Data fusion for traffic flow estimation at intersections

Axel Wolfermann
Postdoctoral Research Fellow, Department of Civil Engineering, Nagoya University
axel.wolfermann@trafficdata.info

Babak Mehran
Postdoctoral Research Fellow, Institute of Industrial Science, The University of Tokyo
babak@iis.u-tokyo.ac.jp

Masao Kuwahara
Professor, Tohoku University, Graduate School for Information Sciences

ABSTRACT

The efficient design and operation of intersections, particularly signalized intersections, depends on the availability of detailed traffic flow data. Information on the turning ratio and lane occupation is, however, commonly scarce or unreliable. Combining Floating Car Data (FCD) with other data sources offers the opportunity to fill this gap. As opposed to existing data fusion concepts, detector data is not directly used to deliver information on the traffic flow on a specific link, but it is evaluated in connection with the FCD to estimate the penetration rate of the traffic with probe vehicles. This information is combined with historic data and information contained in network connections. The data fusion itself is achieved by using a Kalman Filter (KF). By using FCD, the output of the data fusion process is up to date and can be used to take incidents and other fluctuations of traffic volumes into account. The elasticity of the filter to different input reliabilities and flow variations is shown. The concept works well and promises to be a useful tool for cities with sufficient FCD available. The performance depends in the first place on a good calibration, which can be achieved by extensive tests based on simulations and based on manual counts as part of an implementation.

Submitted 2nd May 2011

INTRODUCTION

Motivation, objective, limitations

The efficient design and operation of intersections, particularly signalized intersections, depends on the availability of detailed traffic flow data. The collection and processing of traffic information, however, is a major challenge for transportation agencies. While cross-sectional data for links and intersection approaches is commonly available from validated flow models and detector data, information on the turning ratio and lane occupation is commonly scarce or unreliable. Combining information from probe vehicles (Floating Car Data, FCD) with other data sources (detector data, historic data etc.) offers the opportunity to fill this gap.

While FCD is increasingly used for the determination of origin-destination (OD) matrices in networks and for the estimation of travel times in networks, the application to directional or lane based flow data at intersection level is still a more or less blank space in the research landscape.

The presented methodology will help to close this gap. The objective is to generate reliable OD matrices on intersection level. The methodology can easily be adjusted to deliver also lane based results. While the output is based primarily on FCD in combination with detector data, other data sources, namely historic data and the information available from the network context, is incorporated to improve the quality of the output data and step into the breach left by insufficient FCD. The methodology can easily be extended to other data sources. The data is aggregated to observation intervals. The more accurate the available data is, the shorter this interval might be.

As opposed to existing data fusion concepts, detector data is not directly used to deliver information on the traffic flow on a specific link, but it is evaluated in connection with the FCD to estimate the penetration rate of the traffic with probe vehicles. In this way the location of the detectors is of minor importance.

The data fusion itself is achieved by using a Kalman Filter (KF). Hence, it is assumed that all data sources deliver inaccurate data with the error as white noise (i.e. normally distributed). How much the filter relies on the different data sources depends on the error involved, which is subject to the calibration of the filter.

The proposed procedure relies heavily on the availability of representative FCD. As long as the penetration rate is low, the lack of probe information might be balanced by historic data, but the advantage of the methodology is thus reduced. Incidents can only be detected by online information. However, one outcome of the Kalman Filter is not only the traffic flow information, but also the reliability of this information. Even if no high quality input data is available, the filter will deliver results, but with less accuracy – and the filter will name the level of accuracy. It is important to note, that not only the judgment of this accuracy by the filter, but also the traffic flow estimation itself depends on the correct estimation of the variation of the input data from the different data sources.

Many applications of FCD suffer from the lack of representativeness of the probes for the total population (e.g. when using taxi or bus FCD). This drawback can be overcome to some extent by the presented methodology by computing the penetration rate at different locations based on detector counts.

Outline

This article is arranged in five major sections. After a brief literature review on data fusion and FCD, the methodology for the implementation of the Kalman Filter to estimate the volumes of the vehicular movements at isolated intersections is described. The data sources used in the filter are elucidated in the next section. The data fusion methodology can be extended to adjacent intersections or networks, as is described in a separate section. The performance of the filter is analysed subsequently for simulated scenarios, before closing the article with conclusions and the outlook towards future improvements.

LITERATURE REVIEW

Data fusion (DF) has been used to tackle a variety of problems in transportation engineering ranging from conventional problems in transportation planning to the most recent applications in ITS [1-3]. El Faouzi et al. [4] provide a comprehensive survey of how DF is used in different areas of ITS. The major fields to be mentioned here are advanced traveller information systems (ATIS), automatic incident detection (AID), advanced driver assistance (ADAS), network control, crash analysis and prevention, traffic demand estimation, traffic forecast and monitoring and accurate position estimation. Each of these applications can make use of different information sources and combine them using DF techniques to achieve better results.

Floating Car Data (FCD) becomes a powerful data source when used as an input to DF techniques. The research on FCD dates already back to the last century. The major impediment to the application of FCD is still the modest availability due to low penetration rates or privacy issues. Nevertheless, many field experiments have been conducted and the technology is increasingly established as a complement to other data sources.

FCD to estimate travel times

Data from cameras, FCD or cell phone tracking are increasingly used in order to supplement the information provided by conventional measurement techniques and improve the accuracy of travel time estimates [6-8]. Dailey et al. [9] give a detailed description of a current DF within an ITS project and the presents a new quantitative data fusion algorithm to estimate speeds from volume and occupancy measurements. El Faouzi et al. [10-12] proposed estimation frameworks to fuse conventional detector data and FCD for travel time estimation based on a weighted means method and the evidence theory. Choi [13] and Choi and Chung [14] have tackled the problem of generating travel time from loop detectors, probe vehicles and video-camera sources. They proposed a fuzzy logic based approach with its evaluation on a theoretical example.

Automatic incident detection

Recently, there has been renewed interest in incident detection algorithms partly because of the availability of new sensors and data sources. One of these sources is probe vehicles. Applications of several data fusion techniques to traffic management to support incident detection have been reported in the literature, and the data fusion algorithms used includes Dempster–Shafer inference, Bayesian inference, and voting logic. Most of these applications have explored the use of probe vehicles data with the conventional traffic data for incident detection purposes [15-18].

Traffic demand estimation

One of the most important problems in the field of transportation planning and control is the origin–destination estimation from link counts. Several researchers proposed schemes to derive the OD matrices by combining data from different sources. The initial efforts to address this problem were made by Cremer and Keller [32, 33]. Further developments were pursued later by many authors [34, 35]. The Kalman filtering is of common practice in this class of problems. Ben-Akiva and Morikawa [36] have explored the OD estimation methods that combine different data sources (stated preference data and traffic measurements) and more recently, Lundgring et al. [37] described a method for adjusting time-dependent travel demand information with respect to link flow observations. They utilized the structure of the given OD-matrix, which is compounded from different sources, for making simple overall adjustments.

Traffic forecasting and monitoring

Many prediction schemes of traffic flow were obtained by means of classic autoregressive models, especially time series techniques. Some authors have tackled this problem by using Bayesian frameworks [38]. Some others used Kalman filtering [39] or neural networks and system identification [40], and more recently a nonparametric paradigm was adopted via kernel techniques [41]. None of these proposals allow one to achieve highly accurate predictions except in some special situations (for some network configuration and/or with high detector coverage). This is caused to some extent by the dynamics of traffic flow which cannot be formalized by a single procedure. Granger [42] showed that the linear combination of several predictors from a single data set can outperform the individual predictors. In traffic forecasting with heterogeneous data sources, El Faouzi provided a methodical framework to combine various forecasts of the same quantity [42, 43]. In [44], the integration problem of in-vehicle information and data provided by loop detectors was studied. The core of the integration step was the extended Kalman filtering. More recently, Sau et al. [45] investigated a traffic monitoring problem with data from various sources by applying a particle filter. Choi [46] examined the problem of missing data estimation and proposed a framework for missing data inference based on evidential reasoning.

