EJTIR

Issue **8**(4)
December 2008
pp. 333-354
ISSN: 1567-7141
www.ejtir.tbm.tudelft.nl

A New Method for Travel Time Estimation on Long Freeway Sections

Albert Steiner
Institute of Data Analysis and Process Design, Zurich University of Applied Sciences[1]

Beate Sick
Institute of Data Analysis and Process Design, Zurich University of Applied Sciences[2]

Received March 2008; accepted July 2008

The knowledge of travel times on road networks is of vital importance for network operators as well as for drivers. Operators can use travel time information to improve control on their networks. Drivers or transport companies can choose their optimal route based on the traffic information available and their individual preferences. The presented approach focuses on travel time estimation. The method only requires the time stamps and vehicle lengths captured at subsequent detector stations. The distance between the stations is up to 13 km. The examined network is a two-lane freeway with three unobserved on- and off-ramps each. In order to investigate the influence of measurement errors and to have true data for comparison, simulated detector data based on an existing Swiss freeway section equipped with loop detectors was used. The method shows very good performance for the investigated scenarios. All relevant characteristics of the travel time process were detected and estimation errors were within a well acceptable range. Compared to existing travel time estimation methods, the presented approach considerably extends the maximum distance for which travel time estimations can be carried out.

Keywords: freeway traffic; image processing; pattern recognition; travel time estimation

1. Introduction

Travel time is well accepted as a measure to assess road conditions. The knowledge of travel times on road networks is of vital importance, for network operators as well as for drivers. Operators can use travel time information (current and/or predicted) to improve control on their networks as part of Advanced Traffic Management Systems. Drivers can choose their optimal route, either pre-trip or en route, provided that traffic information from an Advanced Traveller Information System is available together with the drivers' individual preferences. For transport companies the knowledge of travel time can help to improve their service quality. Moreover, they

[1] P.O. Box 805, Rosenstrasse 3, 8401 Winterthur, Switzerland, T: +41589347801, F: +41589357801, E: albert.steiner@zhaw.ch
[2] P.O. Box 805, Rosenstrasse 3, 8401 Winterthur, Switzerland, T: +41589347813, F: +41589357813, E: beate.sick@zhaw.ch

can choose their routes dynamically according to the current and predicted state of traffic and thus increase their efficiency and reliability.

Before discussing travel time estimation, we need to explain the terms "travel time estimation" and "travel time prediction". Definitions of these topics can be found in van Lint (2004) or in Bovy and Thijs (2000). Based on these definitions we can say that *travel time estimation* calculates travel times based on known quantities up to the current point in time. *Travel time prediction* forecasts travel times up to a defined point in time in the future.

In this research, we focus on travel time estimation, i.e. we estimate how long it takes vehicles to traverse a defined freeway route when they arrive at the downstream station. However, we do not know the traffic conditions between the detector stations, but we measure vehicle lengths and passage times at the detector stations.

Travel time estimation on a freeway route can be performed either by direct or indirect estimation approaches. Introductions on *travel time estimation* are provided by Bovy and Thijs (2000), Lindveld et al. (2000) or van Lint (2004).

The techniques of *direct travel time estimation* can generally be divided into three major groups. Firstly, vehicles that are equipped with a GPS module and a transmission interface (e.g. GSM) can send information regarding their position, speed etc. to a service centre (e.g. Quiroga and Bullock (1998), Yim and Cayford (2001), Li and McDonald (2002) or Brockfeld et al. (2007)). The advantage of this approach is that accurate vehicle positions and, provided an appropriate sampling rate, accurate travel time estimates become available. However, due to expensive car-side investments and data transmission costs, the diffusion rate of this technology is still very limited. Secondly, vehicles equipped with a cellular phone or module can be tracked by cell location (e.g. Yim (2003), Yim and Cayford (2006), Pathirana et al. (2006), Bar-Gera (2007)). Although the accuracy of determining the position depends, amongst other things, to a large extent on the size of the cellular phone network cell and the transmission quality, the high diffusion rate of GSM together with the low transmission costs makes this technology an attractive alternative. However, the accuracy of the position and thus of the travel time is usually lower than by using GPS/GSM technologies. A third way is provided by local detection of unique vehicle features (e.g. ID, license plate number). Common technologies in this area are AVI (Automatic Vehicle Identification; e.g. Dion and Rakha (2006)), license plate recognition (e.g. Anagnostopoulos et al. (2006)) or communication technologies like tag reading (e.g. RFID) or Bluetooth (e.g. c.c.com (2006)). AVI technologies have usually low error rates in vehicle identification and thus are well applicable for travel time estimation. However, due to expensive additional investments in road- and/or car-side infrastructure they have limited use so far. Furthermore, especially for the license plate recognition (but also for other technologies), there exist privacy issues, which additionally limit its widespread use.

