Computer-Aided Civil and Infrastructure Engineering **25** (2010) 596–612

A Robust and Efficient Method for Fusing Heterogeneous Data from Traffic Sensors on Freeways

J.W.C. van Lint* & Serge P. Hoogendoorn

Department of Transport & Planning, Faculty of Civil Engineering and Geosciences, Delft University of Technology, The Netherlands

Abstract: *Fusing freeway traffic data such as spot speeds and travel times from a variety of traffic sensors (loops, cameras, automated vehicle identification systems) into a coherent, consistent, and reliable picture of the prevailing traffic conditions (e.g., speeds, flows) is a critical task in any off- or online traffic management or data archival system. This task is challenging as such data differ in terms of spatial and temporal resolution, accuracy, reliability, and most importantly in terms of spatiotemporal semantics. In this article, we propose a data fusion algorithm (the extended generalized Treiber-Helbing filter [the EGTF]) which, although heuristic in nature, uses basic notions from traffic flow theory and is generic in the sense that it does not impose any restrictions on the way the data are structured in a temporal or spatial way. This implies that the data can stem from any data source, given they provide a means to distinguish between free flowing and congested traffic. On the basis of (ground truth and sensor) data from a microsimulation tool, we demonstrate that the EGTF method results in accurate reconstructed traffic conditions and is robust to increasing degrees of data corruption. Further research should focus on validating the approach on real data. The method can be straightforwardly implemented in any traffic data archiving system or application which requires consistent and coherent traffic data from traffic sensors as inputs.*

1 INTRODUCTION

Traffic data collection and archiving systems are essential tools for both online (and real-time) intelligent transportation systems (ITS) applications, such as travel time prediction (Dharia and Adeli, 2003; Van Lint et al., 2005), incident detection (Karim and Adeli, 2003), dynamic speed limits, and ramp metering (Hegyi et al., 2005; Wang et al., 2006), as well as offline applications, such as policy/measure evaluation studies and the development, calibration, and validation of the tools necessary to perform these tasks (e.g., traffic simulation models). Furthermore, the advance of scientific research relies heavily on the availability of large amounts of detailed and reliable empirical data. In the last decades, the amount of empirical data becoming available for both online and offline use has increased steeply, particularly in terms of the wide range of sensor technologies applied to obtain these data, ranging from inductive loop detectors, radar, microwave, and ultrasonic sensors to infrared cameras and in-vehicle GPS/GSM receivers/transmitters ("floating car data"), to name a few.

However, data from a multitude of different traffic sensors do not necessarily mount up to consistent, coherent, and meaningful information. Data from such different sensors are typically characterized by different formats, semantics, temporal and spatial resolution, and accuracy, and also differ in availability and reliability both as a function of location, time, and circumstances. Both from technical and methodological points of view, the integration of such heterogeneous data into comprehensive and consistent information on the state of traffic in a network (e.g., in terms of speed, flow, or density) is a complex and challenging task. This article focuses on the methodological challenge, that is, on the mathematical and statistical tools for fusing traffic data from multiple sensors.

Fusion of data from a multitude of sources (Dailey et al., 1996; Varshney, 1997) generally leads to the following:

*To whom correspondence should be addressed. E-mail: *j.w.c.vanlint@tudelft.nl*.

DOI: 10.1111/j.1467-8667.2009.00617.x

- increased confidence and accuracy and reduced ambiguity;
- increased robustness, as one sensor can collect information where others are unavailable, inoperative, or ineffective;
- enhanced spatial and temporal coverage, for instance because one sensor can work when or where another sensor cannot;
- and (more tentatively) decreased costs because (a) a suite of "average" sensors can achieve the same level of performance as a single, highly-reliable sensor at a significantly lower cost, and (b) fewer sensors may be required to obtain a (for a particular application) *sufficient* picture of the system state.

With these arguments in mind, data fusion techniques provide an obvious solution for traffic state estimation. Most approaches to traffic state estimation (e.g., Wang et al., 2006; Van Lint et al., 2008), however, consider only a *single* data source (e.g., 1-minute aggregate flows or averaged speeds from local detectors), which is due to the constraints of the solution approach chosen (this will be elaborated in Section 2). Of the (few) studies which *do* consider data from various traffic sensors (e.g., Dailey et al., 1996; Kikuchi et al., 2000), most are concerned with limited traffic networks or corridors (e.g., single traffic links). Recent contributions (Herrera and Bayen, 2008; Van Lint and Hoogendoorn, 2007), however, demonstrate that with well-known techniques such as first-order traffic flow models and extended Kalman filters, fusing data from local detectors and probe vehicles is possible and both technically and economically feasible, given the data can be appropriately aligned. In this article, we will propose an alternative and much simpler solution for traffic data fusion, which extends and generalizes the spatiotemporal traffic data filter first proposed in Treiber and Helbing (2002). Our contribution, the so-called extended generalized Treiber-Helbing filter (EGTF), is generic in the sense that it does not impose any restrictions on the way the data are structured in a temporal or spatial way, given they provide a means to distinguish between free flowing and congested traffic.

The article is structured as follows. In Section 2, we will first briefly introduce some basic notions in multi-sensor data and information fusion. Next, we will recall some of the (spatiotemporal) characteristics of traffic data and discuss why common dynamic state estimation techniques (such as Kalman filters) cannot be easily applied for fusing semantically different traffic data. In Section 3, the alternative method proposed in this article is outlined. Section 4 then describes the experimental setup with which we will provide evidence for the validity and robustness of the method. Note that to this end, *synthetic data* from a micro-simulation model are used, as validation of the EGTF method requires both sensor data as well as the underlying ground-truth data (high-resolution trajectories of all vehicles on a large freeway stretch during a significant time period). The latter was not available for actual traffic in this study. Section 5 then presents the results, and in the final section (6), conclusions are drawn and recommendations for further research into (and application of) the proposed EGTF method are discussed.

2 FRAMEWORK FOR TRAFFIC DATA FUSION

2.1 Multi-sensor data fusion—a brief overview

Multi-sensor fusion, which can be defined as the "synergistic combination of data or information from multiple sensors" (Luo et al., 2002), has been studied and applied extensively in many fields of science and engineering, such as robotics, medical diagnosis, image processing, air traffic control, remote sensing, and ocean surveillance (see e.g., Yager, 2004; Xiong and Svensson, 2002; Varshney, 1997; Piella, 2003; Linn and Hall, 1991). There are many ways to classify fusion techniques. Most differentiate between (hierarchical) levels of fusion, for example, between fusion of raw sensor *data* (Level 1); fusion of *features* extracted from the raw data (Level 2); and fusion of *objects or events* deduced from these features (Level 3) (Varshney, 1997; Luo et al., 2002). Examples for these three fusion levels in traffic science and practice include, for instance, speed/density estimation from raw sensor data (Level 1), determining queue locations and lengths from these estimated speeds (Level 2), and incident detection on the basis of the previous two (Level 3, see e.g., Ivan and Sethi (1998)). Another way of categorizing is by means of the (families) of solution techniques which have been proposed for multi-sensor fusion. For example, Luo et al. (2002) distinguish between the following:

Estimation methods including weighted averaging, least squares estimation, and recursive approaches such as Kalman or particle filtering.
Classification methods such as clustering algorithms and unsupervised techniques such as LVQ and self-organizing maps.
Inference methods such as graphical (Bayesian belief) networks, Dempster-Shafer methods.
Artificial intelligence methods including expert systems, fuzzy inference engines, and adaptive neural networks, to name a few.

Clearly, there is a strong relation between these categorizations. Estimation methods such as Kalman

filtering are especially suitable for Level 1 fusion, that is, fusion of noisy (raw) sensor data; classification methods are typically tailored for feature extraction; whereas higher level inference systems combine such features to derive evidence for the occurrence of particular events and objects or to support decision making on the basis of fused data and/or features. In a typical multi-sensor fusion application, different combinations of techniques are employed to go from raw data to for example decisions on the basis of inferred events. In this article, we will focus on the first fusion category, that is, Level 1 fusion of raw traffic sensor data (speeds, vehicle speeds and positions, travel times), using (state) estimation and spatiotemporal averaging methods. To this end, we will first briefly recall some basic characteristics of raw traffic data and discuss the limitations of typical state estimation techniques used in traffic applications with respect to these different types of data.

2.2 Characteristics of traffic data and consequences for data fusion

Data from traffic sensors come in many forms and qualities, but can essentially be subdivided along two dimensions. The first relates to their spatiotemporal semantics, that is, do the data represent *local* traffic quantities (flow, time headway(s), etc.) or do the data reflect quantities over space (journey speed, space mean speed, travel time, and trajectories). The second relates to the degree of aggregation, where data may represent an aggregate or average over fixed time periods (e.g., 1-minute aggregate flows or averaged speeds), or a single event (vehicle passage, travel time, full trajectory). Table 1 overviews this classification with a few examples. In this section, we briefly discuss some characteristics of local and spatial data and (in the final subsection) which constraints these pose on data fusion and state estimation techniques.

In the next two sections, we will briefly discuss some properties of local and spatial traffic data relevant to the data fusion algorithm proposed in this article.

2.2.1 Local traffic data. The basic macroscopic traffic variables used to describe a traffic state include flow q (veh/h), (space) mean speed u (km/h), and vehicular density ρ (veh/km), which are related to one another by $q = \rho u$. From its definition, vehicular flow *can* be measured with local traffic sensors, whereas density and space mean speed cannot (see Table 1). However, with local detection equipment, the space mean speed can be approximated by the local *harmonic* mean speed, that is,

$$u_S \approx u_H = \left(\frac{1}{N} \sum_i \frac{1}{v_i} \right)^{-1}$$

The approximation is exact under the assumption of homogeneity and stationarity on the road section and time period of interest. The local arithmetic (or simply *time*) mean speed $u_T = \frac{1}{N} \sum v_i$ provides a biased approximation of the space mean, due to the fact that in a time sample, *faster* observations are over represented. Especially in congestion (where speeds are typically below 50 km/h), this bias is significant. Time averaging under these congested conditions may lead to overestimates of speed of 20% and more. A typical example of a traffic monitoring system in which these biased time mean speeds are collected is the MoniCa system, operational on the motorway network in the Netherlands (around 6,600 km). MoniCa dual induction loops are installed at every 500 m for around 1/3 of the network, whereas another 1/3 has only limited traffic monitoring, and roughly 1/3 has nothing at all. Besides biased speeds and the limited spatial coverage in some areas, a third major issue within MoniCa is that of the available data (aggregate flows and time mean speeds) on average 5–10% is missing or otherwise deemed unreliable, with regular extremes to over 25 or 30%.