Accurate positioning estimation

Inertial navigation systems (INS) are one of the earliest forms of navigation techniques. INS, which functions on the principle of dead-reckoning, has a potential problem of ‘‘integration drift” which is the accumulation of small errors in the measurement of acceleration and angular velocity into progressively larger errors in velocity, which are compounded into still greater errors in position. On the one hand, GPS data are accurate though when the satellite signals are blocked by tall buildings GPS outage occurs. DF can effectively be used to combine the advantages of both techniques and tackle their drawbacks. One of the earliest approaches in GPS/INS integration is to use Kalman Filtering [47, 48]. Also different types of neural networks have been used to combine GPS and INS information [49-51]. Such systems reconstruct vehicle dynamics by training artificial intelligence modules when GPS signals are available.

Conclusion

The Kalman Filter is by now a well established procedure in data fusion. Though it is based on the assumption of normally distributed (white) noise, applications where this condition was not fulfilled showed many advantages of the filter over comparable techniques and a sufficient quality.

FCD becomes more widely used. Numerous applications have been developed in the past. Most of them are directed at either the estimation of OD matrices in networks or on travel time predictions. Using FCD on intersection level to determine travel volumes for different flow directions is yet a new field for FCD.

DATA FUSION METHODOLOGY FOR INDIVIDUAL INTERSECTIONS

The Kalman Filter

The Kalman Filter (KF) uses measurements and their accuracy to correct estimates for the desired output (in the given case: traffic flow volumes). It consists for each time interval of a prediction step and one or more update steps. The basic assumption of the KF is that the true state of the system (here the traffic flow) evolves from the last step as

$$x_k = F_k x_{k-1} + B_k u_k + w_k \tag{1}$$

with *x* being the true state, *u* is some optional control input, *w* is the process noise, and *k* depicts the time interval. F_k is the state transition model which is applied to the previous state and B_k is the control input model, for each time interval k. The true state (and its estimate accordingly) is a vector containing the traffic flow for the commonly twelve traffic movements for each considered intersection as illustrated in Figure 1 (shown for left-side traffic). For lane based outputs the vector will contain the traffic volumes for all approach lanes. Together with the state itself the filter computes the reliability *P* (covariance matrix) of the estimate from Eq. (2).

$$P_{k|k-1} = F_k P_{k-1|k-1} F_k^T + Q_k \tag{2}$$

where P_k is the a posterior error covariance matrix (a measure of the estimated accuracy of the state estimate), F_k^T is transpose of F_k and Q_k is the covariance of the process noise. If the measurements are distorted by white noise (i.e. the error is normally random distributed) and the error is accurately known, the KF leads to a maximum likelihood estimate of the system state.

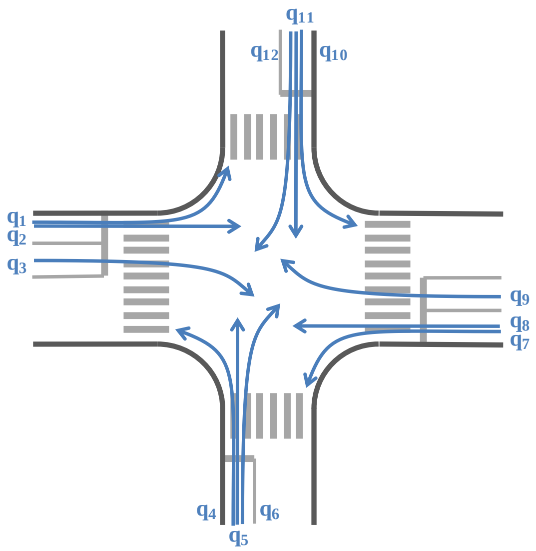

Figure 1 Typical two way intersection with twelve movements

For the traffic flow the control input could be seen in the traffic demand. But precisely this demand is not known, but can only be estimated by measurements and historic data. Hence, the prediction step of the KF consists only of adding noise to the output of the last time step and assuming otherwise an unchanged traffic state. The state transition matrix *F* is, hence, a unity matrix.

During the update step(s) the KF computes an innovation *y* (difference between the prediction and the measurement) from the measurements *z* (e.g. probe observations per movement) and the observation model *H* (in this case containing the penetration rates).

$$y_k = z_k - H_k x_{k|k-1} \quad (3)$$

with its accuracy (covariance) S based on the measurement noise R and the error covariance of the prediction P:

$$S_k = H_k P_{k|k-1} H_k^T + R_k \quad (4)$$

How much the innovation influences the a posteriori estimate (estimate of the state/traffic flow after the update) is determined by the Kalman Gain K. The optimal Kalman Gain used here is based on the assumption of white noise process and the measurement noise (Normal Distribution with zero mean).

$$K_k = P_{k|k-1} H_k^T S_k^{-1} \quad (5)$$

The more reliable the measurement is in relation to the accuracy of the prediction (error covariance determined by the noise terms), the more the innovation influences the estimate. The new estimate is derived from the innovation and weighted by the Kalman Gain K:

$$x_k = x_{k|k-1} + K_k y_k \quad (6)$$

Finally the accuracy P of the new state estimate is determined.

$$P_{k|k} = (I - K_k H_k) P_{k|k-1} \quad (7)$$

The number of update steps depends on the number of available measurements. In the given methodology, three update steps are used:

- FCD scaled by the penetration rate (based on detector counts)
- historic data
- network context

The "measurement" vector contains FCD, historic data, and flow estimates from adjacent intersections respectively, as will be explained in the following sections.

SOURCES USED FOR THE DATA FUSION

Floating Car Data (FCD)

The data fusion is based in the first place on FCD, i.e. the number of probe vehicles passing the intersection(s) under scrutiny. It is assumed that the raw data of the probes is already converted into link based information by using, for instance, map matching algorithms. The information has to contain the entry and exit link (the approach and exit lanes, if lane based information is aspired) with the desired time resolution (e.g. five minute intervals) for each probe. Furthermore, the passing time of all probes at the detector locations included in the data fusion process has to be known. Errors might result from invalid mapping of probe locations to links and from gaps or artefacts in the probe data.

The probe volumes observed during the current time interval for the respective intersection are contained in the measurement vector. The diagonal of the observation model H is filled with the penetration rates, which are determined by using detector counts, as will be explained in the subsequent subsection. The noise of the FCD has to be estimated. The accuracy of the detectors used for the determination of the penetration rate and the variation of the penetration rate in the network (the representativeness of the probe vehicles) has to be considered. The accuracy of the probe information has to be carefully calibrated for the filter to achieve the best results.

Detector data

The FCD delivers information on the number of probes passing an intersection. Probe vehicles are only a sample of the total vehicle population. The probe information has to be converted into total vehicle volumes. This conversion is only possible, if the penetration rate of the traffic with probes is known. To estimate this penetration rate, count data from detectors is used. By comparing the number of vehicles passing a detector with the number of probes passing the same location the penetration rate is estimated. For the observed link this penetration rate is directly used. For links without detectors, the average penetration rate of all detector locations is assumed, as long as no better estimate on the distribution of the probes in the network with reference to the total traffic is known.