Indirect travel time estimation approaches can be subdivided broadly in three groups: estimation of section travel times by means of local measurement of aggregated data (e.g. flow, speed), reidentification approaches, which make use of individual vehicle features (e.g. length), and, as a third approach, determining vehicle positions from a sequence of captured aerial images. Once the vehicle positions are available from the images one can compute the travel times of the vehicles.

Firstly, if the distance between detector stations is not too large (i.e. in the range of a few hundred metres), the estimation of section travel times is very common. They usually rely on aggregated data (e.g. speed, flow, cumulative number of vehicles) collected at local detector stations. Examples can be found in Bovy and Thijs (2000), Lindveld et al. (2000), Nam and Drew (1996), Westerman and Immers (1992) or Oh et al. (2002). The major advantages of this method are that it makes use of already existing infrastructure and that it works fully anonymous. However, as already mentioned, one drawback is the relatively small length of the sections.

Secondly: As loop detectors are in many regions still the most widely used detector technology, many of the published reidentification approaches deal with data from loop detectors. In the following we consider two groups of approaches in detail: methods that are based on voltage signatures (or other physical quantities) and especially procedures that are based on noisy length measurements. There is a variety of other detector technologies that provide vehicle features. A number of surveys can be found in Nam and Drew (1996), Michalopoulos and Hourdakis (2001) or van Lint (2004). Furthermore, the industry together with research institutions is currently making endeavours to develop alternatives to loop detectors, e.g. wireless magnetic sensors (see Haoui et al. (2008)).

The raw (sampled) voltage signatures (or other physical quantities) from a loop detector provide an "electro-magnetic characterisation" of the vehicles passing a detector. Pfannerstill (1984), Kühne and Immes (1993), Abdulhai and Tabib (2003) or Kwon (2006) made use of these signatures to classify and/or reidentify single vehicles. The advantages of this technology are that it uses the raw signature information and thus provides valuable information that characterises a vehicle. This information can be used subsequently by matching algorithms. At the same time, the approach works fully anonymous. However, such signatures are usually not available from standard loop detectors. Moreover, depending on the sampling rate, a considerable amount of data is generated, which requires additional and fast transmission equipment as well as external storage capacities. A potential drawback is that the signatures can be distorted if a vehicle changes lanes while crossing the loops.

Vehicle length measurements provided by detectors rely, amongst other things, on the signatures discussed before. Their accuracy is often very limited due to low sampling rates. In this field, Coifman and co-workers (e.g., Coifman (1998a, 1998b, 2002), Coifman and Cassidy (2002), Coifman and Ergueta (2003), Coifman and Krishnamurthy (2007)) have made important contributions during the last decade with the development of algorithms to estimate travel times, trajectories and road conditions by using vehicle length.

The method presented in this paper makes also use of the vehicle lengths provided by double inductive loop detectors. The only information required is the vehicle length for every vehicle at the upstream and downstream detector station together with the corresponding time stamp. The major difference of the proposed method in comparison to existing reidentification algorithms is that we do not strictly want to reidentify single vehicles but rather try to find groups (platoons) of vehicles with similar arrival times and at the same time similar travel times over the road network.

The presented approach has the following pros and cons: Provided that single car data are available online and in sufficient quality (i.e. small length measurement error, no or small bias), no additional infrastructure investments are required. Another advantage of this method is that it works fully anonymous, i.e. no conclusions can be drawn on a particular vehicle. Moreover, the presented approach considerably extends the maximum distance for which travel time estimations can be carried out. In addition, the method can be easily extended to new technologies or even use data from mixed sources, provided again, that the data fulfil some quality requirements. Finally, extensions to more sophisticated detection technologies that provide additional vehicle features are straightforward, but will not be investigated in this work. However, as for the use of the vehicle signatures, one potential drawback is that lane changes can distort the measured vehicle length and thus might lead to wrong estimates.

