring of traffic systems. The main idea of the GTL is the creation of a stable tool dedicated to collection, statistical analysis and interpretation of real traffic data. The tool was initially developed with the project as a collaboration between INRIA/Gipsa-lab, Networked Control Systems (NeCS) team, Karrus-ITS (an Inria spin-off), and local traffic authorities, in particular, the Direction Interdépartementale des Routes Centre-Est (DIR-CE) [1] [Oje14]. In this chapter we present a renewed version of the platform published in [CDW+15], [Lad+17], we describe all introduced changes and potentials of the new software architecture. The website can be publicly accessed at `http://gtl.inrialpes.fr/status`.

7.2 Principal changes in GTL

The project GTL-v2 is a renewed version of GTL which consists in a set of modifications and new functionalities introduced with respect to the last version published in [CDW+15]. The

[1] www.dir-centre-est.fr

objective of these modifications for this platform were:

1. Deploy a new graphical interface with more insightful visual analytics for users.
2. Develop a stable set of traffic indicators and pannels classified according to specific user profiles.
3. Develop an open software architecture that reduces the complexity of implementation and deployment of new algorithms.

Three type of data are stored within the GTL. The first one *Raw* data indicates information measured directly from the system (see Section 7.3), *Indicator* data which is information processed by the platform. The type of processing includes imputation of missing samples, filtering, computation of new variables among others. Finally *Prediction* data is another type of processed data produced based on *Indicator* data and that involves time dynamic information.

In order to achieve the aforementioned objectives a new structure at the software layer was proposed. The modifications involve two main components. First, a new structure at the level of databases related to *raw* data, *indicator* and *predictions* was developed, the new version as it will be explained later stores data in a new relational database which facilitates the access to large queries of historical data in a structured and predefined manner with relative rapid access to large amounts of data. Last, we consider a new architecture for data processing, which is based on an event based scheme that allows the execution of instructions in an asynchronous way as well as the development of sequential data processing which is optimal for the application under study.

In order to present formally the new architecture we organize the content of this chapter into three main sections. Section 7.3 presents the collection mechanisms, data flow, and characteristics on the raw data. Section 7.4 presents the way data is stored within the database scheme and the relational content inside the application, this section also presents a series of tools developed in order to access the data from a user point perspective. Finally, section 7.5 summarizes the structure of the GTL platform, focusing on block functionalities and explaining capabilities for future developments.

7.3 Physical layer

The south ring of Grenoble (a.k.a N87 [2]) is a highway enclosing the southern part of the city from A41 to A480. It consists in two carriageways joint with two lanes and the direction east-west is considered in this research. The piece of road taken into account consists of 10 on-ramps and 8 off-ramps starting from Meylan (N 45.205 31°, E 5.783 53°) town and ending at Rondeau (N 45.158 64°, E 5.703 84°) (See Fig. 7.1). The total length is about 10.5 km with an average travel time of 7 min at 90 km/h in a free flow condition. The data collection is achieved

[2]Route nationale 87

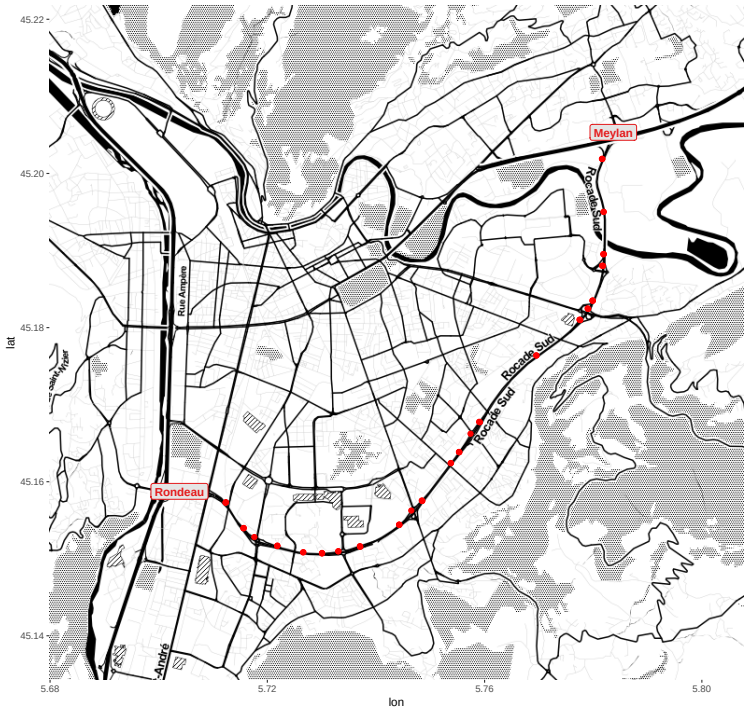


Figure 7.1: Spatial distribution of Station / Collection Points within the Rocade Sud in the direction east-west. Red points correspond to SCP positions where sensors are placed within the map. Map source (Stamen)[3]

throughout sensor stations installed within the highway in each one of the lanes (See Fig. 3.2 in Ch. 3). The highway is managed by the DIR-CE and data is collected within a sensor network that is then recovered in a server centralizing the information. The freeway sensing and actuation equipment includes: 130 wireless magnetic sensors, 4 junctions with in-ramp queue measurements, and 7 variable speed limit electronic panels (70-90 km/h). The instrumentation is composed by *Sensys Networks* VDS240 wireless magneto-resistive embedded within the pavement along fast and slow lanes as well as on/off ramps. Information is measured by pair of sensors separated around 4.8 m and is finally retrieved in Station / Collection Point (SCP) and gathered by Access Point / Station Point (AP). The network architecture consists on 22 SCPs and 19 APs. Fig. 7.2 depicts the sensor names within the network and their physical distribution. The exact positions of the sensors are presented in table (7.1) as well as the asociated SCP. Fig. 7.3 provides a graph representation of the network infrastructure, meaning how each sensor sensor is wired to an AP and additional elements involved in the data transmission process.

Id	`Lane_id`	Location (SCP)	Lane	(Latitude) [DD]	(Longitude)[DD]
01	`Meyl_Fa`	Meylan	Fast	45.205334559197	5.783698046627
02	`Meyl_On`	Meylan	On-ramp	45.205378385709	5.7835613411045
03	`Meyl_Sl`	Meylan	Slow	45.205356504093	5.7836590932541
04	`A41__Fa`	A41	Fast	45.19502	5.7819389533729
05	`A41__On`	A41	On-ramp	45.195015945071	5.7818862944774

Id	Lane_id	Location (SCP)	Lane	(Latitude) [DD]	(Longitude)[DD]
06	A41__Sl	A41	Slow	45.201881	5.781696
07	Tail_Fa	Taillat	Fast	45.2019	5.781605
08	Tail_Sl	Taillat	Slow	45.201891	5.78165
09	DuEx_Fa	Dom U [4] (Exit)	Fast	45.189534219352	5.7819627055226
10	DuEx_Of	Dom U (Exit)	Off-ramp	45.189538	5.781843046627
11	DuEx_Sl	Dom U (Exit)	Slow	45.189536	5.781914682209
12	DuEn_Fa	Dom U (Ent.)	Fast	45.188018	5.781806
13	DuEn_On	Dom U (Ent.)	On-ramp	45.188030164439	5.781608635582
14	DuEn_Sl	Dom U (Ent.)	Slow	45.188022	5.781756
15	GPEx_Fa	Gabriel Peri (Exit)	Fast	45.183487	5.779938
16	GPEx_Of	Gabriel Peri (Exit)	Off-ramp	45.183515	5.779847
17	GPEx_Sl	Gabriel Peri (Exit)	Slow	45.1835	5.77989
18	GPE1_Fa	Gabriel Peri (Ent. 1)	Fast	45.182434	5.779082
19	GPE1_On	Gabriel Peri (Ent. 1)	On-ramp	45.18249	5.778956
20	GPE1_Sl	Gabriel Peri (Ent. 1)	Slow	45.182446	5.779049
21	GPE2_Fa	Gabriel Peri (Ent. 2)	Fast	45.180982	5.777604
22	GPE2_Mi	Gabriel Peri (Ent. 2)	Middle	45.181001	5.777569
23	GPE2_On	Gabriel Peri (Ent. 2)	On-ramp	45.18104	5.777474
24	GPE2_Sl	Gabriel Peri (Ent. 2)	Slow	45.181012	5.7775369999999
25	SMH__Fa	SMH [5]	Fast Lane	45.176375	5.769542
26	SMH__Sl	SMH	Slow	45.176403	5.769505
27	SMEx_Fa	SMH (Exit)	Fast	45.167724163428	5.7591250233136
28	SMEx_Of	SMH (Exit)	Off-ramp	45.167777272384	5.7589927288361
29	SMEx_Sl	SMH (Exit)	Slow	45.167736326857	5.7590894110451
30	SMQu_Qu	SMH (Queue)	Queue	45.166233	5.7574099999999
31	SMEn_Fa	SMH (Ent)	Fast	45.163826	5.755355
32	SMEn_On	SMH (Ent)	On-ramp	45.163866	5.755275
33	SMEn_Sl	SMH (Ent)	Slow	45.16385	5.755315
34	SMHC_Fa	SMH (Ov.)	Fast	45.162407	5.7538079999999
35	SMHC_Sl	SMH (Ov.)	Slow	45.162428	5.7537620000001
36	EyEx_Fa	Eybens (Exit)	Fast	45.157505271532	5.748534728836
37	EyEx_Of	Eybens (Exit)	Off-ramp	45.157560380151	5.7484297521496
38	EyEx_Sl	Eybens (Exit)	Slow	45.157527	5.7484916822091
39	EyQu_Qu	Eybens (Queue)	Queue	45.156261828612	5.7465042797508
40	EyEn_Fa	Eybens (Ent.)	Fast	45.154393	5.744317
41	EyEn_On	Eybens (Ent.)	On-ramp	45.154437	5.744254
42	EyEn_Sl	Eybens (Ent.)	Slow	45.154415	5.744285
43	EcEx_Fa	Echirolles (Exit)	Fast	45.151549	5.7371000000001
44	EcEx_Of	Echirolles (Exit)	Off-ramp	45.151638	5.737041
45	EcEx_Sl	Echirolles (Exit)	Slow	45.151581	5.737079
46	EcQu_Ql	Echirolles (Queue)	Queue	45.150953	5.733072

[4] Domaine Universitaire
[5] Saint Martin d'Heres

Id	Lane_id	Location (SCP)	Lane	(Latitude) [DD]	(Longitude)[DD]
47	EcQu_Qr	Echirolles (Quue)	Queue	45.150985	5.733065
48	EcEn_Fa	Echirolles (Ent.)	Fast	45.150685729034	5.730069317791
49	EcEn_On	Echirolles (Ent.)	On-ramp	45.150753487749	5.730067
50	EcEn_Sl	Echirolles (Ent.)	Slow	45.150714891613	5.7300683411045
51	Echi_Fa	Echirolles (Ov.)	Fast	45.1508224	5.7265848
52	Echi_Sl	Echirolles (Ov.)	Slow	45.15085727098	5.7266347202377
53	EGEx_Fa	Etats Généraux (Exit)	Fast	45.151641433673	5.7218179766865
54	EGEx_Of	Etats Généraux (Exit)	Off-ramp	45.15171	5.7218529766865
55	EGEx_Sl	Etats Généraux (Exit)	Slow	45.151671433677	5.7218339766864
56	EGQu_Qu	Etats Généraux (Queue)	Queue	45.152818596515	5.717605635582
57	EGEn_Fa	Etats Généraux (Ent.)	Fast	45.153959596724	5.715594635582
58	EGEn_On	Etats Généraux (Ent)	On-ramp	45.154001077783	5.7156597901783
59	EGEn_Sl	Etats Généraux (Ent.)	Slow	45.153979108496	5.7156265423279
60	LiEx_Fa	Liberation (Exit)	Fast	45.157283402671	5.7123043877316
61	LiEx_Of	Liberation (Exit)	Off-ramp	45.157336782786	5.7123912711639
62	LiEx_Sl	Liberation (Exit)	Slow	45.157304891394	5.7123336122685
63	LiEn_Fa	Liberation (Ent.)	Fast	45.158711543265	5.7076070000001
64	LiEn_On	Liberation (Ent.)	On-ramp	45.158782814915	5.7076029766865
65	LiEn_Sl	Liberation (Ent.)	Slow	45.158750488944	5.7076056588955
66	Rond_Le	Rondeau	Left	45.158632325951	5.7038303411045
67	Rond_Mi	Rondeau	Middle	45.15867038028	5.7038303411045
68	Rond_Ri	Rondeau	Right	45.158749543271	5.703827

Table 7.1: Network sensor ID's and position details

7.3.1 Data collection

Once data is collected and transmitted to the central server it is stored for further analysis and processing. Data is measured by each pair of sensors and collected by the SCP. A wireless transmission within the frequency of 2.4 GHz is used in this task for AP within a distance of 45 m. For longer distances repeaters are installed to ensure the correctness on the data transmission. The data measured by the sensors individually is called Individual Raw Data (IRD). This data was measured between a non continuous period of time between April, 2013 and September, 2014. From that time to the present, the network has been setup to collect Aggregated Raw Data (ARD) which is characterized in the table 7.2. In the remainder of this chapter, we avoid details in the description of Individual Raw Data, however for further information we invite the reader to follow the discussion in Chapter 3 of [Oje14].

The configuration of the network is such that single detections of vehicles (count, speed, occupancy) measured by the magnetometers are aggregated within the SCPs and then transmitted to the central server through the APs. The aggregation time in this case is 15 s. The communication between the repeaters and the AP is regularly done via fiber optics which

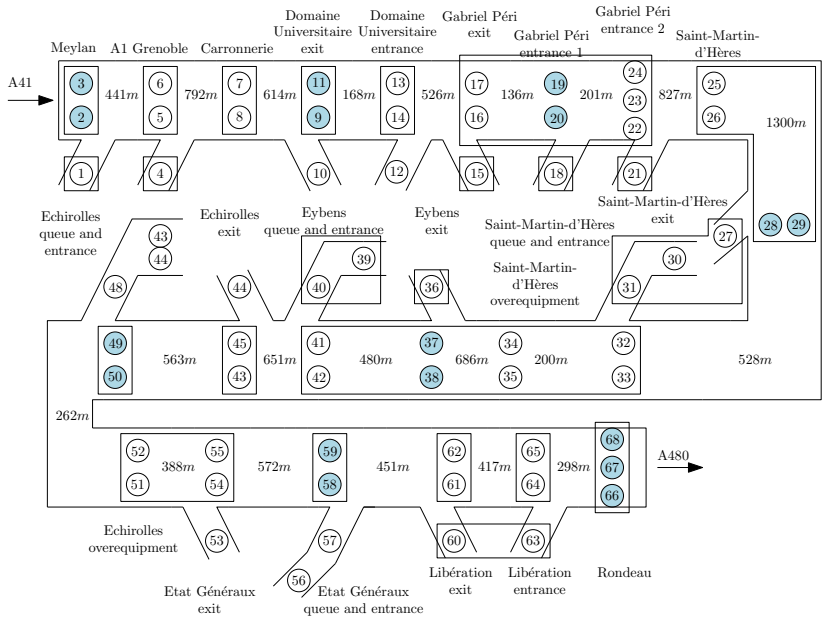


Figure 7.2: Detailed distribution of sensors within the Rocade Sud in the direction east-west. The positions of the 68 fixed sensor points are shown in the stylized map [LCK16].

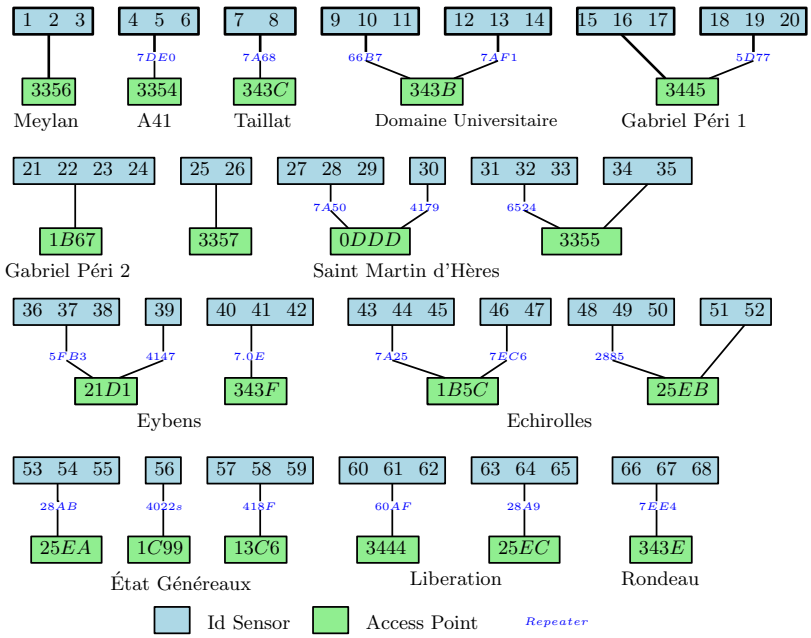


Figure 7.3: Network representation of Station / Collection Points and sensors within the Rocade Sud in the direction east-west. Each sensor identified with a number is grouped in a collection of sensors denominated SCP. Each SCP is connected to an AP in a direct form or throughout a repeater

ensures safety in the data transmission process. In some cases like in the queue `EGQu`, the connection is done with General Packet Radio Service (GPRS). The process of upgrading the full network to fiber technology is up to the date in current implementation. The stability and quality of data received rely on the network architecture and its power infrastructure. GPRS transmission depends on battery supply which is regularly a source of missing data as examined in Chapter 3 in section 3.4.1.

