Research Article

Multisource Data Framework for Road Traffic State Estimation

Josep Maria Salanova Grau ¹, **Evangelos Mitsakis** ¹, **Panagiotis Tzenos** ¹,
Iraklis Stamos ², **Luigi Selmi**³, and **Georgia Aifadopoulou** ¹

¹Centre for Research and Technology Hellas, Hellenic Institute of Transport, 6th km Charilaou-Thermi Rd.,
57001 Thermi, Thessaloniki, Greece

²IRU Projects ASBL, 71 Avenue de Cortenbergh, 1000 Brussels, Belgium

³Fraunhofer IAIS, Schloss Birlinghoven, 53757 Sankt Augustin, Germany

Correspondence should be addressed to Evangelos Mitsakis; emit@certh.gr

Received 14 December 2017; Revised 13 April 2018; Accepted 23 May 2018; Published 19 June 2018

Academic Editor: Gonalo Homem de Almeida Correia

Copyright 2018 Josep Maria Salanova Grau et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents a framework for data collection, filtering, and fusion, together with a set of operational tools to validate, analyze, utilize, and highlight the added value of probe data. Data is collected by both conventional (loops, radars, and cameras) and innovative (Floating Car Data, detectors of Bluetooth devices) technologies and refers to travel times and traffic flows on road networks. The city of Thessaloniki, Greece, serves as a case study for the implementation of the proposed framework. The methodology includes the estimation of traffic flow based on measured travel time along predefined routes and short-term forecasting of traffic volumes and their spatial expansion in the road network. The proposed processes and the framework itself have the potential of being implemented in urban road networks.

1. Introduction

Emerging technological advances have lately revealed new transport and mobility data collection means, enabling cost-efficient ways of both pullulating the actual amount of collected data as well as enriching their quality (in terms of type, format, and content). The broader society's new eye on privacy and sharing of personal information is also with doubt to be attributed with contributing to the latter, as people increasingly need to be "connected". To this end, the travellers (as both drivers and passengers) have shifted from sole information receivers (for meeting their own needs) to active information transmitters, becoming thus key players of the data collection and sharing process. Similar to advances and developments in the way to deal with large quantities of data in other fields, new data sources in the transport domain necessitate the development of innovative methodologies and processes for their utilization.

The challenge nowadays is not merely about the collection of this data itself; it is rather about processing and managing the potential that is inherent in big data. There is therefore a

need for developing tools that can filter, fuse, and process data from various sources at real time, in order to provide reliable Advanced Traveller Information Services.

The aim of this paper is to aid research in this direction, by explicitly laying out current practice regarding the collection, filtering, fusion, analysis, and utilization of big data, with the city of Thessaloniki serving as a case study. The remaining of the paper is structured as follows; in Section 2 big data sources for the transportation and mobility sector are presented and discussed, while Section 3 illustrates the big data collection and processing framework for the city of Thessaloniki. The paper concludes with Section 4, where recommendations for future research directions are presented.

2. Review on Transport and Mobility Data Sources

2.1. Conventional Traffic Data. Conventional data sources mostly aim at collecting aggregated data through a large variety of intrusive (data recording sensors placed on or in the

road) and nonintrusive sensors (remote observations), such as pneumatic tubes, infrared and acoustic devices, inductive loops, radars, or cameras, mainly measuring traffic flow, speed, and occupancy of the road infrastructure. Reference [1] presents the limitations of conventional traffic sensors, concluding that the main limitations are related to installation and calibration costs, low coverage, and low performance under adverse weather conditions, while [2] presents their lifetime duration and costs, concluding that intrusive sensors have an average lifetime of 5 years and low maintenance costs, while nonintrusive sensors have double lifetime and high purchase and installation costs.

2.2. Probe Data. Probe data refers to data being collected directly from individual travellers or vehicles, which are usually aggregated during the processing phase. This allows for increased capabilities of data processing but at the same time requires larger processing capabilities and computing resources. Probe data can be classified into stationary and floating.

Stationary probe data is collected at fixed locations within a transport network by tracking unique characteristics of individuals or vehicles, such as MAC (Media Access Control) identities of Bluetooth devices, or through Automatic Plate Number Recognition (APNR) systems. These detectors are often associated with high costs, calibration needs, low performance under high traffic flow conditions, or adverse weather conditions and/or low market penetration rate.

