From Finance to ITS: Traffic Data Fusion based on Markowitz' Portfolio Theory

Thorsten Neumann, Rüdiger Ebendt, and Günter Kuhns*

November 9, 2015

This is the accepted version of the following article: Neumann, Thorsten; Ebendt, Rüdiger; Kuhns, Günter (2015): *From Finance to ITS: Traffic data fusion based on Markowitz' Portfolio Theory*, Journal of Advanced Transportation, DOI: 10.1002/atr.1351, which has been published in final form at `http://onlinelibrary.wiley.com/doi/10.1002/atr.1351/abstract`.

Abstract

Traffic data fusion has much to do with combining available or considered data sources in the best possible way. In this, it is very similar to optimizing a portfolio of financial assets in regard of return and risk. This article draws the analogy between these two mostly different scientific worlds, i.e. finance and engineering. Similarities and differences in context of weighted-mean data fusion based on numerical traffic flow measurements such as travel times or speeds are discussed. This, in particular, includes guessing the potential benefit of negative weights. Optimal weights are derived following a strict mathematical theory based on assumptions (parameters) about systematic bias and correlations of the considered data sources. Moreover, a specific way of reducing the systematic bias of the fusion results is proposed and compared to common methods. The whole approach is demonstrated based on position data from two independent vehicle fleets in Athens, Greece. In this context, the problem of parameter calibration is solved by applying an advanced tool for such floating car data systems, called "self-evaluation". The experiments show that the proposed methods reliably reduce the systematic bias and variance of the fusion results with regard to the original data as well as in comparison to the naïve fusion approach that uses equal weights for all data sources.

Keywords: Data fusion, linear model, weighted mean, travel time, variance, systematic bias.

1 Introduction

Data fusion has become "an inevitable tool" [1] in connection with intelligent transportation systems (ITS) over the last decades. By combining data of multiple sources, the quality of traffic information can be improved significantly. This finally enables new and better services for road users [2]. The recent survey articles by El Faouzi *et al.* (see [1, 3]) give a comprehensive overview about existing practices for a number of ITS applications. In this context, the authors distinguish between statistical methods (e.g. weighted mean), probabilistic approaches (e.g. Bayesian inference, Kalman filtering) and techniques based on artificial intelligence (e.g. neural networks). Moreover, the relevant literature (cf. [4, 5]) also mentions several layers of data fusion ranging from basic refinements of measurement signals to higher-level aggregation of information in order to fully describe the current state of a larger system, e.g. the overall efficiency of the transport network of a city.

Within this extensive framework, the present article focuses on the integration of pre-processed traffic data of one type, e.g. mean travel times or mean speeds, at the level of single road sections, thereby

*All authors are with the Institute of Transportation Systems, German Aerospace Center (DLR), Berlin, Germany (e-mail: thorsten.neumann@dlr.de; ruediger.ebendt@dlr.de; guenter.kuhns@dlr.de).

assuming that appropriate local subsystems are already in place for the data assimilation over time (by e.g. an extended Kalman filtering technique). Using the terminology of the unified framework given by Li *et al.* [6], such a fusion is called estimate fusion. The sources of such data can be conventional stationary detectors (e.g. induction loops), floating car data (FCD) systems, tracking of Bluetooth devices, automatic vehicle identification (AVI) or video surveillance (cf. [7–9]). Other means are possible as well including those which provide linkwise traffic state information based on suitable models only. While the present fusion approach requires all local estimates to be of the same type, homogeneity is not required at the sensor level, as long as every local subsystem transforms the local sensor data to the one agreed quantity. As is well-known, each source will have its own (context-dependent) quality in terms of a potential systematic bias and a specific error range (i.e. variance). Moreover, there may be correlations between some or all of the considered sources that need to be taken into account when fusing their data. This is in order to avoid giving too much weight to some parts of the available information.

As can be seen from literature, the combination of data based on diverse, e.g. on conventional and more recent detector technologies is of particular interest in context of traffic data fusion. Nanthawichit *et al.* [10] or Cipriani *et al.* [11], for instance, integrate loop data with floating car data by applying Kalman filter techniques with state equations based on a macroscopic traffic flow model. Mehran *et al.* [12] combine probe and fixed sensor data, implementing and enhancing a solution proposed by Daganzo [13, 14], based on the kinematic wave theory. In addition to that, Kong *et al.* [15] propose a model for the fusion of data from loop detectors and floating cars that uses evidence theory in order to increase the accuracy and robustness of mean speed information for urban road networks. Alternatively, the fuzzy regression model by Choi and Chung [16] can be implemented for the fusion of link travel times estimated from floating car data and loop detectors. But also the fusion of data for the same quantity to be estimated, and/or from multiple sensors of the same type has been addressed for traffic state estimation and in other areas, e.g. in signal or image processing, radar tracking, and portfolio optimization.

For traffic forecast, El Faouzi [17] discusses the use of constrained or unconstrained regression to combine l predicting models at a given time t for the same uncertain variable y_{t+h}. Yuan *et al.* [18] use a discretized macroscopic traffic flow model formulated in Lagrangian (vehicle number/time) coordinates (which move with the traffic stream), providing a set of observation equations to deal with floating car data. An extended Kalman filter (EKF) is used to combine the model predictions with the sensor observations. Arguments are given in favor of Lagrangian approaches which offer benefits in terms of both estimation accuracy and computation in comparison to a state estimator based on the same model formulated in Eulerian (space/time) coordinates which are fixed in space. In order to be able to integrate Eulerian sensors (loops, cameras, radar), suitable observation models (for local sensors) have to be derived, appropriately addressing the fact that the coordinates are no longer fixed but moving with vehicles.

Regarding other areas of application besides traffic data fusion, Kolosz *et al.* [19] proposed a combination of Analytical Hierarchy Process and Dempster-Shafer theory for prioritizing and fusing sustainability measures in context of ITS. Moreover, in the finance world, Markowitz' portfolio theory [20] calculates the optimal shares of financial assets in terms of minimizing the risk of the investment in total (in terms of minimum variance) while making sure of an optionally defined expected target return. Interestingly, a nice analogy can be drawn between this approach and traffic data fusion as is explained further in Section 2.2.

In their linearly constrained least squares (LCLS) approach for multisensor data fusion, Zhou *et al.* [21] combine sensory information $\underline{x}(t)$ in order to obtain a good consensus on the (one) signal $s(t)$. In contrast to a minimum variance solution like in [20], the expected power of the fused information is minimized. It is shown that this solution converges to the minimum variance solution when the number of measurements tends to infinity. A problem with the approach however is that a Gaussian noise environment is assumed, which does not generally hold in practical applications. Xia *et al.* [22] addresses this problem in a so-called cooperative learning algorithm for data fusion: a different objective function is used, the approach minimizes the absolute deviation of the fusion estimate from the original random signal. This is done following a problem formulation as a cooperative neural network, and the occurring ordinary differential equations are solved with the well-known Euler method. Sun *et al.* [23] give a decentralized Kalman filter where every sensor subsystem has a local optimal Kalman filter and independently estimates the states, respectively. The sensors are assumed to have correlated noises. The first layer of a two-layer fusion structure has a netted parallel structure to recursively determine the cross covariance between every pair of sensors at each time step. The second layer is the fusion center that fuses the estimates and variances of all local subsystems, and the cross covariance among the local subsystems from the first fusion layer to determine the optimal matrix weights and yield the optimal (i.e. linear minimum variance) fusion filter. Moreover, Xiao *et al.* [24] proposed a robust algorithm for the problem of fusion in a distributed network of sensors with dynamically changing topology. A more detailed comparison of the present fusion approach with the closest related techniques is given in Section 3.

Probably, the most common statistical method for fusing speed or travel time information (or other numerical measurements) is computing a weighted mean of the input data x_i where $i = 1, \ldots, n$ as provided by n given sources (cf. [16, 25]), i.e.

$$\hat{x} := \sum_{i=1}^{n} w_i x_i \tag{1}$$

with suitable weights $w_i \in \mathbb{R}$. Clearly, the question arising then is about the optimal weights in terms of minimizing the systematic bias and variance of the fusion result. In context of so-called meta-analysis, Brockwell and Gordon [26] give the corresponding answer in case of Gaussian distributed measurement errors and unknown Gaussian bias for all x_i. Moreover, El Faouzi [25] discusses the situation even without any assumption about specific distributions.

Interestingly, the whole topic is mathematically strongly related to the basic concepts of modern portfolio theory (cf. [20, 27]) in finance. This, however, has never been recognized in the ITS literature so far. For this reason, the next section starts with a very short review of the principles of portfolio optimization (see Section 2.1) and then draws the analogy between finance and traffic data fusion (see Section 2.2). As a result, this formal correspondence generates a new level of understanding of what happens in context of weighted-mean data fusion as in (1). This also includes guessing the benefit of possible negative weights w_i. Additionally, Section 2.3 derives an alternative formulation for dealing with biased input data while discussing the drawbacks of the nearest known approach as described by El Faouzi in [25]. The practical calibration of the weights w_i is part of Section 2.4 that, in particular, adapts the concept of so-called self-evaluation of floating car data (see [28]) for estimating systematic bias and variance of the input data. Finally, Section 3 gives a comparison to related fusion approaches.

Section 4 exemplarily shows the results of a prototypical implementation of the proposed algorithms that were applied to integrating the data of two complementary FCD fleets in Athens, Greece. The article ends up with some conclusions (see Section 5) including a short discussion of the fundamental difference of combining biased or unbiased input data in terms of the achievable accuracy (i.e. variance) of the fusion result.