7.4 Database scheme

The database contains information of multiple types, in general the database scheme is a set of tables describing two different sets of data. The first set gathers all samples from the Aggregated Raw Data (ARD) which is information directly sent from the sensors without modification. Second set corresponds to Aggregated Indicator Data (AID), it is formed by all types of processed data obtained from the first set, this include all types of derived indicators (See. section 7.5.2 for dependencies details). Among the AID there are two main sub groups. First subgroup is the Instantaneous Aggregated Indicator Data (IAID) which consists in the set of variables that requires only information at the current time instant to be computed, the second subgroup is the Predicted Aggregated Indicator Data (PAID) which consists in the set of indicators that can be computed based on historical information and information at the current time instant. The table 7.2 contains detailed information about the parameters, and summarize the set of variables involved in each case.

All the database tables presented in the Table 7.2 are implemented on MySQL. Next sections will discuss how the graphical interface displays the multiple indicators (ARD, AID and PAID), data interactions among several components and how the workflow of processing is done to obtain the indicators and predictions in real time.

The primary key for storing information is the sample time of the day (date/hour), the secondary key is associated to geographical location (section / sensor name / origin-destination) and finally the tertiary key is the horizon of time which denotes dynamic information of a specific indicator.

7.5 Software architecture

The new version of this application is designed to process in an independent way multiple tasks regardless the moment of execution. Among the requirements for the design there were the following clear needs:

Historical data processing: Relative big sets of historical data should be processed in

[6] % denotes the percentage of time in which the sensor is occupied during the last 15 s
[7] We refer to section indicators, it means a single value between two SCP
[8] % denotes the percentage with respect to a maximum registered during the last year.

Table- Group `TableName`	**First/ Second/ Third key**	**Sampling**	**Variable**
Raw Data - (ARD) `RawData`	TimeStamp/ SensorID	15 s	Speed [$\frac{\text{km}}{\text{h}}$] Count [$\frac{\text{veh}}{\text{15s}}$] Occupancy [%] [6]
Quality Data - (AID) `MissingSampleHistory`	TimeStamp/ SensorID	30 min	Correctly [$\frac{\%}{\text{h}}$] Partially [$\frac{\%}{\text{h}}$] Not received [$\frac{\%}{\text{h}}$]
Indicator Data - (AID) `SectionIndicator`	TimeStamp/ SectionID	15 s	Section TT [min] [7] Speed [$\frac{\text{km}}{\text{h}}$] [7] Count [$\frac{\text{veh}}{\text{15s}}$] Occupancy [$\frac{\%}{\text{15s}}$] Density [$\frac{\text{veh}}{\text{km}}$] Acceleration [$\frac{\text{m}}{\text{s}^2}$] CO_2 [$\frac{\text{kg}}{\text{km}}$] NO_x [$\frac{\text{g}}{\text{km}}$] Pollution [%][8]
Statistical Data - (AID) `AggregatedNetworkIndicator`	TimeStamp	1 min	DTT [min] FI [$\frac{\%}{\text{min}}$] TSC [min] CL [km] TTT [h veh] TTD [km veh] Avg. Speed [$\frac{\text{km}}{\text{h}}$]
Real Data - (PAID) `OriginDestinationRealIndicator`	TimeStamp/ OriginDestinationID	1 min	RTT [min]
Section Predictions - (PAID) `SectionPrediction`	TimeStamp/ SectionID/ Horizon	1 min	DTT [min] Speed [$\frac{\text{km}}{\text{h}}$] Density [$\frac{\text{veh}}{\text{km}}$]
OD Predictions - (PAID) `OriginDestinationPrediction`	TimeStamp/ SectionID| Horizon	1 min	DTT [min]

Table 7.2: Summary of database schema and variables within the GTL

small periods of time. Multiple functions should be defined over an historical data set such as: Imputation, Filtering and Time aggregation among others.

Realtime data processing: A realtime [9] application capable of processing measurements from sensors is desired. The real state of the network should be measured at a particular time and this should be an input to determine current traffic conditions. Functionalities in this case involve: Data qualification, Imputation, Index computation, Time aggregation, Prediction and Statistics.

Generalized open architecture: Implementation of generic functional blocks for particular tasks is desired. A general framework providing inputs and outputs would allow easy modifications within the maintenance of the platform.

In addition to the previous needs there are some particular functionalities defined for the GTL that contribute to the multiple independent stages and perform several tasks such as:

Quality validation: The quality of Aggregated Raw Data is verified indicating the distribution of received/non received samples during a certain period of time.

Imputation: For those non received samples this tasks corresponds to the mechanisms used to reconstruct missing data within a particular event or series of events.

Indicator estimation: It corresponds to the computation of inferred information concentred in Aggregated Indicator Data from Aggregated Raw Data.

Statistical information: Summary of statistics regarding Aggregated Indicator Data are executed by this task. In particular, computation of mean/median/variance values.

Prediction: It is the process of forecast information of a particular variable and its dynamic behavior in a future time interval.

In order to process the sensor information multiple panels have been recreated.

7.5.1 Graphical interface

Before giving a component and functionality explanation we provide a brief introduction on the User Graphical Interface of the GTL for the sake of comprehension. The GTL graphic interface can be classified in 6 main panels, each one displaying a specific functionality or particular content of the network.

1. *Network current state*: The purpose of this page is to display current status or historic status of the network during a single day. A plot of this page can be seen in Fig. 7.4. Aggregated information of the highway is displayed in the upper part containing:

[9] Realtime is considered as the processing of a series of events in less than the time available to receive consecutive events. For the case of the current application, it is 15 s

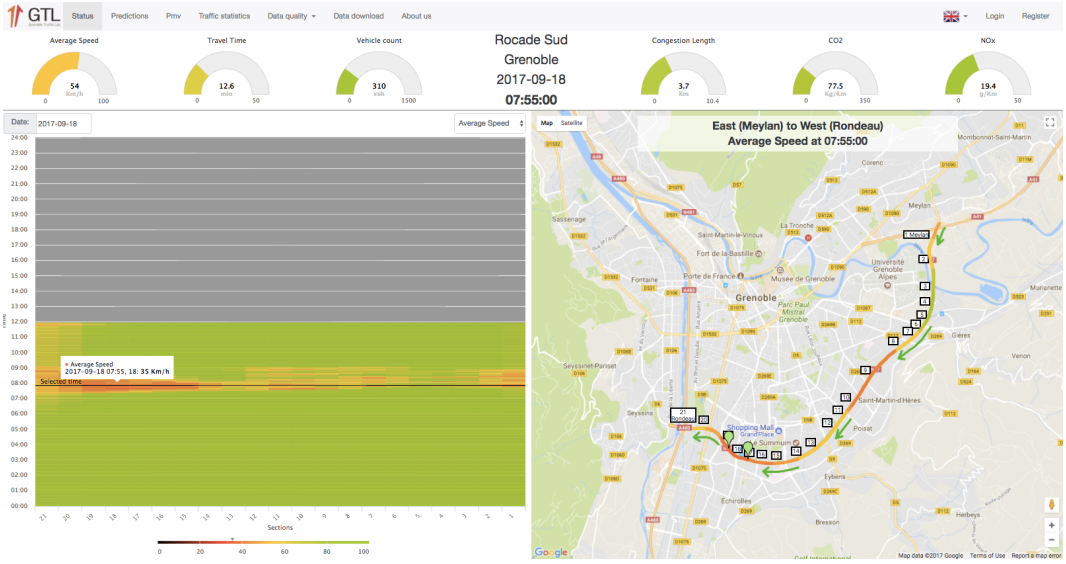


Figure 7.4: Grenoble Traffic Lab Network current status. In the upper part a set of gauges describe general network indicators. Left bottom section displays time-space speed information and the right side depicts a geographical illustration of the network

- Average Speed (AS): It corresponds to the average speed at each location at a fixed time. The average speed is computed as:

$$v^{\text{avg}}(k) = \frac{1}{N}\sum_{i=1}^{N} v(x_i, k) \tag{7.1}$$

  where $v(x_i, k)$ corresponds to the speed at location x_i and time instant k and it is given by the expression in (3.10).

- DTT: Correspond to predicted travel time in minutes according to (3.21), considering the speed of a vehicle measured from the sensors at different times and points in the space. The DTT is computed along the full journey in the highway starting at `Meylan` and finishing at `Rondeau`. We refer the reader to the section 3.5 in chapter 3 for further discussion on this computation.

- Vehicle Count: It is a count of the total number of vehicles that crossed the highway during the last time interval of measurement. Given the failures in the sensors and all the process of flow reconstruction throughout imputation algorithms this indicator is an estimator. Consider an imputed sample of count of vehicles of each one of the sensors then the total count at each SCP denoted *loc* is given by

$$f^{\text{loc}}(k) = \sum_{j\in\text{loc}}^{N} f_i(k). \tag{7.2}$$

  Finally the total count considered within the computation of flow is given by:

$$VC(k) = \sum_{loc} f^{loc}(k). \tag{7.3}$$

The interval of time is normalized such that the vehicle counts are given for the last interval of time, thus 15 s.

- Congestion Length (CL): It is an estimation of the length of congestion created within the Rocade. The length of the congestion can be exploited from the relationship (7.14). We can split computations in two regions based on the state of the cell (i.e. free-flow, congested as:

$$CL_i(k) = \begin{cases} \frac{f_i(k)}{v_i(k)} l_i l^{\text{avg}} & v_i(k) \geq v^{\text{free}} \\ \frac{w_i \rho_i^{\max} l_i l^{\text{avg}}}{v_i(k)+w_i} & v_i(k) < v^{\text{free}}, \end{cases} \tag{7.4}$$

where l_i corresponds to the length of the section v^{free} is the highway speed limit and l^{avg} is the average length of a vehicle within a network. The later one can be obtained from histograms of lengths of vehicles measured by the data, and it was found that the average length is about $7m$ each vehicle considering inter vehicle distance. It is worth to consider that this factor ignores the headspace between vehicles, in that sense the parameter is an underestimation of a more accurate value. By considering the total sum of congestion a general highway indicator can be computed as

$$CL(k) = \sum_i CL_i(k). \tag{7.5}$$

$\rho_i^{\max}$ and w_i are parameters that can be tunned from the fundamental diagram. We provide details on how this calibration can be performed in A

- CO_2: Corresponds to an aggregate estimation of the CO_2 emissions computed from aggregated information within the Rocade. General emissions are found to be proportional with respect to the fuel consumption. (See Chapter 20, section 20.1.3 in [TK13]). Due to the chemistry of the combustion there exists a direct unitary proportional relationship between the fuel consumption and the emission of CO_2. Specifically the amount of mass emitted is 44 (atomic units) over 12 (atomic units) times the mass of carbon contained within the fuel ($44/12 \approx 3.7$). More accurately, the combustion of octane follows this reaction:

$$2\,C_8H_{18} + 25\,O_2 \longrightarrow 16\,CO_2 + 18\,H_2O$$

This relationship considers the conversion of all carbon into CO_2 which is not completely the case in reality. Other derivates and products emerge from the combustion reaction (such as CO or soot). Typical conversion factor for Gasoline 98 RON[10], which is a common octane rating for fuel dispensers within Europe is $\kappa = 2.39\,\text{kg}CO_2/\text{l}$. In the case of Diesel $\kappa = 2.69\,\text{kg}CO_2/\text{l}$. These relationships are usually used to transform Fuel Consumption (FC) in terms of CO_2 emissions.

The Fuel Consumption denotes an estimate of the amount of energy expressed as distance traveled obtained from a particular amount of fuel. It can be computed based on estimation of instantaneous consumption extracted from the power used to thrust a vehicle. The required power is adjusted within the fuel consumption

[10]RON corresponds to the Research octane number denoting the octane rating of the fuel

Parameter	Meaning	Car	HDV[11]
P_0	Idle power	3000 kW	1000 kW
A	Effective cross section	2 m^2	1000 m^2
m	Vehicle mass	1590 kg	25 000 kg
δ	Air density	1.3 $\frac{kg}{m^3}$	
c^d	Drag coefficient	0.3	
w^{cal}	Fuel energy density	39.6×10^6 $\frac{J}{l}$	
μ	Friction coefficient	0.2	
g	Gravity	9.81 $\frac{m}{s^2}$	

Table 7.3: Vehicle and physical parameters for Fuel Consumption estimation

rate through the *energy density* of the fuel defined as $w^{\text{cal}} = \Delta W^{\text{chem}}/\Delta C$ which denotes the relationship between the change in energy of the fuel ΔW^{chem} and a particular volume ΔC. This energy is transformed to mechanical energy through an engine of certain efficiency expressed as $\gamma = \Delta W^{\text{mech}}/\Delta W^{\text{chem}}$. Both components allow the expression of the power:

$$P = \frac{dW^{\text{mech}}}{dt} = \gamma w^{\text{cal}} \frac{dC}{dt}. \tag{7.6}$$

It is possible to calculate Fuel Consumption normalized each 100 km when driving the whole stretch of road in the same operating mode (e.g. continuous acceleration). For each one of the cells the consumption is given by

$$FC_i(k) = 100\text{km}\frac{dC}{dx} = 100\text{km}\frac{P}{v_i(k)\gamma w^{\text{cal}}}. \tag{7.7}$$

In the previous expressions P represents the engine power which designates the power required to overcome the idle state of the vehicle P_0 and the dynamic component given by the state of the vehicle. The overall term of power can be expressed as a function of the speed and the acceleration

$$P(v,a) = P_0 + \max\big(0, vF(v,a)\big), \tag{7.8}$$

where

$$F(v,a) = ma + (\mu + \phi)mg + \frac{1}{2}c^d\delta A v^2,$$

Finally combining (7.7), (7.8) it is possible to obtain the fuel consumption as,

$$FC_i(k) = \frac{P_0 + \max\left(mv_i(k)a_i(k) + \mu mg + \frac{1}{2}c^d\delta A v_i^3(k), 0\right)}{v_i(k)\gamma w^{\text{cal}}} \tag{7.9}$$

where $v_i(k), a_i(k)$ denotes the corresponding speed, acceleration between two SCP computed as 3.6. The parameters are given in table 7.3. In this case we make use of a discrete approximation for the term a which corresponds to the convective derivative of the speed term which allows to transform the acceleration from eulerian framework to a lagrangian coordinate framework [Bat00, p. 72]

$$a_i(k) = \frac{v_j(k) - v_j(k-1)}{T} + v_j(k)\frac{v_{j+1}(k) - v_j(k)}{l_j}. \tag{7.10}$$

	β_1	β_2	β_3	β_4	β_5	β_6
Petrol car						
$a \geq 0.5 \frac{\text{m}}{\text{s}^2}$	6.19e-04	8.00e-05	-4.03e-06	-4.13e-04	3.80e-04	1.77e-04
$a < 0.5 \frac{\text{m}}{\text{s}^2}$	2.17e-04	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00
Diesel car						
$a \geq 0.5 \frac{\text{m}}{\text{s}^2}$	2.41e-03	-4.11e-04	6.73e-05	-3.07e-03	2.14e-03	1.5e-03
$a < 0.5 \frac{\text{m}}{\text{s}^2}$	1.68e-03	-6.62e-05	9.00e-06	+2.50e-0.4	2.91e-04	1.2e-04
HDV						
	8.42e-03	0.00e+00	0.00e+00	0.00e+00	0.00e+00	0.00e+00

Table 7.4: NO_x emissions calibration parameters. Taken from [IBL06]

In order to obtain an average of production in all the highway we cumulate information as follows

$$\text{CO}_2 = \frac{1}{N}\sum_{i=1}^{N} \text{CO}_{2i}, \quad \text{CO}_{2i}(k) = FC_i(k)l_i \tag{7.11}$$

In Figs. 7.5a, 7.5b an illustration of the CO_2 production is illustrated for two particular periods. First we analyze temporal behavior during week days and similarly on weekends. As expected the amount of emissions is directly related to traffic congestion. Fig. 7.5e illustrate this effect also when correlating increments on the CO_2 index in terms of the Fluidity Index (FI). The FI as it will be explained later intends to present the amount of congestion within the highway (See equation (7.15)). This test analyzes 6 months of data between June, 2015 and ending November, 2015.

