A CASE STUDY ON COOPERATIVE CAR DATA FOR TRAFFIC STATE ESTIMATION IN AN URBAN NETWORK

L. Montero[1,2*], M. Pacheco[1], J. Barceló[1,2], S. Homoceanu[3], J. Casanovas[1,2]

[1] **inLab FIB**
[2] **Department of Statistics and Operations Research**
*** Corresponding author**

Universitat Politècnica de Catalunya – BarcelonaTECH - UPC
Jordi Girona 1-3 08034-Barcelona (Spain)

(lidia.montero, jaume.barcelo)@upc.edu
(meritxell.pacheco, josepk)@fib.upc.edu
Tel: +34934017033

[3] **Volkswagen AG**

Brieffach 1777
D-38440-Wolfsburg (Germany)

silviu.homoceanu@volkswagen.de
Tel: +4953619993811

Word count: 4748 words + 9 tables/figures x 250 words (each) + 27 References = 6998 (MAX 7000 words, **without references**)

Revised from original version

ABSTRACT

The use of Floating Car Data (FCD) as a particular case of Probe Vehicle Data (PVD) has been the object of extensive research for estimating traffic conditions, travel times and Origin to Destination trip matrices. It is based on data collected from a GPS-equipped vehicle fleet or available cell phones. Cooperative Cars with vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication capabilities represent a step forward, as they also allow tracking vehicles surrounding the equipped car. This paper presents the results of a limited experiment with a small fleet of cooperative cars in Barcelona's Central Business District (CBD) known as L'Eixample. Data collected from the experiment were used to build and calibrate the emulation of cooperative functions in a microscopic simulation model that captured the behavior of vehicle sensors in Barcelona's CBD. Such a calibrated model allows emulating fleet data on a large scale that goes far beyond what a small fleet of cooperative vehicles could capture. To determine the traffic state, several approaches are developed for estimating traffic variables based on extensions of Edie's definition of the fundamental traffic variables with the emulated data, whose accuracy depends on the penetration level of the technology.

Keywords: Probe vehicle, Lagrangian sensing, Traffic Flow, Simulation, Traffic State Estimation, Cooperative Car data

INTRODUCTION

There is a wide consensus about the brand new family of services that will be enabled by advances in inter-vehicular communications. From the early considerations of equipped vehicles as a network of mobile sensors *(1)* to more recent surveys *(2), (3)* , Vehicular Sensor Networks (VSN) have been studied to analyze their capabilities for effectively monitoring not only the equipped vehicles, but also the physical world surrounding it. In urban areas, vehicles equipped with onboard sensors are expected to reach high concentrations in the near future. Furthermore, the pervasive penetration of smart-phones, with internet connections and their sensing capabilities (e.g., location, accelerometer, etc.), makes them another category of mobile sensors to be added to the proper vehicular onboard sensors.

However, most analyses have been conducted from the perspective of driver safety (one of the primary objectives) or the continuous monitoring of the vehicle state, while some others have considered the potential guidance offered by such mobile sensoring. But, most of the work has been conducted from the perspective of a telecommunications network, which consequently deals with the efficient design of communications architectures. A few analysts have taken into account that mobile sensing platforms provide a means for collecting, processing and accessing sensor data. Of the various novel applications that this offers, we should also consider traffic flow estimation, monitoring ride quality and supporting proactive traffic surveillance, among others.

The work done by *(4)* addresses the combination of both telecommunications and traffic networks in a simulated scenario, proving the feasibility of the main applications, the capture of data and their potential use for management and evaluation. It neglects to go into the technical details (from a traffic modeling perspective) regarding the quality of the generated data. The same simulation platform is used by *(5)* in a typical example of one of the most studied application: how Vehicle-to-Vehicle (V2V) communications can be used to identify congestion ahead and alert other equipped vehicles. This is a relevant approach when safety is the primary objective, and it may also provide guidance for circumventing congestion if no other information is available.

In a parallel line, other researchers have been motivated by advancements in vehicle positioning (one of the main onboard sensors) and in wireless communications supporting the V2V and Vehicle to Infrastructure (V2I) applications. They have investigated how the collected data could be used to generate travel time information, which is some of the traffic information that is most relevant for Advanced Traveler Information Systems (ATIS) (see *(6)*). This research could be considered complementary to a domain of research that has deserved substantial attention in recent years: the estimation of links or path travel times from GPS probe vehicles, also called Floating Car Data. Summarily speaking, two complementary approaches have been explored. *(7)* presents an overview of the statistical approaches, which later on was improved by *(8)*, who combined the GPS data with data from other travel time measurement sources, such as those supplied by License Plate recognition (LPR) technologies. A variant of these statistical models exploiting the GPS data is analyzed in *(9)* to identify paths in the network, whose travel times are estimated in *(10)*. This paper also addresses a relevant aspect concerning the quality of the information provided by ATIS, such as travel time distributions. Here, they assess the reliability of the estimates, which is a key point from the user's perspective.

Many other authors address the travel time prediction on basis to the Ensemble Kalman Filter *(11)* using as core model a Cell Transmission Model (CTM) to emulate the dynamics of the traffic flows. This has been particularly studied by participants in the Berkeley Mobile Millennium *(12)* and in the Stockholm Mobile Millennium *(13)*. The primary description of the approach can be found in *(14)* and in *(15)*. An example of calibration that is assisted by complementary ICT measurements is described in *(16)*.

Nevertheless, although sound estimates of travel times provide valuable information to travelers, traffic management systems need something else, namely procedures to estimate the traffic state and to predict its short term evolution, whenever possible.

Recent approaches exploiting GPS data have been based on Extended Kalman Filtering (EKF) *(17)*. Based on Lagrangian observations, it is used for traffic state estimation on freeways and was validated by simulation using a fleet of GPS-equipped cars and a boundary detector. Other approaches have combined probabilistic models with nonlinear traffic models *(18)*.

In the context of exploiting mobile data, one emerging trend *(19)* explores the use of handover information in cellular networks in order to estimate the traffic level of service. The correlation between cell phone handovers and traffic volumes was investigated in order to understand the relationship between cellular activity and vehicular traffic, although they note important limitations in the overall performance. But to move from counting processes to generating initial estimates is not enough; samples must be expanded and their reliability improved upon and checked against other traffic measures. In *(20)*, an approach is proposed that combines the counting processes with ad hoc traffic assignment models.

