Unscented Kalman filter for urban link travel time estimation with mid-link sinks and sources

Ré-Mi Hage, David Bétaille, François Peyret, Dominique Meizel

2012 15th International IEEE
Annual Conference on Intelligent Transportation Systems
Hilton, Anchorage, AK, USA 16 Sep - 19 Sep 2012

Unscented Kalman filter for urban link travel time estimation with mid-link sinks and sources

R.-M. Hage, D. Bétaille, *Member, IEEE*, F. Peyret, D. Meizel

Abstract— **To estimate the link travel time, the classical analytical procedure uses vehicles counts at upstream and downstream locations. This procedure is vulnerable in urban networks mainly due to significant flow to and from mid-link sinks and sources. One of the important developments recently done on this topic has yielded to the CUPRITE methodology. This method is derived from the classical analytical procedure. It integrates probe vehicle data to correct deterministically the upstream cumulative plot to match the information of probe vehicles travel times, whilst the downstream cumulative plot is kept unchanged. The algorithm proposed and validated in this research estimates urban links travel times based on an unscented Kalman filter (UKF). This algorithm integrates stochastically the vehicle count data from underground loop detectors at the end of every link and the travel time from probe vehicles. The proposed methodology, which can be used for travel time estimation in real-time, is compared to the classical analytical procedure and to the CUPRITE method in case of mid-link perturbation. Along to its lower sensitivity than CUPRITE, the UKF algorithm makes it possible detection and exclusion of outliers from both data sources.**

I. Introduction

Travel time information is an important parameter that can be used to identify and assess operational problems as well as to measure the effectiveness of transportation systems. Travel time in excess (delay) causes indirect costs to drivers in terms of lost time, discomfort and frustration, and direct costs in terms of fuel consumption. Travel time information is easy to be perceived by users and has the potential to reduce congestion on both temporal and spatial scales. By reducing congestion, it also reduces vehicle emissions and energy consumption, and finally the effect of transportation on the global warming phenomenon. As a result, it maximizes the efficiency and capacity of the road network.

Different techniques are used to estimate travel time on roads. These techniques depend on the type of system (i.e. fixed or mobile sensors) used to collect traffic data. Fixed sensors, such as inductive loop detectors, are the oldest and most widely used for traffic data sources. They provide temporal traffic state information, though only point based data. Under certain assumptions, researchers have proposed methodologies that can be characterized by deterministic [1-4] or stochastic approaches [5-6].

Mobile sensors, such as probe vehicles, are vehicles equipped with vehicle-tracking equipment (e.g. Global Positioning System). They provide trajectory data (time stamp and position coordinates) and hence probe vehicle travel time. They represent random sample from the population of the vehicles in the network. Therefore the accuracy of travel time estimation with probe vehicles is related to the number of the latter. Researchers focused on determining the minimum number of probe vehicles required for statistically significant travel time estimation [7].

The properties of these two data sources are complementary. Hence, they can be harnessed by developing a solution that merges multi-sensor data for the problem of estimating travel time in urban areas.

In this context, El Faouzi [8] provides an overview of the application of data fusion techniques in road traffic engineering. El Faouzi and Lefevre [9] have developed a method based on the evidence theory that provides a relevant theoretical basis when dealing with incomplete and inaccurate information. Choi and Chung [10] have applied a Bayesian pooling method to fuse data from detectors (space-mean speed using Dailey's equation [5]) and probe vehicles (using fuzzy regression). However, these methods have not dealt with traffic signals, which affect link travel time of probe vehicles, neither with the flow to and from mid-link sink and source. The CUPRITE methodology [11] addressed these problems by redefining and correcting the upstream cumulative number of vehicles. These numbers (or cumulative plots) are deterministically corrected using data from the probe vehicles. CUPRITE corrects the upstream cumulative plot at the minute-ceiled instant of occurrence of the probe vehicle, as well as prior and post this instant. In a real-time context, prior correction would not have been useful. In fact, any correction on travel times before the occurrence of a probe vehicle is questionable with regard to the "a posteriori" use of this information. Moreover, this methodology is sensitive to the noise on travel time probe vehicles.