There are also other ways to estimate density and speed from local detector equipment using (local) sensor occupancy *Occ*, which is the ratio of time a sensor is occupied by vehicles in a certain time period. However, also then, bias due to time averaging is a major concern (see e.g., Papageorgiou and Vigos, 2008; Hellinga, 2002).

Table 1
Classification of data from traffic sensors with some examples

	Event-based	*Aggregate*
Local	Vehicle passage, speed, length	Traffic flow, time and harmonic mean speed
Spatial	Vehicle travel, time, journey speed, trajectory samples	Space mean speed, mean travel time, mean journey speed

2.2.2 Travel time and trajectory data. Travel time can be measured by means of, for example, automated vehicle identification (AVI) systems, which identify vehicles at two consecutive locations A and B at time instants t_A and t_B and deduce the *realized* travel time with $\overline{TT}_t^{AB} = t_B - t_A$ (Figure 1). AVI systems may employ cameras and license plate detection technology, or may match vehicles through induction footprints, tolling tags, or otherwise. Methodologically, the most important characteristics of travel time are the following.

1. Travel time can only be measured for realized trips, that is, after a vehicle has finished it. Realized travel time $\overline{TT}_t$ of a vehicle arriving at time t hence represents something entirely different

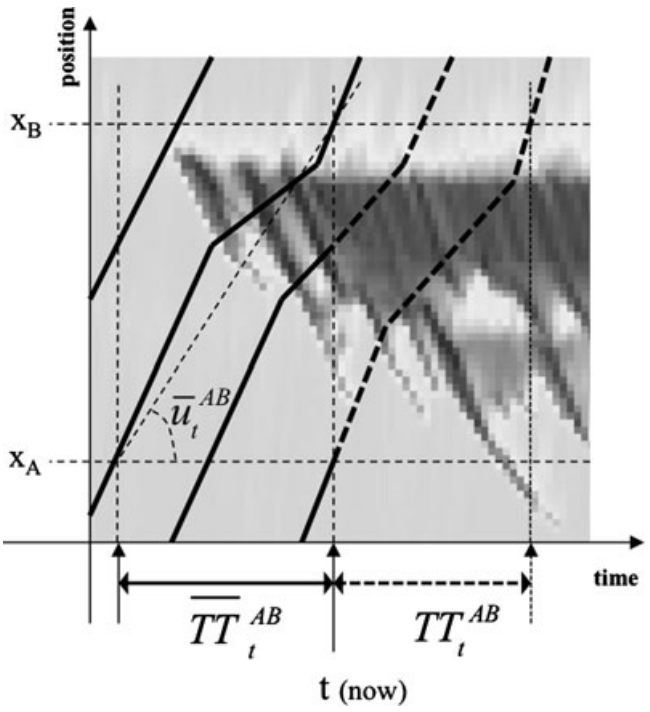

Fig. 1. Relationship between vehicle trajectories (the thick solid and dotted lines), realized and actual travel time ($\overline{TT}_t$ and TT_t), average journey speed ($\overline{u}_t$), and the underlying speeds (represented by a speed contour plot, where dark areas represent low speeds).

than the so-called *actual* travel time TT_t of a vehicle departing at t, which must be *predicted* by definition (Figure 1).

2. Realized travel time (or its reciprocal average realized journey speed $\overline{u}_t = L/\overline{TT}_t$) is a *scalar* representation of the traffic conditions (e.g., the speed $v(s, x), s \leq t$) a vehicle encountered during its trip over a route with length L. There are following two immediate consequences.
   a. The relationship between travel time and the underlying traffic conditions is underdetermined, that is, it *is* possible to estimate travel time from local speeds (e.g., Ni and Wang, 2008; Van Lint and Van der Zijpp, 2003), but conversely, it is not possible to estimate local speeds from travel time, unless other sources of information are available. Underdeterminedness is a common problem in traffic state estimation, for example, in estimating origin-destination flows, where additional information is also needed to solve the problem (see e.g., Castillo et al., 2008; Castillo et al., 2008).
   b. A travel time measurement provides information related to a fixed spatial entity (a route or road segment) but over a variable time period, which duration in fact equals the travel time measurement itself.

We will discuss the consequences of these properties of travel time for traffic state estimation in Section 2.3. By sampling data (location and/or speed) from instrumented vehicles (e.g., through GPS or GSM) at consecutive time instants, vehicle trajectories can also be measured. The relationship between vehicle trajectories, travel time, and the underlying speeds is illustrated in Figure 1. Clearly, when *all* vehicle trajectories are sampled at high frequencies, the traffic state (prevailing speeds, flows, densities, travel times, etc.) can be completely deduced from these so-called *floating car data* (FCD). However, at penetration rates far below 100%, FCD at best provide a proxy for average speed on the road segments from which these data are collected.

2.2.3 Spatiotemporal traffic patterns. Before discussing traffic state estimation techniques, let us first illustrate some of the typical spatiotemporal traffic patterns observed in practice on the basis of real traffic data (speeds[1]). Figure 2 illustrates these patterns with space-time contour maps of the speeds (top graph) and flows (bottom graph) along a 20 km 3-lane freeway section in the Netherlands (the A12 between Gouda and Utrecht). Figure 2 shows the following.

1. Under free-flow conditions, perturbations (e.g., changes in average speed) propagate in the same direction of traffic flow with a (positive) speed c_{free}, which appears remarkably constant and is in the same order of magnitude as the average traffic speed.
2. Under congested conditions, a number of different patterns can be observed, which are described and discussed elsewhere (e.g., in Kerner (2004); Helbing and Treiber (2002); Kesting and Treiber (2008)). Regardless of the precise number of different congested patterns and their traffic flow theoretical interpretation (for an overview see e.g., Hoogendoorn and Bovy (2001)), in congested traffic, perturbations propagate in the *opposite* direction of traffic, that is, with negative speeds c_{cong} (on freeways between -15 to -20 km/h (Treiber and Helbing, 2002)). Figure 2 illustrates how some congested patterns may travel over 15 km in the upstream direction.

The main consequence of the spatiotemporal characteristics of traffic patterns discussed above is that data from local sensors cannot be averaged or interpolated orthogonally (which is what often happens in

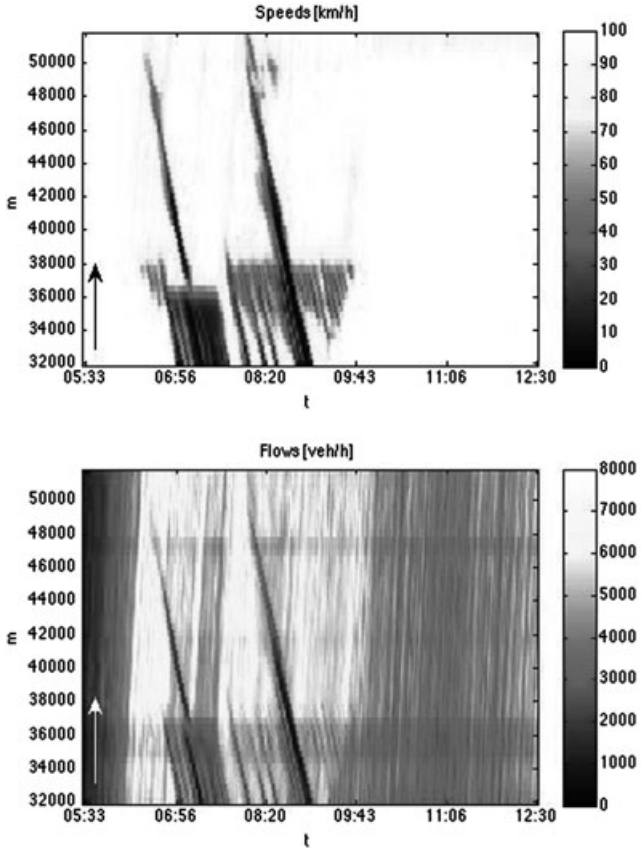

Fig. 2. Speeds (km/h) and flows (veh/h) obtained from inductive loop detectors along a typical densely used 3-lane freeway section in the Netherlands.

practice) as a means to "estimate" the traffic conditions between measurements or consecutive time periods. Instead, these data must be extrapolated over space and time along the characteristic curves mentioned above, which differ in congested and free-flow operations. This is the basis for the spatiotemporal filter introduced by Treiber and Helbing (2002), the extensions to this filter proposed in this article, and also for the model-based traffic state estimators discussed in the next subsection.

2.3 Traffic state estimation techniques

The generic structure of a traffic state estimation/data fusion approach typically includes three components:

1. data from whatever sensors available (see e.g., Table 1);
2. some mathematical model which formalizes the relationship of the data with the underlying dynamic phenomena (e.g., an analytical traffic flow model);
3. data assimilation techniques which combine the former two (data + model).