On the contrary, floating probe data is collected along the trajectories of moving objects (individual travellers or vehicles) equipped with a smart device able to track its own trajectory through GPS or A-GPS (assisted GPS). They are associated with accuracy considerations (related to the devices' exact GPS location), which are often relaxed by map-matching the position to a representation of the road network and to the frequency of the data collection.

A few examples of the use of mobile data for understanding activity chains and trajectories are presented in [3, 4] while Bluetooth and Floating Car Data are used for the same purpose in [5, 6], respectively.

2.3. Data from Social Media. Social media is a new data source with high potential, due to its relation with the daily lives of individuals, but at the same time it is the most difficult to process due to its unstructured format (free text) and its lack of geo-reference (tweets) or low granularity (Facebook check-in events). The major drawback associated with social media data collection is the biased sample used for processing, since the content generators are not representative of the whole population or activities, especially for dedicated services such as the check-in service of Facebook.

3. Methodology

3.1. Thessaloniki's Large-Scale Mobility Model. The urban mobility model for Thessaloniki includes 47807 intersections (with detailed information about the geometry and control type) and 137854 directed road segments (with length,

number of lanes, free flow speed, capacity, direction, allowed transportation modes, existence of dedicated lanes, and parking regulation). Links are classified into 6 categories, depending on their capacity (which has been proportionally reduced, based on the use of bus lanes and/or parking spaces) and free flow speed. Each road category has a unique Volume-Delay Function (VDF), based on the BPR (Bureau of Public Roads [7]).

The network consists of 359 traffic analysis zones (306 describing the metropolitan area of Thessaloniki and the rest representing external zones) and 3508 connectors. The demand side consists of 24 hourly Origin-Destination matrices for private vehicle trips, estimated using household and road side surveys ([8, 9]).

3.2. Collection of Big Transport and Mobility Data. Thessaloniki is the largest city of Northern Greece and the second in the whole Greece, with a total of more than 1 million citizens in its greater area, which covers 1500 km². It has an average density of 665 inhabitants per km² and the total number of vehicles exceeds 777000, including private cars, heavy vehicles, and motorcycles [8]. Conventional sensors (loops, radars, and cameras), probe data (stationary and floating), and social networks derived content are utilized in the city for processing mobility-related indicators, as well as for providing Advanced Traveller Information Services (real-time travel time and vehicle speed estimation).

3.2.1. Conventional Traffic Data Sources. There are three sets of conventional mobility data sensors in Thessaloniki (Figure 1):

- (i) The surveillance system of the Peripheral Ring Road consisting of 9 cameras (CAM in Figure 1) monitoring 100000 vehicles per day (as of 2011) in both directions.
- (ii) The Thessaloniki's Urban Mobility Management System, consisting of 12 cameras and 7 radars (CAR in Figure 1), installed at the city's busiest arterial, monitoring 50000 vehicles per day (as of 2012) in one direction.
- (iii) The traffic lights management system, consisting of 503 Inductive Loop Detectors (ILD in Figure 1) installed at traffic light controlled intersections in the city, monitoring 250000 vehicles per day.

3.2.2. Stationary Probe Data. The Bluetooth (BT) detectors network is comprised of 43 roadside devices, installed at selected major intersections [10] throughout the road network of the city, as shown in Figure 2.

More than 100000 BT-equipped devices are detected every day, generating on average a total of 300000 detections at the 43 locations.

3.2.3. Floating Probe Data. The floating probe data are derived from a 1200 taxi vehicles' fleet, circulating 14 hours on average per day and generating pulses every 6 seconds, which contain their location and instantaneous speed. The

TABLE 1: Relation between detection and traffic flow for different filters.

Time interval used for data filtering	Without filtering	5min filter	15min filter	60min filter
Correlation coefficient	0.3682	0.2179	0.1972	0.0442
R^2	0.8947	0.9193	0.9337	0.8594
Largest differences (absolute value and percentage ranges)	-401 / 623 vehicles -26% / 75%	-410 / 437 vehicles -23% / 61%	-336 / 389 vehicles -22% / 57%	-536 / 767 vehicles -35% / 79%

FIGURE 1: Mobility sensors in the city of Thessaloniki.

data collected and processed reaches on average 2500 pulses per minute, with daily totals at approximately 1.5 million.