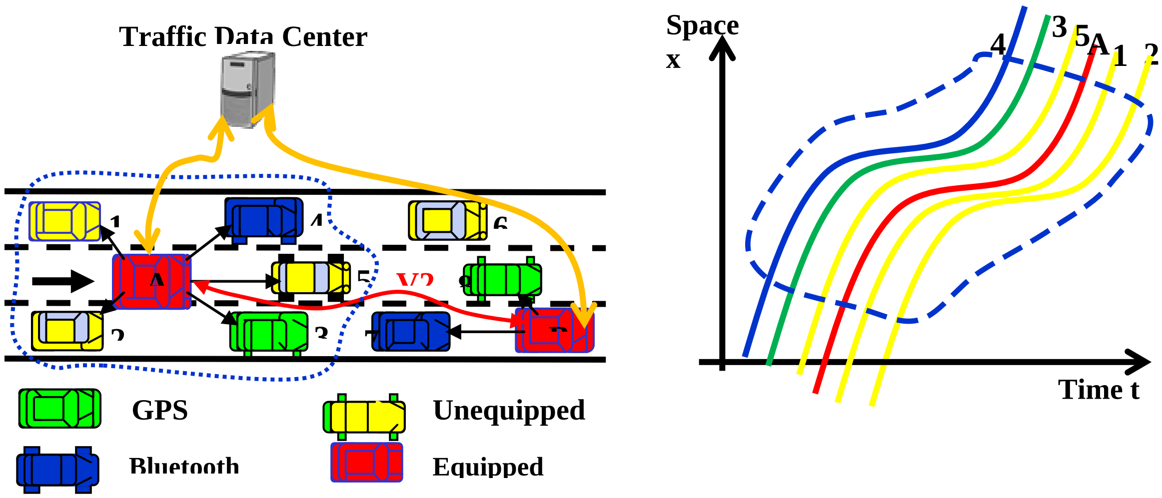


FIGURE 1. V2V equipped vehicles "tracking" a number of surrounding unequipped vehicles (left) and Trajectory reconstruction (right)

However, since tracking equipped vehicles enables the reconstruction of their trajectories, more direct procedures have been developed, such as estimating the traffic state by directly estimating the values of the fundamental traffic variables. This is done by reconstructing the trajectories according to the fundamentals of traffic flow theory. This approach has been explored in *(21)*, which estimates flow, density and speed variables based on the spacing data collected by probe cars from leader vehicles. It also uses a classical conceptualization derived from Edie's generalized definitions *(22)*.

The approach taken in this paper expands on that of *(21)*. FIGURE 1 assumes that the V2V technology is onboard the red cars in A and B. Not only does this allow communication between the equipped vehicles and the Traffic Data Center, but car A also communicates data from those numbered 1, 2, 3, 4 and 5 encircled by the blue dotted line, because the equipped vehicle is "aware" of a number of surrounding vehicles.

The Traffic Data Center is then able to track the equipped vehicle and reconstruct its trajectory (red vehicle and trajectory A in FIGURE 1) in a similar way to some of the proposals in the above referenced papers, which can reconstruct the trajectories of GPS-equipped vehicles. However, the approach is extended by employing the data of the surrounding vehicles captured by the equipped vehicle. Not only is it possible to reproduce the approach in *(21)* and reconstruct the trajectory of the leader vehicle (number 5 in FIGURE 1), but the trajectories of vehicles 1, 2, 3 and 4 can also be reproduced.

The objective then is to estimate the values of the fundamental traffic variables by applying Edie's theory to the enriched set of trajectories. However, due to current limitations on the available technologies, some of the trajectories of the unequipped vehicles can only be partial.

The accuracy of probing data depends of course on the accuracy of the measurements supplied by the individual sensors. However, on the system level, it is a function of the number of probes that can be collected in a time unit per road length unit. Accuracy therefore depends on the coverage of traffic flow by the probe cars as well as on the sensor sampling time. Taking into account that there is no available fleet of equipped vehicles with the required size, the study achieves this objective through simulation. Specifically, we have developed a procedure for emulating the data captured from equipped vehicles.

The structure of this paper is as follows. First, we describe the field experiment with a small fleet of equipped vehicles and the data collected through filtering and analysis. We then follow with a discussion on the data used to calibrate the emulation of equipped vehicles through our microscopic traffic simulator. A further section describes the simulation experiments for large fleets of equipped vehicles according to penetration rates. The next section generalizes the application of Edie's theory in *(21)* to the simulated scenarios, provides an analysis of the results and ends with conclusions.

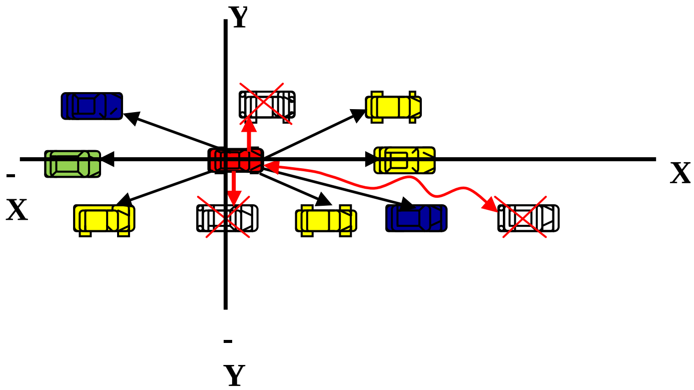
BARCELONA DATA COLLECTION STUDY

The main goal of this paper is to determine how to estimate the values of the local traffic variables and the prevailing traffic conditions by using enriched sets of vehicle trajectories that are reconstructed from the data captured by the equipped vehicles. The work was divided into two stages. In the first stage, a set of experiments were designed and conducted to collect physical data from a fleet of 3 equipped vehicles. The collected data were analyzed and cleaned to remove any outliers and incomplete observations. The filtered data were then used to design and implement procedures to reconstruct the partial trajectories of vehicles.

The second stage addressed the problem of how to extend the procedure to a case where we can get significant results for thousands of equipped vehicles rather than small fleets. According to the research results of *(23)*, the threshold for penetration rates of equipped vehicles should lay between 7% and 10%. Traffic microsimulation in Aimsun *(24)* was used to emulate the collection of data for a large fleet of connected vehicles. The simulation model of the selected scenario and the developed APIs to emulate data captures from equipped vehicles were calibrated using real data collected from the field test consisting of 3 equipped vehicles.

Generic Description of Data Collection

To explicitly take into account market evolution, the physical experiment considered only basic vehicle sensors that were already available on Volkswagen vehicles. To be specific, data were collected by one Volkswagen Golf-Generation-7 and two Audi-A3s equipped with GPS, side and front radars. FIGURE 2 presents the position of the sensors as well as the functions they normally serve.