2 Portfolio theory and traffic data fusion

Assume that there is someone having a fixed amount of money to be invested in buying stocks. But which stocks should he or she buy and what are the best ratios given the available assets where each of them has its own expected return and risk? Modern portfolio theory (cf. [27]) answers this question by calculating the optimal shares in terms of minimizing the risk of the investment in total while making sure of an optionally defined expected target return. In this context, the basic principles reach back to the year 1952 when Harry M. Markowitz – who was awarded the Nobel Prize in Economics in 1990 for his findings – published his pioneering article [20] about “Portfolio Selection”.

2.1 Principles of portfolio optimization

Given n (possibly correlated) different assets, let X_i be the random return of asset i where $i = 1, \dots, n$. Moreover, for all $i, j = 1, \dots, n$, denote the expected return of asset i by $\mu_i := \mathbb{E}(X_i)$ and the covariance of the assets i and j by $\sigma_{ij} := \mathrm{Cov}(X_i, X_j)$. In particular, $\sigma_i := \sqrt{\sigma_{ii}} = \mathrm{Var}(X_i)^{\frac{1}{2}}$ then represents the risk of asset i for all $i = 1, \dots, n$. Finally, let $C := (\sigma_{ij})_{i,j=1,\dots,n}$ be the corresponding symmetric $n \times n$ covariance matrix.

Any portfolio consisting of some or all of the assets is then defined by a vector $\underline{w} := (w_1, \dots, w_n)^T \in \mathbb{R}^n$ of relative shares where $\underline{w}^T \underline{1} = 1$ with $\underline{1} := (1, \dots, 1)^T \in \mathbb{R}^n$. Note, when so-called short selling (shorting) is allowed, w_i is not necessarily restricted to the interval $[0, 1]$, but may also be negative or greater than 1 for some $i = 1, \dots, n$. Regardless of that, the portfolio return $\hat{X}$ in each case is given by

$$\hat{X} = \sum_{i=1}^{n} w_i X_i. \tag{2}$$

Hence, minimizing the portfolio risk without defining an expected target return is equivalent to minimizing the variance of $\hat{X}$, i.e.

$$\mathrm{Var}(\hat{X}) = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j \sigma_{ij} = \underline{w}^T C \underline{w} \quad \longrightarrow \quad \min_{\underline{w} \in \mathbb{R}^n} ! \tag{3}$$

subject to $\underline{w}^T \underline{1} = 1$. In other words, by introducing the Lagrangian multiplier λ, the equation

$$\nabla_{\underline{w}}\, h(\underline{w}, \lambda) = \underline{0} \tag{4}$$

for the first derivative of h with regard to $\underline{w}$ has to be solved where

$$h(\underline{w}, \lambda) := \underline{w}^T C \underline{w} + \lambda \left(\underline{w}^T \underline{1} - 1 \right) \tag{5}$$

and $\underline{0} := (0, \dots, 0)^T \in \mathbb{R}^n$. One obtains

$$2C\underline{w} + \lambda \underline{1} = \underline{0} \tag{6}$$

which finally yields

$$\underline{w} = -\frac{\lambda}{2} C^{-1} \underline{1} \tag{7}$$

given that the inverse C^{-1} of the covariance matrix C exists. As can be shown, this is always the case if there is no riskfree combination of the considered assets.

The value of λ is then derived from the constraint $\underline{w}^T \underline{1} = 1$ or its (by transposition) equivalent form $\underline{1}^T \underline{w} = 1$, respectively, by plugging in (7) so that

$$\lambda = -\frac{2}{\underline{1}^T C^{-1} \underline{1}}. \tag{8}$$

Hence, the optimal (also called "minimum-variance") portfolio – given that no target return is defined – is determined by the vector

$$\underline{w}^* := (w_1^*, \dots, w_n^*)^T := \frac{C^{-1} \underline{1}}{\underline{1}^T C^{-1} \underline{1}} \tag{9}$$

and has the expected return

$$\hat{\mu}^* := \sum_{i=1}^{n} w_i^* \mu_i = (\underline{w}^*)^T \underline{\mu} \tag{10}$$

where $\underline{\mu} := (\mu_1, \dots, \mu_n)^T$.

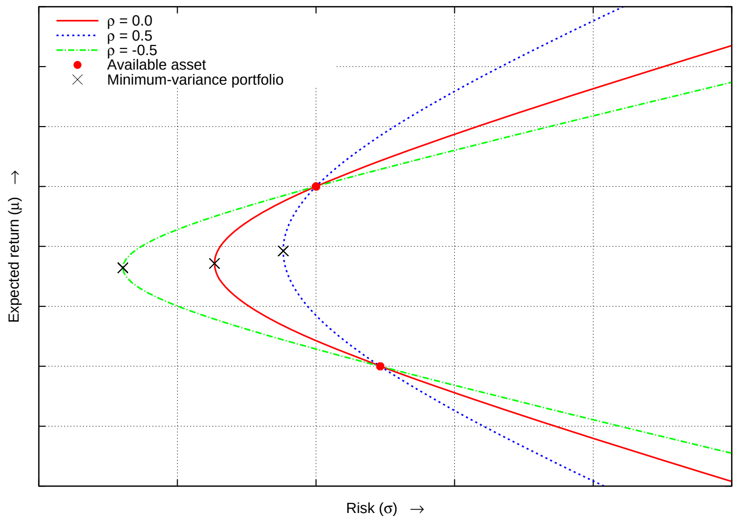


Figure 1: Schematical return-risk-diagram with optimal portfolio curves in case of two available assets with different correlations ρ.

Fig. 1 schematically shows the location of the minimum-variance portfolios in a return-risk-diagram (also called μ-σ-diagram) given two fixed assets with different correlations $\rho := \frac{\mathrm{Cov}(X_1,X_2)}{\sigma_1\sigma_2}$. As can be seen, these portfolios indeed reduce the overall risk compared to each single asset given that no expected target return is defined. In general, each optimal portfolio in terms of having the lowest possible risk for a given expected return $\hat{\nu}$ is located along one of the depicted curves that can be derived in a similar way as the optimal shares $\underline{w}^*$ from (9). In fact, one just needs to add the constraint $\hat{\nu} = \underline{w}^T \underline{\mu}$ to the optimization problem in (3). See the Appendix for a detailed description of the solution of this extended problem.

Finally, achieving a larger expected return than the maximum of all expected returns given by the single assets always induces a higher risk (cf. Fig. 1) and is possible only if shorting is allowed. The same holds if a lower expected return than the lowest one of all single assets shall be realized. Of course, this is not of interest from a finance point of view but may be important when transferring the whole concept to traffic data fusion further below. Needless to say, for real investments, only the upper branch of the optimal portfolio curves from Fig. 1 (also called "efficient frontier" [27]) is relevant as it obviously yields higher returns than the corresponding portfolios on the lower branch without changing the overall risk.

2.2 Adaption to traffic data fusion

Now, what is the analogy between portfolio theory as described above and traffic data fusion? As can be found, optimizing a portfolio directly corresponds to searching for the optimal weights $\underline{w} = (w_1, \ldots, w_n)^T$ when combining the (random) measurements x_i from a given set of n data sources (e.g. detectors or models) for $i = 1, \ldots, n$ according to

$$\hat{x} := \sum_{i=1}^{n} w_i x_i \tag{11}$$

as in (1) at a certain instant t of time and a fixed location ξ of the road network. At this point, x_i can be seen as a realization of a random variable X_i that is characterized by its expectation $\mu_i = \mathbb{E}(X_i)$ and its standard deviation $\sigma_i = \mathrm{Var}(X_i)^{\frac{1}{2}}$ for all $i = 1, \ldots, n$.

Why is that reasonable? And what is the physical meaning of σ? Consider the example of FCD used for estimating the mean speed (e.g. 1-minute-aggregates) at time t for the road cross-section ξ. Furthermore, for simplicity, assume that just one floating car has passed ξ during the relevant aggregation interval at time t so that the corresponding measurement x depends on the trajectory of this vehicle only. Then, following the common FCD approach (cf. [7, 29]), x could be the average travel speed, i.e. the inverse of travel time, between the nearest transmitted vehicle positions $\xi^{(1)}$ and $\xi^{(2)}$ of which one is located upstream and the other one downstream of ξ. Hence, x depends on the true mean speed $\hat{\nu}$ at ξ, of course, but is also affected by the rest of the vehicle trajectory that again is influenced by numerous additional random factors, namely the spatio-temporal variation of traffic flow on the driven route between $\xi^{(1)}$ and $\xi^{(2)}$.

That means, x is the sum of $\hat{\nu}$ and a random error term that may entail significant deviations in the measurements and may also be responsible for some possible systematic bias. In particular regarding σ, two different influencing factors are found that are hardly to be separated in practice, namely the independent "real" measuring errors due to the general degree of accuracy of the measurement devices on the one hand and the variations in traffic as explained above on the other hand.

Interestingly, the spatio-temporal patterns of traffic does not only reenforce the measurements X_i for $i = 1, \ldots, n$ to be random variables with non-zero variance, but may also induce correlations between them depending on how far each data source is affected by possibly overlapping surrounding traffic conditions. In contrast to FCD, the measurements of local detectors such as induction loops are mostly determined by what happens at the specific location where they are installed, for instance. So, the measurement error is more or less independent of the surrounding traffic. On the other hand, Bluetooth detection or automatic vehicle identification (AVI) may be examples that have a very similar behavior as FCD regarding this aspect. Consequently, let $C = (\sigma_{ij})_{i,j=1,\ldots,n}$ be the covariance matrix for the measurements X_i where $\sigma_{ij} = \mathrm{Cov}(X_i, X_j)$ for all $i, j = 1, \ldots, n$, and thus $\sigma_i = \sqrt{\sigma_{ii}}$ for all $i = 1, \ldots, n$.