- NO_x: A second source of emissions which is highly important is NO_x [Yan+16]. During combustion process, the nitrogen bound in the fuel is released as a free radical and ultimately forms free N_2, or NO. Fuel NO_x can contribute as much as 50% of total emissions when combusting oil and as much as 80% when combusting coal. In this case, there is no direct proportion between the production of NO_x and the combustion reaction and it depends on how radicals react within the environment. For this case, we follow the approach studied in [IBL06]. In the study a set of vehicles with determined qualified standards was tested under particular conditions of speed and acceleration, and particular piecewise polynomials are defined to determine the consumption

$$\begin{aligned}\text{NO}_{x,i}(k) = \max\big(0,&\beta_1 + \beta_2 v_i(k) + \beta_3 v_i^2(k) + \\ &\beta_4 \dot{v}_i(k) + \beta_5 \dot{v}_i^2(k) + \beta_6 v_i(k)\dot{v}_i(k)\big),\end{aligned} \tag{7.12}$$

where the parameters are given in table 7.4. Thus the aggregated value of NO_x can be computed as:

$$\text{NO}_x(k) = \sum_{i=1}^{N} \text{NO}_{x,i}(k) TT_i(k) \tag{7.13}$$

In this case the calibration of the model is given in [g/s], we consider a multiplication factor TT to quantify the amount of NO_x that will be emitted for a particular section. Finally the total emission is the sum of all values produced on each one of the sections.

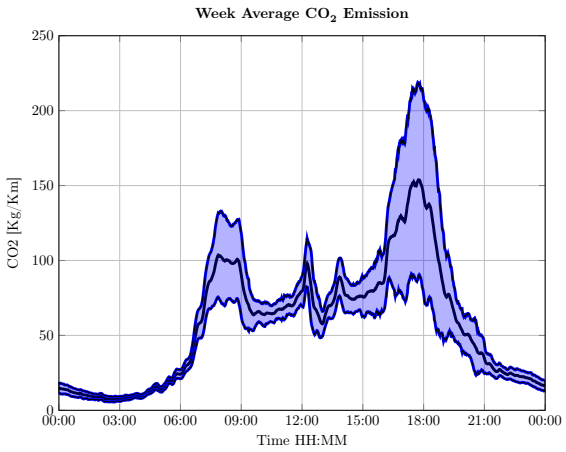
For the sake of illustration in Figs. 7.5c, 7.5d we depict the average of NOx emissions computed for 6 months of data (June, 2015 and ending November, 2015. It can be seen the strong correlation between the NOx production and the congested regimes. In order to illustrate that we consider the correlation between the FI (7.15) index representing the fluidity of a highway and the NOx emission, it can be seen that a high peak is produced with small variations from of the free flow speed which is regularly achieved in nominal traffic conditions.

In the left bottom part (in Fig. 7.4) a heat map indicator displays space-time distribution of multiple variable of interest. In this particular case we consider the following variables:

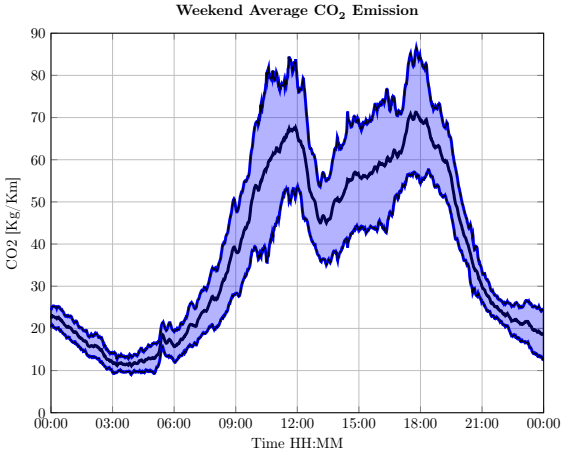
- Raw (Speed / Count/ Occupancy): It displays measurements of each one of the variables as received collected at each one of the Access Point / Station Point. In cases where the samples are not received no information a particular gray color highlight this feature. Information is aggregated each 2 min for displaying purposes.
- Section Speed: Speed at each section is determined by relationship (3.10). This indicator is aggregated each 2 min as well.
- Flow: Flow at each one of the sections is computed.
- Density: Density is computed as the relation between average flow within a section and average speed.
$$\rho_i = \frac{f_i}{v_i} \tag{7.14}$$
- CO_2: Emissions of carbone dioxide CO_{2i} are computed for each section (See. (7.11))
- NOx: Emissions of nitrogen oxides NO_{xi} are computed for each section (See. (7.12))

Finally the right part of the screen displays a map containing the highway and the city. Over the map the highway is highlighted giving colors according to the actual state of, the variable displayed corresponds to the Average Speed.

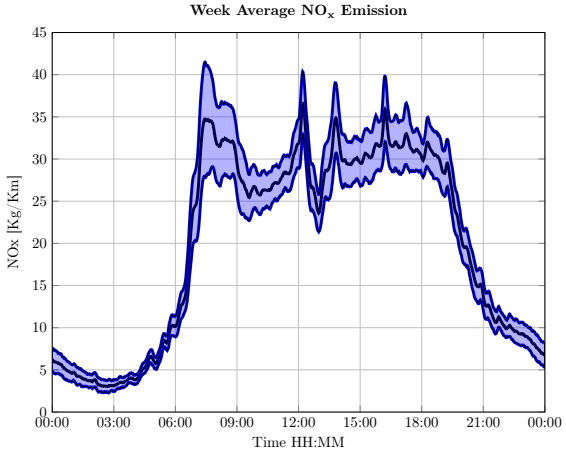
2. *Prediction*: The module is in charge of computing the predicted DTT between all possible combinations of origin destination within the network. It performs the computation according to the multi clustered prediction explained in chapter 5. The module is composed by two main stages. The first stage predicts speed for each one of the sections with the multi clustered prediction and fusion, this information is used to compute the DTT and finally this time is considered as current measurement that will be combined with historical information in order to produce the desired forecast. Figure 7.7 explains the details on what is displayed on the right bottom panel of the interface and the relationship with the left panel of the interface. Figure 7.6 displays the graphical interface used to show the forecasts. Three main windows can bee seen on this tab. On the top left window, input information for the prediction parameters such as prediction date, time (Departure time / arrival time), origin (From) and destination (To) are provided.



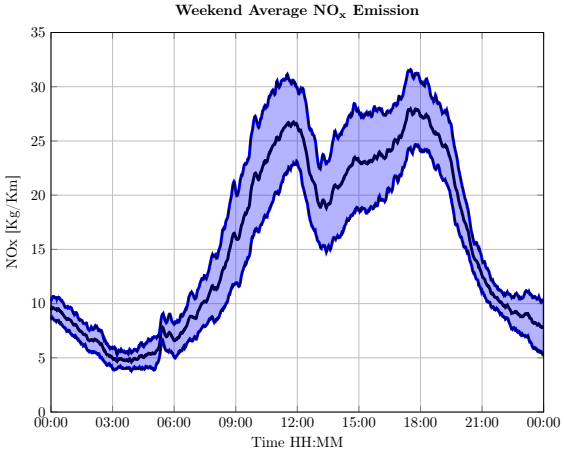
(a) Mean/Standard deviation - week CO_2



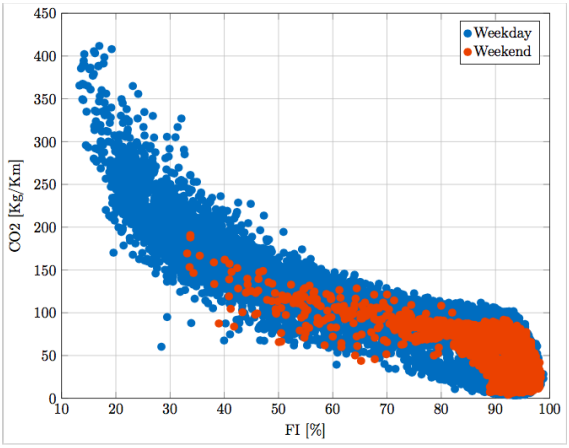
(b) Mean/Standard deviation - weekend CO_2



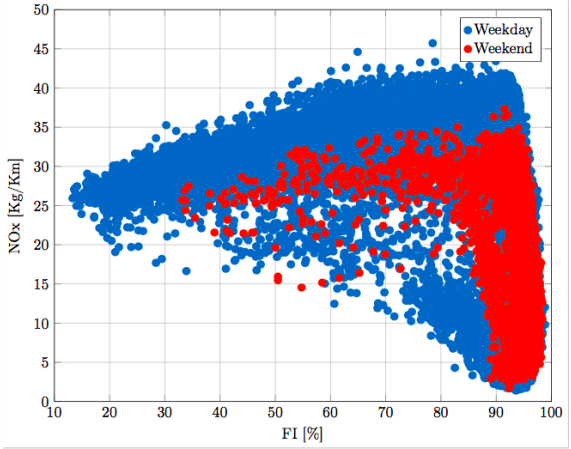
(c) Mean/Standard deviation- week NOx



(d) Mean/Standard deviation - weekend NOx



(e) Correlation CO_2 vs FI



(f) Correlation NOx vs FI

Figure 7.5: CO_2, NOx emissions data analysis

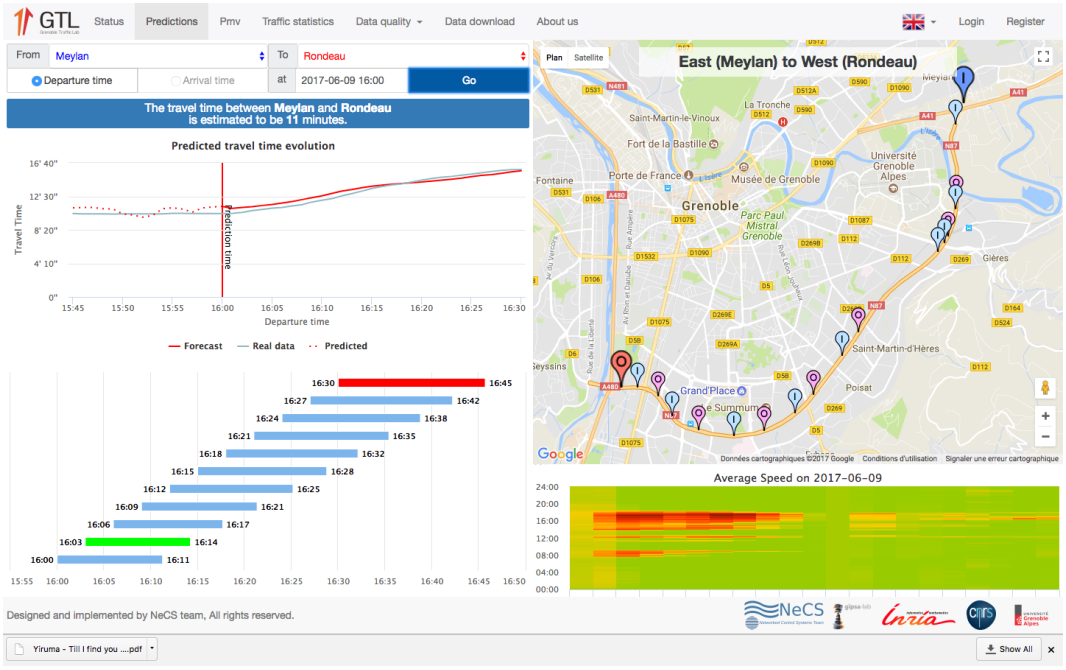


Figure 7.6: Prediction graphic interface. In this case a particular prediction is launched at 09 June, 2017 - 16h00

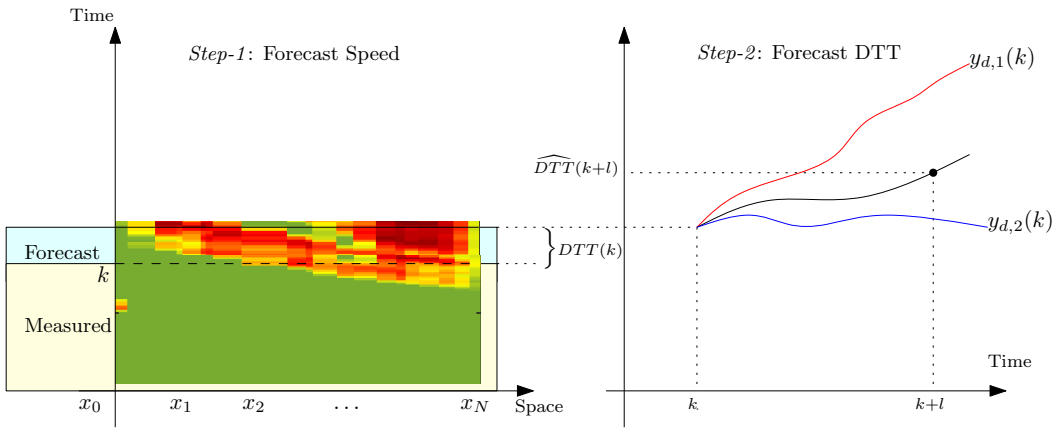


Figure 7.7: DTT prediction and its relationship between the information panels within the graphical interface. Left side displays the speed profile during the current day. Right side shows the DTT prediction based on current day data and historical data

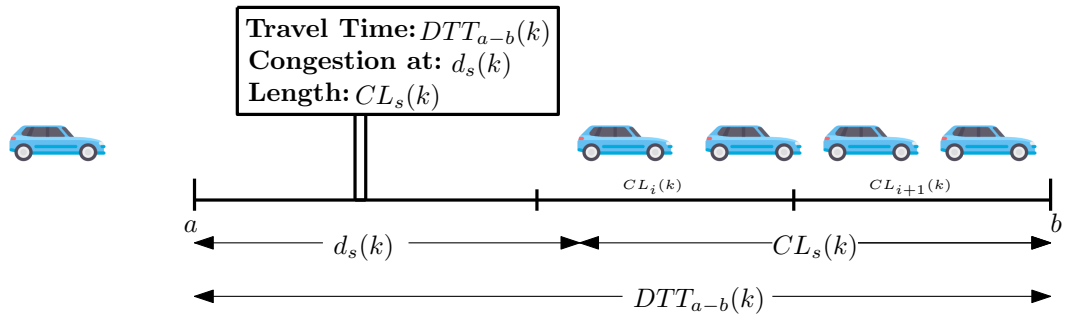


Figure 7.8: Speed panel indicators and displayed information. The panels display three main indicators DTT_{a-b}: travel time from the nearest origin to destination, d_s: and CL_s which represents the length of the congestion for a specific subset of sections

Once the prediction is executed, the result of the forecast can be visualized on the left bottom window. Two graphs are displayed, the first contains a time series proyection in future time. This short term forecast is by default computed in a horizon of 45 min from the current time. For longer horizons the combination of the real time information and historical information is not executed and only the historical mean of the corresponding day of the the week is computed and the long term prediction (time horizon greater than 45 min) is displayed. A second bar graph depicts the DTT for different departure times and provides in colors (green,red) to suggest the best and worst departure times, its purpose is to provide a decision support tool for users of the highway. Finally on the right bottom side a speed profile containing the measured speed and the forecasted speed is depicted (See Fig. 7.7 for details).

3. *Variable Speed Panels*: The speed panels are created for traffic operators. The objective of this functionality is to provide simple and readable information of current state of the highway about DTT, CL and to provide decision support tools for institutions and operators. The panel displays information regarding three indicators as seen in figure 7.8. The first indicator corresponds to DTT between the nearest origin a up to the end b of the highway, the second indicator corresponds to the length of the segment CL_s indicating the length of the downstream congestion and finally $d_s$the distance to the congestion. In cases where there is no congestion the values of CL_s and d_s are omitted and the message "circulation fluide" is displayed. In cases where the panel is embedded in the middle of the congestion only d_s is omitted. It is possible to check measurements at 10 different spots along the highway which correspond in general to the places where onramps are located. The motivation for this development comes from an inquiry received from the DIR-CE in order to suggest the optimal installation of new panels acquired recently by the institution.

