

Observability based data-fusion cascading filtering for urban network flow estimation

Rinaldi, Marco; Viti, Francesco

DOI

[10.1109/MT-ITS49943.2021.9529302](https://doi.org/10.1109/MT-ITS49943.2021.9529302)

Publication date

2021

Document Version

Final published version

Published in

2021 7th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)

Citation (APA)

Rinaldi, M., & Viti, F. (2021). Observability based data-fusion cascading filtering for urban network flow estimation. In *2021 7th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS): Proceedings* Article 9529302 IEEE. <https://doi.org/10.1109/MT-ITS49943.2021.9529302>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

Green Open Access added to TU Delft Institutional Repository

'You share, we take care!' - Taverne project

<https://www.openaccess.nl/en/you-share-we-take-care>

Otherwise as indicated in the copyright section: the publisher is the copyright holder of this work and the author uses the Dutch legislation to make this work public.

Observability based data-fusion cascading filtering for urban network flow estimation

1st Marco Rinaldi

*Department of Transport & Planning
Faculty of Civil Engineering and Geosciences, TU Delft*
Delft, The Netherlands
m.rinaldi@tudelft.nl

2nd Francesco Viti

*Department of Engineering
University of Luxembourg*
Esch-sur-Alzette, Luxembourg
francesco.viti@uni.lu

Abstract—In this work we extend our previously proposed cascading Kalman filtering technique, applied to the problem of urban network flow estimation, to adopt heterogeneous traffic data sources. Both static infrastructure detection (double induction loops) and Floating Car Data are collected from a given transportation network, and employed within separate stages of the cascading technique. The proposed approach relies upon notions of traffic flow inference (observability) to both i) determine the optimal set of locations in which sensors should be installed and ii) provide enhanced covariance information within the estimation technique. The impact of both penetration rates of Floating Car Data and sensor selection procedure is evaluated empirically, through a microscopic simulation software (SUMO) generating experimental data on a simple grid-like network.

Test results showcase that the proposed extension to the cascading framework is indeed beneficial in reducing the overall estimation error on network segments where static infrastructure is unavailable. Furthermore, the importance of observability-based sensor locations is clearly demonstrated.

Index Terms—Traffic Flow Estimation, Observability, Floating Car Data, Urban Networks

I. INTRODUCTION & LITERATURE REVIEW

In the last decades, development of Intelligent Transportation System applications has witnessed a dramatic rise, supported by novel information and communication technologies. These advanced approaches, ranging e.g. from Travel Time and Route Advisory Systems to Dynamic Traffic Management, operate under the assumption that complete knowledge of the traffic state is available, in order to correctly determine the appropriate action to be taken at any given time.

Collecting the required data entails however sizeable costs: on the one hand, a considerable trade-off arises between reliability and precision of measurements and cost of the required technological infrastructure, as high-precision sensors are both expensive to install and to maintain [1]; on the other hand, ensuring complete coverage through measurements alone is economically impractical, due to the size and complexity of transportation networks, especially at the urban level [2]. In order to strike a reasonable trade-off between data requirements and collection costs, state estimation techniques such as

Kalman Filtering (KF) have been widely adopted in the past, so to reconstruct information on portions of the network which are not directly measured [3]–[8], leveraging appropriate modelling techniques in order to estimate traffic propagation phenomena. Furthermore, multiple sources of traffic data, such as double induction loop detectors, V2X technologies (Floating Car Data (FCD), Bluetooth data) and Automatic Number Plate Recognition (ANPR) have been combined through data fusion approaches in order to infer a comprehensive network state [4], [9]–[11]. Such techniques have found considerable success both in freeway and arterial/urban corridor state estimation.

When extending these approaches towards urban network flow estimation, careful consideration must however be taken with respect to the additional complexities arising due to the effects of congestion. Rerouting dynamics introduce nonlinearities and additional dependencies between topologically distant portions of the network [12], which might lead propagation-based predictions astray, as well as violate the assumptions required to ensure that the error minimising properties of Kalman Filtering are maintained.

In this work we continue our recent efforts [13] in attempting to adapt Kalman Filtering techniques to the complex topological landscape of urban networks. As in our previous manuscript, we rely on the concept of network flow observability [14], [15] in order to guide both the optimal selection of candidate locations for fixed sensing infrastructure (double induction loop detectors) and provide enhanced correlation information for the filtering techniques.

