Estimation of fundamental diagrams in large-scale traffic networks with scarce sensor measurements

Olga Lucía Quintero Montoya, Carlos Canudas de Wit

Estimation of fundamental diagrams in large-scale traffic networks with scarce sensor measurements[*]

O. L. Quintero Montoya[1] and Carlos Canudas-de-Wit[2] [‡]

September 14, 2018

Abstract

The macroscopic fundamental diagram (MFD) relates space–mean flow density and the speed of an entire network. We present a method for the estimation of a "normalized" MFD with the goal to compute specific Fundamental Diagram in places where loop sensors data is no available. The methodology allows using some data from different points in the city and possibly combining several kinds of information. To this aim, we tackle at least three major concerns: the data dispersion, the sparsity of the data, and the role of the link (with data) within the network. To preserve the information we decided to treat it as two-dimensional signals (images), so we based our estimation algorithm on image analysis, preserving data veracity until the last steps (instead of first matching curves that induce a first approximation). Then we use image classification and filtering tools for merging of main features and scaling. Finally, just the Floating Car Data (FCD) is used to map back the general form to the specific road where sensors are missing. We obtained a representation of the street by means of its likelihood with other links within the same network.

1 Introduction

Estimation problems in large-scale networks with scarce sensors sets have been studied during the last decade in order to obtain accurate models of the urban traffic networks for simulation and control purposes. At the link level, fundamental diagrams (FDs) describe the relationship between flow, density, and

[†1]Mathematical Modeling research Group at Universidad EAFIT, Medellin, Colombia, oquinte1(at)eafit.edu.co.

[‡2] CNRS, GIPSA-Lab, Grenoble, France. carlos.canudas-de-wit(at)gipsa-lab.fr

speed. The FD relating flow and density is a key element of the traffic models. Although FD can be easily estimated in simple linear roads by collecting flow and velocity in a single probe station, there is no available theory, in how to estimate the FD in a large traffic network when the numbers of sensor stations is scarce (much less than the number of roads where the FD needs to be reconstructed). Alternatives are: to interpolate the FD from measures to roads where sensors are not present or to average the density/flow relation in a some network area. The latter leads to the so called macroscopic fundamental diagram (MFD) which relates the space-mean flow, the density, and the speed of an entire network see [1, 2, 3]. In some cases, it is also possible to fuse data from different sources to make such approaches feasible and numerically efficient.

There is not much theoretical support other than experimental evidences that the Fundamental Diagram (FD) also holds in 2D spaces. A particular difficulty is how to include traffic lights and turns ratios together with the geometry of the network in the construction of the MFD. Even experimentally, there are still many open issues involved when it comes to define and to measure the MFD [3].

The *Aggregated Traffic Relationship (ATR)* is a relationship between average traffic variables (like speed, density, occupancy or flow) for a specific network that do not necessarily have a well defined shape that can be called an MFD. In particular, networks with an uneven and inconsistent distribution of congestion may exhibit traffic states that are well below the upper bound of an ATR and much too scattered to lie along an MFD. For urban networks, some works derives some upper envelopes relating average flow and average density, and then suggest conditions under which the bound should be tight; i.e., under which the bound is an approximate MFD [4]. The connection between network structure and a network's MFD for urban neighborhoods is strongly influenced at least in part by traffic signals. The problem then can be associated to the scarcity of data and the natural non-ideal situations in non-ideal networks. Consequently, the lack of simplicity [5] represents a challenge for the proper understanding and generalization of a model.

Recent findings show that the MFD shape of a heterogeneous network includes many scatter points when the transportation system becomes congested, which brings uncertainties and inaccuracies to the traffic state estimation [6]. Experimental results reported in the literature mostly used data from fixed sensors, such as inductive loops to estimate the MFD, which can cause biased estimation in urban networks. To overcome this limitation, recent literature reported on the use of trajectory data obtained from probe vehicles. However, these studies were conducted with simulated data sets; few works have discussed the limitations of real data sets and their impact on MFD estimation. In [7], the authors compare two methods for estimating traffic state variables of signalized arterial sections: a method based on cumulative vehicle counts (CUPRITE) and one based on vehicle trajectory from taxi GPS logs.

The current trajectory data have limitations in quantity (i.e., the penetration rate), because of which the traffic state variables tend to be underestimated.

Another issues is when data come from different kinds of source. For in-

stance, in [8] the authors proposed methods to estimate and calibrate the FD by using inductive loops, Floating Car Data (FCD), wireless manetometers, and average velocity from FCD sources. Experiments were reported by the implementation of the proposed solution using real fixed sensor measurements from the Grenoble Traffic Lab [9].

A methodology for estimating the MFD based on both data sources simultaneously is the Data fusion algorithm for macroscopic fundamental diagram estimation [10]. To that end, authors defined a fusion algorithm that separates the urban network into two sub- networks, one with loop detectors (LDD) and one without loops with FCD. The LDD and the FCD are then fused taking into account the accuracy and network coverage of each data type. Related to the use of floating car data for estimation purposes, the work of [11] introduced a novel framework to estimate travel times, traversed paths, and missing values over a large-scale road network using sparse GPS traces. No MFD aproximation to the problem was performed. In [6] a partitioning algorithm is proposed as a way to solve the problem of missing or low quality traffic data for MFD construction with control purposes.

Some authors have been working on the discretized partial differential equation of the macroscopic model, solved numerically with low computational effort. From that perspective the spatial and temporal resolution of the discretization are tuning parameters to balance computational effort with model accuracy [5].

The ultimate goal is a universal recipe to estimate FD that can be used for all type of networks. However, recognizing that networks are complex structures described by many variables [12], we shall be satisfied with an approximation that uses as few of these variables as possible; a proper parameterization and description of heterogeneity that would yield good estimates for most practical applications.