To clarify the procedure, Figure 2 provides a simple example. Using the flow volumes of floating cars $q_{FC,A,j}$ and $q_{FC,B,j}$ and the detector counts $q_{Det,i}$ at the three detectors D shown in Figure 2, three individual penetration rates p_i (one for each detector) are computed according to Eq. (8).

$$p_i = \frac{\sum_j^n q_{FC,A,j}}{q_{\mathrm{Det},i}} \text{ and } p_i = \frac{\sum_j^n q_{FC,B,j}}{q_{\mathrm{Det},i}} \text{ respectively} \tag{8}$$

The average penetration rate can be approximated by

$$\bar{p} = \frac{\sum_i^k p_i}{k} \tag{9}$$

with k depicting the number of detectors. For the streams crossing a detector, the individual penetration rate for this detector is used. For all remaining streams the average penetration rate $\bar{p}$ is assumed, as long as no more accurate information is known. The flow volumes of floating cars are used as the measurements z_k for the FCD update step. These measurements z_k divided by the penetration rate (contained in the observation model matrix H) are compared with the prior flow estimates x_k to give the innovation y_k. of this update step.

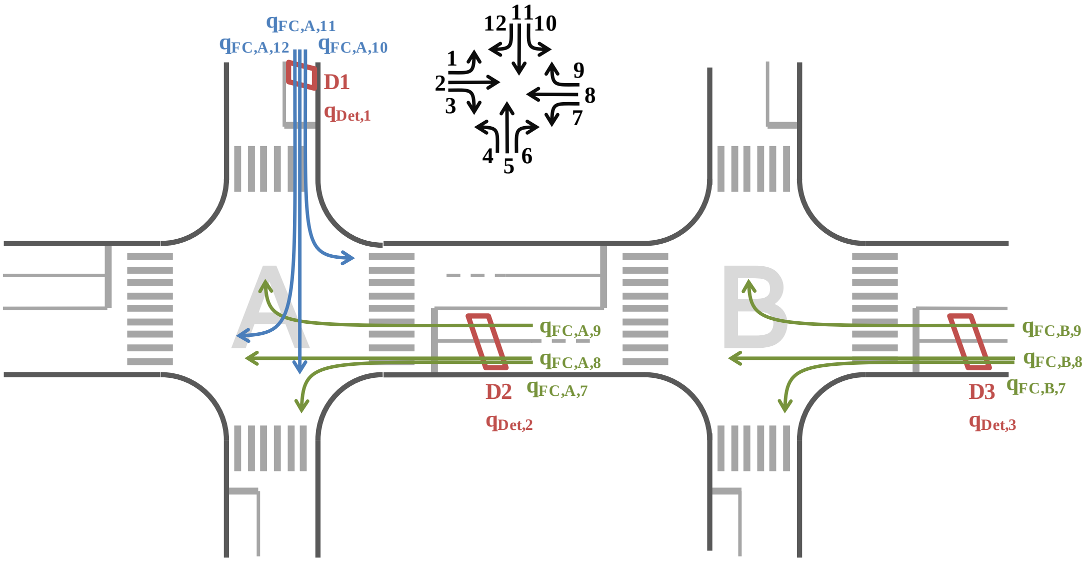

Figure 2 Illustration for the determination of the penetration rate from detector counts

This procedure works particularly well, if the penetration rate is constant in the network. An error is introduced if the probe flows are not representative for the whole vehicle population and the number of detectors is low. If the deviation is systematic and known, it can be incorporated in the data fusion by correction factors on certain links. The methodology will deliver more reliable results, if many detectors are used. The penetration rate will be accurate at locations where only one vehicle movement is detected (e.g. exclusive turning lane) given accurate probe and detector counts. In every case an error might result from inaccurate detectors. Unreliable detectors should be excluded from the data sources.

Historic data

Commonly the traffic flow at a certain time of the current day will be very similar to the flow at the same time on a similar day (e.g. working day, Tue-Thu). If no sufficient FCD is available and no unexpected changes are likely or no online information is desired, the flow estimation can be based on historic data. Historic data in this context depicts data gathered on past similar days. The accuracy of the historic data depends largely on the quality of the data used for generating the historic data and on the clustering of data to *similar* days (see below). Once the filter has been established, the historic database can be continuously updated with the filter outputs. This requires, of course, a good pattern recognition algorithm to detect incidents and special days.

It is crucial for the performance of the filter that the reliability of not only the historic data is correctly judged. The less reliable the historic data is, the more the filter will rely on the other data sources. The accuracy of the historic data principally depends on the variation of the traffic flow. This variation can be reduced by an intelligent clustering process. Historic data cannot predict anomalies like incidents. The online information from probes should consequently be preferred to historic data.

To use historical data, first a database containing traffic flow patterns for different days has to be established. The traffic flow on a link follows a particular pattern over time, i.e., for a whole year, and can be very well described by typical and standardized patterns [52]. To investigate traffic flow patterns, it is necessary to create clusters of typical traffic days [53]. Categories such as month of the year, week of the month, day of a week and hour of a day can be considered. Furthermore, conditions such as weather (e.g. dry, wet) and type of the day (typical day/holiday/weekend) can be included for detailed clustering. Given the historical data and the clustering scheme, one can categorize archived data and develop traffic flow patterns for each type of roads [54]. Using developed models, it is easy to generate an artificial, but rather reliable, traffic flow pattern for a particular link.

IMPLEMENTATION ON CORRIDORS AND NETWORKS

Data fusion on corridors

If the traffic has to be estimated at two or more adjacent intersections, the link between these intersections can be used as a source of information. Namely the vehicles leaving one intersection in direction of the next intersection will arrive there after their travel time. Only if significant entries, exits or on-street parking facilities exist on the connecting link, the volumes will change.

The network only offers an insight into the sum of vehicles leaving one intersection. Commonly no information is given on the assignment of these vehicles to different streams at the next intersection. The update consists, thus, of a scaling of the input and output streams of the respective adjacent intersections relative to each other. The scaling factor is determined by the sum of the inflows in relation to the sum of the respective outflows. The distribution of the inflow among different streams is left unchanged.

Figure 3 illustrates the procedure for two intersections A and B. The procedure described here assumes short travel times in comparison to the aggregation interval and no entries/exits between A and B, i.e. the traffic volumes leaving A towards B ($q_{A,2}$, $q_{A,6}$, $q_{A,10}$) are assumed to equal the traffic volumes reaching B from A ($q_{B,1}$, $q_{B,2}$, $q_{B,3}$) during the same interval (blue) and vice versa (green).

