Assessment of GPS-enabled smartphone data and its use in traffic state estimation for highways

by

Juan Carlos Herrera

INGEN (Universidad Catolica de Chile) 2002
M.S. (University of California Berkeley) 2005

A dissertation submitted in partial satisfaction of the requirements for the degree of
Doctor of Philosophy

in

Engineering – Civil and Environmental Engineering

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:
Professor Alexandre M. Bayen, Chair
Professor Carlos F. Daganzo
Professor Laurent El Ghaoui

Spring 2009
The dissertation of Juan Carlos Herrera is approved:

________________________________________________________________________

Chair                     Date

________________________________________________________________________

Date

________________________________________________________________________

Date

University of California, Berkeley

Spring 2009
Abstract

Assessment of GPS-enabled smartphone data and its use in traffic state estimation for highways

by

Juan Carlos Herrera

Doctor of Philosophy in Engineering – Civil and Environmental Engineering

University of California, Berkeley

Professor Alexandre M. Bayen, Chair

GPS-enabled cell phones provide new opportunities for location-based services and traffic estimation. When traveling on-board vehicles, these phones can accurately provide position and velocity of the vehicle, and therefore can be used as probe traffic sensors. The focus of this thesis is to assess the feasibility of using GPS-enabled cell phones for traffic monitoring purposes.

This thesis first presents a field experiment nicknamed Mobile Century, which was conceived as a proof of concept of a GPS-enabled cell phone based traffic monitoring system. Mobile Century included 100 vehicles carrying a GPS-enabled Nokia N95 phone driving loops on a 10-mile stretch of I-880 near Union City, California, for 8 hours. The data obtained in the experiment was processed in real-time and successfully broadcast on the internet, demonstrating the feasibility of the proposed system for real-time traffic monitoring. Results presented in this thesis suggest that a 2-3% penetration of cell phones in the driver population is sufficient to provide accurate velocity of the traffic flow.

The second part of the thesis proposes and assesses methods to perform traffic state estimation in the presence of data provided by GPS-enabled cell phones. Traffic state estimation uses the measurement collected to produce estimates for locations for which no measurement is available. Three methods to incorporate mobile probe measurements into highway flow models are proposed. The first one is based on Kalman filtering, the second is a heuristic method derived from the first method, and the third technique is an extension of a technique used in oceanography called Newtonian relaxation. The three methods assume the
knowledge of the fundamental diagram and the conditions at both boundaries of the section of interest. Data from intermediate ramps is not required, since mobile sensors are expected to provide data to infer the state of the system at intermediate locations. The performance of the methods is assessed using two different datasets. The first dataset consist of NGSIM data, in which all vehicle trajectories are known. The second dataset consists of the *Mobile Century* data. It is shown how the accuracy in the estimation depends on the number of measurements available and their accuracy. For the cases investigated, one observation per mile-lane per minute provides sufficient data to identify most of the congestion.

Professor Alexandre M. Bayen
Dissertation Committee Chair
I want to dedicate this work to my family: Carolina, Beatriz, and Matilda. They have given me the strength to endure hard times and the peace to enjoy the good moments.

I also dedicate this work to my parents. They made the foundations of what I am today.
# Contents

List of Figures iv
List of Tables viii

## 1 Introduction 1
1.1 Motivation 1
1.1.1 The need for traffic information systems 1
1.1.2 Traffic sensors 2
1.2 Goals and contributions 5
1.3 Dissertation organization 6

## 2 Background 8
2.1 The use of mobile sensors in traffic 8
2.2 Traffic state estimation 11

## 3 A proof of concept: the *Mobile Century* experiment 15
3.1 Experimental design 15
3.1.1 Operational design 16
3.1.2 Data collection 18
3.2 Experimental results 19
3.3 Final comments 34

## 4 Traffic state estimation using Lagrangian data 36
4.1 Background 36
4.1.1 Continuous flow model 36
4.1.2 Discretization method 37
4.2 Proposed methods 39
4.2.1 Definitions: state, estimated, and observed variables 40
4.2.2 Kalman filtering based method 41
4.2.3 Filter-based heuristic method 45
4.2.4 Newtonian relaxation method 49
4.3 Final comments 51
### 5 Evaluation of the methods with traffic data

#### 5.1 Implementation with NGSIM data
- **5.1.1 Scenarios investigated**
- **5.1.2 Comparison of actual vs. observed velocity and density**
- **5.1.3 Results**

#### 5.2 Implementation with Mobile Century data
- **5.2.1 Scenarios investigated**
- **5.2.2 Results**

### 6 Conclusions

### Bibliography

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>54</td>
</tr>
<tr>
<td>55</td>
</tr>
<tr>
<td>57</td>
</tr>
<tr>
<td>60</td>
</tr>
<tr>
<td>74</td>
</tr>
<tr>
<td>75</td>
</tr>
<tr>
<td>75</td>
</tr>
<tr>
<td>91</td>
</tr>
<tr>
<td>97</td>
</tr>
</tbody>
</table>
List of Figures

1.1 Annual total cost of congestion (in billions of 2005$), total delay (in billions of hours), and total fuel wasted (in billions of gallons) from 1982 to 2005. Data extracted from [42]. ................................................................. 2
1.2 Inductive loop detectors deployment. The six squares in the pavement are the loop detectors (courtesy PATH). ................................................................. 3
3.1 Stretch of highway I-880, CA, used in the Mobile Century experiment. .... 17
3.2 Live traffic feed at 10:52am on February 8, 2008, after an accident on the NB direction of I-880 occurred, provided by 511.org (left) and the proposed system (right). Numbers in circles correspond to speed in mph. ............. 19
3.3 Vehicle trajectories in NB direction extracted from the data stored by 50% of the cell phones. The propagation of the shockwave from the accident can clearly be identified from this plot. ................................................................. 21
3.4 Loop detector station locations for the NB direction. Numbers indicate post-miles. Traffic flows from left to right, and loop detector stations have been numbered sequentially from 1 (upstream) to 17 (downstream). ............. 21
3.5 Velocity field in (mph) using a) 17 loop detector stations, b) vehicle trajectories and Edie’s speed definition, c) 17 VTLs at the loop detector station locations, and d) 30 VTLs equally spaced. ................................................................. 23
3.6 Travel time (in minutes) between Decoto Rd. and Winton Ave. Dots correspond to individual vehicle travel times (4,268 in total), collected manually using video cameras at both ends of the section of interest. ............. 24
3.7 Velocity histograms at four different locations: a) detector 3, b) detector 6, c) detector 8, and d) detector 12. Graphs at the top (x.1) correspond to VTL measurements. Graphs at the bottom (x.2) correspond to loop detector station measurements. $v_f$ and $s_f$ are the mean and standard deviation for free flow observations, while $v_c$ and $s_c$ are those for congested observations. .... 29
3.8 Loop detector station versus VTL velocity data collected at detectors a) 1, b) 4, c) 17, and d) 7. Subfigure e) shows the penetration rate at these four locations during the day. ................................................................. 30
3.9 Loop detector station vs. VTL velocity measurements (all locations). Dotted lines are the $\pm 5$ mph thresholds. ................................................................. 32
3.10 Penetration rate map (in %). ................................................................. 33
3.11 Average penetration rate (in %) over time at existing detector station locations during the a) morning and c) afternoon. The range is one standard deviation below and over the mean. Traffic flows from left to right. Histogram of the penetration rate including all the 17 locations during the b) morning and d) afternoon. ................................................. 34

4.1 Sketch of the implementation of the filter-based heuristic method. ............... 47

5.1 Highway US101S used for the NGSIM dataset. Gray area represents the section used in the implementation. .......................... 55

5.2 Schematics of the sampling strategy on equipped vehicle n. ..................... 56

5.3 Lagrangian observations for scenarios a) 2 and b) 7. Each dot corresponds to one observation. ................................. 57

5.4 Ground truth vs. observed velocity (top) and density (bottom) for scenarios 2 (left), 6 (center), and 10 (right). Ground truth quantities have been computed using Edie’s generalized definitions. ..................................... 58

5.5 Density-flow points on cell 8 of the section of interest, computed using Edie’s generalized definitions. The triangular fundamental diagram adjusted for this section is also shown. ................................. 59

5.6 PoI in the vehicle accumulation (in %) for different values of L and number of observations. a) \( k^o \) is computed as described in Section 4.2.1, b) \( k^o \) is assumed to include no error and therefore \( k^o = k \). Note that the scale of both graphs is different. ................................................................. 61

5.7 Vehicle accumulation (vehicles per cell) using the Filter-based method. (a) Ground truth, (b) EDO, (c) Scenario 2, (d) Scenario 3, (e) Scenario 5, and (f) Scenario 10. ...................................................... 64

5.8 PoI (in %) as the number of Lagrangian measurements changes (Sce: scenario) using the Filter-based method. The blue line computes observed density by inversion of the fundamental diagram (i.e. with error), while the red line uses the actual density computed from vehicle trajectories as the observed one (i.e. without error). ...................................................... 65

5.9 Total vehicle accumulation on the entire section using the Filter-based method. 66

5.10 Vehicle accumulation (vehicles per cell) using the Newtonian relaxation method. (a) Ground truth, (b) EDO, (c) Scenario 2, (d) Scenario 3, (e) Scenario 5, and (f) Scenario 10. ...................................................... 68

5.11 PoI (in %) as the number of Lagrangian measurements changes (Sce: scenario) using the Newtonian relaxation method. The blue line computes observed density by inversion of the fundamental diagram (i.e. with error), while the red line uses the actual density computed from vehicle trajectories as the observed one (i.e. without error). ...................................................... 69

5.12 Total vehicle accumulation on the entire section using the Newtonian relaxation method. ................................................................. 70

5.13 PoI (in %) for each of the 20 realizations for scenarios a) 1, b) 3, c) 8, and d) 11. ................................................................. 71
5.14 Total vehicle accumulation on the entire section using the Newtonian relaxation method. 

5.15 PoI (in %) as the number of Lagrangian measurements changes (See: scenario) for the special case of the Newtonian relaxation method. The blue line computes observed density by inversion of the fundamental diagram (i.e. with error), while the red line uses the actual density computed from vehicle trajectories as the observed one (i.e. without error). 

5.16 a) Density field (in vpmpl) collected by PeMS. b) Travel times (in minutes) computed using the velocity measurements collected by PeMS and the method described in the Section 3.2. 

5.17 Density field (in vpmpl) for the scenarios investigated (using the Filter-based method). a) scenario 1, b) scenario 2, c) scenario 3, d) scenario 4. For each scenario, the boundary data is provided by loop detector measurements. 

5.18 Travel times (in minutes) between Decoto Rd. (PM 21) and Winton Ave. (PM 27.5) for the scenarios investigated (using the Filter-based method). For each scenario, the boundary data is provided by loop detector measurements. 

5.19 Density field (in vpmpl) for the scenarios investigated (using the Filter-based method). a) scenario 1, b) scenario 2, c) scenario 3, d) scenario 4. For each scenario, the boundary data is provided by GPS measurements (VTL), i.e. no PeMS loop detector data is used. 

5.20 Travel times (in minutes) between Decoto Rd. (PM 21) and Winton Ave. (PM 27.5) for the scenarios investigated (using the Filter-based method). For each scenario, the boundary data is provided by GPS measurements (VTL), i.e. no PeMS loop detector data is used. 

5.21 Absolute difference in the density field (in vpmpl) between the case when $k_{FF} = 25$ vpmpl and a) $k_{FF} = 5$ vpmpl, b) $k_{FF} = 15$ vpmpl, and c) $k_{FF} = 35$ vpmpl. The results were obtained for scenario 3 and using the Filter-based method. The boundary data is provided by loop detector measurements. 

5.22 Difference in the mode between the case with $k_{FF} = 25$ vpmpl and a) $k_{FF} = 5$ vpmpl, b) $k_{FF} = 15$ vpmpl, and c) $k_{FF} = 35$ vpmpl. The results were obtained for scenario 3 and using the Filter-based method. The boundary data is provided by loop detector measurements. 

5.23 Density field (in vpmpl) for the scenarios investigated (using the Newtonian relaxation method). a) scenario 1, b) scenario 2, c) scenario 3, d) scenario 4. For each scenario, the boundary data is provided by loop detector measurements. 

5.24 Travel times (in minutes) between Decoto Rd. (PM 21) and Winton Ave. (PM 27.5) for the scenarios investigated (using the Newtonian relaxation method). For each scenario, the boundary data is provided by loop detector measurements. 

5.25 Density field (in vpmpl) for the scenarios investigated (using the Newtonian relaxation method). a) scenario 1, b) scenario 2, c) scenario 3, d) scenario 4. For each scenario, the boundary data is provided by GPS measurements (VTL), i.e. no PeMS loop detector data is used.
5.26 Travel times (in minutes) between Decoto Rd. (PM 21) and Winton Ave. (PM 27.5) for the scenarios investigated (using the Newtonian relaxation method). For each scenario, the boundary data is provided by GPS measurements (VTL), i.e. no PeMS loop detector data is used.  

5.27 Travel times (in minutes) between Decoto Rd. (PM 21) and Winton Ave. (PM 27.5) for scenario a) 1, b) 2, c) 3, and d) 4. Each line corresponds to one method: green, Kalman filtering; blue, filter-based heuristic; red, Newtonian relaxation. For each case, the boundary data is provided by loop detector measurements.
List of Tables

3.1 Features of the loops used in the experiment. ........................................ 18
3.2 Mean and standard deviation (both in mph) of 5-minute VTL velocity measurements \( (v_{\text{VTL}}) \). Velocity classification is defined according to VTL velocity \( (v_{\text{VTL}}) \). The number of observations (i.e. 5-minute periods) on each class is also provided. ............................................................ 26
3.3 Mean and standard deviation (both in mph) of 5-minute loop detector station velocity measurements \( (v_{\text{loop}}) \). Velocity classification is defined according to loop detector station velocity \( (v_{\text{loop}}) \). The number of observations (i.e. 5-minute periods) on each class is also provided. ................................. 27
3.4 Mean and standard deviation of the absolute difference between 5-minute VTL and loop detector station velocity measurements. Velocity classification is defined according to VTL velocity \( (v_{\text{VTL}}) \). The number of observations (i.e. 5-minute periods) on each class is also provided. ........................................... 31
5.1 Parameters of the fundamental diagram for the NGSIM site. ................. 55
5.2 Twelve scenarios investigated in the NGSIM dataset. ............................. 56
5.3 Average MAE and its standard deviation between the ground truth and the observed velocity and density. ................................................................. 60
5.4 Value of \( L \) as a function of the number of observations per 100 cell per time step and the accuracy of each observation. ............................ 63
5.5 PoI for each scenario when compared with the EDO case (using the Filter-based method). .......................................................... 65
5.6 PoI for each scenario when compared with the EDO case (using the Newtonian relaxation method). ......................................................... 69
5.7 Parameters of the fundamental diagram for the Mobile Century site. .... 74
5.8 Scenarios investigated using data from the field experiment. .................. 75
5.9 Number of observations per mile-lane per minute during free flow and congested regimes. ................................................................. 89
Acknowledgments

I want to thank my advisor, Professor Alexandre Bayen, for his support and understanding throughout all these years. His dedication is very much appreciated. I also thank Professor Carlos Daganzo and Professor Laurent El Ghaoui for serving on my dissertation committee.

This work was possible with financial support from the Center for Future Urban Transport at UC Berkeley (a Volvo International Center of Excellence) and from the California Department of Transportation (Caltrans).

The *Mobile Century* experiment would have not been possible without the help of numerous fellow students, staff from the California Center of Innovative Transportation (CCIT), and researchers from the Nokia Research Center Palo Alto. Thank you very much to all of you!

Thanks also to my fellow transportation students. They have provided me with their company and constructive feedback during these years. I am going to miss you all.

I also thank all my Chilean friends in Berkeley. They contributed significantly to make all these years here unforgettable.

Last but not least, all my gratitude to my beloved family. My wife Carolina and my daughters Beatriz and Matilda have shaped my life in ways I would have never expected. Thanks you for that! I love you.
Chapter 1

Introduction

1.1 Motivation

1.1.1 The need for traffic information systems

Traffic congestion occurs whenever the number of vehicles wishing to use a certain facility – or demand – is greater than the number of vehicles the facility can serve – or supply. This could be the result of an increase in the demand (for instance, during peak hours or for special events) or a decrease in the supply (such as incidents that block one or more lanes, adverse weather, or a work zone among others). Regardless of its causes, traffic congestion increases travel times, causing delays and an inefficient operation of the transportation system. In the US alone, traffic congestion caused a $78 billion drain on its economy in 2005, 5 billion more than the year before. In total, drivers lose 4.2 billion hours and waste 2.9 billion gallons of fuel during that year because of traffic congestion [42]. Figure 1.1 shows how these numbers have evolved since 1982.

Several solutions have been proposed over the years to address traffic congestion. These strategies aim to close the gap between demand and supply, whenever demand exceeds supply. By closing this gap, the performance of the system is improved. Among the benefits of a better system performance are: drivers spending less time on the road, drivers experiencing less stress and fatigue, less air/noise pollution, and less energy consumption.

The common belief years ago in bridging this gap was to increase the supply by adding more infrastructure (e.g. to construct a new highway or add a new lane). Given the high cost incurred and physical space needed to add more infrastructure, newer traffic
management systems intend to close the gap between demand and supply by influencing not only the supply, but also the demand. These applications are part of the Intelligent Transportation System (ITS) paradigm, in which information and communication technologies are added to vehicles and transportation infrastructure to better manage the system. Traffic management applications such as congestion pricing, ramp metering, coordination of traffic lights, high occupancy vehicles (HOV) lanes on freeways, and provision of information to drivers (through variable message signs or newer technologies) are examples of the strategies that encourage the efficient utilization of current infrastructure.

Before any strategy is implemented, an assessment of the system is needed in order to determine where congestion happens, its causes and consequences, and possible strategies to alleviate the problem. For this purpose, traffic monitoring systems have been deployed over transportation networks around the world.

1.1.2 Traffic sensors

Traffic monitoring systems require sensors. The most common way to monitor traffic nowadays (especially on freeways) is the use of inductive loop detectors. These sensors are embedded in the pavement and collect data from all vehicles as they pass over it (see Figure 1.2). Loop detectors are installed per lane, and the collection of loop detectors
from all the lanes at a given location is known as a loop detector station.

The performance of loop detectors is degraded by pavement deterioration, improper installation, aging, and weather-related effects. Thus, a good loop installation, acceptance testing, repair, and maintenance program is required to maintain the operational status of the detector [26]. In summary, the main drawbacks of the loop detector technology include its cost (installation, maintenance, operation and repair cost), and the fact that they are prone to error and malfunctions (currently in California, 30% of 25,000 detectors do not work properly [2]).

![Inductive loop detectors deployment.](image)

Figure 1.2: Inductive loop detectors deployment. The six squares in the pavement are the loop detectors (courtesy PATH).

As any other stationary sensor, loop detectors allow the operator to see and measure how traffic is flowing at a particular location, which – on freeways – is roughly half a mile apart from its neighbor detectors. Ideally, an operator would like to know and measure how traffic is flowing everywhere over the network at all times. It would be prohibitively expensive, however, to install a loop detector on every road at a reasonable resolution (freeways and surface streets).

Other stationary sensors used for monitoring purposes – but less popular than loop detectors – include magnetic sensors, video cameras, microwave radars, infrared sensors, ultrasonic sensors, and passive acoustic sensors. Reference [26] provides a complete review of different traffic sensors in use.

Alternative ways to monitor traffic make use of information provided by individual vehicles. Examples of mobile sensors capable of providing such information are radio-
frequency identification (RFID) transponders used to pay road tolls electronically such as FasTrak in California or EZ-Pass on the East Coast, global positioning systems (GPS) receivers, and cellular or mobile phones.

RFID transponders can be used to compute travel time between consecutive locations, which requires the installation of post readers along the road. This technology, however, is not universal and only provides point-to-point information. GPS-equipped vehicles reporting traffic conditions are also used for traffic monitoring purposes. Their penetration, however, is usually not high enough to provide sufficient spatio-temporal coverage of the transportation network.

Mobile phones can also be used to provide travel time. When a cell phone is handed over from one cell to another, an area in which the cell phone is located is obtained. The main advantage of mobile phones is its spatial coverage and penetration in the population. By the end of 2007, the penetration rate of mobile phones in the population was over 50% in the world, ranging from 30-40% in developing countries (with an annual growth rate greater than 30%) to 90-100% in developed countries [49]. The main drawback of this technology, however, is its low position accuracy.