Such a consideration leads to the core idea pursued here, which is to seek for a general methodology to build first a "*Normalized Macroscopic Fundamental Diagram*" (NMFD) representative of the whole network using data from several collection points. Then, the NFD will be used as a normalized structure to retrieve the Fundamental Diagram of all other links within the network using a limited amount of information collected from the network geometry and available physical information (speed limits, number of lines, etc.).

The proposed strategy is based on the determination of the similarity between some links within the network from their measured data; later on the understanding of this similarity is used to build a representation of a general behavior. In that sense, the NMFD arises by using the reconstructed relevant information from measured data, via clustering. Our method is based on mapping the original data into two-dimensional signals. Image classifications preserve data veracity until the last steps instead of first matching curves that induce a first approximation. By means of data treated as images, their classification and processing allows then to use just FCD to map back the cluster family towards the links/roads that have missing sensors. This paper is organized as follows: section 2 addresses the problem formulation, introduces the proposed method and the selected case study; section 3 explains the data transformation to im-

age domain and the processing strategies. Section 4 introduces the clustering by image similarity concepts, section 5 presents the merging strategy, the α-scaling and Normalized Macroscopic Fundamental Diagram and finally section 6 the reconstruction of the Fundamental Diagram of the street from floating car and validation results. Concluding remarks will be presented in section 7.

2 Problem formulation, proposed method and selected study-case

2.1 Problem formulation MFD and FD

From any given traffic network we assume that there is a number of roads (links) with available measurements, and there exist an important number of other links without sensing capabilities. Besides, as a direct measurement of density is not always available, so the relationship between densities and flows must be in some cases obtained indirectly from occupancy. Similar issues arise concerning the velocities in each link because the lack of instrumentation of main large scale networks, nevertheless there are FCD from which an average of speeds for streets can be obtained. It is well known that flow Φ can be expressed as:

$$\Phi(\bar{v}, \bar{w}, \rho) = \{\bar{v}\rho, \forall \rho \leq \rho_c\} \tag{1}$$

and

$$\Phi(\bar{v}, \bar{w}, \rho) = \{\bar{w}(\rho - \rho_{max}), \forall \rho > \rho_c\} \tag{2}$$

where $\bar{v}$ and $\bar{w}$ are averages of free-flow speed and backward (congestion) wave speed of the MFD, respectively and ρ is the density of traffic. The maximal density is ρ_{max}. The above equation describes the simplest triangular form of the Fundamental Diagram characterized by the three parameters: $\bar{v}$, $\bar{w}$, and ρ_{max}. The value of critical density ρ_c at Φ_{max} can be assumed to be estimated from:

$$\bar{v}\rho_c = \bar{w}(\rho_c - \rho_{max}), \forall \bar{v}, \bar{w} > 0 \tag{3}$$

It is straightforward to obtain

$$\rho_c = \frac{\bar{w}\rho_{max}}{\bar{v} + \bar{w}} \tag{4}$$

and consequently,

$$\Phi_{max} = \bar{v}\rho_c = \frac{\bar{v}\bar{w}\rho_{max}}{\bar{v} + \bar{w}} \tag{5}$$

Since the maximal density, ρ_{max} can be properly estimated from the geometry of the roads, the estimation of the FD boils down to the identification of the two remaining parameters $\bar{v}$, $\bar{w}$. When road sensor are available (as magnetic loops), such parameters can be easily estimated by the collection of flow and

speed (eventually also using the occupancy). The question under study here is how to reconstruct the FD in places where loop detectors are not available, but eventually average speed can be sensed via FCD.

2.2   4-steps method

What we propose is to identify patterns from raw data by means of the similarity between links on the network and characterize them in order to obtain a representative behavior. Later, a normalized representation can lead to a general space in which the network topology properties can be represented; we called this "*Normalized Macroscopic Fundamental Diagram*" (NMFD). A NMFD is proposed as general explanation of the entire network.

Our method is based on the transformation of data into a two dimensional signal and then will be treated as an image. Image analysis will preserve data veracity until the last steps instead of first matching curves that induce a first approximation.

We use image resizing, and filtering; later groups are merged based on similarity measurement and we introduce an α-scaling procedure to map from (*occupancy*, Φ) space to (ρ, Φ) space to define the NMFD. Afterwards we assume the FCD can be used to help mapping back the cluster family towards the links/roads that have missing sensors.

The main steps are summarized as follows:

1. Data processing through an image based procedure: transformation of raw data to a rounded matrix, image construction, image resizing and image filtering for outliers removal.

2. Strategy of information retrieval via spatial distribution of similarity within the network links based on their attributes (capacity of the street, name, number of lines) and Groups merging.

3. Normalized Macroscopic Fundamental Diagram NMFD construction: Alpha scaling to density, flows space (ρ, Φ) and normalization procedure that allows for the formulation of new functions useful to be mapped to an specific link.

4. Street Fundamental Diagram SFD definition from Floating Car Data Velocity under NMFD conditions

2.3   Study-case

Our experimental setting is the Grenoble Traffic Lab (GTL) [9, 13], a network of sensors deployed for monitoring and research purposes along the "Rocade Sud", a 12 km long freeway encircling the town of Grenoble, France. The network consists of 135 magnetometers buried in the asphalt on both lanes of the main line and on each onramp and offramp, totalling 68 sensing locations. On each sensing location and every 15 seconds, the system provides flow, occupancy