We claim the following original contributions for this work: i) we extend our previously introduced cascading state estimation approach to include Floating Car Data in both the estimation of link travel times and, more importantly, node turning fractions; ii) we evaluate the impact of road sensor locations and observability information on the estimation error, with and without the employment of FCDs.

Validation is carried out on synthetic data obtained through microscopic traffic simulation (DLR SUMO, [16]), based upon a simple grid-like network exhibiting significant congestion and rerouting patterns. Test results showcase major improvement in terms of estimation quality thanks to the addition of FCD data, successfully reducing estimation error throughout the transportation network. Moreover, the impact

This research is supported by the University of Luxembourg's IRP ConTraNet AGR-3132.

978-1-7281-8995-6/21/\$31.00 ©2021 IEEE

of observability information on estimation quality, in terms of both locations and amount of sensors, is clearly illustrated.

The rest of the paper is structured as follows: in Section II we introduce the improved cascading Kalman Filtering framework, specifically detailing the travel-time estimation module (T-Model) and the node turning fraction estimation module (TF-Model), as well as the required variations on the overall cascading scheme. Section III introduces the experimental setup and discusses the test case results. Finally, we draw conclusions and remarks for future research in Section IV.

II. METHODOLOGY

In this Section we detail the updated components of our cascading filtering framework, as well as the required adaptations to the scheme's structure. For the sake of clarity we adopt the notation used in our previous approach [13], briefly summarized as follows. We represent a given transportation network through a directed graph $G(N, L)$ with N the set of nodes and L the set of arcs connecting said nodes. For each link $l \in L$ we collect four key measurements for each simulation time step k : two stationary detection measurements, link flows $f(k) \in \mathbb{R}^{|L| \times 1}$ [veh/h] and (harmonised) link speeds $v(k) \in \mathbb{R}^{|L| \times 1}$ [km/h], and two measurements based on vehicular Floating Car Data, FCD vehicle-link matching positions [veh/link] and FCD link speeds $v_{FCD}(k) \in \mathbb{R}^{|L| \times 1}$ [km/h]. FCD vehicle-link matching positions are converted to FCD-based link flow estimates $f_{FCD}(k) \in \mathbb{R}^{|L| \times 1}$ [veh/h] by re-sampling the FCD link counts appropriately, considering the (assumed known and fixed) FCD polling rate. Penetration rates are assumed unknown, and both T-Model and TF-Model are designed to operate independently from this parameter. Topological characteristics of each link are also assumed known, such as free flow speed v_{lff} [km/h], capacity c_l [veh/h] and length s_l [km]. To distinguish between measured, predicted and estimated variables we adopt the following symbols respectively: $\hat{\square}$ (measured variable), \square^- (predicted variable), $\hat{\square}$ (estimated variable).

A. Link travel-time estimation: the updated T-Model

We extend our previously proposed T-Model by employing FCD speed measurements $\tilde{v}_{FCD}(k)$ alongside harmonised link speeds $\tilde{v}(k)$ in a sensor fusion KF. Specifically, we enrich measured harmonised link speeds with free flow information, and combine this information with FCD measurements in the KF's prediction step. We begin by computing an estimate of link travel times $\hat{\tau}_l(k) = A^\tau \cdot \hat{\tau}_{lff} + B \cdot \tilde{\tau}_l(k)$ where $\hat{\tau}_{lff} = s_l/v_{lff}$ is the topological free-flow link travel time and $\tilde{\tau}_l(k) = s_l/\tilde{v}_l(k)$ the current measured link travel time based on loop detected data. As in our previous work, matrix $A^\tau \in \mathbb{N}^{|L| \times |L|}$ bears elements $a_{ll} = 1$ if link l is unmeasured and zero otherwise, while matrix $B \in \mathbb{N}^{|L| \times |L|} = (I - A)$ operates in the exactly opposite fashion. The a-priori link travel time estimate is therefore based on harmonised link speeds collected by loop detectors, where available, while assuming free flow travel times elsewhere. The prediction step of the T-Model is as follows:

$$\tau_l^-(k) = 0.5 \cdot \hat{\tau}_l(k) + 0.5 \cdot \tilde{\tau}_{lFCD}(k) \quad (1)$$

$$P_K^{-,\tau} = P_{K-1}^\tau + Q^\tau \quad (2)$$

where matrix P_K^τ is the predicted filter's covariance matrix, initially set to $P_0^\tau = 0.1 \cdot I$. This step essentially performs a 50-50 split sensor fusion between stationary detected harmonic link speeds, where available, and FCD-based average link speeds.