Therefore, some of the technologies presented before have potentially good penetration in the driving population but low accuracy (such as cell phones), while others have low penetration but higher accuracy. All of them, however, do not require a huge investment in the infrastructure.

GPS-enabled cellular phones as sensors: low cost, high penetration and high accuracy. In the era of multimedia convergence, communication and sensing platforms, GPS-enabled smartphones are becoming an essential contributor to location-based services. These devices combine the advantages of mobile sensors mentioned earlier: low investment cost, high penetration, and high accuracy achieved by GPS receivers. Also, GPS-enabled smartphones are able to accurately provide not only position, but also velocity and the direction of travel. Note that phones not only can send but also can receive information. Therefore, (personalized) traffic information can be delivered through this channel.

This new sensing technology can greatly impact the transportation field, creating a novel and rich source of traffic data. Maintenance problems common with stationary sensors (mainly loop detectors) do not appear with this technology. For public transportation agencies, there is no cost in deploying this new sensing technology because it is market
driven. Personal devices (smartphones, in this case) that people already own act as sensors, which has brought private companies into a field that has historically been monitored and managed by public agencies (encouraging partnerships between the public and private sector). Given the penetration of mobile phones across the population, this new sensing technology can potentially provide an exhaustive spatial and temporal coverage of the transportation network when there is traffic. Moreover, it has the potential of being deployed almost everywhere – particularly in developing countries, where there is a lack of resources for traffic monitoring infrastructure systems.

1.2 Goals and contributions

As mentioned before, loops detectors cannot be deployed globally at an acceptable resolution. They allow the operator to see and measure traffic flow only at specific locations. Based on this information, statistical models, linear interpolation, or traffic flow models can be used to estimate traffic conditions at locations without detectors. This process is known as traffic state estimation, also referred to as data assimilation or inverse modeling.

Similarly, not all vehicles are equipped with mobile sensors such as onboard GPS-enabled smartphones. Moreover, depending on the sampling strategy, there will be periods of time and locations from which no data is being collected. Since the amount of data is not sufficient, traffic state estimation is also needed.

The problem of interest consists in finding algorithms to make use of the data collected from GPS-enabled smartphones – in addition to the data collected from stationary detectors – to estimate the traffic state on roads.

The main goals of this thesis are the following:

- Practical goal: assess the feasibility of a traffic monitoring system based on GPS-enabled cell phone as sensors at the right penetration rate. This technology has never been tested and assessed on the field to determine its real capabilities with realistic penetration rates.

- Methodological goal: investigate how the data collected can be used to perform traffic state estimation on freeways. In particular, how to incorporate mobile sensing information into traffic flow models.

\footnote{Privacy issues and technology limitations (smartphone battery life and bandwidth usage) prevent the collection of all the data generated by a GPS traveling onboard vehicles.}
The specific contributions are driven by the goals above, and they include:

- A large-scale field experiment aimed to assess the feasibility of a traffic monitoring system based on GPS-enabled smartphones as traffic sensors. The experiment:
  - allows the evaluation of the quality of GPS-enabled smartphone data.
  - has generated a rich dataset of vehicle trajectories that can be used in further studies.
- The analysis and documentation of methods to estimate traffic state using GPS data.
- An implementation of the proposed methods, using two datasets:
  - An extensive dataset – the NGSIM data – for which ground truth is known.
  - A new field dataset from a large scale field experiment, and a cross validation with PeMS data.

The proposed methods provide better information about vehicle accumulation, velocities, travel times and/or delays. Therefore, operators and users can make more informed decisions. In the case of the operator, these decisions include how to meter an on-ramp or how to program traffic lights; in the case of users, those decisions include when to start a trip, which route to take and the expected travel time.

In summary, the methods provide real-time information to implement some of the strategies or traffic management applications mentioned at the beginning of this chapter and are aimed at closing the gap between demand and supply, which would improve the system’s performance by alleviating traffic congestion.

1.3 Dissertation organization

The rest of this dissertation is organized as follows: Chapter 2 discusses the use of mobile phones in traffic and provides some background information about the traffic state estimation problem. Chapter 3 describes and presents the main results of a field experiment conceived as a proof of concept of a traffic monitoring system based on GPS-enabled cell phones. The methods proposed to perform traffic state estimation in the presence of GPS data are discussed in Chapter 4. They were implemented and tested using...
two databases, and the main results are presented in Chapter 5. Finally, Chapter 6 states the main conclusions drawn from this study.
Chapter 2

Background

Before the era of the mobile internet, characterized in particular by the emergence of location based services heavily relying on GPS, the traffic monitoring infrastructure has mainly consisted of dedicated equipment, such as loop detectors, cameras, and radars. Installation and maintenance costs prevent the deployment of these technologies for the entire arterial network and even for highways in numerous places around the world. Moreover, inductive loop detectors are prone to errors and malfunctions.

For this reason, the transportation engineering community has looked for new ways to collect traffic data to monitor traffic. Electronic devices traveling onboard cars are appealing for this purpose, as they usually provide a cost-effective and reliable way to collect traffic data.

2.1 The use of mobile sensors in traffic

FasTrak and EZ-Pass are RFID transponders that can be used to obtain individual travel times based on vehicle re-identification [51, 6]. Readers located on the side of the road keep a record of the time the transponder (i.e. the vehicle) crosses that location. Measurements from the same vehicle are matched between consecutive readers to obtain the travel time. The fundamental limitations of this system is the cost of installing the infrastructure (readers), its limited coverage, and the fact that only the travel time between two locations can be obtained.

*Global Positioning System* (GPS) devices found in the market can compute position and instantaneous velocity readings of a vehicle with a high accuracy, which can be
used to obtain traffic information. Sanwal and Walrand in [41] addressed some of the key issues of a traffic monitoring system based on probe vehicle reports (position, speeds, or travel times), and concluded that they constitute a feasible source of traffic data. Zito et al. also investigated the use of GPS devices as a source of data for traffic monitoring in [56]. Two tests were performed to evaluate the accuracy of the GPS as a source of velocity and acceleration data. The accuracy level found was good, even though the selective availability\(^1\) feature was still on. The main drawback of this technology is its low penetration in the population, which is not sufficient to provide an exhaustive coverage of the transportation network. Dedicated probe vehicles equipped with a GPS device represent an added cost that cannot be applied on a global scale. An example of such a program on a small scale is HICOMP\(^2\) in California, which uses GPS devices in dedicated probe vehicles to monitor traffic for some freeways and major highways in California. However, as pointed out by Kwon et al. in [28], the penetration of HICOMP is low and the collected travel times are not as reliable as other systems such as PeMS.

Other approaches have investigated the possibility of using dedicated fleets of vehicles equipped with GPS or automatic vehicle location (AVL) technology to monitor traffic [34, 43, 10], such as FedEx, UPS trucks, taxis, buses or dedicated vehicles. While industry models have been successful at gathering significant amounts of historical data using this strategy (for example, Inrix), the use of dedicated fleets always poses issues of coverage, penetration, bias due to operational constraints and specific travel patterns. Nevertheless, it appears as a viable source of data, particularly in large cities.

In the era of mobile internet services, and with the shrinking costs and increased accuracy of GPS, probe based traffic monitoring has become very appealing to industries working in the field of mobile sensing. The increasing penetration of mobile phones in the population makes them attractive as traffic sensors, since an extensive spatial and temporal coverage could potentially soon be achieved. Traffic monitoring systems based on GPS-enabled cellular phones are particularly suitable for developing countries, where there is a lack of resources for traffic monitoring infrastructure systems, and where the penetration rate of mobile phones in the population is rapidly increasing.

---

1Selective availability is the intentional inclusion of positioning error in civilian GPS receivers. It was introduced by the Department of Defense of the U.S. to prevent these devices from being used in a military attack on the U.S. This feature was turned off on May 1, 2000.

2Highway COngestion Monitoring Program.
http://www.dot.ca.gov/hq/traffops/sysmgtp/HICOMP/index.htm
Multiple technological solutions using cell phones exist to the localization problem. Historically, the seminal approach chosen for monitoring vehicle motion using cell phones (prior to the rapid penetration of GPS in cellular devices) uses cell tower signal information to identify a handset’s location. This technique usually relies on triangulation, trilateration, tower hand-offs, or a combination of the three. Several studies have investigated the use of mobile phones for traffic monitoring using this approach (see for example [50, 52, 33, 17, 7]). The fundamental challenge in using cell tower information for estimating the position and motion of vehicles is the inherent inaccuracy of the method, which poses significant difficulties to the computation of speed. Several solutions have been implemented to circumvent this difficulty, in particular by the company Airsage, which historically developed its traffic monitoring infrastructure based on cell tower information [32, 29, 24].

Based on the time difference between two positions, average link travel time and speed can be estimated. Yim and Cayford conducted a field experiment to compare the performance of cell phones and GPS devices for traffic monitoring [54]. The study concludes that GPS technology is more accurate than cell tower signal for tracking purposes. In addition, the low positioning accuracy of non-GPS based methods prevents its massive use for monitoring purposes, especially in places with complex road geometries. Also, while travel times for large spatio-temporal scales can be obtained from such methods, other traffic variables of interest, such as instantaneous velocity are more challenging to obtain accurately.

A second approach is based on GPS-enabled smartphones, leveraging the fact that increasing numbers of smartphones or personal devices come with GPS as a standard feature. This technique can provide more accurate location information, and thus more accurate traffic data such as speeds and/or travel times. Additional quantities can potentially be obtained from these devices, such as instantaneous velocity, acceleration, and direction of travel. Fontaine and Smith [17] use cell phones for traffic monitoring purposes, and mention the need of having a GPS-level accuracy for position to compute reasonable estimates of travel time and speed. Yim and Cayford [54] and Yim [53] conclude that if GPS-equipped cell phones are widely used, they will become an attractive and realistic alternative for traffic monitoring.

GPS-enabled mobile phones can potentially provide an exhaustive spatial and temporal coverage of the transportation network when there is traffic, with a high positioning accuracy achieved by a GPS receiver. Demers et al. [15] investigated the deployment of 200 vehicles for an extended period of three months and the potential data that can be gathered
from it. As it appears, in light of this study, one of the main issues in experiments or pilot tests is the problem of penetration, i.e. the percentage of vehicles equipped vs. the total number of vehicles on the road.

Some concerns regarding this technology include the need for a specifically designed handset, and the fact that the method requires each phone to send information to a monitoring system, which could potentially increase the communication load on the system and the energy consumption of the handset\(^3\) as pointed out by Rose [40] and Qui et al. [38]. Another issue is the knowledge of vehicle position and velocity provided by this technology, which needs to be used in a private non-intrusive manner.

The impact of these concerns (communication load, handset energy consumption, and privacy) can be handled with the appropriate sampling strategy. Sampling GPS data in the transportation network can be done in at least two ways:

- **Temporal sampling**: equipped vehicles report their information (position, velocity, etc.) at specific time intervals \(T\), regardless of their positions.

- **Spatial sampling**: equipped vehicles report their information (time, velocity, etc.) as they cross some spatially defined sampling points. This strategy is similar to the one used by inductive loop detectors or RFID transponders, in which data is obtained at fixed locations. It has the advantage that the phone is forced to send data from a given location of interest.

From the traffic information systems standpoint, it is desirable to have a significant amount of information available. Therefore, with a satisfying GPS accuracy, a small \(T\) or very closely placed fixed measurements would yield more accurate estimates of traffic. However, these objectives conflict with the communication load constraints and the concern for privacy preservation.

### 2.2 Traffic state estimation

Regardless of the type of sensors being used for monitoring purposes, it is not possible to collect data from everywhere at all times. Therefore, with the limited information provided by the sensors, we need to estimate traffic flow on a given section or area of interest.

---

\(^3\)With the advent of the 3G network and the rapid growth of data and bandwidth intensive applications, this concern has become less important since the end of 2008.
This process is known as traffic state estimation, where the traffic state corresponds to the accumulation (or density) of vehicles or the velocity on a given section in most cases.

Traffic state estimation requires data and a model of the system. Depending on the type of sensors used, different types of data can be collected. Numerous freeways around the world, in particular in the US, are equipped with loop detector stations that are embedded in the pavement to collect traffic data. For each lane, these detectors aggregate information at a given sample time (usually 20 or 30 seconds). Vehicle counts, occupancy and speed are among the information that a detector can collect. The measurements provided by these static sensors are traditionally referred to as Eulerian measurements, which means that the detector measures flow through a fixed control screen. On the other hand, Lagrangian sensors collect measurements of the system along a particle trajectory (i.e. the motion of a car). Examples of mobile (or Lagrangian) sensors are RFID transponders, GPS devices, and cell phones onboard vehicles which provide position and/or velocity.

Different modeling approaches can be used for traffic reconstruction. The first class includes models that do not make use of a traffic flow model, and the estimation is based on statistics computed from current data and sometimes historical data as well. Real-time traffic reports are usually based on these types of statistical models. These models have also been a common practice in studies that use cell phones as traffic sensors, in which the main goal has been to find the link speed or travel time estimation [41, 50, 52, 7, 27]. Note that the first four aforementioned studies use cell phone antennas to obtain cell phone (i.e. vehicle) position, which is less accurate than GPS positioning. Krause et al. [27] have investigated the use of machine learning techniques to reconstruct travel times on a graph based on sparse measurements collected from GPS devices embedded in cell phones and automobiles.

The second class of models includes traffic flow models based on the physics of the traffic flow. Lighthill and Whitham [31], and Richards [39] independently proposed in the 1950’s a first order partial differential equation (referred here as the LWR PDE) to describe traffic evolution over time and space. The LWR PDE is a first order scalar hyperbolic conservation law based on the conservation of vehicles, which relates density on the road and its flow. Extension of this model includes a second equation accounting for the fact that vehicles do not accelerate/decelerate instantaneously, which are known as second order models, such as in [55] and [5]. Numerical schemes, such as the Godunov scheme [19], can be used to discretize these continuous models.
Daganzo in [13] and [14] proposes the *Cell Transmission Model* (CTM), which is a discretization of the LWR PDE\(^4\). The CTM divides the highway in “cells” of length \(\Delta x\) and computes the state of the system (vehicle accumulation or density) every \(\Delta t\) units of time in each cell, according to the conservation of vehicles principle. The CTM transforms the nonlinear flux function into a nonlinear discrete operator. Modifications to the CTM can be found in the literature. Muñoz et al. in [36] use a hybrid system framework to develop the *Switching-Mode Model* (SMM), which combines discrete event dynamics estimation (mode identification) with nonlinear continuous dynamics state estimation (density estimation). Gomes and Horowitz [20] modify the original merge rule of the CTM and propose the *asymmetric CTM* (ACTM).

For traffic state estimation, dynamic flow models of the system can be combined with data collected by sensors, a process known as data assimilation. However, there are various techniques to perform data assimilation. Gazis and Knapp [18], Szeto and Gazis [47], and Sun et al. [46] have used Eulerian measurements to perform data assimilation using Kalman filtering techniques. While [18] and [47] used the conservation of vehicles as their model, [46] uses the SMM. Chu et al. [12] and Nanthawichit et al. [35] have also performed data assimilation using Kalman filtering techniques, but with simulated Lagrangian data. In [12], the authors use the conservation of vehicles equation and Kalman filtering techniques to combine point detection data and probe vehicle data into the travel time estimation, while a second order model is used in [35] to perform data assimilation.

Treiber and Helbing in [48] propose a method that filters Eulerian data collected from loop detectors (density or velocities) to reconstruct traffic flow. The filter is such that in free flow, traffic information propagates downstream, while in congestion, it travels upstream. Even though this method takes into account the way in which information propagates in traffic streams, it does not make use of any flow model.

Data assimilation using Lagrangian data to estimate the state of the system is common in other fields such as meteorology and oceanography. In these fields, data assimilation methods range from simple, suboptimal techniques such as direct insertion, statistical correction, and statistical interpolation to more sophisticated, optimal algorithms such as inverse modeling, variational techniques and a family of methods based on Kalman filtering [37]. The extrapolation of these techniques to the transportation engineering problems

\(^4\)The numerical scheme in [13] is a special case of the Godunov scheme when the fundamental relation between flow and density is assumed to be triangular.
appears to be very promising.

A known and simple method used in oceanography is the Newtonian relaxation (or nudging) method [4], which we use for the present work. The Newtonian relaxation method relaxes the dynamic model of the system towards the observations. To this end, a source term proportional to the difference between the predicted and observed state is included in the constitutive equation. The weighted factor of this difference is called the nudging factor, and it vanishes away from the measurement location and after the measurement time. Therefore, close to where and when the observation is made, this factor nudges the solution towards the observations.

Most of the data assimilation methods use boundary conditions at the boundary of the physical domain of interest. They are usually assumed to be known at some specific locations (usually at the boundaries of the computational domain). Thus, these methods make use of Eulerian data as well.
Chapter 3

A proof of concept: the *Mobile Century* experiment

As documented in the transportation engineering literature [52, 53, 38, 27], field tests are needed to assess the effectiveness of new technologies. As shown in the previous chapter, test deployments to assess the potential GPS-enabled cell phones as traffic sensors have in general lacked the sufficient proportion of equipped vehicles in the total flow to obtain meaningful results (or penetration rate).

The present chapter presents the results of a large scale field experiment conducted in the San Francisco Bay Area, California. The experiment aimed at assessing the feasibility of a traffic monitoring system using GPS-enabled mobile phones for highways. Its specificity is the penetration rate achieved during the test, which is believed to be representative of the upcoming GPS equipped phones’ penetration in the population within a few months from the experiment. The performance of the system was sustained for a long enough period of time to show the feasibility of such a monitoring system.

The main goal of this chapter is to demonstrate that a traffic monitoring system based on GPS-enabled cell phones is feasible in the short term, and to analyze the quality of the data collected by such a system.

### 3.1 Experimental design

The experiment was conceived as a proof of concept of a traffic monitoring system based on GPS-enabled mobile phones. It was designed with two fundamental goals:
Goal 1: Assess the feasibility of a traffic monitoring system based on GPS-enabled mobile phones. The data collected was shown to provide sufficient and accurate enough data to deliver precise travel time and velocity estimations.

Goal 2: Evaluate speed measurements accuracy from GPS-enabled mobile phones under both free flow and congested traffic conditions. Therefore, the section of highway was chosen to encompass both free flow and congested conditions. Good detector stations coverage was also required for comparison purposes.

For both goals, to maintain a specific penetration rate of equipped vehicles in the total flow throughout the experiment is desirable. This feature of the experiment is a fundamental difference with previous work, and necessary for the proper testing of traffic flow reconstruction algorithms.

3.1.1 Operational design

Nicknamed the Mobile Century experiment, the February 8, 2008, field experiment involved 100 vehicles carrying GPS-enabled Nokia N95 phones. All rented vehicles were driven by 165 UC Berkeley students in shifts. The section of freeway chosen is on I-880 near Union City in the San Francisco Bay Area, California (see Figure 3.1). This section of highway has four (and sometimes five) lanes, the leftmost one being a high occupancy vehicle or HOV lane. It presents interesting traffic properties, which include alternating periods of free-flow and congestion throughout the day (which thus satisfies the requirements of Goal 2). In particular, the northbound (NB) direction presents a recurrent and severe bottleneck between Tennyson Rd. and CA92 during the afternoon. Moreover, on the day of the experiment, there was an accident during the morning, which activated a non-recurrent bottleneck at this same location. The section is also well covered with an unusually high number of existing loop detector stations – 25 between Stevenson Blvd. and Winton Ave. in the NB direction – feeding into the PeMS system [2].

The vehicles repeatedly drove loops of six to ten miles in length continuously for eight hours (also shown in Figure 3.1). The goal of the looping behavior is to achieve and maintain a desirable 2%-5% penetration rate of the total volume of traffic on the highway during the experiment. Given that GPS is becoming an standard feature in cell phones, this

---

1The HOV lane is active between 5am and 10am in the morning and 3pm and 7pm in the afternoon.
level of penetration is seen as realistically achievable in the near future. Note that previous studies have reported that data coming from no more than 5% of the total flow is sufficient to obtain accurate estimates of the travel time [41, 50, 54]. Given that the total flow expected on the section of interest is approximately 6,000 vehicles per hour (obtained from PeMS) and the number of equipped vehicles is 100, the required cycle time to achieve the desired rate is 20 minutes. Knowing the expected speed throughout the day and cycle time is sufficient to determine the length of the loops throughout the day. In the NB direction of the section of interest, free flow conditions are historically expected during the morning until 2-3pm, when the recurrent bottleneck between Tennyson Rd. and CA92 activates. Free flow is expected during most of the day in the southbound direction. For this reason, long loops (or AM) loops were designed during the morning and short (or PM) loops were used during the afternoon. The change was scheduled to start at 1:30pm. Table 3.1 presents the main features of the loops used during the experiment. Three different loops of almost the same length were used during the experiment to prevent oversaturating any of the ramps being used.
Table 3.1: Features of the loops used in the experiment.