and speed measurements. For further details see [13, 14, 15] and see Figure 1. We selected three sets of sensors specially located from the highway A31
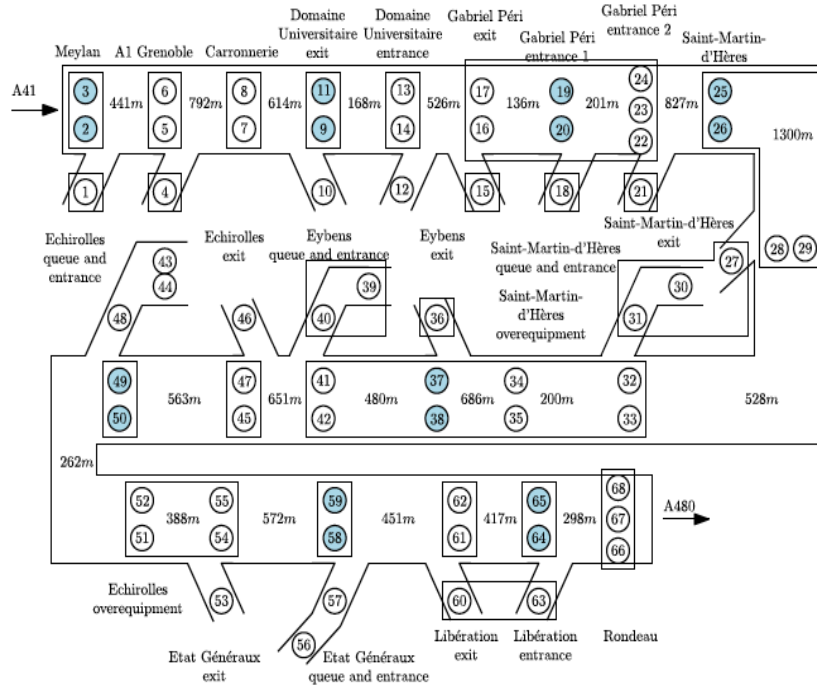

Figure 1: "Rocade Sud", a 12 km long freeway encircling the town of Grenoble, France

(north-west) to A480 (south) in Grenoble France. The criteria for selection of the measurement points was to analyze the behavior of MFD for fast, slow and third lanes and the effect of exits and entrances (on ramp and off ramp) in between them. In Figure 2 it can be seen the kind of real data retrieved from sensors without processing stages, illustrating the type of available data which looks more like an ATR than a MFD. In the left graph, it can be seen that the approximation of density vs. measured flows (ρ, Φ) is not suitable to characterize the entire network; illustrating the different profiles obtained for estimating density. Figure in the right side depicts the relationship obtained from plotting raw data of occupancy and flows (occ, Φ) which looks promising for our analysis.

We extracted a heterogeneous *calibration* set for the development of our proposal. As depicted in Figure 3, it can be observed that fast line dataset a (in black) does not behave equally to fast dataset b (blue) and c (cyan). Also the so called slow lanes are not necesarily fully overlapped. On the other hand, the

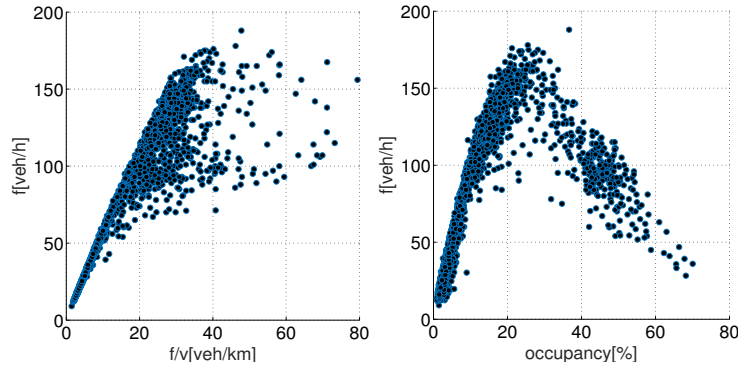Figure 2: Real raw data visualization. From the left, it can be seen that not necesarily preserves the idealized MFD type for the approximation of density (ρ, Φ). The right figure depicts the relationship obtained from plotting raw data of occupancy and flows (occ, Φ)

test dataset is used to see if our proposed NMFD can sucessfully represent the network. In Figure 4 we can observe the test dataset for fast (a and b) and slow lanes (c and d). A simple inspection of the data can lead to several questions about the scarcity of the measurements, also regarding the spread flows for the fast line a on the calibration dataset. This illustrates real case data and allows the reader to understand the complexity of the network. It is obvious that the construction of the MFD is not a straightforward task. The next section will introduce the detailed methodology for the proper treatment of this type of data.
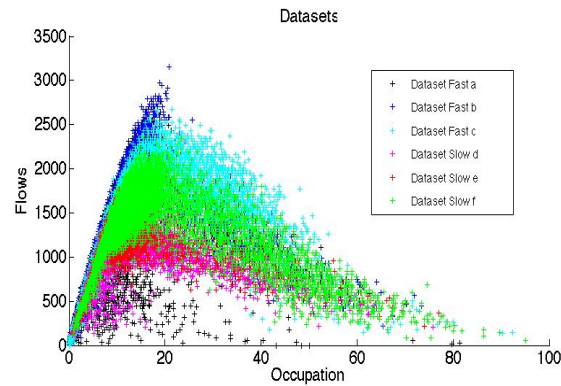

Figure 3: Calibration data plot from sensors. We have three different stations with two types of lines (slow and fast)
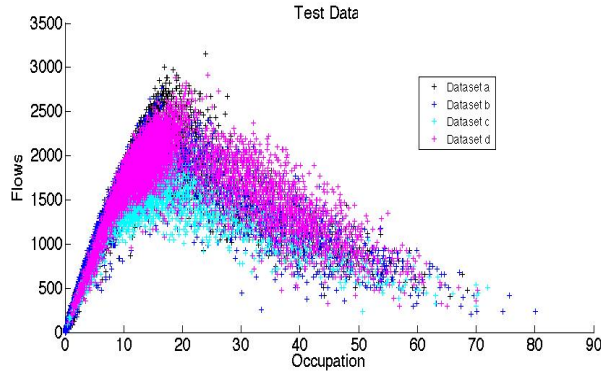
Figure 4: Test data from the "Rocade Sud" [13, 14, 15]

3 Image transformation, filtering and resizing

3.1 Raw data transformation into images

Raw data for a link i usually has the form of the pairs occupancy and flows $(occ, \Phi)_t$ for each sample time, $D_i = [(occ_i, \Phi_i)_1, (occ_i, \Phi_i)_2, \ldots, (occ_i, \Phi_i)_m]$, being m the number of measurements. The transformation allows to map from D to a new three dimensional signal in (x, y) being x integer values of occupancy and y integer values of flows, with values z proportional to the number of data points repetition. To this aim, we processed the raw sensors data by rounding the sensor values for their posterior matrix localization. Each pixel represents directly the posicion of the data point from the ρ, Φ space and its intensity provides a data-frequency based construction of a new space. Here the effects of sample time over the MFD shape dissapear. As example, we can compare Figure 2 with Figure 5; the calibration dataset can be seen as an image. The benefit of using image based solution over statistical models relies on the indirect incorporation of the measure frequency in image construction instead of first matching curves that induce a first approximation.