The Kalman Filter gain and estimate filter's covariance matrices can then be computed as follows:

$$K_K^\tau = \frac{P_K^{-,\tau}}{P_K^{-,\tau} + R^\tau} \quad (3)$$

$$P_K^\tau = (I - K_K^\tau) P_K^{-,\tau} \quad (4)$$

Finally, the error correction step can be performed as follows:

$$\hat{\tau}_l(k) = \tau_l^-(k) + K_K^\tau \cdot (\tilde{\tau}_{lFCD}(k) - \tau_l^-(k)) \quad (5)$$

The updated T-Model adjusts therefore the predicted link travel times $\tau_l^-(k)$ assuming that the average speeds measured through Floating Car Data offer a higher level of precision with respect to the data collected by the sparsely available double loop detectors in the network. The prediction process noise's covariance matrix is chosen as $Q^\tau = 10^4 \cdot I$. To populate the measurement noise's covariance matrix R^τ we employ link-to-link incidence information such that $r_{ij} = 1 \forall (i, j) : \exists n \in N : l_i \in out(n) \cup l_j \in in(n)$, that is an element (i, j) of R^τ is set to 1 if there is a node n such that link l_i is part of its incoming link set $in(n)$ and l_j is part of its outgoing link set $out(n)$. The order of magnitude difference between the two covariance matrices' elements has been chosen to imply we trust the collected data considerably more than the model's prediction.

B. Node turning ratios estimation: the updated TF-Model

Compared to our previous contribution, in this work we rely on the FCD-based link flow measurements $\tilde{f}_{FCD}(k)$ to estimate the node turning ratios at each intersection. To achieve this objective, we employ a sensor fusion KF to augment the measured link flows $f_m(k), \forall m \in M$ with FCD measurements, where M is the set of links $l \in L$ for which a loop detector is installed and $U = L \setminus M$ the unmeasured link set.

As in our previous work, we begin by deriving the observability matrix Ψ :

$$\begin{pmatrix} \tilde{f}_m(k) \\ f_u^-(k) \end{pmatrix} = \begin{pmatrix} I & 0 \\ \Psi & 0 \end{pmatrix} \cdot \begin{pmatrix} \tilde{f}_m(k) \\ f_u^-(k) \end{pmatrix} \quad (6)$$

We classify four potential sources of link flow information for each link: measured link flows $\tilde{f}_m(k)$, FCD-estimated flows on directly measured links $\tilde{f}_{mFCD}(k)$, FCD-estimated flows on unobserved links $\tilde{f}_{uFCD}(k)$ and observability-based link flow estimates $f_u^-(k) = \Psi \cdot \tilde{f}_m(k)$. Each source of information employed for data fusion is assigned a level of trust

are collected once the Dynamic User Assignment *dualterate* tool has completed simulation. Alongside these fixed sensing infrastructure measurements, vehicular Floating Car Data, containing vehicle-link matches (trajectories) and speeds, are collected at a refresh rate of 30s. The cascading KF procedure is implemented in MathWorks® MATLAB™.

Before their inclusion in the cascading filtering framework, FCD outputs are pre-processed in order to convert vehicle-link matches to link flows and speeds. Specifically, vehicle counts are accumulated on a link-by-link basis at the FCD data's original refresh rate, the resulting flows in [veh/30s] are then smoothed through a five-timestep Moving Average (MA) filter and finally undersampled in order to align the sampling rate to that of loop detectors. Similarly, FCD-based vehicle speeds are averaged on a link-by-link basis, before MA filtering and undersampling. Loop detector outputs are equally smoothed through a five-step MA filter. Finally, estimated demand $\hat{d}(k)$ is computed considering the total demand as uniformly distributed throughout the complete simulation horizon of 2h. In order to ensure consistency of route choice, a total number of 100 iterations is utilised.