The majority of the above researches is limited to freeways and cannot be applied in urban networks, where the travel time estimation is more challenging mainly due to significant proportions of flow from and to mid-link sources

R-M. Hage is a PhD student in the Geopositioning team - Institut français des sciences et technologies des transports, de l'aménagement et des réseaux IFSTTAR, route de Bouaye CS4 - 44344 Bouguenais Cedex (phone : +33240845647, e-mail : re-mi.hage@ifsttar.fr).

D. Bétaille is with Geopositioning team - Institut français des sciences et technologies des transports, de l'aménagement et des réseaux IFSTTAR, route de Bouaye CS4 - 44344 Bouguenais Cedex (e-mail : david.betaille@ifsttar.fr).

F. Peyret is with Geopositioning team - Institut français des sciences et technologies des transports, de l'aménagement et des réseaux IFSTTAR, route de Bouaye CS4 - 44344 Bouguenais Cedex (e-mail : francois.peyret@ifsttar.fr).

D. Meizel is with DMI Mathematics and Computer Sciences 123 XLIM, avenue Albert Thomas - 87060 Limoges Cedex, (e-mail : meizel@ensil.unilim.fr)

and sinks. Moreover, most of these methods cannot be applied to real-time travel time estimation. The present study bridges this gap by using a UKF-based statistical filtering approach.

II. Unscented Kalman Filter

The most common way of applying the Kalman filter (KF) to a nonlinear system is in the form of the extended Kalman filter (EKF). In the EKF, the probability distribution function is propagated through a linear approximation of the system around the operating point at each instant of time. In doing so, the EKF needs the Jacobian matrices. However, these matrices can be sometimes difficult and complicated to obtain. Further, the linear approximation of the system at a given time instant may introduce errors in the state, which may lead the state to diverge over time. In other words, the linear approximation may not be appropriate for some systems. In order to overcome the drawbacks of the EKF, other nonlinear state estimators have been developed such as the unscented Kalman filter (UKF). The UKF uses a deterministic sampling technique known as the unscented transform to pick a minimal set of sample points (called sigma points) around the mean. These sigma points are then propagated through the non-linear functions, from which the mean and covariance of the estimate are then recovered. In addition, this technique removes the requirement to explicitly calculate Jacobian matrices, which for complex functions can be a difficult task in itself. More details can be found in [12].

III. UKF Travel Time Estimation Modeling

Link travel time for a vehicle is the time needed to travel from the upstream point to the downstream point in the link. This research focuses on estimating the average travel time for all the vehicles that depart downstream (also called "experienced" travel time).

Fig. 1 illustrates the studied urban link for travel time estimation with mid-link sink/source. The mid-link infra-structures, such as a side streets, parking lots, private properties etc., acting as sink or source or both, is simply represented here by a mid-link.

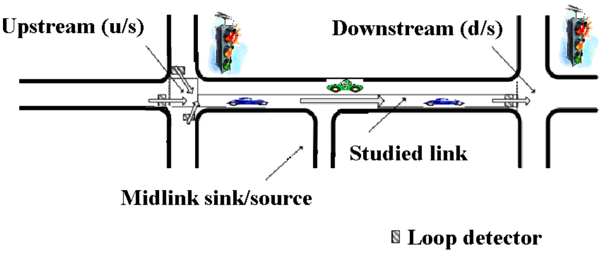

Figure 1. Studied link.

The state vector, evolution, observation and algorithm of the proposed UKF filtering are based on a state space model that we propose to detail below.

A. State vector

Suppose that we have a loop detector at the end of each link. We would like to estimate both the travel time and the number of vehicles that enter/exit the studied link from/to the mid-link source/sink, with a sampling time T_S of 1 minute. Therefore, for a given studied link (Fig. 1), the state vector contains:

- TT: the travel time.
- Nd: the cumulative count of vehicles at the downstream of link k.
- qu: the flow at the upstream of link k.
- Nu: the cumulative count of vehicles at the upstream of link k.
- p: the number of vehicles that enter/exit the link from/to the mid-link source/sink.
- The history of the cumulative count of vehicles at the upstream of link k, i.e. previous Nu, which is also the cumulative count of vehicles at the downstream of the link k-1. This history tabulates a fixed number h of past counts, this number being an "a priori" parameter of our modeling.