The measured streams, however, will differ from this equation due to measurement errors. The difference can be expressed as a factor f calculated as the ratio of inflow q_{in} to outflow q_{out} at cross-section A-B. Four factors are calculated for each connecting link, one for each inflow and outflow. Eq. (10) exemplifies the determination of the scaling factor for the traffic from A to B. The two factors for the traffic from B to A are calculated along the same lines.

$$f_{B,in} = \frac{\sum q_{A,out,i'}}{\sum q_{B,in,i}} = \frac{q_{A,2}+q_{A,6}+q_{A,10}}{q_{B,1}+q_{B,2}+q_{B,3}} \qquad (10)$$
$$f_{A,out} = \frac{\sum q_{B,in,i}}{\sum q_{A,out,i}} = \frac{q_{B,1}+q_{B,2}+q_{B,3}}{q_{A,2}+q_{A,6}+q_{A,10}}$$

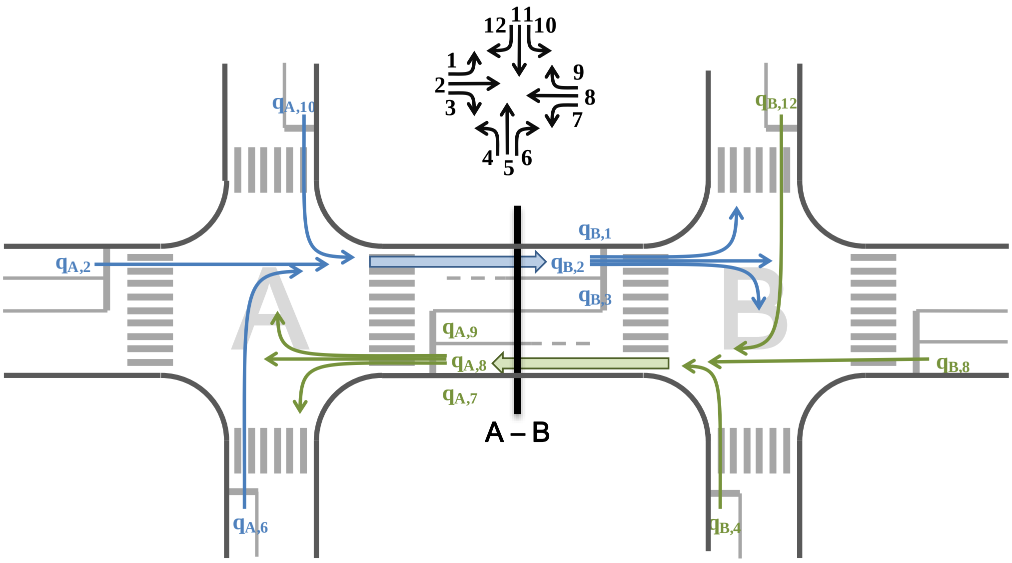

Figure 3 Illustration of using the information contained in the network

All twelve streams involved in this update are converted into "measurements" by scaling the original estimates. The scaled stream volumes are used as the measurement input for the Kalman Filter as shown in Eq. (11).

$$z_{in,i} = f_{in} x_{in,i} \qquad z_{out,i} = f_{out} x_{out,i} \qquad (11)$$

with z_{in} and x_{in} representing the streams to be updated (i.e. for A→B: $q_{B,1}$, $q_{B,2}$, $q_{B,3}$ for the inflow and $q_{A,2}$, $q_{A,6}$, $q_{A,10}$ for the outflow).

The measurement error R depends on the state estimate error covariance P of the streams used for the scaling. If the traffic volumes at intersection A, for instance, are more reliable due to better availability of FCD or detector counts, more trust is put into the outflow of A than into the inflow of B (for A→B). The Kalman Filter automatically weights the measurements z in relation to this reliability. The determination of the measurement error R is exemplified in Eq. (12) for the inflow into B.

$$R_{B,1} = R_{B,2} = R_{B,3} = P_{A,2} + P_{A,6} + P_{A,10} \qquad (12)$$

The Kalman Gain determines the extent to which the scaled estimates as measurement input change the current state estimate; cf. Eq. (5) and (6). If the error is zero, the outflow and inflow sums have to be equal, or the other way round, if the sums are not equal, the measurement error has to be at least as big is the difference between the two sums.

Consideration of travel times and entries/exits along links

The travel time of the vehicles on the connecting link between two adjacent intersections can only be neglected, if it is significantly shorter than the observed time interval. The more similar travel times and observation time interval become, the higher the involved error resulting from the update step using the network context will be.

A procedure taking these travel times into account has to derive the measurement inputs z not only from the current time interval, but also from the recent time intervals. The extent to which the earlier intervals should be used depends on the travel times and its distribution. The calibration of the filter requires, thus, careful consideration. The variation of the travel times should also be reflected in the measurement noise matrix *R*.

If entries and exits on a link connecting two adjacent intersections exist, the sum of vehicles reaching the next downstream intersection has to be adjusted accordingly. This is again connected to some error due to the varying number of vehicles leaving or entering the link and lacking knowledge about their volumes, which has to be considered in the measurement error. Alternatively, the entry/exit can be considered as another intersection.

FILTER PERFORMANCE

The Kalman Filter was tested in a controlled environment. Representative traffic data was generated for two intersections connected by a link as illustrated in Figure 3. By changing the variation of the traffic flow, varying the reliability of detector data and FCD, and varying the penetration rate of the traffic with probe vehicles the sensitivity of the filter to these conditions could be tested. Here the root mean squared error (RMSE) is used as the primary performance measure. Maximum and average deviations of estimated traffic volumes from the true values are given for comparison.

Calibration variables

The performance of the filter depends on the correct weighting of the measurements, i.e. the accurate or suitable judgment of the noise terms in relation to each other. Because the accuracy of the different measurements will not be exactly known in an application context, the filter has to be calibrated by varying the different noise terms and comparing the output with a control sample. The noise is contained in the measurement and process noise matrices *Q* and *R*.

Data generation

For a performance analysis artificial data has been generated for a pair of intersections. A flow pattern for one day in 15 minute intervals was defined. The assumed traffic flow was generated based on this pattern, but varied following a Normal Distribution. 100 data sets generated in this way served as the historic data. One set was excluded and used as the current flow. The FCD and detector counts deviate from the assumed current flow following again a Normal Distribution with given standard deviation. The historic data set and the randomly selected current flow are shown in Figure 4 with the variation parameters provided in Table 1.

Table 1 Parameters used for base scenario data generation

Parameter	Value	Explanation
Flow variation	10 %	Variation of traffic volumes around given mean
Probe variation	10 %	Variation of penetration rate
Penetration Rate	10 %	Penetration rate of total traffic with probes
Detector Variation	10 %	Variation of detector counts around true number
Monte Carlo Runs	20	

Different penetration rates, probe vehicle variation (representativeness of the probes for the total population), fluctuations of traffic flow rates around the defined pattern, and accuracies of the detector measurements were used to analyse the performance of the filter with reference to the base scenario as in Table 1. The noise terms of the filter have been based on the known variation of the different inputs (FCD, detector counts, historic data).

The flow variation describes how much the traffic volumes vary around a fixed pattern from day to day. For this analysis it is assumed that the average of the historic data describes a typical flow pattern. The individual flow follows this pattern with certain accuracy (the flow variation). Incidents are excluded from this analysis.