3.2 Image segmentation for outliers cancellation

Images analysis will preserve data veracity until the last steps. Image segmentation is the process of partitioning an image into parts or regions. This division into parts is often based on the characteristics of the pixels in the image. For example, one way to find regions in an image is to look for abrupt discontinuities in pixel values, which typically indicate edges. These edges can define regions. Another method is to divide the image into regions based on color values. We remove outliers by means of simple segmentation of the image. The segmentation of the images can be performed with No threshold and Multithreshold application based on the Otsu method [16].
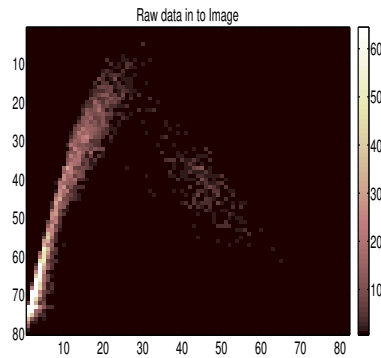
Figure 5: Example of raw data transformed into image, intensity of the image allows to reach the sense of the measurement repeteability that leads to the representability of the undelying information

In Figure 6 we show a comparison between the calibration datasets processed without threshold and multithreshold method.

Compared with the previous set of figures 3, 4, and 5 several new features appear.

Outliers removal rate is proportional to parameters in Otsu Method. The template window used was the default by Matlab implementation. The location of central points as main characteristics of the datasets will allow for the identification of the clusters where the average speeds required for a MFD can be found.
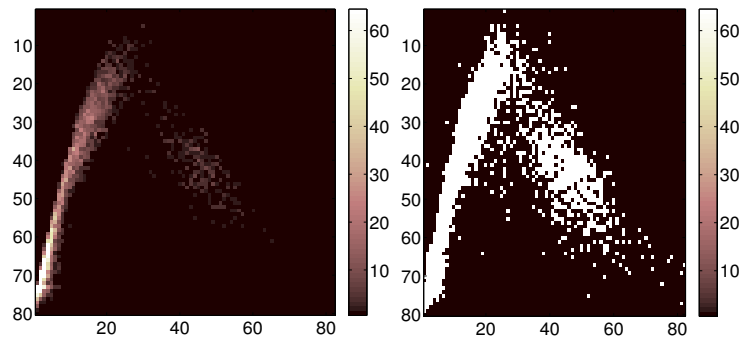

Figure 6: Calibration matrix data from Figure 5 without threshold (left) and multithreshold (right)

3.3 Image Resizing

The goal of image resizing is to build a comparable set of images. It is a typical procedure in image processing and provides a good framework for pattern recognition.

While the final arbiter of image quality is the human viewer, efforts have been made to create objective measures of quality. This can be useful for many applications. Many objective measures of quality require the existence of a distortion-free copy of an image, called the reference image, that can be used for comparison with the image whose quality is to be measured.

The dimensions of the reference image matrix and the dimensions of the degraded image matrix must be identical. An implicit procedure is also the grayscale transformation of the matrix. In Figure 7 we show a comparison between the measured datasets processed without threshold. Also Figure 8 presents the measured data resized and processed with the multithreshold method. Square matrixes of 100x100 pixels are obtained even if the raw data showed differences in maximum values for occupancy.
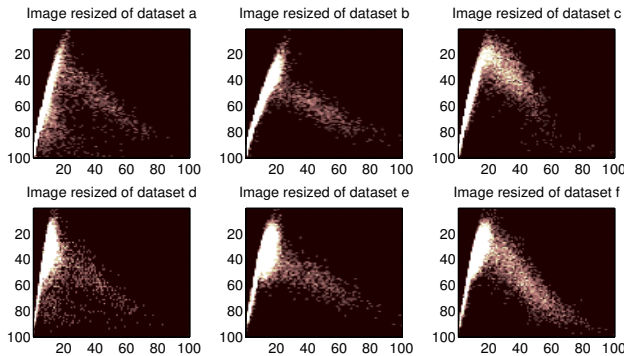

Figure 7: Processed calibration data resized without threshold in section 3.2. No filtering allows the data dispersion patterns to be evident.

4 Merging by image similarity

4.1 Similarity measures

For Similarity measurement between images we selected the IMage Euclidean Distance (IMED) [17]. Unlike the traditional Euclidean distance, IMED takes into account the spatial relationships of pixels. Therefore, it is robust to small perturbation of images. We argue that IMED is the only intuitively reasonable Euclidean distance for images. IMED is then applied to image recognition. The embedding is rather efficient by involving a transformation referred to as the Standardizing Transform (ST) [17] demonstrated that ST is a
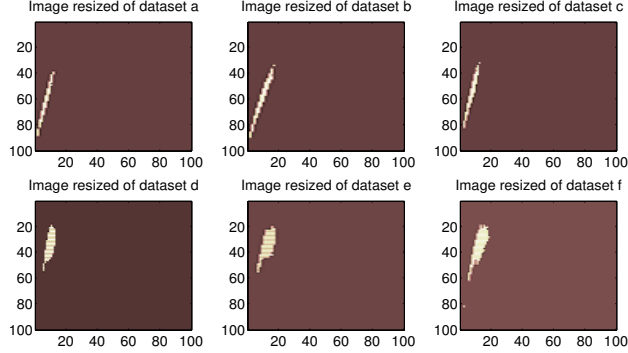
Figure 8: Processed calibration data resized with multitreshold in section 3.2. Only the relevant data points remain.
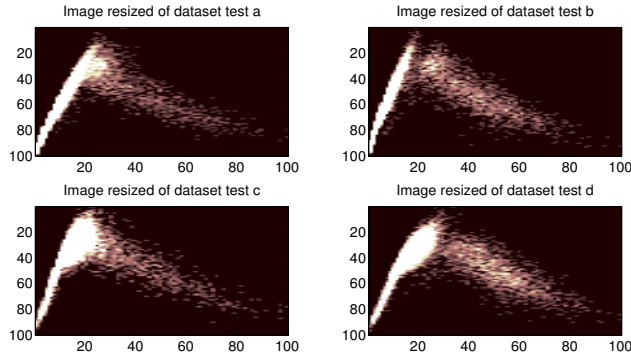