Sensor Technologies	Functions in Series
GPS	• Navigation
Front Camera	• Lane Keeping Assistant • Traffic Sign Recognition • Dynamic Light Assistant • Adaptive Cruise Control
Radars	• Front Assistant • City Emergency Brake • Adaptive Cruise Control • Side Assistant

FIGURE 2. Sensors, functions and surrounding "awareness" of probe cars – depicted here in red. Seven vehicles detected, 4 in front and 3 behind.

Each of the vehicles continuously measures and stores its position, heading, speed, and acceleration, as well as distances and relative speeds to surrounding vehicles captured by the radars. Given the particularities of each sensor, not all surrounding vehicles will be visible to the equipped car, as shown in FIGURE 2: cross-marked vehicles are not detectable by the available sensors. Furthermore, motionless vehicles are not detected by the current technology. The front radar has a radius of about 180 meters and it can simultaneously track multiple vehicles driving in front of the equipped car. However if the line of sight to one vehicle is continuously obstructed (for instance by another vehicle), the respective vehicle will not be detected at all. Since sensors are mounted in the front and rear ends of the car, vehicles driving parallel to the equipped car are not detected.

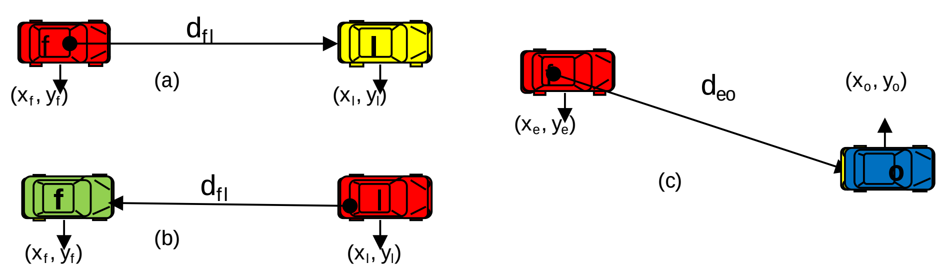


FIGURE 3. Local traffic related data collected by an equipped car (red).

Adopting traffic flow theory terminology, for each pair of vehicles surrounding an equipped car denoted by *e*, a leader is denoted by *l*, a follower is denoted by *f* and, more generally, an observed vehicle is denoted by *o*. For each pair of equipped and surrounding cars, the data gathered by the equipped car at time *t* includes the following measures (presented in FIGURE 3):

- The relative distance d_{eo} between the equipped and the observed car
- The relative speed v_{eo} between the equipped and the observed car
- The map-matched position (x_e,y_e) and speed v_e of the equipped car.

Together with the GPS position data, which is available for both the equipped and the observed vehicles, volume-related information is required to estimate traffic state variables. This is the case for the probe vehicles presented here, since the relative distance between the equipped and the observed vehicles is provided and thus also the volume-related information in terms of spacing measurement data.

Data collection design

Data was collected on three cars driving eight hours/day (including rush-hour intervals) during five weekdays in November 2014. To capture the origin-destination behavior, 12 itineraries were established. Professional drivers were given precise instructions on which itinerary to start and at which time, in order to cover all itineraries multiple times at rush-hour intervals.

For example, Itinerary 2 (depicted in FIGURE 4, middle image) is 10.6 km long, and one lap takes 25-28 min in non-peak conditions. Itinerary 2 incorporates Gran Via Avenue (heading north) and Aragó Street (heading south), both one-way urban corridors in the characteristic grid of L'Eixample District. Some touristic attractions lay within or are very close to this itinerary. Aragó Corridor is the one selected for traffic state estimation. It has between four and six lanes. There are traffic lights every 150m at each intersection, and merging/diverging lanes.

Data Analysis

After loading data, exploratory analysis was performed to identify semantically meaningful trends, atypical behavior and outliers in the data. The analysis covered the sensor data of the equipped as well as o bserved vehicles (position, heading, speed and acceleration). Headways, detection time-length and number of observed vehicles per equipped vehicle were also considered.

Overall, the speeds of the equipped vehicles were consistently within the city limits, with a percentage of relative stopped time (speed less than 3 km/h) of between 36% and 53%, depending on the itinerary. The acceleration typically ranged between -0.36 m/s^2 and 0.32 m/s^2. In this case, some outliers can be recognized, with a maximum positive acceleration higher than 6 m/s^2 and a negative one lower than -4 m/s^2. Taking a closer look at such values, the maximum acceleration value was registered during a stop-and-go situation: first the vehicle reduced speed when coming to a stop; after a series of 54 seconds at 0 km/h speed, the vehicle is reported to be 150m further down the street 0.5 seconds later and driving at 45 km/h. Such abnormal behavior is obviously due to a temporary malfunction of the measurement devices and has been filtered.

DESIGN OF SCENARIOS FOR SIMULATION STUDIES

Model Description

The selected microsimulation scenario was Barcelona's CBD (FIGURE 4), comprising 7.46 km^2 with more than 250,000 inhabitants. Its Aimsun *(24)* model consists of 2,111 sections, 1,227 nodes, 120/130 generation and attraction centroids and 877 non-zero OD pairs. The horizon study is 30 min, accounting for a total number of 20,700 vehicle trips.

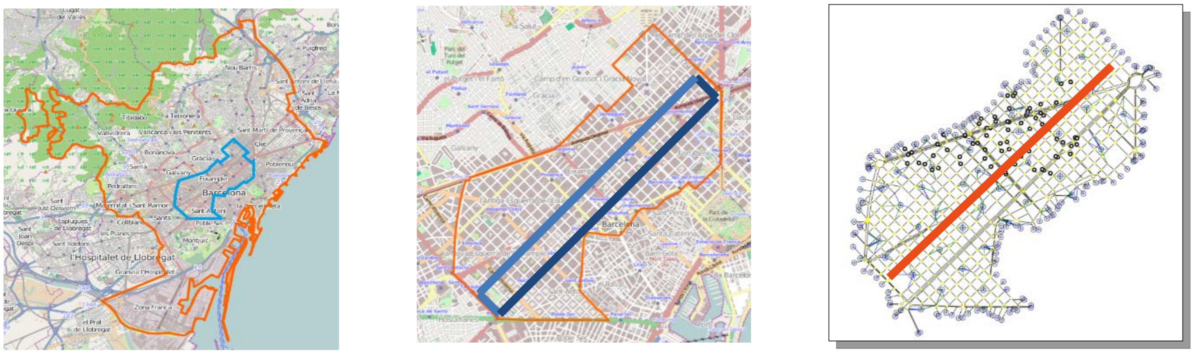


FIGURE 4. L'Eixample in Barcelona: simulation test-site and Aimsun model. Itinerary 2 in blue (central) Selected *Aragó Corridor* in orange (right)

Emulated probe car data were used as input for the proposed traffic state estimation method. The multi-lane urban corridor along *Aragó Street* is controlled by traffic lights and merging/diverging segments. *Aragó Corridor* constitutes a part of Itinerary 2, and data from probe cars were available for several periods. Although the scenario modeled by Aimsun was calibrated several years ago, changes and roadworks do not affect *Aragó Corridor*, because it is one of the more stable arterials in L'Eixample. Furthermore, no loop detector data were available to the authors.