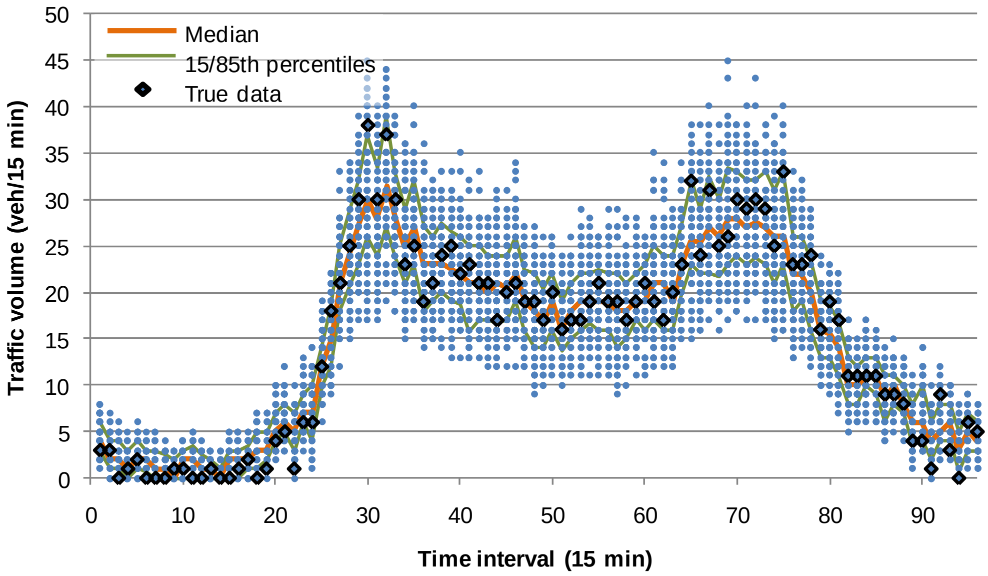

Figure 4 Sample generated data sets and the randomly selected dataset to represent real world data

Performance on intersection level

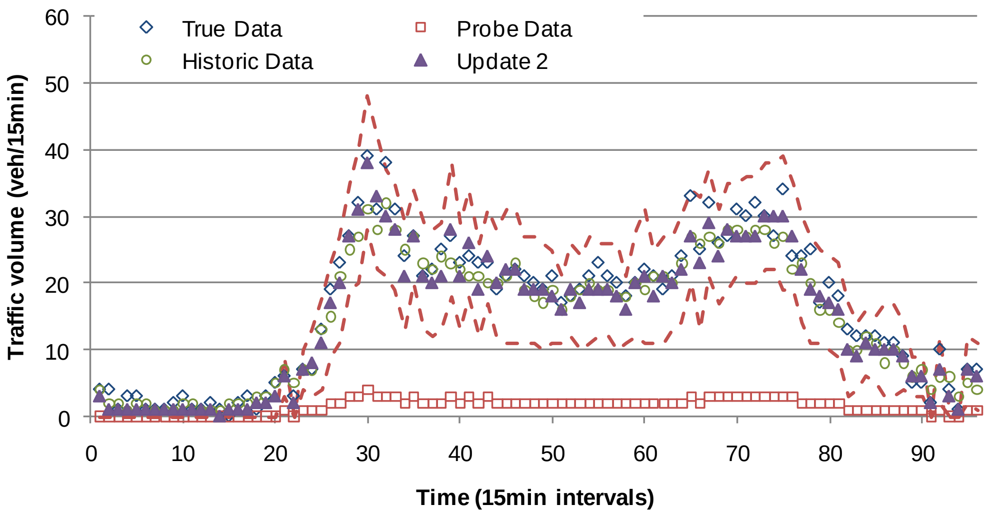

Figure 5 Filter output for single intersection, stream qB,1

Figure 5 shows the result of the filter for stream $q_{B,1}$ a penetration rate of 10 %, a flow variation of 10 % and a detector reliability of 10 %. The dashed lines represent the 90 % confidence interval. It can be seen to which extent the filter relies on the historic data and the FCD. Due to the comparably high penetration rate, the FCD is more relied upon, which is crucial in order to detect incidents and make the filter independent from prior knowledge of the traffic flow, which is often not available.

Figure 6 shows the RMSE of the filter as a function of the differing inputs (pair-wise varied). a) and b) show the sensitivity of the model to the penetration rate of probe vehicles averaged over all streams and for a single stream (tream $q_{A,3}$ in Figure 3), respectively. As expected the estimation error decreases as the penetration rate increases. The trend is non-linear. Figure 6 c) and d) represent the model sensitivity to the random variation of traffic counts at each stream. While the error increases as the flow variation increases, the error increases for the single stream as traffic volumes at a single stream are much more sensitive to the variations.

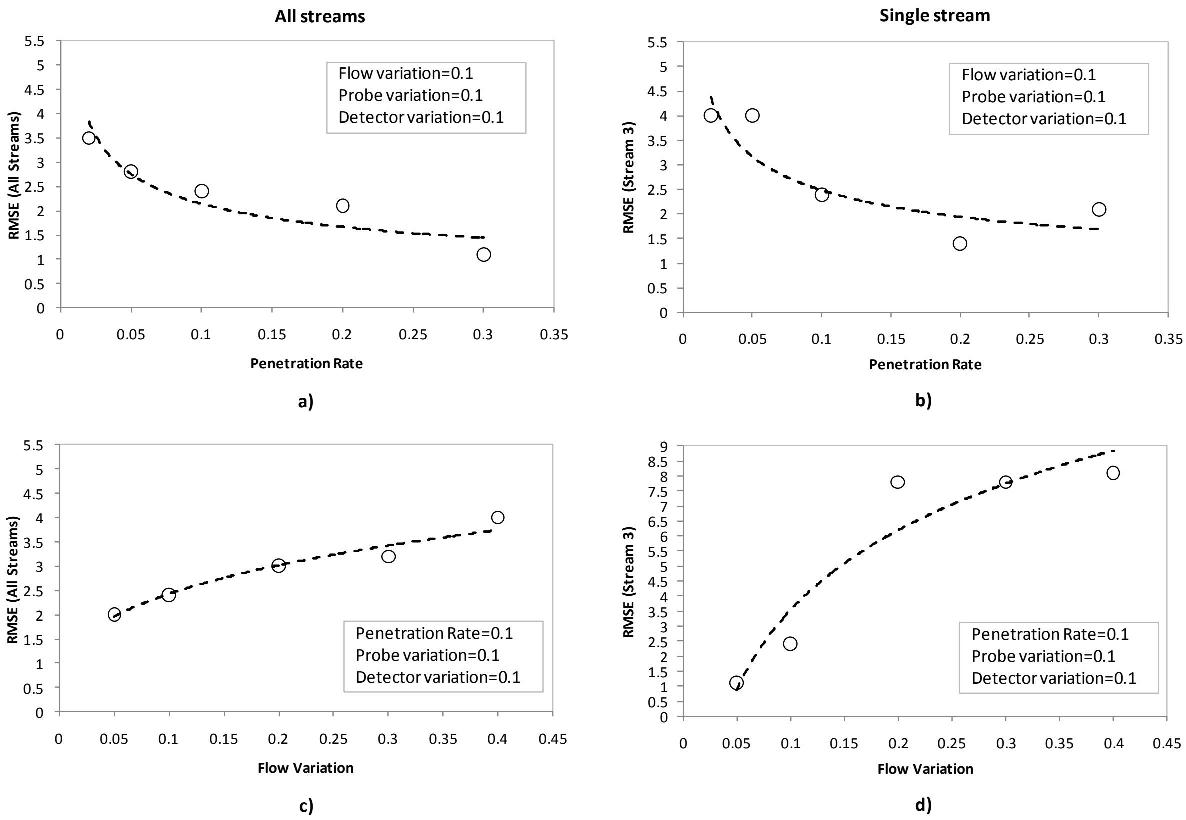

Figure 6 Elasticity of Kalman Filter to different penetration rates and flow variations given by the Root Mean Squared Error (RMSE)

Probe vehicles are randomly sampled from the whole population of the vehicles. As a result, the number of observed probe vehicles in a particular stream may not perfectly represent the whole number of vehicles in that stream. Such representativeness is expressed by the probe variation. The elasticity of the model to the variations in the number of probe vehicles is investigated in Figure 7 a) and b). When probe variation increases, that is the probe data represents the whole population of the vehicles less accurately, the estimation error increases. The estimation error for a single stream depends on the total traffic volume at that particular stream. As for the case of this comparison, estimation errors for stream $q_{A,3}$ seem to be less sensitive to changing variations. Finally, Figure 6 c) and d) show the impact of detector inaccuracies on the estimated traffic volumes.

