Density/Flow reconstruction via heterogeneous sources and Optimal Sensor Placement in road networks

Enrico Lovisari, Carlos Canudas de Wit, Alain Kibangou

▶ To cite this version:

HAL Id: hal-01375928

https://hal.archives-ouvertes.fr/hal-01375928

Submitted on 3 Oct 2016

Density/Flow reconstruction via heterogeneous sources and Optimal Sensor Placement in Road Networks

E. Lovisari[a,1], C. Canudas de Wit[b,c,d,*], A.Y. Kibangou[b,c,d]

[a] *Volvo Cars Corporation, Göteborg, 405 31, Sweden*
[b] *Univ. Grenoble Alpes, Gipsa-Lab, F-38000 Grenoble, France*
[c] *Inria Grenoble Rhône-Alpes, F-38000 Grenoble, France*
[d] *CNRS, Gipsa-Lab, F-38000 Grenoble, France*

---

[*]Corresponding author. *Email:* `carlos.canudas-de-wit@gipsa-lab.grenoble-inp.fr`
[1]Formerly with Univ. Grenoble Alpes, Gipsa-Lab, and CNRS, Gipsa-Lab, and Inria Grenoble Rhône-Alpes, F-38000 Grenoble, France.

List of Figures

2

List of Tables

Abstract

This paper addresses the two problems of flow and density reconstruction in Road Transportation Networks with heterogeneous information sources and cost effective sensor placement. Following a standard modeling approach, the network is partitioned in cells, whose vehicle densities change dynamically in time according to first order conservation laws. The first problem is to estimate flow and the density of vehicles using as sources of information standard fixed sensors, precise but expensive, and Floating Car Data, less precise due to low penetration rates, but already available on most of main roads. A data fusion algorithm is proposed to merge the two sources of information to estimate the network state. The second problem is to place sensors by trading off between cost and performance. A relaxation of the problem, based on the concept of Virtual Variances, is proposed and solved using convex optimization tools. The efficiency of the designed strategies is shown on a regular grid and in the real world scenario of Rocade Sud in Grenoble, France, a ring road 10.5 km long.

Keywords

1. Introduction

The last decades have witnessed a considerable increase of traffic volumes, especially due to urbanisation in big metropolis, which was not matched by a comparable extension of road infrastructures. As a consequence, crucial freeways, highways and arterial roads have been steered to a state of near saturation, and experience on daily basis periods of congested traffic (Papageorgiou et al., 2007). In turn, congestion causes increased travel times and stop-and-go phenomena, leading to decreased safety, economical losses, and environmental and psychological hazards in terms of pollution and road rage (Bilbao-Ubillos, 2008). Increasing road capacity by extending road infrastructures, such as construction of new arterial roads, has been the standard way to cope with congestion problems, but it is infeasible when existing roads lie on built-in areas. Intelligent Transportation Systems (ITSs), on the contrary, are expected to provide robust techniques for real-time monitoring, prediction and actuation of traffic networks, and to better integrate with road and rail public transportation, by leveraging recent technological and theoretical advancements in distributed computation and communication.

(Please place Figure 1 about here.)

The present paper investigates the two interconnected problems of sensor network design and estimation in traffic networks. As illustrated in Fig. 1, we propose three procedures to address the problems of Optimal Sensor Placement, and of Traffic Reconstruction, via the two modules of Fundamental Diagram calibration and Density and Flow Estimation.

Standard devices to obtain information on the state of the network are fixed sensors such as induction loops and magnetometers. Placed over a section of road, they provide rich information on the vehicles that cross such a section over a pre-fixed period of time: 1) their number, or flow, 2) their average speed, and 3) their average density, or more precisely their occupancy (see Section 2). Current technology allows for very precise measurements, with relative errors of measured quantities against ground truth often being below $1 \sim 2\%$. However, deployment and maintenance of a sensing network requires considerable investment and manpower, and consequently sensing networks are usually designed to be as sparse as possible. In this paper we formulate and provide a solution to the Optimal Sensor Placement problem, that is, positioning sensors on the cells of the network given partial information on the system and in such a way to trade off between performance and cost. To this aim, we assume that traffic managers know the splitting ratios of the network, namely, the turning percentages at each road intersection, an information that can be obtained via standard surveys using optical count or radar sensors.

Optimal Sensor Placement is a ubiquitous problem that has received a high degree of attention in several communities due to its importance for network design. In Transportation Systems, it is of interest both in the dual-problem of best placement of hubs for cost-efficient transportation of goods and people (Shahabi and Unnikrishnan, 2014) and Origin-Destination coverage (Ehlert et al., 2006; Hu and Liou, 2014; Antoniou et al., 2016). In these works, and dif-

ferently from the present paper, the problem is cast as a mixed integer problem which corresponds to determining the minimal set of locations from which the flows on the whole network can be determined, and sensor measurements are assumed to be perfect. Finally, recent contributions focused on Origin-Destination flows and travel time estimation from an information theoretical point of view and under several possible cost functions (Zhou and List, 2010; Xing et al., 2013), and on probabilistic approaches to the sensor placement problem that take into account sensors failures and other random events Fei et al. (2013); Danczyk et al. (2016).

Once the sensing network has been designed an implemented, flow and density measurements are available. Sensors are not, however, the only source of information that we exploit. In fact, the recent spread of wireless devices allows sensing and communication capabilities unforeseeable up to few years ago. Limiting the attention to traffic applications, vehicles equipped with positioning devices (such as GPS) and able to communicate with an ITS monitoring system can act as a probes in the traffic and provide Floating Car Data (FCD), namely, information on the vehicles' positions and speeds. The collected data can be used to estimate the speed in the network, thus offering a second source of information. Due to privacy reasons, single vehicles traces are usually not directly used, but rather aggregated as average speed of vehicles in segments of road. Advanced methodologies ensure fine spatial partitions of the network, with segments as short as 250 meters (INRIX, 2014). Compared to fixed sensors, a service based on this technology can only make use of information coming from her customers, which are a fraction of the total vehicles on the road (the *penetration rate* of the system). This implies that speed measurements are less precise and flow measurement are unavailable. On the other side, since it exploits existing communication systems it is relatively inexpensive and, more important, already covers all major traffic networks.

Fixed sensor measurements and Floating Car Data provide rich information that we employ to address the problem of estimating road usage in terms of density and flow of vehicles in a traffic network. The latter are commonly considered a good representation of the state of the system, providing more information than average speed alone. In particular, they are of crucial importance for 1) forecasting travel time and traffic evolution, along with historical data; 2) informing in real-time drivers about the state of the network through navigation systems; 3) providing public authorities with statistical data to monitor the state of the network and predict dangerous scenarios; 4) computing and actuating control actions through traffic lights, ramp metering and speed limits, or, in the future, lane change and semi-autonomous routing and navigation (Papageorgiou et al., 2003, 1991; Pisarski and Canudas de Wit, 2012; Como et al., 2013).

Traffic models for analysis and control synthesis date back to the first half of the 20th century. The most celebrated macroscopic model is the PDE based Lighthill-Whitham and Richards (LWR) model (Lighthill and Whitham, 1955), which, based on fluid kinematics, is able to reproduce crucial phenomena such

as traffic shock waves. Discretization of the LWR-PDE is not straightforward but stable numerical schemes have been proposed, the most well known being the Cell Transmission Model (CTM) (Daganzo, 1994, 1995). Huge efforts have been put in the last 15 years to calibrate the CTM (Muñoz et al., 2006) and to unveil its system-theoretical properties (Morbidi et al., 2014). Fusion of flow, density and speed measurements has also been addressed, even though mostly considering single vehicles traces. Approaches range from signal processing techniques such as the generalized Treiber-Helbing filter (van Lint and Hoogendoorn, 2010), nonlinear versions of the Kalman filter in the context of Lagrangian sensing (Work et al., 2010), stochastic versions of the three-detector model (Deng et al., 2013), and Extended Kalman Filter Nantes et al. (2016). Alternative approaches do not rely on discretization of the LWR-PDE model and allow to cast problems of estimation and control as convex problems (Li et al., 2014). Finally, an extremely recent contribution (Wright and Horowitz, 2016) addresses the density estimation problem from a point of view very similar to ours and making use of particle filtering strategies.

In the first part of this paper we propose an algorithm that aims to reconstruct the traffic density and flow by fusing fixed sensors measurements and Floating Car Data. We employ a macroscopic model, partitioning the network in cells and assigning to each cell a density of vehicles. The latter evolves dynamically according to a first order mass-conservation law, similarly to the Cell Transmission Model. We inherit from the CTM the assumption that the inflow in a cell is a fixed linear combination of the outflows of the preceding cells. Differently from CTM, however, inflows and outflows in all the cells are estimated on the basis of the available flow measurements only. In addition, using the concept of Fundamental Diagram and the speed measurements, we compute an instantaneous (namely, only based on the latest available measurements) pseudo-measurement of the density. These quantities are then the inputs for the density observer. As an intermediate step, we propose a gradient descent method to calibrate the Fundamental Diagram which can be run offline and only requires fixed sensor measurements.

In the second part of the paper we address the problem of Optimal Sensor Placement, namely, the problem of finding the best location where to physically place sensors. This is based on trading off between two contrasting objectives: the first, to maximize the performance of state reconstruction; the second, to minimize the total economic cost of the network. The performance of the state reconstruction is usually related to the ability to properly estimate the density of vehicles in the roads. Unfortunately, nonlinearity and complexity of traffic systems make it hard to evaluate the performance of nonlinear observers. In order to simplify the setting, we consider the related problem of reconstruction of flows in a static setting. In particular, we consider as performance metric the error covariance of the Best Linear Unbiased Estimator of the cumulative flows in the network over a long period of time under a simple linear measurement model. The resulting cost has an analytical formulation and can be easily studied. The rationale of such a choice is the following: In the proposed den-

sity estimation algorithm availability of speed measurements via FCD makes the density estimation directly dependent on the quality of flows estimation. As such, sensor placement for optimal flow reconstruction is also expected to yield good performance for density estimation. The Optimal Sensor Placement problem can be then seen as trading off between the performance of such a flow estimator, and a cost that depends on the dimension of the sensing network. Since this is a combinatorial problem, we relax it using a method that we call Virtual Variance algorithm, based on the idea to associate to each sensor a virtual variance which is large when the sensor is not relevant for good reconstruction of the flow vector. The only input that the algorithm needs is the matrix of splitting ratios, that prescribes how vehicles split at each junction, and the nominal noise variance of each sensor. Furthermore, we discuss in detail two extensions of the proposed algorithm dealing with important scenarios. In the first, Optimal Sensor Placement with geographical constraints, we address the scenario in which sensors cannot be placed in a subset of cells of the network. In the second, Optimal Sensor Placement with Number of Sensors constraints, we deal with the case in which the maximum number of sensors is pre-specified, for example due to budget limitations.

To summarize, the contributions of this paper are the following: 1) we formulate the problem and we design an easily implementable approach to data fusion of fixed sensors measurements and Floating Car Data; 2) we propose a gradient descent calibration algorithm of the underlying macroscopic model; 3) we formulate the problem of Optimal Sensor Placement in terms of positions of sensors in a network when sensors are noisy and we provide an approximate solution using the concept of Virtual Variances; 4) we show the prowess of the devised Optimal Placement procedure on a regular grid, for which we offer a comparison between the solution found with our approach and the true optimal placement, found by exhaustive search; 5) we illustrate the performance of the Optimal Placement procedure and of the Reconstruction algorithm through extensive numerical experiments in the real-world scenario of Grenoble Traffic Lab (GTL) (Canudas de Wit et al., 2015), a sensing network deployed along the freeway "Rocade Sud" in Grenoble, France, with FCD provided by INRIX, one of the most well known traffic solutions companies.

The remainder of the paper is organized as follows: after setting up the notation, Section 2 describes the model for a Road Transportation Network. Section 3 formulates the problem of flow and density reconstruction and describes the proposed nonlinear observer, while the problem of optimal sensor placement is formulated and a solution based on the heuristic Virtual Variance algorithm is presented in Section 4. Finally, Section 5 illustrates the solutions on a regular grid and on the real world scenario of the freeway Rocade Sud in Grenoble, and Section 6 draws the conclusions and provides several future research directions.

1.1. Notation

The symbol $\mathbb{R}^{\mathcal{A}}$ ($\mathbb{R}^{\mathcal{A}}_+$) for a finite set $\mathcal{A}$ denotes the set of real vectors (real positive vectors) indexed by elements of $\mathcal{A}$. For $x \in \mathbb{R}^n$, $x \geq 0$ is meant

component-wise. The identity matrix of suitable dimensions is denoted I, and $|\mathcal{A}|$ is the cardinality of the set $\mathcal{A}$. The symbol $\mathbb{E}$ denotes the expected value.

A graph $\mathcal{G}$ is a pair $(\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ is called the set of nodes and $\mathcal{E}$ the set of edges. For an edge $e \in \mathcal{E}$, $t(e) \in \mathcal{V}$ and $h(e) \in \mathcal{V}$ are the tail note and the head node of e, so that $e = (t(e), h(e))$. Denote by $\mathcal{E}_e^+ := \{j \in \mathcal{E} : h(e) = t(j)\}$ and $\mathcal{E}_e^- := \{j \in \mathcal{E} : h(j) = t(e)\}$ the set of edges that follow or precede e, respectively. A path of length $n \geq 2$ is a sequence of edges $e_1, \ldots, e_n$ such that $e_{i+1} \in \mathcal{E}_{e_i}^+$ for all $i = 1, \ldots, n-1$. A path of length 1 is a path made of a single link. The matrix $L \in \mathbb{R}^{\mathcal{V} \times \mathcal{V}}$ is a sublaplacian of $\mathcal{G}$ if $L_{ej} > 0$ only if $(e, j) \in \mathcal{E}$, $e \neq j$, and $\sum_j L_{ej} \leq 0$.

Table 1 presents the notation that will be used throughout the paper.

(Please place Table 1 about here.)

2. Road Transportation System Model

We adopt a macroscopic approach by partitioning the lanes of the roads in a traffic network in cells. Cells that lie on the same section of a road and on different lanes are said to be parallel one each other (see Figure 2) We interpret each cell as an edge $e \in \mathcal{E}$ in a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, which models the whole network. In such a graph, nodes $v \in \mathcal{V}$ represent intersections between cells, for example, junctions at which roads intersect. Long roads might be partitioned into shorter segments, and are thus represented as a series of consecutive cells. Among the set of cells $\mathcal{E}$, we denote by $\mathcal{R}^i$ and $\mathcal{R}^o$ the set of onramps and offramps, respectively. In this paper, we shall call onramp (offramp) any gate, be it a real ramp, a connector, a secondary road, etc, which lets vehicles enter into (exit from) the network.

(Please place Figure 2 about here.)

We make the following connectivity assumption, which formalizes the mild requirement that any cell can be reached from an onramp and that vehicles from any cell can exit from the network.

Assumption 1. *For any cell e in $\mathcal{E}$, there is at least one onramp $j \in \mathcal{R}^i$ and one offramp $k \in \mathcal{R}^o$ such that e is an edge of a path from j to k.*

Time is discrete and slotted in intervals of duration $T > 0$. On each cell $e \in \mathcal{E}$, denote by $\rho_e(t)$ the density of vehicles, in number of vehicles per km^2, at time t, and let $\rho(t) = [\rho_e(t)]_{e \in \mathcal{E}}$, which we call the *state* of the network. The density of vehicles in a cell changes dynamically in time according to the following mass-conservation first-order model

$$\rho_e(t+1) = \rho_e(t) + \frac{1}{\ell_e}(f_e^{\text{in}}(t) - f_e^{\text{out}}(t)), \qquad \forall e \in \mathcal{E} \tag{1}$$

---

[2]While we employ SI units for simplicity and for coherence with the data in the Grenoble Traffic Lab, the presentation of our results would obviously be unchanged if other systems of measurements, such as the imperial system, were used.

where ℓ_e is the length of cell e, and $f_e^{\mathrm{in}}(t)$ and $f_e^{\mathrm{out}}(t)$ are the inflow and the outflow at cell e during the t-th time slot. As such, $f_e^{\mathrm{in}}(t)$ and $f_e^{\mathrm{out}}(t)$ represent a number of vehicles, not, as more common, a number of vehicles per unit of time, e.g., per hour.