3.3. Processing of Big Data. Figure 3 depicts the framework for real-time data processes and services for the city of Thessaloniki. It is composed of the data collection from the real world as well as the offline and real-time processes, which culminates in the provision of real-time mobility services.

The data collected by the above-mentioned sensors and systems require filtering and processing, in order to accurately estimate mobility indicators and provide Advanced Traveller Information Services in real time. This requests a trade-off between the quality of the analyses (time consuming) and the real-time provision of the services. The main traffic and mobility indicators estimated for the city of Thessaloniki are travel times along the main routes of the city, traffic flows in the prioritized road network, and traffic congestion levels.

The individual components that have been developed and are presented in this paper are the following:

- (i) Traffic flow estimation based on stationary probe data
- (ii) Travel time estimation based on stationary probe data
- (iii) Traffic flow estimation based on travel time
- (iv) Short-term traffic flow prediction
- (v) Spatial expansion of traffic flow
- (vi) Real-time traffic conditions estimation

3.3.1. Traffic Flow Estimation from Stationary Probe Data.

Figure 4 depicts unfiltered traffic flow measurements (dashed line) and BT detections (straight line) in one of the busiest (in terms of traffic flow) intersections of the city for a 7-day period. It can be observed that the detections fit well with the traffic flow modifications, with a small deviation for large values of traffic flow, which may be improved by filtering properly the data. Details of this analysis can be found in [11].

Traffic flow is extrapolated directly from aggregated detections of devices equipped with BT. The reliability and accuracy of this methodology largely depend on the penetration of the Bluetooth technologies to travellers and vehicles. Statistical and comparative analyses conducted for the case of Thessaloniki reveal an R^2 deviation between traffic flow measurements and BT detections equal to 0.91, reflecting the methodology's potential.

Figure 5 presents the relation between detected BT devices and measured traffic flow at a single intersection. As observed, there is a strong correlation between the two with an R^2 relation equal to 0.89 (prior to data filtering). In addition, 37% of the traffic flow is detected by BT detectors. It can be observed that the relation is different in low and in high values of traffic flow volume, which is also observed in Figure 4.

After filtering the BT detections and removing double entries (related to nonmoving objects) during predefined time intervals, the corrected correlation between detected and measured vehicles is approximately 20%. Table 1 summarizes the impact of using different time intervals for filtering the data (time duration during which detected objects are classified as nonmoving) for the same intersection used for obtaining Figures 4 and 5 during a single day. The best fit is obtained with the 15-minute filter.

3.3.2. Travel Time Estimation. Floating probe data is used in a similar way for estimating travel times [12] and traffic states ([13–21]). Two consecutive pulses from moving objects generate 100-meter long trajectories, which are matched with virtual locations within the network, in order to generate “detections” and match them for estimating travel times. Only a part of the filters applied to the stationary probe data is

FIGURE 2: BT detectors network in the city of Thessaloniki.

FIGURE 3: Framework for real-time data processes and services in Thessaloniki.

FIGURE 4: Detected BT devices and measured traffic flow versus time of day.

FIGURE 5: Relation between detected BT devices and measured traffic flow without filtering the detections at a single intersection.

applied, since the sample is “controlled” (all the data is related to vehicles).

Alternatively, individual travel times are estimated by matching the detected MAC address identities at various locations of the road network and comparing the timestamps, in order to obtain individual travel times, which are aggregated and filtered to obtain average travel times. The aggregation of the individual travel times allows filtering them, eliminating outliers and noise, such as en-route stops, trip ends between detectors, detours deviating from the predefined paths, or records from other transportation modes, such as pedestrians, bicycles, and users of the public transport system as well as atypical vehicles such as couriers or delivery vehicles. The methodological flowchart can be found in [22].

3.3.3. Traffic Flow Estimation Based on Travel Time. The estimated travel times along various routes are converted to link travel times and to link traffic flows, according to the process depicted in Figure 6.