4. *Statistics*: Several statistics are computed in the system, these contains a summary of the network state during several days within the past. The main purpose of the statistics module is to present Aggregated Indicator Data computed in the *Network current state*. In particular a calendar containing profiles of multiple indicators is available. Three additional indicators are presented in this section, the first one is the FI defined as

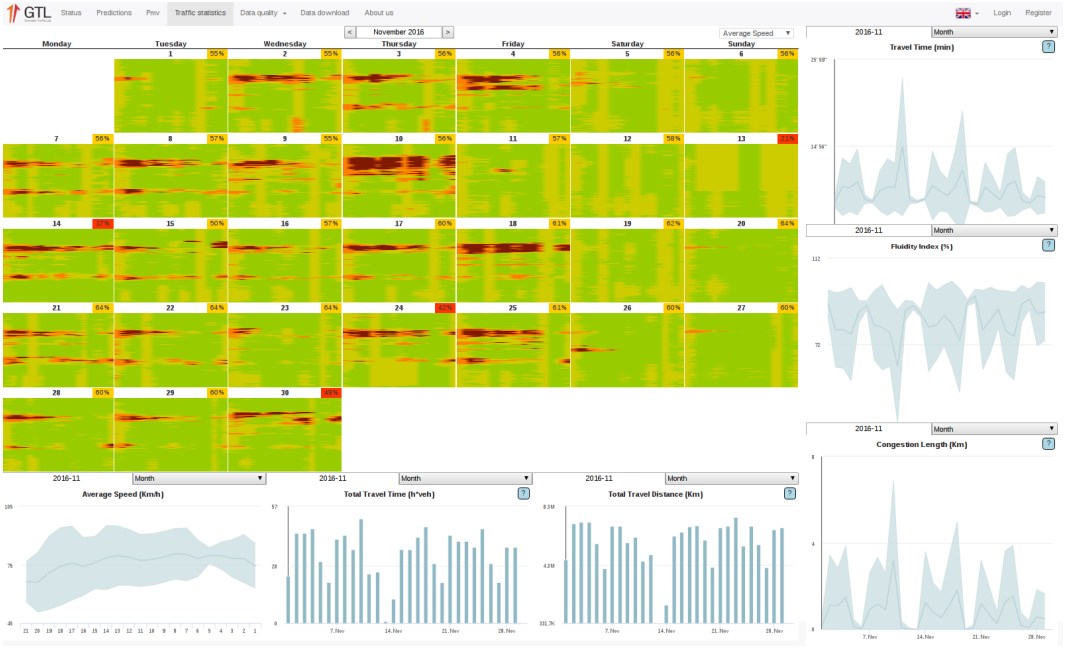


Figure 7.9: Statistics graphic interface. The calendar profile displays miniatures of the speed profiles in time-space domain. Right column contains indicator such as FI, TTT, TTD

$$FI = \frac{v^{\text{avg}}(k)}{v^{\text{free}}} 100\% \tag{7.15}$$

which represents the level of fluidity within the highway. The second indicator is the TTD given by

$$TTD = \sum_k \sum_i \rho_i(k). \tag{7.16}$$

Finally we provide the TTT computed as

$$TTT = \sum_k \sum_i f_i(k). \tag{7.17}$$

Each one of the previous indicators is computed for a single day $1 \leq k \leq 5760$ for all the sections. These indicators are displayed on the right side of the graphic interface as it can be seen in 7.9. On left part the website a calendar of heatmaps is displayed, the information shown contains indications for each one of the sections in particular, speed, flow, density and emissions (CO_2 and NOx). The right column panel displays indicators of CL, DTT, FI which can be visualized during a day/week or a specific month. The values for week and month represent the average values of the samples during the day. The panel in the bottom part of the page shows , section speeds, TTT, TTD which can be visualized during multiple periods of times (day/days/weeks/month).

5. *Data Quality*: This module performs an evaluation of the quality of data within the network. Multiple data errors can be detected in the network (See. 3.4.1). Nevertheless, errors are difficult to be traced when no specific failure condition is notified. In order to reduce uncertainty, Aggregated Raw Data received in the network is classified based on

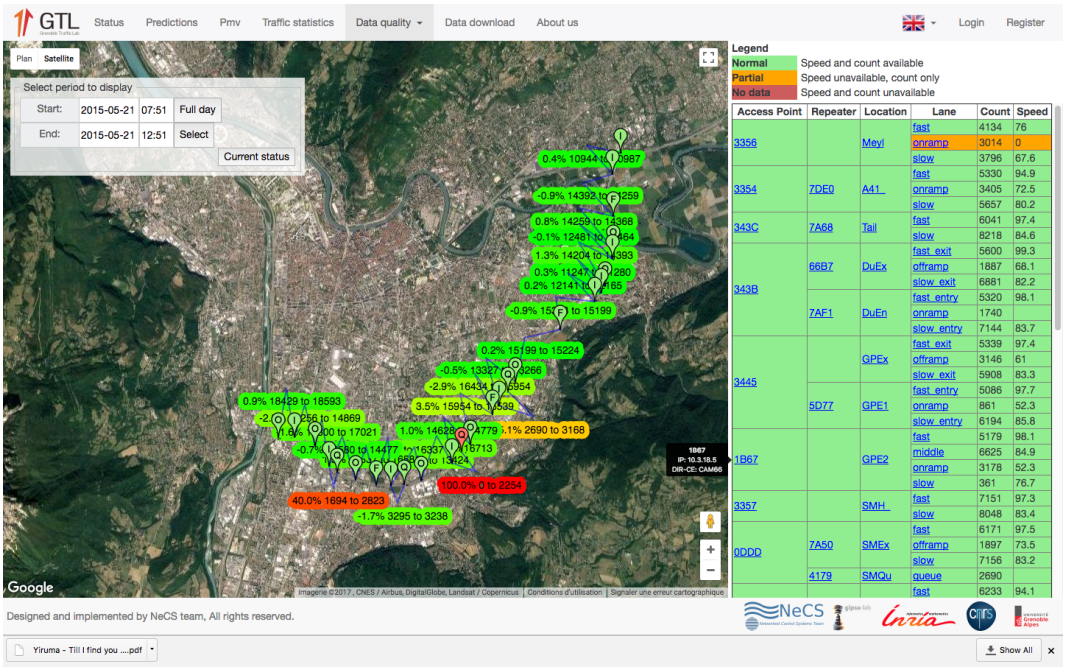


Figure 7.10: Flow conservation data quality graphic interface. Flow balance is performed over a particular period of time. In this case the day 21 May,2015 was considered for examination.

Figure 7.11: Sensor data quality graphic interface. A single or a group of sensors can be examined for a range of dates or a particular day in order to study the characteristics of the data. In this case raw measurements are studied during the month November, 2015

Sample type	Characteristic	Speed	Count	Occupation
Correctly received	Correct	> 0	> 0	> 0
Partially received	No cars No speed	< 0 < 0	0 > 0	0 ≥ 0
Not received	No data	< 0	< 0	< 0

Table 7.5: Error code for raw data in GTL. The table display the range of values of each variable and how data is classified

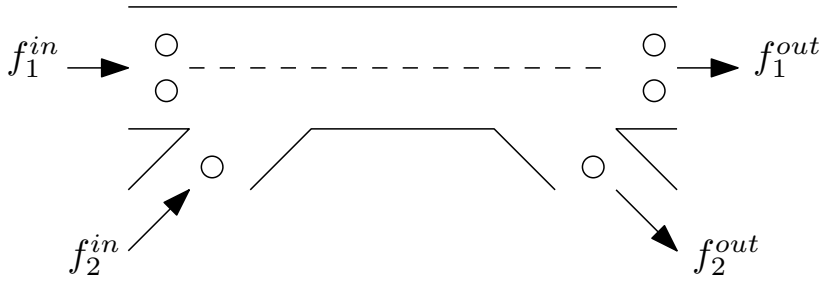


Figure 7.12: Single cell flow conservation. Averaged flow is taken into account at the input of the section and the output of the section, over a sufficient long period of time flow should be preserved

an error code presented in table 7.5. The error code classifies the samples among three big groups which represent a criteria for decision to decide imputation of values. Among quality statistics, it is computed the percentage of *Correctly received*, *Partially received* and *Not received* samples. This classification allows to detect and classify operating sensors sensors within the network. Three main panels contain statistical information and details regarding the quality of the data, the first one contains the flow conservation (See. Fig. 7.10). In this one, the Station / Collection Point and the sensors are placed in the map, and indication of the percentage of conserved flow is displayed during a particular period of time.

Qualification of data is measured then in terms how precise the flow measurements are. To provide a more accurate idea let us examine the figure the figure 7.12. In this case f_i^{in} denotes all incoming average flows to the cell. Over a sufficiently long period of time the cumulated flow should preserve conditions of mass conservation since no vehicles are stored within the network. The computation of data flow conservation expressed in terms of in flow and out flow to the section as

$$E_i = \frac{f^{\sum_k \text{in}(k)} - \sum_k f^{\text{out}}}{\sum_k f^{\text{out}}(k)} \tag{7.18}$$

The flow conservation indicator is important in order to perform validation of quality from the sensor to the server. Thanks to the classification given in table 7.5 errors can

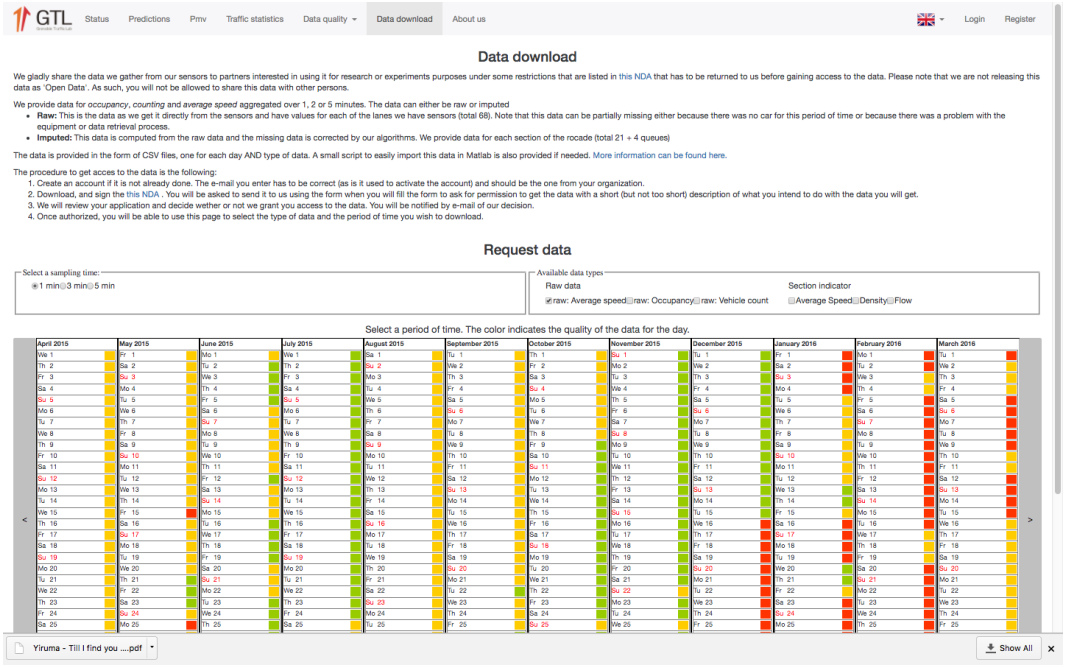


Figure 7.13: Download data quality graphic interface. A single day or a set of dates can be downloaded directly from the website. This page gives a color indicator and the percentage of *correctly received* samples in order to give an idea of the quality of the data measured during a specific date

be classified at the level of a sensor or the level of an AP. The table on the right side of figure 7.10 displays the availability of sensors marking them with colors. For the period time of analysis the table will take the three possible colors based on the majority of data received within the time interval (Green for *Correctly received*, Orange for *Partially received*, Red for *Not received*). In addition to this interface, a second panel is displayed in Fig. (7.11). In this case the analysis over a single sensor or a group of them can be performed, the interface allows to select up to 5 sensors on the right side table. Once sensors a re selected three main graphics are displayed. The first one (top-left) based on a particular period of time (day/weeks/month) access to the sample classification within that period, the second one displays raw data comparisons for a specific selected date and variable and finally the plot on the bottom part shows sensors availability including windows containing up to 6 months of data.

6. *Data Download*: This section is dedicated to data retrieval. Request of data can involve several types of data and several aggregation times. In this case the platform has been designed so the user can get raw data (speed, count, occupancy) or processed data (section speed, section density, flow) with time aggregations of 1 min, 3 min or 5 min. The system will ask to create an account, the process involves approval from the NeCS team. The figure 7.13 displays the content of this page.

The main set of pages and computations have been explained. For the remainder, the

software architecture of the system is detailed providing information on the implementation and execution of computation in realtime.

7.5.2 Components

The Grenoble Traffic Lab is implemented on MATLAB 2015b and it makes use of a license of the Statistics and Machine Learning Toolbox. Fig. 7.14 represents the exchange of data among the different databases. In order to process data the information is retrieved from the remote server, the process is as follows: once information is measured by the sensors it is stored in files within a server located at the DIR-CE. The information is pushed into servers placed at INRIA where there exits a single database containing multiple data tables as indicated in table 7.2. The sample received in the system is provided to the realtime MATLAB application and at same time stored in the *Raw* data table (`RawData`). The sample is imputed in case it is necessary and the value of the imputed samples is stored in *Imputed* data table (`SectionIndicator`). Starting from a clean sample, the Aggregated Indicator Data is computed, and it is also stored within the same data table. At the beginning of the the day the system recovers historical data from the *Imputed* data table for the past 6 months and stores the content in memory. Based on this information and real time information a prediction for (speed, density and DTT) is computed within the system each minute and stored in the *Prediction* data table (`SectionPrediction`,`OriginDestinationPrediction`).

A realtime indicator for the real measured DTT is computed based on speed measurements with one hour of delay in order to be compared with predictions of the corresponding variable, the information is stored in the *Real* data table (`OriginDestinationRealIndicator`). Each 30 min quality of the system is computed from raw data, however, in order to facilitate the access to historical data quality this information is also replicated within the *Quality* data table (`MissingSampleHistory`), the same procedure is carried out when extracting statistics of some indicators. In this case the information is stored in the *Statistics* data table(`AggregatedNetworkIndicator`).

The realtime application is implemented as a system processing a series of events. A general component denominated `TimeController` is responsible of determining the historical time window or time instant the data is being extracted or analyzed, as well as sending orders to the corresponding components which store these data. On the other hand, components denominated `History` and `Indicators` [12] are created to manipulate historical data or current day data correspondingly. `Qualification` determines the quality of the data, `Imputation` is a special component designed to impute data while `Prediction` is a component designed to perform short-term forecasts [13].