To relate inflows and outflows we resort to the standard concept of splitting ratios. Indeed, denote by $R_{ek} \geq 0$ the fraction of vehicles that turn into cell k when they exit from cell e, which is the splitting ratio of e towards k. Clearly, $R_{ek} = 0$ if e and k are not consecutive, and $\sum_k R_{ek} = 1$, if $e \notin \mathcal{R}^o$. Such a cell change might correspond to continuing on the same lane, lane change in the same road, or taking a different road, as exemplified in Figure 2. From this moment on, we make the following assumption.

Assumption 2. *The set of splitting ratios* $\{R_{ej}\}_{(e,j)\in\mathcal{E}\times\mathcal{E}}$ *is known.*

Splitting ratios establish the relation $f_e^{\mathrm{in}}(t) = \sum_{j\in\mathcal{E}} R_{je} f_j^{\mathrm{out}}(t)$, for all $t \geq 0$ and for any cell $e \notin \mathcal{R}^i$, while $f_e^{\mathrm{in}}(t) = \lambda_e(t)$ for $e \in \mathcal{R}^i$, where $\lambda_e(t)$ is an exogenous external inflow. We set for sake of convenience $\lambda_e(t) = 0$ for $e \notin \mathcal{R}^i$. By stacking inflows and outflows into vectors $f^{\mathrm{in}}(t)$ and $f^{\mathrm{out}}(t)$, respectively, we can rewrite the previous relation in matrix form as

$$f^{\mathrm{in}}(t) = R^T f^{\mathrm{out}}(t) + \lambda(t) \tag{2}$$

where the matrix $R = [R_{ej}]_{e\in\mathcal{E}, j\in\mathcal{E}\setminus\mathcal{R}^i}$ is the *matrix of splitting ratios*. In the present paper it is assumed that the matrix of splitting ratios is fixed, pre-determined, and known. Its calibration, closely related to the estimation of Origin-Destination pairs, can be performed on single-lane freeways with on-ramps and offramps by taking ratios of flows on main line and ramps (Muñoz et al., 2006). We plan to extend this setting to networks by casting the problem as an optimization problem in future research.

2.1. On modelling of cell flows

Macroscopic models such as the CTM postulate that the flow $f_e^{\mathrm{out}}(t)$ that exits from cell e at time t is a deterministic function of the densities in the cells around e. In the simplest case, where e and j are two consecutive cells and $\mathcal{E}_e^+ = \{j\}$ and $\mathcal{E}_j^- = \{e\}$, then in the CTM

$$f_e^{\mathrm{out}}(t) = \min\{d_e(\rho_e(t)), s_j(\rho_j(t))\}$$

where $d_e(\rho_e(t))$ and $s_j(\rho_j(t))$ are the *demand* of cell e and the *supply* of cell j, and represent the maximum outflow from e and the maximum inflow into cell j, respectively. The resulting system is a Godunov scheme for discretization of the LWR-PDE model, and can be extended to the network case in various ways (Daganzo, 1995; Lovisari et al., 2015; Coogan and Arcak, 2014). While these models reproduce important phenomena that must be taken into account when modelling and controlling traffic networks, such as the movement of shockwaves, each of them can only partially represent traffic dynamics in networks.

For this reason, we will avoid to explicitly model the relation between flows and density, and we will limit to the standard

$$f_e^{\text{out}} = \rho_e v_e, \forall e \in \mathcal{E} \tag{3}$$

namely that the volume that exits from a cell in a period is proportional to the density of the vehicles in the cell, and to their speed. Again, we leave unmodelled the relation between these two quantities, because, as we will make clear in the following, we assume to have a direct measurement of the average speed in each cell (in the form of Floating Car Data). In conclusion, we consider from now on the dynamics of the real system to be dictated by Eqs. (1)-(3), where v_e, $e \in \mathcal{E}$, is an unmodelled quantity which depends on the local state of the network.

While in the model of our network we do not use an explicit relation between flows and densities, we shall use it for data fusion and estimation purposes. In particular, we write

$$\varphi_e = \varphi_e(\rho_e), \forall e \in \mathcal{E} \tag{4}$$

where φ_e is the flow of vehicles *at the sensor locations*, which we shall always assume to be at the end of the cell. The graph of the function $\varphi_e(\cdot)$, which is the *Fundamental Diagram* on cell e, is a concave function with $\varphi(0) = \varphi(\rho^{\text{jam}}) = 0$, where ρ^{jam} is the jam density, at which vehicles are too close one each other to move. The value of ρ^{jam} varies from 150 to 300 vehicles per km, and is approximately related to average vehicle length by the formula $\rho^{\text{jam}} = 1/\ell_{\text{ave}}$, where ℓ_{ave} is measured in km.

In this paper, and consistently with (1), we make the steady assumption that φ_e measures the flow of vehicles through a certain section of road *per sampling time T*, as opposed to standard choices such as flow per minute or per hour. This is done without loss of generality, as the two functions only only by a multiplicative constant, and solely for sake of notation and to simplify the discussion in the numerical section, where all relevant variables will be normalized to the Grenoble Traffic Lab sampling time $T = 15$ seconds.

2.2. Available measurements

In this paper we consider two heterogeneous sources of information: flow and density measurements from sensors, and Floating Car Data.

2.2.1. Flow and density measurements

Standard measurement devices for traffic are loop detectors or magnetometers, radar traffic detectors, or video detection systems. They are positioned at fixed and predefined positions in the network, monitor a section of road, and are able to detect and assign a timestamp to the event "a vehicle crossed the section". Information is then aggregated in time slots. For sake of simplicity and without loss of generality, we assume that such time slots correspond to the time discretization of the system (1). As such, measurements of the following two quantities are available at every time slot of duration T:

- Flow of vehicles $\varphi_e(t)$, by counting the number of vehicles crossing the section during the t-th time slot;

- Density of vehicles $\rho_e(t)$ over the section. The quantity that is actually measured by the aforementioned devices is the occupancy of vehicles $o_e(t)$, defined as the percentage of time any vehicle was standing over the section in the t-th time slot. Since occupancy is approximately related to the density of vehicles by $\rho_e \approx \frac{o_e}{100\ell_{\mathrm{ave}}}$, where ℓ_{ave} is the average length of a vehicle in km, we assume from now on that density can be directly measured.

Measurements are unavoidably noisy, with sources of noise ranging from temporary inability to detect changes of the magnetic field, too fast or too slow vehicles, blurred videos, etc. We adopt a simple additive noise model

$$\begin{aligned}
\varphi_e^m(t) &= \varphi_e(t) + \omega_e^\varphi(t), \quad e \in \mathcal{E}^m \\
\rho_e^m(t) &= \rho_e(t) + \omega_e^\rho(t), \quad e \in \mathcal{E}^m
\end{aligned} \tag{5}$$

where $\varphi_e^m(t)$ and $\rho_e^m(t)$ are flow and density measurements at time t, and $\omega_e^\varphi(t)$ and $\omega_e^\rho(t)$ are measurement errors whose stochastic properties depend on the performance of the sensor as well as on road and weather conditions, and $\mathcal{E}^m \subseteq \mathcal{E}$ is the set of cells equipped with sensors. Due to installation and maintenance costs, usually $|\mathcal{E}^m| << |\mathcal{E}|$.

Remark 1. *It should be mentioned that measurement systems based on magnetometers can be used to measure the average speed of vehicles crossing the section they monitor, in addition to flow and density, simply by deploying them in pairs monitoring sections that are at a fixed and known distance. The two consecutive instants at which the same vehicle crosses the two sections provides then a measurement of its speed. This type of installation is however not standard and more expensive due to additional hardware and software. For this reason, and to show that our approach does not need this additional information, we shall assume that no speed measurement from static sensors is available.*

2.2.2. Speed measurements

As already mentioned, recent technological advancements provide public and private companies with speed measurements in the form of Floating Car Data (FCD). In this paper we assume that the traffic network is partitioned in segments, let $\mathcal{S}$ be the set of all segments, and that aggregate speed measurements are available on each segment $s \in \mathcal{S}$. In particular, such an aggregate information is the *average* speed of the vehicles on each segment, where averaging is required for privacy reasons. While some classes of vehicles, such as taxi and buses, can indeed be traced, we do not employ this additional information, and leave the possibility to use it as future research. Notice that segment partitioning is, in general, different from cell partitioning, since the former depends on communication and sensing capabilities, while the latter on modeling choices. In this paper we assume that FCD segments are coarser than cells, so that each segment usually consists of several cells, as in Figure 3.

(Please place Figure 3 about here.)

Floating Car Data are less expensive and require less maintenance effort with respect to fixed sensors since they exploit existing communication architectures. For the same reason, they are already accessible almost everywhere once a data collecting mechanism is deployed. Despite such advantages, FCD also have drawbacks. Aside from the already mentioned possibly low penetration rate, information provided via FCD is usually averaged over a relatively long period of time. As an example, within the Grenoble Traffic Lab fixed sensors yield flow and density (and speed) measurements every $T = 15$ seconds. The FCD provided by INRIX are instead aggregated per minute, with standard practice ranging between 5 and 10 minutes. A comparison between sensors speed measurements from the GTL and FCD is provided in Figure 4, in which we plot FCD measurement and average of GTL measurements for slow and fast lanes at location Taillat on the Rocade Sud in Grenoble from 07:00:00 to 11:00:00 on April 24th, 2014 (see Paragraph 5.1.2 for details), both sampled at a rate of one measurement every 15 seconds (left panel) and averaged with a 5 minutes rate (right panel). The left plots clearly illustrates the difference between the high measurement rate of sensors and the effect of temporal averaging of FCD, the latter yielding a smoother signal. Such a phenomenon is much less apparent in the right plot, where the sampling time is the same. As a further comparison, we plot both speed measurements for all the cells on the main line at 09:00:00 on Figure 5, showing again that on average the measurements are in agreement. In fact, the average difference between sensor and FCD speed measurement is $m_e = -5.8$, with standard deviation $s_e = 13.5$, whereas the average speed on the Rocade is 82 km/h (according to FCD data).

(Please place Figure 4 about here.)

(Please place Figure 5 about here.)

We provide now a more formal description of the FCD speed measurements. We partition time into intervals of duration N time slots (NT seconds), that is, into *FCD acquisition intervals* $[0, N-1], [N, 2N-1], \ldots$. At the beginning of each interval, namely at time $t = N, 2N, 3N, \ldots$, new aggregate data are available in the form of the average speed during the whole last interval and over each segment (which, recall, can consist of several cells). For a segment $s \in \mathcal{S}$, we denote such aggregate datum by $v_s^m(k)$, where k refers to the k-th FCD acquisition interval. These speed measurements are kept for the whole subsequent FCD acquisition interval (namely, for N time slots), when they are again updated. During the first interval, namely, for $t = [0, N-1]$, no data are available, and hence the speed measurement can be set to an arbitrary value. In this paper we set it to the freeflow speed, which is a reasonable choice if the first FCD acquisition interval falls during night time, when traffic is very low. In formulae, the speed measurement on cell e at time t is

$$v_e^m(t) = \begin{cases} v_e^{\text{ff}}, & t \in [0, N-1] \\ v_{s(e)}^m(k), & t \in [kN, (k+1)N - 1] \end{cases} \tag{6}$$

where

- $v_e^{\text{ff}} > 0$ is the freeflow speed on cell e, namely, the speed of vehicles in low density regime.

- $v_{s(e)}^m(k)$, the aggregate datum, is given by

$$v_{s(e)}^m(k) = \frac{1}{N|s(e)|} \sum_{j \in s(e), \tau \in \mathcal{I}_t} v_j(\tau) + \omega_{s(e)}^m(k)$$

where $s(e)$ denotes the segment of which e is one of the cells (see Figure 3), $\omega_{s(e)}^m(k)$ is a measurement error whose stochastic properties depend on the performance of the sensor as well as on road and weather conditions, and $\mathcal{I}_t = \{\tau : \lfloor \frac{t}{N} \rfloor - 1 \le \frac{\tau}{N} < \lfloor \frac{t}{N} \rfloor\}$.

(Please place Figure 6 about here.)

3. Flow and Density Reconstruction

The first problem we address in this paper is Flow and Density Reconstruction on the basis of heterogeneous sources of information. Our aim is to build an observer for the densities of vehicles in all cells of the network given static sensor measurements and Floating Car Data.

We start by observing that Eqs. (1)-(2) cannot be directly used to observe the system except for the ideal scenario in which ideal measurements of the outflows $f_e^{\text{out}}(t)$, for all $e \in \mathcal{E}$ and for all times $t \ge 0$, and of the initial conditions of the system, are available. Such a naive observer would however be very sensitive to noise, as notice that while errors in the initial conditions correspond to offsets during the evolution of the system, noises in the flow measurements are integrated by the system's dynamics, thus possibly producing unbounded and/or unrealistic results. Since real systems are never error free, Eqs. (1)-(2) cannot be directly used to observe the system.

We address this difficulty by considering the following standard Luenberger-like observer

$$\hat{\rho}(t+1) = \hat{\rho}(t) + \mathbf{L}^{-1}(\hat{f}^{\text{in}}(t) - \hat{f^{\text{out}}}(t)) + \Gamma(\tilde{\rho}(t) - \hat{\rho}(t)) \tag{7}$$

where

- $\hat{\rho}(t)$ is the stacked vector of $\{\hat{\rho}_e(t)\}_{e \in \mathcal{E}}$, the density estimates on cells $e \in \mathcal{E}$ at time t;

- $\hat{f}^{\text{in}}(t)$ and $\hat{f}^{\text{out}}(t)$ are the stacked vectors of $\{\hat{f}_e^{\text{in}}(t)\}_{e \in \mathcal{E}}$ and $\{\hat{f}_e^{\text{out}}(t)\}_{e \in \mathcal{E}}$, the inflow and outflow estimates on cells $e \in \mathcal{E}$ at time t. Inflow and outflow estimates are to-be-designed functions of the flow measurements $\varphi^m(t) = [\varphi_e^m(t)]_{e \in \mathcal{E}^m}$, as detailed in the next section;

- $\tilde{\rho}(t)$ is the stacked vector of $\{\tilde{\rho}_e(t)\}_{e \in \mathcal{E}}$, which are density pseudo-measures on cells $e \in \mathcal{E}$ at time t. Density pseudo-measures are functions of both flow measurements and speed measurements $\varphi^m(t)$, $v^m(t) = [v_e^m(t)]_{e \in \mathcal{E}}$, as detailed in the next section;

- Γ is a tunable gain trading off between flow and density pseudo-measure;

- $\mathbf{L}$ is the diagonal matrix of cell lengths, with $\mathbf{L}_{ee} = \ell_e$ for all $e \in \mathcal{E}$.

We are interested in the following:

Problem 1 (Flow and Density Reconstruction using Heterogeneous Sources): Design the maps

$$\hat{f}^{\text{in}} = \{\hat{f}_e^{\text{in}}\}_{e \in \mathcal{E}} : \mathbb{R}_+^{\mathcal{E}^m} \to \mathbb{R}^{\mathcal{E}}$$
$$\hat{f}^{\text{out}} = \{\hat{f}_e^{\text{out}}\}_{e \in \mathcal{E}} : \mathbb{R}_+^{\mathcal{E}^m} \to \mathbb{R}^{\mathcal{E}}$$
$$\tilde{\rho} = \{\tilde{\rho}_e\}_{e \in \mathcal{E}} : \mathbb{R}_+^{\mathcal{E}^m} \times \mathbb{R}_+^{\mathcal{E}} \to \mathbb{R}^{\mathcal{E}}$$

to minimize the absolute errors with respect to real flows and densities

$$a^\rho(t, e) = |\hat{\rho}_e(t) - \rho_e(t)|$$
$$a^\varphi(t, e) = \left| \hat{f^{\text{out}}}_e(t) - f_e^{\text{out}}(t) \right| \tag{8}$$

3.1. A nonlinear observer for traffic networks

In this subsection we describe the proposed solution to Problem 1. It consists in an offline calibration procedure and an online filtering step.