The most widely applied data assimilation technique applied to traffic state estimation problems is the Kalman Filter (KF) and/or its many variations (extended and unscented KF, particle filters). For example, Wang and Papageorgiou (2005) and Wang et al. (2006) combine a so-called second-order macroscopic traffic flow model with the extended KF to estimate vehicular density and speed from sparse local data from induction loops. Van Lint and co-workers demonstrate a multi-class first-order macroscopic traffic flow model and a dual extended KF for the same purpose (Van Lint et al., 2008). Such KF approaches exploit the fact that analytical traffic flow models can be numerically solved, in which case they can be expressed in state-space form (k depicts discrete time steps of duration Δt):

$$\mathbf{x}_{k+1} = f(\mathbf{x}_k, \mathbf{w}_k, \mathbf{u}_k) + \eta_k \tag{1}$$

$$\mathbf{y}_k = h(\mathbf{x}_k, \mathbf{v}_k) + \zeta_k \tag{2}$$

Equation (1) depicts the process equation, which describes the dynamics of state $\mathbf{x}_{k+1}$ (density and/or speed) as a function of $\mathbf{x}_k$ and external disturbances $\mathbf{u}_k$ (for example traffic demand at network boundaries) plus an error term η_k, reflecting errors in the process model (e.g., model misspecification). The function f contains (possibly time-varying) parameters $\mathbf{w}_k$, for example the kinematic wave speed or road capacity. Equation (2) depicts the observation equation h which relates the system state to (observable) outputs $\mathbf{y}_k$. Also h may contain (possibly time-varying) parameters $\mathbf{v}_k$. The error term ζ_k depicts errors in either the observation model h and/or the observations themselves. The fundamental diagram of traffic flow $q = Q^e(\rho)$, or $u = U^e(\rho)$, relating speed or flow to density, is a good example of such an observation equation.

Details on KF algorithms can be found in many textbooks (e.g., Simon, 2006), let us just make a few remarks on their applicability to traffic state estimation and data fusion. The key advantage of KF-based state estimation approaches is that they provide a convenient and principled approach to recursively correct state estimates by balancing the errors (uncertainties) in the process and observation model and in the data ($\mathbf{y}_k^{\text{obs}}$) with

$$\mathbf{x}_k = \mathbf{G}_k\left(\mathbf{y}_k^{\text{obs}} - \mathbf{y}_k\right) \tag{3}$$

The so-called Kalman gain $\mathbf{G}_k$ in (3) can be informally understood as

$$\mathbf{G}_k = \frac{\text{variance process model}}{\text{variance observation model and data}} \times \text{sensitivity observation model to state}$$

This implies that (1) the more uncertain the data are (conceived), the more weight is put on the model

predictions and vice versa, and (2) that the KF adjusts $\mathbf{x}_k$ proportional to the sensitivity of the observation model to changes in the state variables. For example, under free-flow conditions, the relationship between traffic density and speed is very weak, which would imply only small corrections in state variables ($\mathbf{x}_k$) even if the speeds measured with sensors ($\mathbf{y}_k^{\text{obs}}$) differ largely from those predicted by the observation model ($\mathbf{y}_k$). This intuitive structure can be easily explained to traffic operators and professionals using such state estimation tools. Moreover, the same process and observation model can be subsequently used for prediction and control purposes, given proper predictions of the boundary conditions (traffic demand, turnfractions, and capacity constraints) and estimates of the model parameters are available.

There is, however, one major drawback of KF approaches, which relates to the spatial and temporal alignment of the data. For every data source used, an (additional) observation Equation (2) is required. This is not necessarily a problem, as long as the spatial and temporal discretization of the data (detector locations x_d and time periods of size ΔT) can be aligned with the discretization used in the model (road segments of length Δx and time periods of size Δt). For example, some spatial data can be transformed fairly easily into local measurements, such as sampled floating car data (Herrera and Bayen, 2008; Van Lint and Hoogendoorn, 2007). This, however, is not the case for, for example, travel time or journey (segment) speeds. These are available after trips are realized, that is, after a time period equal to the measured travel time. As a result, a (realized) travel time observation equation has the following general form:

$$\mathbf{y}_k = h(\mathbf{x}_k, \mathbf{x}_{k-1}, \ldots, \mathbf{x}_{k-TT^{\max}}) + \vartheta_t \tag{4}$$

where the output variable $\mathbf{y}_k$ now depicts (realized) travel time, and $TT^{\max}$ is the maximum observed travel time on the route of interest. Using Equation (4) hence necessitates maintaining state variables $\mathbf{x}_k$ (speeds, densities) for as many discrete time instants as required to deduce the longest possible realized travel time. To illustrate that this leads to practical and computational problems, consider a traffic corridor constituted of N segments (of typically 500 m) on which state variables (e.g., densities) are estimated, which are corrected every ΔT time units (of typically 1 minute). To facilitate observation Equation (4), the number of state variables increases to $N^{\text{aug}} = N \times TT^{\max}/\Delta T$. On a congested corridor of say 20 km, $TT^{\max}$ may become as high as 60 minutes or more, yielding an augmented state vector with 1,000 variables or more. As in the KF procedure also an error covariance matrix is maintained of size $(N^{\text{aug}})^2$, this approach would lead to very unwieldy computations for any realistic traffic network or corridor.

We conclude that in the case of combining data which cannot be straightforwardly aligned, and/or which require augmenting the state vector with multiple time periods, it makes sense to use a simpler approach for data fusion, that is, one with far less degrees of freedom. In Ou et al. (2008), such travel times have been successfully combined with loop data, where they were used as a *constraint* on travel times estimated from loop detector data. In the current article, it is our aim to show that travel times can also be *directly* fused with other types of data. In the next section, we will present this approach.

3 PROPOSED METHODOLOGY FOR TRAFFIC DATA FUSION

3.1 The generalized Treiber–Helbing filter (GTF) for a single data source

Let us first consider a single data source j. For filtering these data over space and time, we use the approach of Treiber and Helbing (2002) stating that the quantity z at (t, x) can be reconstructed as follows:

$$z(t,x) = w(t,x) z_{\text{cong}}(t,x) + (1 - w(t,x)) z_{\text{free}}(t,x) \tag{5}$$

In Equation (5), the reconstruction involves a convex (weighted) combination of two reconstructions of the signal. The first ($z_{\text{cong}}(t,x)$) assumes congested traffic operations and the second ($z_{\text{free}}(t,x)$) free-flow conditions. Further below, we will discuss the weighting mechanism in detail. Let us first note that in both free flow and congestion, it is assumed that information (measured with traffic sensors) travels along the characteristics with constant characteristic speeds c_{cong} or c_{free}, respectively. The underlying assumption is that traffic can be represented sufficiently well by a bilinear fundamental diagram, which according to Newell (1993a, 1993b) is "a not too unreasonable postulate." This bilinear fundamental diagram reads

$$q = Q^e(\rho) = \begin{cases} c_{\text{free}}\rho & \rho \le \rho_c \\ q_c + c_{\text{cong}}\rho & \text{otherwise} \end{cases} \tag{6}$$

According to the kinematic wave theory (Richards, 1956; Lighthill and Whitham, 1955), the assumption that the characteristic speeds $c(\rho)$ are indeed constant for all traffic states (ρ) yields that the resulting trajectories of traffic perturbations are in fact straight lines (moving downstream under free-flow conditions and upstream under congested conditions).

Proof: Consider the conservation of vehicles equation

$$\frac{\partial \rho}{\partial t} + \frac{\partial q}{\partial x} = 0. \tag{7}$$

Substituting (6) in (7) yields

$$\frac{\partial \rho}{\partial t} + c(\rho)\frac{\partial \rho}{\partial x} = 0 \tag{8}$$

with

$$c(\rho) = \frac{\partial}{\partial \rho} Q^e(\rho) = \begin{cases} c_{\text{free}} & \rho \leq \rho_c \\ c_{\text{cong}} & \text{otherwise.} \end{cases} \qquad \square$$

Now consider that we have traffic sensors at locations (t_i, x_i), providing data z_i (e.g., speeds, flows, etc.). Whereas Treiber and Helbing (2002) assume that these data points stem from sources that provide data in an equidistant rectangular grid (e.g., local detectors producing averaged speeds over fixed time intervals), in this article, we will relax this constraint and assume that the data may stem from any point (t_i, x_i). This implies that we also consider event-based data sources, such as trajectory data from moving vehicles (FCD - floating car data), or for example realized travel times from AVI systems. How these are processed by the generalized Treiber-Helbing filter (GTF) will be explained further below. Let us first define similar to Treiber and Helbing (2002) the following anisotropic filter kernel:

$$\begin{aligned} \phi_{\text{cong}}(t, x) &\equiv \phi_0\left(t - \frac{x}{c_{\text{cong}}}, x\right), \text{ and} \\ \phi_{\text{free}}(t, x) &\equiv \phi_0\left(t - \frac{x}{c_{\text{free}}}, x\right) \end{aligned} \tag{9}$$

with

$$\phi_0(t, x) = \exp\left(-\frac{|x|}{\sigma} - \frac{|t|}{\tau}\right) \tag{10}$$

In (10), σ and τ are parameters of the filter kernel (9) which describe the width and time window size of the "influence" region around the points (t, x) where the traffic quantity z is estimated. The larger the distance $(x_i - x, t_i - t)$ between (t, x) and some data points (t_i, x_i), the smaller the weight of the data z_i in the estimation of $z(t, x)$. With (9) the weight of a single data point z_i in the estimation of $z(t, x)$ equals

$$\begin{aligned} \beta^i_{\text{cong}}(t, x) &= \phi_{\text{cong}}(x_i - x, t_i - t), \text{ and} \\ \beta^i_{\text{free}}(t, x) &= \phi_{\text{free}}(x_i - x, t_i - t) \end{aligned} \tag{11}$$

Loosely speaking, the weight is determined by the distance between the points (t, x) where we want to estimate $z(t, x)$ and the data points (t_i, x_i), considering the speed at which information moves through the flow under free-flow or congested conditions. As the chosen kernel in (10) decays exponentially for large values of $x_i - x$ and $t_i - t$, it is favorable to limit the distance over space and time for which data points are still used in the estimation of $z(t, x)$. Define $A(t, x)$ as the area around points (t, x) from which data are used to estimate $z(t, x)$, and $N_A(t, x)$ as the number of data points $z_i \in A(t, x)$. The estimation for the two condition-dependent estimates $z_{\text{cong}}(t, x)$ and $z_{\text{free}}(t, x)$ from Equation (5) then becomes

$$\begin{aligned} z_{\text{cong}}(t, x) &= \frac{\sum_{i=1}^{N_A(t,x)} \beta^i_{\text{cong}}(t, x) z_i}{\sum_{i=1}^{N_A(t,x)} \beta^i_{\text{cong}}(t, x)}, \text{ and} \\ z_{\text{free}}(t, x) &= \frac{\sum_{i=1}^{N_A(t,x)} \beta^i_{\text{free}}(t, x) z_i}{\sum_{i=1}^{N_A(t,x)} \beta^i_{\text{free}}(t, x)} \end{aligned} \tag{12}$$

Before using these equations for fusing data from different data sources, we will first discuss two factors which are important in the estimation of $z(t, x)$: the area $A(t, x)$, which governs the amount of data used for each estimation point, and the weight factor $w(t, x)$.