<table>
<thead>
<tr>
<th>Loop type</th>
<th>North end</th>
<th>South end</th>
<th>One-way distance (miles)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AM (long)</td>
<td>Winton Ave</td>
<td>Thornton Rd.</td>
<td>9.4</td>
</tr>
<tr>
<td></td>
<td>CA92 (San Mateo Br.)</td>
<td>Mowry Ave.</td>
<td>8.6</td>
</tr>
<tr>
<td></td>
<td>Tennyson Rd.</td>
<td>Stevenson Blvd.</td>
<td>9.3</td>
</tr>
<tr>
<td>PM (short)</td>
<td>Winton Ave.</td>
<td>Alvarado-Niles</td>
<td>4.5</td>
</tr>
<tr>
<td></td>
<td>CA92 (San Mateo Br.)</td>
<td>Alvarado Blvd.</td>
<td>5.2</td>
</tr>
<tr>
<td></td>
<td>Tennyson Rd.</td>
<td>Decoto Rd.</td>
<td>5.4</td>
</tr>
</tbody>
</table>

3.1.2 Data collection

The data was collected in two ways during the experiment. First, each cell phone stored its position and velocity log every 3 seconds, which allows for the computation of every equipped vehicle trajectory. This data was gathered locally on the phones for analysis purposes. It became available only after the experiment was finished, and is useful in testing the accuracy of the sampling strategy (Goal 2) a posteriori. This data was also transmitted in real time to a team of safety officers, which were able to monitor the vehicles’ locations on their laptops during the experiment.

The second way to collect data is based on the concept of virtual trip lines (VTLs). Each VTL consists of two GPS coordinates which define a virtual line drawn across a roadway of interest. The coordinates are downloaded into the mobile handset (dynamically), and whenever it crosses a VTL, a position and speed update is triggered and sent to the system. That is, the VTL sampling strategy is a spatial strategy. Sampling in space through VTLs is privacy aware and facilitates a distributed monitoring architecture that processes only anonymous location updates (i.e. no single infrastructure entity possesses identity and accurate location information [23]). A total of 45 VTLs were deployed between Stevenson Blvd. and Winton Ave. (each VTL covered both travel directions). This data was used to produce real-time travel time and speed estimates.

Finally, high resolution video cameras located on the over-crossing at Winton Ave., Decoto Rd., and Stevenson Blvd. (stars in Figure 3.1) recorded traffic in the NB direction. Each This video data is accurate enough to provide exact travel time of individual vehicles through license plate re-identification. Two cameras were used at each over-crossing in order to achieve a sufficiently high resolution to read the plates.
3.2 Experimental results

This section analyzes the main results derived from the experiment. Unless otherwise noted, the rest of this section focusses on the highway segment covered by the afternoon loops in the northbound (NB) direction. The section consists of the portion of highway between Decoto Rd. to the south – postmile 21 – and Winton Ave. to the north – postmile 27.5.

Goal 1: Assessment of the feasibility of a smartphone-based traffic monitoring system

The data obtained in the experiment consists of VTL, and was processed in real-time. We deployed 30 VTLs during the experiment in the section of interest. Information collected by these VTLs was used to produce real-time travel time and velocity estimates, which were broadcasted for eight hours. Figure 3.2b illustrates the interface used to broadcast travel time and speed during the day. The figure shows traffic at a time after an accident occurred in the NB direction between Tennyson Rd. and CA92. Figure 3.2a shows the 511.org traffic display at the same time.

Figure 3.2: Live traffic feed at 10:52am on February 8, 2008, after an accident on the NB direction of I-880 occurred, provided by 511.org (left) and the proposed system (right). Numbers in circles correspond to speed in mph.

As can be seen from the two subfigures in Figure 3.2, the extent of congestion based on the GPS cell phone data match closely to the one broadcasted on the 511.org website,
which used a combination of techniques for velocity and travel time calculation including loop detector data, FasTrak-equipped vehicles, and speed radars. Comparisons with the 511.org speed map at other times during the experiment showed similar results, which confirm that the GPS cell phone based technique can produce reasonable speed estimates for the section of interest, at least for the experiment day.

**Goal 2: Assessment of the accuracy of the probe data**

This subsection analyzes the data stored in each phone and the type of information that can be collected using VTLs. By nature of the test site, it provides an assessment of GPS data quality in suburban freeways, and of this data’s values for highway traffic estimation.

**Trajectory data**

Each phone stored its position (latitude and longitude) and a velocity log every three seconds. We refer to this data as *trajectory* data since vehicle trajectories can be reconstructed from it.

Trajectory data was processed after the experiment, in order to conduct a more detailed analysis of the quality of the data collected by the GPS-enabled smartphones. Figure 3.3 shows 50% of the gathered trajectories between Stevenson Blvd. (postmile 17) and Winton Ave. (postmile 27.5) in the NB direction. The transition from the AM loops to the PM loops that occurs at 1:30pm can be clearly seen in the figure, as well as the fact that different vehicles were using different ramps to get in and out of the highway (as shown in Figure 3.1). The propagation of the shockwave generated by the accident is clearly identified from this plot as well.

The information about the propagation of shockwaves can be used to infer parameters of the fundamental diagram (assuming a triangular relationship), as well as flows and densities that mobile sensors are not able to capture directly. This information can be useful in the absence of loop detectors, since it relates the sample that provides GPS data with the total driving population.

Using these trajectories, a velocity field can be reconstructed and compared with the PeMS velocity field using data from the 17 loop detector stations deployed in the section
Figure 3.3: Vehicle trajectories in NB direction extracted from the data stored by 50% of the cell phones. The propagation of the shockwave from the accident can clearly be identified from this plot.

of interest\(^2\) (loop detector station locations are shown in Figure 3.4).

Figure 3.4: Loop detector station locations for the NB direction. Numbers indicate post-miles. Traffic flows from left to right, and loop detector stations have been numbered sequentially from 1 (upstream) to 17 (downstream).

Loop detector stations compute the \textit{temporal mean speed} (TMS) every 30 seconds (considering all lanes); every 5 minutes the average of the 30-second observations collected during this time is computed. PeMS velocity field is shown in Figure 3.5a. The method used in PeMS associates an influence area with each detector station. The assumption is that

\(^2\)Only loop detectors on lanes 1 and 2 at detector station 3 were not working properly as reported by PeMS.
measurements for this area are provided by the corresponding detector station. The size of the influence area depends on the proximity of neighboring detector stations. Therefore, the closer the neighboring detectors are, the smaller the influence area and the better the estimates that can be obtained using this method. Note that the vehicles used in the experiment were not allowed to drive on the HOV lane after 3pm. For this reason, data from loop detector stations only includes the three general purpose lanes after 3pm. For the rest of the time, four lanes are considered.

Since equipped-vehicle trajectories are known, the velocity field is computed using Edie’s generalized definition [16]. The corresponding result is shown in Figure 3.5b. The qualitative agreement between subfigures a) and b) is evident – in terms of bottlenecks location, and the spatial and temporal extent of the queue. Less than 5% of the trajectories are able to obtain a velocity field (or congestion map) comparable to the one computed using the 17 detector stations.

VTL data

In addition to the trajectory data stored by each phone, VTL data was collected during the experiment. As was mentioned earlier, 30 VTLs were deployed during the experiment in the section of interest. Note that since all vehicle trajectories can be reconstructed, it is possible to artificially recreate VTL data off-line at different locations. This proves to be very useful because it allows a better analysis of the VTL data by not restricting its locations to the 30 locations deployed during the experiment.

By placing VTLs on existing loop detector station locations (17 in total), velocity measurements collected by a loop detector station every 5 minutes can be compared to the ones provided by a VTL at the same location. For comparison purposes, VTL measurements are also aggregated in 5-minute periods, and the TMS is computed for each period. Using data from the 17 VTLs, the velocity field is reconstructed using the same method described before for the loop detector stations (see Figure 3.5c). The velocity map captures the same main features captured by the loop detector velocity field. Even though both sensors provide qualitatively similar information, there is some discrepancy in the velocity value they report (suggested by the difference in colors observed at certain times and locations). The field in Figure 3.5d was constructed using the 30 VTLs deployed during the experiment, and it is shown here just for comparison. The different level of granularity among the plots
Figure 3.5: Velocity field in (mph) using a) 17 loop detector stations, b) vehicle trajectories and Edie’s speed definition, c) 17 VTLs at the loop detector station locations, and d) 30 VTLs equally spaced.

is explained by the different number of detector stations deployed in each case (17 loop detector stations/VTLs vs. 30 VTLs).

Since ground truth velocity is not known, the accuracy of VTL velocity measurements cannot be directly assessed. Note that loop detector measurements are usually
considered as ground truth. However, it is known that they sometimes include significant errors. For this reason, it was decided not to use them as ground truth. Instead, ground truth travel times between Decoto Rd. and Winton Ave from 10:45am to 5pm are extracted from high definition video cameras using license plate re-identification. A total of 4,268 vehicles were re-identified between Decoto Rd. and Winton Ave., which account for 10% of the flow at these two locations between 11am and 5pm.

Velocity fields constructed with VTL and loop detector measurements can be integrated to compute travel time\(^3\), which can be used to assess which measurements are more likely to be closer to ground truth. Figure 3.6 shows ground truth travel times, and the travel times computed by integrating both the VTL and loop detector velocity fields. Note that at 3pm, the left most lane becomes a HOV lane, which explains the points traveling the section faster than the rest of the traffic after 3pm.

![Figure 3.6: Travel time (in minutes) between Decoto Rd. and Winton Ave. Dots correspond to individual vehicle travel times (4,268 in total), collected manually using video cameras at both ends of the section of interest.](image)

Both estimates replicate the main trend observed in the travel time during the

\(^3\)This a-posteriori travel time estimation method is also known as dynamic travel time or the walk the speed matrix method.
day. The VTL estimates, however, also adequately reproduce the value of travel times on general purposes lanes. Loop detector estimates tend to underestimate travel times (they tend to overestimate velocities).

Travel times computed with the VTL velocity field are in better agreement with ground truth travel times than loop detector travel times. This suggests that the VTL velocity field is more likely to be closer to the actual velocity experienced by the vehicles, and therefore more accurate, than the loop detector velocity field. That is, the accuracy of this technology is such that a low proportion of equipped vehicles can often provide more accurate measurements of velocity than loop detectors – which sample (eventually) all vehicles. This has to be kept in mind when loop detector data is considered as ground truth, especially for an assessment of alternative data sources.

Because of the previous considerations, loop detector station measurements are not considered as ground truth in this study. A data analysis is carried out only to observe the main features of both types of measurements and not to determine the accuracy of measurements.

For each location, the velocity measurements have been classified into three classes according to the VTL or loop detector station velocity reported in the corresponding 5-minute period ($v_{\text{VTL}}$ or $v_{\text{loop}}$, respectively): congested conditions ($v_{(i)} \leq 40$ mph), free flow conditions ($v_{(i)} \geq 55$ mph), and the transition between both ($40 < v_{(i)} < 55$). Tables 3.2 and 3.3 show the mean velocity, its standard deviation, and the number of measurements used to compute these quantities in each class, for each location, and for both the VTL and loop detector station measurements, respectively.

The means and standard deviations computed for both sensors show significant differences in some cases. For a significant number of observations, loop detector stations tend to provide higher velocity measurements, suggesting a bias of the detector, but with a smaller variability than VTLs, which may be explained by the 5-minute aggregation performed in each case. As expected, higher variability in velocity is experienced during congestion.

Given the differences observed in the mean and standard deviation, the distribution of velocity measurements is not always similar between VTL and the corresponding loop detector station. Figure 3.7 shows velocity histograms at four different locations – detectors 3, 6, 8, and 12. Histograms at the top correspond to VTL velocity measurements, while the ones at the bottom correspond to loop detector station velocity measurements. Both the
Table 3.2: Mean and standard deviation (both in mph) of 5-minute VTL velocity measurements ($v_{VTL}$). Velocity classification is defined according to VTL velocity ($v_{VTL}$). The number of observations (i.e. 5-minute periods) on each class is also provided.

<table>
<thead>
<tr>
<th>Location</th>
<th>$v_{VTL} \leq 40$ mph</th>
<th>$40 &lt; v_{VTL} &lt; 55$ mph</th>
<th>$v_{VTL} \geq 55$ mph</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std.Dev</td>
<td># obs.</td>
</tr>
<tr>
<td>1</td>
<td>20.7</td>
<td>8.6</td>
<td>33</td>
</tr>
<tr>
<td>2</td>
<td>17.4</td>
<td>7.1</td>
<td>39</td>
</tr>
<tr>
<td>3</td>
<td>24.5</td>
<td>6.1</td>
<td>39</td>
</tr>
<tr>
<td>4</td>
<td>30.0</td>
<td>4.5</td>
<td>38</td>
</tr>
<tr>
<td>5</td>
<td>28.7</td>
<td>7.4</td>
<td>29</td>
</tr>
<tr>
<td>6</td>
<td>28.5</td>
<td>5.8</td>
<td>39</td>
</tr>
<tr>
<td>7</td>
<td>17.1</td>
<td>7.7</td>
<td>52</td>
</tr>
<tr>
<td>8</td>
<td>14.2</td>
<td>5.6</td>
<td>64</td>
</tr>
<tr>
<td>9</td>
<td>28.6</td>
<td>7.6</td>
<td>69</td>
</tr>
<tr>
<td>10</td>
<td>28.1</td>
<td>8.5</td>
<td>62</td>
</tr>
<tr>
<td>11</td>
<td>22.9</td>
<td>6.6</td>
<td>86</td>
</tr>
<tr>
<td>12</td>
<td>20.3</td>
<td>7.1</td>
<td>90</td>
</tr>
<tr>
<td>13</td>
<td>29.9</td>
<td>6.4</td>
<td>91</td>
</tr>
<tr>
<td>14</td>
<td>32.7</td>
<td>9.5</td>
<td>43</td>
</tr>
<tr>
<td>15</td>
<td>29.2</td>
<td>7.9</td>
<td>16</td>
</tr>
<tr>
<td>16</td>
<td>36.6</td>
<td>1.5</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>32.0</td>
<td>7.8</td>
<td>7</td>
</tr>
<tr>
<td>All</td>
<td>24.6</td>
<td>9.0</td>
<td>799</td>
</tr>
</tbody>
</table>
Table 3.3: Mean and standard deviation (both in mph) of 5-minute loop detector station velocity measurements ($v_{\text{loop}}$). Velocity classification is defined according to loop detector station velocity ($v_{\text{loop}}$). The number of observations (i.e. 5-minute periods) on each class is also provided.

<table>
<thead>
<tr>
<th>Location</th>
<th>$v_{\text{loop}} \leq 40$ mph</th>
<th>$40 &lt; v_{\text{loop}} &lt; 55$ mph</th>
<th>$v_{\text{loop}} \geq 55$ mph</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev</td>
<td># obs.</td>
</tr>
<tr>
<td>1</td>
<td>31.7</td>
<td>3.5</td>
<td>32</td>
</tr>
<tr>
<td>2</td>
<td>26.1</td>
<td>4.0</td>
<td>39</td>
</tr>
<tr>
<td>3</td>
<td>36.3</td>
<td>2.0</td>
<td>34</td>
</tr>
<tr>
<td>4</td>
<td>38.4</td>
<td>1.0</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>31.1</td>
<td>3.7</td>
<td>29</td>
</tr>
<tr>
<td>6</td>
<td>29.7</td>
<td>3.1</td>
<td>37</td>
</tr>
<tr>
<td>7</td>
<td>18.8</td>
<td>6.4</td>
<td>57</td>
</tr>
<tr>
<td>8</td>
<td>21.9</td>
<td>4.2</td>
<td>61</td>
</tr>
<tr>
<td>9</td>
<td>32.8</td>
<td>5.5</td>
<td>62</td>
</tr>
<tr>
<td>10</td>
<td>31.3</td>
<td>5.4</td>
<td>73</td>
</tr>
<tr>
<td>11</td>
<td>26.9</td>
<td>4.3</td>
<td>90</td>
</tr>
<tr>
<td>12</td>
<td>27.4</td>
<td>4.1</td>
<td>92</td>
</tr>
<tr>
<td>13</td>
<td>30.2</td>
<td>6.8</td>
<td>98</td>
</tr>
<tr>
<td>14</td>
<td>37.3</td>
<td>2.3</td>
<td>10</td>
</tr>
<tr>
<td>15</td>
<td>-</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>-</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>39.1</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>All</td>
<td>28.2</td>
<td>6.5</td>
<td>628</td>
</tr>
</tbody>
</table>
free flow and congested mode can be identified at some locations.

Some of the 17 VTLs deployed generate similar velocity profiles as loop detector stations, but some other exhibit significant differences. Figure 3.8 shows a time-series of loop detector station and VTL velocity measurements for four different locations with changing proportion or penetration rates during the day\(^4\) (see subfigure e). Locations on the figure correspond to detectors 1, 4, 7, and 17.

Both loop detector station and VTL measurements differ from each other, and the level of discrepancy varies with time, location, penetration rate, and traffic conditions (i.e. velocity). Differences between both measurements can be explained by several factors, among them:

- Loop detectors station and VTLs compute instantaneous velocity in different ways, and they have different measurement errors. While (double) loop detector stations use the travel time between dual coils to compute the speed for each vehicle, VTLs use GPS-computed velocity. Also, different aggregation methods explain some of the difference in the variability of the measurements (loop detector stations compute the TMS every 30 seconds, and the ten values computed during 5 minutes are then averaged again to obtain the 5-minute average velocity; VTLs compute the TMS using the data collected during each 5-minute period).

- VTLs collect velocity from a proportion of all vehicles crossing that location, while loop detector stations collect data from (eventually) all the vehicles. If this proportion is too small, it might not be statistically representative of the entire population.

- Drivers selected for the experiment might exhibit a specific bias, either because of their selection or because of their looping behavior\(^5\). Drivers hired for the experiment are not necessarily a proper statistical sample of the population. The 165 drivers were UC Berkeley students over 21, which may constitute a biased sample of the driving population. A specific bias can be observed at some locations close to the off-ramps used during the experiment, where VTL velocity measurements are always lower than the loop detector station velocity measurements (Figure 3.8c). However, this bias is not observed at some other locations (Figure 3.8d). Therefore, the discrepancy is

\(^4\)The penetration rate is the proportion of GPS equipped vehicles in the total flow. The next subsection describes how this rate is computed.

\(^5\)This reason applies only for this experimental case.
Figure 3.7: Velocity histograms at four different locations: a) detector 3, b) detector 6, c) detector 8, and d) detector 12. Graphs at the top (x.1) correspond to VTL measurements. Graphs at the bottom (x.2) correspond to loop detector station measurements. $v_f$ and $s_f$ are the mean and standard deviation for free flow observations, while $v_c$ and $s_c$ are those for congested observations.
most likely due to \( i \) bias in the detector, or \( ii \) test driver dynamics before exiting the mainline of the highway.

Table 3.4 shows the mean and standard deviation of the absolute difference between the loop detector station and VTL velocity measurements. For each location, measurements have been classified as before (and according to the VTL velocity \( v_{VTL} \) reported).

The discrepancy between VTL and loop detector station measurements is higher
Table 3.4: Mean and standard deviation of the absolute difference between 5-minute VTL and loop detector station velocity measurements. Velocity classification is defined according to VTL velocity ($v_{VTL}$). The number of observations (i.e. 5-minute periods) on each class is also provided.