This analysis underlines the importance of a sufficient penetration rate. For a moderate flow variation of 10 %, a penetration rate of at least 10 % is desirable. Of course, the required penetration rate depends on the local situation and the aspired accuracy of the filter. But it can be seen, that the accuracy of the filter changes only significantly for low penetration rates. If a critical penetration rate is exceeded, good results can be expected.

A major influence on the filter accuracy is exerted by the traffic volumes and the aggregation interval. The error of the FCD multiplies by the reciprocal of the penetration rate. And since traffic volumes take discrete values, the precision reduces with low traffic volumes (one vehicle presents a higher percentage of the total volume for low traffic volumes). The filter is less reliable for minor approaches and off-peak time. This means on the other hand, that for the situations commonly of major concern (major approaches during peak time), the filter performs well. The decision has to be met between higher accuracy with lower precision (long aggregation interval), and lower accuracy, but higher precision (short aggregation interval).

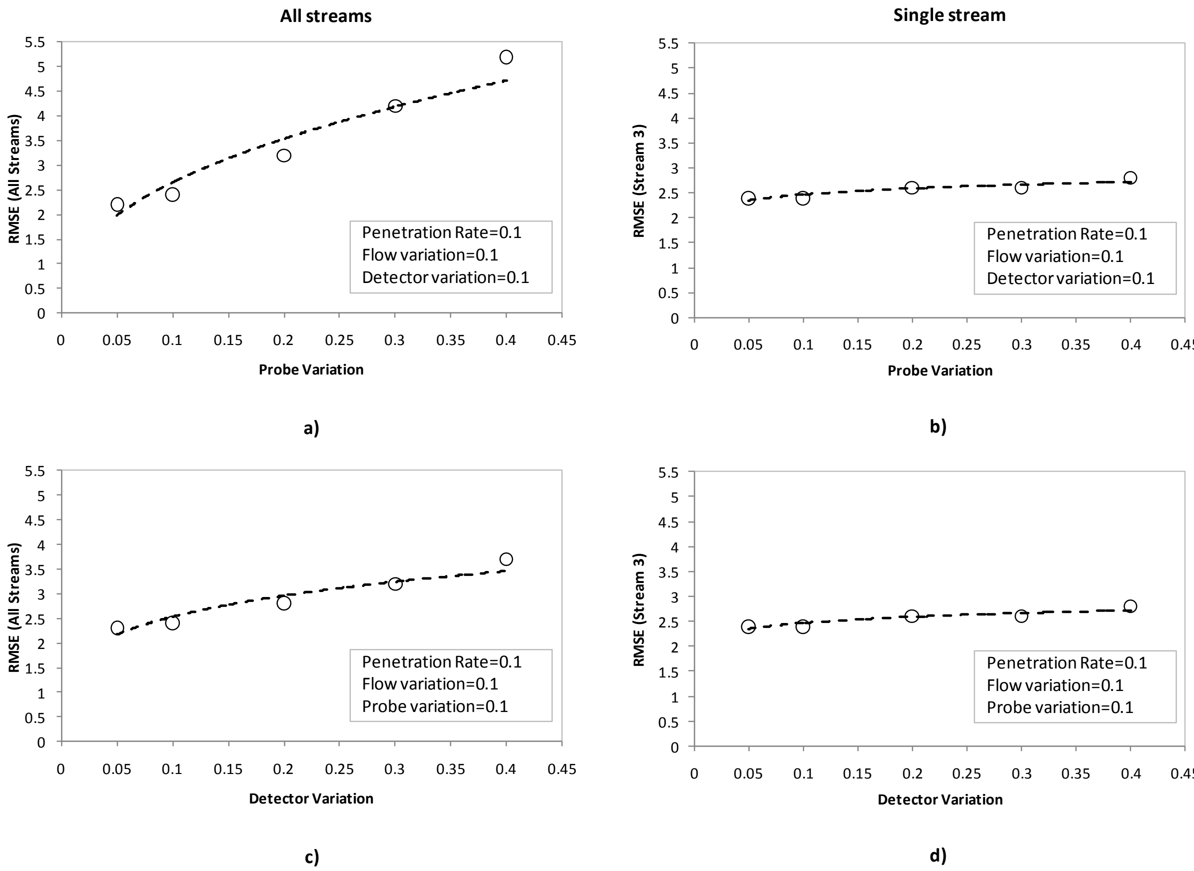

Figure 7 Elasticity of Kalman Filter to different probe and detector variations

Performance on network level

If the data fusion is extended to network level, the link volumes can be used to scale the outflows and inflows of adjacent intersections as described before. This procedure is particularly useful, if the data quality varies markedly between adjacent intersections. Figure 8 illustrates the effect of this third update step (after FCD and historic data). Shown are the true (blue empty diamonds) and estimated traffic volumes of stream $q_{B,8}$ distinguished for the three update steps (dark green: FCD only; light green: FCD and historic data, HD; purple dots: FCD, HD, and network update) together with the cumulated total error.

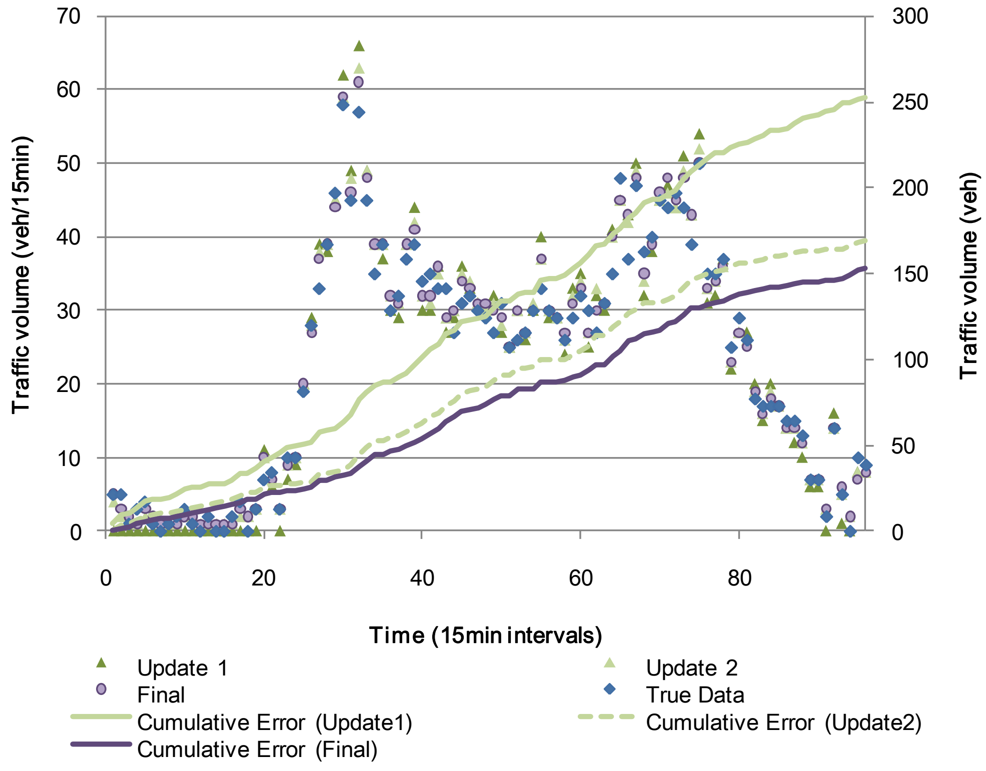

Figure 8 Filter output for three update steps on one stream (penetration rate: 10 %)

To avoid a bias by the random data generation, a Monte Carlo Simulation has been run with 20 different random seeds. The average performance (RMSE, minimum, maximum and average difference between true data and data fusion estimates) is shown in Table 2, distinguished for the three update steps and for a stream with low traffic volumes ($q_{B,1}$) as compared to a stream with high traffic volumes ($q_{A,2}$).