As observed, the measurements of route travel times (based on probe data) and individual link travel times (from conventional sensors/detectors) are used for the correction of the link travel times which belong to those routes. The link traffic flow measurements (based on both probe data and conventional sensors/detectors) are also used for the estimation of the link travel times. Finally, the link travel times are converted to link traffic flows (with the use of volume-delay functions) providing a unique relation between traffic volume and speed on a specific link for determining the spatial evolution of traffic flows and merged with the links having traffic flow measurements (stand-alone links).

The conversion from route travel time to link travel time for the links which constitute the route is conducted using the following mathematical minimization problem:

$$\begin{aligned} \min \quad & \delta_1 \sum (Ax - b) + \delta_2 \sum \left(\frac{(x - v_0)}{v_0} \right) \\ \text{S. t.} \quad & x_i \geq t_{0i}, \quad \forall i \in I \\ & x_j = t_j, \quad \forall j \in J \end{aligned} \quad (1)$$

where

δ_1 and δ_2 are weights for the minimization problem (0.6 and 0.4, respectively),

A is the link-path incidence matrix (α_{ij} is 1 if the link j is part of path i),

x is the link travel times vector (x_i is the travel time of link i),

b is the paths travel times vector (b_j is the travel time of path j),

v_0 is the initial link travel times vector derived from the traffic assignment model (v_{0i} is the initial travel time of link i as in the traffic assignment model),

t_{0i} is the free flow travel times of link i ,

FIGURE 6: Traffic flow estimation procedure from path travel times.

t_j is the observed travel times of link j (if measurements of traffic flow exist, the conversion takes place based on the VDF),

I is the set of all links which constitute the paths,

J is the subset of I of the links for which observations exist.

The objective function consists of two parts: the first part contains the estimation of the difference between the path travel time and the sum of the link travel times which define the path. In the second part, the percentile difference is estimated between the link travel times derived from the traffic assignment model and the estimated travel time, in order to avoid burdening links with very small travel times, which would excessively increase the traffic flow.

The first constraint ensures that the minimum link travel time cannot be smaller than the free flow link travel time. The second constraint ensures that the link travel time used for links for which travel times are measured is equal to the measured travel time.

3.3.4. Short-Term Traffic Flow Prediction. Various studies have been developed for forecasting traffic state, flow, and patterns ([23–29]) but most of them are applied to freeways.

For the short-term traffic flow prediction used in this framework, a linear autoregressive (AR of degree up to 3, depending on the time series created at each location) model is used. The degree of each model depends on the time series of the previous time periods and is formulated as follows:

$$\varphi_k^{i+1} = \bar{\varphi}_k + \sum_{j=1}^N \beta_k^j * (\varphi_k^{i+1-j} - \bar{\varphi}_k) \quad (2)$$

where

φ_k^i is the traffic flow of link k for time period i ,
 $\bar{\varphi}_k$ is the average traffic flow of link k for the time period of the last two days,

β_j is the serial correlation coefficient j of link k ,

N is the degree of the autoregressive model.

3.3.5. Spatial Expansion of Traffic Flows. Various data fusion techniques have been developed for fusing multisource traffic data ([30, 31]). In the proposed framework, the spatial expansion of traffic flows is based on the Data Expansion Algorithm (DEA) ([32]). The implementation for Thessaloniki's road network differs in that the link traffic flows are used as a constraint and the optimization regards the minimization of the differences of the distributions of vehicles on intersections (the original DEA does it in the opposite way). The mathematical formulation is presented below:

$$\begin{aligned} \min \quad & (D - p * C_n) * x - (p * C_c * c) \\ \text{s.t.} \quad & x_i > lb_i \quad \forall i \in I \\ & x_i < ub_i \quad \forall i \in I \\ & x_j = f_j \quad \forall j \in J \end{aligned} \quad (3)$$

where

D is diagonal unitary matrix (with dimensions equal to the total number of links),

p is the vector of the percentages of traffic flows on intersection (junction) level which is distributed at every link (p_{ij} is the percentage of traffic at intersection i derived to link j),

C_n is the adjacency table at link level (C_{nij} is 1 if the link j precedes link i),

x is the vector of traffic flows (x_i is the traffic flow at link i),

C_c is the incidence matrix (C_{cij} is 1 if the connector i ends or starts at intersection j),

c is the vector of traffic flows found at the connectors (c_i is the traffic flow at connector i),

lb is vector of the minimum traffic flow values (lb_i is the minimum traffic flow at link i),

ub is the vector with the maximum traffic flow values (ub_i is the maximum traffic flow at link i),

f is the vector with the measured traffic flows (f_i is the measured traffic flow at link i),

I is the set of all links,

J is the subset of I for which traffic flow measurements exist.