3.1.1. Offline calibration

This paragraph is devoted to providing a solution for calibrating the Fundamental Diagram. We employ a gradient descent strategy, similarly to (Qu et al., 2015), for which we do not require a CTM formulation as in (Munoz et al., 2004).

Recall that the Fundamental Diagram is the graph of the function $\varphi_e(\cdot)$ that is the flow of vehicles at the point where sensors are placed on cell e and whose argument is the density of vehicles on cell e. As such, it can be only estimated on cells $e \in \mathcal{E}^m$, where measurements of flow and density are available. For a cell $j \in \mathcal{E} \setminus \mathcal{E}^m$, we assume that the Fundamental Diagram can be estimated by extending by linear interpolation the parameters on the cells in $\mathcal{E}^m$ that are close to j, with coefficients that depend on the mutual distance between the cells and on the type of road j belongs to.

The profile of Fundamental Diagram that we consider in the present paper is the following

$$\varphi_e(\rho) = \begin{cases} v_e^{\text{ff}} \rho, & \rho \le \rho_e^c \\ a_e \rho^2 + b_e \rho + c_e, & \rho_e^c < \rho \le \rho^{\text{jam}} \end{cases}$$

where

16

- ρ_e^c is the *critical density*. It partitions the set of densities $[0, \rho^{\mathrm{jam}}]$ into the *freeflow* low-density region $[0, \rho_e^c)$, in which the mutual influence of vehicles is small, from the high-density *congested* region $(\rho_e^c, \rho_e^{\mathrm{jam}}]$, in which speed decreases with density due to interaction of close vehicles;

- ρ^{jam}, the jam density, is assumed to be known a priori;

- $v_e^{\mathrm{ff}} > 0$ is the freeflow speed on cell e; the value $C_e = v_e^{\mathrm{ff}} \rho_e^c$ is the *capacity* of the section or road, namely, the maximum number of vehicles that can flow through it during a period T;

- we assume that the Fundamental Diagram congested region is a quadratic function of the density. The following relations among the parameters a_e, b_e and c_e hold for consistency

$$\begin{cases} a_e \left(\rho_e^c\right)^2 + b_e \rho_e^c + c_e = v_e^{\mathrm{ff}} \rho_e^c \\ a_e \left(\rho_e^{\mathrm{jam}}\right)^2 + b_e \rho_e^{\mathrm{jam}} + c_e = 0 \\ a_e \geq 0 \end{cases}$$

Remark 2. *Recall that we defined $\varphi_e(\cdot)$ to be the flow of vehicles through a section during a sample time T. As such, the units of the speed v_e^{ff} are km/T.*

Remark 3. *A standard choice for the Fundamental Diagram is the a triangular Fundamental Diagram, for which in the congested region*

$$\varphi_e(\rho_e) = \omega_e(\rho_e^{\mathrm{jam}} - \rho_e)$$

where ω_e is the wave speed *at section e. Clearly, our model recovers the latter with $a_e = 0$, $b_e = -\omega_e$, and $c_e = \omega_e \rho_e^{\mathrm{jam}}$. The choice of a quadratic Fundamental Diagram in congestion regime has been driven by the empirical observation, based on data on our experimental setting, that the triangular diagram tends to overestimate the flow in congestion, as it will be shown in Section 5. For this reason, we call it linear-quadratic Fundemental Diagram.*

An alternative appealing solution which fits our data is the inverted-λ fundamental diagram (Hall et al., 1986). However, the number of parameters to be estimated is higher in the latter case, and the resulting model is more complex as it involves hysteresis.

Since it is well known that deterministic Fundamental Diagrams are in any case only a rough approximation of the relation between flow and density, we chose the quadratic profile because it is simple to calibrate and to use.

We describe now the procedure for the calibration of the Fundamental Diagram. Let $e \in \mathcal{E}^m$ and let $\{(\rho_{e,k}^m, \varphi_{e,k}^m)\}_{k \in \mathcal{K}}$, $\mathcal{K} = \{1, \ldots, K\}$, the set of K density and flow measurements used as learning set and obtained via the fixed sensor on cell e. For sake of notation, and since all variables refer to cell e only, let us write from now on ρ_k^m and φ_k^m instead of $\rho_{e,k}^m$ and $\varphi_{e,k}^m$, and the same for the parameters of the Fundamental Diagram.

The proposed calibration procedure requires two steps

- Estimation of ρ^c and $C = v^{\mathrm{ff}}\rho^c$: the first step consists in estimating the critical density and the capacity C of the cell. We consider the standard least square estimation, which results into the following non-linear and non-convex minimization problem: given the set of measurements $\{(\rho_k^m, \varphi_k^m)\}_{k \in \mathcal{K}}$, $\mathcal{K} = \{1, \ldots, K\}$, solve

$$
\begin{aligned}
\min_{(\rho^c, C)} \quad & V_{(\rho^c, C)} = \tfrac{1}{2} \sum_{k=1}^{K} (\varphi_k - \varphi_{(\rho^c, C)}(\rho_k))^2 \\
\text{s.t.} \quad & 0 < \rho^c < \rho^{\mathrm{jam}} \\
& C > 0 \\
& \varphi_{(\rho^c, C)}(x) = \begin{cases} \frac{C}{\rho^c} x, & x \le \rho^c \\ \frac{C(\rho^{\mathrm{jam}} - x)}{\rho^{\mathrm{jam}} - \rho^c}, & x > \rho^c \end{cases}
\end{aligned} \tag{9}
$$

We aim to solve (9) by the following gradient descent with diminishing stepsize algorithm

- Basic step: initialize ρ_0^c, C_0. A reasonable choice is $\rho_0^c = 20$, which corresponds to the vehicles influencing one each other when the average distance among them is less than 50 meters, and $C_0 = v_e^{\mathrm{limit}} \rho_0^c$, where v_e^{limit} is the speed limit on cell e normalized by the sampling time T;

- n-th step: let (ρ_n^c, C_n) descend along the gradient of the cost, namely

$$
\rho_{n+1}^c = \rho_n^c - \frac{\delta}{n} \nabla_{\rho^c} V_{(\rho^c, C)}
$$

$$
C_{n+1} = C_n - \frac{\delta}{n} \nabla_C V_{(\rho^c, C)}
$$

with

$$
\nabla_{\rho^c} V_{(\rho^c, C)} = \sum_{k \in \mathcal{I}^{\mathrm{FF}}(\rho_n^c)} \left(\varphi_k - \varphi_{(\rho_{n-1}^c, C_{n-1})}(\rho_k) \right) \frac{C_{n-1}}{(\rho_{n-1}^c)^2} \rho_k
$$

$$
- \sum_{k \in \mathcal{I}^{\mathrm{C}}(\rho^c)} \left(\varphi_k - \varphi_{(\rho_{n-1}^c, C_{n-1})}(\rho_k) \right) C_{n-1} \frac{\rho^{\mathrm{jam}} - \rho_k}{(\rho^{\mathrm{jam}} - \rho_{n-1}^c)^2}
$$

$$
\nabla_C V_{(\rho^c, C)} = - \sum_{k \in \mathcal{I}^{\mathrm{FF}}(\rho_n^c)} \left(\varphi_k - \varphi_{(\rho_{n-1}^c, C_{n-1})}(\rho_k) \right) \frac{\rho_k}{\rho_{n-1}^c}
$$

$$
- \sum_{k \in \mathcal{I}^{\mathrm{C}}(\rho^c)} \left(\varphi_k - \varphi_{(\rho_{n-1}^c, C_{n-1})}(\rho_k) \right) \frac{\rho^{\mathrm{jam}} - \rho_k}{\rho^{\mathrm{jam}} - \rho_{n-1}^c}
$$

$$
\mathcal{I}^{\mathrm{FF}}(\rho^c) = \{ k \in \mathcal{K} : 0 < \rho_k \le \rho^c \}
$$

$$
\mathcal{I}^{\mathrm{C}}(\rho^c) = \{ k \in \mathcal{K} : \rho_k > \rho^c \}
$$

where the gradients $\nabla_{\rho^c} V_{(\rho^c, C)}$ and $\nabla_C V_{(\rho^c, C)}$ are computed at the point $(\rho^c, C) = (\rho_{n-1}^c, C_{n-1})$, and $\delta > 0$ is a fixed initial step size. Notice that if $\rho_n^c = 0$ then $\mathcal{I}^{\mathrm{FF}}(\rho^c) = \emptyset$, and conversely when $\rho_n^c = \rho^{\mathrm{jam}}$

18

then $\mathcal{I}^{\mathrm{C}}(\rho^c) = \emptyset$, and thus the previous summations are always well defined. Nonetheless, for numerical reasons, additional care should be taken in order to avoid $\rho^c < 0$ or $\rho^c > \rho^{\mathrm{jam}}$, for example projecting at each step ρ^c_{n+1} into $[0, \rho^{\mathrm{jam}}]$ after the gradient update.

– Stopping criterion: stop if $\left\| \begin{bmatrix} \rho^c_n \\ C_n \end{bmatrix} - \begin{bmatrix} \rho^c_{n-1} \\ C_{n-1} \end{bmatrix} \right\| < \varepsilon$ for some small threshold $\varepsilon > 0$.

• Calibration of the quadratic function in the congested region: the problem of calibrating the quadratic function for the congested region is cast into the quadratic problem: given the set of measurements $\{(\rho^m_k, \varphi^m_k)\}_{k \in \mathcal{K}}$, $\mathcal{K} = \{1, \ldots, K\}$, solve

$$
\begin{aligned}
\min_{(a,b,c)} \quad & \sum_{k \in \mathcal{I}^{\mathrm{C}}(\rho^c)} (\varphi_k - (a\rho_k^2 + b\rho_k + c))^2 \\
\text{s.t.} \quad & a\,(\rho^c)^2 + b\rho^c + c_e = C \\
& a\,(\rho^{\mathrm{jam}})^2 + b\rho^{\mathrm{jam}} + c = 0 \\
& a \geq 0
\end{aligned}
\tag{10}
$$

The problem (10) is computationally very simple and can be solved using off-the-shelf tools.

Notice that as side products of the previous procedure we can compute the freeflow speed as $v^{\mathrm{ff}} = C/\rho^c$ and, in case a bilinear Fundamental Diagram is also needed, the wave speed as $\omega = -C/(\rho^{\mathrm{jam}} - \rho^c)$, for each cell.

Finally, we remark that, in principle, one could add in (9) an upper bound on the capacity, for example $C^{\mathrm{max}} = v^{\mathrm{limit}}\rho^c$. Such an additional constraint can be included in the cost, and hence in the gradient iterates, via a standard barrier function. However, such a choice might be in practice too restrictive (for example, drivers tend to exceed the speed limit). Furthermore, since the algorithm aims to fit the (density, flow) pairs, it is to be expected that the resulting capacity does not exceed the maximum measured flow, which can be verified *a posteriori*. A capacity upper bound is therefore a modeling choice, which for the aforementioned reasons has not been implemented in the present paper.

3.1.2. Online density reconstruction algorithm

We assume from now on that Fundamental Diagrams have either been calibrated or extended on the whole network, and that the matrix of splitting ratios has been pre-specified or estimated on the basis of field surveys.

We propose the following online algorithm for Density Reconstruction

• at the beginning of the t-th time slot, a centralized computation unit

– receives measurements $\{\varphi^m_e(t)\}_{e \in \mathcal{E}^m}$;

– *flow estimation*: estimates the vector of outflows $\hat{f}^{\text{out}}(t)$ by solving the following minimization problem

$$\min_{\hat{f}^{\text{out}}} \quad ||(I - R^T)\hat{f}^{\text{out}}||^2 + \gamma \sum_{e \in \mathcal{E}^m}(\hat{f}_e^{\text{out}} - \varphi_e^m(t))^2$$
$$\text{s.t.} \quad \hat{f}^{\text{out}} \geq 0 \tag{11}$$

Problem (11) aims to a) match outflows and measurements where available, by penalizing the squared difference between the two, and b) to balance outflows according to the splitting ratios. The latter term provides the estimate of flows on cells in which no measurement is available, and is performed "as if" the network were at steady state, which is a simplifying assumption due to absence of a dynamical model for flows. The tunable parameter γ selects whether more weight is given to matching estimated outflows and measurements (high γ), or to estimate the flows as if the network were at steady state (low γ). Once Problem (11) is solved, the vector of estimate of the inflows is easily computed according to Eq. 2, by setting $\hat{f}_e^{\text{in}}(t) = \sum_{j \in \mathcal{E}_e^-} R_{je}\hat{f}_e^{\text{out}}(t)$.

– receives the measurements $\{v_e^m(t)\}_{e \in \mathcal{E}}$ when available, or holds the last speed measurements received;

– For each cell e, computes the two possible densities ρ_e^1 (freeflow) and ρ_e^2 (congested) corresponding to flow $\hat{f}_e^{\text{out}}(t)$ assuming that the local flow $\varphi_e = \varphi_e(\rho_e)$ is exactly determined by the Fundamental Diagram;

– For each cell e, computes the two velocities $v_e(\rho_e^1) = \frac{\hat{f}_e^{\text{out}}}{\rho_e^1}$ (freeflow) and $v_e(\rho_e^2) = \frac{\hat{f}_e^{\text{out}}}{\rho_e^2}$ (congested);

– Selects
$$\tilde{\rho}_e(t) = \arg \min_{i=1,2}\{|v_e(\rho_e^i) - v_e^m(t)|\}$$

as a rough estimate of the density. This estimate is only based on the actual measurements of flow and speed, and is generally very noisy, especially when the cell is in congestion. Therefore, the algorithm does not directly uses it;

– *density estimation*: for each cell $e \in \mathcal{E}$, lets the density estimate evolve according to the observer equation (7).

3.2. Observer dynamic analysis

This section is dedicated to the analysis of the estimation error dynamics for the observer (7). Let $e(t) = \rho(t) - \hat{\rho}(t) \in \mathbb{R}^{\mathcal{E}}$ be the estimation error at time t. By (1) and (7), the error evolution is given by

$$e(t + 1) = \rho(t) + \mathbf{L}^{-1}(f^{\text{in}}(t) - f^{\text{out}}(t))$$
$$- \hat{\rho}(t) - \mathbf{L}^{-1}(\hat{f}^{\text{in}}(t) - \hat{f}^{\text{out}}(t)) - \Gamma(\tilde{\rho}(t) - \hat{\rho}(t)) \tag{12}$$
$$= (I - \Gamma)e(t) - \mathbf{L}^{-1}(R^T - I)\Delta f^{\text{out}}(t)) - \Gamma \Delta \rho(t)$$

20

where we defined

$$\Delta f^{\text{out}}(t) := f^{\text{out}}(t) - \hat{f}^{\text{out}}(t), \qquad \Delta\rho(t) := \tilde{\rho}(t) - \rho(t)$$

the flow and density errors. The quantities $\Delta f^{\text{out}}(t)$ and $\Delta\rho(t)$ can be interpreted as noises, but we observe that they 1) depend on the measured quantities, and 2) do not follow a Gaussian distribution. For these reasons, a straightforward application of Kalman filter is not theoretically justified, and would not be optimal in the minimum variance sense. One could resort to employing extensions of Kalman filter based on system linearization, such as extended Kalman filter, but such strategies would be more complex than the proposed linear estimator, and their optimality would not be proved anyway. Finally, particle filters might provide better performance than both our algorithm and the extended Kalman filter, as they run in parallel a number of scenarios and iteratively select the most informative trajectories, but would be much more complex and computationally requiring than the simple linear observer proposed.