Figure 9: Test data resized and processed without threshold in section 3.2

domain-smoothing transform. Generally, we call a Euclidean distance IMage Euclidean Distance (IMED) if the metric satisfies three conditions which leads to appealing properties. A Euclidean distance $d(x,y) = \left[(x-y)^T G(x-y)\right]^{1/2}$, $G = (g_{ij})_{MN x MN}$ is said to be an IMED if three conditions are satisfied [17].

The following conditions are satisfied:

1. The metric coefficient g_{ij} depends on the distance between pixels P_i and P_j. Let f represent this dependency:

$$g_{ij} = f(|P_i - P_j|), \forall i,j = 1,2,\ldots,MN \tag{6}$$

2. f is continuous and g_{ij} decreases monotonically as $|P_i - P_j|$ increases

3. The functional dependency f is an universal function. That is, it is not restricted for images of a particular size or resolution.
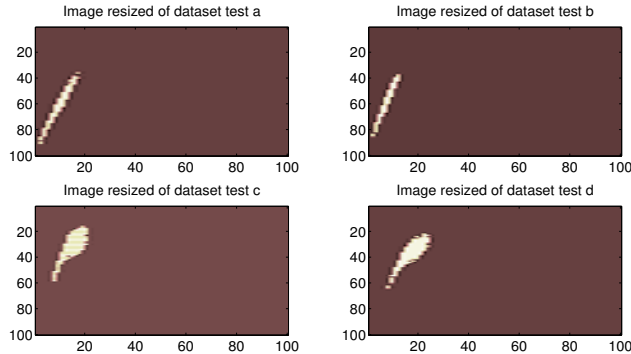
Figure 10: Test data resized and processed by multitreshold in section 3.2

The results of the IMED calculated between the calibration datasets (fast and slow lanes) in order to establish the similarity extracted from sensors are presented in Figures 11 and 12 for no threshold and multithreshold applications respectively. It can be seen that the use of the thresholding method has consequences on the IMED obtained. It will be expected that the IMED will be lower for similar road networks. A merging process is carried out on the images with high similarity indicating they "presumably" belong to the same type of road within the network and can be characterized by the same Representative Macroscopic Fundamental Diagram. From Figure 11 can be seen that within the Fast subset there is less IMED distance between the sets a and b than the a and c or b and c. Similarly, for the subset of slow lanes the d and e datasets exhibit to be more close to each other. So the merging process must be carried out between the closest datasets to build a common representation obtaining as many representations as natural merged datasets appear. The processes carried out on the raw data allows for the determination of clusters that seems to be relevant for the dynamic identification of the roads.

5 Clustering, α-scaling and normalization

5.1 Clustering process

The purpose of clustering is to identify natural groupings of data from a large data set to produce a concise representation of a system's behavior. Simple procedure allows to the use of the representative data, in the sense of repeatability, to build a Normalized Macroscopic Fundamental Diagram NMFD. As an example, simple k-means clustering will retrieve the point for the location of the velocities in free flow v and backward wave speed w of the merged sets in the classic MFD triangular approach. The decision to choose an appropriate point to calculate the main features of the MFD is now based on the data itself and not from the designer *a priori* selection. We can expect to have two clusters,
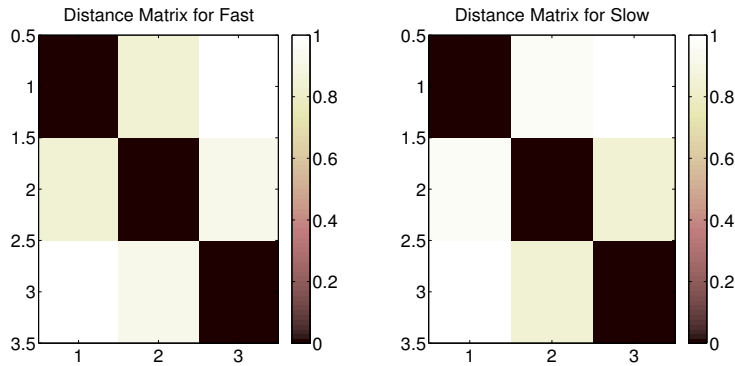
Figure 11: Distances from development dataset measured without thresholding in section 3.2 axis indicates the dataset a, b and c for fast lines
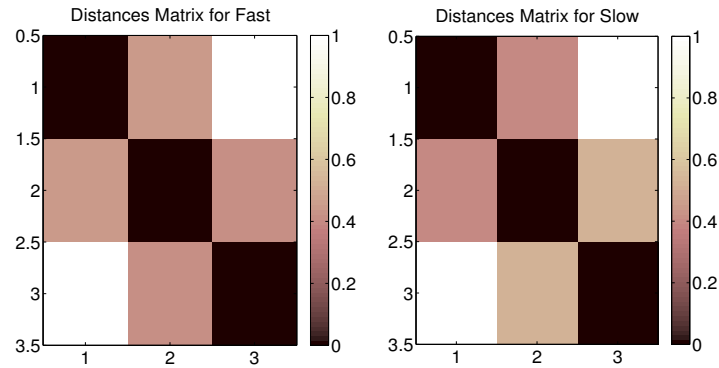

Figure 12: Distances from development dataset processed with multithreshold in section 3.2 axis indicates the dataset d, e and f for slow lines

one for each regime (free flow and congested). As clustering is an exploratory analysis tool, it is convenient to perform several replications. Five replications allowed us to reach our results. k-means++ algorithm was used for initialize the seeds. k-means++ algorithm uses an heuristic to find centroid seeds for k-means clustering [18]. k-means++ improves the running time of Lloyd's algorithm, and the quality of the final solution [18]. While the distance function used in the cluster was the euclidean distance. IMED as cluster metric will be explored in future works.