The objective function consists of two parts: in the first part the adjacency matrix is estimated at link level. In the adjacency matrix, all the percentages of traffic flows from every link to all other links are included. In the second part, the traffic flow of the connector links is estimated.

The first and second constraints ensure the minimum and maximum traffic flow for each link, while the third constraint ensures that the flow on links for which measurements exist is the measured flow.

3.3.6. Real-Time Traffic Conditions Estimation. Based on the data and processes described above, traffic congestion is estimated at real time and forecasted in a short-term basis. Both travel time measured at selected routes using BT detectors and floating probe data are converted into traffic flow by using volume-delay functions. All measurements and estimations are forecasted by an autoregressive model for 15, 30, 45, and 60 minutes and the process is repeated, in order to obtain traffic flows for the forecasted scenarios. Finally, all traffic flows are merged into a nonlinear mathematical program by means of a modified DEA, resulting in link flows, travel times, and average speed values for the entire network (with the use of link-based volume-delay functions).

3.4. High Performance Computing in Thessaloniki. Transport and mobility data sets coming from multiple innovative and conventional data sources like those described in Section 2 consist of both structured and unstructured data that usually are relatively large and also are being produced at very high rates. On the other hand, the corresponding monitoring and processing applications typically utilize complex algorithms that require the acquisition, transformation, combination, and processing of those data sets in almost real time in an efficient way. As the technology evolves, these requirements lead to a need of constant improvement of the software infrastructure, procedures, and data pipelines used by the applications.

In this context, it becomes clearer that mature and cutting-edge technologies can be used together, in order for those data sets and streams to be handled efficiently. In the current architecture, a set of applications filters and transforms the raw data coming from different data sources and the result is stored in a relational database management systems and exposed via unified, compatible, and secure APIs. As many new open source software frameworks are made available to process big datasets, in batches or streams, and store the result in distributed storage systems, the architecture

is being evolved towards a microservice architecture. The aim of the evolution is on one side to address the scalability and availability issues of the overall architecture and on the other side to implement the most recent software engineering best practices for the development and deployment of software components leveraging containerization technologies such as Docker. Evolving a software architecture involves both the technologies and the people who use them. As it is common in different research and production environments, new software tools and technologies must be integrated with some of the existing components and software packages.

A new architecture for the ingestion, processing, and storage of the traffic data is being developed within the Big Data Europe project. The project aims at developing a platform where big data frameworks can be easily integrated to build distributed data processing pipelines. The platform usage will be demonstrated in all the seven societal challenges that have been addressed by the European Commission's H2020 programme among which the fourth is about smart, green, and integrated transport. The architecture is based on a messaging system through which data producers and consumers can communicate using different channels. A data producer reads the data from a source, applying a transformation in order to use a common format and semantics that is known to the consumers, and writes the data into a channel. A consumer, subscribed to that channel, reads the data and applies some specific transformations before sending the result to another channel for further processing or to a sink. Producers and consumers are jobs that can be deployed in different nodes in order to parallelize the computation. The architecture is based on Apache Kafka for the message passing, on Apache Flink, Rserve, and Postgis for processing real time data streams and historical data, and on Elasticsearch for the storage of the results. All the components are deployed as Docker containers in a distributed environment. The platform is being tested using the Floating Car Data from Thessaloniki. The raw data containing the timestamp, the location of the vehicles, orientation, and speed is transformed and map matched to the road segments and then aggregated in time windows to compute the average speed and traffic flow.

During the deployment phase of the transport pilot, some of those components will be externalized in virtual machines, relying upon cloud based High Performance Computing provided by GRNET, the national research infrastructure provider.

4. Conclusions and Further Work

This paper presented the framework for big transport and mobility data collection and processing applied in Thessaloniki. Recent technological advances allow for shifting from aggregated to disaggregated data collection at individual traveller level, which increases significantly on the one hand the needs for data processing and on the other hand the reliability of the Advanced Traveller Information Services.