We observe now that the error equation (12) is essentially a low pass filter of the total noise signal $n(t) = \mathbf{L}^{-1}(R^T - I)\Delta f^{\text{out}}(t)) + \Gamma\Delta\rho(t)$ with stability and bandwidth dependent on the eigenvalues of Γ. Let in fact $\Lambda = I - \Gamma$. Stability, a strictly necessary condition, requires that $|\lambda| < 1$ for any eigenvalue λ of Λ. Within this range, eigenvalues of Λ close to 0 yield large bandwidth, hence a faster filtering but a more noisy error, and viceversa eigenvalues close to 1 imply small bandwidth, thus less influence of the total noise $n(t)$ but slower decaying rate of $e(t)$ to zero (in absence of $n(t)$).

Further analysis can be carried on under different assumptions on the noise terms $\Delta f^{\text{out}}(t))$ and $\Delta\rho(t)$. We shall consider for sake of simplicity the scenario in which $\Gamma = gI$, so that stability requires $g \in (0, 2)$. We first present a deterministic analysis based on Lyapunov theory and the concept of ultimate boundedness (Khalil, 2002). Assume that $||\mathbf{L}^{-1}(R^T - I)\Delta f^{\text{out}}(t))||_\infty \leq M^f$ and $||\Delta\rho(t)||_\infty \leq M^\rho$, namely, that the noise components of the error equation are bounded. Notice that such M^f and M^ρ are surely finite by definition of $\Delta f^{\text{out}}(t)$ and $\Delta\rho$. Let us consider the discrete time Lyapunov equation

$$(1-g)^2 p - p = -q \quad \implies \quad p = \frac{q}{g(2-g)}$$

where $q > 0$, and the candidate Lyapunov function $V(e) = p||e||^2$. Tedious but straightforward computations[3] yield

$$V(e(t+1)) - V(e(t))$$
$$\leq q\left(-||e(t)||_2^2 + \frac{2(M^f + gM^\rho)|1-g|\sqrt{N}}{g(2-g)}||e(t)||_2 + \frac{N(M^f + gM^\rho)^2}{g(2-g)}\right)$$

---

[3]Here we use that for $x \in \mathbb{R}^N$ it holds true $||x||_1 \leq \sqrt{N}||x||_2$.

We can now establish uniform boundedness. Since $q > 0$, the right hand side is a concave parabola in $||e(t)||_2$ with roots $\frac{(\pm 1 + |1-g|)(M^f + gM^\rho)}{g(2-g)}$. As a consequence, after possibly an initial transient, the error trajectory will be bounded within the sphere of $e(t)$ such that $||e(t)||_2 \le r(g) := \frac{(1 + |1-g|)(M^f + gM^\rho)}{g(2-g)}$, since outside such a sphere $V(t)$, and hence by construction $||e(t)||_2$, cannot but decrease. A reasonable tuning choice is to minimize the radius of such a sphere, namely to minimize $r(g)$ as a function of g. Straightforward computations yield that the minimum radius is $r_{\min} = M^f + M^\rho$, obtained for $g_{\min} = 1$. By this choice we make the filter as fast as possible and the ultimate boundedness radius as small as possible.

A second, probabilistic, analysis is based on a statistical description of $\Delta f^{\text{out}}(t))$ and $\Delta \rho(t)$. In particular, assume that $\Delta \rho(t)$ and $\Delta f^{\text{out}}(t)$ are two jointly independent and identically distributed stochastic processes with $\overline{\Delta f} = \mathbb{E}(\mathbf{L}^{-1}(R^T - I)\Delta f^{\text{out}}(t)))$, $\sigma_f^2 = \mathbb{E}||\mathbf{L}^{-1}(R^T - I)\Delta f^{\text{out}}(t)) - \Delta \bar{f}(t)||^2$, $\overline{\Delta \rho} = \mathbb{E}\Delta \rho(t)$, $\sigma_\rho^2 = \mathbb{E}||\Delta \rho(t)) - \Delta \bar{\rho}(t)||^2$. These values can be obtained numerically as empirical average and empirical second order moment of $\Delta f^{\text{out}}(t))$ and $\Delta \rho(t)$. We remark once again that these assumptions do not hold true in the framework under analysis, where the two processes are not independent and where their statistical properties depend in a complex way on the state of the system - this is one of the reasons for which directly employing a Kalman filter strategy is not justified. Nonetheless, they are instrumental for the tuning strategy presented in this section. Let $m(t)$ and $\sigma_e^2(t)$ be the expected value and the second moment of $e(t)$, which evolve according to $m(t+1) = (1-g)m(t) + \overline{\Delta f(t)} + g\overline{\Delta \rho(t)}$ and $\sigma_e^2(t+1) = (1-g)^2 \sigma_e^2(t) + \sigma_f^2 + g^2 \sigma_\rho^2$. Standard analysis establishes that asymptotically $m(t) \to \bar{m}(g) = \overline{\Delta \rho(t)} + \frac{\overline{\Delta f(t)}}{g}$ and $\sigma_e^2(t) \to \bar{\sigma}_e^2(g) = \frac{\sigma_f^2 + g^2 \sigma_\rho^2}{g(2-g)}$. Tuning the gain g thus involves trading off between minimizing the expected value of the error $\bar{m}_e(g)$, for which the best choice is $g_m^* = 2$ (the maximum possible), and minimizing the error variance $\bar{\sigma}_e(g)$, for which the best choice can be found to be $g_\sigma^* = \frac{-1 + \sqrt{1 + 4\sigma_\rho^2/\sigma_f^2}}{2\sigma_\rho^2/\sigma_f^2}$.

The two analysis of the error dynamics proposed in this section are based on simplifying assumptions, but provide qualitative insights on how to tune the parameter g. Additional information can be derived from the actual implementation of the algorithms. In our numerical experiments, for example, we chose a small g since we observed that the density estimation algorithm with $g \ge 1$ was very noisy and would sometimes provide negative, and hence physically inconsistent, density estimates.

4. Optimal Sensor Placement

The second problem we tackle in this paper is Optimal Sensor Placement, namely, the problem of deciding the position of sensors yielding a good trade off between performance and cost. Since assessing in a theoretical way the performance of algorithms for density reconstruction is difficult due to the nonlinear

nature of the system, we simplify the setting and limit our attention to estimation of cumulative flows, namely, of the total outflows from the cells. The resulting, static, problem is then considered as a proxy for the more complicated problem built on the *dynamic* density model.

We start by deriving some properties of cumulative outflows, and we proceed describing a simple linear model for cumulative flows estimation. This will help us formalizing the Optimal Placement problem.

4.1. Linear flow constraints

Let $f_e := \sum_{k=t_0}^{t_1-1} f_e^{\text{out}}(k)$ be the *cumulative outflow* from cell e, namely, the total flow through the cell over the period of time $[t_0, t_1]$, and let $f = \begin{bmatrix} f_1 & \dots & f_{|\mathcal{E}|} \end{bmatrix}^T \in \mathbb{R}_+^{\mathcal{E}}$ be the vector of cumulative outflows. By integrating the system's dynamics we have

$$\ell_e(\rho_e(t_1) - \rho_e(t_0)) = \sum_{j \in \mathcal{E}} R_{je} f_j - f_e, \qquad e \in \mathcal{E} \setminus \mathcal{R}^i$$

Assume now that $[t_0, t_1]$ is a period of time whose duration is high enough, and that at both times t_0 and t_1 the number of vehicles in the network is low, for example, assume that t_0 and t_1 correspond to two consecutive midnights. Then the magnitude of the vector of differences of vehicles $\{\ell_e(\rho_e(t_1) - \rho_e(t_0))\}_{e \in \mathcal{E} \setminus \mathcal{R}^i}$ is small if compared to the cumulative flows in the network, and the following relation holds approximately

$$\bar{L} f \approx 0, \tag{13}$$

where $\bar{L} \in \mathbb{R}^{\mathcal{E} \setminus \mathcal{R}^i \times \mathcal{E}}$, is the matrix obtained by removing from $L = R^T - I$ the rows corresponding to onramps. This imposes a linear constraint on the cumulative flows that we shall exploit in the next subsection.

4.2. Linear measurement model and the Optimal Sensor Placement problem

In this subsection we study the performance of a linear estimator of the cumulative outflows and we show how to formulate the problem of Optimal Sensor Placement.

Let $\mathcal{E}^m \subseteq \mathcal{E}$ be a set of cells in which sensors are placed. We assume the following simple linear measurement model

$$y = H_{\mathcal{E}^m} f + \omega^f \tag{14}$$

where

- y_s is the measurement of the s-th sensor, namely $f_e + \omega_s^f$ if the s-th sensor is located on cell e;

- $H_{\mathcal{E}^m} \in \{0, +1\}^{p \times n}$, $p = |E^m|$, $[H_{\mathcal{E}^m}]_{se} = 1$ if the s-th sensor is located on cell e, and $[H_{\mathcal{E}^m}]_{se} = 0$ otherwise, so that $H_{\mathcal{E}^m} \mathbf{1} = \mathbf{1}$ and $\mathbf{1}^T H_{\mathcal{E}^m} \mathbf{1} = p$;

- ω^f is a random noise vector with zero mean and covariance matrix Σ_{nom}, related to the measurement noise ω^φ described in the previous sections. For sake of simplicity, we shall often assume that the components of the noise, one for each sensor, are independent with same variance σ^2_{nom}, so that $\Sigma_{\text{nom}} = \sigma^2_{\text{nom}} I$.

Let now $n = |\mathcal{E}|$ and $r = \text{rank}\{\bar{L}\}$, and consider a matrix $V \in \mathbb{R}^{n \times r}$ whose columns are an orthonormal basis of the right kernel of $\bar{L}^T$, i.e., $\bar{L}^T V = 0$ and $V^T V = I$. From (13) we get (approximately) $f = Vz$ for some $z \in \mathbb{R}^r$, so that the measurement model can be rewritten as

$$y = H_{\mathcal{E}^m} V z + \omega^f . \tag{15}$$

Given y, consider a linear estimator of z, $\hat{z} = K_z y + q_z$, where $K_z \in \mathbb{R}^{r \times p}$ and $q_z \in \mathbb{R}^r$. The Best (minimum variance) Linear Unbiased Estimator of z corresponds to the solution to

$$\begin{aligned}
\min_{K_z, q_z} \quad & \mathbb{E}[(z - \hat{z})(z - \hat{z})^T] \\
\text{s.t.} \quad & \mathbb{E}[z - \hat{z}] = 0 \\
& \hat{z} = K_z y + q_z
\end{aligned} \tag{16}$$

We provide an equivalent formulation and a solution of (16) in Lemma 1 in Appendix. An immediate consequence of Lemma 1 is that the Best Linear Unbiased Estimator (BLUE) of f is

$$\hat{f} = K_f y = V(V^T H_{\mathcal{E}^m}^T \Sigma_{\text{nom}}^{-1} H_{\mathcal{E}^m} V)^{-1} V^T H_{\mathcal{E}^m}^T \Sigma_{\text{nom}}^{-1} y$$

and its error covariance is

$$V_p(\mathcal{E}^m) = \mathbb{E}[(f - \hat{f})(f - \hat{f})^T] = V(V^T H_{\mathcal{E}^m}^T \Sigma_{\text{nom}}^{-1} H_{\mathcal{E}^m} V)^{-1} V^T .$$

The quantity $V_p(\mathcal{E}^m)$ depends on a) the (right kernel of the) matrix of splitting ratios via the matrix V and the nominal variance of the noise ω^f, two parameters that are assigned, and b) on the locations of the sensors, which is the set $\mathcal{E}^m$, via the matrix $H_{\mathcal{E}^m}$. For this reason, we will take the magnitude of $V_p(\mathcal{E}^m)$, measured via its trace, as our metric to measure the performance of a sensor network placed on the cells $\mathcal{E}^m$.

Clearly, with no additional constraints the optimal placement is to equip every cell with sensors. This is straightforward as equipping all cells means setting $H_{\mathcal{E}} = I$, and from $H_{\mathcal{E}}^T H_{\mathcal{E}} = I \geq H_{\mathcal{E}^m}^T H_{\mathcal{E}^m}$ immediately descends $V_p(\mathcal{E}) \leq V_p(\mathcal{E}^m)$, for any $\mathcal{E}^m \subseteq \mathcal{E}$.

Each device has however a non-negligible purchase and maintenance cost, which has to be considered when designing the sensor network. For sake of simplicity, in this paper we make the simplifying assumption that the cost of a network over its lifetime is proportional to its number of sensors via a coefficient $c > 0$, so that the cost of deploying sensors on $\mathcal{E}^m$ is $c|\mathcal{E}^m|$.

We thus consider the following

Problem 2 (Optimal Sensor Placement)**:** Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a traffic network with splitting ratios R and cumulative flows noise variance σ_{nom}^2. Find $\mathcal{E}^m$ which solves

$$\min_{\mathcal{E}^m} \quad \text{trace}\{V_p(\mathcal{E}^m)\} + c|\mathcal{E}^m| \tag{17}$$

The optimal $\hat{\mathcal{E}}^m$, solution to (17), trades off between the network performance, which is measured by the trace of the estimator error covariance, and the total cost of the network. Clearly, the two have a contrasting effect on the number of deployed sensors. It is however inherently combinatorial, the optimal position of the sensor being in general hard to find and requiring an exhaustive search among all the possibilities, which is intractable even for relatively small network dimensions.

We approach the problem by proposing an heuristic that relaxes it into a convex problem. Such a strategy is described after a brief discussion on the minimum required number of sensors.

Remark 4. *The choice of measuring the performance via the trace of the estimator error covariance is one of several possible choices to* scalarize *the problem of minimizing the minimum covariance itself. Another interesting choice is the logarithm of the determinant of the estimator error covariance, proposed and studied for example in (Zhou and List, 2010; Xing et al., 2013), which also yields a convex optimization problem following the ideas proposed in the following sections. A further possibility is the $||\cdot||_2$ norm of the covariance. A comparison between these possible choices, in light of their geometrical interpretations, is due in future research.*

4.3. Minimum number of sensors

Before proposing our method for solving (17), we prove that there exists a lower bound on the number of sensors $|E^m|$ in order the trace $\{V_p(\mathcal{E}^m)\}$ to be finite. As it will be proven, below such number, which corresponds to the number of onramps of the system, the problem of reconstruction of flows admits infinite solutions.

To this aim, relabel the cells in such a way that onramps are the first $1, \ldots, |\mathcal{R}^i|$ cells so that we can partition $\bar{L}$ as

$$\bar{L} = \begin{bmatrix} L_{on} & L_{nn} \end{bmatrix}$$

where L_{nn} models the mutual influence of flows on non-onramp cells, and L_{on} models the influence of onramps on non-onramp cells.

It can be shown, and we provide a formal proof in Appendix, that the minimum number of sensors for the flow estimation problem to be solved is equal to the number of onramps $|\mathcal{R}^i|$.

We offer two interpretations on this fact:

- Ideal measurement scenario: assume $p = |\mathcal{E}^m|$ ideal measurements at sensors $s(1), \ldots, s(p)$ are available, $y_{s(i)} = f_{s(i)}$, $i = 1, \ldots, p$. Then a

25

solution to the system of equations

$$
\begin{cases}
\bar{L}\hat{f} = 0 \\
\hat{f}_{s(i)} = f_{s(i)}, i = 1, \ldots, p
\end{cases}
$$

is a candidate vector of cumulative flows. Then, if $p \geq |\mathcal{R}^i|$ the system has a unique solution which is the true vector of flows. Conversely, if $p < |\mathcal{R}^i|$, the system is undetermined;

- Noisy measurement scenario: assume $p = |\mathcal{E}^m|$ noisy measurements at sensors $s(1), \ldots, s(p)$ are available and assume to adopt the Best Linear Unbiased Estimator presented above to estimate the flows. If $p = |\mathcal{E}^m| < |\mathcal{R}^i|$, then rank $\left\{ H_{\mathcal{E}^m}^T H_{\mathcal{E}^m} \right\} < |\mathcal{R}^i| = \text{rank}\{V\}$, which implies that rank $\left\{ V^T H_{\mathcal{E}^m}^T H_{\mathcal{E}^m} V \right\} < |\mathcal{R}^i|$. However, $V^T H_{\mathcal{E}^m}^T H_{\mathcal{E}^m} V \in \mathbb{R}^{|\mathcal{R}^i| \times |\mathcal{R}^i|}$, so the matrix is singular, and therefore the trace of the error covariance is unbounded. Conversely, if $p \geq |\mathcal{R}^i|$ then rank $\left\{ V^T H_{\mathcal{E}^m}^T H_{\mathcal{E}^m} V \right\} = |\mathcal{R}^i|$ and the trace of the error covariance is bounded.