Therefore the state vector resumes as follows:

$$x(t) = \begin{bmatrix} TT(t) \\ N_d(t) \\ N_u(t) \\ q_u(t) \\ p(t) \\ N_u(t - T_S) \\ .... \\ N_u(t - h \times T_S) \end{bmatrix}$$

B. Evolution model

We suppose that state at time t derives from state at time $t\text{-}T_S$ as follows:

- The classical analytical procedure states that the cumulative number of vehicles at link entrance shall be equal to the cumulative number of vehicle at link exit after an average travel time TT at time t. Thus the travel time at time t is equal to difference between time t and the corresponding time t_1 when $N_u(t1)=N_d(t)$.
- The cumulative number at the downstream at time t is equal to the cumulative number at the upstream at time t-TT, incremented by p the number of mid-link sink/source vehicles.
- The cumulative number at the upstream at time t is the cumulative number at the upstream at time $t\text{-}T_S$, incremented by the flow at the upstream multiplied by T_S.
- The flow at the upstream at time t is stable, as well as the mid-link cumulative vehicles number.

Therefore the evolution model is as follows:

$$\begin{aligned} &TT(t) = t - N_u^{-1}\left(N_d(t)\right) \\ &N_d(t) = N_u(t - TT) + p(t) \\ &N_u(t) = N_u(t - Ts) + q_u(t) \times Ts \\ &q_u(t) = q_u(t - Ts) \\ &p(t) = p(t - Ts) \end{aligned} \tag{1}$$

The key point in this evolution model is that it depends on the state vector itself (second equation) therefore it is impossible to explicit the Jacobian matrix needed in the EKF, which justifies the use of an UKF.

Observation model

The considered observations are both the travel time from probe vehicles and the number of vehicles from loop detectors. The data from the probe vehicles contain vehicle ID, GPS position coordinates, time, and eventually speed, moving direction, etc. To estimate individual travel times, a map-matching process needs to be made. This is an important step in the process; its accuracy will directly affect the final results. Map-matching algorithms may adopt either a geometric or a topologic or both approach [13].

The observation equation depends on the available measurement.

Case 1: a travel time issued from a probe vehicle is modeled as:

$$z_t = H_1 x_t + v_{1t} \tag{2}$$

where: H_1 is [1 0 0 0 … 0] and v_{1t} is the observation noise assumed to be zero mean Gaussian white noise with covariance R_{1t}.

Case 2: reading the counter associated to a loop detector yields:

$$z_t = H_2 x_t + v_{2t} \tag{3}$$

where: H_2 is resp. [0 1 0 0 … 0] and [0 0 1 0 … 0] for downstream and upstream counters and v_{2t} is the observation noise assumed to be zero mean Gaussian white noise with covariance matrix R_{2t}.

These covariances are obviously different whether one considers counters or probe vehicles travel time. As for the last, it should characterize possible errors in the process of map-matching GPS positions. GPS errors, and the consecutive map-matching errors, will be fixed depending on the location of the link: in a dense city center, the order of magnitude of those errors is some tens of meters, whereas in an open area, it is only a few meters. In a very first approximation, we will fix travel time observation errors to a maximum of 10 seconds down to few seconds.

D. Algorithm

The filter estimates travel time with 1 minute sample time. Data from detectors are aggregated each minute whereas data from probe vehicles are available between two consecutive minutes t and t+1. In order to use information at its exact time, an intermediate step is made between t and t+1. Fig. 2 summarizes the UKF algorithm.

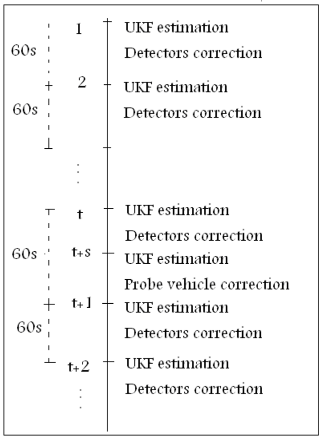

Figure 2. UKF algorithm.