Of course, the conformity of the notation here and in the previous sections is not quite accidental, but underlines the analogy between portfolio theory and traffic data fusion. In this way, the available data sources are the "assets", each of them providing a random measurement X_i ("return") where $i = 1, \ldots, n$. Moreover, the portfolio return $\hat{X}$ from (2) becomes the fusion result where w_i for $i = 1, \ldots, n$ is the weight (or "share") of the ith data source in the corresponding "detector portfolio".

Hence, all formulas from Section 2.1 (and the Appendix) can directly be applied to traffic data fusion, too. In this context, for all $i = 1, \ldots, n$, the standard deviations σ_i as well as $\hat{\sigma} = \mathrm{Var}(\hat{X})^{\frac{1}{2}}$ can be treated as the "risk" of observing measurements far from the corresponding expectation values $\mu_i = \mathbb{E}(X_i)$ and $\hat{\mu} = \mathbb{E}(\hat{X})$, respectively. In other words, poor data sources in terms of large variances are the "risky assets" that nevertheless might prove beneficial.

In order to demonstrate that, consider a hypothetic example with two uncorrelated data sources ("assets"). Moreover, let $\mu_1 = \mu_2 = \hat{\nu}$ where $\hat{\nu}$ is the true reference value (e.g. true link travel time in seconds). That is, none of the sources has a systematic bias. Regarding Fig. 1, that means the depicted curves become degenerated in such a sense that all possible portfolios (including the optimal portfolio and the assets themselves) lie on a horizontal line. Given $\sigma_1 = 1$ and $\sigma_2 = 2$, the minimum-variance portfolio, i.e. the optimal fusion then has a reduced standard deviation $\hat{\sigma}^* \approx 0.9$ following (9) and (3). Interestingly, the weights $\underline{w}^*$ are proportional to the reciprocals of the variances of each data source. This, by the way, holds whenever C is diagonal, i.e. in case of pairwise uncorrelated assets (cf. [16, 25, 26]). In particular, note that $w_i^* \geq 0$ for all $i = 1, \ldots, n$ in this situation.

Now, take a third data source that also has no systematic bias (i.e. $\mu_3 = \hat{\nu}$) but is much poorer in terms of its standard deviation $\sigma_3 = 6$. Moreover, let X_1 and X_3 be correlated with $\rho = 0.7$ (i.e., $\mathrm{Cov}(X_1, X_3) = 4.2$) while X_2 and X_3 are uncorrelated, i.e. $\mathrm{Cov}(X_2, X_3) = 0$. Based on (9), one obtains the weights for the minimum-variance fusion, namely $\underline{w}^* = (0.96, 0.14, -0.1)^T$, resulting in a significant reduction of about 18% regarding $\hat{\sigma}$ compared to fusing X_1 and X_2 only. That is, (3) yields $\hat{\sigma}^* \approx 0.74$ while still $\hat{\mu}^* = \hat{\nu}$. Obviously, negative weights ("shorting") are an important instrument for improving the quality of the fusion result.

So far, all X_i in this section have been unbiased measurements, i.e. $\mathbb{E}(X_i) = \mu_i = \hat{\nu}$ for all $i = 1, \ldots, n$

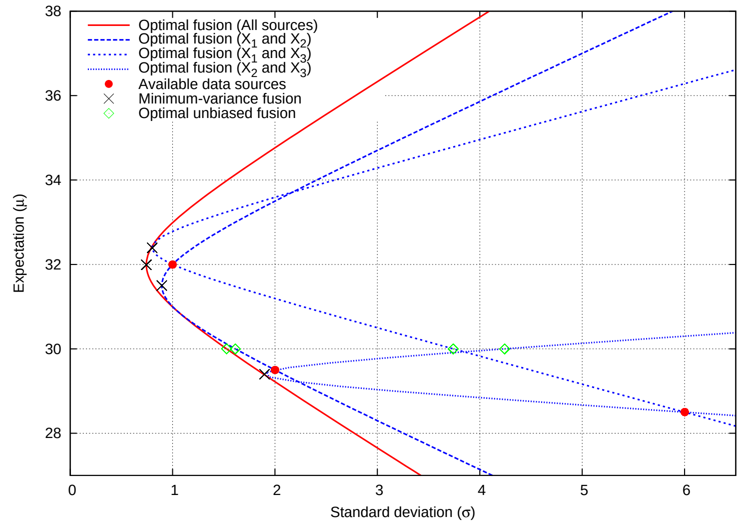


Figure 2: Optimal fusion curves in a μ-σ-diagram in case of three data sources.

where $\hat{\nu}$ is the true reference. For, in this case, the minimum-variance fusion automatically is an unbiased estimator of $\hat{\nu}$ because of $\underline{w}^T\underline{1} = 1$. In other words, according to (2), one obtains

$$\mathbb{E}(\hat{X}) = \mathbb{E}\Big(\sum_{i=1}^{n} w_i X_i\Big) = \sum_{i=1}^{n} w_i \mu_i = \hat{\nu} \sum_{i=1}^{n} w_i = \hat{\nu}. \tag{12}$$

But what happens if $\mathbb{E}(X_i) \neq \hat{\nu}$ for some or all $i = 1, \ldots, n$? Of course, it is still possible to compute the minimum-variance fusion as in (9). However, $\hat{X}$ will typically have a systematic bias in this case.

For instance, consider again the 3-assets-example from above, but now with $\mu_1 = 32$, $\mu_2 = 29.5$ and $\mu_3 = 28.5$ while all other values remain constant. Moreover, let $\hat{\nu} = 30$. Fig. 2 shows the corresponding optimal "portfolio curves" (cf. Section 2.1), i.e. the locations of the optimal pairwise combinations of the measurements as well as the optimal fusion of all three "assets" in a μ-σ-diagram as computed according to the formulas in the Appendix. Then, the minimum-variance fusion based on all three sources yields an expected value $\hat{\mu}^* \approx 32.0$ and thus a systematic bias of about 2.0. On the other hand, the optimal unbiased fusion has a standard deviation $\hat{\sigma}^* \approx 1.53$. Finally, Fig. 2 also shows that – in terms of minimizing the "risk" – fusing all measurements X_i for $i = 1, \ldots, n$ is always superior to combining just a (small) subset of them even if there are sources with very large standard deviations compared to others.

2.3 Handling of systematic bias

One of the major goals of data fusion is the reduction of the systematic error in case of biased input data. Note that it does not matter here whether the input is biased because of inaccurate sensor measurements or because of data pre-processing in case of high-level fusion (cf. "registration problem" [4]). The discussion above showed that Markowitz' portfolio theory provides all necessary tools for solving this task. For, it is always possible to make $\hat{X}$ as in (2) to be an unbiased estimator of the true value $\hat{\nu}$ ("target return") by adding the constraint $\underline{w}^T\underline{\mu} = \hat{\nu}$ (cf. Section 2.1).

From a practical point of view, that means it is not sufficient to guess the covariances σ_{ij} for all $i, j = 1, \ldots, n$ from historical measurements when calibrating the weights $\underline{w}$. But, one also needs knowledge about the expectations $\mu_i = \mathbb{E}(X_i)$ for all $i = 1, \ldots, n$ as well as the true value $\hat{\nu}$ based on some reference data that allow the offline computation of the optimal weights in advance. Afterwards, of course, these weights may be used as an approximate setting for fusing further online data, too, assuming that μ_i for

$i = 1, \ldots, n$ and $\hat{\nu}$ do not vary too much over time. For this purpose, it might be useful to define time slices with typically similar traffic conditions (cf. Section 2.4), for instance.

However, one of the drawbacks of this method is that μ_i for $i = 1, \ldots, n$ as well as $\hat{\nu}$ also differ from one road section to another even if all variations over time are neglected. Hence, in respect of area-wide data fusion, one would need an enormous amount of reference data during the calibration process which is not realistic. Consequently, the only option would be to assume that – for simplicity – the same values of μ_i for $i = 1, \ldots, n$ and $\hat{\nu}$ hold for a large number of roads in parallel.

As this is unrealistic as well, El Faouzi (see [17, 25]) proposes a slightly different approach for dealing with biased measurements in context of aggregative fusion schemes as in (1). He defines an estimator of $\hat{\nu}$ according to

$$\hat{X}' := w_0 + \sum_{i=1}^{n} w_i X_i \tag{13}$$

where $w_0 \in \mathbb{R}$ is an additional weight used for bias correction. Moreover, he drops the corresponding normalizing condition $\underline{w}_+^T \underline{1}_+ = 1$ with $\underline{w}_+ := (w_0, \ldots, w_n)^T$ and $\underline{1}_+ := (1, \ldots, 1)^T \in \mathbb{R}^{n+1}$ so that the weights $\underline{w}_+$ have no longer to sum up to 1. Finally, he computes $\underline{w}_+$ from some reference data via common regression methods.