3.2 Filter design considerations and computational efficiency

In contrast to KF-based approaches, the complexity of the GTF scales linearly with the size of the vector of quantities (speeds, flows, densities) which needs to be estimated. There are two factors which determine how much computations are needed for the estimation. The coarser the estimation grid is chosen (i.e., the larger Δx and Δt), the fewer estimation points $z(t, x)$ need to be calculated. The second factor affecting computational complexity is the area $A(t, x)$, which may be specific for each estimation point (t, x) in terms of both shape and size. The larger $A(t, x)$ is chosen, the more data points are considered in the estimation of $z(t, x)$ and the smoother (and possibly more accurate) the estimation becomes. On the other hand, the larger $A(t, x)$ becomes (in the limit $A(t, x) = \mathcal{A}$ (the entire space time grid considered)), the more computationally demanding the algorithm (Equation (12)) will be. As a pragmatic choice, we will consider a circular area around each estimation point with a radius $r_A(t, x)$ defined by

$$r_A(t, x) = \sqrt{(\Delta T_{\max})^2 + (\Delta L_{\max})^2} \tag{13}$$

with

$$N_A(t,x) \geq 2$$

where $\Delta T_{\max}$ and $\Delta L_{\max}$ are constants determining the farthest distance over time and space for which data points are still considered in the estimation of $z(t,x)$. $A(t,x)$ necessarily needs to be large enough such that at every (t,x) at least two data points z_i are near enough to be used in the estimation of $z(t,x)$. In many practical cases, for example, with $\Delta T_{\max}$ and $\Delta L_{\max}$ set to constants of an order larger than σ and τ (a few km and 10–15 minutes, respectively), $N_A(t,x)$ will be larger than 2 (in the order of 10 or more).

A second important filter design choice is the weight factor $w(t,x)$ used in (5). This factor describes to which degree the conditions in points (t,x) are dictated by free-flow traffic operations ($w(t,x)=0$); by congested traffic operations ($w(t,x)=1$); or a combination of both ($0<w(t,x)<1$). Treiber and Helbing (2002) propose to use speed data u_i for this purpose and use the following expression to calculate this weight factor:

$$w(t,x) = \frac{1}{2}\left[1+\tanh\left(\frac{V_c - u^*(t,x)}{\Delta V}\right)\right] \qquad (14)$$

with

$$u^*(t,x) = \min(u_{\text{cong}}(t,x), u_{\text{free}}(t,x)) \qquad (15)$$

In (15), $u_{\text{cong}}(t,x)$ and $u_{\text{free}}(t,x)$ are calculated with (12), so that $u^*(t,x)$ effectively represents a lowerbound of the speed $u(t,x)$ given the available data points u_i. Note that the hyperbolic tangent function in (14) is an arbitrary filter design choice. Any crisp, smooth, or even fuzzy function which is able to discriminate between free-flowing and congested traffic operations based on whatever data are available (density/occupancy, speed) would in principle do.

3.3 The EGTF for multiple data sources

The GTF formulated above is able to reconstruct traffic conditions on the basis of data stemming from a particular data source j which provides data from sensors at points $(t_i^{(j)}, x_i^{(j)})$. Let $z^{(j)}(t,x)$ denote the considered traffic value as reconstructed from data source j with (in that order) Equations (15), (14), (13), (12), and (5). To fuse data from multiple data sources into one estimate $z(t,x)$, we propose the following linear combination:

$$z(t,x) = \frac{\sum_j \alpha^{(j)}(t,x) \sum_{i=1}^{N_A^{(j)}(t,x)} \phi_i^{(j)}(t,x) z^{(j)}(t,x)}{\sum_j \alpha^{(j)}(t,x) \sum_{i=1}^{N_A^{(j)}(t,x)} \phi_i^{(j)}(t,x)} \qquad (16)$$

3.3.1 Data source specific weights. In Equation (16), two weight factors ($\alpha^{(j)}(t,x)$ and $\phi_i^{(j)}(t,x)$) are used. The first (dynamic) weight factor $\alpha^{(j)}(t,x)$ in (16) is data source specific and can be interpreted as a dynamic indicator of the *reliability* of the data from source j at (t,x). It could for example be determined on the basis of *a priori* estimates of the measurement accuracy of data source j. We make two assumptions in setting $\alpha^{(j)}(t,x)$.

1. We assume $\alpha^{(j)}(t,x) \propto 1/\Theta^{(j)}(t,x)$, where $(\Theta^{(j)})^2$ depicts the error variance of data measured with data source j.
2. We assume that $\Theta^{(j)}(t,x)$ is constant in free-flow operations and constant under congested conditions, where the latter is smaller than the former. Specifically, let $\Theta_0^{(j)}$ represent the standard deviation of the measurement error of data source j under congestion ($u(t,x) \ll V_c$) and $[1+\mu^{(j)}]\Theta_0^{(j)}$, with $\mu^{(j)} \geq 0$, the standard deviation under free-flowing conditions ($u(t,x) \gg V_c$).

We can then reuse the datasource specific weight $w^{(j)}(t,x)$ calculated with (14) to compute $\alpha^{(j)}(t,x)$ as follows

$$\alpha^{(j)}(t,x) = \frac{1}{\Theta_0^{(j)}[1+\mu^{(j)}(1-w^{(j)}(t,x))]} \qquad (17)$$

The second weight factor in (16), $\phi_i^{(j)}(t,x)$, is data source and data point specific and reflects the fact that data from different sources will result in a different spatiotemporal distinction between congested and free-flow conditions. This is due to, for example, the fact that one data source will have more data points available around a certain estimation point (t,x) than other data sources. Using Equations (11) and (14) we propose

$$\phi_i^{(j)}(t,x) = w^{(j)}(t,x)\cdot \beta_{\text{cong}}^{i,(j)}(t,x) + (1-w^{(j)}(t,x))\cdot \beta_{\text{free}}^{i,(j)}(t,x) \qquad (18)$$

We will return to these data source-specific weights in Section 4.2. In the next subsection, we will discuss how to use the data fusion filter for travel time data.

3.3.2 Fusing realized travel times with local traffic data. Let us now consider realized travel times $\overline{TT}_{\text{AB}}(t)$ measured with AVI systems at consecutive locations x_A and x_B. Such a realized travel time then provides (ex post) a proxy for the average speed in a space-time area $\{(t-\overline{TT}_{\text{AB}}(t), x_A), (t, x_B)\}$ which equals

$$\overline{v}_{\text{AB}}(s) = \frac{L_{\text{AB}}}{\overline{TT}_{\text{AB}}(t)}, \text{ with } t-\overline{TT}_{\text{AB}}(t) \leq s \leq t \qquad (19)$$

Clearly, a new estimate $\overline{v}_S(s)$ becomes available with every vehicle finishing its trip over AB. Now consider discrete time periods $k=(t_{k-1}, t_k]$ of typically 1, 3, or 5

minutes. The average travel time of vehicles *departing* during such periods equals $TT_{\text{AB}}(k) = \sum TT_{\text{AB}}(t), t \in (t_{k-1}, t_k]$. Let us denote the average travel time of vehicles *arriving* in k by $\overline{TT}_{\text{AB}}(k)$. Note that $\overline{TT}_{\text{AB}}(k) \neq TT_{\text{AB}}(k - \overline{TT}_{\text{AB}}(k))$ as vehicles arriving in k may have departed at time instants t spread over several discrete departure time periods. Nonetheless, nothing prevents one from using $\overline{TT}_{\text{AB}}(k)$ to approximate the space mean speed over AB *for all vehicles arriving in period k*, which analogously to (19) reads

$$\bar{u}_{\text{AB}}(k) = \frac{L_{\text{AB}}}{\overline{TT}_{\text{AB}}(k)} \tag{20}$$

On the basis of (20) we can translate the travel time measurement into a series of speed measurements at discrete time points $\{t_k - \overline{TT}_{\text{AB}}(k), t_k - \overline{TT}_{\text{AB}}(k) + 1, \ldots, t_k\}$ along the line

$$\Upsilon = \{(t, x) \mid t \in (t_{k-\overline{TT}_{\text{AB}}(k)}, t_k] \text{ and } x = x_A + \bar{u}_{\text{AB}}(k) \cdot (t_k - t)\} \tag{21}$$

The line Υ in Equation (21) describes an average vehicle trajectory between AVI stations with a slope equal to $\bar{u}_{\text{AB}}(k)$, such that we can treat consecutive travel time measurements as FCD samples, on the basis of which we can use the same formula for the dynamic data source-specific weight $\alpha^{(j)}(t, x)$ as in Equation (17). Figure 3 illustrates this process schematically and shows that the trajectory described by Υ may lead to both under- and overestimated speeds on different locations and time instants. $\Theta_0^{(j)}$ should hence be chosen proportional to the distance between the consecutive AVI stations, reflecting the fact that the reconstructed vehicle trajectory with (20) and (21) becomes a coarser (and possibly more biased) approximation of the real average vehicle trajectory with a larger AVI station spacing.