Table 2 Data fusion performance (Penetration rate: 10 %, 20 Monte Carlo runs)

Stream		FCD (veh/15min)	FCD&HD (veh/15min)	Filter output (veh/15min)
$q_{A,2}$	RMSE	4.1	3.0	2.8
	Min	-16	-15	-14
	Max	9	5	5
	Mean	-0.4	-0.2	-0.2
$q_{B,1}$	RMSE	2.9	2.3	2.3
	Min	-9	-8	-9
	Max	6	3	3
	Mean	-1.1	-1.2	-1.3

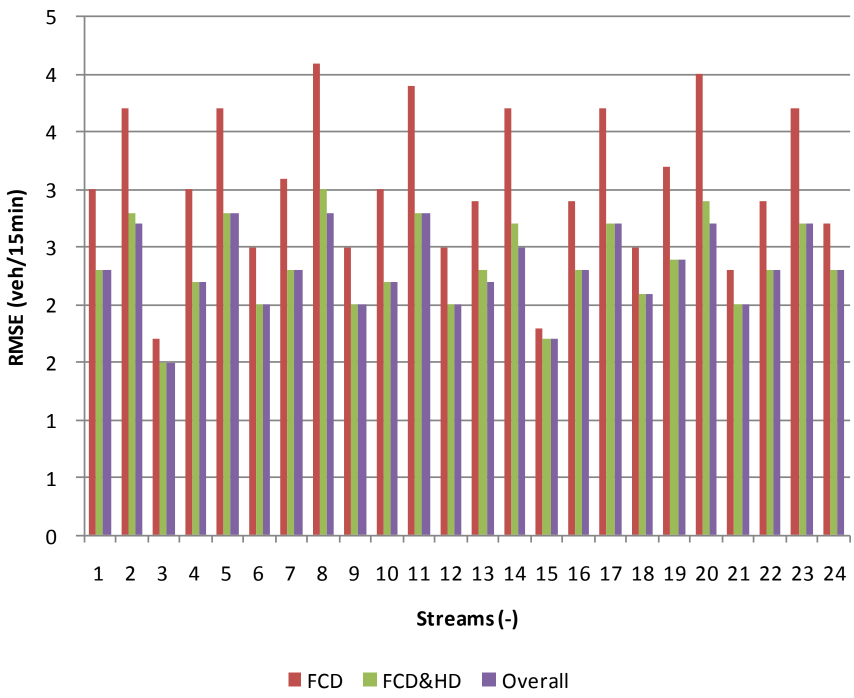

Figure 9 Root mean squared error (RMSE) of all streams (A: 1-12, B: 13-24) for a Monte Carlo Simulation with 20 runs

CONCLUSIONS AND OUTLOOK

The presented methodology drafts a new application of FCD for the generation of OD matrices on intersection level. These matrices (flow direction or lane based) are crucial for efficient signal control, intersection design, and further planning applications. The output of the data fusion process is up to date and can be used to take incidents and other fluctuations of traffic volumes into account.

Data from probe vehicles is fused with detector counts using a Kalman Filter. Historic data can be incorporated, if no current FCD is available or the penetration rate is low. The filter can further be extended by taking the information contained in the network itself into account. The flow relationships between adjacent intersections can be used as another input for the filter. The performance of the filter depends – apart from the penetration rate of floating cars, which should be more than 10 % – to large extent on the accurate estimation of the error involved in the different inputs. The Kalman Filter uses the given error terms together with the FCD, detector counts, and optionally the historic data and the network context as an input to compute a maximum likelihood estimate of the traffic volumes (Normal Distribution of the error assumed). The elasticity of the filter to different input reliabilities and flow variations has been shown. The concept works well and promises to be a useful tool for cities with sufficient FCD available. The performance depends in the first place on a good calibration, which can be achieved by extensive tests based on simulations and based on manual counts as part of an implementation.

The presented methodology is a starting point for more powerful data fusion. Ways to improve it are the consideration of different traffic volumes on different detectors (so far all detectors are equally weighted regardless of the traffic volumes), more sensibly dealing with constraints (so far negative volumes are set to zero), and the consideration of cross-correlations (so far all measurements are seen as independent, which they are not necessarily). Also the methodology for the network context might lead to feedback loops in closed networks.

REFERENCES

1. J. Sussman, Introduction to Transportation Problems, Artech House, Norwood, Massachusetts, 2000.
2. Federal Highway Administration. Data fusion for delivering advanced traveller information services, US DoT and ITS joint Program, 2003, pp. 247–254.
3. N.-E. El Faouzi, Data Fusion: Concepts and Methods, Monograph: INRETS, 2000 (INRETS Edition).
4. N.-E. El Faouzi, H. Leung, A. Kurian, Data fusion in intelligent transportation systems: Progress and challenges – A survey, Information Fusion 12 (2011) 4–10.
5. Federal Highway Administration. Data fusion for delivering advanced traveller information services, US DoT and ITS joint Program, 2003, pp. 247–254.
6. A. Tarko, N. Rouphail, Travel time fusion in ADVANCE, in: Pacific Rim TransTech Conference: A Ride into the Future, July 25–28, 1993, Seattle, Washington, USA.
7. P. Thakuriah, A. Sen. Data fusion for travel time prediction: a statement of requirements, Advance Working Paper Series, vol. 22, 1993, Task TRF-TT-07, Urban Transportation Center, University of Illinois, Chicago.
8. S. Berka, X. Tian, A. Tarko, Data fusion algorithm for ADVANCE, Advance Working Paper 48, 1995.
9. D.J. Dailey, P. Harn, P.-J. Linet, ITS data fusion, Research project T9903, Task 9: ATIS/ATMS Regional IVHS Demonstration, Report# WA-RD 410.1, Washington State, WADOT, 1996.
10. N.-E. El Faouzi, J.-B. Lesort, Travel time estimation on urban networks from traffic data and on-board trip characteristics, in: Proceedings of the Second World Congress on ITS, Yokohama, Japan, 1995, pp. 88–93.
11. N.-E. El Faouzi, L.A. Klein, O. De Mouzon, Improving travel time estimates from inductive loop and toll collection data with Dempster–Shafer data fusion, Transportation Research Record: Journal of the Transportation Research Board 2129 (2009) 73–80.
12. N.-E. El Faouzi, Travel time estimation via evidential data fusion, Recherche Transports Sécurité 68 (2000) 15–30.
13. K. Choi, Data fusion for generating the link travel time with insufficient data sources, in: Proceedings of the I.V.H.S., 1997.
14. K.Y. Choi, Y. Chung, Travel time estimation algorithm using GPS probe and loop detector data fusion, in: Proceedings of the 80th TRB Annual Meeting (CDRom), Paper No. 01-0374. Washington DC, Transportation Research Board, 2001.
15. F. Koppelman, V. Sethi, J. Ivan. Calibration of data fusion algorithm parameters with simulated data, ADVANCE Project, Technical Report, TRF-ID-152, 1994.