But what does his approach mean to the optimal unbiased combination of the measurements in terms of minimizing the variance ("risk") of the fusion result? Obviously, the definition of $\hat{X}'$ yields

$$\mathbb{E}(\hat{X}') = w_0 + \sum_{i=1}^{n} w_i \mu_i \tag{14}$$

so that each unbiased estimator in (13) must satisfy

$$w_0 = \hat{\nu} - \sum_{i=1}^{n} w_i \mu_i \tag{15}$$

where $\hat{\nu}$ is the true reference again. Now, $\mathrm{Var}(\hat{X}')$ is to be minimized among all $\underline{w}_+ \in \mathbb{R}^{n+1}$. This is equivalent to

$$\begin{aligned} \mathrm{Var}(\hat{X}') &= \mathrm{Var}\Big(w_0 + \sum_{i=1}^{n} w_i X_i\Big) \\ &= \mathrm{Var}\Big(\sum_{i=1}^{n} w_i X_i\Big) \quad \longrightarrow \quad \min_{\underline{w}_+ \in \mathbb{R}^{n+1}}! \end{aligned} \tag{16}$$

Since $\mathrm{Var}\left(\sum_{i=1}^{n} w_i X_i\right) \geq 0$ for all $\underline{w}_+ \in \mathbb{R}^{n+1}$, the solution is trivial, namely $w_1^* = \ldots = w_n^* = 0$. Moreover, one obtains $w_0^* = \hat{\nu}$ based on (15).

Consequently, in terms of variance minimization, the optimal estimator $(\hat{X}')^* = w_0^* + \sum_{i=1}^{n} w_i^* X_i = \hat{\nu}$ is a simple constant that – except for its offline calibration – does not depend on any measurements and thus is not even a form of data fusion any more. Thus, the practical utility of such an approach is very limited.

However, keeping El Faouzi's idea of bias correction in mind, one may use the same formula as in (13), but with retaining the original constraint $\underline{w}^T \underline{1} = 1$ for $\underline{w} = (w_1, \ldots, w_n)^T$. In other words, the normalization constraint is relaxed for w_0 only while all other weights still have to sum up to 1. As can be seen, finding the weights $\underline{w}_+^*$ for the optimal unbiased combination of the measurements X_i with $i = 1, \ldots, n$ is then equivalent to a 2-step-approach with computing the "minimum-variance portfolio" according to (9) first and correcting the resulting systematic bias (cf. Section 2.2) afterwards by adding the term $w_0^* := \hat{\nu} - \sum_{i=1}^{n} w_i^* \mu_i$ (cf. (15)).

Obviously, here is the major difference between data fusion and portfolio theory. While manipulating a fusion result (i.e. its expectation $\hat{\mu}$) after minimizing its variance is very easy, it is impossible to change the expected return of a fixed portfolio in finance. That means, data fusion allows more flexibility in a certain sense. However, also the last described 2-step-approach has the same drawbacks as the basic idea from the beginning of Section 2.3 (i.e. adding the contraint $\underline{w}^T\underline{\mu} = \hat{\nu}$) for avoiding biased results. Namely, since the systematic error of the fusion results varies from one location to another, area-wide bias correction would again require an unrealistic amount of reference data covering all road sections. Consequently, one had to assume that the true reference $\hat{\nu}$ is constant for a large number of roads. This, of course, is mostly unrealistic as already discussed above.

For this reason, a slightly different approach (cf. [30]) is proposed here that allows for varying $\hat{\nu}$ but (given that $\mu_i \neq 0$ for all $i = 1, \dots, n$) instead assumes that $\overline{p}_i := \frac{\hat{\nu}}{\mu_i}$ for all $i = 1 \dots, n$ is (more or less) constant for all considered roads (or at least for suitable known sets of roads, cf. Section 2.4). In other words, let each data source have a fixed relative error regarding its expectation $\mu_i = \mathbb{E}(X_i)$.

From a practical point of view, it is possible then to guess $\overline{p}_i$ based on a sample set of m explicit measurement values $c_i^{(k)}$ of source i and m corresponding reference values $o_i^{(k)}$. For, let $p_i^{(k)} := o_i^{(k)}/c_i^{(k)}$ where $k = 1, \dots, m$ and $i = 1 \dots, n$, and define the sample mean $\overline{p}_i^* := \frac{1}{m}\sum_{k=1}^m p_i^{(k)}$ as an estimator for $\overline{p}_i$. Alternatively, $\overline{p}_i$ may be approximated based on average values of the measurements via

$$\overline{p}_i \approx \overline{p}_i^* := \frac{\frac{1}{m}\sum_{k=1}^m o_i^{(k)}}{\frac{1}{m}\sum_{k=1}^m c_i^{(k)}} \tag{17}$$

for $i = 1, \dots, n$ (cf. Section 2.4).

Given that $\overline{p}_i$ really is a fixed number and $\overline{p}_i^* = \overline{p}_i = \frac{\hat{\nu}}{\mu_i}$ for all $i = 1, \dots, n$, one then obtains

$$\mathbb{E}(\overline{p}_i^* X_i) = \frac{\hat{\nu}}{\mu_i}\mathbb{E}(X_i) = \hat{\nu}. \tag{18}$$

Hence, $Y_i := \overline{p}_i^* X_i$ is an unbiased random measurement of the true value $\hat{\nu}$. That means, the "minimum-variance portfolio" consisting of $Y_1, \dots, Y_n$ is an unbiased estimator of $\hat{\nu}$, too, and (9) can directly be applied for computing the optimal weights $\underline{\tilde{w}}^* = (\tilde{w}_1^*, \dots, \tilde{w}_n^*)^T$ for the fusion of all Y_i where $i = 1, \dots, n$, without taking care of any "target return". However, note that the modified covariance matrix $\tilde{C} := (\tilde{\sigma}_{ij})_{i,j=1,\dots,n}$ has to be used instead of C where

$$\tilde{\sigma}_{ij} := \mathrm{Cov}(Y_i, Y_j) = \overline{p}_i^*\overline{p}_j^*\mathrm{Cov}(X_i, X_j) = \overline{p}_i^*\overline{p}_j^*\sigma_{ij} \tag{19}$$

for all $i, j = 1, \dots, n$. Finally, the optimal (non-normalized) weights for combining the original measurements X_i for $i = 1, \dots, n$ according to (1) are given by the vector $\underline{w}^* = (\overline{p}_1^*\tilde{w}_1^*, \dots, \overline{p}_n^*\tilde{w}_n^*)^T$.

2.4 Parameter calibration

This section describes the approach chosen to calibrate the parameters for the present method of traffic data fusion in case of interpreting travel time data from two FCD systems, including the handling of systematic bias as described in Section 2.3. There are two families of parameters that must be estimated. The first is the set of parameters needed for the correction of the systematic bias, i.e. the $\overline{p}_i^*$ for $i = 1, \dots, n$. The second is the vector of optimal weights $\underline{\tilde{w}}^* = (\tilde{w}_1^*, \dots, \tilde{w}_n^*)^T$ for the actual fusion of the unbiased random measurements Y_i which, in particular, includes guessing their covariances $\tilde{\sigma}_{ij}$ for $i, j = 1, \dots, n$.

The first family of parameters, $\{\overline{p}_i^*\}_{i=1,\dots,n}$, is estimated with an advancement of a method called "self-evaluation" (cf. [28]). The approach followed here (cf. [30]) relies on two basic assumptions. Firstly, the approach of [28] is based on the assumption that the observed actual travel times for individual vehicle trajectories can be used as a ground truth for the mean link travel times computed by a FCD system.

Secondly, to extend this approach to periodically computing systematic biases for each link of interest, another assumption has to be made. It is assumed that it is possible to define corresponding periods (time slices) with typically similar traffic conditions.

If the requirements of the first assumption are met, then, for a particular observation period, the absolute systematic bias can be computed as the difference of two mean values, namely the mean actual trajectory travel time and the mean travel time on these trajectories computed by the FCD system (see [28]). More precisely, the first mean value is that of m observed actual travel times $o^{(k)}$ for $k = 1, \ldots, m$ for individual vehicle trajectories (denoted $\overline{o}$), and the second mean value is that of the travel times $c^{(k)}$ for $k = 1, \ldots, m$ computed by the FCD system along the same trajectories at the time of observation (denoted $\overline{c}$). The travel times $c^{(k)}$ for $k = 1, \ldots, m$ are computed by summing up the mean link travel times computed by the FCD system at the respective periods of travel on a trajectory, for all links constituting the respective individual trajectories [28].

It is of note that, in the scope of self-evaluation, the mean link travel times are computed without use of the link travel times observed for the vehicle that generated the respective trajectory [28]. In other words, yet it is computed as usual as the arithmetic mean of the travel times of all individual vehicles observed on that link during the respective period, but the link travel time of the vehicle that drove the ground truth trajectory is excluded from this arithmetic mean. This is done in order to avoid any circular reasoning that would be introduced by comparing an observation, namely the actual trajectory travel time, with a computed value (partly) based on exactly this observation.

The relative systematic bias then, of course, is the ratio of the absolute systematic bias and the mean observed actual trajectory travel time, given as percentage $\frac{\overline{c}-\overline{o}}{\overline{o}} \cdot 100\%$. Also notice that the method described in [28] yields only one global value for the overall systematic bias of the FCD system per observation period (e.g., one hour), and that only one data source (i.e., one vehicle fleet) is considered.

A practical implementation can use a digital road map. In such a map, links of the road network are usually tagged with constructional attributes like e.g. speed limits. According to Section 2.3, then assume that similar relative systematic biases are in effect on links with identical attributes in corresponding periods. That means, no further distinction needs to be made between such links, and, for every set of corresponding periods and every such set of links, the same value of the relative systematic bias can be used for the computations in the following.