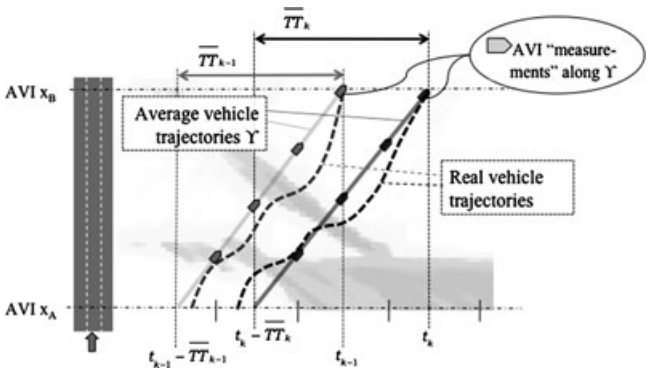

Fig. 3. Schematic outline of the "conversion" of realized travel times $\overline{TT}$ into discrete FCD speed measurements.

4 EXPERIMENTAL SETUP

To test the feasibility of the proposed EGTF method and analyze its properties, we will use data generated by a traffic micro-simulation tool FOSIM[2]. One reason for using micro-simulation data is that this provides ground-truth data (space mean speeds in this case) over a long freeway section, which are not readily available in practice. Secondly, it enables us to freely emulate any possible sensor configuration under many different scenarios (e.g., spacing, reliability), and finally, although drivers in micro-simulation tools behave differently than real-world drivers, the problem to be solved is essentially still the same, that is, the reconstruction of complex traffic patterns from heterogeneous and unreliable sensor data.

4.1 Freeway corridor and basic sensor configuration

To generate both ground-truth data (u_S) as well as all detector data, a FOSIM implementation of the 19 km A13 southbound freeway section between Delft and Rotterdam is used as shown in Figure 4. This corridor contains six on-ramps, five off-ramps, and three weaving sections (combined on/off-ramps). The underlying origin-destination (OD) matrix was estimated on the basis of loop detector data collected on a weekday afternoon peak in 2006 and tuned rudimentary such that a realistic traffic pattern resulted, with congestion occurring at the two bottlenecks where also in reality congestion sets in first.

To make a comparison between different scenarios possible, a single simulation was run, from which all data (ground-truth and sensor data) were derived. As the EGTF effectively removes the high frequency fluctuations (smaller than τ (s) and σ) (m) from the data, the reconstructed traffic conditions on the basis of the various data sources are compared against *filtered* ground-truth data (with the same filter settings as will be used for the data fusion itself—see below), such that the differences in performance solely reflects error variance due to using different data sources. In our case, the ground-truth data depict the space mean speeds $u_S(k, m)$ over equidistant space time regions (k, m) of size 30 $(s) \times 100$ (m). Figure 5 shows a speed contour map of both raw (top) and filtered (bottom) ground-truth data. The drive direction is from bottom to top.

In the ensuing the sensor data will be filtered and fused in the same space-time discretization as the ground-truth data, that is a grid of 30 $(s) \times 100$ (m). Three data sources (sensor types) are considered:

1. FCD, where a certain percentage P of all vehicles communicate their current location

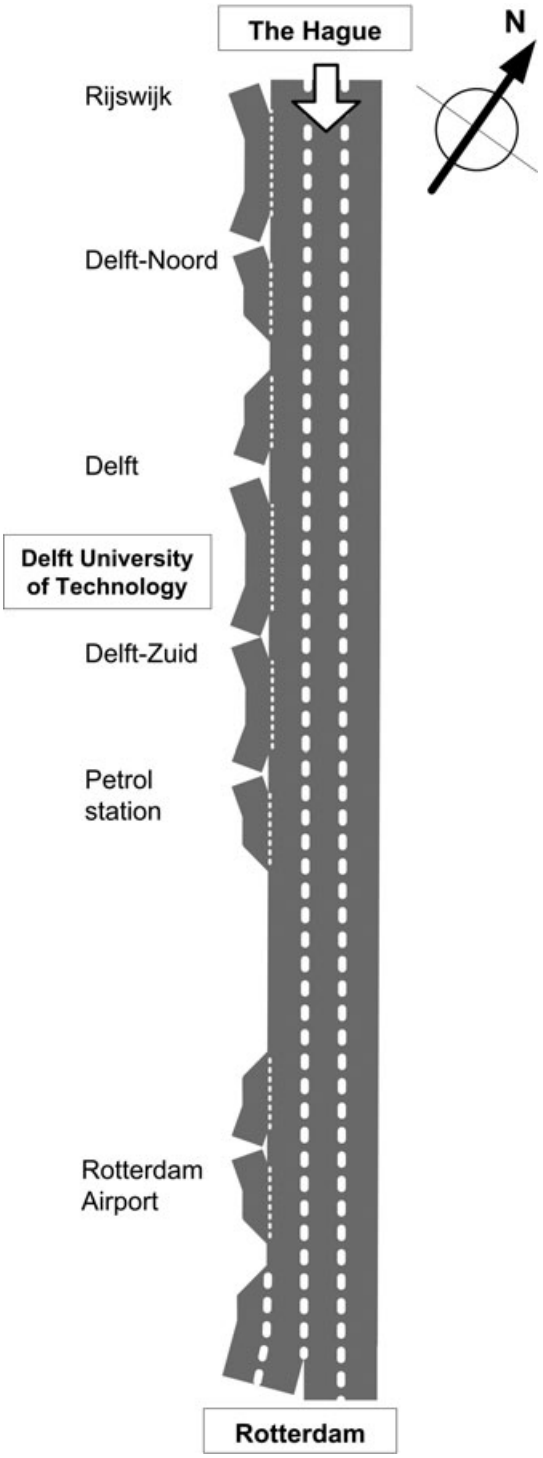

Fig. 4. Schematic drawing of the 19 km A13 freeway corridor from The Hague to Rotterdam.

and speed at fixed time intervals of τ_{fcd} seconds. The first data transmission for each probe vehicle is assumed to occur at the moment of entering the freeway section.

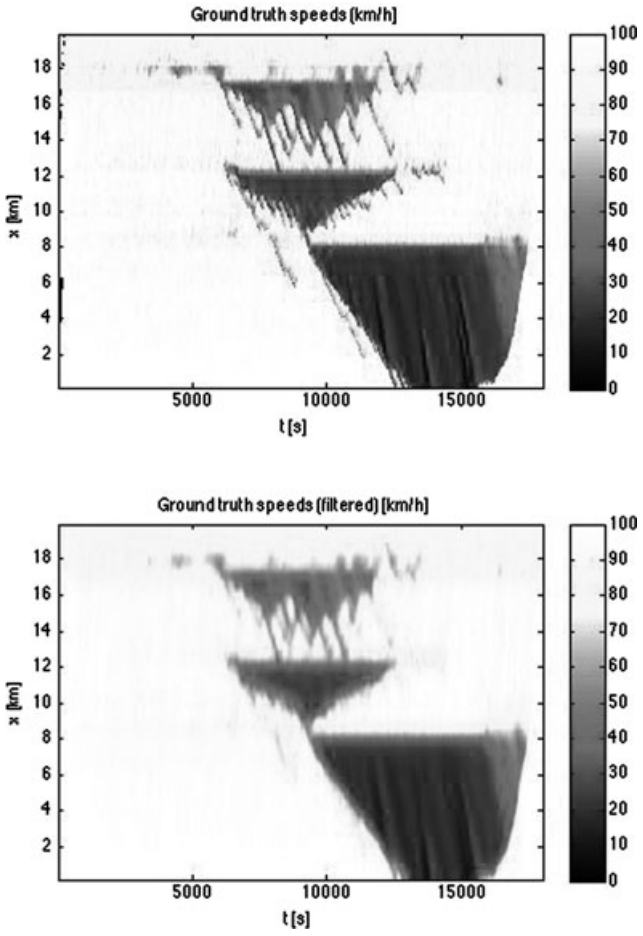

Fig. 5. Space-time contour map of raw (top) and filtered (bottom) ground-truth speeds. The drive direction is from bottom to top.

2. One-minute time mean speeds u_T and harmonic mean speeds u_H from induction loops (or any other type of local detector), spaced evenly at a distance of Δ_{loop} meters, for which we will use multiples of 500 meters (500, 1000, 1500 meters).
3. Average realized travel times $\overline{TT}(k)$ from AVI stations spaced at a distance of Δ_{avi} meters for which we will use multiples of 1,500 m (1,500 and 3,000 m).

4.2 Filter parameter settings and weights

Applying the EGTF to fuse data from different sources requires setting a number of parameters. To deduce the characteristic speeds (c_{cong} and c_{free}), an optimization method was used (the standard line search algorithm implemented in fmincon.m, Matlab 7.5) in which the data fusion algorithm was run with ground-truth speeds as inputs (so $z(t, x) = u_S(t, x)$), such as to minimize the difference between the average filtered and

Table 2
Data source weight settings

Data source	$\Theta_0^{(j)*}$	$\mu^{(j)}$
u_T^{loop}	4	2
u_H^{loop}	3	1.5
v^{fcd}	1	3
$\bar{u}_T^{\text{avi}}$	$\frac{1}{500}\Delta_{\text{avi}}$	1

* in (km/h)

original data (i.e., the structural error or *bias*: $\langle z \rangle - \langle u_S \rangle$)[3]. On the basis thereof, the following values (which differ slightly from the ones proposed for real traffic data in Treiber and Helbing (2002)) are used

$$c_{\text{cong}} = -25 \text{ km/h}$$
$$c_{\text{free}} = 80 \text{ km/h}.$$

The other core filter parameters are chosen on the basis of Treiber and Helbing (2002), that is, $\sigma = 300$ (m); $\tau = 30$ (s); $\Delta V = 10$ (km/h); and $V_c = 80$ (km/h). For setting the parameters to calculate the data source specific weights in (17), the following rationale was used.