We want to mention here, that the proposed approach is on estimating mean travel times and not on estimating individual travel times. Furthermore, in this work we focus on estimating mean arrival travel time, i.e. travel times related to arrival times at the destination. The step from mean arrival travel times to mean departure travel times is, however usually not straightforward. As mentioned also in section 6 (Conclusions and future work), one of the next steps is thus to derive

mean departure travel time estimates, which are an important basis for travel time prediction algorithms.

The third group of indirect travel time estimation includes approaches to determine the vehicle position on a sequence of aerial images and from this to compute the travel time (e.g. Coifman et al. (2006), Toth and Grejner-Brzezinska (2006), Reinartz et al. (2006)). One of the advantages of this approach is that not only a few equipped vehicles are detected (tracked) but a whole traffic situation can be captured. This is of great importance when it comes to determine the traffic state within a long freeway section or network area. The drawbacks of this technology are, that the data collection task is usually very costly, that the quality depends to a large extent on the sensing technology (incl. camera system) and that the technology can not be applied to non-visible areas (e.g. tunnels).

The final part of the document is structured as follows: In section 2 we present the different steps of our approach and in section 3 we briefly explain the settings of the application. In section 4 we introduce two investigated scenarios, in section 5 we present and discuss the results, and in section 6 we conclude with a summary of the main achievements and give an outlook on further research.

2. Estimation approach

The method presented here aims to estimate the travel time between subsequent detector stations under the assumption that only noisy length measurements of individual vehicles and the associated time stamps as the vehicles pass the detectors, are available. The error of the length measurements is assumed to be known and constant over time. Section 2.1 gives a rough overview on the method and its steps, whereas section 2.2 describes the single steps more detailed.

2.1 The overall procedure

Given the noisy length measurements from the loop detectors together with the time stamps of the vehicles, we first compute the so-called scores for all possible vehicle-pairs (upstream and downstream). These scores can be regarded as a similarity measure of the lengths between two vehicles. Afterwards, the scores are written into the score matrix, with the idea to cumulate the scores of pairs of vehicles with similar downstream arrival time and similar travel time.
The subsequent post-processing steps together with the filtering step all intend to reveal the travel time information from a noisy environment. The successive computation of the most feasible path, i.e. the path with maximal costs, within the image finally leads to the travel time estimate.

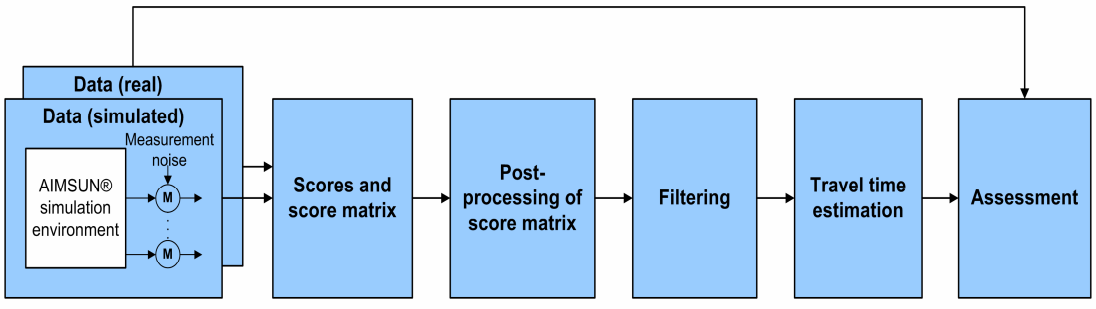
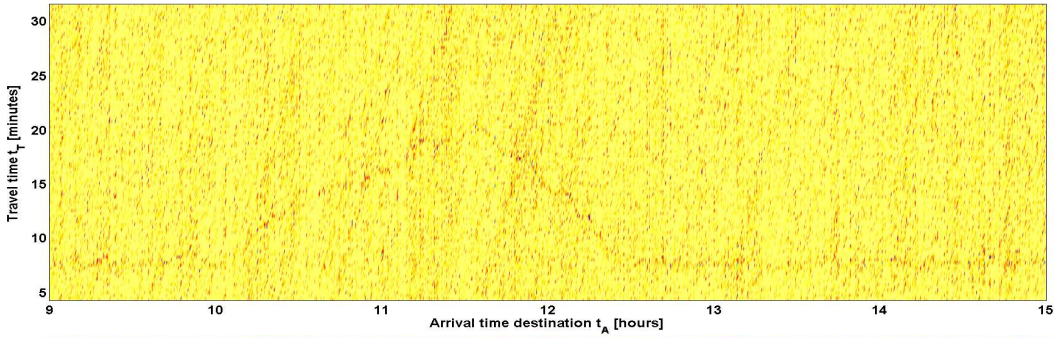


Figure 1. Overall procedure of the method.