To assess the impact of sensor locations on the proposed estimation framework, we adopt two random and one deterministic location selection approaches: a *pure random* approach for which both the amount and the locations of sensors to be placed on the network are drawn from discrete uniform distributions $U\{1, 36\}$; a *random loc.* approach, for which while the locations of sensors are drawn from a uniform distribution $U\{1, 36\}$, the amount of draws is assumed constant and equal to that known to yield full observability and, finally, the *observability* deterministic approach, based on [17], which yields the sensor locations shown in green in Figure 1. A total of 10 random draws are performed. When considering randomly generated locations, other than adapting the data sources themselves, unobtainable observability information is omitted from both the TF-Model (Ψ is assumed an all-zeroes matrix of appropriate dimensions) and F-Model (the structure of the filter design matrix Q^F is adapted accordingly i.e. a diagonal matrix of appropriate dimensions).

In order to showcase the impact of the newly included FCD data, as well as the importance of appropriate sensor location selection, we compute the Absolute Percent Error (APE) in terms of link flows for all *unmeasured* links $l_u \in U$. APE distributions are represented through box-plot graphs for the sake of immediacy. The APE measurements collected for randomly generated sensor locations pertain to the cumulative distribution from all draws.

B. Test results

We begin this Section by comparing the results of our previously developed TF-Model [13] with those of the newly proposed sensor fusion based approach of Section II, on a selected subset of unmeasured links, shown respectively in Figures 2 and 3.

These results, obtained considering an unrealistic FCD penetration rate of 100%, showcase indeed how the proposed

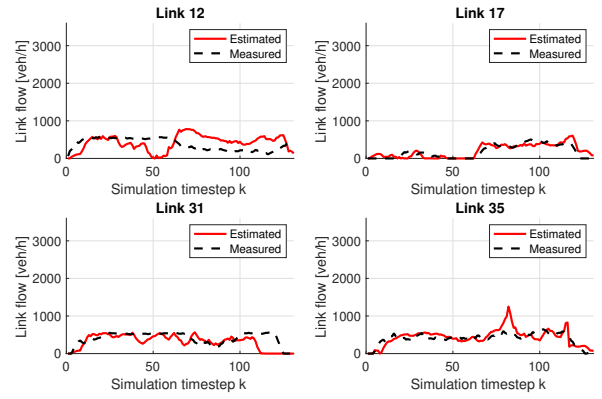


Fig. 2. Estimated vs Ground-truth flows on selected links. Observability-based TF-Model.

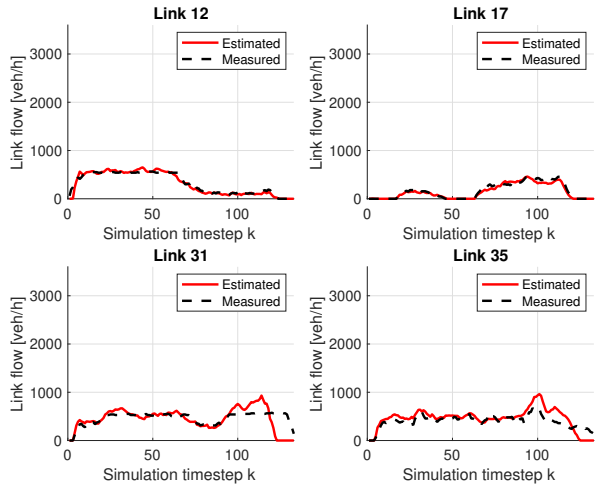


Fig. 3. Estimated vs Ground-truth flows on selected links. Proposed data fusion TF-Model.

data fusion approach is capable of considerably improved estimation of node turning ratios and, thus, link flows. Naturally, as the penetration rate decreases we can expect the gains in estimation quality to reduce substantially, reverting eventually to the results obtained by the cascading scheme when no such data is available. An analysis showcasing this effect is reported in Figure 4, in terms of Absolute Percent Error distribution, for penetration rates of 0%, 50% and 100%.