1. `TimeController`: It is an object dedicated to regulate time and synchronize and launch

[12] The component has been implemented with an extension `ProcessedIndicator` which consists in a set of functions to compute sophisticated indicators FC,CO_2 ,NOx

[13] By typing the command `help` and the name of the object the user can access to full documentation of the properties and methods of each class

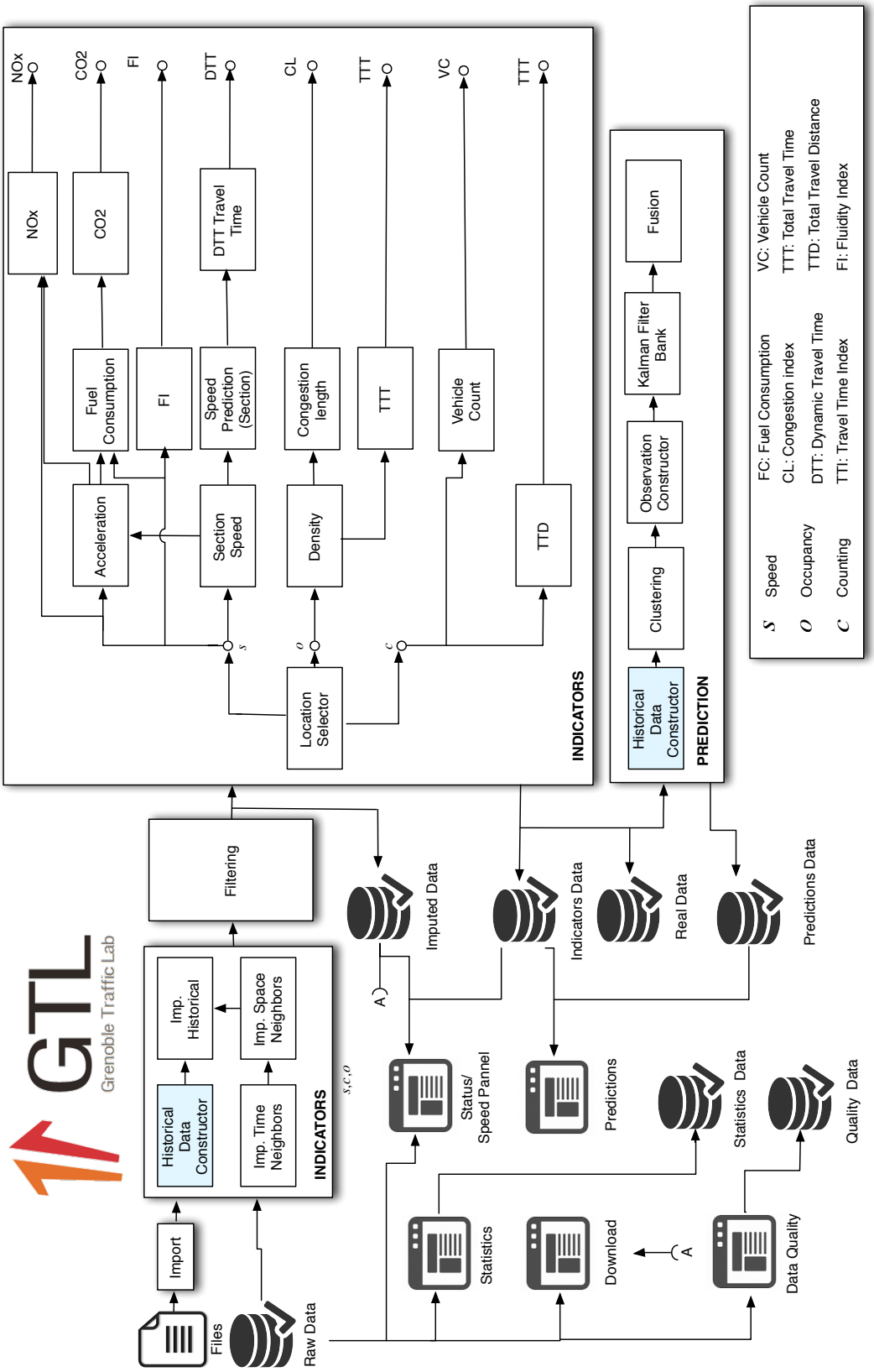


Figure 7.14: General schema of GTL and interaction with databases

a set of events within the application. The component contains a series of properties and methods to validate the execution of an event with the GTL. The events within the system can be classified as synchronous events with different time executions 15 s, 1 min, 30 min, 1 h, 24 h. At each time step $k = 15$ s this component verifies the tasks to be executed and commands to the corresponding components in order to launch the computations. The table 7.6 joins all the events and order of execution along the day.

2. `Qualification`: This component is in charge to count and compute the amount of *correctly, partially and non received* samples within the network for certain periods of time. It points stores a pointer of information to the raw data. When indicated by the `TimeController` the object computes the number of correct, partially and non received samples.

3. `History`: This object constructs historical data information. Within the GTL historical data is represented as a multidimensional array in which the rows correspond to the temporal day information, columns correspond to spatial information and the third dimension the historical information given in days. The component is able to construct representations of historical data for the following variables

   Raw data: Speed, count or occupancy

   Section data: speed, count, density

   Origin destination: DTT

4. `Indicators`: The object constructs single day data information. Within the GTL single day data information corresponds to information that can be computed at current time instant based on current time information. The indicators computed by this component are listed as follows

   Section data: speed, acceleration, count, density, CO_2, NO_x, FC

   Global network information: Average Speed, TTT, TTD, FI

5. `Imputation`: The technique of imputation described in Chapter 3 is implemented within this component. The information requires imputed historical data contained in an `History` object. The information to be imputed is contained in a `Indicator` object if the imputation is performed in real time or a `History` object if the data to be imputed is done as a batch process.

6. `Prediction`:The technique of data clustering explained in Chapter 4 and the multi clustered based technique explained in Chapter 5 are implemented within this component. The predicted indicators for this case correspond to:

   Section prediction: Speed, density

   Origin destination: DTT

Event	Time Execution	Component	Variable Computed
Load History	24 h	`History`	Speed Count Occupancy DTT
Sample arrival	15 s	`TimeController`	Speed Count Occupancy
Sample Imputation	15 s	`Imputation` `Indicators`	Speed Count Density
Indicators	15 s	`Indicators`	Section Speed Count Occupancy Density CO_2 NOx CL
Predictions	1 min	`Prediction` `Indicators`	Section Speed Section Density DTT
Real Indicators	1 min	`Indicators`	DTT CL FI
Quality	30 min	`Qualification`	# Correct # Partial # Missing
Statistics	30 min	`Indicators`	FI [14] CL TTT TTD

Table 7.6: Real time event list within GTL

7.6 Final comments on the chapter

The chapter has presented the main architecture of the GTL. The main description contains new changes with respect to the previous version published in [CDW+15]. We provide further details in computations and a new system to process events in real time as fast as they arrive to the system. The new architecture is maintained within the INRIA servers based on GitLab a web based repository manager. Further information and details on the application can be accessed within the MATLAB documentation of the application.

We propose an event based processing scheme which allows fast computations and guarantees integrity on the data processing. The dependencies between components helps to identify sources of failure and makes easy the development of future capabilities of the platform. This event based schema can be applied to the historical or the current day data, either for one or multiple variables within the system at same time.

A fully functional interface is publicly open `http://gtl.inrialpes.fr/status` and multiple users can make use of the portfolio of functionalities such as travel time predictions, indicators and computations. The platform is also intended to be a real case scenario and a pedagogical platform for traffic systems.

Chapter 8

Conclusion and perspectives

8.1 General summary

In this thesis we addressed the problem of TSP in highway traffic networks and TSE in urban networks. We have proposed solutions that differ each other from methodological, computational and operational point of view. We started the discussion by positioning this dissertation in the actual state of the art and we have centered the discussion in the short-term DTT forecasting at the beginning. At the end a density reconstruction problem is examined in order to exploit fusion integration of heterogeneous sources of traffic into the TSE problem. The proposed research intends to provide better decision support tools for users of the traffic infrastructure.

The discussion on the DTT short-term forecasting problem is detailed in the framework of a real-time application. Data workflow from the network sensor up to the cleaning process was detailed and computation of the DTT indicator is provided. Side effects of the intermediate process are analyzed and their impact over the DTT computation is also presented.

It was shown experimentally that traffic regimes are observable in historical patterns which are a key tool in order to perform short-term forecasts. Furthermore, a suitable technique (e.g. K-means) can lead to a separation of this regimes and a correct assignment. A full parametrization of the technique is provided in terms of initialization, and detection of the number of partitions in a particular historical dataset.

Two short-term forecasting methods are proposed for DTT. The proposed methodologies perform the prediction based on the fusion of individual forecasts obtained from clustered time series. The study is focused on a real time application where new measurements of data can be received in short periods of time ([15sec $\sim$ 1 min] in the presented case), therefore the main benefit of this work is to perform low computational workload to obtain online DTT forecasts under real time traffic conditions.

The approach followed is based on data-driven models which are considered better to describe micro dynamics and stop and go waves relevant in the data. In this case more shorter step heads for the forecasts (1min step) are provided with respect to previous approaches [Oje14]. From Figs. 5.9, 5.10 and tables 5.2, 5.3 it can be observed that PSFM provides better results than ECFM. We provide general comparisons with a naive historical mean predictor since it can be a general pattern of reference for future research works. Based on

the review [VKG14] it was found a big amount of contribution and criteria to select particular techniques for comparison are not easy to define.

The full real-time platform GTL was renewed in terms of appearance and functionalities. During this research, robust mechanisms were established in order to detect data leakages due to network failures. We rely also in an averaging neighbors technique in order to reconstruct missing data which in practice reduce computational cost and furthermore allows a good recovery of data when PLS are not higher than 40%. A new set of indicators and their computations are provided in order to measure the traffic conditions and their environmental impact.

Motivated by the existence of new sources of information accessible at a reasonable cost, we center the discussion into the joint flow/density and flow reconstruction, the approach intended to provide a fusion mechanism for multiple data sources while considering the reconstruction of states and internal flows within the RTN in a simultaneous way. We simplify the problem from a non linear optimization framework to a quadratic constrained optimization easily solvable by most of the solvers nowadays.

8.2 Review of the contributions

The framework of development of this dissertation was the SPEEDD project. The project was intended to provide tools for scalable-proactive-event-driven-decision making. The objective of the project was to develop mechanisms for anticipation of events and correctly decide based on the predicted conditions. For that reason, the research was initially focused on the TSP problem.

The project had a scenario case concerning traffic systems, this dissertation continued the research developed by [Oje14] where forecasting techniques were explored for highway networks. The main contributions in this work are related to the fusion mechanisms implemented to obtain the forecasts. In [OKC13b], short-term forecasts were obtained in multi-steps ahead of 5min. We have proposed a time step reduction for a prediction with the objective to better capture fast dynamics appearing in stop and go waves. We have introduced also fusion mechanisms to reduce the impact of APE for long term-horizons. As a results we have obtained errors up to 35% in 90% of the cases for horizons of 25 mins (25-steps ahead). In this case we remark that the propagation of the error augments with the future horizon step. Making direct comparison is not evident since parametrization of both techniques is of different nature, for instance, clustering approach in the former technique is fixed along the day while we promote dynamic clustering and number of partitions along the day.

In order to make accessible the tool we have developed an open accessible platform for traffic systems. The main objective was to deploy a stable monitoring tool for multiple kind of users. In this case traffic operators can use the platform to accurately assess the state of the highway in real-time and with precise indicators such as CL, AS or TT. Moreover the PSFM technique developed in this research has been integrated into the application, so

regular drivers of the highway have now public access to a practical tool in order to plan their commuting times within the Grenoble South Ring. The GTL has been also renewed to attract researchers in the development of techniques for TSE and TSP, data is accessible via web and a new computational architecture was developed in order to efficiently compute indicators from the network.

Installation of the same network at the Grenoble city would be an expensive and demanding project. However, some infrastructure for measuring traffic conditions based on MLD can be found in some cities. Networks are particularly designed to partially cover important zones in those cities were congestions are observed by human perception. A new contribution in this dissertation is the proposition of an estimation approach for joint flow/density reconstruction with capabilities to integrate multiple sources of data. The main contribution in this case is the integration of an multiple measurements of different nature (MLD & FCD) and the relaxation of the nonlinear nature of the traffic problem into an estimation problem expressed as a convex quadratic constrained optimisation problem. It was found that three different criteria can be integrated into a single optimization problem, in particular the joint flow/density reconstruction problem constitutes an simultaneous input/state estimation problem in the control community. Results for these problem propose multi-stage filters that produce estimates of the input before state-estimation [YZF16]. Optimality of these filter techniques have been tackled for linear systems, however the non-linear case remains an open question. Our approach follows a different alternative, we have relaxed the problem transforming the demand supply paradigm into a set of equalities that should be fulfilled by the system. Here we rely on a reconstruction performed in a time scale so small that the network can be considered at the steady state. This assumption allows the outgoing flow recovery through the solution of a linear system (See (6.23)). This equilibria condition can be found in a general family of urban traffic networks in particular in Manhattan grid like networks.

8.3 Future research perspectives

The efficacy of the prediction methodologies based on data-driven models is studied in terms of the prediction error along the future step horizon under a cross validation schema. It is found a growing behavior of the error with the prediction horizon. The proposed scheme outperforms the historical mean and provides interesting results when *a priori* knowledge on the cluster to which belongs the time series to be predicted is available. Future directions for this work may include the analysis of fusion mechanisms in which the weights are controlled by alternative parameters (e.g. number of elements within a cluster), the study of statistical bayesian based mixtures approaches and further studies on the effect of the horizon in relationship with the clustered time series and the forecasting strategy. Other alternative approach is to study soft clustering techniques or hierarchical clustering meaning that time series for a particular day in the history may belong to different clusters, or they may be assigned to a cluster based on decision three. For this case traffic regimes may not be clearly visualized but on the other hand they offer alternative sources of historical classification that may lead to improvements in the network.

For the joint flow/density reconstruction problem a single validation scenario based on synthetic data was provided. An extension to a more realistic scenarios based on micro simulation are the following step. In those cases noise can be characterized and a model is required for the noise terms in the observation equation (6.22) as a function of the penetration rate (See. Appendix B). This test could provide more accurate results in a more realistic scenario. The formal problem of observability is still an open problem to be tackled in urban traffic networks. Given local interactions between the variables and the supply paradigm of the traffic model, distributed versions of the proposed algorithm can be studied in order to simplify the scalability of the solution.

Appendix A

Fundamental diagram calibration

The parameters w_i for each one of the sections, can be obtained from real data within the platform. From occupancy and flow measurements it is possible to construct the fundamental diagram. The occupancy relates to the density in a quasi linear relationship [KH04]. This result can be seen also from experimental data in Fig. A.1. In this case a linear model of the form $\hat{\rho} = \Gamma o$ was obtained for the density in terms of the occupancy.

This information can be used to formulate the following parameter estimation problem in which the parameters v_i^{free} and w_i can be retrieved. Let us consider a set of points $S = (\rho, \bar{f}(\rho))$ denoting measured values, the approximation for the density function given by

$$\hat{f}(\rho) = \begin{cases} \theta_1 \rho & \rho \leq \rho_c \\ \theta_2 \rho + \theta_3 & \rho > \rho_c, \end{cases} \tag{A.1}$$

where $\rho_c = \operatorname{argmax} \bar{f}(\rho)$ and the vector parameter $\theta = \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 \end{bmatrix}$ is unknown. Among the set of point let us separate the points in two sets defined $\mathcal{S}_1 = \{(\rho, y_1) | \rho \leq \rho_c, y_1 = \bar{f}(\rho)\}$, $\mathcal{S}_2 = \mathcal{S} \setminus \mathcal{S}_1$. It is possible to formulate the parameter estimation problem as

$$\begin{aligned} \min_{\theta} \quad & f(\theta) = \|b - A\theta\|_2^2 \\ \text{s.t} \quad & C\theta = d \\ A = \quad & \begin{bmatrix} \rho & 0 & 0 \\ 0 & 1 & \rho \end{bmatrix}, \quad b = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \\ C = \quad & \begin{bmatrix} 0 & \rho_c & 0 & 0 \\ 0 & 0 & 1 & \rho_c \end{bmatrix} \\ d = \quad & \begin{bmatrix} b_0 \end{bmatrix}. \end{aligned} \tag{A.2}$$

By expressing the lagrangian function of $f(\theta)$ with lagrangian multipliers z, the solution can be explicitly expressed as (See.[BV04, Chapter 10])

$$\begin{bmatrix} \hat{\theta} \\ z \end{bmatrix} = \begin{bmatrix} 2A^T A & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} 2A^T b \\ d \end{bmatrix}. \tag{A.3}$$

A graphical representation of this solution is depicted in Fig. A.2 where a linear fundamental diagram of density and flow is calibrated. In general, the same procedure is applied for all sections and the parameters are calibrated a single time, based on data from the 26, November 2016.