In contrast to the method of [28], the present approach aims at a data fusion, and therefore n data sources (e.g., FCD from n vehicle fleets) are considered instead of only one. Therefore, the computation of systematic biases is done for each of the n data sources separately, that is, for $i = 1, \ldots, n$, percentage systematic biases are calculated as $\frac{\overline{c}_i-\overline{o}_i}{\overline{o}_i} \cdot 100\%$.

Moreover, whereas the previous method was used for the *diagnosis* of the (global) systematic bias only, the present approach also aims at *correcting* the systematic bias in the data. For this purpose, correction factors $\overline{p}_i^* = \frac{\overline{o}_i}{\overline{c}_i}$ for every data source $i = 1, \ldots, n$ are calculated alongside the systematic biases as in (17). To see that the $\overline{p}_i^*$ with $i = 1, \ldots, n$ as given here have in fact the property of correcting the systematic bias estimated from the m sample trajectories for data source i, or more precisely, from the m computed travel times $c_i^{(1)}, \ldots, c_i^{(m)}$, and the m observed actual travel times $o_i^{(1)}, \ldots, o_i^{(m)}$, notice that in analogy to (18)

$$\frac{1}{m}\sum_{k=1}^{m} \overline{p}_i^* c_i^{(k)} = \frac{\overline{p}_i^*}{m}\sum_{k=1}^{m} c_i^{(k)} = \frac{\overline{o}_i}{\overline{c}_i} \cdot \overline{c}_i = \overline{o}_i \tag{20}$$

holds.

For $i = 1, \ldots, n$, the trajectory data of source i is used for the calculation of $\overline{p}_i^*$. Due to the typically rather low penetration rates for FCD and the resulting lack of sufficient amounts of tracking data, it will often not be possible to do this separately for every individual link and for every period of interest. For this reason, the approach followed here calculates the correction factors with regard to L sets of links with identical constructional attributes, and T sets of corresponding time periods, respectively. Thereby it relies on the validity of the second assumption. The result is a separate set of parameters $\{\overline{p}_i^*\}_{i=1,\dots,n}$ for each of the L link sets and each of the T period sets, i.e. there are correction factors $\overline{p}_i^{(l)}(t)^*$ for all $l = 1, \dots, L$ and $t = 1, \dots, T$ where $i = 1, \dots, n$ (cf. (21)).

More precisely, for each data source i where $i = 1, \dots, n$, for each set L_l of links with identical constructional attributes where $l = 1, \dots, L$, and for each set T_t of corresponding periods where $t = 1, \dots, T$, a separate estimation of the true average travel times $\hat{\nu}^{(l)}(t)$ on the links in L_l for the periods in T_t is done, using trajectory data of source i. Thereby, $\hat{\nu}^{(l)}(t)$ is estimated as the average trajectory travel time $\overline{o}_i^{(l)}(t)$ for trajectories of source i on links in the particular set L_l, and during observation periods in T_t. In doing so, it is also assumed that the observed travel time along a trajectory can be allocated to individual links without introducing any further systematic bias. In other words, one assumes that a reintroduction of any significant systematic bias during the necessary arithmetic decomposition of the total trajectory travel time to individual links can be avoided by appropriate means (cf. [29]).

Then, separate estimations of the expected average travel time $\mu_i^{(l)}(t)$ for each data source i and the aforementioned links and periods are done, as the average travel time $\overline{c}_i^{(l)}(t)$ on links of trajectories of data source i that are also in the particular set L_l, and observed during periods in T_t, as computed by the FCD system, using tracking data from source i. The final correction factor used for data source i where $i = 1, \dots, n$, for each set L_l of links with identical constructional attributes where $l = 1, \dots, L$, and for each set T_t of corresponding periods where $t = 1, \dots, T$, is

$$\overline{p}_i^{(l)}(t)^* := \frac{\overline{o}_i^{(l)}(t)}{\overline{c}_i^{(l)}(t)}. \tag{21}$$

This is an estimator for $\hat{\nu}^{(l)}(t)/\mu_i^{(l)}(t)$ (cf. Section 2.3).

Notice that the approach only uses the trajectories of data source i when calculating the estimator $\overline{o}_i^{(l)}(t)$ for $\hat{\nu}^{(l)}(t)$. This is done in order to match the degree of data coverage on individual links during the computation of $\overline{c}_i^{(l)}(t)$ for data source i, respectively. For this reason, there are n estimators of the true average travel times $\hat{\nu}^{(l)}(t)$, namely $\overline{o}_i^{(l)}(t)$, one for every data source $i = 1, \dots, n$.

When estimating the second family of parameters, $\{\tilde{w}_i^*\}_{i=1,\dots,n}$ or as vector, $\underline{\tilde{w}}^* = (\tilde{w}_1^*, \dots, \tilde{w}_n^*)^T$, the following assumption is made: it is assumed that the covariances $\{\sigma_{ij}\}_{i,j=1,\dots,n}$ of the (original) data sources can be estimated appropriately as sample covariances, using historical measurements as a sample. The magnitude of the measurement error of the data sources will typically be affected by possibly overlapping, surrounding traffic conditions (cf. Section 2.2), and therefore the estimated covariances reflect how much the measurement errors of the sources change together and whether they have similar or opposite behavior.

Now, the key to estimating the covariances is to define time slices with typically similar traffic conditions as before: in the present approach, periods are considered as corresponding if and only if they define the same time of day (TOD), and the same day of week (DOW). This choice is based on the assumption that on the same day of week and at the same time of day, similar traffic conditions can be expected on the same road section. Then, measurements based on corresponding time slices can be collected for e.g. several months. For example, one could expect similar traffic conditions on a certain fixed road segment during all periods from 09:15 a.m. to 09:30 a.m. (i.e. time slices of 15 minutes) on all Wednesdays in the collected data of three months, resulting in a total number of $T = 24 \cdot 4 \cdot 7 = 672$ sets of corresponding time periods with 12 elements (time slices) per set if assuming that every month had exactly 4 weeks.

During calibration, the weights $\{\tilde{w}_i^*\}_{i=1,\dots,n}$ for fusing the bias-corrected measurements Y_i are determined following (9), in which, for all $i, j = 1, \dots, n$, the covariances σ_{ij} are replaced by the modified (i.e., bias-corrected) terms $\tilde{\sigma}_{ij}$ from (19). Notice that the assumed fixed relative errors $\overline{p}_i^*$ required here have already been estimated during the previous phase of calibration according to (21) as inspired by the approach of [28] where the dependency on the link sets (index l) and the sets of corresponding time periods (index t) is suppressed in the notation here. In fact, for all data sources $i, j = 1, \ldots, n$, the covariances $\tilde{\sigma}_{ij}$ are estimated as sample covariances from the aforementioned historical measurements. More precisely, given the period set T_t and the link set L_l, let $N := |L_l| \cdot |T_t|$ be the product of the number of elements of L_l and T_t. Then, the $\tilde{\sigma}_{ij}$ are calculated as

$$\tilde{\sigma}_{ij} = \frac{1}{N-1} \sum_{k=1}^{N} (y_i^{(k)} - \overline{y}_i)(y_j^{(k)} - \overline{y}_j) \tag{22}$$

where $\overline{y}_i = \frac{1}{N} \sum_{k=1}^{N} y_i^{(k)}$ and $\overline{y}_j = \frac{1}{N} \sum_{k=1}^{N} y_j^{(k)}$, respectively, and the $y_i^{(k)}$ with $k = 1, \ldots, N$ are the unbiased historical measurements belonging to the considered period and link set, that is, the bias-corrected realizations $y_i^{(k)} = \overline{p}_i^* x_i^{(k)}$ of the random variable $Y_i = \overline{p}_i^* X_i$ (indices l and t suppressed). The $x_i^{(k)}$ with $k = 1, \ldots, N$ are the original historical measurements of data source i, of course, that belong to the considered period and link set of interest as well (analogously, for $k = 1, \ldots, N$, the $y_j^{(k)}$ are the bias-corrected historical measurements of data source j).

The result of (22) then are separate covariance matrices $(\tilde{\sigma}_{ij})_{i,j=1,\ldots,n}$ for each of the T sets of corresponding time periods and for each of the L link sets. Thus, there are also different fusion weights $\{\tilde{w}_i^*\}_{i=1,\ldots,n}$ for each period set and each link set. Note that, in such a sense, the proposed data fusion is not static over time but accounts for varying traffic conditions even if there is no explicit dynamical model as it is used by other fusion approaches based on Kalman filtering, for instance (cf. Section 3).

3 Related Work

This section gives a comparison of the presented fusion approach to the closest related fusion methods [17, 18, 21–25] outlined in Section 1. A general difference is that the present approach proposes estimate fusion rather than sensor fusion, thereby assuming that appropriate local subsystems are already in place for the data assimilation over time (by e.g. an extended Kalman filtering technique). An advantage is the use for fusion of the data provided by already existing subsystems of independent technology partners, each such data collection and processing subsystem with their own characteristics with respect to the used technology and quality. This is possible with an only loosely-coupled (and therefore quick) technical setup since the fusion center of the resulting fusion system does not require any input from or change at the sensor level of the participating systems.

In contrast to previous approaches which assume unbiased measurements (with one notable exception, i.e. the works by El Faouzi [17, 25]), the present approach handles the case of biased measurements with a linear correction for every spatio-temporal "regime" of corresponding periods (time slices, e.g. in terms of DOW/TOD), and of corresponding links with identical constructional attributes. By that, the bias correction is handled via a discretized, time-varying linear transformation, using the relative error. Since fusion is at the level of estimates, a potential dependence of the relative error on the dynamics of traffic flow is not modelled explicitly. Nonetheless, the relative error (and also the covariance matrix) is assumed to be fixed for a particular spatio-temporal regime only, that is, it is still assumed to change over time, and with the constructional attributes of the considered links. This is motivated by the assumption that there are periodically repeating traffic flow patterns that result in similar sensor qualities for corresponding periods on all links with identical constructional attributes (cf. Section 2.4).