Indeed, the distribution of APE changes considerably as the penetration rate of FCD-equipped vehicles rises in the network. It is however important to remark that the rate itself is not involved in any computational aspect of the approach, only affecting its outcome exogenously.

In the final set of tests we aim to assess the impact of sensor locations on estimation quality, by comparing the APE distribution resulting from randomly generated locations with those pertaining to observability-based solutions. Figures 5, 6 and 7 detail the impact of sensor locations for the three instances of, respectively, 0% FCD penetration rate, 50%

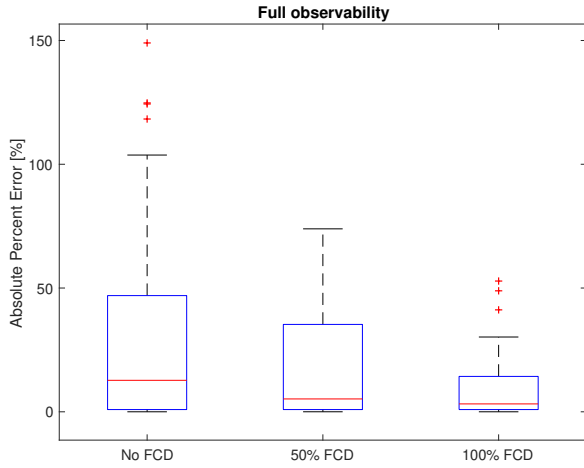


Fig. 4. Distribution of link flow APE with respect to varying degrees of FCD penetration rates.

penetration rate and 100% penetration rate. It is important to remark that the three figures do not share the same Y-axis scales, for the sake of representability.

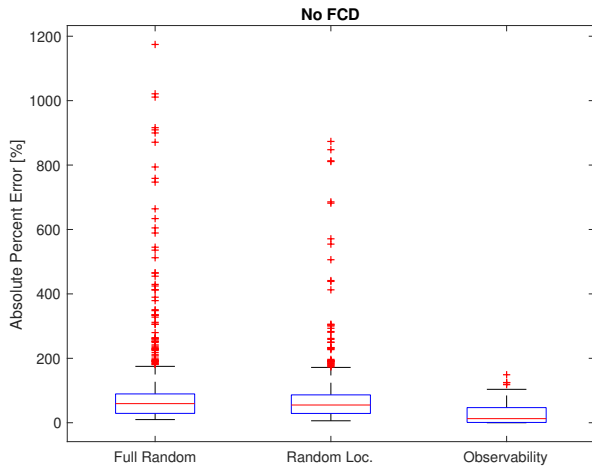


Fig. 5. Distribution of link flow APE with respect to sensor location policy. 0% FCD penetration rate.

Two key considerations arise from these results. Firstly, the proposed cascading framework appears to be strongly susceptible to sensor locations, and seems indeed capable of exploiting the topological inference information obtained through the observability matrix Ψ to considerably improve the overall quality of the performed estimation, whereas less topologically significant locations lead to considerable losses in estimation performance. Secondly, the impact of poorly selected locations apparently transcends the penetration rates of Floating Car Data, suggesting that observability principles could in fact play a solid role in complementing the application of data fusion techniques in transportation, exploiting the

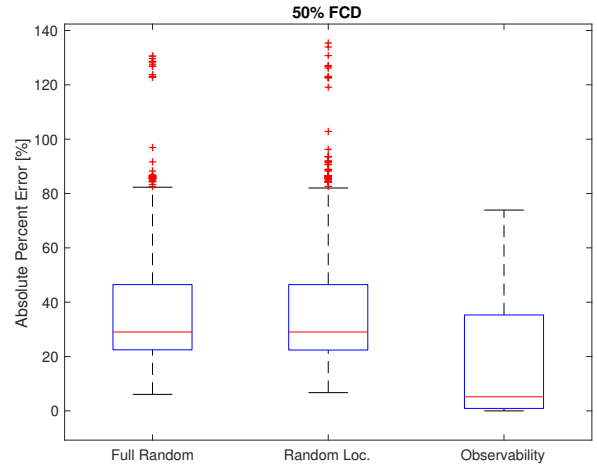


Fig. 6. Distribution of link flow APE with respect to sensor location policy. 50% FCD penetration rate.