Instead of focusing on the calibration of the whole scenario, we concentrated on *Aragó Corridor*. Regarding Itinerary 2 along *Aragó Corridor*, we used the simulator to faithfully reproduce the percentage of stopped time as one of the analyzed performance indicators. However, the simulated speed range was slightly different from that collected in the field experiment. Thus, the global simulation parameters for speed acceptance were tuned to reproduce a wider dispersion of speeds along the corridor.

Methodological Proposal

The following conditions are assumed in the emulation of probe car data collection:

- Probe car data measurements have no error for position, heading and speed for the same probe car, and they are collected for each time-stamp (half second).
- For cars inside the visibility polygon of a probe car, each time-stamp is provided with relative position, heading and observed car speed. All of them are estimated with a negligible measurement error.
- The trajectory for probe cars is identified without error.
- Probe cars represent a fixed percentage of OD demand in the study area, which is spread uniformly among the OD pairs.
- Probe cars perform as floating cars, since their behavior is the same as the other private vehicles.

Once *Aragó Corridor* was properly calibrated in the basic scenario, an API was developed to calculate the probe car radar visibility polygon for observed cars. This was done by adapting a ray-casting detection algorithm to obtain observed objects according to the radar detection range, the location of radars in the probe cars *(25)* and the surrounding conditions. The polygon was generated as a robust α-shape profile, calculated by using the package *alphahull (26)* in R *(27)*.

The algorithm was validated from three different perspectives at Itinerary 2. It was done first by comparing the collected and emulated data according to the robust α-shape profile for the sample of points defined by the positions relative to the probe car. Second, we compared the collected and emulated distributions for the number of observed vehicles per time-stamp and time detection length per vehicle. Third, we checked the collected versus emulated traffic variable distributions, such as inter-vehicle frontal and rear distances and speeds for probes and observed cars. Corresponding comparative plots are in FIGURE 5.

Percentile plots (QQPlot) for comparison of emulated versus field test observed data for several variables was performed. A R^2 coefficient of determination of 0.9882 for the regression line fitting percentiles of emulated and collected speeds in Itinerary 2 was obtained. The null hypothesis stated as the slope equals to 1 for the fitted line has a p-value of 0.36. Thus emulated speed has been successfully validated. The same holds for inter-vehicular distances and accelerations.

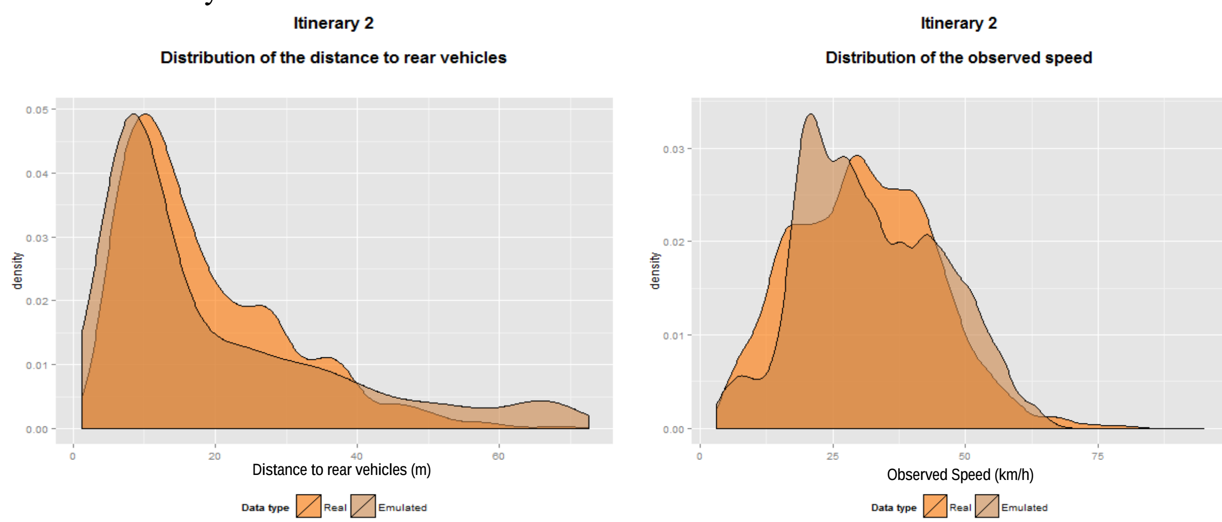


FIGURE 5. Distributions of real *vs.* emulated rear inter-vehicle distance (left) and observed speed (right) from probe cars in Itinerary 2

A dynamic user equilibrium (DUE) was obtained for the base scenario, with a relative gap of less than 1%. Optimal OD paths and OD path proportions were stored. The behavioral assumptions were those of DUE for route choice in probe cars and regular cars in the simulation experiments.

Simulation Experiments

The design of the simulation experiments was targeted to assess the availability of traffic data when a large fleet of probe cars circulate in a realistic scenario. Considered factors were:

- Penetration rate of probe cars providing traffic data for themselves and surrounding vehicles: 5% and 10%.
- Estimated time-window length of traffic variables: 1.5 min, 3 min and 6 min (multiples of the cycle length in L'Eixample).

A first analysis of the results was provided by each experiment in terms of maps showing coverage performance indicators. Coverage indicators were defined at the lane and link level for the global horizon of study (30 min) and for each time-window. FIGURE 6 shows a map consisting of a gradient of colors indicating the percentage of the total number of circulating vehicles that were captured by probe cars for each road segment. We establish that 3-5 replications were enough to estimate the overall coverage, with an absolute error of less than 1% (confidence interval 90%). For each replication, statistical data was stored in the Aimsun database.