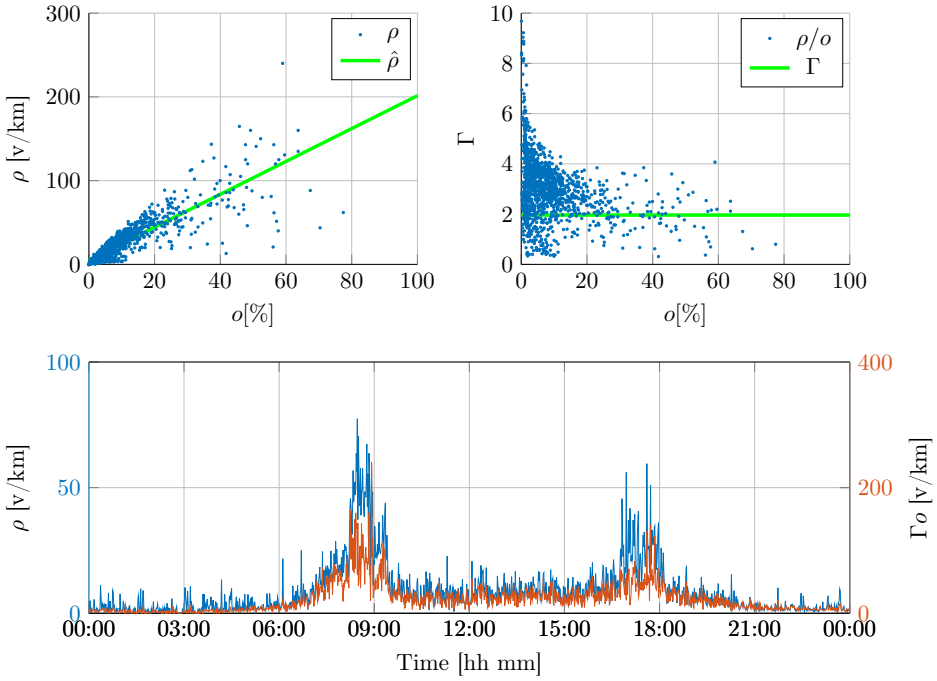


Figure A.1: Density - occupancy data relationship. Section `Meylan - A41`. Analyzed date: 21, May 2015

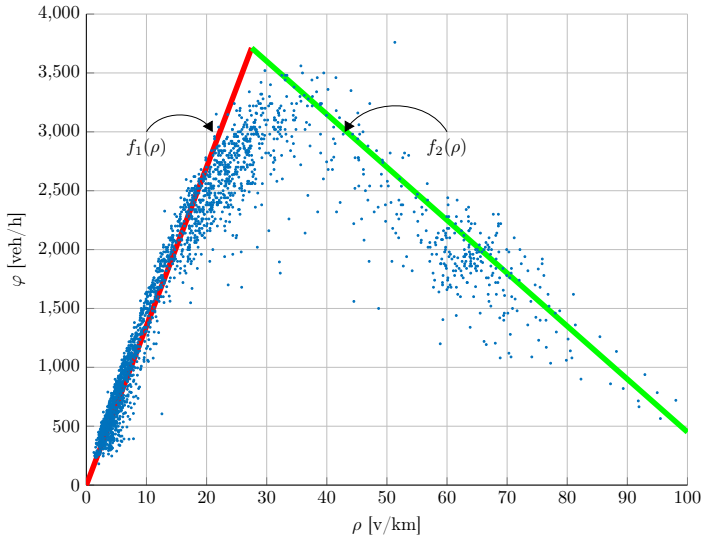


Figure A.2: Density- flow data relationship. Section `Meylan | - A41`. Analyzed date: 21, May 2015

Appendix B

Aggregation characterization of FCD

In chapter 3 we have established basic connections between microscopic variables and macroscopic variables. However, in real applications the accuracy of FCD is affected by the amount of information collected from the network. For very few number of vehicles equipped with a GPS technology the traffic conditions cannot be easily retrieved. In addition, aggregation time of this information may affect this behavior. We conducted an experiment to validate the importance of these two aspects at level of micro-simulation.

We consider a road segment with varying demand between 100 veh/h and 4000 veh/h (See Fig. B.1). Downstream congestions are generated by introducing vehicles in an through an onramp. We measure speeds of probe vehicles by considering a random sampling with probability p denoting the *penetration-ratio*. Fig. B.2 exhibits multiples scatter plots and the general behavior of the fundamental diagram. Two main points can be highlighted from the observations.

When diminishing the value of p the amount of noise in the congested part tends to increase for low aggregation times (See Fig. B.2). On the other hand aggregation over long periods of times tends to smooth the effect of the noise. From the point of view of the Traffic State Estimation this two aspects are crucial at the moment of constructing an estimation approach. The integration of noise considerations in the estimator as well as the side effects introduced by the aggregation should be considered when modeling the function $\Psi(\rho)$. Even though when sampling times are different in nature for the FCD and MLD in this dissertation we consider a synchronous process to concentrate the efforts on the estimation problem.

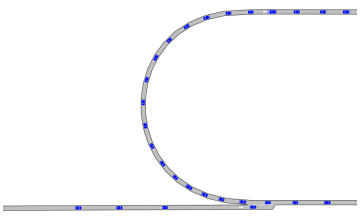
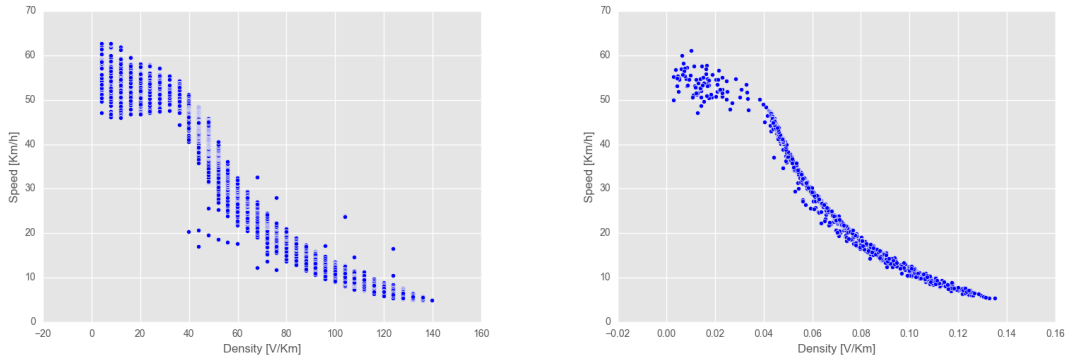
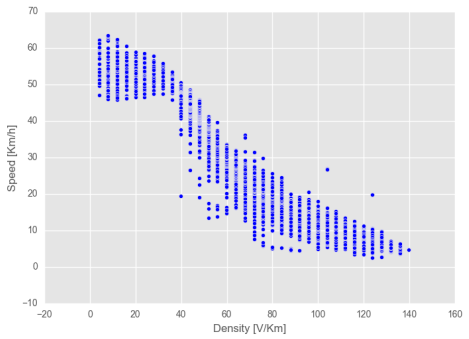


Figure B.1: Road segment micro-simulation

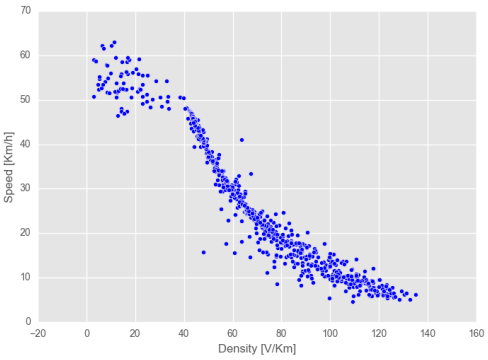


(a) 100% - 1s

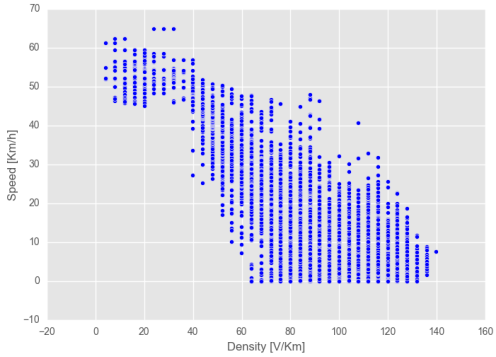
(b) 100% - 15s



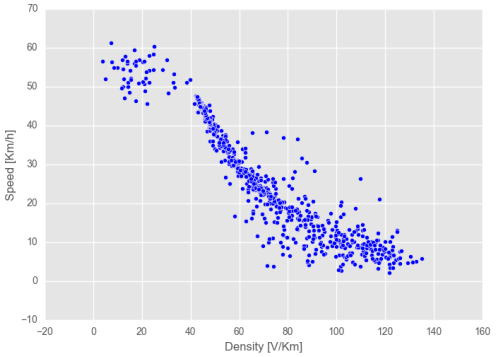
(c) 60% - 1s



(d) 60% - 15s



(e) 10% - 1s



(f) 10% - 15s

Figure B.2: Speed - Density fundamental diagram obtained from FCD. Scatter plots for multiple aggregation times and penetration rates. Data obtained from Aimsun micro-simulation

Bibliography

[ACH14] Anderson, Leah, Canepa, Edward S, and Horowitz, Roberto. "Optimization-based queue estimation on an arterial traffic link with measurement uncertainties." In: *Transportation Research Board 93rd Annual Meeting. January 12-16, Washington, D.C.* (2014) (cit. on p. 4).

[AR00] Aw, A and Rascle, Michel. "Resurrection of" second order" models of traffic flow." In: *SIAM journal on applied mathematics* 60.3 (2000), pp. 916–938 (cit. on p. 17).

[Ath65] Athol, Patrick. *Interdependence of certain operational characteristics within a moving traffic stream.* Tech. rep. 1965 (cit. on p. 27).

[AV07] Arthur, David and Vassilvitskii, Sergei. "k-means++: the advantages of careful seeding." In: *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms.* Association for Computing Machinery, 2007, pp. 1027–1025 (cit. on pp. 58, 62).

[Baj+11] Bajwa, Ravneet, Rajagopal, Ram, Varaiya, Pravin, Kavaler, Robert, and Street, Ninth. "In-Pavement Wireless Sensor Network for Vehicle Classification." In: *Proceedings of the 10th ACM/IEEE International Conference on Information Processing in Sensor Networks.* Chicago, USA, 2011, pp. 85–96 (cit. on pp. 21, 35, 37).

[Bat00] Batchelor, George. *An Introduction to Fluid Dynamics.* Cambridge University Press, 2000, p. 615 (cit. on p. 116).

[Bla+12] Blandin, Sébastien, Couque, Adrien, Bayen, Alexandre, and Work, Daniel. "On sequential data assimilation for scalar macroscopic traffic flow models." In: *Physica D: Nonlinear Phenomena* 241.17 (2012), pp. 1421–1440 (cit. on p. 26).

[BLG14] Balcan, Maria-Florina, Liang, Yingyu, and Gupta, Pramod. "Robust Hierarchical Clustering." In: *Journal of Machine Learning Research* 15 (2014), pp. 4011–4051 (cit. on p. 57).

[BLRP16] Bekiaris-Liberis, Nikolaos, Roncoli, Claudio, and Papageorgiou, Markos. "Highway Traffic State Estimation With Mixed Connected and Conventional Vehicles." In: *IEEE Transactions on Intelligent Transportation Systems* 17.12 (2016), pp. 3484–3497 (cit. on p. 25).

[BM99] Brackstone, Mark and McDonald, Mike. "Car-following: a historical review." In: *Transportation Research Part F: Traffic Psychology and Behaviour* 2.4 (1999), pp. 181–196 (cit. on p. 16).

[BNP06] Bretti, Gabriella, Natalini, Roberto, and Piccoli, Benedetto. "Numerical approximations of a traffic flow model on networks." In: *Networks and Heterogeneous Media* 1.1 (2006), pp. 57–84 (cit. on pp. 17, 86, 89).

[Box+15] Box, George EP, Jenkins, Gwilym M, Reinsel, Gregory C, and Ljung, Greta M. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015 (cit. on p. 18).

[Boy11] Boyles, Stephen. "A Comparison of Interpolation Methods for Missing Traffic Volume Data." In: *Proceedings of 90th Transportation Research Board*. Washington, 2011, pp. 11–3757 (cit. on p. 25).

[Bra+17] Brandi, Alberto, Ferrara, Antonella., Sacone, Simona, Siri, Silvia, Vivas, Carlos, and Rubio, Francisco. "Model predictive control with state estimation for freeway systems." In: *2017 American Control Conference (ACC)* July (2017), pp. 3536–3541 (cit. on p. 5).

[BV04] Boyd, S.P. and Vandenberghe, Lieven. *Convex optimization*. Cambridge University Press, 2004, p. 716. arXiv: `1111.6189v1` (cit. on p. 135).

[Can16] Candy, James V. *Bayesian signal processing: classical, modern, and particle filtering methods*. Vol. 54. John Wiley & Sons, 2016 (cit. on p. 20).

[CC16] Canepa, Edward S and Claudel, Christian G. *Networked traffic state estimation involving mixed fixed-mobile sensor data using Hamilton-Jacobi equations*. 2016. arXiv: `1606.03332` (cit. on p. 25).

[CDW+15] Canudas De Wit, Carlos, Morbidi, Fabio, Ojeda, Luis Leon, Kibangou, Alain Y, Bellicot, Iker, and Bellemain, Pascal. "Grenoble Traffic Lab: An Experimental Platform for Advanced Traffic Monitoring and Forecasting." In: *IEEE Control Systems* 35.3 (2015), pp. 23–39 (cit. on pp. 3, 5, 9, 10, 23, 35, 37, 47, 62, 105, 130).

[CGP05] Coclite, Giuseppe Maria, Garavello, Mauro, and Piccoli, Benedetto. "Traffic flow on a road network." In: *Society for Industrial and Applied Mathematics* 36.6 (2005), pp. 1862–2886 (cit. on pp. 17, 89).

[Cha+12] Chang, H., Lee, Y., Yoon, B., and Baek, S. "Dynamic near-term traffic flow prediction: system-oriented approach based on past experiences." In: *IET Intelligent Transport Systems* 6.3 (2012), p. 292 (cit. on p. 28).

[Che+03] Chen, Chao, Kwon, Jaimyoung, Rice, John, Skabardonis, Alexander, and Varaiya, Pravin. "Detecting Errors and Imputing Missing Data for Single-Loop Surveillance Systems." In: *Transportation Research Record: Journal of the Transportation Research Board* 1855.January (2003), pp. 160–167 (cit. on pp. 20, 25).

[Che04] Chen, Chao. "Travel times on changeable message signs: Pilot project." In: *California Partners for Advanced Transit and Highways (PATH)* (2004) (cit. on p. 26).

[CHM58] Chandler, Robert E., Herman, Robert, and Montroll, Elliott W. "Traffic Dynamics: Studies in Car Following." In: *Operations Research* 6.2 (1958), pp. 165–184 (cit. on p. 16).

[Chr+04] Chrobok, R., Kaumann, O., Wahle, J., and Schreckenberg, M. "Different methods of traffic forecast based on real data." In: *European Journal of Operational Research* 155.3 (2004), pp. 558–568 (cit. on pp. 27, 28, 65).

[Coi03] Coifman, Benjamin. "Estimating density and lane inflow on a freeway segment." In: *Transportation Research Part A: Policy and Practice* 37.8 (2003), pp. 689–701 (cit. on p. 25).

[CSS09] Caligaris, Carlo, Sacone, Simona, and Siri, Silvia. "Model predictive control for multiclass freeway traffic." In: *2009 European Control Conference, ECC 2009* (2009), pp. 1764–1769 (cit. on p. 5).

[Dag94] Daganzo, Carlos F. "The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory." In: *Transportation Research Part B* 28.4 (1994), pp. 269–287 (cit. on pp. 17, 87).

[Dag95a] Daganzo, Carlos F. "Requiem for second-order fluid approximations of traffic flow." In: *Transportation Research Part B: Methodological* 29.4 (1995), pp. 277–286 (cit. on p. 17).

[Dag95b] Daganzo, Carlos F. "The cell transmission model, part II: Network traffic." In: *Transportation Research Part B* 29.2 (1995), pp. 79–93 (cit. on pp. 17, 87).

[DAK14] De Almeida, André LF and Kibangou, Alain Y. "Distributed large-scale tensor decomposition." In: *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE. 2014, pp. 26–30 (cit. on p. 24).

[DG16] Dimitriou, Loukas and Gkania, Vana. "Dynamic Short-Term Projections of Travel Time Distributions in Urban Signalized Networks utilizing Composite Information of Traffic Characteristics." In: *IFAC-PapersOnLine* 49.3 (2016), pp. 237–242 (cit. on p. 32).

[DLZ13] Deng, Wen, Lei, Hao, and Zhou, Xuesong. "Traffic state estimation and uncertainty quantification based on heterogeneous data sources: A three detector approach." In: *Transportation Research Part B: Methodological* 57 (2013), pp. 132–157 (cit. on p. 25).

[DPK12] Du, Lili, Peeta, Srinivas, and Kim, Yong Hoon. "An adaptive information fusion model to predict the short-term link travel time distribution in dynamic traffic networks." In: *Transportation Research Part B: Methodological* 46.1 (2012), pp. 235–252 (cit. on p. 28).

[DY17] Duret, Aurélien and Yuan, Yufei. "Traffic state estimation based on Eulerian and Lagrangian observations in a mesoscopic modeling framework." In: *Transportation Research Part B: Methodological* 101 (2017), pp. 51–71 (cit. on p. 21).

[ECR14] Elhenawy, Mohammed, Chen, Hao, and Rakha, Hesham A. "Dynamic travel time prediction using data clustering and genetic programming." In: *Transportation Research Part C: Emerging Technologies* 42 (2014), pp. 82–98 (cit. on p. 53).

[Edi61] Edie, Leslie C. "Car-following and Steady-state Theory for Noncongested Traffic." In: *Operations Research* 9.1 (1961), pp. 66–76 (cit. on p. 16).