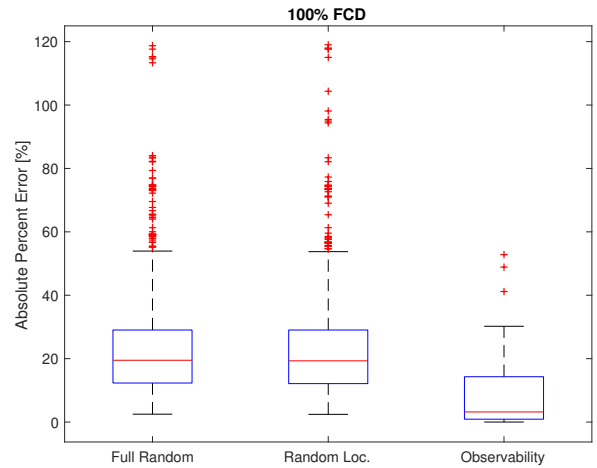


Fig. 7. Distribution of link flow APE with respect to sensor location policy. 100% FCD penetration rate.

additional information embedded in topological connectivity.

IV. CONCLUDING REMARKS

In this paper we developed a cascading data fusion filtering technique applied to network traffic state estimation. Loop detector and Floating Car Data are combined in order to produce reliable estimates of both link flow travel times and node turning ratios. The impact of sensor locations on the estimation accuracy is evaluated, comparing the effect of observability-based sensor placement with randomly generated locations. Test results showcase that the inclusion of FCDs leads to substantially reduced estimation errors as penetration rates increase, and that, indeed, sensor locations play a key role in enabling the highest possible levels of precision in measurement reconstruction. Further testing is warranted on real-life networks in order to both validate whether the pro-

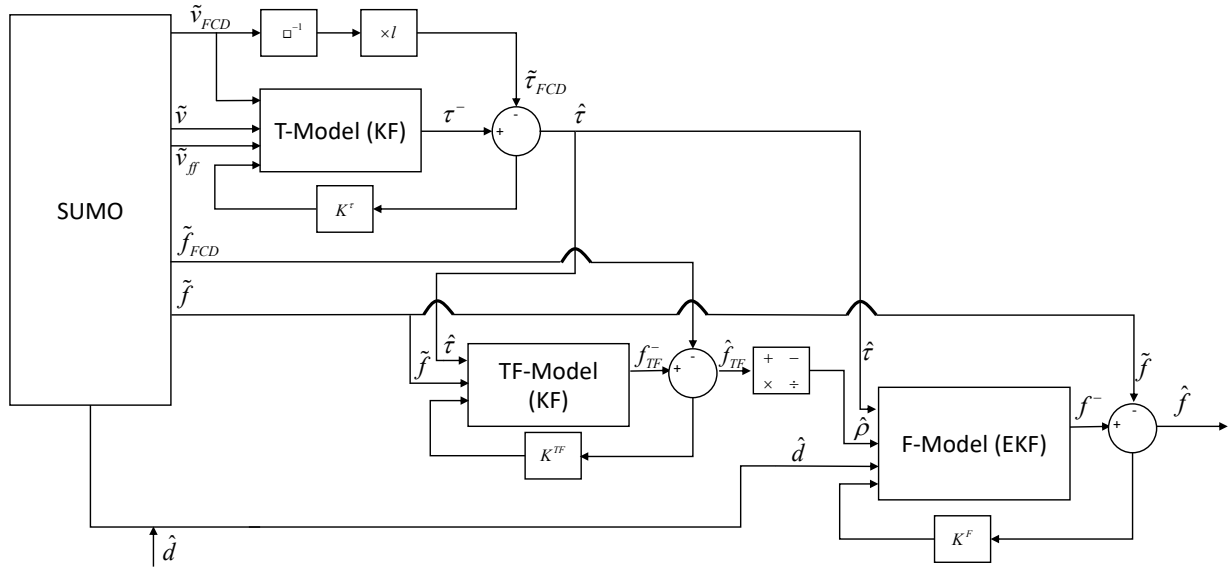


Fig. 8. Updated cascading Kalman Filtering scheme.

posed approach is effectively capable of maintaining desirable performance in large scale networks bearing complex connectivity patterns, as well as to include explicit consideration of the impact of traffic lights and the resulting queuing and travel time dynamics occurring at a signal cycle level.