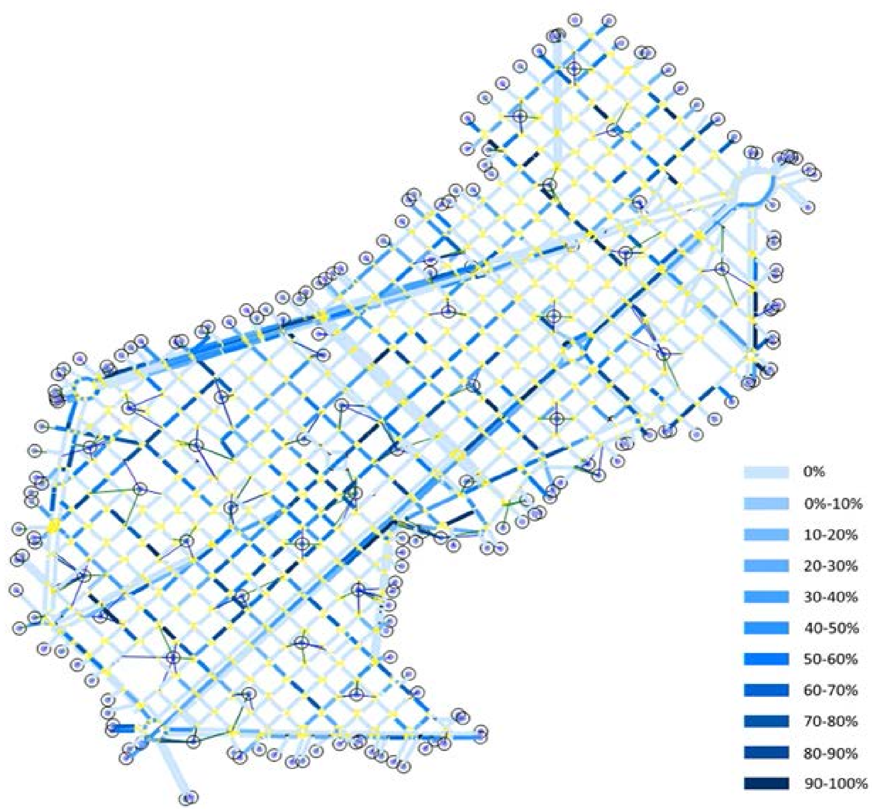


FIGURE 6. Percentage of total vehicles monitored by probe cars in a simulation experiment, with 5% penetration rate and a 3' time-window.

ESTIMATION OF TRAFFIC VARIABLES FROM SIMULATED PROBE CAR DATA

Probe vehicles equipped only with GPS positioning devices cannot provide volume-related information. Nevertheless, on-board devices that recognize surrounding vehicles as those in our experiment may be able to supply these data. Based on this idea, the researchers in *(21)* developed a method to estimate flow, density and speed from both GPS positioning and spacing measurement data. The latter were provided by cameras located on the dashboards of the vehicles.

The estimation method described in *(21)* consists of two main steps. The first discretizes the targeted time-space domain into discrete time-space regions, and the second uses probe vehicle data to apply Edie's generalized definitions for estimating the traffic-state variables for each region. It is worth noting that this method is for traffic variables in a one-way urban freeway with multiple lanes, merging and diverging sections, and whose topology allows for ad hoc discretizations. However it cannot be applied in a straightforward way to an urban scenario whose conditions must be considered in order to adapt the procedure, since the topology induces specific discretizations that account for the space length between two consecutive junctions, thus keeping the regions from being equally sized. Moreover, as opposed to freeways, the interruptions originating from the traffic light system prevent the time from being continuous. Thus, in our case, the length of the intervals employed to discretize the time horizon depends on the common cycle length in *Aragó Corridor*.

The traffic state variables in a time-space region **A** (FIGURE 7) specifically for the flow $q(\mathbf{A})$, the density $k(\mathbf{A})$ and the speed $v(\mathbf{A})$, were defined by Edie as follows:

$$q(\mathbf{A}) = \frac{d(\boldsymbol{A})}{|\boldsymbol{A}|}, \quad k(\mathbf{A}) = \frac{t(\boldsymbol{A})}{|\boldsymbol{A}|}, \quad v(\mathbf{A}) = \frac{d(\boldsymbol{A})}{t(\boldsymbol{A})},$$

where $d(\boldsymbol{A})$ is the total distance traveled by all vehicles in region **A**, $t(\boldsymbol{A})$ is the total time spent by all vehicles in region **A** and $|\mathbf{A}|$ the time-space area of the region **A**.

This formulation generates the estimators based on the probe vehicle data by considering the subset of vehicles comprising probe vehicles in each region. These estimators are:

$$\hat{q}(\mathbf{A}) = \frac{\sum_{n \in \boldsymbol{P}(\boldsymbol{A})} d_n(\boldsymbol{A})}{\sum_{n \in \boldsymbol{P}(\boldsymbol{A})} |a_n(\boldsymbol{A})|}, \quad \hat{k}(\mathbf{A}) = \frac{\sum_{n \in \boldsymbol{P}(\boldsymbol{A})} t_n(\boldsymbol{A})}{\sum_{n \in \boldsymbol{P}(\boldsymbol{A})} |a_n(\boldsymbol{A})|}, \quad \hat{v}(\mathbf{A}) = \frac{\sum_{n \in \boldsymbol{P}(\boldsymbol{A})} d_n(\boldsymbol{A})}{\sum_{n \in \boldsymbol{P}(\boldsymbol{A})} t_n(\boldsymbol{A})},$$

with $\boldsymbol{P}(\boldsymbol{A})$ being the set of probe vehicles within region **A**, $d_n(\boldsymbol{A})$ the total distance traveled by vehicle n within region **A**, $t_n(\boldsymbol{A})$ the total time spent by vehicle n within region **A** and $|a_n(\boldsymbol{A})|$ the area of the time-space region between probe vehicle n and its leading vehicle in region **A**.

The implementation of this methodology becomes straightforward by using the leader vehicles as the subset of observed vehicles. Although the type of vehicle is not provided, as is the case with an equipped vehicle (leader, follower or other), it can be inferred according to FIGURE 3.

However, when considering only leader observations, a preliminary trajectory analysis for each time-space region showed some blanks along the trajectory. Some time-stamps exist in which leader data is not gathered. In FIGURE 7, the trajectories for three equipped vehicles in time-space region **A** are illustrated. For each equipped vehicle, its corresponding leader is plotted (left), showing some missing data at some time-stamps. Since motionless vehicles are not detected, some imputations of the data regarding motionless vehicles at traffic lights are required.