<table>
<thead>
<tr>
<th>Location</th>
<th>$v_{VTL} \leq 40$ mph</th>
<th>40 &lt; $v_{VTL}$ &lt; 55 mph</th>
<th>$v_{VTL} \geq 55$ mph</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std.Dev</td>
<td># obs.</td>
</tr>
<tr>
<td>1</td>
<td>12.2</td>
<td>5.8</td>
<td>33</td>
</tr>
<tr>
<td>2</td>
<td>9.4</td>
<td>5.2</td>
<td>39</td>
</tr>
<tr>
<td>3</td>
<td>12.3</td>
<td>5.3</td>
<td>39</td>
</tr>
<tr>
<td>4</td>
<td>17.0</td>
<td>5.9</td>
<td>38</td>
</tr>
<tr>
<td>5</td>
<td>5.8</td>
<td>5.3</td>
<td>29</td>
</tr>
<tr>
<td>6</td>
<td>4.0</td>
<td>3.6</td>
<td>39</td>
</tr>
<tr>
<td>7</td>
<td>4.3</td>
<td>3.4</td>
<td>52</td>
</tr>
<tr>
<td>8</td>
<td>8.9</td>
<td>4.4</td>
<td>64</td>
</tr>
<tr>
<td>9</td>
<td>6.0</td>
<td>4.4</td>
<td>69</td>
</tr>
<tr>
<td>10</td>
<td>4.8</td>
<td>3.7</td>
<td>62</td>
</tr>
<tr>
<td>11</td>
<td>5.0</td>
<td>3.6</td>
<td>86</td>
</tr>
<tr>
<td>12</td>
<td>7.5</td>
<td>5.0</td>
<td>90</td>
</tr>
<tr>
<td>13</td>
<td>18.6</td>
<td>5.0</td>
<td>91</td>
</tr>
<tr>
<td>14</td>
<td>11.1</td>
<td>6.4</td>
<td>43</td>
</tr>
<tr>
<td>15</td>
<td>27.2</td>
<td>8.9</td>
<td>16</td>
</tr>
<tr>
<td>16</td>
<td>20.9</td>
<td>5.1</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>22.8</td>
<td>7.8</td>
<td>7</td>
</tr>
</tbody>
</table>
for lower velocities. Differences observed among locations suggest that some detectors are either biased or not computing the velocity properly.

Finally, Figure 3.9 plots the VTL versus the loop detector station velocity measurements for all the observations collected at the 17 locations.

![Figure 3.9: Loop detector station vs. VTL velocity measurements (all locations). Dotted lines are the ±5 mph thresholds.](image)

For velocities below 40 mph, 31% of those observations have an absolute difference of less than 5 mph. This number reaches 70% for high velocities (over 55 mph). As aforementioned, loop detector station velocity measurements tend to be higher than VTL measurements, which explain the smaller travel times computed with the loop detector station velocity field – shown before in Figure 3.6.

**Penetration rate.** The penetration rate refers to the proportion of equipped vehicles in the total flow. This proportion can be computed by placing VTLs on each of the 17 existing loop detector station locations and dividing the VTL count by the loop detector station count every 5 minutes.

During the experiment, the penetration rate changes over time and space, as shown
in the penetration rate map in Figure 3.10. Locations that are traveled by vehicles from the three loops – such as between Decoto Rd. (postmile 21) and Tennyson Rd. (postmile 26) in the morning and between Alvarado-Niles Rd. (postmile 23.3) and Tennyson Rd. in the afternoon – experience the highest proportion of equipped vehicles during the day. Locations at the ends of the section – such as between CA92 (postmile 27) and Winton Ave. (postmile 27.5) during the whole day – are traveled by only one third of the equipped vehicles and thus present the lowest proportions during the day.

Penetration rates for locations between Decoto Rd. and Winton Ave. can be seen in Figure 3.11. Circles in part a) and c) of the figure represent the average penetration rate along the section of interest during the morning and the afternoon, respectively. The range corresponds to one standard deviation below and over the mean. The histograms in part b) and d) include all the 17 locations for the morning and afternoon periods, respectively.

During the morning, less than 3% of the 5-minute periods have no observations, and in the afternoon, that number goes down to less than 1%. In addition, 50% of the periods in the morning have a penetration rate of at least 2%, while in the afternoon only 35% of the periods meet this condition. This suggests that a continuous flow of equipped vehicles was achieved, which means most of the 5-minute periods contain at least one vehicle crossing each location.
Figure 3.11: Average penetration rate (in %) over time at existing detector station locations during the a) morning and c) afternoon. The range is one standard deviation below and over the mean. Traffic flows from left to right. Histogram of the penetration rate including all the 17 locations during the b) morning and d) afternoon.

3.3 Final comments

The *Mobile Century* field experiment presented in this chapter was conceived as a proof of concept of a traffic monitoring system based on GPS-enabled mobile phones. It has been shown that small penetration rates are sufficient to collect valuable and meaningful information. The sampling strategy during the experiment is based on the concept of VTLs, which allows a detailed analysis of the accuracy of the data computed by GPS devices. It also provides enough data for traffic monitoring purposes while managing the privacy of participants.

The experiment demonstrates the feasibility of the proposed system for real-time traffic monitoring, in which GPS-enabled mobile phones can be used as traffic sensors,
providing their velocity at different points on the freeway.

The way in which the experiment was conceived allows the comparison of the velocity measurements collected by both VTLs and loop detector stations, as well as the computation of the penetration rate achieved during the day. The results highlight the tremendous potential of this type of data: even without fixed detectors, this data has the potential to supply transportation agencies with loop detector quality (and potentially better) data, at no additional public infrastructure cost.

Ground truth in terms of velocity is not known. However, a comparison of travel times generated by VTL and travel times derived from loop detector station velocity measurements suggests that VTL measurements are more likely to be closer to the actual velocities observed on the field. For this reason, loop detector data was not used as a benchmark, and only a comparison with travel times was carried out to assess the accuracy of the data. The comparison suggests the presence of some bias in the velocity estimation for some loop detector stations, showing sometimes significant differences from the VTL measurements. Because of the different 5-minute aggregation methods used, VTL measurements exhibit more variability than loop detector station measurements.

An average penetration rate between 2-3% was achieved during the experiment, which is viewed as realistic, considering that GPS-enabled cell phones are expected to penetrate the market rapidly in the near future. In addition, the quality of measurements will increase with the evolution of GPS technology itself, thus opening new opportunities for smartphone-based monitoring systems.
Chapter 4

Traffic state estimation using Lagrangian data

The previous chapter has demonstrated the feasibility of using GPS-enabled cell phones to monitor traffic on freeways. In this type of system, sensors send sporadic observations in time and space, depending on the sampling strategy. The present chapter describes methods to use these observations (or Lagrangian data) for traffic state estimation purposes. It describes three methods developed as part of this thesis, which were implemented and for which numerical results are subsequently presented in the next chapter.

4.1 Background

4.1.1 Continuous flow model

In the 1950’s, Lighthill and Whitman [31] and Richards [39] independently proposed a continuous flow model to describe the evolution of traffic on an infinite road. The first order partial differential equation proposed (referred here as the LWR PDE) is a scalar hyperbolic conservation law based on the conservation of vehicles, which relates density on the road and its flow:

$$\frac{\partial}{\partial t}k(x,t) + \frac{\partial}{\partial x} q(k(x,t)) = 0$$  \hspace{1cm} (4.1)

The function $k(x,t)$ represents the vehicle density in vehicles per unit length and $q(k(x,t))$ is the flux function in vehicles per hour (or the fundamental diagram, which is assumed to be triangular in this work) at location $x$ and time $t$. 

For a finite road, the boundary conditions are given at the upstream and downstream ends of the section \((x = a\) and \(x = b\), respectively). Ideally, they would be directly prescribed as values \(k_a(t)\) and \(k_b(t)\) of the density.

\[
k(a, t) = k_a(t) \quad \text{and} \quad k(b, t) = k_b(t)
\] (4.2)

The initial condition corresponds to the density along the road at the beginning of the period of analysis:

\[
k(x, 0) = k_0(x), \quad \text{with} \quad x \in (a, b)
\] (4.3)

For a proper characterization of the solution to equation (4.1), weak boundary conditions are required, and strong boundary conditions in (4.2) cannot be used as such. Boundary conditions only apply on the boundary of the section where characteristics are entering the computational domain. The weak boundary conditions for the specific case of the LWR PDE can be found in [9] for general hyperbolic conservation laws and in [44] for the specific context of highways:

\[
\begin{cases}
  k(a, t) = k_a(t) & \text{or} \\
  q'(k(a, t)) \leq 0 \quad \text{and} \quad q'(k_a(t)) \leq 0 \quad \text{or} \\
  q'(k(a, t)) \leq 0 \quad \text{and} \quad q'(k_a(t)) \geq 0 \quad \text{and} \quad q(k(a, t)) \leq q(k_a(t))
\end{cases}
\]

and

\[
\begin{cases}
  k(b, t) = k_b(t) & \text{or} \\
  q'(k(b, t)) \geq 0 \quad \text{and} \quad q'(k_b(t)) \geq 0 \quad \text{or} \\
  q'(k(b, t)) \geq 0 \quad \text{and} \quad q'(k_b(t)) \leq 0 \quad \text{and} \quad q(k(b, t)) \leq q(k_b(t))
\end{cases}
\]

These conditions state that the upstream boundary condition is relevant (and thus can be applied in the strong sense) only when a free flow condition is observed at that point. Otherwise, the boundary condition is irrelevant and conditions are dictated by downstream traffic. The opposite is true for the downstream boundary, in which the boundary condition is relevant only if congestion is observed at that point.

In practice, these boundary conditions are implemented using ghost cells. These cells correspond to the input and output cells as proposed in [13], and will be explained next.

### 4.1.2 Discretization method

For implementation purposes, the LWR PDE needs to be discretized. To this end, the freeway section is divided into \(I\) cells (each one of length \(\Delta x\) distance units, and indexed
by \(i\). Time is divided into \(H\) time steps (each one of length \(\Delta t\) time units, and indexed by \(h\)). In order to meet the Courant-Friedrichs-Lewy (CFL) stability condition [30], which states that a vehicle traveling at the free flow speed \(v_f\) cannot traverse more than one cell in one time step, the condition \(\Delta t \cdot v_f \leq \Delta x\) should be met. At every time step, the model estimates the density in each cell according to the following expression:

\[
k_i^{h+1} = k_i^h - r \left( q_{i+1}^h - q_i^h \right) \quad i = 1, 2, \ldots, I \text{ and } h = 0, 1, \ldots, H - 1.
\]

(4.4)

The parameter \(r\) is the inverse of the speed needed to travel one cell in exactly one time step (i.e. \(r = \frac{\Delta t}{\Delta x}\)), while \(k_i^h\) is the density in cell \(i\) at time step \(h\). The variable \(q_i^h\) is the flow into cell \(i\) between time \(h\) and \((h + 1)\), and depends nonlinearly on the density of cells \((i - 1)\) and \(i\).

The cell transmission model proposed by Daganzo [13], in the context of highways, or the Godunov scheme [19], for general scalar hyperbolic first order conservation laws, can be used to solve equation (4.4), since both yield the same results when the fundamental diagram is assumed to be triangular. In the latter case, the flow is defined as \(q_i^h = q_G(k_i^{h-1}, k_i^h)\), where \(q_G\) is the Godunov flux, defined as follows:

\[
q_G(k_1, k_2) = \begin{cases} 
q(k_2) & \text{if } k_c < k_2 < k_1, \\
q(k_c) & \text{if } k_2 < k_c < k_1, \\
q(k_1) & \text{if } k_2 < k_1 < k_c, \\
\min(q(k_1), q(k_2)) & \text{if } k_1 \leq k_2.
\end{cases}
\]

(4.5)

The parameter \(k_c\) is the critical density, which corresponds to the density at the maximum flow.

As explained earlier, weak boundary conditions are required for a proper characterization of the solution of the LWR PDE in (4.1). In the implementation, one ghost cell is inserted at each boundary of the section. The ghost cells contain the boundary conditions, and they allow the first and last cells of the computational domain to be updated depending on the existing traffic conditions (free flow, congested, or a combination of the two). This idea was proposed in [13], and it was successfully implemented and tested with traffic data collected from loop detectors in an earlier work [44]. The equations for the ghost cells \((i = 0\) and \(i = I + 1)\) and for the initial conditions \((h = 0)\) are given by equation (4.6) and (4.7),
respectively:

\[
k^h_0 = \frac{1}{\Delta t} \int_{(h-1)\Delta t}^{h\Delta t} k_a(t) dt \quad \text{and} \quad k^h_{I+1} = \frac{1}{\Delta t} \int_{(h-1)\Delta t}^{h\Delta t} k_b(t) dt \quad h = 1, 2, \ldots, H \quad (4.6)
\]

\[
k^0_i = \frac{1}{\Delta x} \int_{x_{i-1}}^{x_i} k_0(x) dx \quad i = 1, 2, \ldots, I. \quad (4.7)
\]

### 4.2 Proposed methods

In order to determine the state of the traffic (vehicle accumulation or density, in this case) along the section of interest using Lagrangian data, specific data assimilation methods are developed. If accurate traffic data from all locations was available at all times, traffic state estimation would not be needed since the state can be directly inferred from the data. In practice, however, data is not always available from all the locations and is, in general, very sparse.

As explained in Chapter 2, different approaches can be used to obtain estimates for locations without measurements. The approaches adopted here use traffic flow models, based on vehicle accumulation (or density), instead of statistical techniques that look for spatio-temporal correlation of the data.

Two methods are proposed and discussed in the following sections, which rely on the following assumptions:

- GPS-enabled mobile phones traveling on the section of interest are sampled in time. They report their position and velocity at specific time intervals \(T\).

- Boundary conditions are known. That is, the density at both ends of the section of interest are available. This data can be provided by loop detectors. However, as we shall see in the next chapter, the availability of loop detector data at the boundaries is not a critical requirement.

- The fundamental diagram is assumed to be triangular and known.

- Information from intermediate ramps is not required. In fact, this constitutes one of the main features of the methods proposed since, in reality, information from ramps is rarely available.
4.2.1 Definitions: state, estimated, and observed variables

The state variable characterizes the state of a dynamical system. In the present case, it corresponds to vehicle density, and it is denoted by \( k(x, t) \). By definition, it satisfies the physical model, which is the LWR PDE in (4.1) in the present case.

The state estimate (or estimated variable) is the result of the state estimation process, and it is distinguished from the model variable by the use of a hat. We denote \( \hat{k}(x, t) \) the state estimate.

To perform the estimation, it is assumed that some variable is measured. In the present case, the available data consists of position and velocity measurements from GPS-equipped vehicles at different times. Therefore, a relationship between the velocity and the density is needed in order to relate the measured or observed variable and the estimated variable. This issue is addressed in the next paragraphs.

The observed position and velocity of an equipped vehicle labeled \( j \) at time \( t \) are denoted by \( s^o_j(t) \) and \( v^o_j(t) \), respectively (superscript \( o \) denotes observation). It is assumed that the observed velocity at the corresponding location and time is provided by the individual probe velocity measurement. Using \( v(x, t) \) to denote the average velocity field on the freeway, we defined the observed velocity at time \( t \) and location \( s^o_j(t) \) by:

\[
v^o (s^o_j(t), t) = v^o_j(t)
\]  

This assumption may not be appropriate when dealing with multi-lane freeways, where different lanes might have different speeds. In these cases, this problem requires a specific treatment, which should use more sophisticated models, and are out of the scope of this work\(^1\). As a first approach, we treat the freeway as a single traffic stream.

The fundamental diagram relates the flow, the density and the velocity of the flow on a section of road. In particular, this relationship can be used to infer the density from the velocity. That is, using the fundamental diagram, \( v^o(s^o_j(t), t) \) can be converted into the observed density at the point of measurement, denoted by \( k^o(s^o_j(t), t) \). In reality, \( k^o(s^o_j(t), t) \) is not the observed density but an estimate based on the speed measurement and the fundamental diagram. Since the fundamental diagram is only a model, and in reality flow-density points (or traffic states) do not necessarily lay on a line, this conversion

\(^1\)In multi-lane freeways, more than one vehicle could send a measurement from the same location \( s_j(t) \) at the same time \( t \). In this case, the observed velocity is the average of all the measurements from location \( s_j(t) \) at time \( t \).
is expected to introduce error in the observed density.

If a triangular fundamental diagram is used, a problem arises when the speed \( v^o_j(t) \) is high. In fact, if the measured speed \( v^o_j(t) > v_f \), where \( v_f \) is the free flow speed, different combinations of flow and density have the same free flow speed. Indeed, under free flow conditions it is not possible to observe the local density through speed measurements in the way described in the previous paragraph. For these cases, free flow conditions will be assumed and a “free flow density” value (denoted \( k^{FF} \)) will be used. The value for \( k^{FF} \) can change in time and space, and can be obtained from historical data. This seems reasonable considering that our main interest is to obtain accurate density estimates specifically when congestion arises.

Therefore, the observed density is given by:

\[
k^o(s^o_j(t), t) = \begin{cases} 
  k^{FF}, & \text{if } v^o(s^o_j(t), t) > v_f \\
  \frac{w-k_f}{v^o(s^o_j(t), t)+w}, & \text{if } v^o(s^o_j(t), t) \leq v_f 
\end{cases}
\]  

(4.9)

Parameters \( k_f \) and \( w \) are the jam density and the slope of the right branch of the triangular fundamental diagram, respectively. The observed velocity \( v^o(s^o_j(t), t) \) comes from equation (4.8).

### 4.2.2 Kalman filtering based method

Kalman filtering is a recursive method used to estimate the state of a discrete process governed by a linear stochastic dynamical system [8] in the presence of noisy measurements. The method assumes that the way in which the state of the system (density in this case) evolves is linear and known, and is referred to as the dynamics or state equation (which includes a process noise). Noisy measurement of the output the system (i.e. observed density) are available, and the measurement or observation equation relates the output and the state of the system. Knowing the covariance of both the process and the measurement error, the method obtains the best estimate of the state of the system in the sense of the least squares.

Kalman filtering techniques have been proposed to perform traffic state estimation in the presence of loop detector data [45, 46]. Given the nonlinearity of the model in (4.4), conventional Kalman filtering can only be applied to subsets of this dynamic, in which it is linear. As shown in [21], this technique can be extended to cases in which data is provided by mobile sensors, which is described in the next subsections.
State space representation

The state space representation of the system consists of two equations: the dynamics (or state) equation and the measurement (or observation) equation.

The dynamics equation describes how the state of the process (density) evolves over time and space. Since the constitutive equation in (4.4) is nonlinear, it needs to be linearized first. The hybrid system framework used in [36] is adopted for this purpose. Discrete event dynamics estimation is performed to identify the traffic condition or mode on the section of interest. That is, the mode of each cell (i.e. free flow or congested) needs to be determined at the beginning of each time interval.

Previous studies [36, 46] have proposed different ways to identify the mode on a short section of highway. For longer sections, the number of possible modes increases, adding complexity to the mode identification.

One alternative to identify the mode is to use the state estimates (and indirectly, the Lagrangian measurements) as in [22]. At the end of time interval \( h \), density estimates for every cell are available. These estimates consider all the observations collected until time step \( h \) (inclusive) and are referred to as the *a-posteriori* estimates at time step \( h \). The *a-posteriori* estimate at cell \( i \) at time step \( h \) is denoted by \( \hat{k}_{i}^{+,h} \). The mode chosen for each cell will depend on the value of \( \hat{k}_{i}^{+,h} \). In other words, if \( \hat{k}_{i}^{+,h} > k_c \), cell \( i \) is congested; otherwise, cell \( i \) is in free flow (\( k_c \) is the critical density). Note that the mode is being identified by using the state estimates and not by measuring the actual state of the system.

Once the mode for each cell has been identified, the flow into cell \( i \) between time step \( h \) and \( (h + 1) \), \( q_{i}^{h} \), becomes linear in the densities in cells \( (i - 1) \) and \( i \). Therefore, equation (4.4) can be written as follows:

\[
k_{h+1} = A_h \cdot k_h + B_h \cdot u_h + B_{h}^{J} \cdot k_J + B_{h}^{Q} \cdot q_{\text{max}} + w_{h} \tag{4.10}
\]

Bold letters represent a matrix or vector notation. The scalars \( k_J \) and \( q_{\text{max}} \) are the jam density and the maximum flow, respectively. The state vector at time \( h \) is defined as \( k_h = [k_{1}^{h} \ k_{2}^{h} \ \ldots \ k_{I}^{h}]^{T} \), where \( k_{i}^{h} \) is the density on cell \( i \) at time step \( h \). The vector \( u_h \) is the input vector at time \( h \), which includes the density at the boundaries of the domain, and \( w_{h} \) is the process error (caused by the fact that not all the entry and exit counts are available, for instance). Equation (4.10) is linear in the state \( k_h \). The matrices \( A_h, B_h, B_{h}^{J} \) and \( B_{h}^{Q} \) depend on the traffic conditions or the mode of the section, which shows the
importance of the mode identification step. The algebraic expression of these matrices for specific cases can be found in [36].

The measurement or observation equation projects the state vector into the measurements provided by Lagrangian sensors, $y_h$, with the one predicted by the model:

$$y_h = C_h \cdot k_h + v_h \quad (4.11)$$

The vector $v_h$ represents the measurement noise. The matrix $C_h$ is time dependent, and its size and elements depend on the location of the measurements. The matrix $C_h$ only contains zeros and ones, and the position of these numbers in the matrix $C_h$ depend on the locations at which the measurements are taken. The time dependency of $C_h$ is a major challenge and is directly linked to the Lagrangian aspect of the measurements.