IV. UKF Testing

This section presents the results of the classical analytical procedure, the CUPRITE model and our UKF-based model. Each algorithm is implemented and tested on simulated data. The simulation is made with AIMSUN on a 600 meters long one-lane link (see Fig. 1). The free flow speed is 36 km/h, the maximum density is 140 veh/km, and the critical density is observed at 1/5 of this maximum, therefore the flow is limited to 1008 veh/h. This calculation is based on the fundamental diagram applied in the center of Nantes by Nantes-Metropole traffic management center.

For an hour of simulation the demand flow i.e. the flow that enters the link is as follows: for the first 15 min the average flow is 500 veh/h, for the second 15 min the average flow is 900 veh/h, for the third 15 min the average flow will increase to 1400 veh/h, which is greater than the maximum flow (1008 veh/h), and for the final 15 min the average flow will decrease to 500 veh/h. Sinks and sources are defined as the percentage of vehicles that are lost into the sink and gained from the source (perturbation). In this analysis 1% and 5% of mid-link sinks and sources were considered. Moreover, probe vehicles are a random sample from the total population of vehicles. 1%, 5%, and 10% were considered as probe vehicles. For each probe vehicle, we simply use its corresponding travel time. Finally, the average travel time for each T_S interval (denoted further: reference TT) is the sum of the travel time of all the vehicles that exit the link between t and t+T_S divided by the total numbers of vehicles that enter the link.

A. Sensitivity of CUPRITE

The classical analytical procedure gives an unbiased estimation of the travel time when the vehicles in the studied link are conserved. In such case, the CUPRITE correction should remain zero. But in the eventuality when the deviation between the reference TT and the TT of the considered probe vehicle is significant, this correction will bias the travel time estimation later on. With the UKF, this deviation has mainly an effect at the probe vehicle instant of correction. After this correction, the UKF evolution and detectors correction will overcome the previous resulting effect, whereas CUPRITE remains biased as long as no new probe vehicle passes. Fig. 3 illustrates the sensitivity of CUPRITE with a biased probe vehicle travel time.

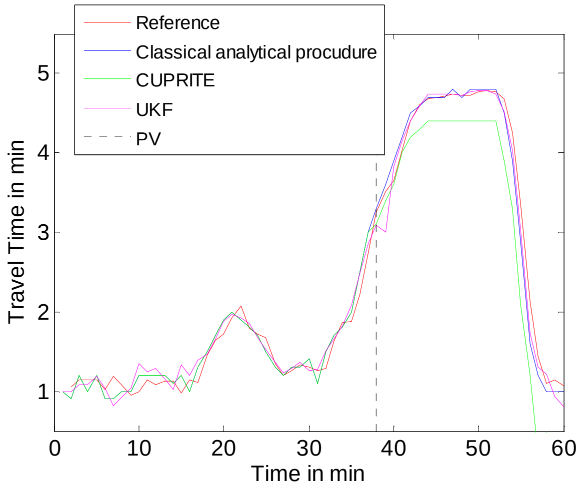

Figure 3. Demonstration of the sensitivity of CUPRITE with 1 vehicle.

B. Model performance testing

The following statistics are used to qualify each of the classical analytical procedure, the CUPRITE and UKF methodologies:

$$error_i = actual_i - estimated_i \qquad (4)$$

$$MAPE = \frac{\sum_{i=1}^{n}\left(\left|\frac{actual_i - estimated_i}{actual_i}\right|\right)}{n} \qquad (5)$$

$$accuracy\,(\%) = (1 - MAPE) \times 100 \qquad (6)$$

where:

- $Actual_i$, $estimated_i$, and $error_i$ and are respectively the actual average travel time, the estimated travel time, and the relative error for the *i*th estimation interval.
- MAPE stands for the mean absolute percentage error.
- Accuracy indicates the mean exactitude in %.
- n is the total number of estimation interval.

V. Results and Analysis

Table I summarize the average accuracy in percentage with 1% and 5% mid-link sinks/sources, for 10 simulations, for the classical analytical procedure, CUPRITE and UKF with 1%, 5%, and 10% of probe vehicles. It is obvious that both UKF and CUPRITE correct the bias in the classical analytical procedure. As expected, their accuracies increase with probe vehicle percentage.