[Edi63] Edie, L.C. "Discussion of traffic stream measurements and definitions." In: *Proceedings of the 2nd International Symposium on the Theory of Traffic Flow* (1963). cited By 130, pp. 139–154 (cit. on pp. 23, 32).

[FH89] Fix, Evelyn and Hodges, J. L. "Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties." In: *International Statistical Review / Revue Internationale de Statistique* 57.3 (1989), p. 238 (cit. on p. 57).

[FSS15] Ferrara, Antonella, Sacone, Simona, and Siri, Silvia. "Event-triggered model predictive schemes for freeway traffic control." In: *Transportation Research Part C: Emerging Technologies* 58 (2015), pp. 554–567 (cit. on p. 5).

[GGCDW16] Grandinetti, Pietro, Garin, Federica, and Canudas-De-Wit, Carlos. "Towards scalable optimal traffic control." In: *Proceedings of the IEEE Conference on Decision and Control*. Vol. 2016-Febru. IEEE, 2016, pp. 2175–2180 (cit. on pp. 5, 86).

[GHP16] Garavello, Mauro, Han, Ke, and Piccoli, Benedetto. *Models for vehicular traffic on networks*. American Institute of Mathematical Sciences, 2016, pp. 1–471 (cit. on pp. 14, 17, 89, 90, 92).

[GKF17] Goulart, JH de M, Kibangou, AY, and Favier, G. "Traffic data imputation via tensor completion based on soft thresholding of Tucker core." In: *Transportation Research Part C: Emerging Technologies* 85 (2017), pp. 348–362 (cit. on p. 24).

[GKS15] Gabel, Moshe, Keren, Daniel, and Schuster, Assaf. "Monitoring least squares models of distributed streams." In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM. 2015, pp. 319–328 (cit. on p. 39).

[GP06] Garavello, Mauro and Piccoli, Benedetto. *Conservation laws on complex networks*. 1st ed. American Institute of Mathematical Sciences, 2006, p. 243 (cit. on pp. 4, 14).

[GR96] Godlewski, Edwige and Raviart, Pierre-Arnaud. *Numerical Approximation of Hyperbolic Systems of Conservation Laws*. Vol. 118. Applied Mathematical Sciences. New York, NY: Springer New York, 1996 (cit. on p. 17).

[Gre35] Greenshields, Bruce D. "A study of traffic capacity." In: *14 Annual Meeting of the Highway Research Board Proceedings*. 1935, pp. 448–477 (cit. on p. 24).

[HA14a] Higgs, Bryan and Abbas, Montasir. "Development of an emotional car-following model." In: *Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on*. IEEE. 2014, pp. 2972–2977 (cit. on p. 16).

[HA14b] Hyndman, Rob J and Athanasopoulos, George. *Forecasting: principles and practice*. OTexts, 2014 (cit. on p. 66).

[HAMS95] Hamed, Mohammad M., Al-Masaeid, Hashem R., and Said, Zahi M. Bani. *Short-Term Prediction of Traffic Volume in Urban Arterials*. 1995 (cit. on p. 18).

[HB10] Herrera, Juan C and Bayen, Alexandre M. "Incorporation of Lagrangian measurements in freeway traffic state estimation." In: *Transportation Research Part B: Methodological* 44.4 (2010), pp. 460–481 (cit. on pp. 23, 25).

[HC16] Habtemichael, Filmon G. and Cetin, Mecit. "Short-term traffic flow rate forecasting based on identifying similar traffic patterns." In: *Transportation Research Part C: Emerging Technologies* 66 (2016), pp. 61–78 (cit. on p. 28).

[Her] Here. *Here Wego*. `https://wego.here.com`, mendeley-groups = PhD/Web Content, url = https://wego.here.com, urldate = 17-10-2017, (cit. on p. 7).

[Her+10] Herrera, Juan C, Work, Daniel B, Herring, Ryan, Ban, Xuegang Jeff, Jacobson, Quinn, and Bayen, Alexandre M. "Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment." In: *Transportation Research Part C: Emerging Technologies* 18.4 (2010), pp. 568–583 (cit. on p. 23).

[HHB12] Hofleitner, Aude, Herring, Ryan, and Bayen, Alexandre. "Arterial travel time forecast with streaming data: A hybrid approach of flow modeling and machine learning." In: *Transportation Research Part B: Methodological* 46.9 (2012), pp. 1097–1122 (cit. on pp. 24, 27).

[Hon+11] Hong, Wei-Chiang, Dong, Yucheng, Zheng, Feifeng, and Lai, Chien-Yuan. "Forecasting urban traffic flow by SVR with continuous ACO." In: *Applied Mathematical Modelling* 35.3 (2011), pp. 1282–1291 (cit. on p. 28).

[HSL02] Haj-Salem, H and Lebacque, J. "Reconstruction of false and missing data with first-order traffic flow model." In: *Transportation Research Record: Journal of the Transportation Research Board* 1802 (2002), pp. 155–165 (cit. on p. 25).

[Hun+13] Hunter, Timothy, Hofleitner, Aude, Reilly, Jack, Krichene, Walid, Thai, Jerome, Kouvelas, Anastasios, Abbeel, Pieter, and Bayen, Alexandre. "Arriving on time: estimating travel time distributions on large-scale road networks." In: Figure 1 (2013). arXiv: `1302.6617` (cit. on p. 27).

[IBL06] Int Panis, Luc, Broekx, Steven, and Liu, Ronghui. "Modelling instantaneous traffic emission and the influence of traffic speed limits." In: *The Science of the total environment* 371.1-3 (2006), pp. 270–85 (cit. on p. 117).

[ID16] Iannini, M. L. L. and Dickman, Ronald. "Kinetic theory of vehicular traffic." In: *American Journal of Physics* 84.2 (2016), pp. 135–145 (cit. on p. 16).

[INR16] INRIX. *Traffic Congestion in Europe: INRIX France Traffic Scorecard Provides Revealing Look at Traffic Congestion in Cities Nationwide | INRIX*. 2016 (cit. on pp. 13, 32).

[Jai10] Jain, Anil K. "Data clustering: 50 years beyond K-means." In: *Pattern Recognition Letters* 31.8 (2010), pp. 651–666. arXiv: `0402594v3 [arXiv:cond-mat]` (cit. on pp. 52, 57, 58).

[Kal60] Kalman, R.E. "A new approach to linear filtering and prediction problems." In: *Journal of Fluids Engineering* 82.1 (1960), pp. 35–45 (cit. on p. 69).

[KH04] Kim, Youngho and Hall, Fred. "Relationships Between Occupancy and Density Reflecting Average Vehicle Lengths." In: *Transportation Research Record: Journal of the Transportation Research Board* 1883.1 (2004), pp. 85–93 (cit. on p. 135).

[Kha+16] Khaleghi, Azadeh, Ryabko, Daniil, Mary, Jérémie, and Preux, Philippe. "Consistent Algorithms for Clustering Time Series." In: *Journal of Machine Learning Research* 17 (2016), pp. 1–32 (cit. on p. 54).

[KV15] Kumar, S. Vasantha and Vanajakshi, Lelitha. "Short-term traffic flow prediction using seasonal ARIMA model with limited input data." In: *European Transport Research Review* 7.3 (2015), pp. 1–9 (cit. on p. 19).

[Lad+16a] Ladino, Andres, Kibangou, Alain Y., Fourati, Hassen, and Canudas de Wit, Carlos. *Speed, counts, occupancy dataset (Grenoble Traffic Lab)*. 2016 (cit. on p. 76).

[Lad+16b] Ladino, Andres, Kibangou, Alain Y., Fourati, Hassen, and Canudas de Wit, Carlos. "Travel time forecasting from clustered time series via optimal fusion strategy." In: *European Control Conference 2016*. 2016 (cit. on pp. 9, 46, 72).

[Lad+17] Ladino, A., Kibangou, A. Y., Canudas de Wit, C., and Fourati, H. "A real time forecasting tool for dynamic travel time from clustered time series." In: *Transportation Research Part C: Emerging Technologies* 80 (2017), pp. 216–238 (cit. on pp. 9, 105).

[Lad+18] Ladino, Andres, Canudas de Wit, Carlos, Kibangou, Alain Y., and Fourati, Hassen. "Density and Flow Reconstruction in Urban Traffic Networks using Heterogeneous Data Sources." In: *European Control Conference 2018*. 2018 (cit. on p. 10).

[LB11] Leclercq, Ludovic and Bécarie, Cécile. "A Meso LWR Model designed for Network Applications." In: *TRB 2012 Annual Meeting* (2011) (cit. on pp. 16, 21).

[LCK15] Lovisari, Enrico, Canudas de Wit, Carlos, and Kibangou, Alain Y. "Data fusion algorithms for density reconstruction in road transportation networks." In: *2015 54th IEEE Conference on Decision and Control (CDC)*. Osaka: IEEE, 2015, pp. 2804–2809. arXiv: `1507.07093` (cit. on pp. 25, 65).

[LCK16] Lovisari, E., Canudas de Wit, Carlos, and Kibangou, Alain Y. "Density/Flow reconstruction via heterogeneous sources and Optimal Sensor Placement in road networks." In: *Transportation Research Part C: Emerging Technologies* 69 (2016), pp. 451–476 (cit. on pp. 25, 110).

[Leb96] Lebacque, Jean-Patrick. "The Godunov scheme and what it means for first order traffic flow models." In: *International Symposium on Transportation and Traffic Theory* (1996), pp. 647–677 (cit. on p. 87).

[LHC16] Lu, Yonggang, Hou, Xiaoli, and Chen, Xurong. "A novel travel-time based similarity measure for hierarchical clustering ustering." In: *Neurocomputing* 173 (2016), pp. 3–8 (cit. on p. 60).

[Lin04] Lint, Johan Willem Christiaan van. "Reliable travel time prediction for freeways: bridging artificial neural networks and traffic flow theory." PhD thesis. TU Delft, 2004 (cit. on pp. 3, 27).

[Liu+14] Liu, Shen, Mcgree, James, White, Gentry, and Dale, Wayne. "Transport mode identification by clustering travel time data." In: *ANZIAM Journal* 56 (2014), pp. 95–116 (cit. on p. 58).

[LLL13] Li, Li, Li, Yuebiao, and Li, Zhiheng. "Efficient missing data imputing for traffic flow by considering temporal and spatial dependence." In: *Transportation Research Part C: Emerging Technologies* 34 (2013), pp. 108–120 (cit. on p. 24).

[Llo82] Lloyd, S. "Least squares quantization in PCM." In: *IEEE Transactions on Information Theory* 28.2 (1982), pp. 129–137 (cit. on pp. 57, 58).

[LR02] LeVeque and Randall. *Finite Volume Methods for Hyperbolic Problems*. Cambridge University Press, 2002, p. 558 (cit. on p. 17).

[LW55] Lighthill, M. J. and Whitham, G. B. "On Kinematic Waves. II. A Theory of Traffic Flow on Long Crowded Roads." In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 229.1178 (1955), pp. 317–345 (cit. on pp. 16, 17, 35, 87).

[LZ03] Lint, J. van and Zijpp, N. van der. "Improving a Travel-Time Estimation Algorithm by Using Dual Loop Detectors." In: *Transportation Research Record* 1855.1 (2003), pp. 41–48 (cit. on pp. 26, 32).

[LZT08] Lint, J. W C van, Zuylen, Henk J. van, and Tu, H. "Travel time unreliability on freeways: Why measures based on variance tell only half the story." In: *Transportation Research Part A: Policy and Practice* 42.1 (2008), pp. 258–277 (cit. on p. 27).

[MGF11] Montazeri-Gh, M. and Fotouhi, A. "Traffic condition recognition using the k-means clustering method." In: *Scientia Iranica* 18.4 B (2011), pp. 930–937 (cit. on p. 53).

[Mih+12] Mihaylova, L., Hegyi, A., Gning, A., and Boel, R. K. "Parallelized Particle and Gaussian Sum Particle Filters for Large-Scale Freeway Traffic Systems." In: *IEEE Transactions on Intelligent Transportation Systems* 13.1 (2012), pp. 36–48 (cit. on p. 26).

[MJW13] Mu, Tingting, Jiang, Jianmin, and Wang, Yan. "Heterogeneous Delay Embedding for Travel Time and Energy Cost Prediction Via Regression Analysis." In: *IEEE Transactions on Intelligent Transportation Systems* 14.1 (2013), pp. 214–224 (cit. on p. 28).

[MNV09] Mahajan, Meena, Nimbhorkar, Prajakta, and Varadarajan, Kasturi. *WALCOM: Algorithms and Computation*. Ed. by Das, Sandip and Uehara, Ryuhei. Vol. 5431. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 274 –285 (cit. on p. 58).

[Mor+14] Morbidi, Fabio, Ojeda, Luis Leon, Canudas de Wit, Carlos, and Bellicot, Iker. "A new robust approach for highway traffic density estimation." In: *2014 European Control Conference (ECC)*. Strasbourg, France: IEEE, 2014, pp. 2575–2580 (cit. on p. 25).

[MP69] Munjal, P. and Pahl, J. "An analysis of the Boltzmann-type statistical models for multi-lane traffic flow." In: *Transportation Research* 3.1 (1969), pp. 151–163 (cit. on p. 16).

[MRRR12] Mohan Rao, Amudapuram and Ramachandra Rao, Kalaga. "Measuring Urban Traffic Congestion, A Review." In: *International Journal for Traffic and Transport Engineering* 2.4 (2012), pp. 286–305 (cit. on p. 4).

[MS16] Motie, Mohammad and Savla, Ketan. "Throughput analysis of a horizontal traffic queue under safe car following models." In: *Decision and Control (CDC), 2016 IEEE 55th Conference on*. IEEE. 2016, pp. 6771–6776 (cit. on p. 16).

[Muñ+03] Muñoz, Laura, Sun, Xiaotian, Horowitz, Roberto, and Alvarez, Luis. "Traffic density estimation with the cell transmission model." In: *American Control Conference, 2003. Proceedings of the 2003*. Vol. 5. IEEE. 2003, pp. 3750–3755 (cit. on p. 25).

[New93] Newell, G.F. "A simplified theory of kinematic waves in highway traffic, part II: Queueing at freeway bottlenecks." In: *Transportation Research Part B: Methodological* 27.4 (1993), pp. 289–303 (cit. on pp. 16, 87).

[Oje14] Ojeda, Luis Leon. "Short-term multi-step ahead traffic forecasting." PhD thesis. Grenoble INP, Grenoble - France, Apr. 2014, p. 134 (cit. on pp. 3, 5, 32, 58, 65, 105, 109, 131, 132).

[OKC13a] Ojeda, Luis Leon, Kibangou, Alain Y, and Canudas de Wit, Carlos. "Adaptive Kalman Filtering for Multi-Step ahead Traffic Flow Prediction." In: *American Control Conference (ACC)*. Washington, 2013, pp. 4731–4736 (cit. on pp. 28, 65, 69, 74).

[OKC13b] Ojeda, Luis Leon, Kibangou, Alain Y, and Canudas De Wit, Carlos. "Online Dynamic Travel Time Prediction using Speed and Flow Measurements." In: *European Control Conference*. Zurich: IEEE, 2013, pp. 4045–4050 (cit. on pp. 27, 68, 132).

[OS84] Okutani, Iwao and Stephanedes, Yorgos J. "Dynamic prediction of traffic volume through Kalman filtering theory." In: *Transportation Research Part B* 18.1 (1984), pp. 1–11 (cit. on p. 69).

[PA60] Prigogine, I. and Andrews, F. C. "A Boltzmann-Like Approach for Traffic Flow." In: *Operations Research* 8.6 (1960), pp. 789–797 (cit. on p. 16).

[Pap+08] Papageorgiou, Markos, Diakaki, Christina, Dinopoulou, Vaya, Kotsialos, Apostolos, and Wang, Yibing. "Review of Road Traffic Control Strategies." In: *Proceedings of the IEEE* 91.May (2008), pp. 2043–2067 (cit. on p. 5).

[Pap+10] Papageorgiou, Markos, Papamichail, Ioannis, Messmer, Albert, and Wang, Yibing. "Traffic Simulation with METANET." In: *Fundamentals of Traffic Simulation*. Ed. by Barceló, Jaume. New York, NY: Springer New York, 2010, pp. 399–430 (cit. on p. 3).

[Pay71] Payne, Harold J. "Models of freeway traffic and control." In: *Mathematical models of public systems* (1971) (cit. on p. 16).

[PDN04] Pham, D. T., Dimov, S. S., and Nguyen, C. D. "An Incremental K-means algorithm." In: *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* 218.7 (2004), pp. 783–795 (cit. on p. 61).