In summary, the dynamics (or state) equation and the measurement (or observation) equation of the system are given by equation (4.10) and (4.11), respectively.

**Kalman filtering**

The following notation is used:

- $\hat{k}_h$: a-priori state estimate of $k_h$, where $\hat{k}_h = [\hat{k}_1^h \ \hat{k}_2^h \ \ldots \ \hat{k}_I^h]^T$,
- $\hat{k}_h^+$: a-posteriori state estimate of $k_h$, where $\hat{k}_h^+ = [\hat{k}_1^{+,h} \ \hat{k}_2^{+,h} \ \ldots \ \hat{k}_I^{+,h}]^T$,
- $P_h$: a-priori estimate error covariance, where $e_h = k_h - \hat{k}_h$ is the a-priori estimate error,
- $P_h^+$: a-posteriori estimate error covariance, where $e_h^+ = k_h - \hat{k}_h^+$ is the a-posteriori estimate error.

The difference between the a-priori and a-posteriori estimates at time step $h$ is the fact that the a-priori estimates do not take into account the observations collected at time step $h$, while the a-posteriori estimates do. That is, the a-posteriori estimate is an updated version of the a-priori estimate.
Kalman filtering provides a set of recursive equations to estimate the vector state. The equations are as follows:

\[
\begin{align*}
\hat{k}_{h+1} &= A_h \cdot \hat{k}_h + B_h \cdot u_h \quad (4.12) \\
P_{h+1} &= A_h \cdot P_h^+ \cdot A_h^T + Q \quad (4.13) \\
F_{h+1} &= P_{h+1} \cdot C_{h+1}^T \left[ C_{h+1} \cdot P_{h+1} \cdot C_{h+1}^T + R \right]^{-1} \quad (4.14) \\
\hat{k}_{h+1}^+ &= \hat{k}_{h+1} + F_{h+1} \left( y_{h+1} - C_{h+1} \cdot \hat{k}_{h+1} \right) \quad (4.15) \\
P_{h+1}^+ &= (I - F_{h+1} \cdot C_{h+1}) P_{h+1} \quad (4.16)
\end{align*}
\]

Initial conditions \( \hat{k}_0^+ \) and \( P_0^+ \) are assumed to be known. \( Q \) and \( R \) are the covariance matrices of the process and measurement error, respectively. \( F_h \) is known as the Kalman gain at time step \( h \).

**Implementation**

The state vector contains the density in each cell. At the beginning of time step \( (h+1) \), \textit{a-posteriori} estimates at time step \( h \) are available. The traffic conditions on the network at the end of time step \( h \) need to be identified in order to determine which set of matrices \( A_h, B_h, B_J^h \) and \( B_Q^h \) to use. The mode can be identified with the process outlined previously, which indirectly uses the Lagrangian observations collected. Note, however, that the mode identification is the most challenging task in the implementation of this method.

Once the mode has been identified at the beginning of time step \( (h+1) \), equation (4.12) and (4.13) are used to obtain the \textit{a-priori} density estimate and its covariance, respectively. At this point, Lagrangian data becomes available to the model, i.e. the observed local density at time \( (h+1) \) will be known for some cells (the quantity and position of the Lagrangian sensors at \( (h+1) \) will determine how many and for which cells the density is observed). With this information, the observed vector \( y_{h+1} \) and matrix \( C_{h+1} \) can be constructed. Then, the Kalman gain is computed using equation (4.14). Finally, the \textit{a-posteriori} density estimate and its covariance are obtained using equation (4.15) and (4.16), respectively. In the event that no Lagrangian observation is available at time \( (h+1) \), the matrix \( C_{h+1} \) is set equal to zero, which implies that the Kalman gain is also zero. In this case, the \textit{a-priori} and \textit{a-posteriori} density estimates are the same.
4.2.3 Filter-based heuristic method

The Kalman filtering approach described in the previous section requires the mode to be identified at each time step. Ideally, the mode should be directly observed from the observations collected. Since observations are sparse in time and space, however, the mode identification is a challenging task. In addition, the approach assumes knowledge of the process and the measurement error covariances $Q$ and $R$. All this information is required to obtain the Kalman gain, which is optimal in the sense of the least squares.

A heuristic can be developed to circumvent these issues. The difference with respect to the approach presented before is that the gain to be used is not necessarily optimal.

As before, the following notation is used:

- $\hat{k}^h_i$: a-priori density estimate at cell $i$ at time step $h$,
- $\hat{k}^{+,h}_i$: a-posteriori density estimate at cell $i$ at time step $h$,
- $k^{o,h}_i$: observed density at cell $i$ at time step $h$.

The a-posteriori density estimates at time $(h-1)$ are available at the beginning of time step $h$. With the model in (4.4), a-priori estimates at time $h$ are computed. For the cells with no observations at time step $h$, the a-posteriori estimate is the same as the a-priori estimate at the same time step $h$. For a cell $i$ for which an observation is available at time step $h$, the a-posteriori density estimate is given by the following expression:

$$\hat{k}^{+,h}_i = \hat{k}^h_i + L^h_i \cdot (k^{o,h}_i - \hat{k}^h_i)$$  (4.17)

Note the similarity with equation (4.15). The a-posteriori density is a linear combination of the a-priori and observed densities. The second term on the right hand side of equation (4.17) is adding/removing vehicles to/from cell $i$ at time step $h$. As with the Kalman gain, parameter $L^h_i$ is a correction gain that defines how many vehicles are to be added-to/removed-from cell $i$ at time step $h$ and is assumed to be positive. If the a-priori density estimate is underestimating the observed density (i.e. $k^{o,h}_i - \hat{k}^h_i > 0$), vehicles are added. If the difference is negative, vehicles are removed.

Note that $L^h_i = 1$ implies $\hat{k}^{+,h}_i = k^{o,h}_i$. In this case, the a-posteriori density estimate is the observed density. On the other hand, $L^h_i = 0$ implies $\hat{k}^{+,h}_i = \hat{k}^h_i$, which is the case when there is no observation available from $x^o$ at time $t^o$. Depending on the confidence
on the observed density $k_{i}^{o,h}$, $L_{i}^{h}$ ranges between these two values (high confidence: $L_{i}^{h} \sim 1$; low confidence: $L_{i}^{h} \sim 0$).

**Implementation issues**

As mentioned in Section 4.1.2, the freeway section is divided into $I$ cells and time is divided into $H$ time steps for implementation purposes. At time step $h$, the discrete model in equation (4.4) provides the a-priori density at the next time step ($h+1$). Note that at time step $h$, the a-posteriori density for each cell is known and used in the estimation. That is:

$$\hat{k}_{i}^{h+1} = \hat{k}_{i}^{h} - r \left( q_{i+1}^{+,h} - q_{i}^{+,h} \right) \quad i = 1, 2, \ldots, I \text{ and } h = 0, 1, \ldots, H - 1. \quad (4.18)$$

The flow $\hat{q}_{i}^{+,h}$ denotes that the flow between cells ($i-1$) and $i$ at time step $h$ is based on the a-posteriori density. That is, $\hat{q}_{i}^{+,h} = q_{i}(\hat{k}_{i-1}^{h}, \hat{k}_{i}^{h})$.

Denoting by $v_{i}^{a,h+1}$ the observed velocity at cell $i$ at time step ($h+1$), equation (4.9) is used to obtain the corresponding observed density $k_{i}^{a,h+1}$. If more than one observation is available from cell $i$ at time step ($h+1$), $v_{i}^{a,h+1}$ is the average of the observations.

For the cells with an observation available at time step ($h+1$), the a-posteriori density at time step ($h+1$) is computed using equation (4.17). For the cells with no observation available, the a-posteriori density can be assumed to be the same as the a-priori density (i.e. $L_{i}^{h} = 0$ in these cases). Now, the a-posteriori density at time step $h+1$ is used to compute the a-priori density at time step $h+2$, and the process is repeated. The process is sketched in Figure 4.1.

**Computation of $L_{i}^{h}$.** Let us assume an observation is available from cell $i$ at time step $h$. The parameter $L_{i}^{h}$ determines how many vehicles are added-to or removed-from cell $i$ at time step $h$. Therefore, it is expected to depend on the accuracy of the observations and the model. This subsection illustrates how the errors in both the a-priori and observed density should affect parameter $L_{i}^{h}$.

The estimation error at cell $i$ and time step $h$ can be defined as the difference between the corresponding actual density and the a-posteriori density at time step $h$.

$$\tilde{k}_{i}^{h} = k_{i}^{h} - \hat{k}_{i}^{+,h} = k_{i}^{h} - (1 - L_{i}^{h}) \cdot \hat{k}_{i}^{h} - L_{i}^{h} \cdot k_{i}^{a,h} \quad (4.19)$$
The parameter $L^h_i$ can be chosen such that the mean squared estimation error $E \left[ (\hat{k}^h_i)^2 \right]$ is minimized. It is assumed that both the a-priori and observed density are related to the actual density by the following expressions:

\[
\hat{k}^h_i = k_i^h + \xi^h_i \tag{4.20}
\]

\[
k_{i}^{o,h} = k_{i}^h + \phi_i^h \tag{4.21}
\]

Note that for the observed density this assumption may not be realistic in light of equations (4.8) and (4.9). Given these equations, the error in the observed density is not necessarily additive. This simplification, however, allows us to easily compute $L^h_i$ and achieve the goal of this subsection.

The errors $\xi^h_i$ and $\phi^h_i$ are assumed to be two independent random processes that represent the deviation of both the a-priori and observed density from the actual value.

The error for the a-priori density is related to the fact that the conservation of vehicles is not met\(^2\). Model and measurement errors also account for $\xi$.

\(^2\text{This is due to the fact that data from ramps is assumed to be unavailable.}\)
The error in the computation of the observed density includes at least three sources. The first source of error has to do with equation (4.8), when the velocity of an individual vehicle is assumed to correspond to the velocity at that given location. This error may be higher for low velocities, when the variability in speed among vehicles is high. The second source of error is related to the fact that the fundamental diagram is not exact. Thus, the velocity-to-density conversion, as expressed in equation (4.9), introduces error. This error may not be negligible for free flow because of the approximation made for these cases. The third source of error corresponds to the measurement error in the velocity $v_j(t)$, which is expected to be small given the accuracy of GPS.

Substituting equations (4.20) and (4.21) in equation (4.19), we find:

$$\tilde{k}_h^i = -(1 - L_h^i) \cdot \xi_i^h - L_h^i \cdot \phi_i^h$$ (4.22)

Denoting the variances of $\xi_i^h$ and $\phi_i^h$ by $\sigma_{\xi_i^h}^2$ and $\sigma_{\phi_i^h}^2$, respectively, the mean square estimation error is given by:

$$E \left[ \left( \tilde{k}_h^i \right)^2 \right] = \left( 1 - L_h^i \right)^2 \cdot \sigma_{\xi_i^h}^2 + \left( L_h^i \right)^2 \cdot \sigma_{\phi_i^h}^2$$ (4.23)

The value of $L_h^i$ that minimizes equation (4.23) is the following:

$$L_h^i = \frac{\sigma_{\phi_i^h}^2}{\sigma_{\xi_i^h}^2 + \sigma_{\phi_i^h}^2} = \frac{1}{1 + \frac{\sigma_{\phi_i^h}^2}{\sigma_{\xi_i^h}^2}}$$ (4.24)

If the error in the observed density is very low compared to the error in the a-priori density (i.e. $\frac{\sigma_{\phi_i^h}^2}{\sigma_{\xi_i^h}^2}$ is low), $L_h^i \sim 1$. Therefore, for very accurate measurements of the density obtained from GPS phones, $L_h^i$ is expected to be close to one. Note that $L_h^i$ may differ across location and/or time if the error variances change.

Equation (4.24) assumes that an observation is available from every cell and any time step. Therefore, the best value for cell $i$ at time step $h$, $L_h^i$, is only affected by the error in the a-priori and observed densities at cell $i$ and time step $h$. If a given cell $i_N$ has no observation at a given time step $h_N$, the model in equation (4.18) should propagate the effect of neighboring observations into $(i_N, h_N)$. If there are only a few observations, the best possible value of $L_h^i$ for those locations with an observation will be affected. As a consequence, it is expected that $L_h^i$ should not only depend on the accuracy of the observations, but also on the total number of observations per cell per time step.
Unlike the Kalman filtering approach, the heuristic method does not require one to identify the mode and the knowledge of the error covariances. The gain used to correct the estimates using the observations is obtained heuristically, and it is not necessarily optimal. Chapter 5 presents an assessment of this heuristic method, and shows how the gain depends on the number and accuracy of the observations.

4.2.4 Newtonian relaxation method

Both the Kalman filtering approach and the heuristic method presented before assume that an observation obtained from cell \( i \) at time step \( h \) only corrects or updates the density on the corresponding cell and time step. The flow model propagates the effect in time and space. However, if cell \( i \) is congested at time step \( h \) (according to the observation), neighboring cells are expected to also be congested for a given period of time (congestion takes time to vanish). This feature might be desirable specially when the number of observations is low. The method present in this subsection explores this idea.

The Newtonian relaxation (or nudging) method is a simple heuristic method that has been used for data assimilation in the field of environmental fluid mechanics [4]. In oceanography, GPS-equipped drifters are used to estimate the velocity field of rivers, using shallow waters models. The extension of this technique to transportation engineering problems appears to be very promising, since GPS-equipped vehicles are similar to the drifters, and we aim to estimate the state of the freeway in terms of the vehicle density.

The Newtonian relaxation method relaxes the dynamic model of the system towards the observations. To this end, a source term called the nudging term, proportional to the difference between the estimated and observed state, is included in the constitutive equation of the model, which in the present case is the LWR PDE in (4.1):

\[
\frac{\partial \hat{k}}{\partial t} + \frac{\partial}{\partial x} \left( q(\hat{k}) \right) = \sum_{j=1}^{J} \sum_{t_p^j \in S^j_t} \lambda(x - s_j(t_p^j), t - t_p^j) \cdot \left( k^o(s_j(t_p^j), t_p^j) - \hat{k}(s_j(t_p^j), t_p^j) \right) \tag{4.25}
\]

The summation over the index \( j \) in the RHS of equation (4.25) accounts for the \( J \) different vehicles acting as Lagrangian sensors, while the second summation includes all the observations sent by each vehicle \( j \) before the current time \( t \). The expression in equation (4.25) assumes that vehicle \( j \) sends observations at times \( t_p^j \in S^j_t \), where \( S^j_t \) represents the set of times until \( t \) at which measurements from vehicle \( j \) are performed and used for data assimilation (note that necessarily \( t_p^j < t \)).
A common expression for the nudging factor $\lambda(\delta x, \delta t)$ (that is useful for our purposes) can be found in [25] and is given by:

$$\lambda(\delta x, \delta t) = \begin{cases} 
\frac{1}{T_a} \exp\left(-\left(\frac{\delta x}{X_M}\right)^2\right) \exp\left(-\frac{\delta t}{T_M}\right) & \text{if } |\delta x| \leq X_M \text{ and } 0 < \delta t \leq T_M \\
0 & \text{otherwise}
\end{cases}$$

(4.26)

The nudging factor represents the weight of each observation, and it becomes negligible away from the measurement location and after the measurement time. In fact, the factor dies out in time and space based on the time and space scales $T_M$ and $X_M$, respectively. Therefore, close to where and when the observation is made, $\lambda(\delta x, \delta t)$ nudges the solution towards the observations. The parameter $T_a$ has units of time and determines the strength of the nudging factor. Similarly to the parameter $L$ presented before, $T_a$ is expected to depend on the number of observations available and their accuracy.

**Implementation issues**

The discretization of the nudging term in the RHS of equation (4.25) needs to be added to equation (4.4). The final expression for the discretized model is as follows:

$$\hat{k}_{i}^{h+1} = \hat{k}_{i}^{h} - r\left(q_{i+1}^{h} - q_{i}^{h}\right) + \Delta t \sum_{j=1}^{J} \sum_{t_{p}^{j} \in S_{h}^{\Delta t}} \lambda(x_{i} - s_{j}(t_{p}^{j}), h\Delta t - t_{p}^{j}) \left[\hat{k}_{c_{j}p}^{m_{j}p} - \hat{k}_{c_{j}p}^{m_{j}p}\right]$$

$$i = 1, 2, \ldots, I \text{ and } h = 0, 1, \ldots, H - 1.$$  

(4.27)

The notation introduced in equation (4.27) is explained below:

- $x_{i}$: location of the beginning of cell $i$, $x_{i} = x_{0} + (i - 1)\Delta x$, for $i = 1, 2, \ldots, I$, where $x_{0}$ is the beginning of the section of interest,

- $c_{jp}$: cell index corresponding to location $s_{j}(t_{p}^{j})$, which is the location where vehicle $j$ is at time $t_{p}^{j}$ when its $p$-th observation is sent, $c_{jp} = \left\lceil \frac{s_{j}(t_{p}^{j}) - x_{0}}{\Delta x} \right\rceil$, where $s_{j}(t_{p}^{j}) > x_{0}$,

- $m_{jp}$: time step corresponding to time $t_{p}^{j}$, which is the time (in time step units) when the $p$-th observation from vehicle $j$ occurs, $m_{jp} = \left\lceil \frac{t_{p}^{j}}{\Delta t} \right\rceil$.

As before, the flow $q_{i}^{h}$ is computed with the Godunov flux in equation (4.5). Weak boundary conditions are required for a proper characterization of the solution of the LWR
PDE in (4.1). Therefore, the concept of ghost cells is used again. The equations for the ghost cells \((i = 0\) and \(i = I + 1\)) and for the initial conditions \((h = 0)\) are given by equation (4.6) and (4.7), respectively.

4.3 Final comments

**Computation of the observed density.** In all three methods, the observed density is computed using equation (4.9), which is expected to be different from the actual value in the free flow case. Note, however, that if density estimates are used to obtain travel times (by computing the velocity), the value used for \(k^{FF}\) would not affect the travel time estimates significantly. Therefore, travel times estimates might be more accurate than density estimates under free flow conditions.

**Comparison of the filter-based heuristic method and the Newtonian relaxation method.** Both methods are conceptually similar. They add or remove vehicles depending on the difference between an estimated density and the observed density computed using GPS data. They differ, however, in the way this difference is used. In the Filter-based heuristic method, density at location \(x\) at time \(t\) can only be directly affected by observations collected at the same location \(x\) and time \(t\). The effect of the observation is propagated in time and space using the flow model, indirectly affecting other locations. In the Newtonian relaxation method, the density at location \(x\) at time \(t\) is directly affected by any observation collected between locations \((x - X_M)\) and \((x + X_M)\) and between times \((t - T_M)\) and \(t\). Therefore, the effect of each observation is propagated in time and space directly through the nudging term, but also through the flow model (this may be useful when the number of observations is low).

However, even though both approaches are similar, they are not the same in practice and they will yield different results. This is true even for the specific case of the Newtonian relaxation method that is comparable to the first approach (when \(X_M\) and \(T_M\) are small), as explained next.

If the nudging factor is set such that observations from location \(x\) at time \(t\) affect only density at location \(x\) and time \(t\) (i.e. \(X_M\) and \(T_M\) are small), this factor would be similar to the parameter \(L\) in the first approach. In fact, when \(X_M\) and \(T_M\) are small, the
nudging factor in (4.26) acts like a Dirac function. Equation (4.27) becomes:

\[
\hat{k}_{i}^{h+1} = \hat{k}_{i}^{h} - r \left( q_{i+1}^{h} - q_{i}^{h} \right) + \Delta t \cdot \frac{1}{T_{a}} \cdot \left[ k_{i}^{o,h} - \hat{k}_{i}^{h} \right]
\]

\[
i = 1, 2, \ldots, I \text{ and } h = 0, 1, \ldots, H - 1.
\]

(4.28)

Rearranging terms in equation (4.28):

\[
\hat{k}_{i}^{h+1} = \hat{k}_{i}^{h} + \Delta t \cdot \frac{1}{T_{a}} \left( k_{i}^{o,h} - \hat{k}_{i}^{h} \right) - r \left( q_{i+1}^{h} - q_{i}^{h} \right)
\]

\[
i = 1, 2, \ldots, I \text{ and } h = 0, 1, \ldots, H - 1.
\]

(4.29)

The only difference between equation (4.4) and (4.29) is in the dependency of the flux function. In the first approach, this function depends on the a-posteriori density \(\hat{k}_{i}^{h}+\), while this function depends on the estimated density \(\hat{k}_{i}^{h}\) in the second case. That is, the filter-based heuristic method computes a new (or a-posteriori) density for the location from which an observation is available at the corresponding time, while the Newtonian relaxation method does not. Rather, the Newtonian relaxation method uses the difference between the observation and the estimation to add/remove vehicles at the next time step.