1. Harmonic mean speeds are more reliable than time mean speeds.
2. The unreliability of local data (from loops) increases faster with lower speeds than from spatial data. This is first of all due to increasingly smaller samples per time period (in the limit of zero speed the sample size is also zero), and secondly caused by the fact that the measurement equipment itself makes larger errors as vehicles pass over at lower speeds.
3. As in congestion, the speeds of vehicles are strongly correlated, we assume that speeds from FCD under congestion are more reliable than from local sensors.
4. The reliability of vehicle trajectories deduced from averaged realized travel times (see Section 3.3.2) is proportional to the distance between AVI stations.

The data source-specific weights and parameters are listed in Table 2 and were chosen on the basis of rough estimates on a selection of ground-truth data.

4.3 Scenarios: combinations of sensors and data reliability

Below we will assess the performance of the proposed data fusion technique under controlled scenarios which differ with respect to the following.

1. The data sources used: induction loops, FCD, and AVI (see Section 4.1).
2. The spacing of the loop detectors (500; 1,000; 1,500 m) and AVI systems (1,500; 3,000 m).
3. The type induction loop data, that is, harmonic versus time mean speed (u_H and u_T, respectively, see Section 4.1).
4. The percentage loop detector data missing, which is simulated by labeling M% uniformly random selected loop measurements $z^{(j)}(t_i, x_i)$ as missing, and replacing these with null values. These null values are then discarded by the EGTF algorithm. It is assumed that no FCD or AVI data are missing.
5. The percentage FCD, which we vary between 0, 2, 5, and 10% (see Section 4.1).

To limit the total number of scenarios evaluated, not all possible combinations are simulated. Table 3 overviews the in-total 35 data fusion scenarios considered, which are grouped into six scenario categories, listed as rows in Table 3. The first row depicts scenarios in which only harmonic mean speeds from loop detectors spaced 500 m apart are considered and where the amount of randomly missing data was varied. Rows 2 and 3 depict scenarios where different loop data are combined with different percentages FCD, whereas rows 4 and 5 list scenarios which combine loop data and AVI data.

Table 3
Data fusion scenarios assessed

	Induction loops				
No scenarios	*Loop spacing (m)*	*% Missing*	*Data type*	*% FCD*	*AVI spacing (m)*
5 (1 × 5)	500	5, 10, 20, 35, 50	u_H	-	-
9 (3 × 3)	500; 1,000; 1,500	10	u_T	2, 5, 10	-
9 (3 × 3)	500; 1,000; 1,500	10	u_H	2, 5, 10	-
6 (3 × 2)	500; 1,000; 1,500	10	u_T	-	1,500; 3,000
6 (3 × 2)	500; 1,000; 1,500	10	u_H	-	1,500; 3,000

Table 4
Performance measures

RMSE (m/s)	$\sqrt{\frac{1}{N_{km}}\sum_{k,m}^{N_{km}}(z(k,m)-u_S(k,m))^2}$
MAPE* (%)	$\frac{100}{N_{km}}\sum_{k,m}^{N_{km}}\lvert\epsilon^{rel}(k,m)\rvert$
MPE (%)	$\frac{100}{N_{km}}\sum_{k,m}^{N_{km}}\epsilon^{rel}(k,m)$
SPE (%)	$100\times\sqrt{\frac{1}{N_{km}}(\epsilon^{rel}(k,m)-\frac{MPE}{100})^2}$

* $\epsilon^{rel}(k,m)=\frac{z(k,m)-u_S(k,m)}{u_S(k,m)}$ represents the relative error.

4.4 Performance measures

As performance measures, we use the root mean of the squared error (RMSE) and three relative error measures, which, in contrast to RMSE, weigh errors relative to the ground-truth speeds. As it is predominantly errors at low speeds (under congestion) which are relevant for traffic management purposes, these relative error measures are more informative than RMSE. The relative measures (Table 4) include the mean percentage error (MPE) which indicates structural *bias*, the standard deviation of the percentage error (SPE), which is a measure for the variability around the MPE, and the mean absolute percentage error (MAPE) which provides a combined indication of the relative error (bias and variance).

5 RESULTS AND DISCUSSION

5.1 Quantitative results single data sources

Table 5 first lists the results of applying the filter to harmonic mean speeds from loops with increasing percentages of missing data. As can be expected, RMSE, MAPE, and SRE (a proxy for variance) increase with increasing amounts of missing data, albeit in a moderate and below proportional sense. This indicates that the filter provides a graceful deterioration of the resulting state estimate even if up to 50% of the data is (randomly) missing. To further substantiate this claim,

Table 5
Results increasing missing data

Data missing (%)	*RMSE*	*MAPE*	*MPE*	*SPE*
0	0.53	1.71	−0.36	3.81
5	0.58	1.85	−0.36	4.12
10	0.63	1.96	−0.38	4.38
20	0.72	2.24	−0.27	5.02
35	0.86	2.65	−0.25	5.90
50	1.08	3.24	−0.11	7.48

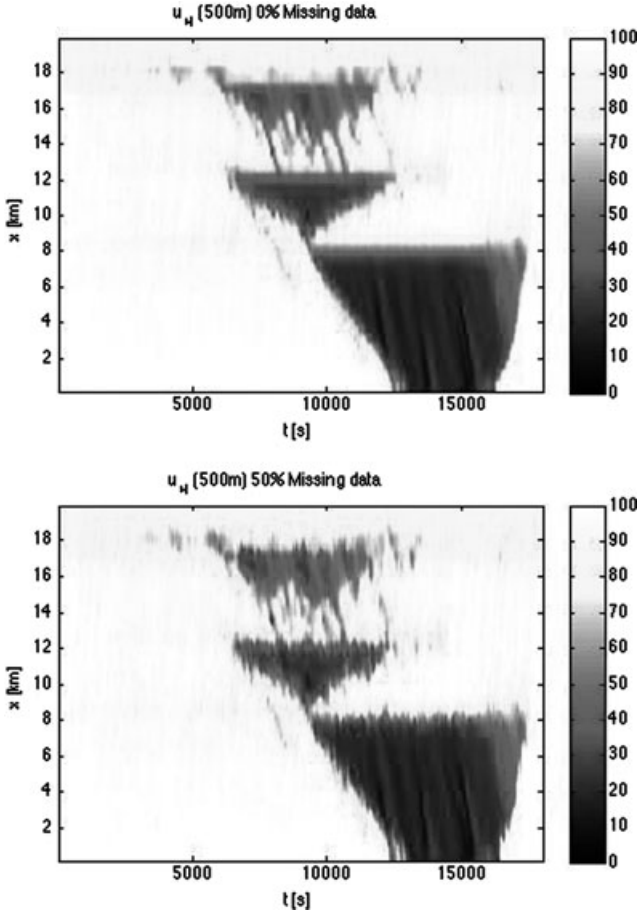

Fig. 6. Space-time contour map of harmonic mean speeds of 500 m spaced loops in case of 0% (top) and 50% (bottom) missing data.

Table 5 shows that the MPE (proxy for bias) does not increase at all and remains well below |1%| under all degrees of missing data. This is qualitatively illustrated by Figure 6, which shows that the resulting speed contour map in case of 0% and 50% randomly missing sensor data are still quite similar.

Table 6 lists the results of filtering each *individual* data source, where a random selection of 10% of the loop data was deliberately dubbed missing/unreliable matching approximately the degree of data corruption corresponding to practice (in the Netherlands). Note that the results differ slightly from 5 due to a different random seed which was used to generate the sensor data for this scenario.

As expected, both bias (MPE) and variance (SRE) as well as the overall error (RMSE and MAPE) increase with wider sensor spacing (both for AVI and loops) and with smaller data samples (i.e., lower percentages FCD).[4] Also the difference in bias between time mean speeds (u_T^{loop}) and harmonic mean speeds (u_H^{loop}) is clearly observable by comparing MPE (and

Table 6
Results individual data sources. Note that 10% of the raw data from loops (u_T^{loop} and u_H^{loop}) is dubbed unreliable/missing

		RMSE	*MAPE*	*MPE*	*SPE*
u_T^{loop}	500 m	0.67	2.63	2.02	4.61
	1,000 m	1.06	3.71	2.48	7.05
	1,500 m	1.80	6.35	2.83	14.38
u_H^{loop}	500 m	0.60	1.94	−0.42	4.42
	1,000 m	1.04	3.22	0.13	7.03
	1,500 m	1.83	5.97	0.41	13.72
FCD	2%	1.94	6.86	−2.98	10.04
	5%	1.46	5.07	−2.18	7.68
	10%	1.29	4.42	−2.25	6.91
AVI	1,500 m	2.32	6.80	−1.60	12.40
	3,000 m	3.49	11.07	−2.06	19.33

MAPE). In the former, this bias is positive (structural overestimation of speeds) and in the latter it is virtually zero. Clearly, on the basis of a single source of data, densely spaced loop detectors are superior for speed estimation over the other data sources.

5.2 Quantitative results of data fusion

Tables 7, 8, and 9 show the MAPE, MPE ("bias"), and SPE ("variance") results of all data fusion scenarios considered. The first observation one can make is that fusing time mean speeds from densely spaced loops ($\Delta_{\text{loop}} = 500$ m) with FCD practically eliminates the structural bias. This is still true for larger detector spacing, in which case also the variance is significantly reduced, compared to the case where no additional data sources are used.