5.2 Transformation from (*occupancy,* Φ) to (ρ, Φ) via α-scaling

The procedure in section 5.1 allows for the location of the main features to be drawn in the percentage of occupancy $occ(\%)$ vs flow Φ space. We need to establish a scaling α parameter to map the real data obtained from sensors to the density ρ vs Φ space. So assuming the free flow speed v_{ff} as the average speed v_0 before reaching the maximum flux Φ_{max} at 10 % of the occupancy rate, we will map the maximum occupancy percentage to the maximum density ρ_{max} to build the Normalized Macroscopic Fundamental Diagram NMFD. Our density mapping starts with the relationship:

$$\Phi = v * occ \tag{7}$$

After a few calculations, it will be obvious that

$$v = \frac{\|\Phi\|}{\Phi * occ} \tag{8}$$

We propose to find a scaling factor α that relates both speeds

$$\alpha = \frac{v}{v_{ff}} = \frac{\frac{\|\Phi\|}{\Phi * occ}}{v_{ff}} \tag{9}$$

in terms of flows Φ and occupancy occ, and if we consider a simple relationship as $\rho = \alpha * occ$, we can express the density as,

$$\rho = \frac{\frac{\|\Phi\|}{\Phi * occ}}{v_{ff}} * occ \tag{10}$$

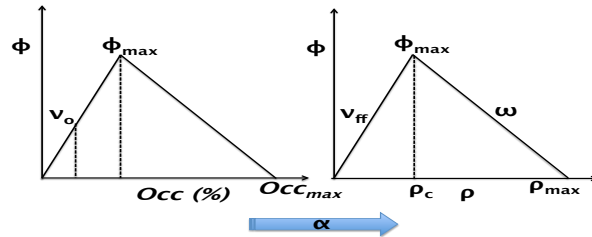Figure 13 shows the simple procedure for the relationships established and the proposed scaling.

Figure 13: Procedure for scaling from (*occupancy,* Φ) to (*density,* Φ) via α-scaling

5.3 Macroscopic Fundamental Diagram Normalization

Once the data is scaled to the density ρ vs Φ space, we can normalize the Representative Macroscopic Fundamental Diagram to finally obtain the desired relationship. Our MFD is considered well defined if the free flow speed v_{ff}, the backward wave speed w and the maximum for Φ are clear. We want to have a spatial distribution of vehicle density and derive an analytical description of the spatial distribution of congestion. Now, normalization of the entire MFD must be performed under ρ_{max} as follows: First we define,

$$\rho_{norm} = \frac{\rho}{\rho_{max}}, \Phi_{norm} = \frac{\Phi}{\Phi_{max}} \tag{11}$$

Then $\rho_{norm} \in [0,1]$ and $\Phi_{norm} \in [0,1]$. Normalization of $\bar{v}$ and $\bar{w}$ take us to the following expressions:

$$\bar{v}_{norm} = \frac{\bar{v}\rho_{max}}{\Phi_{max}} = \mathcal{A}\bar{v}, \forall \rho_{norm} <= \bar{\rho}_{c_{norm}} \tag{12}$$

with $\mathcal{A} = \frac{\rho_{max}}{\Phi_{max}}$, and analogously for $\bar{w}_{norm}$:

$$\bar{w}_{norm} = \frac{\bar{v}+\bar{w}}{\bar{v}} = \beta(\bar{v},\bar{w}), \forall \rho_{norm} > \bar{\rho}_{c_{norm}} \tag{13}$$

Then our Normalized Macroscopic Fundamental Diagram is complete as shown in Figure 14

$$\Phi_{norm} = \{\mathcal{A}\bar{v}_{norm}\rho_{norm}, \forall \rho_{norm} <= \bar{\rho}_{c_{norm}}\} \tag{14}$$

and

$$\Phi_{norm} = \{\beta(\rho_{norm} - 1), \forall \rho_{norm} > \bar{\rho}_{c_{norm}}\} \tag{15}$$
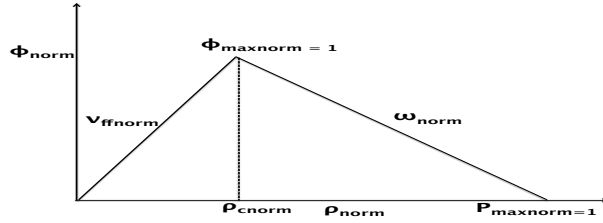

Figure 14: Normalized Macroscopic Fundamental Diagram

6 Reconstruction Mapping

6.1 Using Floating car velocity for FD reconstruction

Once the Normalized Macroscopic Fundamental Diagram is well defined, one can to use floating car data to obtain the Fundamental Diagram of the street

[19]. This way floating car data is used to map back the cluster family towards the links/roads that have missing sensors. Assuming that for the $j - th$ link there is available the speed v_j (or using the previous notation, $\bar{v}_j$), one way can be to obtain the values for $\bar{w}_j$. To this aim one can use the equations (13) and (15) and solve for $\bar{w}_j$:

$$\bar{w}_j = \bar{v}_j \bar{w}_{norm} - \bar{v}_j \qquad (16)$$

and with values for $\bar{v}_j$ and $\bar{w}_j$ obtain:

$$\bar{\rho_{cj}} = \frac{\bar{w}_j \rho_{max}}{\bar{v}_j + \bar{w}_j} \qquad (17)$$

obtaining maximum flow Φ_{maxj} is then straightforward.