Implications on the conservation of vehicles. The proposed methods add/remove vehicles to/from a cell depending on the relative values of the observed and estimated density. If perfect counts were known exactly at every entry and exit point of the network, this addition/removal of vehicles would violate the conservation of vehicles. In practice, however, on- and off-ramps are rarely equipped with loop detectors. Moreover, the measurements are never perfect because of measurement errors. The present methods can thus be used to incorporate Lagrangian data in place of this missing loop detector data. Thus, these methods bypass the modeling of networks: for mainlines with on- and off-ramps, they replace the merge-diverge junctions by Lagrangian data incorporation.

When no loop detector data is available from ramps, numerical simulations are simply under determined, because of the lack of inflow and outflow information. When data from ramps is available, the data might be inconsistent with the model (because of measurement errors), in particular that of vehicle conservation. In any case, the methods steer the state of the model locally (in \(x\) and \(t\)) towards the Lagrangian measurements,
which is a way to reestablish a value of the state that is closer to the actual state of the system. Therefore, this process compensates for the lack of inflow and outflow information or the corresponding inaccuracy.

Because of the considerations discussed above, the conservation of vehicles that incorporates the on- and off-ramps is replaced by the addition/removal of vehicles, which performs the required local adjustments based on measurements.
Chapter 5

Evaluation of the methods with traffic data

The filter-based heuristic method and the Newtonian relaxation method proposed in Section 4.2 are implemented on two datasets to assess their performance. In the first dataset, we have full knowledge of all positions and speeds of all the vehicles during the experiment, which enables an extensive validation of the methods. In the second dataset, we use GPS data obtained from the field deployment described in Chapter 3, and thus demonstrate the applicability of the methods to this novel way of gathering traffic data.

5.1 Implementation with NGSIM data

Traffic data from the Next Generation Simulation (NGSIM) project [1] is used to evaluate the proposed approaches. The NGSIM data has been extracted from video cameras, which provides ground truth trajectories for all vehicles. The data consists of the trajectories of all vehicles entering a stretch of the US Highway 101S in Los Angeles, CA, during a period of 45 minutes (from 7:50am to 8:35am on June 15, 2005). The site is approximately 0.4 miles in length, with 5 mainline lanes (see Figure 5.1). An auxiliary lane exists between the on-ramp and the off-ramp. The section is congested almost all the time (transition from free flow to congestion happens during the first 10-12 minutes of the dataset).

For implementation purposes, the five mainline lanes were considered. The section of interest is 1920 ft in length (gray area in Figure 5.1) and has been divided into $I = 16$
cells of $\Delta x = 120$ ft each, and the total simulation time (2610 seconds) has been divided into $H = 2,175$ time steps of $\Delta t = 1.2$ seconds each. Parameters of the fundamental diagram were extracted from the PeMS system [2] and are shown in Table 5.1 (a triangular fundamental diagram is assumed).

Table 5.1: Parameters of the fundamental diagram for the NGSIM site.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Notation</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum flow</td>
<td>$q_M$</td>
<td>2040 vphpl</td>
</tr>
<tr>
<td>Jam density</td>
<td>$k_J$</td>
<td>205 vpmpl</td>
</tr>
<tr>
<td>Critical density</td>
<td>$k_C$</td>
<td>30 vpmpl</td>
</tr>
<tr>
<td>Free flow speed</td>
<td>$v_f$</td>
<td>68 mph</td>
</tr>
<tr>
<td>Shockwave speed</td>
<td>$w$</td>
<td>-11.7 mph</td>
</tr>
</tbody>
</table>

The section of interest does not include the existing loop detectors. Therefore, the data is processed in order to emulate the loop detector data at both boundaries of the section of interest. Since all vehicle trajectories are known, this can easily be done without the usual measurement error associated with loop detectors. These emulated detectors will provide the boundary conditions.

### 5.1.1 Scenarios investigated

**Penetration rate and sampling strategies.** The implementation of the methods involves selecting a subset of the NGSIM data and treating it as Lagrangian data that is used for the assimilation. Twelve different scenarios were investigated to account for different penetration rates and sampling strategies. The penetration rate $P$ is the ratio of the number of probe vehicles over the total number of vehicles. In the present case, since trajectories of all vehicles are known, a proportion $P$ of trajectories is (randomly) chosen and considered
as probes. These equipped vehicles report their position and speed every time interval $T$. The reported speed is the average speed over the last $\tau$ seconds (Figure 5.2 sketches this sampling strategy).

![Figure 5.2: Schematics of the sampling strategy on equipped vehicle $n$.](image)

The values chosen for $P$, $T$, and $\tau$ determine the total number of Lagrangian measurements created for each case investigated. Five different values of $P$ and two values of $T$ were investigated ($\tau$ was assumed to be 6 seconds in all scenarios). Table 5.2 shows the scenarios investigated and the average number of Lagrangian measurements produced per mile-lane and per minute, and per 100 cells per time step for each one of them.

<table>
<thead>
<tr>
<th>Case</th>
<th>$P$ (%)</th>
<th>$T$ (sec)</th>
<th>$\tau$ (sec)</th>
<th>Average number of observations per mile-lane per minute</th>
<th>Average number of observations per 100 cells per time step</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>150</td>
<td>6</td>
<td>1</td>
<td>0.3</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>150</td>
<td>6</td>
<td>3</td>
<td>0.7</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>150</td>
<td>6</td>
<td>6</td>
<td>1.5</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>150</td>
<td>6</td>
<td>9</td>
<td>2.2</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>150</td>
<td>6</td>
<td>12</td>
<td>2.9</td>
</tr>
<tr>
<td>6</td>
<td>25</td>
<td>150</td>
<td>6</td>
<td>14</td>
<td>3.6</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>10</td>
<td>6</td>
<td>8</td>
<td>2.0</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>10</td>
<td>6</td>
<td>20</td>
<td>4.9</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>10</td>
<td>6</td>
<td>38</td>
<td>9.5</td>
</tr>
<tr>
<td>10</td>
<td>15</td>
<td>10</td>
<td>6</td>
<td>55</td>
<td>14</td>
</tr>
<tr>
<td>11</td>
<td>20</td>
<td>10</td>
<td>6</td>
<td>72</td>
<td>18</td>
</tr>
<tr>
<td>12</td>
<td>25</td>
<td>10</td>
<td>6</td>
<td>87</td>
<td>22</td>
</tr>
</tbody>
</table>

Figure 5.3 shows the location and time of all the Lagrangian observations for scenarios 2 and 7. Each dot corresponds to one observation.
The travel time for the section of interest is around two minutes under congested conditions. Thus, scenarios with $T = 150s$ assume that each equipped vehicle sends only one report while it is traveling the section. Almost continuous tracking for equipped vehicles is assumed for scenarios 7 to 12 ($T = 10s$ and $\tau = 6s$). Because of privacy issues, these scenarios might not be an acceptable solution for future technological developments, which is a problem addressed by Hoh et al in [23]. The scenarios were investigated, however, to evaluate the potential of the proposed methods.

5.1.2 Comparison of actual vs. observed velocity and density

Since in the NGSIM dataset all vehicle trajectories are known, ground truth velocity and density fields can be computed using Edie’s generalized definitions [16]. Ideally, both the observed velocity and observed density (computed by inversion of the fundamental diagram – as described in Section 4.2.1) should match the corresponding ground truth.

In practice, however, the way in which the observed quantities are computed using Lagrangian measurements introduces some error. Graphs at the top of Figure 5.4 show the ground truth velocity versus the observed velocity, computed according to equation (4.8), for scenarios 2, 6, and 10. Ground truth versus observed density, computed according to equation (4.9), for the same scenarios are at the bottom of the figure. Note that velocities rarely exceed 40 mph, implying that congested conditions are predominant in this data.
Figure 5.4: Ground truth vs. observed velocity (top) and density (bottom) for scenarios 2 (left), 6 (center), and 10 (right). Ground truth quantities have been computed using Edie’s generalized definitions.
Low velocities come with large error, which implies a large error for high densities as well. The main reason for this is the high variability among vehicle speeds in congestion. In addition to the variability in speed, the velocity-to-density conversion also introduces some error in the case of the density. In congestion, a given density can be achieved at different velocities (or the same velocity can yield different densities). Indeed, it is well known that the congested branch of the fundamental diagram is not a line of points but a cloud of points for non-stationary traffic [11], as visible in Figure 5.5. The figure shows the flow-density points computed using Edie’s generalized definitions at cell 8 (the triangular fundamental diagram adjusted for this section is also shown).

![Figure 5.5: Density-flow points on cell 8 of the section of interest, computed using Edie’s generalized definitions. The triangular fundamental diagram adjusted for this section is also shown.](image)

If more than one vehicle sends a report from the same cell at the same time step, the observed velocity is the arithmetic average of all velocities. This is the case for approximately only 10% of the observations for scenarios 10, 11, and 12 (for the other scenarios, this proportion is less than 5%). As expected, comparison among the scenarios in Figure 5.4 suggests that more data does not imply more accuracy in the observed quantity, which can be confirmed quantitatively. 20 different realizations\(^1\) have been created for each scenario, and for each one of them, the *Mean Absolute Error*\(^2\) (MAE) was computed. For

\[ \text{MAE} = \frac{1}{M} \sum_{m} |\hat{z}_m - z_m| \]

\(^1\)For the same penetration rate \(P\) and sampling strategy \((T, \tau)\), different realizations consider different vehicles as equipped vehicles and different times when measurements are sent. Therefore, for the same \((P, T, \tau)\), 20 realizations are 20 different sets of observations.

\(^2\)The MAE is defined as: \(\text{MAE} = \frac{1}{M} \sum_{m} |\hat{z}_m - z_m|\) where \(\hat{z}_m\) and \(z_m\) are the quantities to be compared, and \(M\) is the total number of observations (indexed by \(m\)).
each scenario, the 20 MAEs are averaged. Table 5.3 shows the results.

Table 5.3: Average MAE and its standard deviation between the ground truth and the observed velocity and density.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Velocity (mph) MAE</th>
<th>Velocity (mph) Std.Dev.</th>
<th>Density (vpm) MAE</th>
<th>Density (vpm) Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.32</td>
<td>0.27</td>
<td>71.5</td>
<td>6.17</td>
</tr>
<tr>
<td>2</td>
<td>3.42</td>
<td>0.18</td>
<td>72.3</td>
<td>4.97</td>
</tr>
<tr>
<td>3</td>
<td>3.43</td>
<td>0.12</td>
<td>73.0</td>
<td>3.12</td>
</tr>
<tr>
<td>4</td>
<td>3.42</td>
<td>0.14</td>
<td>73.2</td>
<td>1.66</td>
</tr>
<tr>
<td>5</td>
<td>3.42</td>
<td>0.08</td>
<td>72.5</td>
<td>1.27</td>
</tr>
<tr>
<td>6</td>
<td>3.43</td>
<td>0.08</td>
<td>72.7</td>
<td>2.10</td>
</tr>
<tr>
<td>7</td>
<td>3.43</td>
<td>0.15</td>
<td>77.0</td>
<td>2.73</td>
</tr>
<tr>
<td>8</td>
<td>3.47</td>
<td>0.08</td>
<td>76.2</td>
<td>1.93</td>
</tr>
<tr>
<td>9</td>
<td>3.42</td>
<td>0.06</td>
<td>75.6</td>
<td>1.13</td>
</tr>
<tr>
<td>10</td>
<td>3.38</td>
<td>0.05</td>
<td>75.0</td>
<td>0.97</td>
</tr>
<tr>
<td>11</td>
<td>3.37</td>
<td>0.04</td>
<td>74.0</td>
<td>0.74</td>
</tr>
<tr>
<td>12</td>
<td>3.34</td>
<td>0.03</td>
<td>73.7</td>
<td>0.64</td>
</tr>
</tbody>
</table>

The MAE for the velocity and for the density does not change significantly with the total number of observations. In fact, it does not exhibit any specific trend. The variability, however, tends to decrease as the number of observations increases. That is, for each scenario, the 20 MAEs computed for the 20 realizations vary more when the number of observations is low.

5.1.3 Results

The twelve scenarios presented in Table 5.2 were used and tested with the implementation of the filter-based heuristic method and the Newtonian relaxation method described in the previous chapter. For comparison purposes, we use a scenario which only uses information from boundary detectors, which we will refer to as the Eulerian data only (EDO) scenario. This scenario was implemented according to the numerical scheme presented earlier, and does not make use of the ramp counts for estimation. Therefore, this scenario is used to assess the improvement of the estimation as Lagrangian data becomes available.

The MAE of the accumulation of vehicles (i.e. number of vehicles per cell per time step) is used to assess the results. Note that the MAE computed for each scenario corresponds to the average obtained from 20 different realizations. The same 20 realizations were used in both methods.

The true accumulation of vehicles can easily be computed since all vehicle trajecto-
ries are known. The MAE for the EDO case is 2.02, which is used as the benchmark. Therefore, the percentage of improvement in the MAE of the vehicle accumulation when compared with the EDO case is used as a metric for the accuracy of the methods. The *percentage of improvement* for scenario $i$ is denoted by $\text{PoI}_i$, and is computed as $\text{PoI}_i = \frac{\text{MAE}_{\text{EDO}} - \text{MAE}_i}{\text{MAE}_{\text{EDO}}}$, where $\text{MAE}_{\text{EDO}} = 2.02$ and $\text{MAE}_i$ is the MAE of scenario $i$.

**Filter-based heuristic method**

The filter-based (FB) method depends on the parameter $L$, which as was explained earlier acts as a Kalman gain. Since the error in the *a-priori* and *observed* density are not known ($\xi$ and $\phi$ in equation 4.20 and 4.21, respectively), different values of $L$ are tried to determine how they affect the results. For simplicity, a constant value for $L$ over time and space is assumed. As mentioned in Section 4.2.3, the best value of $L$ depends on the accuracy of the observations and the total number of observations per cell per time step.

The values tried for $L$ are 0.2, 0.4, 0.6, 0.8, and 1.0. Note that $L = 0$ corresponds to the EDO case mentioned before. Figure 5.6a shows the average PoI for each scenario and for different values of $L$ (the standard deviation is very small in all cases). The scenarios are defined by the average number of observations per mile-lane per minute they provide, as shown in Table 5.2.

![Figure 5.6: PoI in the vehicle accumulation (in %) for different values of $L$ and number of observations.](image)

The scale of both graphs is different.

Figure 5.6: PoI in the vehicle accumulation (in %) for different values of $L$ and number of observations. a) $k^o$ is computed as described in Section 4.2.1, b) $k^o$ is assumed to include no error and therefore $k^o = k$. Note that the scale of both graphs is different.
For all the scenarios, all the values of $L$ improve the performance of the EDO case, which constitutes the first validation of the method. Any reasonable value of $L$ ($0 < L \leq 1$) provides better results than the EDO case, even though the observations include errors.

The value of $L$ that provides the best results for each of the 12 scenarios is different. This confirms that $L$ depends on factors such as the number of observations available and the accuracy of each observation, which change in time and space.

For cases with 38 or more observations per mile-lane per minute (scenarios 8 to 12), values of $L$ close to 0.4–0.6 are preferred. These values of $L$ are not only due to the fact that observations include errors, but are also due to the number of observations. The number of cells with an observation is sufficient to drive the PoI. Given the error in the observations, these cells are better off with low values of $L$. This explains the dependency of $L$ on the number of observations per cell per time step.

The opposite happens for scenarios with less than 20 observations per mile-lane per minute, where $L = 1$ is always preferred (the PoI for these cases, however, does not change too much for $L \geq 0.8$). In this case, only a few cells have an observation, and the PoI is driven by the other (several) cells without observations. These cells are affected by the observations through the model, which propagates the effect in time and space. If the observation from cell $i$ at time step $h$ does not significantly affect the state of this cell (i.e. when a low $L$ is used), the model cannot propagate this effect too far in time and space, and most of the cells without observations will not be affected at all. Therefore, even though there is some error in the observations, the cells without observations are better off with high values of $L$.

The previous analysis can be confirmed by using the actual density, instead of the observed density which was computed using the procedure outlined in Section 4.2.1. That is, it is assumed that the observed density is error free. The subscript $wo$ will denote that the observations come without error, and we will refer to this case as the without error case (no subscript is used for the original case, referred to as the with error case). Results are shown in Figure 5.6b. As expected, in all cases $L = 1$ is preferred. Table 5.4 summarizes these results, presenting the best value of $L$ depending on the total number of observations per cell per time step\textsuperscript{3} and on the accuracy of the observations.

For the rest of the analysis, the best value of $L$ in terms of the PoI is chosen for

\textsuperscript{3}5 observations per 100 cell per time step correspond to 20 observations per mile-lane per minute.
Table 5.4: Value of $L$ as a function of the number of observations per 100 cell per time step and the accuracy of each observation.

<table>
<thead>
<tr>
<th># observations/100 cell-time step</th>
<th>Low ($\leq 5$)</th>
<th>High ($\geq 5$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No error</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Error</td>
<td>$\geq 0.8$</td>
<td>0.4–0.6</td>
</tr>
</tbody>
</table>

each scenario. That is, $L = 1$ for scenarios 1 to 8, $L = 0.6$ for scenarios 9 and 10, and $L = 0.4$ for the last two scenarios. When the observed density is assumed to be the actual density (the without error case), $L = 1$ is used for all 12 scenarios.

Figure 5.7 shows the ground truth and estimated spatio-temporal evolution of the accumulation of vehicles for one realization of scenarios 2, 3, 5, and 10 for the with error case. Also shown is the field yielded in the EDO case. It can be seen visually that the estimates are able to capture the main shockwaves traversing the section of interest. The intensity of the shockwave, however, is not equally captured by different scenarios.

Figure 5.7 also suggests that a significant difference among the estimations occurs during the first 10-12 minutes (before 8:00am), when some waves emanate from intermediate locations. In fact, the fundamental difference between the EDO scenario and scenarios with Lagrangian measurements, in terms of the MAE, happens during this period of time. The added value of Lagrangian measurements is thus clear, as it enables the method to capture phenomena otherwise not detectable with only loop detectors.

Table 5.5 shows the average PoI for each scenario investigated. The seconds column presents the PoI for the with error cases, while the third column assumes no error in the observed density (without error case, PoI$_{\text{wo}}$). Superscript $FB$ denotes the method used to compute the PoI. The fourth column is the percentage of difference between both PoI’s. Figure 5.8 graphically shows the information in the second and third column of Table 5.5 as a function of the number of observations available for each scenario. The standard deviation for all scenarios is small ($\approx 0.01$ on average), ranging from 0.003 to 0.02.

Figure 5.8 suggests an increasing performance of the algorithm as the number of observations increases. The rate at which the performance increases, decreases as the number of observations increases. When the observations include some error, the PoI$_{FB}$ reaches 35% for a large number of observations. When error-free observations are assumed, the rate at which the performance increases, decreases at a lower rate than the previous

\[ \text{%diff} = \frac{\text{PoI}_{\text{wo}} - \text{PoI}_{FB}}{\text{PoI}_{FB}}. \]
The PoI_{FB}\textsuperscript{w0} of scenario 3 is close to the PoI_{FB} of scenario 5, even though scenario 3 has half the number of observations of scenario 5. A similar comparison exists between scenarios 5 and 10, but in this case, scenario 10 has almost five times the number of observations of scenario 5. This shows the trade-off between the number of observations and their accuracy, and it stresses the importance of having either several observations with
Table 5.5: PoI for each scenario when compared with the EDO case (using the Filter-based method).