Fig. 4 and 7 represent respectively the classical analytical procedure error in minute for 1% and 5% mid-link sink and source. We notice the cumulative effect of the perturbation especially on the last 10 minutes. Fig. 5 and 8 represent respectively the CUPRITE error in minute for 1% and 5% mid-link sink and source. They show that in some simulation, the error reaches ±2 min. Here, we observe again the sensitivity of CUPRITE model. Fig. 6 and 9 represent respectively the UKF error for 1% and 5% mid-link sink and source. In UKF and CUPRITE, the standard deviation of the estimated travel time decreases as the percentage of probe vehicles increases. To conclude, UKF estimation is less noisy then CUPRITE.

TABLE I. AVERAGE ACCURACY (100-MAPE%) IN % OF CLASSICAL ANALYTICAL PROCEDURE, UKF, AND CUPRITE WITH MID-LINK SINK AND SOURCE

		Average accuracy			
		Mid-link sink		**Mid-link source**	
	% of probe vehicles	*1%*	*5%*	*1%*	*5%*
CAP		85	45	86	48
UKF	1	90	80	91	84
	5	93	90	93	91
	10	94	93	94	94
CUPRITE	1	87	77	89	85
	5	90	88	92	90
	10	92	90	92	93

VI. Cumulative Number of Mid-Link Sink/Source

With the UKF filter, we can also estimate the cumulative number of mid-link sink/source without direct measurement of the perturbation. Fig. 10 and 12 illustrate the estimation of the latter with 1%, 5%, and 10% probe vehicles for respectively 1% and 5% mid-link sink/source.

VII. Outlier Detection and Exclusion

Fig. 11 illustrates the effect of a probe vehicle outlier (3 min vs. 2 min) on the travel time estimation, whereas Fig. 13 illustrates the effect of a detector outlier, where the counted downstream value has been fixed for 3 minutes. By means of a chi-square test of the normalized innovation squared (also called the Mahalanobis distance), the UKF model has rejected the outlier, whereas CUPRITE could not, leading to an aberrant travel time estimation.

VIII. Conclusion

The UKF filter developed here provides encouraging results for urban link travel time estimation with mid-link sinks and sources. The evolution model of this algorithm is based on the classical analytical procedure. The observations are vehicle counts from loop detectors located at the end of every link and travel time from probe vehicles after they have been associated to the appropriate link by map-matching.

The main contribution of this article is that the UKF stochastic approach overcomes the sensitivity of the CUPRITE deterministic approach to probe vehicle sampling. Actually, CUPRITE supposes that data are exact whereas the UKF filter offers the possibility to introduce an error model for the travel time obtained by map-matching as well as for loop detectors. Thus, the estimation is smoothed and statistic tests, made possible by the UKF formalism, enable detection and exclusion of outliers, like mis-matched GPS or loop deficiency. Furthermore, UKF can be applied in a real-time context. In this article many simulation were run with variable flow, variable percentage of the vehicles that are randomly selected and considered as probe, and variable percentage of mid-link sink/source. The feasibility of outlier detection and exclusion has been demonstrated, but this should be deepened and the next step is to determine the observation error from real data. Furthermore, the application of this model to an urban network is under development: it is an extension of the proposed model with no information on the turning movement proportion at crossroads.

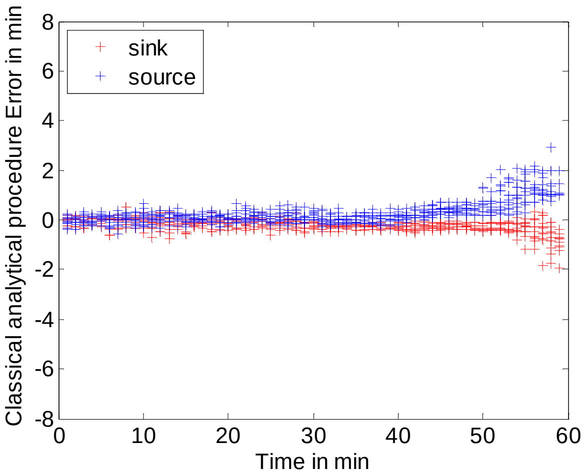

Figure 4. Classical analytical procedure estimation error with 1% mid-link sink/source.