The same holds also for (the in principle more accurate) harmonic mean speeds; also then the addition of FCD data decreases error variance. For example, the state estimate on the basis of just u_H^{loop} spaced at 1,500 m results in a bias (MPE) of 2.8% and a relative standard deviation (SPE) of 14% (see Table 6). After fusing these loop data with 5% FCD, the MPE reduces to around 2% and the standard deviation is reduced to less than 7%. Less dramatic reductions in errors can be observed when combining loop data with the AVI data. Although densely spaced loops producing harmonic mean speeds would still yield the best overall results, wider loop spacing in combination with either FCD or AVI data provides clear added value in terms of performance. Vice versa, (offline) state estimation on the basis of just data from AVI stations spaced at 3,000 m (i.e., realized travel time) can be improved in terms of both bias and variance significantly by adding additional loops (or other local detection equipment), even at a wide spacing of 1,500 m. In that case, for example, the SPE reduces from over 19% to 14% and MPE from −2% to around −0.5%, yielding improvements in variance and bias of 25 and 75%, respectively.

Table 7
MAPE results of data fusion

		FCD			*AVI*	
		2%	5%	10%	1,500 m	3,000 m
u_T^{loop}	500 m	3.40	3.51	3.58	3.57	4.83
	1,000 m	4.37	4.06	3.90	4.39	6.39
	1,500 m	5.13	4.38	4.09	5.05	7.71
u_H^{loop}	500 m	3.51	3.61	3.66	3.26	4.36
	1,000 m	4.44	4.12	3.95	4.20	6.06
	1,500 m	5.21	4.42	4.11	4.94	7.51

Table 8
MPE results of data fusion

		FCD			*AVI*	
		2%	5%	10%	1,500 m	3,000 m
u_T^{loop}	500 m	−0.53	−1.02	−1.58	0.80	1.35
	1,000 m	−1.08	−1.39	−1.83	0.44	1.04
	1,500 m	−1.65	−1.68	−2.01	−0.00	0.31
u_H^{loop}	500 m	−1.47	−1.52	−1.87	−0.37	−0.14
	1,000 m	−1.68	−1.67	−1.98	−0.34	−0.01
	1,500 m	−2.09	−1.88	−2.11	−0.62	−0.52

Table 9
SPE results of data fusion

		FCD			*AVI*	
		2%	5%	10%	1,500 m	3,000 m
u_T^{loop}	500 m	5.57	5.66	5.79	6.47	9.28
	1,000 m	6.97	6.40	6.23	7.87	11.98
	1,500 m	7.97	6.75	6.42	9.24	14.39
u_H^{loop}	500 m	5.60	5.70	5.82	6.01	8.44
	1,000 m	6.99	6.42	6.24	7.58	11.39
	1,500 m	7.99	6.76	6.43	9.09	14.03

5.3 Qualitative results and discussion

To illustrate the quantitative results presented above, Figure 7 shows contour maps of speeds reconstructed by a number of different combinations of data sources. When looked upon from a distance, each of the presented speed contour estimates appears to reproduce the congestion patterns correctly, that is, each combination reproduces congestion at the right place and the right time instants. Figure 7 illustrates that the

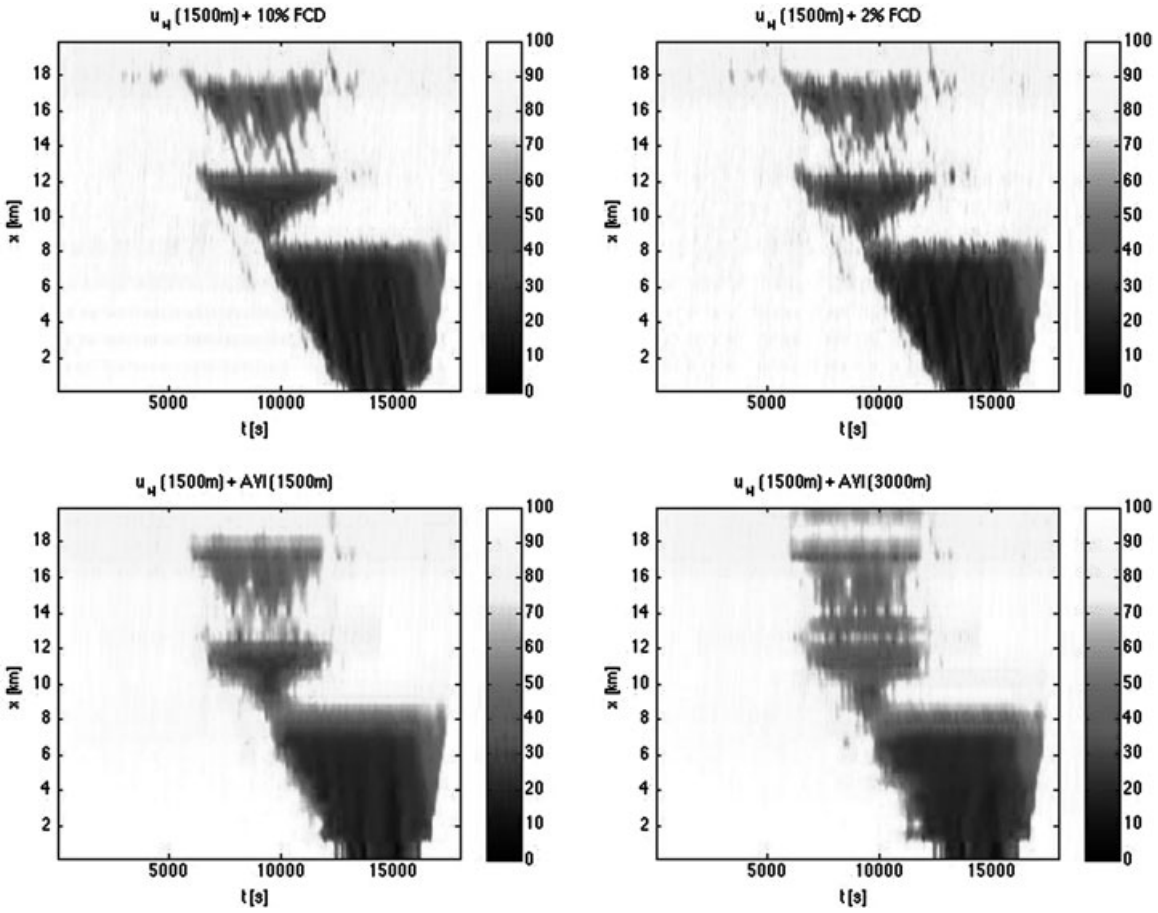

Fig. 7. Space-time contour maps of four example data fusion combinations.

additional bias and variance caused by wider spacing, lower samples FCD, and/or the use of coarser AVI data reflect errors predominantly in the estimates around the edges (head and tail) of the congested regions. A more densely spaced sensor configuration reveals more accurately the trajectories of shock waves and the boundaries between congested and non-congested traffic. Alternatively, fusing different coarsely spaced (or sampled) data sources has the same effect. Data fusion does indeed improve the state estimation/data reconstruction significantly. The results presented above in our view illustrate the robustness and suitability of the proposed filter to this particular application. It enables fusing both statistically and semantically very different data into a consistent and coherent space-time grid of estimates.

There are a number of issues which should be kept in mind when interpreting these results. The proposed approach enables the reconstruction of any traffic quantity which propagates over space time along characteristics (speeds, density, flow, percentage trucks, vehicle composition, etc.), on the basis of any type of sensor data, under the premise that these sensor data provide information on the traffic quantity of interest. The approach can only be applied if the relationship (over space-time) between the quantity of interest and the actual data measured can be quantified, which in not all cases is straightforward or even possible.

In the example in this article, the state estimation quantity chosen is space mean speed. All data sources considered (loops, FCD, and AVI) provide information related to this quantity. If, however, the choice was made to estimate density, the use of FCD data would have been more difficult. Possibly, one could use the fundamental diagram to relate FCD speeds to density under congested conditions and use an estimate of the penetration rate of the equipped vehicles to estimate the density under free-flow conditions. Another possibility could be in the use of *other* FCD data if available, such as time headway or distance gap to predecessor, which could be used as samples of the average flow and density. Either way, in the proposed data fusion method, a low setting of the data source weights could be used to emphasize a possibly weak relation between

FCD data and density. As such, many data sources can be exploited, even if these provide only very coarse and inaccurate information.

Let us finally remark that the results here provide a proof of concept. They are the result of common sense but nonetheless to a degree arbitrary choices made for the filter parameters and the data source weights. More elaborate analysis on the basis of detailed (and real) vehicle trajectory data is needed to assess the impacts of the parameter and weight setting. In the next and final section, we will make some recommendations for further rationalizing these settings and implementing the filter in a real traffic data archiving system.

6 CONCLUSIONS, IMPLICATIONS, AND FURTHER RESEARCH

In this article, a new robust and efficient method is put forward for fusing heterogeneous and semantically different data from different traffic sensors, such as local sensors (induction loops, radar, microwave, etc.), AVI systems, or data from moving vehicles (FCD). The method is generic in that it allows fusion of data from any (fixed or moving) data source. Moreover, the method enables the reconstruction of any traffic quantity which evolves over space and time along traffic characteristics, provided each data source used contains information (however coarse or imprecise) which can be quantitatively related to the traffic quantity of interest.

The main conclusions on the basis of experiments with synthetic data from a validated micro-simulation tool (providing both ground-truth and various sensor data) are as follows:

1. The method is robust to random amounts of missing data. The filter produces an nearly unbiased estimate of the ground-truth data even if up to 50% of the data is randomly missing.
2. The method also works with coarse data, which could be due to either widely spaced sensors, or for example structural failure of sensors. In that case however, estimation bias increases with wider spacing due to estimation errors at the head and tail of congestion.
3. In case the available sensor data are coarse, fusing multiple (coarse) data sources leads to a significant reduction in both estimation bias and variance of up to 75% and 25%, respectively

The proposed data fusion method as presented in this article constitutes a simple but efficient algorithm which could be implemented and used in central traffic data archiving and processing facilities to perform a number of critical tasks:

Data visualization. The method produces equidistant rectangular contour maps of speeds, flows, or whatever traffic quantity is chosen which allow fast and easy visualization (imaging) of the sensor data.