4.4. Relaxation via Virtual Variances

The solution that we propose is based on the observation that cells that are not endowed with sensors can be interpreted as cells in which sensors have infinite noise variance. It turns out that an equivalent formulation of (17) is

$$
\begin{aligned}
\min_{\mathcal{E}^m} \quad & \text{trace} \left\{ V (V^T \Sigma V)^{-1} V^T \right\} + c\sigma_{\text{nom}}^2 \mathbf{1}^T \Sigma^{-1} \mathbf{1} \\
\text{s.t.} \quad & \Sigma_{ee} =
\begin{cases}
+\infty, & e \notin \mathcal{E}^m \\
\sigma_{\text{nom}}^2, & e \in \mathcal{E}^m
\end{cases}
\end{aligned}
\tag{18}
$$

In this formulation

- the term $H_{\mathcal{E}^m}$ is the identity matrix, namely, all cells are endowed with a sensor – except, some of them have infinite noise variance and thus provide no information;

- the second term in the cost corresponds to $c|\mathcal{E}^m|$ as

$$
c\sigma_{\text{nom}}^2 \mathbf{1}^T \Sigma^{-1} \mathbf{1} = c\sigma_{\text{nom}}^2 \left(\sum_{e \in \mathcal{E}^m} \frac{1}{\sigma_{\text{nom}}^2} \right) = c \sum_{e \in \mathcal{E}^m} 1 = c|\mathcal{E}^m|
$$

In other terms, (18) corresponds to assigning a *virtual variance* $\Sigma_{ee} = \sigma_e^2$ to each sensor, and decide for which it should be $\sigma_e^2 = \sigma_{\text{nom}}^2$, the cells in $\mathcal{E}^m$ that actually provide information, and for which it should be $\sigma_e^2 = +\infty$. We call Σ the (diagonal) matrix of virtual variances, and we let the corresponding trace of error covariance be denoted, with an abuse of notation, $V_p(\Sigma) = \text{trace} \left\{ V (V^T \Sigma V)^{-1} V^T \right\}$.

Our approach is then based on the intuitive idea that increasing the variance on the sensors that are not crucial for the solution of (17) should not have a

26

strong effect on the performance term $V_p(\Sigma)$. More formally, we consider the following relaxed version of the previous problem

$$\begin{aligned} \min_{\Sigma \in \mathbb{D}_n} \quad & \text{trace}\left\{V(V^T \Sigma^{-1} V)^{-1} V^T\right\} + c\sigma_{\text{nom}}^2 \mathbf{1}^T \Sigma^{-1} \mathbf{1} \\ \text{s.t.} \quad & \Sigma_{ee} \geq \sigma_{\text{nom}}^2, \forall e \in \mathcal{E} \end{aligned} \tag{19}$$

where $\mathbb{D}_n$ is the set of diagonal matrices of dimension n.

We now slightly rewrite the cost. First of all, by the well known property of trace $\text{trace}\{AB\} = \text{trace}\{BA\}$, and by $V^T V = I$, it follows that $\text{trace}\left\{V(V^T \Sigma^{-1} V)^{-1} V^T\right\} = \text{trace}\left\{(V^T \Sigma^{-1} V)^{-1}\right\}$. Second, we perform the change of variables $\Omega = \Sigma^{-1}$. In this way, we obtain the following problem

$$\begin{aligned} \min_{\Omega \in \mathbb{D}_n} \quad & \text{trace}\left\{(V^T \Omega V)^{-1}\right\} + \eta \mathbf{1}^T \Omega \mathbf{1} \\ \text{s.t.} \quad & 0 \leq \Omega \leq \Sigma_{\text{nom}}^{-1} \end{aligned} \tag{20}$$

where η is a tunable parameter. The choice $\eta = c\sigma_{\text{nom}}^2$ is a natural one due to the previous discussion, but since η influences the relative weight of performance (penalized for low η) and cost (penalized for high η) we leave it as an additional degree of freedom. Notice, finally, that the term $\mathbf{1}^T \Omega \mathbf{1}$ corresponds to the ℓ_1 norm of the inverse of the variances, a term which is commonly used term to sparsify solutions of optimization problems.

The *Virtual Variance algorithm* proceeds as follows:

1. solve (20) and denote by Ω its solution;
2. compute $\Sigma^{-1} = \Omega$
3. discard all cells whose virtual variance is above a fixed discard threshold $\mathcal{T}_d$

As explained above, if the found solution provides high virtual variances at locations where sensors are redundant then this is effectively a way to select the most important cells where to place sensors.

However, this strategy does not, in general, provide good solutions to the problem. Indeed, numerical simulations have shown that in the considered scenario the solution of (20) can be often interpreted as endowing all cells of the network with sensors with average virtual variance, rather than keeping it low in some of them and high in others.

In order to enhance *diversity* between sensors, we enrich the cost of (20) with a term that aims to penalize homogeneity. This is reminiscent of *dissensus* (as opposed to consensus) strategies in multi agent networks, in which each agent possesses a value and the goal of the network is to differentiate as much as possible such values.

In this paper, we make the following simple choice. Let $W \in \mathbb{R}^{n \times n-1}$ be an orthonormal base of the subspace orthogonal to $\mathbf{1}$, that is, $W^* \mathbf{1} = 0$ and $W^* W = I$. We add to the cost in (20) a term that is proportional to $e^{-\mathbf{1}^T W^* \Omega \mathbf{1}}$. Since the columns of W span the orthogonal to $\mathbf{1}$, $W^* \Omega \mathbf{1}$ is high when the element on the diagonal of Ω, which are gathered in the vector $\Omega \mathbf{1}$, are different one with respect to the other, and is low otherwise.

We propose the following optimization problem

$$\begin{aligned}
&\min_{\Omega \in \mathbb{D}_n} \quad \text{trace}\left\{(V^T \Omega V)^{-1}\right\} + \eta \mathbf{1}^T \Omega \mathbf{1} + \kappa e^{-\mathbf{1}^T W^* \Omega \mathbf{1}} \\
&\text{s.t.} \quad\quad 0 \le \Omega \le \sigma_{\text{nom}}^{-2} I
\end{aligned} \tag{21}$$

While the optimization variable in the (21) is Ω, the real optimization variables are clearly its diagonal entries. Observe that a) Ω is an affine function of its diagonal entries; b) $V^T \Omega V$ and $\mathbf{1}^T W^* \Omega \mathbf{1}$ are affine functions of Ω; c) trace $\left\{X^{-1}\right\}$ is convex in the matrix X (Boyd and Vandenberghe, 2004) and e^{-x} is convex in x; d) composition of convex and affine functions, and nonnegative sum of functions, preserves convexity (Boyd and Vandenberghe, 2004). The cost function of (21) is thus a convex function of the diagonal entries of Ω under the natural requirement that the tunable parameters η, the *total variance weight*, and κ, the *discrepancy weight*, are chosen to be nonnegative. Since the inequality constraints are clearly convex in the diagonal entries of Ω, we conclude that (21) is a *convex optimization problem* having as optimization variables the diagonal entries of Ω.

Tuning of the parameters η and κ appearing in the previous optimization problem requires some careful selection so that the resulting virtual variances are strongly separated (in order to easily discard cells not providing sufficient information) while showing sufficiently good performance. We shall provide in the following paragraphs examples of application of the previous optimization problem for different sensor placement scenarios. We anticipate here that, by trial and error, the first parameter we tuned was κ, set to the value 20. This ensures that, as required, the virtual variances solution of (21) are distributed in a strongly bimodal way, with low and high values being different by several orders of magnitude. As a consequence, it is usually easy to distinguish among the two groups and discard cells whose contribution to the performance metric would be negligible. The parameter η was then chosen to show good performance/cost trade off, or according to additional constraints such as the ones described in the next paragraphs. In general, high η penalizes the number of sensors, thus yielding to solutions with higher virtual variances at the expense of poor performance.

We conclude this section observing that an important issue is sensitivity of the Virtual Variance algorithm with respect to the only datum it requires, the matrix of splitting ratios. We present in Appendix the results of a numerical study suggesting that such a sensitivity might be low in some situations.

4.4.1. Optimal Sensor Placement with geographical or budget constraints

In this paragraph we discuss two variations of the previous procedure, which address the additional problems of geographical constraints and of strict budget limitation.

Optimal Sensor Placement with geographical constraints

The first scenario we address is concerned with the scenario in which a) some cells cannot be equipped with sensors, for example for physical reasons,

and/or b) there are subsets of cells for which either all cells are endowed with sensors, or none are. An example of the latter constraints is a multi-lane road which is modelled using parallel cells (on different lanes) and on which the traffic manager can deploy induction loop. The former are buried underground and are usually required to cover the whole carriageway. As such, if i and j are, for example, two parallel cells on the two lanes on a certain section or road, then either i and j are both equipped with a sensor (i.e., the induction loop), or not.

To include this type of constraints in the proposed procedure, let $\mathcal{E}_{am} \subseteq \mathcal{E}$, $|\mathcal{E}_{am}| = m$, be the subset of available cells, and let $H_{\mathcal{E}_{am}} \in \{0, +1\}^{m \times n}$ be built as in Subsection 4.2. Further, let $\mathcal{S} \subseteq \mathcal{P}^{\mathcal{E}}$ be the set of subsets of $\mathcal{E}$ which must be simultaneously equipped, or not, with sensors, where $\mathcal{P}^{\mathcal{E}}$ is the powerset, or set of subsets, of $\mathcal{E}$.

Then the following problem

$$
\begin{aligned}
\min_{\Omega \in \mathbb{D}_k} \quad & \text{trace}\left\{(V^T H_{\mathcal{E}_{am}}^T \Omega H_{\mathcal{E}_{am}} V)^{-1}\right\} + \eta \mathbf{1}^T \Omega \mathbf{1} + \kappa e^{-\mathbf{1}^T W^* \Omega \mathbf{1}} \\
\text{s.t.} \quad & 0 \leq \Omega \leq \sigma_{\text{nom}}^{-2} I \\
& \Omega_{ii} = \Omega_{jj}, \forall i, j \in \sigma, \forall \sigma \in \mathcal{S}
\end{aligned}
\tag{22}
$$

is (21) once we constrain sensors to be placed on cells in $\mathcal{E}_{am}$ only and simultaneous sensor placement to happen according to the constraints specified by the set $\mathcal{S}$. Notice that the diagonal entries of the solution Ω are the inverse of the virtual variances on the cells in $\mathcal{E}_{am}$ only. The matrix W is defined as previously, but with suitable dimension $(m \times m - 1)$. As in the general problem, cells are chosen only if the corresponding virtual variance is below a certain threshold, and clearly cells that are not in $\mathcal{E}_{am}$ cannot be chosen. Notice that the constraint $\Omega_{ii} = \Omega_{jj}$ in (22), or the less requiring $|\Omega_{ii} - \Omega_{jj}| \leq \varepsilon$, for some tunable parameter ε, are convex constraints, so (22) remains convex.

Remark 5. *By the discussion in Subsection 4.3, the minimum number of sensors is $r = |\mathcal{R}^i|$. As such, if $|\mathcal{E}_{em}| < |\mathcal{R}^i|$ the problem (22) is not well posed and the solution will only have very high virtual variances. Clearly, such a solution is not acceptable and should be discarded.*

4.4.2. Optimal Sensor Placement with budget constraints

In this second scenario we discuss budget constraints in the form of constraints on the maximum number of chosen sensors, a very common requirement in real-case applications.

We propose a solution based on the following iterative approach:

- Initialization: set $\eta(0)$ and κ to some prespecified nonnegative values; $t_{\max}$ to the maximum number of iterations; $n_{\max}$ to the maximum number of sensors;

- t-th step

  - The problem (21) is solved with $\eta = \eta(t)$;
  - Let $n(t)$ be the number of sensors in the solution provided by the Virtual Variance algorithm. Then

* If $n(t) \leq n_{\max}$, or if $t \geq t_{\max}$, the procedure stops;
* Otherwise, set $\eta(t + 1) = g(\eta(t))$, where g is a monotonically increasing increasing function, and the procedure iterates.

The rationale behind this procedure is that, as previously discussed, η penalizes a low total sum of the virtual variances. Therefore, by iteratively increasing η the solution to (21) will tend to exhibit higher and higher virtual variances, thus reducing the number of sensors.

Remark 6. *Once again, and related to the discussion in Remark 5, the specified maximum number of sensors cannot be be less than $r = |\mathcal{R}^i|$ by the results presented in Section 4.3. If this is not the case, numerical experiments show that the algorithm either simply iterates until the number of iterations reaches $t_{\max}$, or the found solution exhibits an extremely high* trace $\left\{ (V^T H_{\mathcal{E}_{am}}^T \Omega H_{\mathcal{E}_{am}} V)^{-1} \right\}$ *- due to the fact that the internal matrix is (numerically) almost singular. As in Remark 5, such a solution is not acceptable and should be discarded.*

5. Numerical experiments

5.1. Numerical Experiments for the Optimal Sensor Placement

In this subsection we present the results of two numerical experiments. In the first, we solve the problem of Optimal Sensor Placement in a small (but not trivial) regular grid, in the second, we apply the procedure to the real-world case of the Rocade Sud.

(Please place Figure 7 about here.)

5.1.1. Regular grid

Consider the regular grid composed of 25 cells shown in Figure 7. We want to solve the problem (17) by (20) with parameters $\sigma_{\mathrm{nom}}^2 = 1$, nominal sensor variance, and $c = 1$, cost of a single sensor.

The network is small enough to run an exhaustive search to solve the problem (17). In particular, for each $h = 4, 5, 6, \ldots, 21$, we try all possible combinations of $h = |\mathcal{E}^m|$ sensors, thus finding the one that minimizes $V_p(\mathcal{E}^m)$.

We also run our Virtual Variance algorithm with total variance weight $\eta = 2$ and discrepancy weight $\kappa = 20$. Further, we set the threshold on the virtual variance for discarding a sensor to $\mathcal{T}_d = 100$.

We illustrate the results in Figures 7 and 8. In Figures 7 the 8 cells chosen by the Virtual Variance algorithm are represented as a green dot marks, and the best possible placement with 8 sensors (found by exhaustive search) as red dot marks. It can be noticed that both procedures place most of sensors at the boundaries of the network.

Figure 8 shows instead the total cost of the best placement obtained through exhaustive search for $h = 4, \ldots, 21$, and the total cost found by the Virtual Variance algorithm with 8 sensors. By total cost we mean the sum of estimation performance and network cost $V_p(\mathcal{E}^m) + c|\mathcal{E}^m|$, which, notice, is *not* the metric that is used in the Virtual Variance algorithm. Nonetheless, it is appreciable

that the Virtual Variance algorithm not only provides a solution whose number of sensors is close to the global optimum (which is with 6 sensors), but also that, using 8 sensors, the Virtual Variance algorithm places them almost in the optimal way.

(Please place Figure 8 about here.)

5.1.2. Rocade Sud

Our second experimental setting is the Grenoble Traffic Lab (GTL), a network of sensors deployed for monitoring and research purposes along the Rocade Sud, a peri-urban 10.5 km long freeway connecting the two highways A41 (north-west) to A480 (south) in the town of Grenoble in the south of France, see Figure 9. The network is composed of 135 magnetometers buried in the ground on both lanes along the main line every 250 meters (on average), on each onramp and offramp, and on three connectors from urban roads to three onramps, for a total of 68 *sensing locations*. For our purposes, each sensing location will correspond to one pair of sensors. For a detailed report on the GTL, we refer to (Canudas de Wit et al., 2015). Furthermore, and for sake of simplicity, while in the real network sensors are deployed in pairs, we shall assume from now on that each sensing location has only one sensor (as we shall always discard sensor speed measurements).