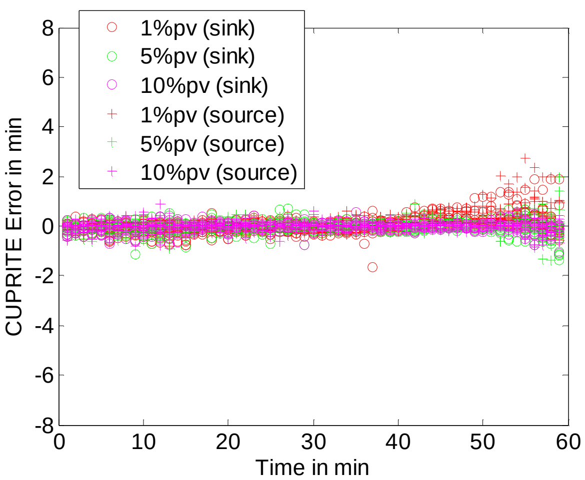

Figure 5. CUPRITE estimation error with 1% mid-link sink and source.

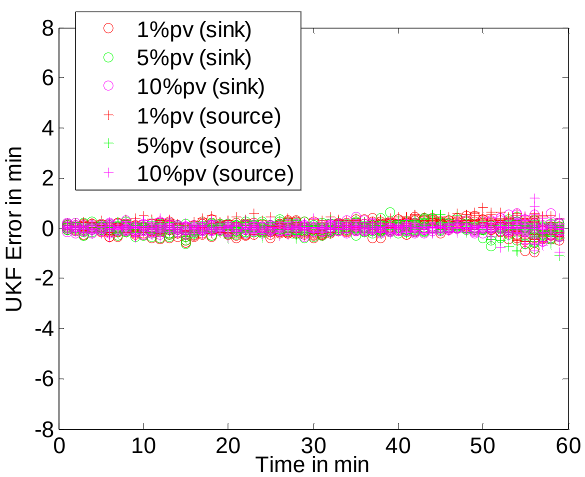

Figure 6. UKF estimation error with 1% mid-link sink and source.

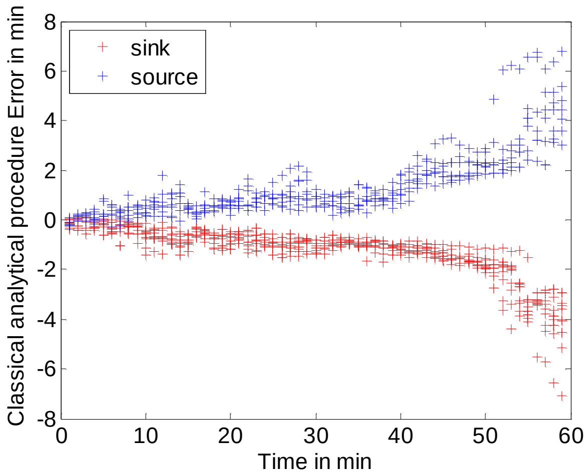

Figure 7. Classical analytical procedure estimation error with 5% mid-link sink/source.

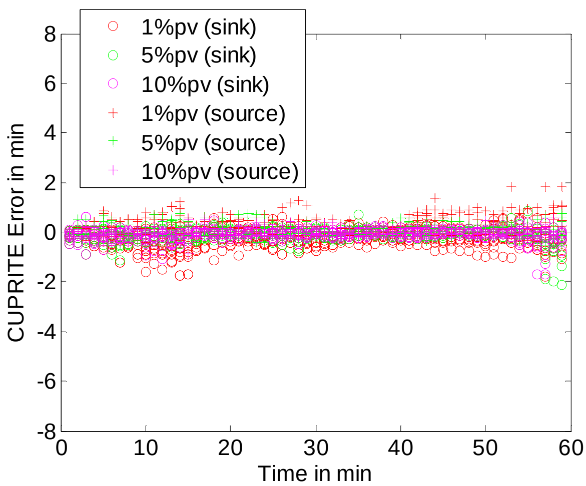

Figure 8. CUPRITE estimation error with 5% mid-link sink and source.