These result in a need for new postprocessing tools, which should be able to perform in real time. Also, new methodologies are needed, in order to take advantage of the technologies and new societal behaviour patterns for

increasing the accuracy and performance of the existing transport related information services and the creation of new traveller services.

Finally, future research directions should cover the topics of multisource mobility data exploitation, as well as fusion with data coming from social networks.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The Bluetooth data is processed at the GRNET infrastructure as part of the “Big Data Warehouse for Mobility (BD W4M)” project funded by the “2007-2013 NSRF programme for development”. Some of the tools for handling and processing the big data used have been developed in the framework of the Big Data Europe project (644564) funded under the H2020 programme.

References

- [1] L. S. Skaszek, *State of the Art, Report on Non-Traditional Traffic Counting Methods*, Arizona Department of Transportation & U.S. Department of Transportation Federal Highway Administration, 2001.
- [2] G. Leduc, “Road traffic data: collection methods and applications,” *Working Papers on Energy, Transport and Climate Change* 55, 2008.
- [3] K. Leetaru, S. Wang, G. Cao, A. Padmanabhan, and E. Shook, “Mapping the global Twitter Heartbeat: The geography of Twitter,” *First Monday*, vol. 18, June 2013.
- [4] M. Zilske and K. Nagel, “A simulation-based approach for constructing all-day travel chains from mobile phone data,” in *Proceedings of the The International Conference on Ambient Systems, Networks and Technologies, ANT-2015, the International Conference on Sustainable Energy Information Technology, (SEIT '15)*, vol. 52, pp. 468–475, June 2015.
- [5] J. Novak, R. Ahas, A. Aasa, and S. Silm, “Application of mobile phone location data in mapping of commuting patterns and functional regionalization: A pilot study of Estonia,” *Journal of Maps*, vol. 9, no. 1, pp. 10–15, 2013.
- [6] İ. Erkan and H. Hastemoglu, “Bluetooth as a traffic sensor for stream travel time estimation under Bogazici Bosphorus conditions in Turkey,” *Journal of Modern Transportation*, vol. 24, no. 3, pp. 207–214, 2016.
- [7] G. Azar David, “Louisiana Department of Highways: Bureau of Public Roads, 1969”.
- [8] E. Mitsakis, I. Stamos, J. M. Grau, E. Chrysochoou, P. Iordanopoulos, and G. Aifadopoulou, “Urban Mobility Indicators for Thessaloniki,” *Journal of Traffic and Logistics Engineering*, vol. 1, no. 2, pp. 148–152, 2013.
- [9] E. Mitsakis, J. M. S. Grau, I. Stamos, and G. Aifadopoulou, “An integrated framework for embedding large-scale dynamic traffic assignment models in advanced traveler information systems,” *Transportation Planning and Technology*, vol. 38, no. 8, pp. 866–877, 2015.
- [10] E. Mitsakis, E. Chrysochoou, J. M. Salanova Grau, P. Iordanopoulos, and G. Aifadopoulou, “The sensor location problem: methodological approach and application,” *Transport*, vol. 32, no. 2, pp. 113–119, 2017.
- [11] J. M. Salanova Grau, I. Stamos, E. Mitsakis, D. Margaritis, P. Tzenos, and G. Aifadopoulou, “Estimation of traffic flow from Bluetooth detections,” in *Transportation Research Board 95th Annual Meeting*, Wash, USA, 2017.
- [12] W. Shi, Q. Kong, and Y. Liu, “A GPS/GIS Integrated System for Urban Traffic Flow Analysis,” in *Proceedings of the 11th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pp. 844–849, Beijing, China, October 2008.
- [13] M. Fountoulakis, N. Bekiaris-Liberis, C. Roncoli, I. Papamichail, and M. Papageorgiou, “Highway traffic state estimation with mixed connected and conventional vehicles: Microscopic simulation-based testing,” in *Proceedings of the 19th IEEE International Conference on Intelligent Transportation Systems, (ITSC '16)*, pp. 1761–1766, November 2016.
- [14] N. Bekiaris-Liberis, C. Roncoli, and M. Papageorgiou, “Highway traffic state estimation with mixed connected and conventional vehicles,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 12, pp. 3484–3497, 2016.
- [15] N. Bekiaris-Liberis, C. Roncoli, and M. Papageorgiou, “Highway traffic state estimation per lane in the presence of connected vehicles,” *Transportation Research Part B: Methodological*, vol. 106, pp. 1–28, 2017.
- [16] M. Fountoulakis, N. Bekiaris-Liberis, C. Roncoli, I. Papamichail, and M. Papageorgiou, “Highway traffic state estimation with mixed connected and conventional vehicles: Microscopic simulation-based testing,” *Transportation Research Part C: Emerging Technologies*, vol. 78, pp. 13–33, 2017.
- [17] J. C. Herrera and A. M. Bayen, “Incorporation of Lagrangian measurements in freeway traffic state estimation,” *Transportation Research Part B: Methodological*, vol. 44, no. 4, pp. 160–481, 2010.
- [18] T. Seo, A. M. Bayen, T. Kusakabe, and Y. Asakura, “Traffic state estimation on highway: A comprehensive survey,” *Annual Reviews in Control*, vol. 43, pp. 128–151, 2017.
- [19] T. Seo, T. Kusakabe, and Y. Asakura, “Traffic state estimation with the advanced probe vehicles using data assimilation,” in *Proceedings of the 18th IEEE International Conference on Intelligent Transportation Systems, (ITSC '15)*, pp. 824–830, Spain, September 2015.
- [20] Z. Shan, Y. Wang, and Q. Zhu, “Feasibility study of urban road traffic state estimation based on taxi GPS data,” in *Proceedings of the 2014 17th IEEE International Conference on Intelligent Transportation Systems, (ITSC '14)*, pp. 2188–2193, October 2014.
- [21] Q.-J. Kong, Q. Zhao, C. Wei, and Y. Liu, “Efficient traffic state estimation for large-scale urban road networks,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 1, pp. 398–407, 2013.
- [22] E. Mitsakis, J. M. Salanova Grau, E. Chrysochoou, and G. Aifadopoulou, “A robust method for real time estimation of travel times for dense urban road networks using point-to-point detectors,” *Transport*, vol. 30, no. 3, pp. 264–272, 2015.
- [23] T. L. Pan, A. Sumalee, R. X. Zhong, and N. Indra-Payoong, “Short-term traffic state prediction based on temporal-spatial correlation,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1242–1254, 2013.
- [24] P. Zhang, L. Rui, X. Qiu, and R. Shi, “A new fusion structure model for real-time urban traffic state estimation by multi-source traffic data fusion,” in *Proceedings of the 18th Asia-Pacific*