Data cleaning and filtering. If no data are available on a particular location/time, the filter estimates these from data which *are* available. By further analyzing the (data source specific) weights (e.g., in respect to those associated with neighboring sensors) also mechanisms can be designed to detect possibly unreliable data in the first place. This is a subject for further research.

(ex post) Incident/bottleneck detection. Causes for traffic congestion (bottlenecks or incidents) can be more accurately pinpointed to a particular time and location, and the consequences (the resulting congestion) can be more accurately estimated and visualized. Further research could focus on automated procedures of detecting frequent or incidental bottleneck locations.

Travel time estimation. The equidistant time-space contours of speeds provide the ideal ingredients for trajectory-based travel time estimation (see e.g., Van Lint and Van der Zijpp, 2003).

The above points all relate to offline use of the filter. There may also be online applications possible. The filter could be used for real-time state estimation, where the state (e.g., speed/density) at a certain time t along a certain route would be continuously updated with the arrival of new data (from whatever source). This state estimate could then be input to many possible real-time applications (ramp metering, travel time prediction, routing algorithms, etc.). The filter may even be used for rough prediction purposes by itself, although it could never predict any state change (e.g., the onset of congestion) if such a change is not already visible in (some of) the sensor data.

Let us finally propose a few directions of further research in terms of the underlying methodology. The filter contains a number of parameters which were set on the basis of common sense or engineering judgement in this article. Further research is needed into principled and preferably automated procedures, which derive these parameters directly from the data (from whichever source), or by combining *a priori* estimates and observations, for example, on the basis of Bayesian approaches.

ACKNOWLEDGMENT

This research is supported by the Dutch Technology Foundation STW, applied science division of NWO, and

the Technology Program of the Ministry of Economic Affairs under grant DCB 7814.

NOTES

1. Corrected for arithmetic averaging on the basis of Van Lint (2004).
2. FOSIM (Freeway Operations Simulation), developed at Delft University of Technology, is a traffic microsimulation package which has been extensively calibrated and validated for traffic operations on Dutch freeways and is used by the Dutch Ministry of Transport, Public Works and Water Management as a freeway design tool for *ex ante* freeway capacity estimation FOSIM (*FOSIM (Freeway OperationsSIMulation)*, 2008; Vermijs and Schuurman, 1994).
3. Minimizing the (root) mean squared error may in fact induce bias, as the filter algorithm deliberately removes the high frequency deviations (smaller than τ (s) and σ) (m) from the data.
4. Note how the increase in the error is more pronounced in the relative error measures than in the RMSE, which is caused by the fact that the RMSE weighs errors at higher speeds more heavily than errors at lower speeds. Relatively, errors at lower speeds are much larger and more important for many applications, for example, travel time estimation.

REFERENCES

Castillo, A. E., Conejo, A. J., Menéndez, J. M. & Jiménez, P. (2008), The observability problem in traffic network models, *Computer-Aided Civil and Infrastructure Engineering*, **23**(3), 208–22.

Castillo, A. E., Menéndez, J. M. & Sánchez-Cambronero, S. (2008), Traffic estimation and optimal counting location without path enumeration using Bayesian networks, *Computer-Aided Civil and Infrastructure Engineering*, **23**(3), 189–207.

Dailey, D. J., Harn, P. & Lin, P.-J. (1996), *ITS data fusion*, Technical Report Research Project T9903, Task 9, Washington State Transportation Commission, Department of Transportation and the U.S. Department of Transportation, Federal Highway Administration.

Dharia, A. & Adeli, H. (2003), Neural network model for rapid forecasting of freeway link travel time, *Engineering Applications of Artificial Intelligence*, **16**(7–8), 607–13.

FOSIM (Freeway Operations SIMulation) (2008), http://www.fosim.nl/IndexUK.html.

Hegyi, A., De Schutter, B. & Hellendoorn, H. (2005), Model predictive control for optimal coordination of ramp metering and variable speed limits, *Transportation Research Part C: Emerging Technologies*, **13**(3), 185.

Helbing, D. & Treiber, M. (2002), Critical discussion of synchronized flow, *Cooperative Transportation Dynamics*, **1**, 2.1–2.24.

Hellinga, B. R. (2002), Improving freeway speed estimates from single-loop detectors, *Journal of Transportation Engineering*, **128**(1), 58–67.

Herrera, J. C. & Bayen, A. M. (2008), Traffic flow reconstruction using mobile sensors and loop detector data, *Transportation Research Board 87th Annual Meeting*, Transportation Research Board, p. 18, 08-1868.

Hoogendoorn, S. P. & Bovy, P. H. L. (2001), State-of-the-art of vehicular traffic flow modeling, *Proc. Institution of Mechanical Engineers*, **215**(1), 283–303.

Ivan, J. N. & Sethi, V. (1998), Data fusion of fixed detector and probe vehicle data for incident detection, *Computer-Aided Civil and Infrastructure Engineering*, **13**, 329–37.

Karim, A. & Adeli, H. (2003), Fast automatic incident detection on urban and rural freeways using wavelet energy algorithm, *Journal of Transportation Engineering, ASCE*, **129**(1), 57–68.

Kerner, B. S. (2004), *The Physics of Traffic: Empirical Freeway Pattern Features, Engineering Applications, and Theory*, Springer, Berlin, New York.

Kesting, A. & Treiber, M. (2008), How reaction time, update time, and adaptation time influence the stability of traffic flow, *Computer-Aided Civil and Infrastructure Engineering*, **23**, 125–37.

Kikuchi, S., Miljkovitc, D. & van Zuylen, H. (2000), Examination of methods that adjust observed traffic volumes on a network, *Transport Research Record*, **1717**, 109–19.

Lighthill, M. & Whitham, G. (1955), On kinematic waves ii: a theory of traffic flow on long crowded roads, *Proceedings of the Royal Society of London Series A*, **229**(1178), 317–45.

Linn, R. J. & Hall, D. L. (1991), A survey of multi-sensor data fusion systems, *Proceedings of the SPIE - The International Society for Optical Engineering*, Orlando, Florida, 13–29.

Luo, R., Yih, C.-C. & Su, K. L. (2002), Multisensor fusion and integration: approaches, applications, and future research directions, *Sensors Journal, IEEE*, **2**(2), 107–19.

Newell, G. (1993a), A simplified theory of kinematic waves in highway traffic: part i. General theory, *Transportation Research Part B: Methodological*, **27**(4), 281–87.

Newell, G. (1993b), A simplified theory of kinematic waves in highway traffic: part ii. Queueing at freeway bottlenecks, *Transportation Research B: Methodological*, **27**(4), 289–303.

Ni, D. & Wang, H. (2008), Trajectory reconstruction for travel time estimation, *Journal of Intelligent Transportation Systems*, **12**(3), 113–25.

Ou, Q., Van Lint, J. & Hoogendoorn, S. (2008), Piecewise inverse speed correction by using individual travel times, *Transportation Research Record: Journal of the Transportation Research Board*, **2049**(-1), 92–102. 10.3141/2049-11.

Papageorgiou, M. & Vigos, G. (2008), Relating time-occupancy measurements to space-occupancy and link vehicle-count, *Transportation Research Part C: Emerging Technologies*, **16**(1), 1–17.

Piella, G. (2003), A general framework for multiresolution image fusion: from pixels to regions, *Information Fusion*, **4**(4), 259–80.

Richards, P. (1956), Shock waves on the highway, *Operations Research*, **4**, 42–51.

Simon, D. (2006), *Optimal State Estimation: Kalman, H-infinity, and Nonlinear Approaches*, John Wiley & Sons, New York.

Treiber, M. & Helbing, D. (2002), Reconstructing the spatio-temporal traffic dynamics from stationary detector data, *Cooper@tive Transport@tion Dyn@mics*, **1**, 3.1–3.24.

Van Lint, J. & Hoogendoorn, S. P. (2007), The technical and economic benefits of data fusion for real-time monitoring of freeway traffic, *World Congress of Transportation Research*, WCTRS, Berkely, California.

Van Lint, J. W. C. (2004), *Reliable travel time prediction for freeways*, TRAIL Thesis series, TUD Technische Universiteit Delft, Delft.

Van Lint, J. W. C., Hoogendoorn, S. P. & Hegyi, A. (2008), Dual ekf state and parameter estimation in multi-class first-order traffic flow models, *Proceedings of the 17th IFAC (International Federation of Automatic Control) World Congress*, Seoul, Korea.

Van Lint, J. W. C., Hoogendoorn, S. P. & Van Zuylen, H. J. (2005), Accurate travel time prediction with state-space neural networks under missing data, *Transportation Research Part C: Emerging Technologies*, **13**(5–6), 347–69.

Van Lint, J. W. C. & Van der Zijpp, N. J. (2003), Improving a travel time estimation algorithm by using dual loop detectors, *Transportation Research Record*, **1855**, 41–48.

Varshney, P. K. (1997), Multisensor data fusion, *Electronics and Communications Engineering Journal* (December), 245–53.

Vermijs, R. G. M. M. & Schuurman, H. (1994), Evaluating capacity of freeway weaving sections and on-ramps using the microscopic simulation model fosim, *Proceedings of the second international symposium on highway capacity*, Vol. 2, Sydney, Australia, 651–70.

Wang, Y. & Papageorgiou, M. (2005), Real-time freeway traffic state estimation based on extended Kalman filter: a general approach, *Transportation Research Part B*, **39**, 141–67.

Wang, Y., Papageorgiou, M. & Messmer, A. (2006), A real time freeway network traffic surveillance tool, *IEEE Transactions on Control Systems Technology*, **14**(1).

Xiong, N. & Svensson, P. (2002), Multi-sensor management for information fusion: issues and approaches, *Information Fusion*, **3**(2), 163–86.

Yager, R. R. (2004), A framework for multi-source data fusion, *Information Sciences*, **163**(1–3), 175–200.