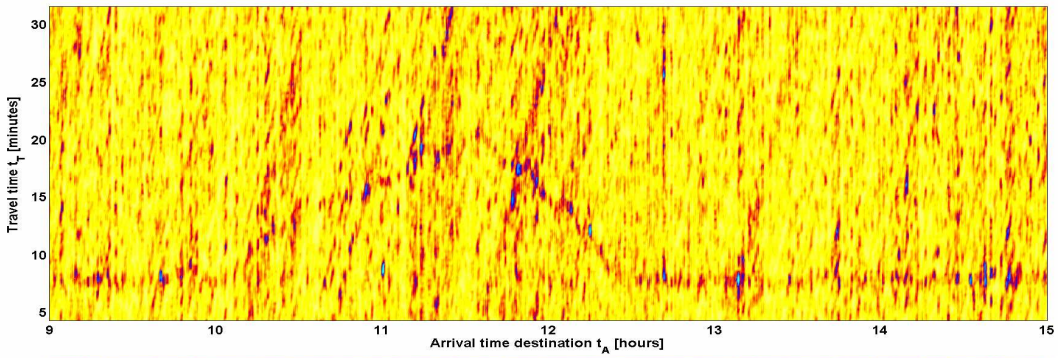
The raw data stem either from real detector stations or, as in our case, are generated artificially by the microscopic simulation environment AIMSUN®. The four major steps are processed sequentially resulting in the travel time estimates. The final comparison with true data allows for an assessment of these travel time estimates.

Figure 2 below shows the score matrix at some important stages during the information extraction process. This shall highlight how we get from only little information (Figure 2a) to the final travel time estimate (Figure 2e). In section 2.2 we will explain this procedure in more detail.

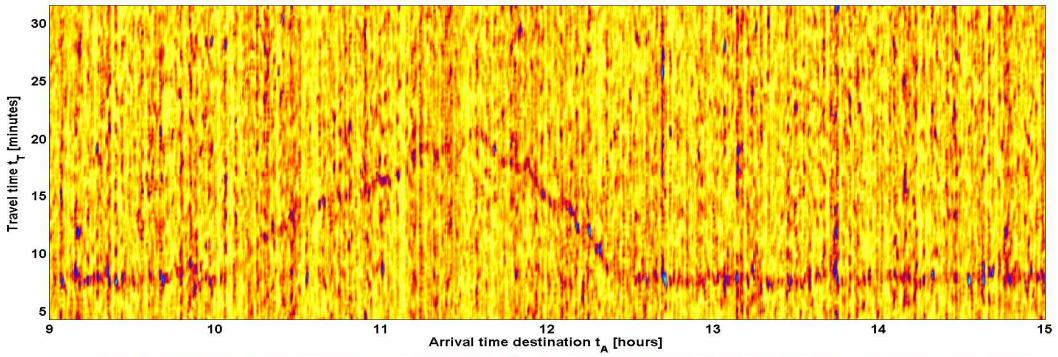
(a)



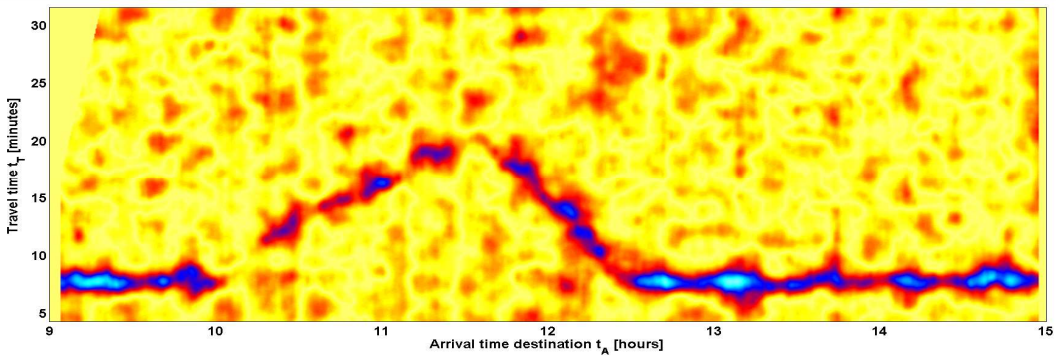
(b)