Figure 9 shows the position of each of the 22 sections of the main line in which there are sensing locations on both slow and fast lanes (and usually a ramp). In the same figure we also show a stylized representation of the freeway, including ramps and queues, the positions of the 68 sensing locations, and the distance between consecutive measurement sections along the main line.

We partition the Rocade in cells in such a way each cell includes one sensor, so that in Figure 9 each numbered circle also corresponds to one cell.

Here, we do not consider onramps and offramp, limiting our attention to the main line of the Rocade Sud. The reason for this choice is that the Rocade has 10 onramps along its main line, which, summed to the two cells in the very first section of road, imply a minimum number of sensor of 12 by the discussion in Subsection 4.3. While this number is not high on its own, numerical experiments not reported in this paper have shown that good estimation performance require a number of sensors which is too high in most realistic (i.e., non academical) implementations.

The corresponding reduced graph representation, essentially made of several groups of parallel edges, consists of 46 cells. We provide a stylized version of it in Figure 10.

(Please place Figure 9 about here.)

The matrix of splitting ratios could be estimated from the data, but we make here a simpler assumption and assume that vehicles split according to the following rule: vehicles on slow lane cells remain on slow lane or turn into fast lane with a 70%-30% rule, and analogously for vehicles on the fast lane. If the next section has three parallel cells (cells 22-23-24 and cells 66-67-68), vehicles spit uniformly in such three cells.

We did not run an exhaustive search due to the relatively high dimension of the network and the consequent relevant computational load. Our comparison is instead with the locations of fixed loops installed by the Government Agency *Centre national d'information routiére (CNIR)* (Centre national d'information routiére (CNIR), 2014), which correspond to cells marked with a red dot in Figure 10. We shall show that the latter are positioned in a way that is in good accordance with the results of the Virtual Variance algorithm, even though our procedure allows for a slightly better design.

We run the Virtual Variance algorithm in three scenarios: 1) unconstrained scenario with total variance weight $\eta = 0.2$ and discrepancy weight $\kappa = 20$; 2) unconstrained scenario with $\eta = 1$ and $\kappa = 20$; 3) constrained scenario with number of sensors at most 10, initial $\eta = 0.2$, and $\kappa = 20$. We assume that $\sigma_{\text{nom}}^2 = 1$ and that the cost per sensor is $c = 1$. In all cases, sensors are constrained either to be present in both lanes on a same section, or to be absent.

The results are summarized in Figure 10 and Table 2. The tables shows the number of sensors in the solution computed via the Virtual Variance algorithm, as well as the trace of the corresponding estimation error covariance $V_p(\mathcal{E}^m)$ and the total cost $V(\mathcal{E}^m) = V_p(\mathcal{E}^m) + c|\mathcal{E}^m|$. In Figure 10, cells found in the unconstrained scenario with $\eta = 0.2$ are marked with a green dot. As can be seen in Table 2, our algorithm requires one less sensors than the network deployed by CNIR and in addition the trace of the error covariance is smaller.

In the constrained scenario and in the unconstrained scenario with high η (which, as explained above, indirectly penalizes the number of sensors), the trace of the error covariance $V_p(\mathcal{E}^m)$ increases, as expected. Furthermore, the chosen cells in the latter two cases are subsets of the cells chosen in the unconstrained case: in particular, in the constrained scenario all cells are kept except 13, 14, 49 and 50, and in the unconstrained scenario with high η the algorithm further discards cells 64 and 65.

(Please place Figure 10 about here.)
(Please place Table 2 about here.)

5.2. Density reconstruction - experimental results

We provide here numerical results of the implementation of the data fusion algorithm for density reconstruction on data from the Rocade Sud.

On each sensing location and every $T = 15$ seconds, the system counts the number φ_e^m of vehicles that crossed the location, their average speed v_e^m, and the average occupancy o_e^m of the location. Since the latter is approximately proportional to the density of vehicles, so we shall assume that sensors can directly measure densities. In addition to fixed sensors, we use Floating Car Data provided by INRIX. Following INRIX schema, the Rocade has been further partitioned into *FCD segments*. One measurement of average speed is available on each FCD segment every 1 minutes. FCD segments partition the whole main line of the Rocade and include most onramps and offramps, but single lanes are not distinguished along the main line. FCD segments are represented in Figure 9 as rectangles encircling several sensing locations/cells.

For our experiments, we employ the sensor configuration obtained in the previous section via the Virtual Variance algorithm with $\eta = 0.2$. In particular, and in order to prove that the method shows good performance even with sparse equipment, we only use the sensors on the main line, which are, for reference, shown in light blue in Figure 9. Further, we don't use any information on flow or speed on the ramps.

Calibration of the Fundamental Diagram was performed via the algorithm described in Paragraph 3.1.1 and using the data from the GTL sensor network from April 10th, 2014, a working day (a Thu.) exhibiting very standard traffic pattern:

- very limited night time traffic;

- a peak of congestion in the morning (8:00 - 10:00), triggered by vehicles exiting towards the city from the Rocade at the offramp of Eybens (cells 37/38) and spilling back until Meylan (cells 2/3), and a second, smaller peak of congestion triggered by vehicles entering in A480 at Rondeau (cells 66/67/68) but blocked by the high traffic on A480, and spilling back until around Libération (cells 61/62);

- a third, smaller, congestion triggered around Eybens around 14:00-15:00;

- in general, medium/heavy but fluid traffic from 10:00 to 16:00

- a second peak of congestion in the afternoon, again triggered by congestion at Rondeau at around 16:00, spilling back on the whole freeway in around 60 minutes, and lasting approximately two hours.

We considered an average vehicle length of $\ell_{\mathrm{ave}} = 5$ meters, close to the value provided in (Treiber and Kersing, 2013). This yields a jam density value of $\rho^{\mathrm{jam}} = 200$ veh/km.

(Please place Figure 11 about here.)

As for the previous section, we consider the following simple rules to set the matrix of splitting ratios

- let e be a fast lane cell. Then 70% of vehicles continue on the fast lane cell and 30% turn into the slow lane cell; if in the following section there are three parallel cells, vehicles split uniformly;

- let e be a slow lane cell. If among the following cells there is not an offramp, then 70% of vehicles continue on the slow lane cell and 30% turn into the fast lane cell; if in the following section there are three parallel cells, vehicles split uniformly. Otherwise, 20% of the flow is directed towards the offramp, and the rest splits as previously specified;

- if e is an onramp cell and j is the following slow ramp cell, then $R_{ej} = 1$.

- if e is a queue cell and j is the following onramp cell, then $R_{ej} = 1$.

In words, vehicles split according to a 70%-30% lane-change rule in the cells on the main line, and at each offramp approximately 10% of vehicles exit from the freeway.

5.2.1. Implementation

To assess our method, we considered the whole month of April 2014 (except April 13th, a Sun., for which FCD measurements were not provided). A typical result of calibration of the Fundamental Diagram is illustrated in Figure 11, which shows in thick black the linear-quadratic Fundamental Diagram, in dashed thick black the corresponding standard linear Fundamental Digram in congestion regime, and as crosses the pairs (density, flow) measured on a day different from that used for calibration, in this case April 24th, 2014. As standard and well known, data in freeflow regime are in good accordance with the linear part, while data in congested regime are much more scattered and more difficult to fit. As it can be noticed, the standard bilinear Fundamental Diagram overestimates the flows in congested regime (the dashed think line is on average higher than the corresponding pairs (density, flow)), while the quadratic curve seems to better capture the average flow-density relation. Nonetheless, it is clear that the so found curve is only a very crude approximation of such a relation, which might be better captured using a stochastic description (Jabari and Liu, 2013). Investigation of the latter possibility will be the focus of future research.

The proposed algorithm was implemented in Matlab on a non dedicated commercial laptop with 2.1 GHz i7-4600U CPU and 8 GB RAM. Optimization problems, required both in offline and online steps, were solved using standard Matlab functions as well as the modelling and optimization system `CVX` (Grant and Boyd, 2013, 2008). The time required for calibration of the Fundamental Diagrams is between 30 and 40 seconds for each cell, while reconstruction of all the samples for a whole day requires less 10 minutes, averaging 100 ms per sample.

(Please place Figure 12 about here.)
(Please place Figure 13 about here.)

Typical results are reported in Figures 12-13. For validation purposes only, density and flow measurements from all GTL fixed sensors are considered ground truth. As such, the left panels show the evolution of the *"true" measured density and flow* in all the cells on the main line, over the whole day, for each day. On the x-axis, the 46 sensing locations along the main line (numbers correspond to the labels in Figure 9), on the y-axis, the 5760 time slots over the whole validation day (one slot every $T = 15$ seconds). The chosen colors range from green (low density/flow) to yellow (medium/critical density/flow) to red (high density/flow). In the right panels, we show, using the same legend, the results of density and flow reconstruction. As it can be observed, the estimation algorithm is able to represent the four congestion events described in the previous section in a reasonably good way, given the limited amount of information employed; in particular, observe that the two small congestions at Rondeau during the morning and at Eybens during early afternoon, when present, are both detected. Notice that the resulting estimate remains good and reasonably close to the real profile, despite the absence of flow measurements on ramps, which in the real system have a strong impact on the traffic dynamics, especially in

34

peak hours. We further observe that the observer is able to capture different regimes in parallel cells. An example of such a phenomenon is the succession of congested/non congested cells 28-29-32-33 at 16:00 (Saint-Martin-d'Héeres – Saint-Martin-d'Héeres entrance), which correspond to slow/fast lanes.

(Please place Table 3 about here.)

(Please place Table 4 about here.)

The performance of the algorithm is quantitatively illustrated via the absolute error between measured and estimated flows and densities (8). The results are summarized in Table 3 and Table 4, which report the maximum absolute error δ on the 75%, 90% and 95% of the pairs (cell, time slots), for all cells on the main line and all samples during a day, for all the considered days, and for both densities and flows. In Table 3 time slots correspond to the original 15 seconds time slots. In Table 4 we consider time slots of 5 minutes and we average in time all quantities. For the original slots, the table shows an average error of less than 11 veh/km for the 75% of pairs and less than 23 veh/km for the 90% of the pairs. In 5% of the (cell, time) pairs the error is higher but still less than around 35 veh/km. The magnitude of the errors is influenced by the high sampling rate of the system, which, as it is well known, increases the noise power. This is confirmed by a comparison with Table 4, for the averaged system, which shows errors on average 40% less for flows and $25 - 30\%$ less for densities.

We conclude with a comparison with an an oracle that knows

- the exact outflow $f_e^{\mathrm{out}}(t)$ for *every* cell e and *every* sample time t (namely, for each 15 seconds time slot);

- whether the cell is in freeflow or in congestion, for *every* cell e and *every* sample time t.

The oracle then runs the same estimation algorithm as described above. Notice that by exploiting its available information, the oracle is able to propagate the flows in the network with zero error (i.e., $\Delta f^{\mathrm{out}} = 0$) and always knows whether a cell's density is low (if the cell is in freeflow) or high (if the cell is congested). Consequently, the only source of noise that affects the oracle is the model uncertainty which is inherent in the Fundamental Diagram. The average error obtained by the oracle is 5.8 for the 75% of the pairs (cell, time) and 12.5 for the 90% of the pairs, even with the high amount of additional (and precise) information available to the oracle. Indeed, this confirms that estimation in traffic systems is a difficult task, and that errors of absolute magnitude around 10-20 veh/km can be acceptable, as they capture the qualitative trend features of the traffic system - such as low density, medium density, high density. On the other side, estimate of flows are rather low, being less than $1 \sim 2$ vehicles for the 75% of the pairs (cell, time), and less than $2 \sim 3$ for the 90% of the pairs.

The biggest difference between estimated and measured densities can be observed at Eybens exit (cells 37-38), where the estimated flow constantly predicts a higher density than the measured one. The explanation is straightforward: Recall that we do not use ramp data in order to show the prowess of our method

even employing a very small number of sensors. However, as mentioned in the description of the data, the exit of Eybens is a critical point whose ramp is selected by a high fraction of vehicles to exit the Rocade towards the town; in addition, the corresponding cells belong to a long FCD segment running from Saint-Martin-d'Hères (cells 32-33) to Eybens entrance (cells 41-42), which provides, during peak time, just one set of rather low speed measurements, which do not distinguish between the stretch of road before Eybens exit (congested and at low speed) against that after Eybens exit (uncongested and at high speed). Due to the so obtained low speed measurement, the algorithm tends to estimate a high number of vehicles along the whole segment, instead of two different regimes before and after the ramp. Analogously, we observe a mismatch between measured and estimated flow in the first and last sections, which are due to unobserved flows from onramp and to offramps. A second discrepancy is the smoothness of the reconstructed density and flow, as compared with the more scattered measurements. The latter is due to the high measurement rate, which during stop-and-go phenomena results in measurements which rapidly oscillate between stopped vehicles and low or medium speed. On the converse, the optimization based flow reconstruction and the first order mass conservation law for densities have a low pass effect therefore producing smoother, more regular patterns. Further research direction will investigate the possibility to detect stop-and-go phenomena and reproduce, at least qualitatively, the resulting irregular patterns.

6. Conclusions

This paper addressed the problems of fusion of heterogeneous sources of information for density estimation in Road Transportation Networks and optimal sensor placement via a heuristic that we called Virtual Variance algorithm. A gradient descent procedure for the calibration of the Fundamental Diagram is also discussed. Efficacy of the proposed solutions is illustrated on a regular grid and on the real world scenario of the Rocade Sud in Grenoble. Future research directions include and are not limited to estimation of statistical properties of measurement noises from real data and development of stochastic models for the relation between flows, speed and densities, optimization of the observer's parameters for minimization of mean-square reconstruction error, calibration of the matrix of splitting ratios, tuning of the weights in the Virtual Variance algorithm, and extension of the optimal sensor placement strategy to maximize density reconstruction performance.

7. Acknowledgements

Appendix

An equivalent formulation for (16) *and its solution*

The following Lemma formulates a problem that is equivalent to (16). The proof is straightforward.

Lemma 1. *The solution* (K_z, q_z) *to* (16) *is given by* $q_z = 0$ *and* K_z *the solution of*

$$\begin{aligned} \min_{K_z} \quad & K_z \Sigma_{\mathrm{nom}} K_z' \\ \mathrm{s.t.} \quad & K_z H_{\mathcal{E}^m} V = I \end{aligned} \tag{23}$$

which is

$$K_z = (V^T H_{\mathcal{E}^m}^T \Sigma_{\mathrm{nom}}^{-1} H_{\mathcal{E}^m} V)^{-1} V^T H_{\mathcal{E}^m}^T \Sigma_{\mathrm{nom}}^{-1} ,$$

with error covariance

$$\mathbb{E}[(z - \hat{z})(z - \hat{z})^T] = (V^T H_{\mathcal{E}^m}^T \Sigma_{\mathrm{nom}}^{-1} H_{\mathcal{E}^m} V)^{-1}$$

Proof of minimum number of sensors

Recall that $\bar{L} = \begin{bmatrix} L_{on} & L_{nn} \end{bmatrix}$ is the lower part of the matrix $L = I - R^T$ once we relabel the cells in such a way that the onramps are the first $|\mathcal{R}^i|$ cells.

Define a dual graph $\mathcal{G}^d = (\mathcal{V}^d, \mathcal{E}^d)$ in which the roles of cells and junctions are reversed, and in particular in which $\mathcal{V}^d = \mathcal{E} \setminus \mathcal{R}^i$ and $(e, j) \in \mathcal{E}^d$ if $e \neq j$ and $[L_{nn}]_{ej} \neq 0$. Then it is easy to see that L_{nn}^T is a sublaplacian of $\mathcal{G}^d$. The following result is adapted from (Lovisari et al., 2014).