6.2 Validation of the results

A validation procedure was carried out for the test data. The natural procedure is to follow the proposed methodology on the calibration data and then, use the test data as input to be processed, measured and IMED distances calculated. Merging process was tested to figure it out if the test data was likely representative of its group. On the other hand, the FCD on the test point was used to find the FD of the street from the NMFD. Test data was collected from sensor located in positions 25, 26 (Fast and slow lanes) in *Saint Martin dHeres*, 51 and 52 (slow and fast lanes) in *Echirolles* from Figure 1, respectively. The difference between them is that the former is located rigth after an entrance to the *Rocade* and the latter rigth before an exit to the city. Even if they can be classified as the same kind of line, the data exhibits a different behavior. Our procedure allows to represent the previous presented test dataset by means of processing the calibration data located in 9, 11, 36, 38, 60 and 62 sensors location from Figure 1. *Domaine Universitaire* sensors are located before an exit (9 and 11), *Eybens* exit is represented by sensor 36 while the fast line is 38 and finally *Liberation* exit sensors are 62 and 60 in Fast and exit lines respectively in [13, 14, 15].
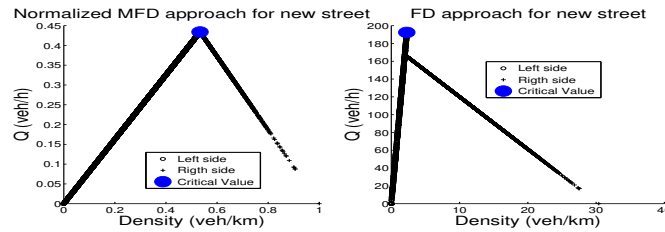

Figure 15: Fundamental diagram for Street at fast lines from calibration data in *Domaine Universitaire*, *Eybens* and *Liberation*

Results depicted in Figure 15 allows to see the Normalized Fundamental Diagram obtained from calibration data. Also the Fundamental Diagram obtained

for one line in test dataset. In Figure 16 a very similar behavior of the FD obtained from Normalized Macroscopic Fundamental Diagram and real measurements is exhibited for fast lines i.e *Saint Martin dHeres* and *Echirolles*. It can be noticed that the Normalized Macroscopic Fundamental Diagram (NMFD) was obtained from the calibration dataset with diversity of information. Also the Fundamental Diagram (FD) of the street was succesfully obtained.
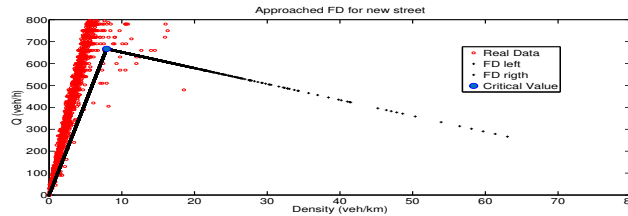

Figure 16: Real data for test processed in fast lines such as sensor 25 in *Saint Martin dHeres* and 52 in *Echirolles*

Our approach is able to determine critical values in order to characterize and use the information for control purposes. It can also be used in such cases the information is retrieved from a network by means of data fusion techniques [10] but our approach provides structure and generalization. On the other hand, for slow lanes the Figure 17 shows a good understanding of the network behavior and a proper approximation of the data measured and processed. The use of Floating car data will allows the entire network representation by means of the NMFD and the easy and understandable calculation of each link FD. Scarcity of data is an unavoidable feature, but without any interpolation or averaging technique we figured it out how to perform a general procedure for description of an entire network without any special constraints.
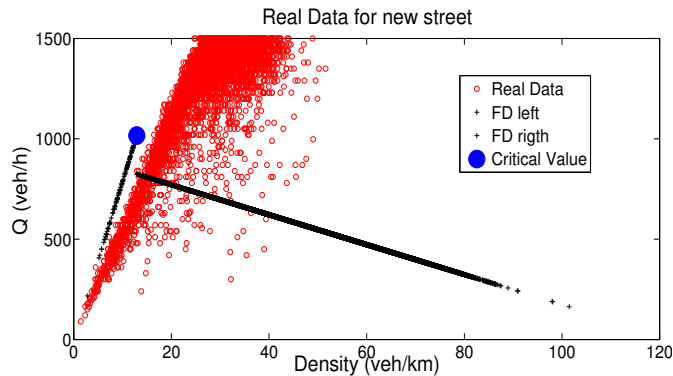

Figure 17: Real data for test processed in slow lines such as sensor 26 in *Saint Martin dHeres* and 51 in *Echirolles*

It is obvious the dependence of the ATR and MFD diagrams to sample time

and scarcity of data. Our method proposes to use the main features of the data and provides a framework of simplification. It means that if you have a network with around 135 sensors and use only 3 sensors (around 5%), you will be able to reconstruct the main dynamics. We suppose that if you use more sensors, the accuracy of the method will increase. Possible errors to our algorithm come from the threshold parameter to outliers cancellation. Patterns obtained from core data allows for proper identification of entire network features for α-scaling. Consequently, FCD will allow reconstructing the FD of any link. The applications of the proposed framework can be to help to propose MFD and FD based control solutions and also to help with the purpose of sensors location reduction taking into account the velocities from FCD for a network monitoring. This way, estimation of dynamic variables will come leading to interesting results in densities and flows. Potential applications for pollutants estimation and subsequent application in a complex network like the city of Medellín Colombia.