Network Operations and Management Symposium, (APNOMS '16), October 2016.

- [25] M. Lippi, M. Bertini, and P. Frasconi, "Short-term traffic flow forecasting: an experimental comparison of time-series analysis and supervised learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 2, pp. 871–882, 2013.
- [26] Y.-S. Jeong, Y.-J. Byon, M. M. Castro-Neto, and S. M. Easa, "Supervised weighting-online learning algorithm for short-term traffic flow prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 4, pp. 1700–1707, 2013.
- [27] T. Thomas, W. Weijermars, and E. Van Berkum, "Predictions of Urban volumes in single time series," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 1, pp. 71–80, 2010.
- [28] B. Ghosh, B. Basu, and M. O'Mahony, "Multivariate Short-Term Traffic Flow Forecasting Using Time-Series Analysis," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 2, pp. 246–254, 2009.
- [29] Z. Hou and X. Li, "Repeatability and Similarity of Freeway Traffic Flow and Long-Term Prediction under Big Data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 6, pp. 1786–1796, 2016.
- [30] E. Cipriani, S. Gori, and L. Mannini, "Traffic state estimation based on data fusion techniques," in *Proceedings of the 15th International IEEE Conference on Intelligent Transportation Systems, (ITSC '12)*, pp. 1477–1482, September 2012.
- [31] N. Zhang, J. Xu, P. Lin, and M. Zhang, "An approach for real-time urban traffic state estimation by fusing multisource traffic data," in *Proceedings of the 10th World Congress on Intelligent Control and Automation, (WCICA '12)*, pp. 4077–4081, July 2012.
- [32] R. Lederman and L. Wynter, "Real-time traffic estimation using data expansion," *Transportation Research Part B: Methodological*, vol. 45, no. 7, pp. 1062–1079, 2011.