Lemma 2. *Let* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ *be a graph and* $J \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{V}|}$ *be a sublaplacian of* $\mathcal{G}$. *Then all the eigenvalues of* J *have negative real part except possibly eigenvalues in* 0. *Moreover, if* $\mathcal{S}$ *is the set of nodes* v *for which* $\sum_u J_{vu} < 0$, *then* J *is invertible if for every* u *there exists a directed path in* $\mathcal{G}$ *from* u *to a node* $v \in \mathcal{S}$.

In the case under analysis, we take $J = L_{nn}^T$ and $\mathcal{S}$ to be the set of cells directly following an onramp, namely $\mathcal{S} = \{e \in \mathcal{E} : \exists j \in \mathcal{R}^i : R_{je} > 0\}$. Recall that by assumption for every cell $e \notin \mathcal{R}^i$ there exists an origin $j \in \mathcal{R}^i$ and a path from j to e, so there must also be a $k \in \mathcal{S}$ and a path from k to e (at most being $k = e$), so the assumptions of Lemma 2 are satisfied. This establishes that L_{nn} is invertible, and therefore that $\bar{L}$ is a full row-rank matrix with rank $|\mathcal{E} \setminus \mathcal{R}^i|$. Since the number of columns of $\bar{L}$ is $|\mathcal{E}|$, it follows that its kernel has dimension $r = |\mathcal{R}^i|$, and thus rank $\{V\} = |\mathcal{R}^i|$.

(Please place Figure 14 about here.)

Sensitivity of the Virtual Variance algorithm with respect to splitting ratios

In this section we present numerical results suggesting that the sensitivity of the Virtual Variance algorithm, namely, the sensitivity of sensor positioning as a function of small variations in the splitting ratios, might be low in some scenarios. We have considered the regular grid presented in Section 5.1.1 and run 100 times the Virtual Variance algorithm with $\eta = 2$, each time adding zero mean gaussian noise with variance $\sigma = 0.1$ to each nonzero entry of the nominal matrix of splitting ratios, and then renormalizing the latter to obtain again

a row-stochastic matrix. The average of the 100 instances of the cost $V(\mathcal{E}_m)$ (namely, the estimator variance) is $m_{\text{cost}} = 6.93$ with variance $\sigma_{\text{cost}} = 0.28246$. The number of chosen cells varies between seven and nine. In Figure 14 we plot the distribution of cell usage. The figure is to be interpreted as follows: Four cells are essentially always used. Six cells are used between half and two-thirds of the times. Four other cells are used sporadically, the rest are never used. This suggests that there is a bulk of around ten cells that is particularly informative for this grid-like structure with splitting ratios specified in the manuscript. Whether this is a generic feature of the Virtual Variance algorithm, or a consequence of the very regular pattern of the chosen grid will be investigated in future research.

References

Antoniou, C., Barceló, J., Breen, M., Bullejos, M., Casas, J., Cipriani, E., Ciuffo, B., Djukic, T., Hoogendoorn, S., Marzano, V., Montero, L., Nigro, M., Perarnau, J., Punzo, V., Toledo, T. and van Lint, H. (2016), 'Towards a generic benchmarking platform for origindestination flows estimation/updating algorithms: Design, demonstration and validation', *Transportation Research Part C: Emerging Technologies* **66**, 79 – 98.

Bilbao-Ubillos, J. (2008), 'The costs of urban congestion: estimation of welfare losses arising from congestion on cross-town link roads', *Transportation Research Part A: Policy and Practice* **42**(8), 1098–1108.

Boyd, S. and Vandenberghe, L. (2004), *Convex Optimization*, Springer.

Canudas de Wit, C., Morbidi, F., Ojeda, L., Kibangou, A., Bellicot, I. and Bellemain, P. (2015), 'Grenoble traffic lab: An experimental platform for advanced traffic monitoring and forecasting [applications of control]', *Control Systems, IEEE* **35**(3), 23–39.

Centre national d'information routiére (CNIR) (2014). Official CNIR website: http://www.bison-fute.gouv.fr/.

Como, G., Lovisari, E. and Savla, K. (2013), Throughput optimal distributed routing in dynamical flow networks, *in* 'Proc. of the 52st IEEE Conference on Decision and Control (CDC'13)', Florence, Italy.

Coogan, S. and Arcak, M. (2014), Dynamical properties of a compartmental model for traffic networks, *in* 'Proc. of the American Control Conference'.

Daganzo, C. F. (1994), 'The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory', *Transportation Research Part B: Methodological* **28B**(4), 269–287.

Daganzo, C. F. (1995), 'The cell transmission model, part II: network traffic', *Transportation Research Part B: Methodological* **29B**(2), 79–93.

Danczyk, A., Di, X. and Liu, H. X. (2016), 'A probabilistic optimization model for allocating freeway sensors', *Transportation Research Part C: Emerging Technologies* **67**, 378 – 398.

Deng, W., Lei, H. and Zhou, X. (2013), 'Traffic state estimation and uncertainty quantification based on heterogeneous data sources: A three detector approach', *Transportation Research Part B: Methodological* **57**, 132157.

Ehlert, A., Bell, M. G. H. and Grosso, S. (2006), 'The optimisation of traffic count locations in road networks', *Transport. Res. B-Meth.* **40**(6), 460 – 479.

Fei, X., Mahmassani, H. S. and Murray-Tuite, P. (2013), 'Vehicular network sensor placement optimization under uncertainty', *Transportation Research Part C: Emerging Technologies* **29**, 14 – 31.

Grant, M. and Boyd, S. (2008), Graph implementations for nonsmooth convex programs, *in* V. Blondel, S. Boyd and H. Kimura, eds, 'Recent Advances in Learning and Control (a tribute to M. Vidyasagar)', Lecture Notes in Control and Information Sciences, Springer.

Grant, M. and Boyd, S. (2013), 'Cvx: Matlab software for disciplined convex programming'. version 2.0 beta, http://cvxr.com/cvx.

Hall, F. L., Allen, B. L. and Gunter, M. A. (1986), 'Empirical analysis of freeway flow-density relationships', *Transport. Res. A-Gen.* **20**(3), 197 – 210.

Hu, S.-R. and Liou, H.-T. (2014), 'A generalized sensor location model for the estimation of network origindestination matrices', *Transport. Res. C-Emer.* **40**(0), 93 – 110.

INRIX (2014). Official INRIX website: http://www.inrix.com/xd-traffic/.

Jabari, S. and Liu, H. X. (2013), 'A stochastic model of traffic flow: Gaussian approximation and estimation', *Transportation Research Part B: Methodological* **47**(0), 15 – 41.

Khalil, H. (2002), *Nonlinear Systems*, 3 edn, Prentice Hall.

Li, Y., Canepa, E. and Claudel, C. (2014), 'Optimal control of scalar conservation laws using linear/quadratic programming: Application to transportation networks', *IEEE Transactions on Control of Network Systems* **1**(1), 28–39.

Lighthill, M. J. and Whitham, G. B. (1955), 'On kinematic waves. ii. a theory of traffic flow on long crowded roads', *Proc. of the Royal Society of London, Series A, Mathematical and Physical Sciences* **229**(1178), 317345.

Lovisari, E., Canudas de Wit, C. and Kibangou, A. Y. (2015), Data fusion algorithms for density reconstruction in transportation networks, *in* 'Proc. of the 54th IEEE Conference on Decision and Control (CDC'14)'.

Lovisari, E., Como, G. and Savla, K. (2014), Stability of monotone dynamical flow networks, *in* 'Proc. of the 53st IEEE Conference on Decision and Control (CDC'14)', pp. 2384–2389.

Morbidi, F., Ojeda, L. L., Canudas de Wit, C. and Bellicot, I. (2014), A new robust approach for highway traffic density estimation, *in* 'Proc. of the 13th European Control Conference, ECC14', Strasbourg, France, pp. 2576–2580.

Muñoz, L., Sun, X., Horowitz, R. and Alvarez, L. (2006), A piecewise-linearized cell transmission model and parameter calibration methodology, *in* 'Proc. of the 85th Annual Meeting of the Transportation Research Board (TRB)', Washington D.C., USA.

Munoz, L., Sun, X., Sun, D., Gomes, G. and Horowitz, R. (2004), Methodological calibration of the cell transmission model, *in* 'Proc. of the American Control Conference, ACC'2004.', Vol. 1, pp. 798–803.

Nantes, A., Ngoduy, D., Bhaskar, A., Miska, M. and Chung, E. (2016), 'Real-time traffic state estimation in urban corridors from heterogeneous data', *Transportation Research Part C: Emerging Technologies* **66**, 99 – 118.

Papageorgiou, M., Ben-Akiva, M., Bottom, J., Bovy, P., Hoogendoorn, S. P., Hounsell, N. B., Kotsialos, A. and McDonald, M. (2007), ITS and traffic management, *in* C. Barnhart and G. Laporte, eds, 'Handbooks in Operations Research and Management Science: Transportation', Elsevier, chapter 11, pp. 715–774.

Papageorgiou, M., Diakaki, C., Dinopoulou, V., Kotsialos, A. and Y., W. (2003), 'Review of road traffic control strategies', *Proc. of the IEEE* **91**(12), 2043–2067.

Papageorgiou, M., Hadj-Salem, H. and Blosseville, J.-M. (1991), 'Alinea: A local feedback control law for on-ramp metering', *Transportation Research Record* pp. 58–64.

Pisarski, D. and Canudas de Wit, C. (2012), Optimal balancing of road traffic density distributions for the cell transmission model, *in* 'Proc. of the 51th IEEE Conference on Decision and Control (CDC'12)', Maui, Hawaii, USA, p. 6969 6974.

Qu, X., Wang, S. and Zhang, J. (2015), 'On the fundamental diagram for freeway traffic: A novel calibration approach for single-regime models', *Transportation Research Part B: Methodological* **73**(0), 91 – 102.

Shahabi, M. and Unnikrishnan, A. (2014), 'Robust hub network design problem', *Transportation Research Part E: Logistics and Transportation Review* **70**(0), 356 – 373.

Treiber, M. and Kersing, A. (2013), *Traffic Flow Dynamics*, Cambridge University Press.

van Lint, J. W. C. and Hoogendoorn, S. P. (2010), 'A robust and efficient method for fusing heterogeneous data from traffic sensors on freeways', *Computer-Aided Civil and Infrastructure Engineering* **25**, 596–612.

Work, D. B., Blandin, S. Tossavainen, O.-P., Piccoli, B. and Bayen, A. M. (2010), 'A traffic model for velocity data assimilation', *Applied Mathematics Research eXpress* **2010**(1), 1 – 35.

Wright, M. and Horowitz, R. (2016), 'Fusing loop and gps probe measurements to estimate freeway density', *IEEE Transactions on Intelligent Transportation Systems* **PP**(99), 1–14.

Xing, T., Zhou, X. and Taylor, J. (2013), 'Designing heterogeneous sensor networks for estimating and predicting path travel time dynamics: An information-theoretic modeling approach', *Transportation Research Part B: Methodological* **57**(2), 66 – 90.

Zhou, X. and List, G. (2010), 'An information-theoretic sensor location model for traffic origin-destination demand estimation applications', *Transportation Science* **44**(2), 254 – 273.
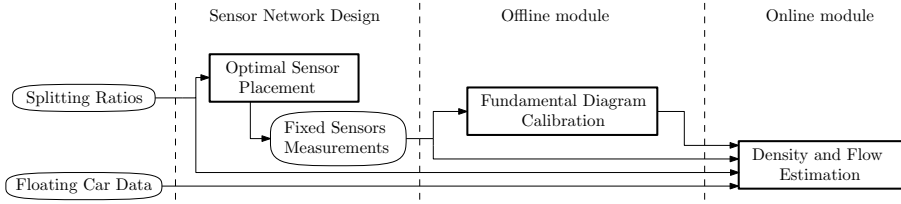
Figure 1: Flowchart of the paper's contributions: Optimal Sensor Placement for sensor network design, Offline Fundamental Diagram Calibration, and Online Density Estimation. Curvy blocks represent information provided to the three procedures.
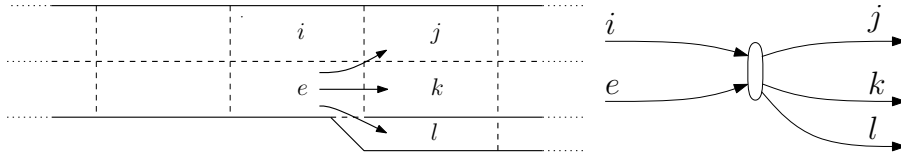

Figure 2: Example of partitioning of a road in cells and of traffic splitting. Vehicles in cell e can continue on the same lane on cell k, change lane on cell j, or exit the road it is travelling on, for example taking an offramp starting on cell l, with splitting ratios R_{ej}, R_{ek} and R_{el}, respectively, satisfying $R_{ej} + R_{ek} + R_{el} = 1$. Similarly for i. The graph topology is captured by a graph where cells correspond to edges and nodes represent the intersection between cells.
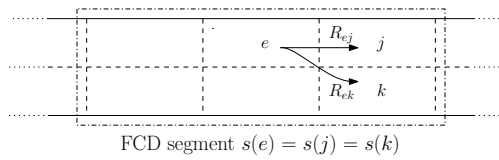

FCD segment $s(e) = s(j) = s(k)$

Figure 3: Roads are partitioned in cells and FCD segments. The depicted FCD segment (in dot-dashed line) includes, among others, the three cells e, j and k, so that $s(e) = s(j) = s(k)$.
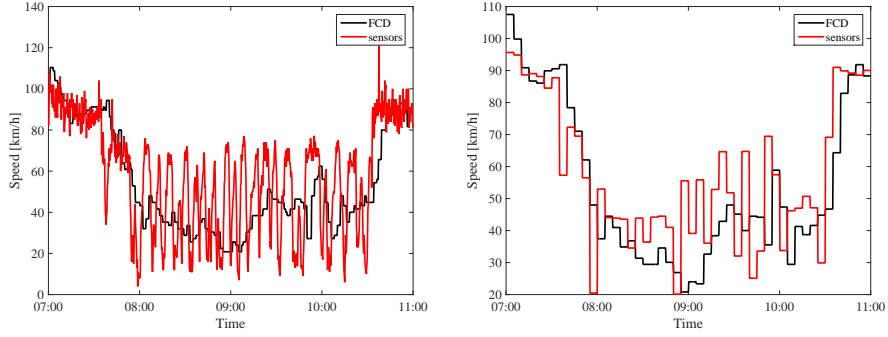
Figure 4: Comparison of speed measurements from sensors and Floating Car Data at location Taillat on Grenoble Rocade Sud, 24/04/2014.
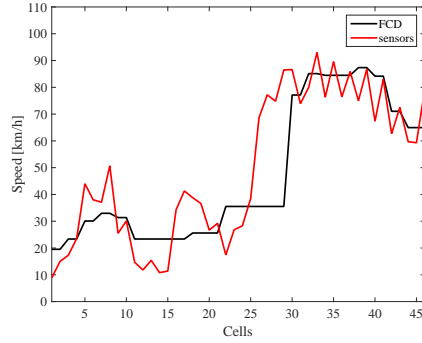

Figure 5: Comparison of speed measurements from sensors and Floating Car Data for all cells on the main line of the Grenoble Rocade Sud at 09:00, 24/04/2014.
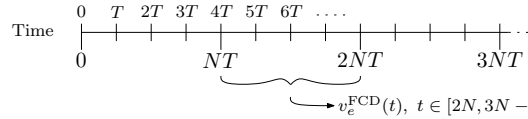

Figure 6: Measurement sampling times on GTL network. A new sensor measurement is available at each cell equipped with sensors every time slot, namely, every $T = 15$ seconds. A new FCD measurement (average speed during the previous N time slot) is available at each cell every N time slots, namely every $NT = 4T = 60$ seconds.