Moreover, the present approach estimates the required a priori information (such as the sensor cross covariances) by *vertical* rather than by *horizontal* sampling of the observed data: this means that covariances for one regime are calculated by sampling the measurements in all corresponding periods. This is different from horizontal sampling where each sample consists of measurements in subsequent periods, as applied in all aforementioned approaches.

More precisely, the multi-sensor optimal information fusion Kalman filter by Sun *et al.* [23] assumes white noises with zero mean. The LCLS approach by Zhou *et al.* [21] assumes Gaussian noises. Both the robust distributed sensor fusion method by Xiao *et al.* [24] and the cooperative learning algorithm by Xia *et al.* [22] mention the possibility of non-Gaussian errors, but they require them to be independent with zero mean. The individual methods of the unified framework by Li *et al.* [6] either assume unbiased measurements or that all biases are known a priori. The only work explicitly addressing the case of biased measurements is the aforementioned work on short-term traffic forecasting by El Faouzi [17] (see also [25]). Section 2.3 already gave a more detailed discussion of this approach.

Regarding the linear correction applied for bias-correction in the present approach: some but not all of the aformentioned methods apply a corresponding linear transformation, but then rather with the idea of expressing the sensor observations as an affine function of the system state (that is, relating the observed sensor measurement to the state of the modelled stochastic system by a linear equation, namely the sum of said linear transformation of the state and a random term for observation noise), and none of them gives a detailed discussion of a time-varying transformation. More precisely, the LCLS approach by Zhou *et al.* [21] does not apply a linear transformation. The same holds for the work on short-term traffic forecasting of El-Faouzi [17]. The multi-sensor optimal information fusion Kalman filter by Sun *et al.* [23] assumes that H, a linear transformation applied to the system state when relating it to a sensor measurement, is time-varying in general, but the approach does not discuss a concrete realization. It is of note that H is not subject to improvement at a transition t to $t+1$ between subsequent points in time, and also that H is often modelled as a constant matrix by other Kalman filtering approaches. When relating the unknown parameter to be estimated to a sensor measurement, Xiao *et al.* [24] also apply a linear transformation that does not change over time. The cooperative learning algorithm by Xia *et al.* [22] assumes a time-invariant vector of scaling coefficients when relating the sensor measurements to the original random signal. Finally, all methods described by the unified framework by Li *et al.* [6] (e.g. BLUE and WLS) assume that a corresponding linear transformation of the quantity to be estimated is not varying in time.

The present approach uses a different covariance matrix for every spatio-temporal regime. In particular, vertical sampling is used to determine an estimation of such a covariance matrix. By contrast, techniques based on Kalman filtering like [11, 23] usually estimate the initial covariances by horizontal sampling, which are then improved (corrected) at every discrete time step, based on appropriate recursive equations. Instead of covariances, the expected power of the fused information is used in the LCLS approach by Zhou *et al.* [21], which is estimated using horizontal sampling. Both the unified framework by Li *et al.* [6] and the robust distributed sensor fusion method by Xiao *et al.* [24] simply assume prior knowledge of the covariances rather than discussing how to estimate them empirically. In a first part of his work on short-term traffic forecasting, El Faouzi [17] uses horizontal sampling to estimate covariances, and later in a second part of the paper (which discusses stationary vs. non-stationary underlying processes), he proposes time-varying covariance matrices. They are estimated using the whole sample, but higher weights are assigned to the more recent observations. Finally, the cooperative learning algorithm by Xia *et al.* [22] does not require an empirical estimate of covariances, because it targets the least absolute deviation of the fusion estimate from the original random signal.

4 Implementation and results

A first prototype of the proposed approach for fusing travel times (or travel speeds) from various sources of traffic information (such as tracking data from FCD) has been implemented during the project SimpleFleet [31]. It has been applied to two floating car data systems in Athens, Greece. The vehicle fleets belong to the Greek telematics and fleet management service providers BK Telematics and Zelitron. They show different characteristics in terms of sampling frequency (BK Telematics: on average 0.22 samples per minute; Zelitron: on average 1.08 samples per minute), fleet size (BK Telematics: about 1,500 vehicles, Zelitron: about 600 vehicles), average number of reporting vehicles per 5-minutes batch of GPS samples

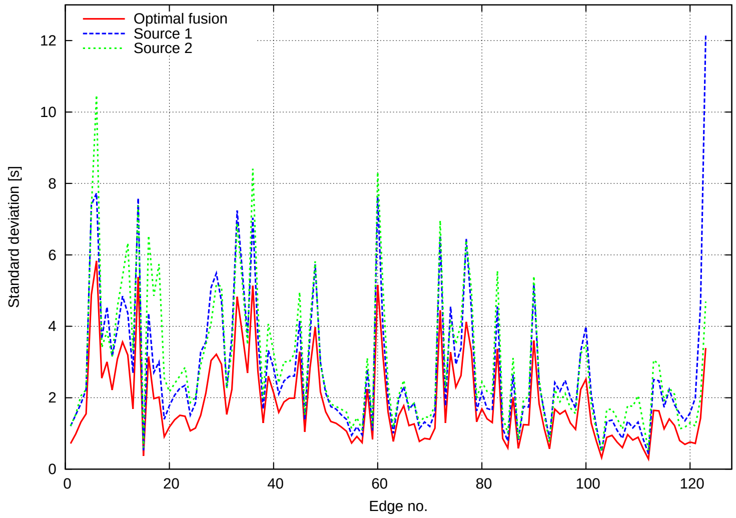
Figure 3: Route in Athens (dotted line), where FCD of two fleets have been collected, and where the proposed data fusion approach then has been used. © OpenStreetMap contributors (CC BY-SA).

(BK Telematics: 424, Zelitron: 399), and, as will be seen later in this section, also in terms of variance of the mean link travel times computed from their tracking data. In the following, results of the data fusion for a selected route in Athens are given in order to demonstrate the effectiveness of the approach. This route is part of the Greek motorway 1 and has a length of 9,471 m passing a motorway junction from southwest to northeast and vice versa, i.e., the road is bidirectional. It is shown in Fig. 3 where the stretch of interest is depicted as a dotted line.

Fig. 4 gives the results of the initial experiments. Here, the weights for the fusion have been derived using the mean link travel times as computed from three months of FCD (from December 2012 to February 2013). The average link length of the used OpenStreetMap digital map was 77.0 m (minimum: 12.3 m, maximum: 353.3 m). Then, in order to model a typical use case of offline computation, the fusion has been applied to the same data of the two sources, as collected during this observation period. In Fig. 4a, the abscissa shows the links of the examined stretch of road in spatial order of subsequent links. The ordinate shows the standard deviation of the mean link travel times in the considered time intervals, that is, the respective standard deviations have been computed based on all considered corresponding time slices (cf. Section 2.4) of the examined period of three months (for the experiments, time slices of 15 min have been used).

Firstly, rising edges, peaks, and falling edges of standard deviation are observed on adjacent links,

(a)



(b)

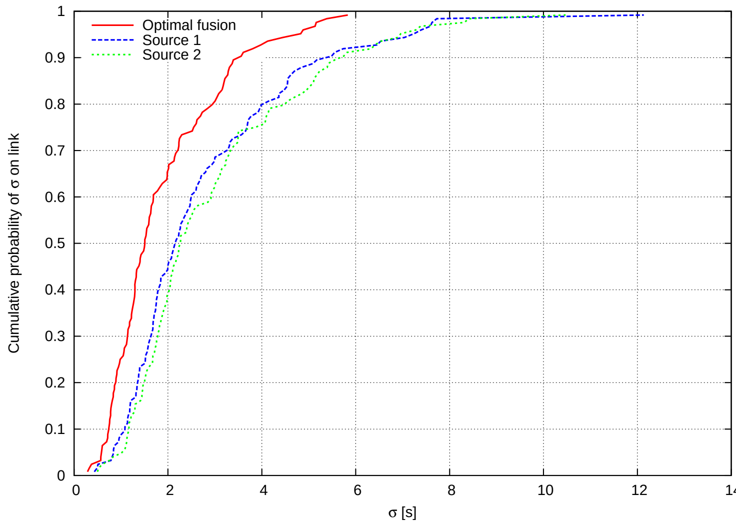


Figure 4: Offline use-case: comparison of standard deviations of the data sources with that of the fusion result (South-North route): (a) per link (b) as cumulative distribution.

representing the stochastic nature of traffic and differences in road condition. Secondly, as predicted, the standard deviation (and thus also the variance) of the fusion result is always smaller on average than those of the two data sources. This can be seen even better in Fig. 4b where the cumulative distribution of the observed standard deviations is shown. Obviously, there is significant shift to the left, i.e. to lower standard deviations in case of data fusion.

It is also interesting to compare the performance of the proposed fusion method to the naïve approach that assigns equal weights of $1/n$ instead of the optimal weights $\underline{\tilde{w}}^* = (\tilde{w}_1^*, \ldots, \tilde{w}_n^*)^T$ to the n data sources. Therefore, the experiment has been repeated, applying the constant weight of $1/2$ to the two data sources. The proposed fusion method with optimal weights reduced the standard deviation on a link by 32.2% on average, when comparing the fusion result to data source 1 (the Zelitron fleet), whereas the naïve approach yielded a respective reduction of only 7.0%. When comparing the results of the proposed fusion method to data source 2 (the BK Telematics fleet), a reduction in standard deviation by 36.5% on average has been achieved, contrasted by only 12.8% for the naïve approach (see Table 1).