By including follower observations in the preceding plot, we realize that some of the aforementioned blanks can be mitigated (see FIGURE 7-right). This consideration leads to the definition of the *Leader-Follower* approach by taking into account follower data when the leader is not available. Then, a couple of new approaches arise for dealing with the estimation of traffic state variables, namely the *Leader approach* as in *(21)* and the *Leader-Follower approach* that extends *(21)* .

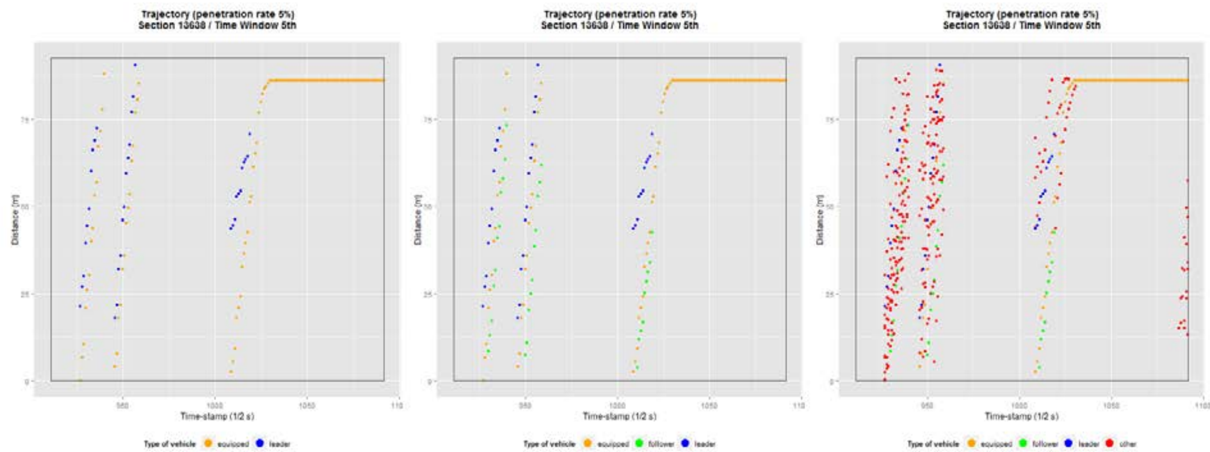


FIGURE 7. Trajectory scatterplot for equipped, leader and follower vehicles in a time-space region A

Going further, we can extend previous approaches into a new method by considering the data from all the vehicles observed in the equipped surroundings. We call this the *Extended approach*. For each lane within the section all consecutive observed vehicles are considered, without loss of generality and regardless of the vehicle type, as pairs of probe vehicle-leader. Then, the *Leader approach* can be applied for each of such pairs, in the sense of how the total travelled distance, the total time spent and the areas are calculated. This approach requires an error-free identification of the observed vehicles. However, due to the original purpose in the radar technology's design, this is not the case. Nevertheless, this limitation can be overcome for the present study because Aimsun supplies a unique identifier for each vehicle.

The described methods were implemented in R, and the obtained results were contrasted against the values provided by Aimsun, which will be taken as ground truth values from now on and can be obtained at any level of aggregation for each model's section and each predetermined time window.

FIGURE 8 compares estimated and ground truth for traffic variables according to the *Extended approach*. This visualization corresponds to an experiment with a 5% penetration rate and 6 min time-window. Several experiments with different combinations of design parameters were implemented (penetration rate 5%-10% and time-window length 1.5'–3.0'–6.0'), but due to space limitations, only an example has been included. Speed is the traffic variable that best reproduces the ground truth phenomena, whereas flow and density tend to be underestimated in all tested scenarios.

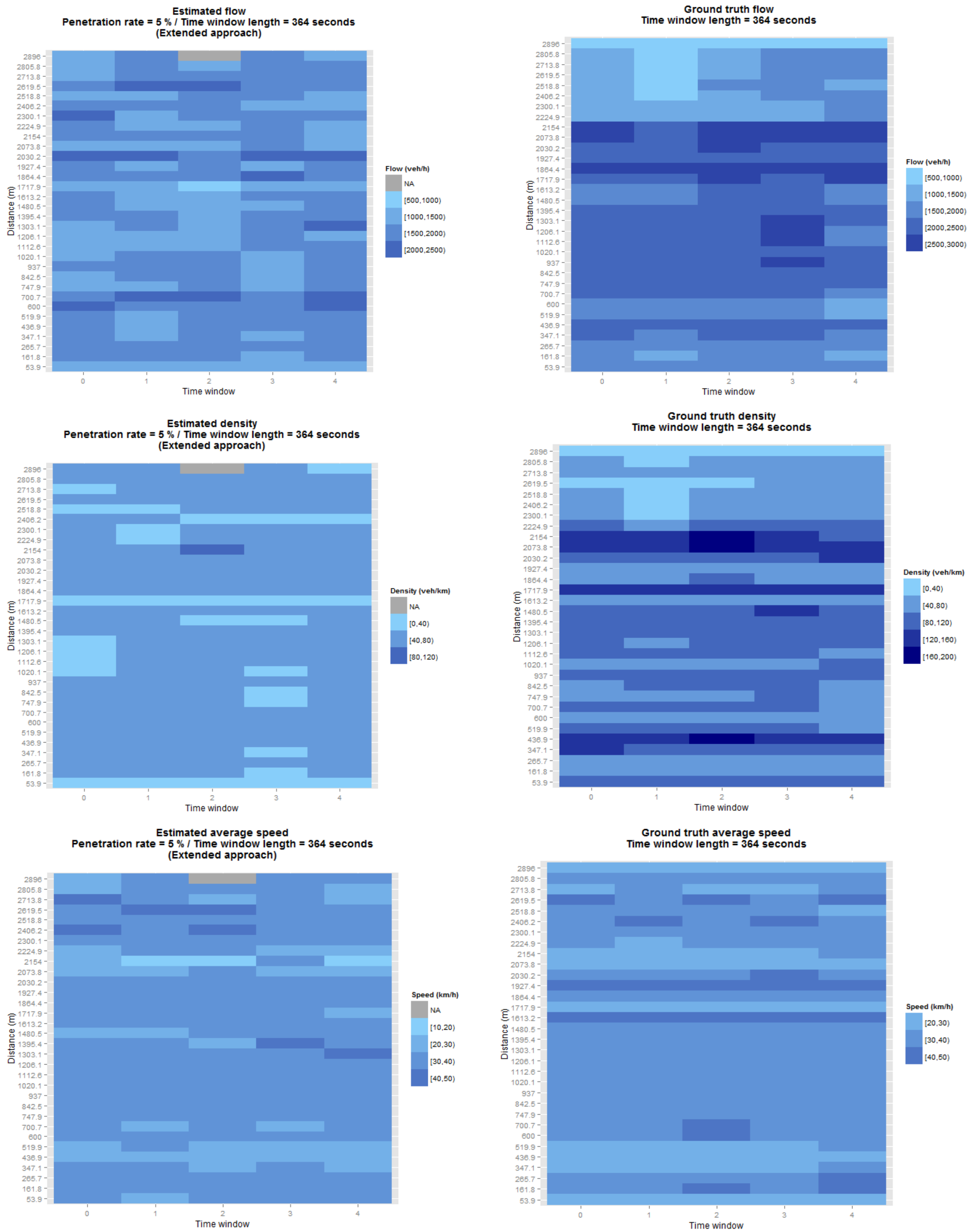


FIGURE 8. Representation of the estimated and ground truth flow, density and speed. Extended approach