7   Conclusions

We developed a method for estimation of a Macroscopic Fundamental Diagram (MFC) "normalized" and the introduction to a Fundamental Diagram of each traffic intersection. We considered three major factors: the data dispersion nature, the sparcity of the data and the role of the link with data within the network. A four-steps method was proposed based on images classifications which preserve data veracity until the last steps instead of first match curves that induce a first approximation, uses image classifications and filtering, floating car data to map back the cluster family towards the links/roads that have missing sensors. Our results are satisfactorily tested on real data from Grenoble Traffic Lab (GTL). Our approach is able to determine critical values in order to characterize and use the information for control purposes. The information retrieved from a network by means of data fusion techniques can also be used, but our approach provides structure and generalization. We developed an approximation that uses as few of these variables as possible, it can be understood as a proper parameterization and description of heterogeneity that would yield good estimates for most practical applications.

8   Acknowledgements

References

[1] C. F. Daganzo, "Urban gridlock: Macroscopic modeling and mitigation approaches," *Transportation Research Part B: Methodological*, vol. 41, no. 1, pp. 49–62, jan 2007. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0191261506000282

[2] C. F. Daganzo and N. Geroliminis, "An analytical approximation for the macroscopic fundamental diagram of urban traffic," *Transportation Research Part B: Methodological*, vol. 42, no. 9, pp. 771–781, nov 2008. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0191261508000799

[3] N. Geroliminis and C. F. Daganzo, "Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings," *Transportation Research Part B: Methodological*, vol. 42, no. 9, pp. 759–770, nov 2008. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0191261508000180

[4] C. F. Daganzo and N. Geroliminis, "An analytical approximation for the macroscopic fundamental diagram of urban traffic," *Transportation Research Part B: Methodological*, vol. 42, no. 9, pp. 771–781, nov 2008. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0191261508000799

[5] E. Thonhofer, T. Palau, A. Kuhn, S. Jakubek, and M. Kozek, "Macroscopic traffic model for large scale urban traffic network design," *Simulation Modelling Practice and Theory*, vol. 80, pp. 32–49, jan 2018. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1569190X17301405

[6] K. An, Y. Chiu, X. Hu, and X. Chen, "A network partitioning algorithmic approach for macroscopic fundamental diagram-based hierarchical traffic network management," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 4, pp. 1130–1139, April 2018.

[7] T. Tsubota, A. Bhaskar, A. Nantes, E. Chung, and V. V. Gayah, "Comparative Analysis of Traffic State Estimation," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2491, no. January, pp. 43–52, 2015. [Online]. Available: http://trrjournalonline.trb.org/doi/10.3141/2491-05

[8] E. Lovisari, C. Canudas de Wit, and A. Y. Kibangou, "Data fusion algorithms for density reconstruction in road transportation networks," in *2015 54th IEEE Conference on Decision and Control (CDC)*. IEEE, dec 2015, pp. 2804–2809. [Online]. Available: http://ieeexplore.ieee.org/document/7402641/

[9] C. Canudas de Wit, F. Morbidi, L. L. Ojeda, A. Y. Kibangou, I. Bellicot, and P. Bellemain, "Grenoble Traffic Lab: An Experimental Platform for

Advanced Traffic Monitoring and Forecasting [Applications of Control]," *IEEE Control Systems*, vol. 35, no. 3, pp. 23–39, jun 2015. [Online]. Available: http://ieeexplore.ieee.org/document/7106643/

[10] L. Ambühl and M. Menendez, "Data fusion algorithm for macroscopic fundamental diagram estimation," *Transportation Research Part C: Emerging Technologies*, vol. 71, pp. 184–197, oct 2016. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0968090X16301267

[11] W. Li, D. Nie, D. Wilkie, and M. C. Lin, "Citywide Estimation of Traffic Dynamics via Sparse GPS Traces," *IEEE Intelligent Transportation Systems Magazine*, vol. 9, no. 3, pp. 100–113, 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7990642/

[12] N. Geroliminis and J. Sun, "Properties of a well-defined macroscopic fundamental diagram for urban traffic," *Transportation Research Part B: Methodological*, vol. 45, no. 3, pp. 605–617, mar 2011. [Online]. Available: http://dx.doi.org/10.1016/j.trb.2010.11.004 http://linkinghub.elsevier.com/retrieve/pii/S0191261510001372

[13] E. Lovisari, C. C. D. Wit, and A. Kibangou, "A Fusion Algorithm for Traffic Density Estimation using Sensors and Floating Car Data," in *22nd ITS World Congress*, no. October, 2015, pp. 5–9. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01185522

[14] E. Lovisari, C. C. D. Wit, and A. Y. Kibangou, "Optimal Sensor Placement in Road Transportation Networks using Virtual Variances," *Conference on Decision and Control (CDC)*, no. Cdc, pp. 2786–2791, dec 2015. [Online]. Available: http://ieeexplore.ieee.org/document/7402641/ http://ieeexplore.ieee.org/document/7402638/

[15] E. Lovisari, C. Canudas de Wit, and A. Y. Kibangou, "Density/Flow reconstruction via heterogeneous sources and Optimal Sensor Placement in road networks," *Transportation Research Part C: Emerging Technologies*, vol. 69, pp. 451–476, aug 2016. [Online]. Available: http://dx.doi.org/10.1016/j.trc.2016.06.019 http://linkinghub.elsevier.com/retrieve/pii/S0968090X16300936

[16] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, jan 1979. [Online]. Available: http://ieeexplore.ieee.org/document/4310076/

[17] L. Wang, Y. Zhang, and J. Feng, "On the Euclidean distance of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1334–1339, aug 2005. [Online]. Available: http://ieeexplore.ieee.org/document/1453520/

[18] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," in *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms.* Society for Industrial and Applied Mathematics, 2007, pp. 1027–1035.

[19] M. Smith and R. Mounce, "A splitting rate model of traffic re-routeing and traffic control," *Transportation Research Part B: Methodological*, vol. 45 ', no. 9, pp. 1389–1409, nov 2011. [Online]. Available: http://dx.doi.org/10.1016/j.sbspro.2011.04.520 http://linkinghub.elsevier.com/retrieve/pii/S1877042811010792 http://linkinghub.elsevier.com/retrieve/pii/S0191261511000634