Table 1: Average Reductions in Standard Deviation

	Offline Approach		Online Approach	
	Proposed	Naïve	Proposed	Naïve
Reduction for source 1 (%)	32.2	7.0	11.3	9.8
Reduction for source 2 (%)	36.5	12.8	13.5	12.0

Table 2: Systematic Bias

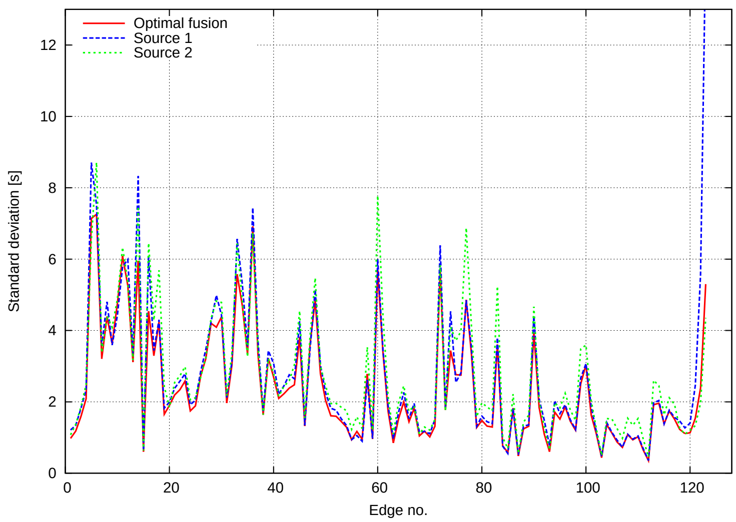
	Offline Approach	Online Approach
Systematic bias of source 1 (%)	2.32	0.11
Systematic bias of source 2 (%)	4.28	6.90
Systematic bias of fusion result (%)	-0.73	2.83

A next subject was to model the use case of an online computation. Calculating the optimal weights in general (i.e., for large n) involves quite complex mathematical operations, including e.g. the inversion of a rather large number of covariance matrices and the complete computation of all bias-correction factors. Therefore, continuously updating the weights online might not be feasible in practice. Instead, the following heuristic approach has been chosen: sets of weights are calculated *offline* for periods of three months (e.g., for each season of the year). Then, the weights calculated for a certain period are used for the corresponding period of the next year, too. That means the weights are not recalculated but are directly used for the *online* calculation of the weighted mean in (1). This approach is based on the assumption that similar traffic conditions can be observed in corresponding periods (e.g. for the winter of 2012/2013 and the winter of 2013/2014).

Thus, for a respective further experiment, the weights calculated during the first experiment (that is, for the period of December 2012 to February 2013, using the mean link travel times computed from these three months of FCD) have been used for applying the proposed fusion approach to the corresponding period of the next year, that is, December 2013 to February 2014. Fig. 5 gives the results of this second experiment (for the same route as in the first experiment). As before, also a naïve fusion approach has been applied to the period in question to allow for a comparison with the proposed method. The online approach reduced the standard deviation on a link by 11.3% on average, when comparing the fusion result to data source 1 (the Zelitron fleet), whereas the naïve approach yielded a smaller respective reduction of 9.8%. When comparing the results of the proposed online method to data source 2 (the BK Telematics fleet), a reduction in standard deviation by 13.5% on average has been achieved, contrasted by a smaller reduction, 12.0%, for the naïve approach (see Table 1).

Notice that in both the offline and the online approach, handling of systematic bias (cf. Section 2.3) has been applied. Table 2 gives the remaining systematic bias for the two (bias-corrected) data sources and the fusion result, for the offline and the online approach, respectively. It can be seen that the systematic bias in the fusion results remains small. For the offline case, it is even smaller than in both data sources. In this context, finally note that the bias in the original data, i.e. the original X_i instead of Y_i where $i = 1, 2$ was significantly larger, e.g. for data source 1 it was -11.04% for the data used in the offline approach, and -9.06% for the data used in the online approach.

(a)



(b)

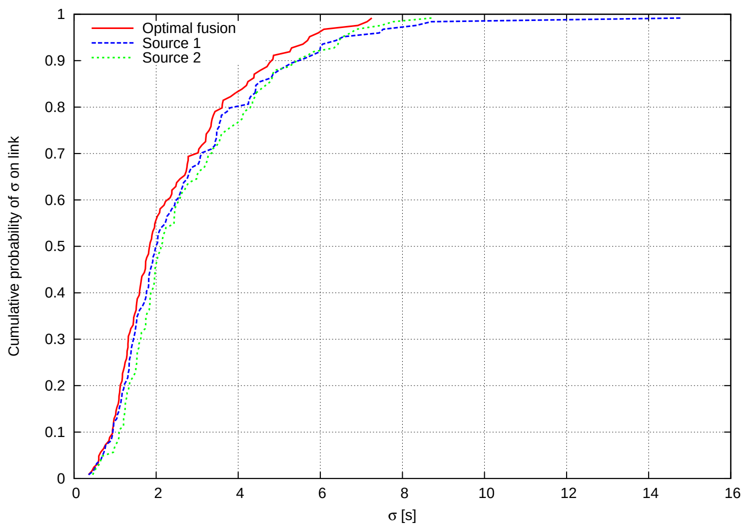


Figure 5: Online use case: comparison of standard deviations of the data sources with that of the fusion result (South-North route): (a) per link (b) as cumulative distribution.

5 Conclusions and further discussion

The experiments above show that the optimized weighted-mean data fusion in connection with the proposed bias correction from Section 2.3 reliably reduces systematic error and variance of the fusion result $\hat{X}$. In case of unbiased (or bias-corrected) measurements X_i, the resulting variance $\hat{\sigma}_i^2$ is in fact lower than the variance of each single X_i for all $i = 1, \dots, n$. For, the minimization in (3) guarantees that

$$\mathrm{Var}(\hat{X}) \leq \mathrm{Var}\Big(\sum_{k=1}^{n} w_k X_k\Big) \tag{23}$$

for all $\underline{w} = (w_1, \dots, w_n)^T \in \mathbb{R}^n$ with $\underline{w}^T \underline{1} = 1$. That means, with $w_k = 1$ for $i = k$ and $w_k = 0$ else for any fixed $i \in \{1, \dots, n\}$, one obtains

$$\mathrm{Var}(\hat{X}) \leq \mathrm{Var}(w_i X_i) = \mathrm{Var}(X_i) = \sigma_i^2 \tag{24}$$

as proposed.

But note that this does not necessarily hold for unbiased measurements when the additional constraint $\underline{w}^T\underline{\mu} = \hat{\nu}$ is used for defining the target expectation value of the fusion as in the original formulation of portfolio theory (cf. Section 2.1 and the Appendix). Let, for instance, X_1 and X_2 be two uncorrelated random measurements with $\mathbb{E}(X_1) = \mu_1 = 10$ and $\mathbb{E}(X_2) = \mu_2 = 15$ while the true reference is $\hat{\nu} = 8$. Moreover, assume that $\mathrm{Var}(X_1) = \sigma_1^2 = 5$ and $\mathrm{Var}(X_2) = \sigma_2^2 = 10$. The relevant optimization problem (cf. Section 2.1 and the Appendix) from (25) then yields $\mathrm{Var}(\hat{X}) = \hat{\sigma}^2 = 11.4$. Thus, indeed $\hat{\sigma}^2 > \sigma_i^2$ for $i = 1, 2$.

That finally means, the proposed bias correction method from Section 2.3 does not only reduce the systematic error of the fusion results, but also helps in avoiding an unwanted increase of the final variances. A second advantage is that the parameter calibration in case of FCD is heuristically possible without any further data due to the described "self-evaluation" (cf. Section 2.4). All in all, this makes the presented approach very attractive for offline and online data fusion depending on the amount of available measurement data.

Nevertheless, future research could try to find even better options for calibrating the fusion parameters, namely the bias correction factors $\{\overline{p}_i^*\}_{i=1,\dots,n}$ and the covariances $\{\sigma_{ij}\}_{i,j=1,\dots,n}$ of the data sources. This may include a better definition of periods with similiar traffic conditions as well as optimizing the considered link sets (cf. Section 2.4). Moreover, it might also be useful to go through the extensive literature about portfolio optimization in finance in order to see how the problem of estimating the correlations between the available assets was solved there. Finally, of course, the explicit integration of other data than FCD within the proposed data fusion framework should be part of further studies.

Appendix

As in Section 2.1, assume that there are n assets with expected returns $\underline{\mu} = (\mu_1, \dots, \mu_n)^T$ and covariance matrix $C = (\sigma_{ij})_{i,j=1,\dots,n}$. The question then is how to find the optimal portfolio in terms of minimizing its risk $\hat{\sigma} := \mathrm{Var}(\hat{X})^{\frac{1}{2}}$ for some given expected target return $\hat{\nu}$ where $\hat{X}$ is defined as in (2). This yields the non-linear optimization problem

$$\mathrm{Var}(\hat{X}) = \underline{w}^T C \underline{w} \quad \longrightarrow \quad \min_{\underline{w} \in \mathbb{R}^n}! \tag{25}$$

subject to $\underline{w}^T\underline{1} = 1$ and $\underline{w}^T\underline{\mu} = \hat{\nu}$ where $\underline{w} = (w_1, \dots, w_n)^T$ is the vector of relative shares of all single assets in the considered portfolio.