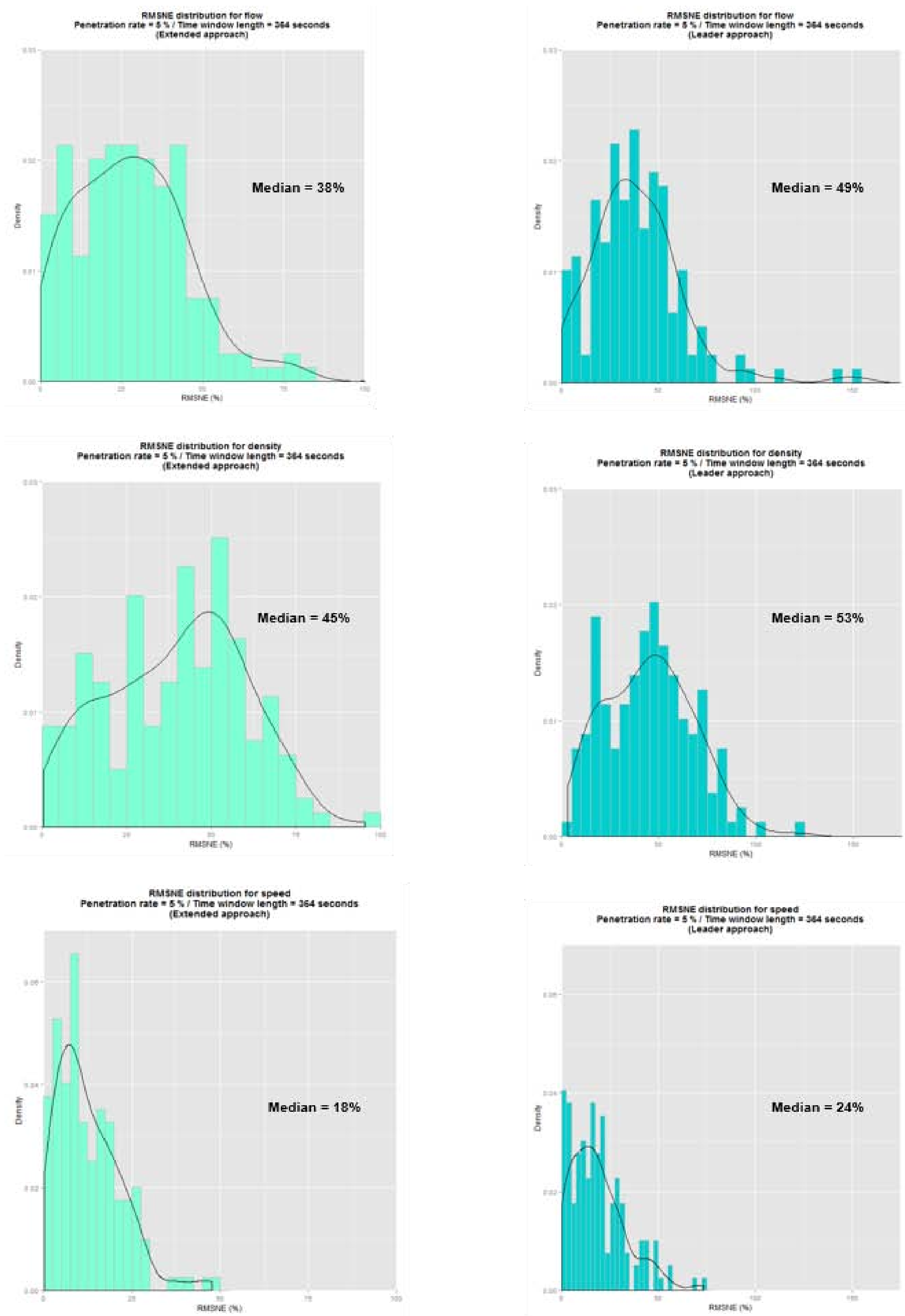


FIGURE 9. Distribution of the RMSNE for flow, density and speed estimates: Extended (left) vs Leader (right) Approaches (5% penetration rate-6' time-window)

FIGURE 9 shows the distribution of the Root Mean Square Normalized Error (RMSNE) for the estimated variables. Speed is that which reports lower errors (medians for the approaches rely around 20%), which makes sense because its estimator is unbiased (see *(21)* for a discussion of statistical properties). Although flow and density present larger RMSNE, mainly density estimates, the *Extended approach* shows an improvement compared to the basic *Leader approach*.

The proposed method under any of the developed approaches reduces RMSNE as t he time-window increases (91 to 182 and 364 s ec). Speed estimates are more reliable than flow estimates and flow more reliable than density estimates according to Theil and RMSNE goodness of fit indicators. As the penetration rate of probe vehicles increases from 5% to 10%, the reliability also increases following the same profile. Low penetration rate means lack of observations and this mainly affects when section length is not large. The results of the *Extended approach* support further developments of the imputation procedures.

CONCLUSIONS AND FUTURE RESEARCH

The undertaken simulation experiments have shown that probe car data open up a promising line of research for traffic state estimation. Probe car data can also be enhanced by data fusion techniques that combine traffic data from several sources. Market penetration in probe car technology is increasing, and this trend points positively to using probe car data to characterize the traffic state at a low cost for traffic authorities. Our results confirm the validity of the approaches in urban arterials, and they further provide guidelines for improving the involved technology for traffic state estimation and forecasting. Nevertheless, the length of the sections involved in the AIMSUN model seems to play an important role in the estimate errors, since they increase as space discretization decreases (shorter sections). A further study will be conducted.