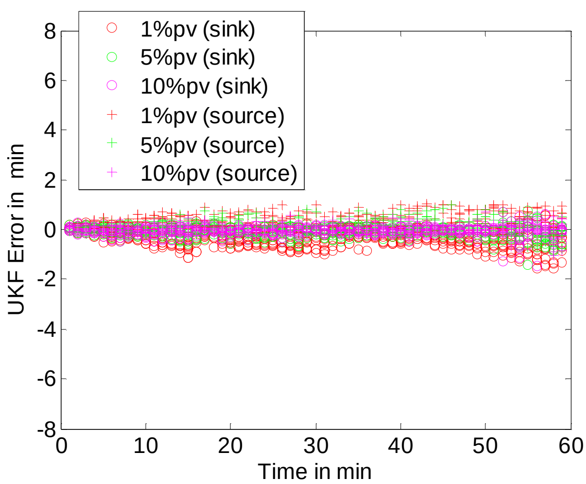

Figure 9. UKF estimation error with 5% mid-link sink and source.

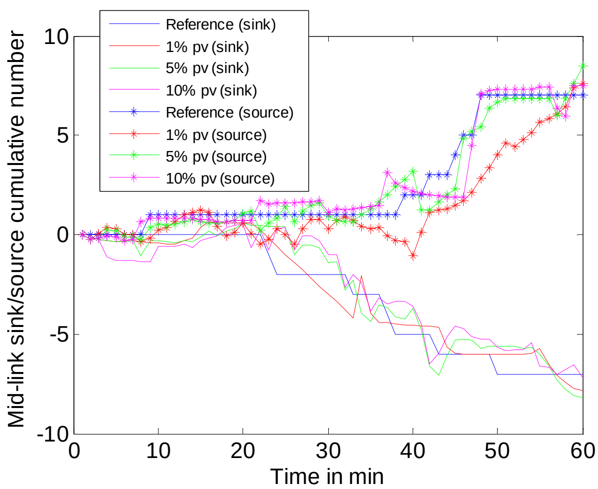

Figure 10. Cumulative number of mid-link sink/source with 1% perturbation.

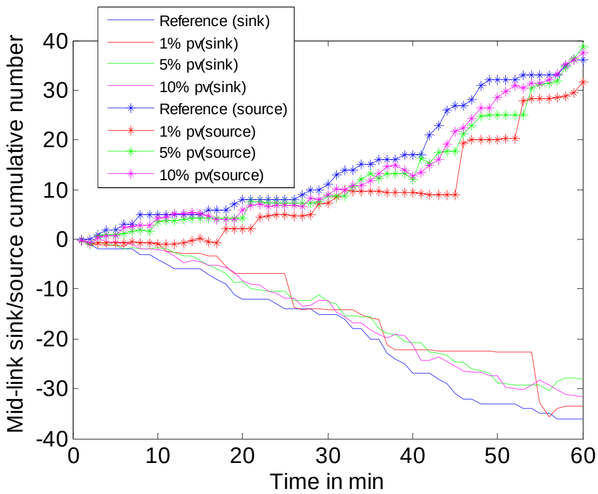

Figure 12. Cumulative number of mid-link sink/source with 5% perturbation.

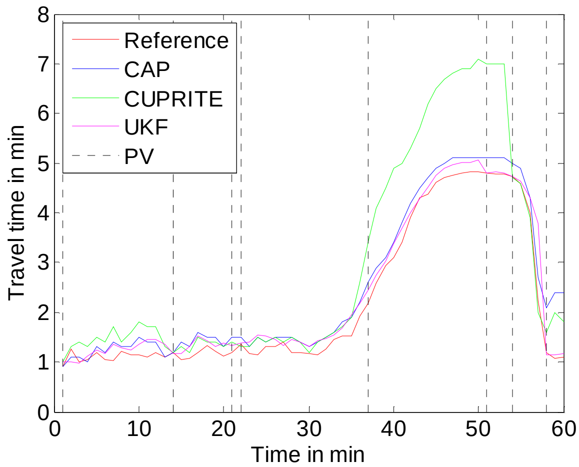

Figure 11. Travel time estimation by classical analytical procedure (CAP), CUPRITE, and UKF with one outlier probe vehicle travel time.