16. J.N. Ivan, J.L. Schofer, F.S. Coppelman, L.L.E. Massone, Real-time data fusion for arterial street incident detection using neural networks, Transportation Research Record 1497 (1995) 27–35.

17. J.N. Ivan, Neural network representations for arterial street incident detection data fusion, Transportation Research C (1997) 247–254.

18. S.C. Byun, D.B. Choi, B.H. Ahn, H. Ko, Traffic incident detection using evidential reasoning-based data fusion, in: Proceedings of the 6th World Congress on ITS, Toronto, Canada, November 1999.

19. R.R. Murphy, Sensor and information fusion for improved vision-based vehicle guidance, IEEE Intelligent Systems (1998) 49–56.

20. S.-C. Pei, G. Liou, Vehicle-type motion estimation by the fusion of image point and line features, Pattern Recognition 31 (1998) 333–344.

21. J. Langheim, A.J. Buchanan, CARSENSE – sensor fusion for DAS, ITS in Europe 2002, Lyon.

22. C. Stiller, J. Hipp, C. Rossig, A. Ewald, Multisensor obstacle detection and tracking, in: Proc. 1998 IEEE Int. Conf. on Intelligent Vehicles, 1998, pp. 451– 456.

23. C. Wang, C. Thorpe, A. Suppe, Ladar-based detection and tracking of moving objects from a ground vehicle at high speeds, in: IEEE Intelligent Vehicles Symposium (IV2003), June 2003.

24. Y. Bar-Shalom, X.R. Li, T. Kirubarajan, Estimation with Applications to Tracking and Navigation: Theory, Algorithms, and Software, Wiley, New York, 2001.

25. J. Mueck, Estimation methods for the state of traffic at traffic signals using detectors near the stop-line, Traffic Engineering and Control, December 2002.

26. Y. Wang, M. Papageorgiou, Real-time freeway traffic state estimation based on extended Kalman filter: a general approach, Transportation Research B 39 (2005) 141–167.

27. B. Friedrich, R. Minciardi, Data fusion techniques for adaptive traffic signal control, in: IFAC Symposium on Transportation Systems, Tokyo, Japan, 2003, pp. 53–55.

28. L. Breiman, Bagging predictors, Machine Learning 26 (1996) 123–140.

29. L. Breiman, Arcing classifiers, Annals of Statistics 26 (1998) 801–849.

30. L. Breiman, Random forests, Machine Learning 45 (2001) 5–32.

31. D.H. Wolpert, Stacked generalization, Neural Networks 5 (1992) 241–259.

32. M. Cremer, H. Keller, Dynamic identification of OD flows from counts at complex intersections, in: Proceedings of the 8th International Symposium on Transportation and Traffic theory (ISTTT), Toronto 1981, University of Toronto Press.

33. M. Cremer, Determining time dependent trip distribution in a complex intersection for traffic responsive control, in: Proceedings on Control in Transportation Systems, IFAC-IFIP-IFORS, Baden-Baden, Germany, 1983.

34. I. Okutani, The Kalman filtering approaches in some transportation and road traffic problems. in: Gartner, Wilson (Eds.), Proceedings of the 10th ISTTT, 1987, pp. 397–416.

35. K. Ashok, M. Ben-Akiva, Dynamic O–D matrix estimation and prediction for real-time management systems, in: Proceedings of the 12th ISTTT, 1993, pp. 465–484.

36. M. Ben-Akiva, T. Morikawa, Methods to combine different data sources and estimates OD matrices, in: Proceedings of the 10th ISTTT, Cambridge, 1987, MA, pp. 459–481.

37. J.T. Lundgren, A. Peterson, S. Tengroth, Methods for pre-adjusting time dependent origin–destination matrices, in: Proc. 10th World Congress and Exhibition on Intelligent Transportation Systems and Services, Madrid, Spain, November, 2003, pp. 16–20.

38. J.F. Harrison, F. Stevens, A Bayesian approach to short-term forecasting, Operational Research Quarterly 22 (1971) 341–362.

39. I. Okutani, J. Stephanedes, Dynamic prediction of traffic volume through Kalman filtering theory, Transportation Research B 18 (1984) 1–11.

40. P.C. Vythoulkas, Alternative approaches to short term traffic forecasting for use in driver information systems, in: Proceedings of the 12th ISTTT, Berkeley, 1993, pp. 485–506.

41. N.-E. El Faouzi, Nonparametric traffic flow prediction using kernel estimator, in: Lyons, J.-B. Lesort (Eds.), Proceedings of the 16th International Symposium on Transportation and Traffic theory (ISTTT), Pergamon, Elsevier, 1996, pp. 41–54.

42. W.J. Granger, Combining forecasts: twenty years later, Journal of Forecasting 8 (1989) 167–173.

43. N.-E. El Faouzi, Combining predictive schemes in short-term traffic forecasting, in: A. Ceder (Ed.), Proceedings of the 17th International Symposium on Transportation and Traffic theory (ISTTT), Jerusalem, 26–28 July 1999, Pergamon, Elsevier, pp. 471–487.

44. M. Cremer, S. Schrieber, Monitoring traffic load profiles with heterogeneous data sources configurations, in: Lyons, J.-B. Lesort (Ed.), Proceedings of the 16th International Symposium on Transportation and Traffic theory (ISTTT), 1996, Pergamon, Elsevier, pp. 41–54.

45. J. Sau, N.-E. El Faouzi, A. Ben Aïssa, O. De Mouzon, Particle filter-based realtime estimation and prediction of traffic conditions, in: Proceedings of the ASMDA, 29 May 2007 to 1 June 2007, CHANIA, CRETE, Greece.

46. D.B. Choi, On multisensor data fusion using attributes association for intelligent traffic congestion information inference, in: Proceedings of the I.V.H.S., 1989.

47. M. Wei, K.P. Schwarz, Testing a decentralized filter for GPS/INS integration, in: Proc. IEEE Position Location and Navigation Symposium, Las Vegas, NV, March 1990, pp. 429–435.

48. W. Li, H. Leung, Constrained unscented Kalman filter based fusion of GPS/INS/ digital map for vehicle localization, in: IEEE 2003 International Conference on Intelligent Transportation Systems, Shanghai, China, October 2003, pp. 1362– 1367.

49. N. El-Sheimy, K. Chiang, A. Noureldin, The utilization of artificial neural networks for multi-sensor system integration in navigation and positioning instruments, IEEE Transactions on Instrumentation and Measurement 55 (2006) 1606–1615.

50. R. Sharaf, A. Noureldin, A. Osman, N. El-Sheimy, Implementation of online INS/ GPS integration utilizing radial basis function neural network, IEEE Aerospace and Electronic Systems Magazine 20 (2005) 8–14.

51. A. Hiliuta, R. Landry, F. Gagnon, Fuzzy corrections in a GPS/INS hybrid navigation system, IEEE Transactions on Aerospace and Electronic Systems 2(2004) 591–600.

52. W. Brilon, Traffic Flow Analysis beyond Traditional Methods, Proceedings of the 4th International Symposium on Highway, Hawaii, (2000).

53. M. Friedrich, I. Katrin, J. Prokop, O. Thomas, S. Johannes, Generating OD Matrices from Mobile Phone Trajectories, Transportation Research Board 89th Annual Meeting, 10-14 Jan 2010, Washington DC, USA.

54. G. Schmidt, Hochrechnungsfaktoren fuer Kurzzeitzaehlungen auf Innerortsstrassen (Estimation of Traffic Flow using Short-Period Traffic Counts for Urban Roads). Straßenverkehrstechnik, No. 11/96, (1996).