ACKNOWLEDGMENTS

This research was funded by TRA2014-52530-C3-3-P of the Spanish R+D National Programs and *Secretaria d'Universitats i Recerca de la Generalitat de Catalunya* under 2014 SGR 1534. Throughout, the authors have benefited from the support of R.Martín, C.Carmona, S.Pedrajas and D.Martínez from inLab-FIB.

REFERENCES

1. Hull, B., V. Bychkovsky, Y. Zhang, K. Chen, M. Goraczko, A. Miu, E. Shih, H. Balakrishnan, and S. Madden. CarTel: A Distributed Mobile Sensor Computer System. *Proceedings of SenSys'06 November 1-3, Boulder, Colorado, USA*, 2006, pp. 1–14.
2. Lee, U., and M. Gerla. A survey of urban vehicular sensing platforms. *Computer Networks*, Vol. 54, No. 4, 2010, pp. 527–544.
3. Bessler, S., and T. Paulin. *Literature Study on the State of the Art of Probe Data Systems in Europe*. 2013.
4. Rondinone, M., J. Maneros, D. Krajzewicz, R. Bauza, P. Cataldi, F. Hrizi, J. Gozalvez, V. Kumar, M. Röckl, L. Lin, O. Lazaro, J. Leguay, J. Härri, S. Vaz, Y. Lopez, M. Sepulcre, M. Wetterwald, R. Blokpoel, and F. Cartolano. ITETRIS: A modular simulation platform for the large scale evaluation of cooperative ITS applications. *Simulation Modelling Practice and Theory*, Vol. 34, 2013, pp. 99–125.
5. Bauza, R., and J. Gozalvez. Traffic congestion detection in large-scale scenarios using vehicle-to-vehicle communications. *Journal of Network and Computer Applications*, Vol. 36, No. 5,

2013, pp. 1295–1307.

6. Rim, H., C. Oh, K. Kang, and S. Kim. Estimation of Lane-Level Travel Times (L2TT) in V2X-based Traffic Information System. *Proceedings of 90th Transportation Research Board Annual Meeting - CD-ROM*, 2011.
7. Jenelius, E., M. Rahmani, and H. N. Koutsopoulos. Travel time estimation for urban road networks using low frequency GPS probes. *Transportation Research Board 91st Annual Meeting Compendium of Papers. CD-ROM*, 2012.
8. Jenelius, E., and H. N. Koutsopoulos. Travel time estimation for urban road networks using low frequency probe vehicle data. *Transportation Research Part B: Methodological*, Vol. 53, 2013, pp. 64–81.
9. Rahmani, M., and H. N. Koutsopoulos. Path inference from sparse floating car data for urban networks. *Transportation Research Part C: Emerging Technologies*, Vol. 30, 2013, pp. 41–54.
10. Rahmani, M., E. Jenelius, and H. N. Koutsopoulos. Non-parametric estimation of route travel time distributions from low-frequency floating car data. *Transportation Research Part C: Emerging Technologies*, 2014.
11. Evensen, G. *Data Assimilation: the Ensemble Kalman Filter*. Springer-Verlag, Heidelberg, 2009.
12. Herrera, J. C., D. B. Work, R. Herring, X. Ban, Q. Jacobson, and A. M. Bayen. Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment. *Transportation Research Part C: Emerging Technologies*, Vol. 18, No. 4, 2010, pp. 568–583.
13. Allström, A., J. Archer, A. M. Bayen, S. Blandin, J. Butler, D. Gundlegard, H. N. Koutsopoulos, J. Lundgren, M. Rahmani, and O.-P. Tossavainen. MOBILE MILLENNIUM STOCKHOLM, 2nd International Conference on Models and Technologies for Intelligent Transport Systems. 2011.
14. Work, D. B., S. Member, O. Tossavainen, S. Blandin, A. M. B. Member, T. Iwuchukwu, and K. Tracton. An Ensemble Kalman Filtering Approach to Highway Traffic Estimation Using GPS Enabled Mobile Devices. *Proceedings of IEEE Conference on Decision and Control*, 2008, pp. 5062–5068.
15. Work, D. B., S. Blandin, O. P. Tossavainen, B. Piccoli, and a. M. Bayen. A Traffic Model for Velocity Data Assimilation. *Applied Mathematics Research eXpress*, Vol. 2010, No. 1, 2010, pp. 1–35.
16. Allström, A., A. Bayen, M. Franssona, D. Gundlegard, A. Patire, C. Rydergren, and M. Sandin. Calibration Framework based on Bluetooth Sensors for Traffic State Estimation Using a Velocity based Cell Transmission Model. *Transportation Research Procedia*, Vol. 3, 2014, pp. 972–981.
17. Yuan, Y., J. W. C. Van Lint, R. E. Wilson, F. Van Wageningen-Kessels, and S. P. Hoogendoorn. Real-time lagrangian traffic state estimator for freeways. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 13, No. 1, 2012, pp. 59–70.
18. Hofleiter, A. *Leveraging geolocalization technologies to model and estimate urban traffic*. University Paris- Est, 2012.
19. Demissie, M. G., G. H. de Almeida Correia, and C. Bento. Intelligent road traffic status detection system through cellular networks handover information: An exploratory study. *Transportation Research Part C: Emerging Technologies*, Vol. 32, 2013, pp. 76–88.
20. Iqbal, M. S., C. F. Choudhury, P. Wang, and M. C. González. Development of origin–destination matrices using mobile phone call data. *Transportation Research Part C: Emerging Technologies*, Vol. 40, Mar. 2014, pp. 63–74.
21. Seo, T., T. Kusakabe, and Y. Asakura. Estimation of flow and density using probe vehicles with spacing measurement equipment. *Transportation Research Part C: Emerging Technologies*, Vol. 53, 2015, pp. 134–150.
22. Edie, L. C. Discussion of traffic stream measurements and definitions in Proceedings of 2nd International Symposium on the Theory of road Traffic Flow. 1963.
23. Huber, W., M. Lädke, and R. Ogger. Extended floating-car data for the acquisition of traffic information. 1999.
24. Transport Simulation Systems. Aimsun Microscopic Simulator (v 8.0). http://www.aimsun.com.
25. Franklin, W. R. Art Gallery Theorems and Algorithms (Joseph O'Rourke). 2. Volume 31, 342–343.

26. Pateiro-l, B., and A. Rodr. Generalizing the Convex Hull of a Sample : The R Package Alphahull. *Journal of Statistical software*, Vol. 34, No. 5, 2010, pp. 1–28.
27. R Development Core Team, R. R: A Language and Environment for Statistical Computing. Volume 1, 409. http://www.r-project.org.