(c)



(d)



(e)

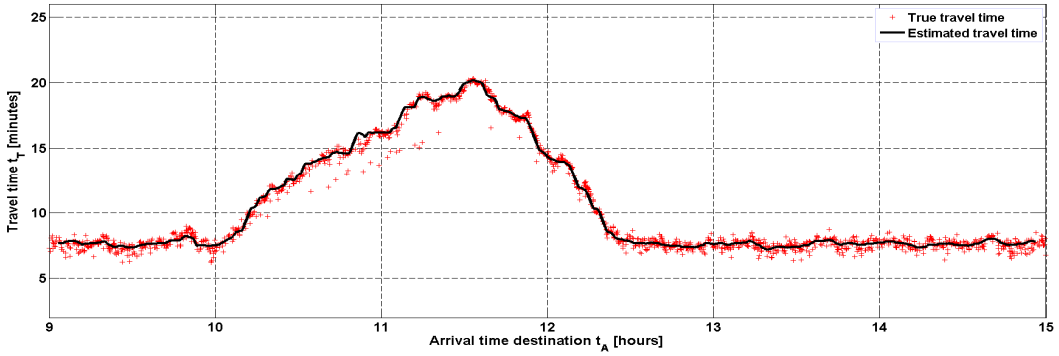


Figure 2. A graphical representation of the score matrix at some important steps within the travel time estimation procedure: (a) Raw score matrix (b) Score matrix after convolution, (c) Score matrix after weighting of the diagonals, (d) Score matrix after resizing and filtering, (e) Estimated travel time after finding the path with maximum costs from (2d), and comparison of estimated travel times (black line) with true travel times (red crosses).

The colour code for Figures 2a and 2d is in the range from yellow (0) to light blue (1). This implicates that yellow areas contain no information whereas red to light blue areas represent relevant travel time information. In Figure 2e, the red crosses indicate true travel times of individual vehicles and the black line represents the estimated travel times as a function of the downstream arrival time.

2.2 The steps in detail

2.2.1 The scores and the score matrix

In a first step, we compute the score for each vehicle-pair between the downstream and upstream vehicle combination. This means that for the investigated time window, the length of every vehicle that passed the downstream detector station is compared with the length of every vehicle at the upstream station.

A score s_{ij} is calculated for each combination of vehicles i and j observed at the up- and downstream station, respectively, as follows:

$$s_{ij} = \max\left(1 - \frac{\left|L_i - L_j\right|}{2 \cdot \Delta L}, 0\right) \tag{1}$$

where L_i and L_j denote the measured vehicle lengths at the upstream and downstream detector stations, respectively, and ΔL is the assumed measurement error (for details see section 3).

Once we have the scores available for each combination of vehicles i and j, we use them to fill a so-called score matrix **S**. In the score matrix each cell includes the cumulated scores of pairs of vehicles, which fall into the same time-bin according to their arrival time and travel time, respectively.

The score matrix is of size $X \times Y$ (width × height), where X and Y denote the number of columns and rows and correspond to the number of time bins in the arrival and in the travel time direction, respectively. Each row represents a travel time bin $b_y = [t_y - \Delta t, t_y)$ and each column represents an arrival time bin $b_x = [t_x - \Delta t, t_x)$. Δt represents the time resolution in both directions, i.e. the size of the bins. Detailed explanations on the size of the bins and its implications can be found in the section 'Post-processing of the score matrix – Convolution' below.