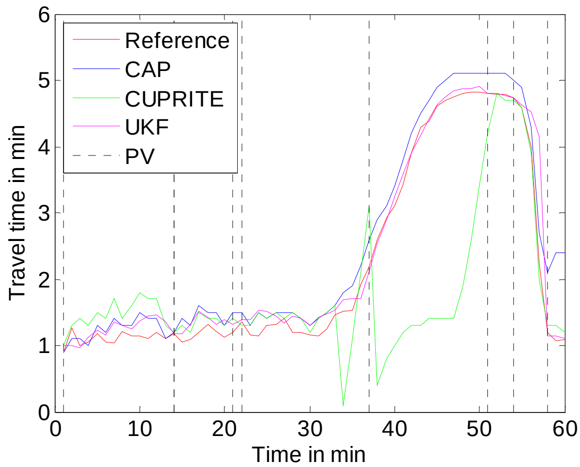

Figure 13. Travel time estimation by classical analytical procedure(CAP), CUPRITE, and UKF with detector outlier.

References

[1] NAM, D.H. & DREW, D.R. "Automatic measurement of traffic variables for intelligent transportation systems applications". In *Transportation Research Part B*: Methodological, 33, 437-457, 1999.

[2] OH, J.S., JAYAKRISHNAN, R. & RECKER, W. "Section travel time estimation from point detection data". In proceedings of the 82th Annual Meeting of *Transportation Research Board*, Washington, D.C., U.S.A. 2003.

[3] COIFMAN, B. & KRISHNAMURTHY, S. "Vehicle reidentification and travel time measurement across freeway junctions using the existing detector infrastructure". In *Transportation Research Part C*: Emerging Technologies, 15, 135-153, 2007.

[4] BUISSON, C. "Analyse bibliographique et première classification des algorithmes de simulation de l'écoulement du trafic routier". *Recherche Transports Sécurité*, 53, 35-48, 1996.

[5] DAILEY, D.J. "A statistical algorithm for estimating speed from single loop volume and occupancy measurements". In *Transportation Research Part B*: Methodological, 33, 313-322, 1999.

[6] HINSBERGEN, C.P. et al. "Fast Traffic State Estimation with the Localized Extended Kalman Filter". In proceedings of the *IEEE ITSC*, 917-922, Madeira, Portugal, 2010.

[7] SRINIVASAN, K.K. & JOVANIS, P.P. "Determination of Number of Probe Vehicles Required for Reliable Travel Time Measurement in Urban Network". In *Transportation Research Record*, 1537, 15-22, Washington, D.C., 1996.

[8] EL FAOUZI, N.E. "Data fusion in road traffic engineering: an overview". In proceedings of the *SPIE - International Society for Optical Engineering*, Orlando, FL, 2004.

[9] EL FAOUZI, N.E. & LEFEVRE, E. "Classifiers and distance based evidential fusion for road travel time estimation". In proceedings of the *SPIE - International Society for Optical Engineering*, Orlando, FL, 2006.

[10] CHOI, K. & CHUNG, A. "Data Fusion Algorithm for Estimating Link Travel Time". In *Journal of Intelligent Transportation Systems*, 7, 235–260, 2002.

[11] BHASKAR, A., CHUNG, E. & DUMONT, A.G. "Estimation of Travel Time on Urban Networks with Midlink Sources and Sinks". In *Transportation Research Record*, 2121, 41-54, Washington, D.C, 2009.

[12] WAN, E.A. & VAN DER MERWE, R. "The Unscented Kalman Filter, In Kalman Filtering and Neural Networks", *Wiley Publishing*, 2001.

[13] QUDDUS, M.A. & OCHIENG, W.Y. "Current map matching algorithms for transport applications: State-of-the-art and future research directions". In *Transportation Research Part C*: Emerging Technologies, 15, 312-328, 2007.