[PDN05] Pham, D. T., Dimov, S. S., and Nguyen, C. D. "Selection of K in K-means clustering." In: *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* 219.1 (2005), pp. 103–119 (cit. on pp. 52, 61, 64).

[PHSB90] Papageorgiou, Markos, Hadj Salem, Habib, and Blosseville, Jean-Marc. "ALINEA a local feedback control law for on ramp metering." In: *Transportation Research record 1320* (1990), pp. 58–64 (cit. on p. 5).

[PJ09] Park, Hae-Sang and Jun, Chi-Hyuck. "A simple and fast algorithm for K-medoids clustering." In: *Expert Systems with Applications* 36.2 (2009), pp. 3336–3341 (cit. on p. 57).

[PL16] Piotaix, Remi; and Ladino, Andres. *Grenoble Traffic Lab - Predictions*. http://necs.inrialpes.fr/pages/grenoble-traffic-lab/results.php. 2016 (cit. on p. 76).

[Ran+15] Ran, Bin, Tan, Huachun, Feng, Jianshuai, Liu, Ying, and Wang, Wuhong. "Traffic speed data imputation method based on tensor completion." In: *Computational Intelligence and Neuroscience* 2015 (2015) (cit. on p. 24).

[Ras02] Rascle, M. "An improved macroscopic model of traffic flow: derivation and links with the Lighthill-Whitham model." In: *Mathematical and computer modelling* 35.5-6 (2002), pp. 581–590 (cit. on p. 17).

[Ric56] Richards, Paul I. "Shock Waves on the Highway." In: *Operations Research* 4.1 (1956), pp. 42–51 (cit. on pp. 16, 35, 87).

[Ro+17] Ro, Jin Woo, Roop, Partha S, Malik, Avinash, and Ranjitkar, Prakash. "A Formal Approach for Modeling and Simulation of Human Car-Following Behavior." In: *IEEE Transactions on Intelligent Transportation Systems* (2017) (cit. on p. 16).

[Seo+17] Seo, Toru, Bayen, Alexandre M., Kusakabe, Takahiko, and Asakura, Yasuo. "Traffic state estimation on highway: A comprehensive survey." In: *Annual Reviews in Control* 43 (2017), pp. 128–151 (cit. on pp. 14, 17, 20).

[SG72] Szeto, Michael W. and Gazis, Denos C. "Application of Kalman Filtering to the Surveillance and Control of Traffic Systems." In: *Transportation Science* 6.4 (1972), pp. 419–439 (cit. on p. 25).

[SKA15] Seo, Toru, Kusakabe, Takahiko, and Asakura, Yasuo. "Estimation of flow and density using probe vehicles with spacing measurement equipment." In: *Transportation Research Part C: Emerging Technologies* 53 (2015), pp. 134–150 (cit. on pp. 23, 25).

[Sum+12] Sumalee, Agachai, Wang, Jiankai, Jedwanna, Krit, and Suwansawat, Suchatvee. "Probabilistic fusion of vehicle features for reidentification and travel time estimation using video image data." In: *Transportation Research Record: Journal of the Transportation Research Board* 2308 (2012), pp. 73–82 (cit. on p. 26).

[SW17] Sun, Ye and Work, Daniel B. "Scaling the Kalman filter for large-scale traffic estimation." In: *IEEE Transactions on Control of Network Systems* (2017), pp. 1–1. arXiv: `1608.00917` (cit. on p. 25).

[Tan+13] Tan, Huachun, Feng, Guangdong, Feng, Jianshuai, Wang, Wuhong, Zhang, Yu-Jin, and Li, Feng. "A tensor-based method for missing traffic data completion." In: *Transportation Research Part C: Emerging Technologies* 28 (2013), pp. 15–27 (cit. on p. 24).

[TK13] Treiber, Martin and Kesting, Arne. *Traffic flow dynamics: Data, models and simulation*. Ed. by Springer. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 1–503 (cit. on pp. 4, 14, 32, 115).

[TYG16] Tong, Zhou, Yuxuan, Li, and Geng, Zhang. "The Drivers' Individual Behavior Difference in Car-Following Model." In: *Intelligent Transportation, Big Data & Smart City (ICITBS), 2016 International Conference on*. IEEE. 2016, pp. 550–553 (cit. on p. 16).

[Urr+09] Urrego, Germán Enrique, Calderón, Francisco Carlos, Forero, Alejandro, and Quiroga, Julián Armando. "Adquisición de variables de tráfico vehicular usando visión por computador." In: *Revista de Ingeniería* 30 (2009), pp. 7–15 (cit. on p. 23).

[VGK04] Vlahogianni, Eleni I., Golias, John C., and Karlaftis, Matthew G. "Short term traffic forecasting: Overview of objectives and methods." In: *Transport Reviews* 24.5 (2004), pp. 533–557 (cit. on pp. 14, 27).

[Viv+15] Vivas, Carlos, Siri, Silvia, Ferrara, Antonella, Sacone, Simona, Cavanna, Giulia, and Rubio, Francisco R. "Distributed consensus-based switched observers for freeway traffic density estimation." In: *2015 54th IEEE Conference on Decision and Control (CDC)*. Vol. 54rd IEEE. IEEE, 2015, pp. 3445–3450 (cit. on p. 25).

[VKG14] Vlahogianni, Eleni I., Karlaftis, Matthew G., and Golias, John C. "Short-term traffic forecasting: Where we are and where we're going." In: *Transportation Research Part C: Emerging Technologies* 43 (2014), pp. 3–19 (cit. on pp. 14, 18, 27, 65, 132).

[Wal07] Wallström, Margot. *Reclaiming city streets for people: chaos or quality of life*. Tech. rep. Brussels: European Commission Directorate-General for the Environment, 2007, pp. 38–40 (cit. on p. 13).

[Wan+09] Wang, Yibing, Papageorgiou, Markos, Messmer, Albert, Coppola, Pierluigi, Tzimitsi, Athina, and Nuzzolo, Agostino. "An adaptive freeway traffic state estimator." In: *Automatica* 45.1 (2009), pp. 10–24 (cit. on p. 26).

[Waz] Waze. *Waze*. `https://www.waze.com/livemap` (cit. on p. 7).

[WB05] Weijermars, W. and Berkum, E. van. "Analyzing highway flow patterns using cluster analysis." In: *Proceedings. 2005 IEEE Intelligent Transportation Systems, 2005.* IEEE, 2005, pp. 831–836 (cit. on pp. 52, 65).

[WF75] Whitham, GB and Fowler, Richard G. "Linear and nonlinear waves." In: *Physics Today* 28 (1975), p. 55 (cit. on p. 16).

[WFW16] Wang, Ren, Fan, Shimao, and Work, Daniel B. "Efficient multiple model particle filtering for joint traffic state estimation and incident detection." In: *Transportation Research Part C: Emerging Technologies* 71 (2016), pp. 521–537 (cit. on p. 26).

[WGH14] Wan, Nianfeng, Gomes, Gabriel, and Horowitz, Roberto. "Prediction on Travel-Time Distribution for Freeways Using Online Expectation Maximization Algorithm." In: *Transportation Research Board 93rd Annual Meeting*. Washington, 2014, pp. 14–3221 (cit. on p. 28).

[WH03] Williams, Billy M. and Hoel, Lester A. "Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results." In: *Journal of Transportation Engineering* 129.6 (2003), pp. 664–672 (cit. on p. 19).

[WH16] Wright, Matthew and Horowitz, Roberto. "Fusing loop and GPS probe measurements to estimate freeway density." In: *IEEE Transactions on Intelligent Transportation Systems* 17.12 (2016), pp. 3577–3590. arXiv: `arXiv:1510.06702v1` (cit. on p. 25).

[WOK12] Wit, Carlos Canudas-de, Ojeda, Luis Leon, and Kibangou, Alain Y. "Graph constrained-CTM observer design for the Grenoble south ring." In: *IFAC Proceedings Volumes* 45.24 (2012), pp. 197–202 (cit. on p. 25).

[Woo+17] Woodard, Dawn, Nogin, Galina, Koch, Paul, Racz, David, Goldszmidt, Moises, and Horvitz, Eric. "Predicting travel time reliability using mobile phone GPS data." In: *Transportation Research Part C: Emerging Technologies* 75 (2017), pp. 30–44 (cit. on p. 27).

[Wor+10] Work, D. B., Blandin, S., Tossavainen, O. P., Piccoli, B., and Bayen, Alexandre. "A Traffic Model for Velocity Data Assimilation." In: *Applied Mathematics Research eXpress* 2010.1 (2010), pp. 1–35 (cit. on p. 17).

[WPM06] Wang, Yibing, Papageorgiou, Markos, and Messmer, Albert. "RENAISSANCE - A unified macroscopic model-based approach to real-time freeway network traffic surveillance." In: *Transportation Research Part C: Emerging Technologies* 14.3 (2006), pp. 190–212 (cit. on p. 26).

[WSH14] Wu, Cheng-ju, Schreiter, Thomas, and Horowitz, Roberto. "Multiple-clustering ARMAX-based predictor and its application to freeway traffic flow prediction." In: *2014 American Control Conference*. Portland: IEEE, 2014, pp. 4397–4403 (cit. on p. 28).

[WVT03] Weisbrod, Glen, Vary, Don, and Treyz, George. "Measuring economic costs of urban traffic congestion to business." In: *Transportation Finance, Economics and Economic Development 2003: Planning and Administration* (2003), pp. 98–106 (cit. on p. 4).

[XT15] Xu, Dongkuan and Tian, Yingjie. "A Comprehensive Survey of Clustering Algorithms." In: *Annals of Data Science* 2.2 (2015), pp. 165–193 (cit. on pp. 57, 58).

[Yan+16] Yang, Liuhanzi, Zhang, Shaojun, Wu, Ye, Chen, Qizheng, Niu, Tianlin, Huang, Xu, Zhang, Shida, Zhang, Liangjun, Zhou, Yu, and Hao, Jiming. "Evaluating real-world CO2 and NOX emissions for public transit buses using a remote wireless on-board diagnostic (OBD) approach." In: *Environmental Pollution* 218 (2016), pp. 453–462 (cit. on p. 117).

[YG13] Yildirimoglu, Mehmet and Geroliminis, Nikolas. "Experienced travel time prediction for congested freeways." In: *Transportation Research Part B: Methodological* 53 (2013), pp. 45–63 (cit. on p. 32).

[YRG15] Yildirimoglu, Mehmet, Ramezani, Mohsen, and Geroliminis, Nikolas. "Experienced travel time prediction for congested freeways." In: *Transportation Research Procedia* 9 (2015), pp. 185–204 (cit. on p. 52).

[Yua+14] Yuan, Yufei, Van Lint, Hans, Van Wageningen-Kessels, Femke, and Hoogendoorn, Serge. "Network-wide traffic state estimation using loop detector and floating car data." In: *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*. Vol. 18. 1. Taylor & Francis Group, 2014, pp. 41–50 (cit. on p. 27).

[YZF16] Yong, Sze Zheng, Zhu, Minghui, and Frazzoli, Emilio. "A unified filter for simultaneous input and state estimation of linear discrete-time stochastic systems." In: *Automatica* 63 (2016), pp. 321–329 (cit. on p. 133).

[Zha+01] Zhang, Michael, Kim, Taewan, Nie, Xiaojian, Jin, Wenlong, Chu, Lianyu, and Recker, Will. "Evaluation of On-ramp Control Algorithms California PATH Research Report." In: (2001) (cit. on p. 5).

[Zha11] Zhang, Yang. "Hourly Traffic Forecasts Using Interacting Multiple Model (IMM) Predictor." In: *IEEE Signal Processing Letters* 18.10 (2011), pp. 607–610 (cit. on p. 28).

[ZL11] Zhang, Yang and Liu, Yuncai. "Analysis of peak and non-peak traffic forecasts using combined models." In: *Journal of Advanced Transportation* 45.1 (2011), pp. 21–37 (cit. on pp. 4, 28, 70).

Résumé — La centralisation du travail, la croissance économique et celle de la population autant que l'urbanisation continue sont les causes principales de la congestion. Lors que les villes s'efforcent pour mettre à jour leurs infrastructures du traffic, l'utilisation de nouvelles techniques pour la modélisation, l'analyse de ces systèmes ainsi que l'integration des mega données aux algorithmes aident à mieux comprendre et combattre les congestions, un aspect crucial pour le bon développement de nos villes intelligentes du XXIe siècle. Les outils d'assistance de trafic spécialement conçus pour détecter, prévoir et alerter des conditions particulières sont très demandés dans nos jours. Les prévisions à court terme et l'estimation de l'état du trafic constituent des éléments importants parmi ces outils et ont ils sont reçus beaucoup d'attention au cours de la dernière décennie.

Cette recherche est consacrée au développement des algorithmes pour l'estimation et la prédiction sur des réseaux de trafic routier. Tout d'abord, nous considérons le problème de prévision à court terme du temps de trajet dynamique (DTT) basé sur des méthodes pilotées par les données. Nous proposons deux techniques de fusion pour calculer les prévisions à court terme. Dans un premiére temps, nous considérons la matrice de covariance d'erreur et nous utilisons ses informations pour fusionner les prévisions individuelles creées á partir de clusters. Dans un deuxiéme temps, nous exploitons les mesures de similarité parmi le signal á prédire et des clusters dans l'histoire et on propose une fusion en tant que moyenne pondérée des sorties des prédicteurs de chaque cluster. Les résultats des deux méthodes on été validés dans le Grenoble Traffic Lab, un outil en temps reel qui permet la récupération de données d'une autoroute d'environ (10.5Km) qui couvre le sud de Grenoble.

Postérieurement nous considérons le problème de reconstruction de la densité / et le débit de façon simultanée à partir de sources d'information hétérogènes. Le réseau de trafic est modélisé dans le cadre de modèles de trafic macroscopique, où nous adoptons l'équation de conservation Lighthill-Whitham-Richards (LWR) avec un diagramme fondamental linaire par morceaux. Le probléme d'estimation repose sur deux principes clés. Dans un premier temps, nous considérons la minimisation des erreurs entre les débits et les densités mesurés et reconstruits. Finalement, nous considérons l'état d'équilibre du réseau qui établit la loi de propagation des flux entrants et sortants dans le réseau. Tous les principes sont intégrés et le problème est présenté comme une optimisation quadratique avec des contraintes d'égalité a fin de réduire l'espace de solution des variables à estimer. Des scénarios de simulation basés sur des données synthétiques pour un réseau de manhattan sont fournis avec l'objectif de valider les performances de l'algorithme proposé.

Mots clés : Estimation et prévision de trafic, prévsision à court terme, reconstruction de densité, clustering, fusion des données.

Abstract — Workforce centralization, population and economic growth alongside continued urbanization are the main causes of congestion. As cities strive to update or expand aging infrastructure, the application of big data, new models and analytics to better understand and help to combat traffic congestion is crucial to the health and development of our smart cities of XXI century. Traffic support tools specifically designed to detect, forecast and alert these conditions are highly requested nowadays.

This dissertation is dedicated to study techniques that may help to estimate and forecast conditions about a traffic network. First, the problem Dynamic Travel Time (DTT) short-term forecast based on data driven methods. We propose two fusion techniques to compute short-term forecasts from clustered time series. The first technique considers the error covariance matrix and uses its information to fuse individual forecasts based on best linear unbiased estimation principles. The second technique exploits similarity measurements between the signal to be predicted and clusters detected in historical data and it performs a fusion as a weighted average of individual forecasts. Tests over real data were implemented in the study case of the Grenoble South Ring, it comprises a highway of 10.5Km monitored through the Grenoble Traffic Lab (GTL) a real time application was implemented and open to the public.

Based on the previous study we consider then the problem of simultaneous density/flow reconstruction in urban networks based on heterogeneous sources of information. The traffic network is modeled within the framework of macroscopic traffic models, where we adopt Lighthill-Whitham-Richards (LWR) conservation equation and a piecewise linear fundamental diagram. The estimation problem considers two key principles. First, the error minimization between the measured and reconstructed flows and densities, and second the equilibrium state of the network which establishes flow propagation within the network. Both principles are integrated together with the traffic model constraints established by the supply/demand paradigm. Finally the problem is casted as a constrained quadratic optimization with equality constraints in order to shrink the feasible region of estimated variables. Some simulation scenarios based on synthetic data for a manhattan grid network are provided in order to validate the performance of the proposed algorithm.

Keywords: Traffic forecasting, short-term forecast, density estimation, clustering, data fusion.

GIPSA-Lab, 11 Rue des Mathématiques
38400 Saint-Martin-d'Héres