43
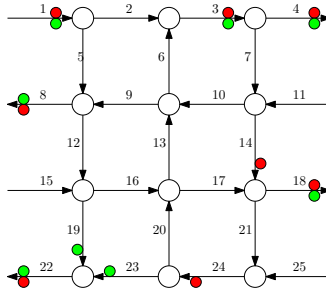
Figure 7: The regular grid network used in the numerical experiment. The green dots correspond to the 8 cells selected by the Virtual Variance algorithm with $\eta = 2$ and $\kappa = 20$. The red dots correspond to cells selected via exhaustive search when the number of possible sensors is 8.
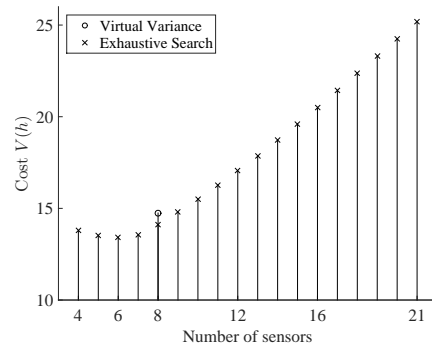

Figure 8: Cost of the best placement via exhaustive search (crosses, minimum possible cost for each number of sensors) and via virtual variance algorithm (circle).
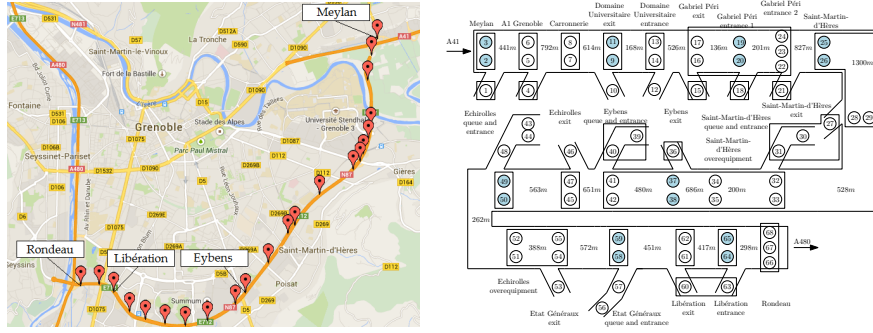
Figure 9: The experimental setting: the town of Grenoble and the Rocade Sud and a stylized version of the freeway. The positions along the main line of the 22 sections of the main line in which sensors have been placed is shown as red pin. The positions of the 68 fixed sensing locations are shown in the stylized map. Each sensing location also corresponds to a cell. Light blue circles denote fixed sensors that are selected by the Virtual Variance algorithm and are used in the implementation of the Density Reconstruction algorithm. Each rectangle represents one FCD segment, often providing average speed measurements over more than one cell.

Figure 10: Stylized representation of the main line part of the Rocade Sud. White ovals represents junctions of the graph. The selected positions of 17 fixed sensors by CNIR are marked by red dots, those of the 16 sensors chosen by the Virtual Variance algorithm by green dots.

45

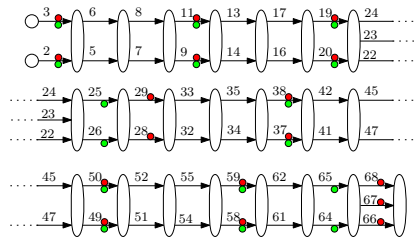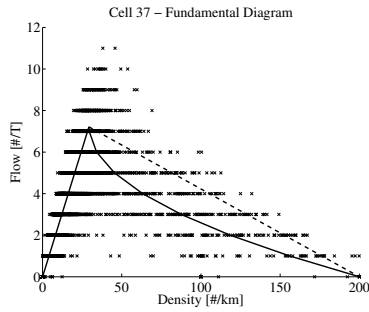Figure 11: Calibration of the Fundamental Diagram on the cell Eybens exit - slow lane. The linear-quadratic Fundamental Diagram, calibrated using data from April 10th, 2014, is shown in thick line. The dashed thick line represents the corresponding linear Fundamental Diagram in congested regime. Each cross is a (flow, density) pair measured on April 24th, 2014, one for each time slot of $T = 15$ seconds during the whole day. Flow is the number of counted vehicles crossing the sensing location, density is the measured density of vehicles, during the time slot.
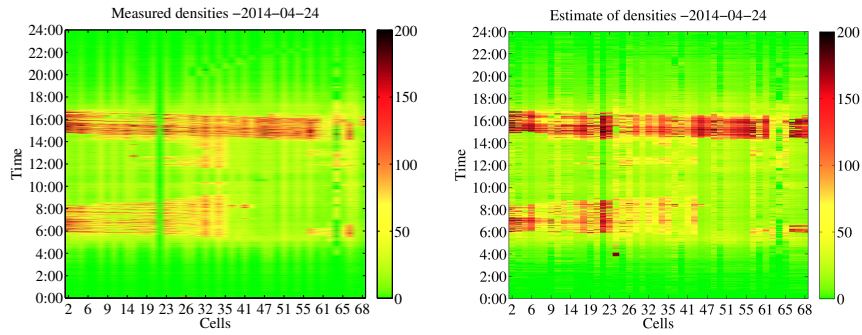
Figure 12: Numerical results of density estimation on all cells: measured densities (left panel) and estimated densities right panel). Day: 24-04-2014

Figure 13: Numerical results of flow estimation on all cells: measured densities (left panel) and estimated densities right panel). Day: 24-04-2014

46

Figure 14: Distribution of cell usage resulting from the virtual variance algorithm applied to 100 different instances of perturbed matrix of splitting ratios.

Graph notation and modeling parameters

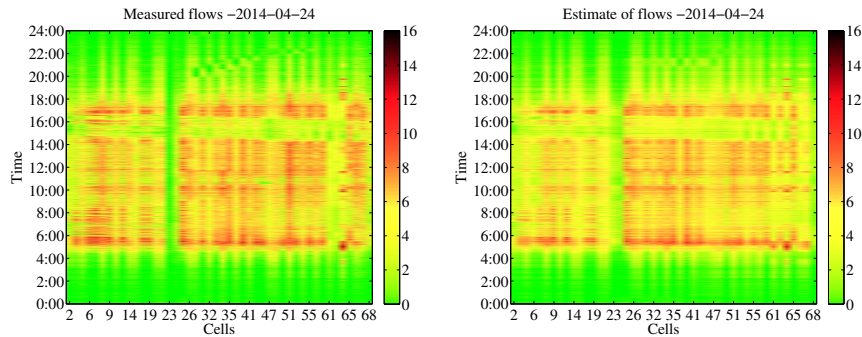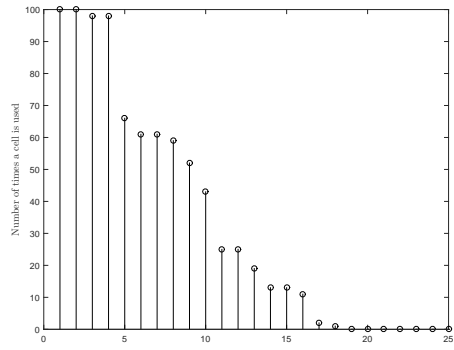|			
$\mathcal{V}$	set of junctions	$\mathcal{E}$	set of cells
$\mathcal{R}^i$	set of onramps	$\mathcal{R}^o$	set of offramps
$\mathcal{E}_e^-$	set of cells preceding cell e	$\mathcal{E}_e^+$	set of cells following cell e
T	sampling time	t	time index
ℓ_e	length of e	ℓ_{ave}	average vehicular length
R_{ek}	splitting ratio from e to k	$\bar{L}$	non-onramp sublaplacian
λ_e	inflow into onramp e	v_e^{limit}	speed limit on e
$d_e(\cdot)$	demand function on e	$s_e(\cdot)$	supply function on e
φ_e	Fundamental Diagram on e	ρ^{jam}	jam density
v_e^{ff}	freeflow speed on e	ρ_e^c	critical density on e
C_e	flow capacity on e	ω_e	wave speed on e

Variables and measurements

ρ_e	density on e	o_e	occupancy on e
f_e^{in}	inflow in e	f_e^{out}	outflow from e
v_e	speed on e	f_e	cumulative outflow from e
φ_e^m	flow measurement on e	ρ_e^m	density measurement on e
v_s^m	speed meas. on segment s	v_e^m	speed measurement on e
$\hat{\rho}_e$	density estimate on e	$\tilde{\rho}_e$	density pseudo-meas. on e
$\hat{f^{\mathrm{in}}}_e$	inflow estimate on e	$\hat{f^{\mathrm{out}}}_e$	outflow estimate on e

Algorithms parameters

Γ	density observer gain	γ	data fitting parameter
$\mathcal{E}^m$	cells equipped with sensors	$V_p(\mathcal{E}^m)$	performance of sensors in $\mathcal{E}^m$
c	cost of a sensor	Σ_{ee}	inverse of Virtual Variance on e
η	total variance parameter	κ	homogeneity parameter

Table 1: List of symbols.

Scenario	η	optimal # sensors	$V_p(\mathcal{E}^m)$	$V(\mathcal{E}^m)$
Fix		17	3.8161	20.8161
Unconstrained	0.2	14	3.6867	17.6867
Unconstrained	1	10	5.1732	15.1732
Constrained, $\# \leq 12$	0.32	12	4.4972	16.4972

Table 2: Results of the four considered scenarios.

Max %(t,e):	$\delta : a^\rho(t,e) \leq \delta$			$\delta : a^\varphi(t,e) \leq \delta$		
	75%	90%	95%	75%	90%	95%
14-04-01 -Tue.	10.4317	19.4079	32.1121	2.4901	3.9824	4.8838
14-04-02 -Wed.	10.058	16.8625	23.4218	2.504	4.0508	4.9457
14-04-03 -Thu.	14.214	39.0677	68.6383	2.401	3.8752	4.765
14-04-04 -Fri.	13.3274	35.6493	67.2034	2.5477	4.0648	4.9543
14-04-05 -Sat.	7.6795	11.6078	14.9678	2.5927	4.138	5.0691
14-04-06 -Sun.	5.6087	7.8254	9.5749	2.1949	3.5373	4.4432
14-04-07 -Mon.	8.8653	14.9656	21.5993	2.2046	3.6666	4.6213
14-04-08 -Tue.	9.4236	16.1212	24.1236	2.134	3.5462	4.4686
14-04-09 -Wed.	10.7057	21.1416	34.5579	2.1496	3.5922	4.5308
14-04-10 -Thu.	10.3578	21.7147	38.2241	2.1529	3.5922	4.5125
14-04-11 -Fri.	10.1152	19.0844	32.2754	1.8778	3.2413	4.1324
14-04-12 -Sat.	5.7334	8.4439	11.1728	2.1883	3.5934	4.5191
14-04-14 -Mon.	9.5789	18.4657	31.2646	2.1349	3.5506	4.5182
14-04-15 -Tue.	11.2095	18.9568	32.7132	2.0675	3.6966	4.6661
14-04-16 -Wed.	12.9045	28.6587	46.8904	2.2881	3.9458	4.9609
14-04-17 -Thu.	17.482	43.2213	62.2287	2.0069	3.5265	4.4383
14-04-18 -Fri.	14.8889	33.0979	48.2054	2.3525	4.0964	5.1424
14-04-19 -Sat.	6.3368	9.8998	12.8105	2.3955	3.9664	4.8847
14-04-20 -Sun.	5.1875	7.1362	8.7217	2.0231	3.2701	4.0776
14-04-21 -Mon.	5.4528	7.433	8.9298	2.1513	3.4238	4.2284
14-04-22 -Tue.	12.0275	26.3377	43.2225	2.2019	3.7727	4.7825
14-04-23 -Wed.	12.9473	32.5637	49.0958	2.0893	3.4793	4.4004
14-04-24 -Thu.	18.222	42.5245	59.2349	2.0391	3.3786	4.282
14-04-25 -Fri.	24.4905	49.1946	65.6698	2.0058	3.3379	4.244
14-04-26 -Sat.	6.0991	8.7496	10.8339	1.922	3.3109	4.1612
14-04-27 -Sun.	5.4655	7.4201	8.93	2.1341	3.3638	4.1301
14-04-28 -Mon.	13.8312	35.9539	57.5468	1.9	3.3551	4.2986
14-04-29 -Tue.	9.7221	17.4268	25.9634	2.2904	3.7966	4.714
14-04-30 -Wed.	12.736	29.0402	47.1253	2.2473	3.7883	4.7363
Average	10.8656	22.3439	34.3882	2.1961	3.6531	4.5694

Table 3: Quantitative measurement of the performance of the proposed algorithm. Maximum magnitude of the absolute error on densities (left) and flows (right) on 75%, 90% and 95% of the pairs (cell, time) between 07:00 and 19:00.

Max %(t,e):	$\delta : a^\rho(t,e) \leq \delta$			$\delta : a^\varphi(t,e) \leq \delta$		
	75%	90%	95%	75%	90%	95%
14-04-01 -Tue.	7.7399	14.6455	23.9397	1.6519	2.8907	3.6042
14-04-02 -Wed.	7.6856	13.4467	19.397	1.6426	2.9	3.7202
14-04-03 -Thu.	9.6351	24.9891	46.7688	1.6201	2.8768	3.5025
14-04-04 -Fri.	8.9069	23.5012	46.4841	1.6842	2.9369	3.6128
14-04-05 -Sat.	5.7186	9.5516	12.4627	1.7893	3.0004	3.6258
14-04-06 -Sun.	4.2724	6.0777	8.0227	1.6371	2.3554	2.9572
14-04-07 -Mon.	6.204	11.438	16.8194	1.2662	1.9248	2.5075
14-04-08 -Tue.	6.6256	12.2475	18.7567	1.1452	1.78	2.2611
14-04-09 -Wed.	7.5039	15.2545	27.0917	1.2161	1.8343	2.3119
14-04-10 -Thu.	7.0607	14.528	28.1098	1.14	1.808	2.3011
14-04-11 -Fri.	6.5122	14.6898	26.9997	1.192	1.8202	2.4124
14-04-12 -Sat.	4.1982	6.0592	8.0033	1.3212	1.9123	2.3261
14-04-14 -Mon.	6.6178	12.9569	23.4577	1.2233	1.8117	2.2992
14-04-15 -Tue.	8.2876	15.1747	26.9744	1.3578	2.2702	3.9263
14-04-16 -Wed.	9.1428	21.1292	36.7874	1.325	2.317	3.6409
14-04-17 -Thu.	11.6003	32.2321	50.8533	1.3217	2.1836	3.4486
14-04-18 -Fri.	10.3143	25.1777	38.3014	1.2932	2.3882	4.1402
14-04-19 -Sat.	4.6936	7.5892	11.6674	1.5063	2.2413	3.4459
14-04-20 -Sun.	4.0264	5.4524	6.7887	1.4632	1.969	2.4225
14-04-21 -Mon.	4.2807	5.822	7.1171	1.5684	2.0967	2.6073
14-04-22 -Tue.	8.4038	19.0648	35.9572	1.2548	1.9906	2.8319
14-04-23 -Wed.	8.3312	23.6232	40.2403	1.1912	1.8248	2.3789
14-04-24 -Thu.	10.1973	31.1337	45.6909	1.1822	1.8056	2.3106
14-04-25 -Fri.	14.1	37.9687	49.7985	1.1321	1.7656	2.3675
14-04-26 -Sat.	4.2046	6.0036	7.7334	1.3413	1.891	2.2467
14-04-27 -Sun.	4.3004	5.6931	6.7654	1.5965	2.0499	2.3881
14-04-28 -Mon.	9.0015	23.7898	43.682	1.2556	1.9867	2.834
14-04-29 -Tue.	6.8045	13.6265	20.584	1.2544	1.9524	2.7057
14-04-30 -Wed.	8.5286	21.3739	37.2919	1.2532	1.9506	2.7559
Average	7.4103	16.3531	26.6395	1.3754	2.1564	2.8929

Table 4: Quantitative measurement of the performance of the proposed algorithm. Maximum magnitude of the absolute error on densities (left) and flows (right) on 75%, 90% and 95% of the pairs (cell, time) between 07:00 and 19:00 downsampled and averaged at a rate of one datum every 5 minutes.