REFERENCES

- [1] U.S. Department of Transportation, "Roadside Detection (RS-D)," <https://www.itskrs.its.dot.gov/costs/adjusted/rsd>.
- [2] E. Castillo, A. J. Conejo, J. M. Menéndez, and P. Jimenez, "The observability problem in traffic network models," *Computer-Aided Civil and Infrastructure Engineering*, vol. 23, no. 3, pp. 208–222, 2008.
- [3] M. Gentili and P. B. Mirchandani, "Survey of Models to Locate Sensors to Estimate Traffic Flows," *Transportation Research Record*, vol. 2243, no. 1, pp. 108–116, Jan. 2011.
- [4] F. Rempe, P. Franek, U. Fastenrath, and K. Bogenberger, "Online Freeway Traffic Estimation with Real Floating Car Data," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, Nov. 2016, pp. 1838–1843.
- [5] M. Saeedmanesh, A. Kouvelas, and N. Geroliminis, "A real-time state estimation approach for multi-region MFD traffic systems based on extended Kalman filter," in *2019 TRB Annual Meeting Online*. Transportation Research Board, 2019, pp. 19–02756.
- [6] Y. Yuan, 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, pp. 59–70, Mar. 2012.
- [7] Y. Wang and M. Papageorgiou, "Real-time freeway traffic state estimation based on extended Kalman filter: A general approach," *Transportation Research Part B: Methodological*, vol. 39, no. 2, pp. 141–167, Feb. 2005.
- [8] H. Yang, P. J. Jin, B. Ran, D. Yang, Z. Duan, and L. He, "Freeway traffic state estimation: A Lagrangian-space Kalman filter approach," *Journal of Intelligent Transportation Systems*, vol. 23, no. 6, pp. 525–540, Nov. 2019.
- [9] A. Nantes, D. Ngoduy, A. Bhaskar, M. Miska, and E. Chung, "Real-time traffic state estimation in urban corridors from heterogeneous data," *Transportation Research Part C: Emerging Technologies*, vol. 66, pp. 99–118, May 2016.
- [10] X. S. Trinh, D. Ngoduy, M. Keyvan-Ekbatani, and B. Robertson, "A comparative study on filtering methods for online freeway traffic estimation using heterogeneous data," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, Oct. 2019, pp. 4386–4391.
- [11] G. Böker and J. Lunze, "Freeway Traffic State Estimation with Floating-Car-Data," *IFAC Proceedings Volumes*, vol. 34, no. 9, pp. 291–296, Jul. 2001.
- [12] M. Rinaldi, C. M. J. Tampère, and F. Viti, "On characterizing the relationship between route choice behaviour and optimal traffic control solution space," *Transportation Research Part B: Methodological*, vol. 117, pp. 892–906, Nov. 2018.
- [13] M. Rinaldi and F. Viti, "A cascading Kalman filtering framework for real-time urban network flow estimation," in *Proceedings of the 23rd IEEE International Conference on Intelligent Transportation Systems*, Rhodes, Greece, Oct. 2020.
- [14] E. Castillo, Z. Grande, A. O. W. Y. Szeto, and H. K. Lo, "A State-of-the-Art Review of the Sensor Location, Flow Observability, Estimation, and Prediction Problems in Traffic Networks," *Journal of Sensors*, vol. 2015, p. e903563, Oct. 2015.
- [15] M. Rinaldi and F. Viti, "Exact and approximate route set generation for resilient partial observability in sensor location problems," *Transportation Research Part B: Methodological*, vol. 105, no. Supplement C, pp. 86–119, Nov. 2017.
- [16] P. Alvarez Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, and E. Wießner, "Microscopic Traffic Simulation using SUMO," in *2005 IEEE Intelligent Transportation Systems Conference (ITSC)*. Maui, USA: IEEE, Nov. 2018.
- [17] F. Viti, M. Rinaldi, F. Corman, and C. M. J. Tampère, "Assessing partial observability in network sensor location problems," *Transportation Research Part B: Methodological*, vol. 70, pp. 65–89, Dec. 2014.