<table>
<thead>
<tr>
<th>Scenario</th>
<th>$\text{PoI}^F_B$ (%)</th>
<th>$\text{PoI}^F_{\text{woo}}$ (%)</th>
<th>% difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.9</td>
<td>6.4</td>
<td>31.6</td>
</tr>
<tr>
<td>2</td>
<td>10.2</td>
<td>13.8</td>
<td>35.4</td>
</tr>
<tr>
<td>3</td>
<td>17.0</td>
<td>23.2</td>
<td>36.8</td>
</tr>
<tr>
<td>4</td>
<td>20.5</td>
<td>28.6</td>
<td>39.8</td>
</tr>
<tr>
<td>5</td>
<td>23.0</td>
<td>33.0</td>
<td>43.4</td>
</tr>
<tr>
<td>6</td>
<td>25.1</td>
<td>36.2</td>
<td>44.5</td>
</tr>
<tr>
<td>7</td>
<td>18.8</td>
<td>27.4</td>
<td>45.6</td>
</tr>
<tr>
<td>8</td>
<td>26.3</td>
<td>40.0</td>
<td>52.4</td>
</tr>
<tr>
<td>9</td>
<td>30.8</td>
<td>48.3</td>
<td>56.8</td>
</tr>
<tr>
<td>10</td>
<td>32.7</td>
<td>52.8</td>
<td>62.6</td>
</tr>
<tr>
<td>11</td>
<td>33.9</td>
<td>56.3</td>
<td>65.8</td>
</tr>
<tr>
<td>12</td>
<td>35.0</td>
<td>59.1</td>
<td>68.9</td>
</tr>
</tbody>
</table>

From Table 5.5, the percentage of difference for scenarios 7 to 12 (almost continuous tracking, i.e. $T = 10s$ and $\tau = 6s$) is greater than that computed for scenarios 1 to 6. In fact, even for cases with a similar number of observations (scenarios 4 and 7), scenario 7 with almost continuous tracking presents a larger improvement than scenario 4 with spo-
radar tracking – although scenario 4 still performs better than 7. This suggests that almost continuous tracking benefits more from the removal of the error in the observations. However, more scenarios that yield the same number of observations but uses different sampling strategies should be analyzed to be more conclusive in this matter.

Figure 5.9 shows the evolution of the true total accumulation of vehicles on the entire section over time and its corresponding estimate with EDO and one run of scenarios 2 and 10. This graph shows the agreement that can be achieved in determining the accumulation of vehicles in the section of interest. It also confirms that significant differences between the ground truth and different scenarios occur during the first 10-12 minutes.

![Figure 5.9: Total vehicle accumulation on the entire section using the Filter-based method.](image)

It is expected that for longer sections between detectors, with more intermediate ramps and probably more waves emanating from intermediate locations, the difference between EDO and scenarios with Lagrangian measurements will be larger. Therefore, the improvement that can be achieved by using Lagrangian data would be even more significant.

**Newtonian relaxation method**

For the Newtonian relaxation (NR) or nudging method, the parameters of the nudging factor are first chosen, and then a similar analysis as the one carried out before will be performed with the NR.
The nudging factor $\lambda(\delta x, \delta t)$ – equation (4.26) – depends on three parameters, which are typically set based on physical considerations. $X_M$ is the distance of the surrounding area that will be directly affected by the observation. $T_M$ is the time beyond which the effect of an observation is negligible. That is, density at location $x$ and time $t$ is directly affected by observations collected from locations between $(x - X_M)$ and $(x + X_M)$ during the last $T_M$ time units. The initial values chosen for $X_M$ and $T_M$ used in this study are 180 ft and 15 seconds, respectively. It is believed that for any larger values (in time and space), an observation should not directly affect the state. This is based on physical considerations and can be adjusted heuristically. The parameter $T_a$ can be seen as a gain that determines the strength of the nudging term. Therefore, it has to be tuned as a control parameter of the model.

Similar to parameter $L$ in the FB method, $T_a$ should also depend on the total number of observations available. In this case, however, the discretization scheme (i.e. $\Delta t$ and $\Delta x$) does not affect the value of $T_a$, which has units of time. Instead, the number of observations per mile-lane per minute and their accuracy affects $T_a$. For the present study, values of 10, 20 and 30 seconds for $T_a$ have provided good results for the scenarios with less than 20 observations per mile-lane per minute. For the last four scenarios, values of 70, 130, 150, and 220 were used for $T_a$. For the same reasons that the parameter $L$ changes among scenarios in the FB method, high values for $T_a$ (i.e. less weight to the nudging factor) are used in scenarios with a large number of observations.

Figure 5.10 shows the ground truth and the estimated spatio-temporal evolution of the accumulation of vehicles for scenarios 2, 3, 5, and 10.

As with the previous case, the method captures the shockwaves traveling the section, but with different intensities. Also, the discrepancy between ground truth and the estimates for scenarios with few observations is significant during the first 10-12 minutes.

Table 5.6 shows the average PoI for each scenario investigated for the with error (PoI$^{NR}$) and without error (PoI$^{NR}_{wo}$) cases. The percentage of difference between both PoI$^{NR}$ and PoI$^{NR}_{wo}$ is computed as before and shown in the fourth column. Figure 5.11 graphically presents the information in the second and third column of Table 5.6 as a function of the number of observations available for each scenario. The standard deviations are small, with an average of around 0.01% again.

For the with error case, scenarios 3 and 6 performs slightly better than scenario 7 and 8, respectively, even though they have fewer observations. This suggests that sporadic
Figure 5.10: Vehicle accumulation (vehicles per cell) using the Newtonian relaxation method. (a) Ground truth, (b) EDO, (c) Scenario 2, (d) Scenario 3, (e) Scenario 5, and (f) Scenario 10.

sampling from several vehicles is preferred over almost continuous tracking of a few vehicles, even though both scenarios produce the same total number of observations.

As expected, PoI increases when the observations are assumed to be perfect. The PoI\textsuperscript{NR} reaches the 25\%, while PoI\textsubscript{w/o} is close to 40\%. For the scenarios with 20 or more observations per mile-lane per minute, the difference between PoI\textsubscript{w/o} and PoI\textsuperscript{NR} is almost constant (∼11\%). Since the PoI\textsuperscript{NR} seems to stabilize around 25\%, this implies that this
Table 5.6: PoI for each scenario when compared with the EDO case (using the Newtonian relaxation method).

<table>
<thead>
<tr>
<th>Scenario</th>
<th>PoI(_{NR}) (%)</th>
<th>PoI(_{NR}^{\text{wo}}) (%)</th>
<th>% difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.2</td>
<td>9.0</td>
<td>45.9</td>
</tr>
<tr>
<td>2</td>
<td>11.0</td>
<td>17.3</td>
<td>57.4</td>
</tr>
<tr>
<td>3</td>
<td>15.2</td>
<td>21.3</td>
<td>40.5</td>
</tr>
<tr>
<td>4</td>
<td>17.8</td>
<td>25.5</td>
<td>43.0</td>
</tr>
<tr>
<td>5</td>
<td>18.8</td>
<td>26.6</td>
<td>41.8</td>
</tr>
<tr>
<td>6</td>
<td>20.3</td>
<td>29.2</td>
<td>44.1</td>
</tr>
<tr>
<td>7</td>
<td>15.0</td>
<td>23.8</td>
<td>58.8</td>
</tr>
<tr>
<td>8</td>
<td>19.7</td>
<td>30.8</td>
<td>56.5</td>
</tr>
<tr>
<td>9</td>
<td>23.4</td>
<td>34.3</td>
<td>46.5</td>
</tr>
<tr>
<td>10</td>
<td>24.7</td>
<td>36.1</td>
<td>42.6</td>
</tr>
<tr>
<td>11</td>
<td>25.4</td>
<td>37.1</td>
<td>46.0</td>
</tr>
<tr>
<td>12</td>
<td>26.7</td>
<td>38.3</td>
<td>43.7</td>
</tr>
</tbody>
</table>

Figure 5.11: PoI (in %) as the number of Lagrangian measurements changes (Sce: scenario) using the Newtonian relaxation method. The blue line computes observed density by inversion of the fundamental diagram (i.e. with error), while the red line uses the actual density computed from vehicle trajectories as the observed one (i.e. without error).

Figure 5.12 shows the evolution of the actual total accumulation of vehicles on the entire section over time and its corresponding estimate with EDO and one run of scenarios 2 and 10. The figure confirms the reasonable accuracy achieved by the method, and that most of the error among ground truth, EDO case, and the different scenarios is again observed.

method (with the parameters chosen) will likely not achieve a PoI greater than 40%, even for large numbers of perfect observations.

Figure 5.12 shows the evolution of the actual total accumulation of vehicles on the entire section over time and its corresponding estimate with EDO and one run of scenarios 2 and 10. The figure confirms the reasonable accuracy achieved by the method, and that most of the error among ground truth, EDO case, and the different scenarios is again observed.
during the first 10–12 minutes of the period of analysis.

Figure 5.12: Total vehicle accumulation on the entire section using the Newtonian relaxation method.

Comments

Since the same 20 realizations were used for both methods, we can compare how both methods perform when using exactly the same data.

Figure 5.13 shows the PoI for all 20 realizations using both methods, for the with and without error cases, and for scenarios 1, 3, 8, and 11. For the first scenario, both methods provide similar results, even though the NR for the without error case performs slightly better than the others. Starting from scenario 3, FB systematically performs better than NR, and the gap between both methods increases with the number of observations. In fact, when the number of observations is high, PoI$^{\text{FB}}$ reaches a similar level as PoI$^{\text{NR}}$.

Figure 5.14a shows the difference of the average PoI between the with and without error cases for each method and for all the scenarios. That is, (PoI$^{\text{FB}}_{\text{wo}}$ − PoI$^{\text{FB}}$) in the blue line and (PoI$^{\text{NR}}_{\text{wo}}$ − PoI$^{\text{NR}}$) in the red line. A positive difference means that observations without error are preferred over those with error, which is always the case. When the blue line is over the red line, FB benefited more by the removal of error from the observations. For scenarios with more than 12 observations per mile-lane per minute, this difference increases
Figure 5.13: PoI (in %) for each of the 20 realizations for scenarios a) 1, b) 3, c) 8, and d) 11.

with the number of observations.

Figure 5.14: Total vehicle accumulation on the entire section using the Newtonian relaxation method.
Figure 5.14b shows the difference of the average PoI between the FB and NR for the with and without error cases and for all the scenarios. That is, \((\text{PoI}^{\text{FB}} - \text{PoI}^{\text{NR}})\) in the blue line and \((\text{PoI}^{\text{FB}_{\text{wo}}} - \text{PoI}^{\text{NR}_{\text{wo}}})\) in the red line. A positive difference means that the FB outperforms the NR, which is the case for the first two scenarios. Note that although the first two scenarios are negative, the difference is small and close to zero, suggesting that both methods perform similarly for that level of observations per mile-lane per minute. When the red line is over the blue line, the performance of the FB compared to the performance of the NR increases when observations come without error (error-free observations have a greater impact in the results). Again, for scenarios with more than 12 observations per mile-lane per minute, this difference increases with the number of observations.

Finally, the special case of the NR discussed in Section 4.3, which is similar to the FB, is investigated. This case assumes small value for \(X_M\) and \(T_M\), which implies \(\lambda = \frac{1}{T_a}\). From equation (4.29), \(L \approx \Delta t \cdot \frac{1}{T_a}\). Since \(L\) is known for each scenario, \(T_a\) can be obtained. When observations include error, for scenarios 1 to 8, \(T_a = 1.2s\); for scenarios 9 and 10, \(T_a = 2s\); for the last two scenarios, \(T_a = 3s\). When observations come without error, \(T_a = 1.2s\) for all cases.

Figure 5.15 shows the average PoI obtained for each scenario for the special case of the NR. Both cases with and without error are depicted in the graph. For comparison purposes, the PoIs obtained before for the FB and NR (with and without error) are also shown.

For scenarios 1 and 2 (low number of observations), the special case of the NR behaves similarly to the FB method. For the scenarios with more observations, the special case of the NR outperforms the original case of the NR. However, it does not reach the FB levels. The same reasons discussed in the selection of parameter \(L\) in Section 5.1.3 explain the behavior observed.

Finally, the results shown here lead us to make the following concluding statements:

- Regardless of the method used, and for a reasonable accuracy in the measurements, the use of Lagrangian data, in addition to the Eulerian data provided by static sensors, improves the accuracy of the estimates of the traffic state.

- The NR performs slightly better than the FB when the number of observations per mile-lane per minute is less that 3. For larger numbers of observations, the FB always outperforms the NR.
Figure 5.15: PoI (in %) as the number of Lagrangian measurements changes (Sce: scenario) for the special case of the Newtonian relaxation method. The blue line computes observed density by inversion of the fundamental diagram (i.e. with error), while the red line uses the actual density computed from vehicle trajectories as the observed one (i.e. without error).

- Both methods perform better when the accuracy of the observations is increased. The FB, however, benefits more than the NR when error is removed from the observations, and the improvement increases with the number of observations.

- For small $X_M$ and $T_M$ in the NR, both the FB and the NR are similar in the treatment of each observation. The performance achieved by the FB, however, is higher than this special case of the NR, and the difference increases with the number of observations when accurate observations are available.

- The original case of the NR considers $X_M = 180$ ft and $T_M = 15$ s. Greater values for these parameters lack a tangible physical interpretation. Given the performance of the special NR case, it is expected that lower values than the values chosen in this case (which may also make physical sense) will not achieve the performance of the special case.

The section of interest is congested almost all the time, and therefore the impact of the simplifications made in Section 4.2.1 to obtain the observed density in free flow conditions cannot be assessed.

In both methods, the main difference between the estimates from the EDO sce-
nario and from scenarios using Lagrangian measurements was observed during the periods in which shockwaves emanate from intermediate points. This result provides us with insight regarding the benefits of using Lagrangian measurements. By collecting data from individual vehicles at different times and locations, we were able to capture the shockwave generated in the middle of the section (between detectors). Therefore, in the presence of Lagrangian data, inter-detector spacing could be increased. Unfortunately, the NGSIM section is too short to test this statement. The next section provides results of the implementation of both the FB and NR on a longer section of highway, which also encompasses free flow and congestion periods (and the transition between both). The data comes from the *Mobile Century* experiment described in detail in Chapter 3.

### 5.2 Implementation with *Mobile Century* data

The data used in this section was described earlier in Chapter 3. For implementation purposes, lanes 1 to 4 (lane 1 being the leftmost lane) on the section between Decoto Rd. and Winton Ave. in the NB direction were considered. The period of time analyzed starts at 10am and ends at 6pm. The 6.5 miles were divided into \( I = 44 \) cells of \( \Delta x = 780 \) ft each, and the 8 hour period of meaningful data was divided into \( H = 3,600 \) time steps of \( \Delta t = 8 \) seconds each. As in the NGSIM case, the parameters of the fundamental diagram were extracted from PeMS and are shown in Table 5.7 (a triangular fundamental diagram is used\(^5\)).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Notation</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum flow</td>
<td>( q_M )</td>
<td>2275 vphpl</td>
</tr>
<tr>
<td>Jam density</td>
<td>( k_J )</td>
<td>152 vpmpl</td>
</tr>
<tr>
<td>Critical density</td>
<td>( k_C )</td>
<td>35 vpmpl</td>
</tr>
<tr>
<td>Free flow speed</td>
<td>( v_f )</td>
<td>65 mph</td>
</tr>
<tr>
<td>Shockwave speed</td>
<td>( w )</td>
<td>-19.4 mph</td>
</tr>
</tbody>
</table>

\(^5\)Since only few observations were available to obtain the jam density, the analysis shown in the rest of this chapter was also performed assuming the same jam density as in the NGSIM case (i.e. \( k_J = 205 \) vpmpl), which produces a shockwave speed of \( w = -13.4 \) mph. The results, however, do not change significantly from the results presented in this section.
5.2.1 Scenarios investigated

As before the scenarios are determined by the values used for $P$, $T$, and $\tau$. The penetration rate $P$ is the ratio of the number of probe vehicles over the total number of vehicles. The probe vehicles report their position and speed every time interval $T$, and the reported speed is the average speed over the last $\tau$ logs.

The penetration rate $P$ is given by the number of cell phones on the road. During the day, this rate mainly moves between 1% and 4% of the total flow, as shown in Chapter 3. Only the values of $T$ and $\tau$ can be changed to construct different scenarios. Table 5.8 shows the scenarios investigated in this case and the corresponding average number of observations. Note that in this case, $\tau$ corresponds to the number of logs over which the average speed is computed (cell phones were storing the position and velocity logs every 3 seconds).

Table 5.8: Scenarios investigated using data from the field experiment.

<table>
<thead>
<tr>
<th>Case</th>
<th>$T$ (sec)</th>
<th>$\tau$ (# of logs)</th>
<th>Average number of observations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>per mile-lane per minute</td>
</tr>
<tr>
<td>1</td>
<td>300</td>
<td>2</td>
<td>0.2</td>
</tr>
<tr>
<td>2</td>
<td>180</td>
<td>2</td>
<td>0.3</td>
</tr>
<tr>
<td>3</td>
<td>60</td>
<td>2</td>
<td>0.9</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>2</td>
<td>1.6</td>
</tr>
</tbody>
</table>

5.2.2 Results

The scenarios presented in Table 5.8 were implemented using both the FB and NR methods. Unlike for the NGSIM dataset, in this case, we only have access to the trajectories of the equipped vehicles, which is a subset of all the vehicles. Therefore, the ground truth is not known for all vehicle trajectories and for vehicle accumulation, but it is known for travel times between Decoto Rd. and Winton Ave. between 10am and 5pm (which was obtained from high definition video cameras).

In addition to the ground truth travel time, information collected by 17 loop detector stations installed along the section of interest is also available from the PeMS system [2]. As shown in Chapter 3, detector measurements contain errors. These measurements cannot be used as ground truth. Thus, a quantitative comparison (using the MAE, for instance) between these measurements and our estimates is not performed. Qualitatively, however, loop detectors provide an accurate picture of the traffic situation during the day. Therefore, they can be used to determine whether the methods can qualitatively replicate traffic con-
ditions. That is, if the methods can accurately identify bottleneck locations, and reproduce the temporal and spatial extent of the queues.

Loop detectors provide occupancy, which can be converted into density using g-factors. As with the velocity, PeMS computes the density field by associating an influence area with each detector station. The assumption is that measurements for this area are provided by the corresponding detector station. Therefore, a density is computed for each one of the 44 cells at each of the 3,600 time steps, and from that, the density field is obtained. Since the length of the cells is known, the computation of the accumulation of vehicles per cell is trivial. The resulting density field is shown in Figure 5.16a.

Figure 5.16: a) Density field (in vpmpl) collected by PeMS. b) Travel times (in minutes) computed using the velocity measurements collected by PeMS and the method described in the Section 3.2.

The density field computed using loop detectors captures the morning accident at postmile 26 (between Tennyson Rd and CA92). It also shows the spatial and temporal extent of the queue that was formed upstream. The situation at this location does not fully recover, and the bottleneck remains active until the evening. A short wave propagates upstream for about 2 miles at 12:30pm. Congestion starts at 2pm, and by 3pm, the entire section between postmile 21 (Decoto Rd) and 26 exhibits a serious level of congestion. Note that the severity of the congestion around postmile 24 (between Whipple Rd. and Industrial Blvd.) is higher than for some other locations, suggesting the presence of a second active bottleneck in series at this postmile.
The estimated density field can be used to compute travel times, which can be compared to the ground truth travel times. The density field is converted into the velocity field using the fundamental diagram, and this field can be integrated to obtain trajectories, and therefore travel times, as done earlier in Chapter 3. In the case of the PeMS density field, its corresponding travel times and ground truth travel times are shown in Figure 5.16b, which are the same as the ones shown in Figure 3.6.

In the computation of the observed density for both methods, $k_{FF} = 25 \text{ vpmpl}$ was used in equation (4.9). A different value for $k_{FF}$ will affect the density estimates, specially during free flow conditions, but it will not affect travel time estimates. The effect of using a different value for $k_{FF}$ is shown later on this chapter.

**Filter-based heuristic method**

The quantitative performance of the method for different values of $L$ cannot be evaluated, since the ground truth of the vehicle accumulation is not known. However, travel times can be obtained from the density estimates (as explained before). In the present case, the value of $L$ was chosen heuristically. The values chosen for each scenario are $L = 1$ for the first two scenarios, $L = 0.8$ for scenario 3, and $L = 0.6$ for the last scenario.

Note that if the number of observations per 100 cells per time step is used to select parameter $L$, values of $L$ for the NGSIM case and for this case are in good agreement. For instance, the average number of observations per 100 cells per time step for scenario 2 in this dataset is 2.3, which is close to the corresponding number for scenario 4 in the NGSIM dataset (2.2). Therefore, both scenarios could used the same (or a close) value for parameter $L$.

Figure 5.17 shows the estimated density field for each scenario in Table 5.8 using the value of $L$ that provides the best agreement between the travel times. The main congested pattern observed from detectors is well captured for all the scenarios investigated. All the scenarios captured the accident at postmile 26 at 10:30am, even though not all the equipped vehicles were on the highway by then (vehicles were released between 10-10:30am). The location and duration of the incident is accurately replicated in each scenario. The short wave at 12:30pm reported in PeMS field is observed on all the scenarios investigated, but with different durations and intensities. The main congestion and its propagation starts at 2pm and can be seen in all the scenarios as well. The more severe congestion around
postmile 24 that is observed from detectors is also observed from Lagrangian data.