By introducing the Lagrangian multipliers λ and κ, one obtains the equivalent form

$$\nabla_{\underline{w}}\, h(\underline{w}, \lambda, \kappa) = \underline{0} \tag{26}$$

where

$$h(\underline{w}, \lambda, \kappa) := \underline{w}^T C \underline{w} + \lambda \left(\underline{w}^T\underline{1} - 1\right) + \kappa \left(\underline{w}^T\underline{\mu} - \hat{\nu}\right). \tag{27}$$

Explicitly computing the first derivative in (26) then yields

$$2C\underline{w} = -\lambda\underline{1} - \kappa\underline{\mu} \tag{28}$$

which is the same as

$$\underline{w} = -\frac{\lambda}{2}C^{-1}\underline{1} - \frac{\kappa}{2}C^{-1}\underline{\mu} \tag{29}$$

whenever C is invertible (cf. Section 2.1). Finally, the Lagrangian multipliers λ and κ are determined by the constraints of the optimization via plugging in (29). Thus, consider the linear system of equations

$$1 = \underline{w}^T\underline{1} = -\frac{\lambda}{2}\underline{1}^T C^{-1}\underline{1} - \frac{\kappa}{2}\underline{\mu}^T C^{-1}\underline{1} \tag{30}$$

$$\hat{\nu} = \underline{w}^T \underline{\mu} = -\frac{\lambda}{2}\underline{1}^T C^{-1} \underline{\mu} - \frac{\kappa}{2}\underline{\mu}^T C^{-1} \underline{\mu}, \tag{31}$$

or in short

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} \lambda \\ \kappa \end{pmatrix} = \begin{pmatrix} 1 \\ \hat{\nu} \end{pmatrix} \tag{32}$$

with the constants $a_{11} := -\frac{1}{2}\underline{1}^T C^{-1}\underline{1}$, $a_{12} := -\frac{1}{2}\underline{\mu}^T C^{-1}\underline{1}$, $a_{21} := -\frac{1}{2}\underline{1}^T C^{-1}\underline{\mu} = a_{12}$ and $a_{22} := -\frac{1}{2}\underline{\mu}^T C^{-1}\underline{\mu}$. In this context, remember that $(C^{-1})^T = C^{-1}$ since C is symmetric. Hence,

$$\begin{aligned} \begin{pmatrix} \lambda \\ \kappa \end{pmatrix} &= \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ \hat{\nu} \end{pmatrix} \\ &= \frac{1}{a_{11}a_{22} - a_{12}^2} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{12} & a_{11} \end{pmatrix} \begin{pmatrix} 1 \\ \hat{\nu} \end{pmatrix} \end{aligned} \tag{33}$$

i.e.

$$\lambda = \frac{a_{22} - a_{12}\hat{\nu}}{a_{11}a_{22} - a_{12}^2}, \tag{34}$$

$$\kappa = \frac{a_{11}\hat{\nu} - a_{12}}{a_{11}a_{22} - a_{12}^2}. \tag{35}$$

Moreover, the optimal portfolio in sense of (25) is obtained by plugging in all these numbers into (29).

Acknowledgment

This work was partially supported by the European Commission through its Seventh Framework Programme during the project "SimpleFleet" (http://www.simplefleet.eu, grant agreement No. FP7-ICT-2011-SME-DCL-296423). The authors would like to thank Zelitron and BK Telematics for providing the floating car data as used in this study. Moreover, Dr. István Pál contributed to the implementation of the necessary software modules at an early stage of the project.

References

[1] El Faouzi NE, Leung H, Kurian A. Data fusion in intelligent transportation systems: Progress and challenges – a survey. *Information Fusion* 2011; **12**: 4–10.

[2] Hu J, Kasparias I, Bell MGH. Current state and future outlook of traffic data fusion in london. *Proceedings 15th International IEEE Conference on Intelligent Transportation Systems (ITSC)*; **2012**: 1483–1488.

[3] El Faouzi NE. Data fusion in road traffic engineering: An overview. *Multisensor, multisource information fusion: architectures, algorithms, and applications*; **2004**: 360–371.

[4] Llinas J, Liggings ME, Hall DL. *Handbook of Multisensor Data Fusion: Theory and Practice.* Boca Raton, FL: Taylor & Francis, 2008.

[5] Forschungsgesellschaft für Straßen- und Verkehrswesen (FGSV). *Hinweise zur Datenvervollständigung und Datenaufbereitung in verkehrstechnischen Anwendungen.* Köln: FGSV Verlag, 2003.

[6] Li X, Zhu Y, Wang J, Han C. Optimal linear estimation fusion – Part I. Unified fusion rules. *IEEE Transactions on Information Theory* 2003; **49**: 2192–2208.

[7] Jie L, van Zuylen H, Chunhua L, Shoufeng L. Monitoring travel times in an urban network using video, gps and bluetooth. *Procedia - Social and Behavioral Sciences* 2011; **20**: 630–637.

[8] Malinovskij Y, Wu YJ, Wang Y, Lee UK. Field experiments on bluetooth-based travel time data collection. *TRB 89th Annual Meeting CD-ROM*. Washington, DC, 2010.

[9] Laureshyn A, Ardö H, Svensson Å, Jonsson T. Application of automated video analysis for behavioural studies: concept and experience. *IET Intelligent Transport Systems* 2009; **3**: 345–357.

[10] Nanthawichit C, Nakatsuji T, Suzuki H. Application of probe vehicle data for real-time traffic state estimation and short-term travel time prediction on a freeway. *TRB 82nd Annual Meeting CD-ROM*. Washington, DC, 2003.

[11] Cipriani E, Gori S, and Mannini L. Traffic state estimation based on data fusion techniques. *Proceedings 15th International IEEE Conference on Intelligent Transportation Systems (ITSC)*; **2012**: 1477–1482.

[12] Mehran B, Kuwahara M, Naznin F. Implementing kinematic wave theory to reconstruct vehicle trajectories from fixed and probe sensor data. *Transportation Research Part C: Emerging Technologies* 2012; **20**: 144–163.

[13] Daganzo C. A variational formulation of kinematic waves: basic theory and complex boundary conditions. *Transportation Research Part B: Methodological* 2005; **39**: 187–196.

[14] Daganzo C. A variational formulation of kinematic waves: solution methods. *Transportation Research Part B: Methodological* 2005; **39**: 934–950.

[15] Kong QJ, Li Z, Chen Y, LiuY. An approach to urban traffic state estimation by fusing multi-source information. *IEEE Transactions on Intelligent Transportation Systems* 2009; **10**: 499–511.

[16] Choi K, Chung Y. A data fusion algorithm for estimating link travel time. *Journal of Intelligent Transportation Systems* 2002; **7**: 235–260.

[17] El Faouzi NE. Combining predictive schemes in short-term traffic forecasting. *Proceedings of the 14th International Symposium on Transportation and Traffic Theory (ISTTT)*; **1999**: 471–487.

[18] Yuan Y, van Lint J, Wilson R, van Wageningen-Kessels F, Hoogendoorn S. Real-time lagrangian traffic state estimator for freeways. *IEEE Transactions on Intelligent Transportation Systems* 2012; **13**: 59–70.

[19] Kolosz B, Grant-Muller S, Djemame K. Modelling uncertainty in the sustainability of Intelligent Transport Systems for highways using probabilistic data fusion. *Environmental Modelling & Software* 2013; **49**: 78–97.

[20] Markowitz H. Portfolio selection. *The Journal of Finance* 1952; **7**: 77–91.

[21] Zhou Y, Leung H. Linearly constrained least squares approach for multisensor data fusion. *Sensor Fusion: Architectures, Algorithms, and Applications*; **1997**: 118–129.

[22] Xia Y, Kamel MS. Cooperative learning algorithms for data fusion using novel l_1 estimation. *IEEE Transactions on Signal Processing* 2008; **56**: 1083–1095.

[23] Sun SL, Deng ZL. Multi-sensor optimal information fusion kalman filter. *Automatica* 2004; **40**: 1017–1023.

[24] Xiao L, Boyd S, Lall S. A scheme for robust distributed sensor fusion based on average consensus. *Proceedings of the 4th International Symposium on Information Processing in Sensor Networks*. 2005.

[25] El Faouzi NE. Data-driven aggregative schemes for multisource estimation fusion: a road travel time application. *Multisensor, multisource information fusion: architectures, algorithms, and applications*; **2004**: 351–359.

[26] Brockwell SE, Gordon IR. A comparison of statistical methods for meta-analysis. *Statistics in Medicine* 2001; **20**: 825–840.

[27] Luenberger DG, *Investment science*. New York: Oxford Univ. Press, 1998.

[28] Kuhns G, Ebendt R, Wagner P, Sohr A, Brockfeld E. Self evaluation of floating car data based on travel times from actual vehicle trajectories. *Proceedings IEEE Forum on Integrated and Sustainable Transportation Systems (FISTS)*; **2011**: 115–120.

[29] Neumann T. Accuracy of distance-based travel time decomposition in probe vehicle systems. *Journal of Advanced Transportation* 2014; **48**: 1087–1106.

[30] Kuhns G, Ebendt R, Brockfeld E, Niebel W. Automated bias-correction for accurate fcd processing systems. *Telematics - Support for Transport* 2014; **471**: 441–449.

[31] SimpleFleet. http://www.simplefleet.eu/ (last access: 20/10/2014).