Figure 5.17: Density field (in vpmpl) for the scenarios investigated (using the Filter-based method). a) scenario 1, b) scenario 2, c) scenario 3, d) scenario 4. For each scenario, the boundary data is provided by loop detector measurements.

The travel times yielded by all the scenarios are shown in Figure 5.18. Dots correspond to the ground truth travel time, where each dot corresponds to one vehicle.

The number of observations for scenarios 1 and 2 is not enough to identify the congestion that occurred during the day entirely. This data did not capture all the congestion at postmile 26 during the morning. For that reason, their estimates of the travel time tend to significantly underestimate the ground truth travel time. The other two scenarios capture the congestion pattern observed during the day better, and therefore produce more accurate estimates of the travel time (which is similar for both scenarios). These estimates, however, still underestimate the main trend observed in travel time, specially between 12:30am and
Figure 5.18: Travel times (in minutes) between Decoto Rd. (PM 21) and Winton Ave. (PM 27.5) for the scenarios investigated (using the Filter-based method). For each scenario, the boundary data is provided by loop detector measurements.

2:30pm.

Note that the estimates shown in Figure 5.17 make use of only two detector stations located at both boundaries of the section. That is, they are 6.5 miles apart from each other, with 11 ramps in between them. However, in the absence of detector data, the results do not change too much. For instance, we can assume that the Eulerian boundary data is also provided by equipped vehicles. To this end, the concept of VTL introduced in Chapter 3 is useful. In our case, every 5 minutes, a VTL at each boundary of the section of interest computes the average speed from GPS measurements from different vehicles. Obviously, the more equipped vehicles there are per interval, the more representative the estimate of the speed should be. However, as shown in Chapter 3, even a low proportion of vehicles yields reasonable velocity measurements. By inversion of the fundamental diagram, the density at the boundaries can be computed, which is the input for the model. Figures 5.19 and 5.20 show the density field and the travel time estimates, respectively, for the same scenarios described in Table 5.8, where boundary data is computed from GPS data as described before.

The similarity between the results using loop detector and VTL measurements
as boundary conditions suggests that the results are mainly driven by the observations collected from the GPS, and perfect knowledge of the boundary conditions is not critical. The average difference of vehicles per cell per time step between both estimates is less than 2 vehicles, and it decreases as the number of observations increases. Therefore, boundary conditions become less and less important as the number of observations increases.

**Influence of \( k^{FF} \).** Without any particular consideration, the value chosen for \( k^{FF} \) was 25 vpmpl. Three different values, in addition to \( k^{FF} = 25 \) vpmpl, were investigated to determine the influence of this choice on the estimates. The values chosen were 5, 15, and 35 vpmpl. Figure 5.21 shows the absolute difference in density between the original results obtained with \( k^{FF} = 25 \) vpmpl and the three values investigated for scenario 3.
In the three cases, significant difference with respect to the original case occurs during the transition between free flow and congested periods. Difference during the free flow periods can only be observed for the comparison with the case with $k_{FF} = 5$ vpmpl. The same trend is obtained for the other three scenarios investigated.

The results indicate that different values for $k_{FF}$ change the density estimates. Most of the difference between different values for $k_{FF}$ occurs only during transition periods. This is specially true for values of $k_{FF}$ greater than 40% of the critical density.

Figure 5.22 shows the difference in mode between the case with $k_{FF} = 25$ vpmpl and the three values investigated for scenario 3. White areas in figure mean that the case with $k_{FF} = 25$ estimates congestion at that location and time, while the corresponding case estimates free flow conditions. Black areas are the opposite: the case with $k_{FF} = 25$ estimates free flow conditions, while the corresponding case estimates congestion. Gray areas imply that both estimates are in agreement in terms of the mode. The cases with higher values for $k_{FF}$ tend to show more congestion.
Figure 5.21: Absolute difference in the density field (in vpmpl) between the case when \( k_{FF} = 25 \) vpmpl and a) \( k_{FF} = 5 \) vpmpl, b) \( k_{FF} = 15 \) vpmpl, and c) \( k_{FF} = 35 \) vpmpl. The results were obtained for scenario 3 and using the Filter-based method. The boundary data is provided by loop detector measurements.
Figure 5.22: Difference in the mode between the case with $k^{FF} = 25$ vpmpl and a) $k^{FF} = 5$ vpmpl, b) $k^{FF} = 15$ vpmpl, and c) $k^{FF} = 35$ vpmpl. The results were obtained for scenario 3 and using the Filter-based method. The boundary data is provided by loop detector measurements.
Newtonian relaxation method

The parameter choice for $X_M$ and $T_M$ is the same as for the NGSIM case (180 ft and 15s, respectively). As before, the criteria in the selection of parameter $T_a$ for each scenario is driven by the corresponding estimate of the travel time. The values chosen are between $T_a = 10s$ for the first three scenarios, and $T_a = 20s$ for the last scenario.

If the number of observations per mile-lane per minute is used to determine the value of $T_a$, values chosen in the NGSIM case and in this case are again in reasonable agreement.

Figure 5.23 and 5.24 show the density field and the travel time estimated, respectively, for each scenario investigated using the NR.

Figure 5.23: Density field (in vpmpl) for the scenarios investigated (using the Newtonian relaxation method). a) scenario 1, b) scenario 2, c) scenario 3, d) scenario 4. For each scenario, the boundary data is provided by loop detector measurements.
As with the FB, the main congested pattern observed from detectors is well captured in most of the scenarios investigated. Scenarios 1 and 2 do not capture the congestion entirely, and therefore they underestimate travel times. Note that the free flow density estimates differ from those computed with the FB (as suggested by the difference in color between postmiles 21 and 23 and before 3pm). This difference, however, is not observed in the travel time estimates.

Figure 5.25 and 5.26 represent the density field and travel time estimates, respectively, when boundary conditions are provided using the concept of VTL. The differences between this case and the case where loop detectors provide boundary conditions are again small. This method is also expected to work well even when Eulerian data is not available.

Finally, the same analysis performed to determine the influence of the choice of $k^{FF}$ with the FB method was performed with the NR method. The results are the same in this case, suggesting that the choice of the value for $k^{FF}$ affect mainly the estimates during the transition periods.
Kalman filtering

Traffic conditions or the mode of traffic (as defined in Section 4.2.2) on the section of interest changes throughout the experiment. The four scenarios shown in Table 5.8 have been implemented in the Kalman filtering method described in Section 4.2.2. The purpose of this implementation is to assess the mode identification process proposed in that section.

Figure 5.27 shows the travel time estimates for each of the four scenarios using each of the three methods: Kalman filtering, filter-based heuristic, and Newtonian relaxation.

The figure confirms the similarity in the FB and NR estimates. For each scenario, these two estimates are almost identical.
Kalman filtering estimates are more accurate than FB and NR estimates for the first two scenarios. This suggests that the mode identification proposed works well in practice, and it is desirable when the number of observations is low. For the other two scenarios, Kalman filtering yields similar estimates to the FB and NR estimates, and thus mode identification is not critical.

Comments

The following can be concluded after implementing both the FB and the NR with the field experiment data:

- For each scenario, both the FB and the NR produce similar density field and travel time estimates. The small number of observations per mile-lane per minute as compared to the NGSIM scenarios might explain this. If more observations are available, both estimates would be expected to differ from each other. Since ground truth in

---

6 The scenario with the most observations per mile-lane per minute in this case (scenario 4) is comparable with the scenario with the least number of observations per mile-lane per minute in the NGSIM case (scenario 1).
Figure 5.27: Travel times (in minutes) between Decoto Rd. (PM 21) and Winton Ave. (PM 27.5) for scenario a) 1, b) 2, c) 3, and d) 4. Each line corresponds to one method: green, Kalman filtering; blue, filter-based heuristic; red, Newtonian relaxation. For each case, the boundary data is provided by loop detector measurements.

terms of vehicle accumulation is not known, a comparison to determine the performance of the methods cannot be done.

• Loop detector stations providing boundary conditions are not required by the meth-
ods. When VTL data is used to provide boundary conditions, the results do not change significantly, specially when the number of observations is high. Therefore, traffic state estimation using either the FB or the NR can be performed on a large section of highway using only Lagrangian data, provided the number of observations per mile-lane per minute is enough to drive the results.

- The mode identification proposed in Section 4.2.2 works well in practice. For an average of one observation per mile-lane per minute, however, the mode identification does not improve significantly the estimates compared to the FB and NR. Therefore, in these situations, FB or NR are preferred because of their simplicity.

- For both methods, different values for $k^{FF}$ were tried in order to determine the sensitivity of the results to this choice. Results suggest that the choice of $k^{FF}$ is not critical since the density estimate does not change significantly.

- Based on the results obtained with both methods, an average of at least one observation per mile-lane per minute is needed to identify most of the congestion and to obtain reasonably accurate estimates of the traffic state. This average is computed using observations collected during free flow and congested conditions. Table 5.9 shows the average number of observations per mile-lane per minute for each condition and for each scenario.

Table 5.9: Number of observations per mile-lane per minute during free flow and congested regimes.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Free flow</th>
<th>Congestion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1</td>
<td>0.3</td>
</tr>
<tr>
<td>2</td>
<td>0.2</td>
<td>0.5</td>
</tr>
<tr>
<td>3</td>
<td>0.5</td>
<td>1.4</td>
</tr>
<tr>
<td>4</td>
<td>1.0</td>
<td>2.6</td>
</tr>
</tbody>
</table>

A vehicle traveling at velocity $V$ mph and sending an observation every $T$ hours, sends $\frac{1}{VT}$ observations in one mile. If a proportion $P$ of the total flow $Q$ per lane (in vphpl) is equipped, the total number of observations per mile-lane per minute, $N_{obs}$ is given by:

$$N_{obs} = \frac{P \cdot Q}{60 \cdot V \cdot T} = \frac{P \cdot K}{60 \cdot T}$$

(5.1)
$K$ is the density per lane in vpmpl. Equation (5.1) provides a relationship between the existing traffic conditions (given by $Q$, $V$, and $K$) and the sampling strategy to be used ($P$ and $T$) in order to achieve a specific number of observations per mile-lane per minute. For instance, knowing the number of observations per mile-lane per minute desirable for a given traffic condition, the period between observations $T$ can be selected for different penetrations $P$. In fact, for a given penetration $P$ and during congestion (i.e. high $K$), a larger period $T$ is allowed.
Chapter 6

Conclusions

Traffic monitoring system based on GPS-enabled cell phones. Traffic monitoring systems are essential to identify congestion and to implement control strategies to alleviate it. Current sensor infrastructure, which mainly consists of loop detectors, is expensive, requires maintenance and repair, and is prone to error and malfunctions. GPS-enabled cell phones can determine their position and velocity accurately. If this information is transmitted to a traffic center, it can provide a cost-effective augmentation of the available traffic data, in particular:

- It requires no infrastructure, maintenance, and/or repair costs for transportation agencies since it is a market driven technology and is maintained by the private sector.

- It achieves higher accuracy than loop detectors for speed measurements.

Since the sensors are moving over the transportation system, a sufficient penetration of mobile phones would provide an extensive spatio-temporal coverage of the network. Therefore, monitoring can be done on freeways and urban streets as well.

*Mobile Century*, a field experiment conceived as a proof of concept of such a monitoring system has been carried out. The experiment was performed on a section of freeway I-880 in the San Francisco Bay Area. In terms of the data collection process and the quality of the data gathered, the experiment showed the feasibility of this type of system with strong implementation potential in the near future. The experimental design enabled us to know the proportion of equipped vehicles in the total flow, or penetration, and to evaluate the accuracy of the data collected. Evidence has been shown that even under low
penetrations data obtained from GPS-enabled cell phones is as accurate as (and sometimes more accurate than) the data collected by loop detectors for speed readings.

A field operational test which extends this type of system to the urban network is underway. The field operational test in its initial phase of development, called *Mobile Millennium*, consists of the free distribution of traffic software to regular commuters, downloadable directly on their phones. This software allows the collection of traffic data during a number of months and will principally cover Northern California in its initial phase [3]. Using data features of cell phones, commuters will be delivered with traffic conditions on their phone.

Given the features of a traffic monitoring system based on GPS-enabled cell phones, this type of system is particularly appropriate for developing countries, which lack resources and monitoring infrastructure, and where the penetration of mobile phones in the population is significant (and rapidly increasing).

**Traffic state estimation.** Traffic management centers need extensive coverage of the transportation network. Eventually, if the amount of data received is large enough, modeling assumptions used to estimate traffic can be relaxed and replaced by data. However, only sporadic data (in time and space) can be provided by monitoring systems, regardless of the sensors used. Traffic state estimation is performed to fill the areas with no data. An extensive literature can be found on how to fill the areas when data is provided by loop detectors. Very little effort, however, has been devoted to analyzing this problem when data is provided by GPS-enabled cell phones (Lagrangian data).

Three methods to incorporate Lagrangian measurements into the traffic state estimation have been proposed: the Kalman filtering based (KF) method, the Filter-based heuristic (FB) method, and the Newtonian relaxation (NR) method. Conceptually, the methods are similar: they add or remove vehicles depending on the difference between the estimates and the observations.

The KF method requires the traffic conditions to be identified at each time step. Ideally, the mode should be directly observed from the observations collected. Since observations are sparse in time and space, however, the mode identification is a challenging task. In addition, the approach assumes the knowledge of the process and measurement error covariances. The FB heuristic method has been proposed to circumvent these issues.

The three methods assume that boundary conditions are provided by loop de-
ectors, and the fundamental diagram is known. In addition to the parameters of the fundamental diagram, the FB method uses one parameter, while the NR method uses three additional parameters. Two of the three parameters of the NR ($X_M$ and $T_M$) can be chosen based on physical considerations, as they can be seen as the influence area (in time and space) of an observation. The values chosen were $X_M = 180$ ft and $T_M = 15$ s, and they have provided good results. The NR would probably perform better with higher values, especially when the total number of observations is low, but they would not make physical sense.

The other parameters, $L$ in the FB method and $T_a$ in the NR method, depend on the number and accuracy of the observations. In fact, when observations include some error, the number of observations per cell per time step is a good indicator to determine $L$, while the number of observations per mile-lane per minute is a good indicator to determine $T_a$. The reason for this difference is because $L$ depends on the discretization scheme, and $T_a$ does not.

The FB and NR methods were implemented using two datasets. In both cases, a temporal sampling strategy has been used, in which vehicles report their position and velocity periodically in time. This does not imply all the vehicles send a report at the same time. It only means they send a report at the same frequency. Different scenarios were constructed for each dataset, which produce a different number of observations per mile-lane per minute.

The first dataset is the NGSIM data. This data includes the trajectory of all vehicles traveling a 0.4 mile section of freeway US-101 during a 45 minute period. The ground truth consists in trajectories of all vehicles and is known. Therefore, this dataset allows an assessment of the accuracy of the methods. The assessment was made in terms of the improvement in the estimates as compared to a base case in which only loop detector data is used. 12 scenarios were investigated, ranging from 1 to 90 observations per mile-lane per minute. The main conclusions drawn from this implementation are the following:

- A conversion is needed to make use of probe data and the methods proposed. A GPS reports its velocity, which is converted to density (using the fundamental diagram). This conversion introduces some additional error in the observed density (estimate of density from the speed). This is important because of the trade-off between the number of observations and their accuracy to achieve a given percentage of improvement.
as compared to the base case.

- Despite the error in the observations, both methods improve the traffic state estimates when compared to the base case. This is true even for a low number of observations per mile-lane per minute. This statement validates the value of both methods for traffic state estimation in the presence of Lagrangian data.

- Accuracy in the estimates increases with the number of observations. The improvement in the estimates reaches a saturation level for a high number of Lagrangian measurements. Marginal benefits are obtained from new observations when the number of measurements becomes too large. Also, the increasing rate is different for both methods, and depends on the accuracy of the data as well. When error-free observations are available, an improvement of up to 60% in the FB method and up to 40% in the NR method can be reached. When error is introduced in the observations, these numbers go down to 35% and 25% for the FB and NR methods, respectively.

- For low numbers of observations per mile-lane per minute (less than three), both methods perform similarly, with the NR performing slightly better than the FB. Otherwise, the FB method significantly outperforms the NR method.

The second dataset used was obtained from the Mobile Century experiment. This dataset considers a longer section of freeway (6.5 miles) and for a longer period of time (8 hours). In this case, only a subset of vehicle trajectories is known. Therefore, the accuracy of the methods cannot be assessed. Ground truth travel times obtained from video cameras and the data collected by 17 loop detector stations deployed on the section of interest are used for a qualitative evaluation of the methods. Four scenarios were analyzed in this case, all of them with less than 2 observations per mile-lane per minute. The following is concluded from this analysis:

- Both the FB method and the NR method yield similar density field estimates. This could be explained by the number of observations per mile-lane per minute for all the scenarios investigated. For more frequent sampling or higher penetration, discrepancies between both estimates would probably be significant.

- The mode identification performed in the Kalman filtering method does not improve the travel time estimates of the FB method and NR method if there is one or more
observations per mile-lane per minute on average. In these cases, FB is preferred for its simplicity. For fewer observations, Kalman filtering yields more accurate the estimates of travel times compared to the other two methods. All the travel time estimates, however, underestimate ground truth travel time specially during the transition from free flow to congested traffic.

- Accurate or low error boundary conditions are not critical for the methods. The concept of virtual trip lines (VTL) introduced in Chapter 3 is used to provide boundary conditions, and the results do not change significantly. Therefore, in the presence of Lagrangian data, inter-detector spacing can be increased, requiring fewer detectors to cover the same length of freeway. In fact, if VTLs are used, both methods are useful even in places where no monitoring infrastructure (like loop detectors) has been previously deployed, such as in developing countries.

- When congestion arises, one observation per mile-lane per minute would allow for its identification. With fewer observations, not all the congestion is identified, and travel times are underestimated. Therefore, for a given penetration of equipped vehicles in the total flow, a larger period between consecutive observations from one vehicle is allowed during congestion – in order to achieve a certain number of observations per mile-lane per minute. During the free flow regime, the period should be shorter. However, given the approximation made in the computation of the observed density, detection of the free flow condition is more important than the value of the actual density.

The methods proposed here are more useful when sufficient data is provided. Given that GPS is becoming a standard feature of cell phones and the given penetration of cell phones in the population, it is likely that sufficient data will be available in the near future, making these methods attractive for traffic estimation purposes.

**Future work.** The two datasets used for validation of the method proposed have different features. The NGSIM dataset provides data on a short section of freeway (0.4 miles) during a short period of time (45 minutes), but the ground truth in terms of vehicle trajectories is known. That is, the vehicle accumulation on the section can be easily computed. The Mobile Century experiment provides data on a longer section of freeway (6.5 miles) during a longer period of time (8 hours). Trajectories, however, are only known for 2-3% of the total
traffic flow. For this reason, the methods cannot be assessed in terms of their capability to estimate vehicle accumulation. A new dataset that combines the advantages of both the NGSIM and Mobile Century datasets is needed in order to better assess the capabilities of the method. This dataset could be provided by microsimulation.

In the present work, GPS data (and sometimes loop detector data as well) is used to perform traffic state estimation. In some places, however, more sources of traffic data are available, such as video cameras, radar, RFID transponders, etc. In these cases, a system that fuses all the sources of data is expected to provide a more accurate estimation of traffic than each of them individually. A possible extension to this work is the inclusion of data coming from different sources in the traffic estimation using data fusion techniques.

The methods proposed here add or remove vehicles from the section of interest, violating the conservation of vehicles principle. Since most of the time, counts from every entry and exit point are not available (and therefore the conservation of vehicles cannot be observed), the addition or removal of vehicles replaces the conservation of vehicles by performing the required local adjustments based on measurements. If all the counts are known, an additional constraint should be added to force the total number of vehicles on the section to remain the same. However, this would never be the case in practice due to measurement errors.

It would be interesting to determine how different sampling strategies affect the number of observations available, their accuracy, and the performance of the methods. Some evidence shown in Chapter 5 suggests that sporadically sampling several vehicles is preferred to frequently sampling fewer vehicles. A possible explanation is because the first strategy provides a better coverage in time and space than the first one. However, further analysis is needed for this topic to be conclusive.
Bibliography


[48] M. Treiber and D. Helbing. Reconstructing the spatio-temporal traffic dynamics from stationary detector data. *Cooperative Tr @nsport@tion Dyn@mics*, 1:3.1–3.24, 2002.


[52] J-L. Ygnace, C. Drane, Y.B. Yim, and R. de Lacivivier. Travel time estimation on the San Francisco Bay Area network using cellular phones as probes. California PATH