Furthermore, t_i and t_j denote the points in time, where vehicles i and j pass the upstream and downstream station, respectively. Thus, if a vehicle j arrives at the downstream station within

time bin b_x and if at the same time the difference of the arrivals ($t_j - t_i$) is within the time bin b_y, then the score s_{ij} is added to cell $S(x,y)$. Formally written, this leads to

$$S(x,y) = \sum_{\substack{t_j \in b_x \\ t_j - t_i \in b_y}} s_{ij} \qquad \forall\, i, j, x, y \tag{2}$$

where $i \in \{1,...,I\}$, $j \in \{1,...,J\}$, $x \in \{1,...,X\}$, and $y \in \{1,...,Y\}$. I and J represent the total number of vehicles that passed the upstream and downstream detector station within the considered time window, respectively.

2.2.2 *Post-processing of the score matrix*

Convolution

High scores s_{ij} indicate that the compared vehicles have similar lengths. From this follows, that cells with high values within the score matrix indicate that several vehicles with similar arrival time and similar travel time have passed the downstream detector station. This score matrix is visualised in Figure 2a according the specified colour code: yellow cells correspond the low scores whereas red to light blue cells correspond to high scores. Since the travel time develops over time in a continuous manner, in the best case we expect to see some light blue curve corresponding to the travel time development in Figure 2a. However, at this point we have only rough information on how the true travel time curve will look like. To make the curve more visible, we firstly perform a smoothing procedure resulting in Figure 2b. This is done by convolution of the score matrix **S** with a box-kernel, i.e. matrix $\mathbf{B} = \mathbf{1}_{[b \times b]}$, where **B** is a matrix with all ones and a boundary length b. In our case we set $b = 2$. Given a time resolution in both directions of $\Delta t = 15$ seconds, this results in a boundary time length of $b \cdot \Delta t = 30$ seconds. Matrix **T** is computed according to

$$\mathbf{T} = \mathbf{S} * \mathbf{B} \tag{3}$$

where $*$ denotes the convolution of two matrices.

Vehicle speeds are correlated for traffic densities $\rho \geq 10$ vehicles per kilometre (e.g. Helbing (2001), Helbing et al. (2002)). Following Neubert et al. (1999, fig. 15), for congested traffic, the speed auto-covariance is ≥ 0.4 for up to $N = 17$ consecutive vehicles. However, even a lower auto-covariance could be considered, leading to a higher number of correlated vehicles. The maximum, meaningful number of correlated vehicles according to Neubert et al. (1999) is around 35-40.

Together with the fundamental diagram according to Helbing (1997, p. 159; with a maximum velocity of 120 km/h), with $N = 17$ and densities between 12 and 50 vehicles per kilometre, $\tau(N,\rho)$ for correlated speeds is in the range of 30 to 50 seconds. This is well in accordance with the dimension of our convolution matrix, which has a boundary length of 2 time intervals Δt, i.e. 30 seconds, since $\Delta t = 15$ seconds. It is important to mention that Δt and b can vary in quite a wide range, i.e. 3 sec. ≤ Δt ≤ 15 sec. and $2 \leq b \leq 18$ as long as their product, i.e. $b \cdot \Delta t$, is within a range of 24 to 54 seconds. Below and above this range the travel time estimates can deviate significantly from the true travel time. As long as $b \cdot \Delta t$ is constant (and within the allowed range), the travel time estimates are more or less identical.

In the explanations so far we worked with $\Delta t = 15$ seconds and $b = 2$. Besides that, we also made some tests with $\Delta t = 3$ seconds, mainly because the range for $b \cdot \Delta t$ could be investigated in smaller steps than with larger Δt. In general, increasing Δt while keeping b constant leads to more blurred images and, by exceeding the upper boundary for $b \cdot \Delta t$, to less accurate travel time estimates.

Besides the use of a simple two-dimensional box-kernel we also performed tests with a two-dimensional Gaussian-kernel. However, the results were almost similar to those with the box-kernel and thus we decided to choose the box-kernel for further investigations because it is simpler and requires less computational performance.

By carrying out the convolution, we "connect" areas within matrix **S** whose speed, and hence travel time, is correlated up to a certain degree. The "interaction" time $\tau(N,\rho)$ is defined by

$$\tau(N,\rho) \approx \frac{N}{V_e(\rho)\rho} \tag{4}$$

where N denotes the number of correlated vehicles, ρ represents the density and $V_e(\rho)$ denotes the equilibrium velocity against ρ. Please note, that N considers all vehicles, independently of their length. By contrast, the method presented here considers only vehicles with length $L \geq 6$m. However, this does not influence the idea of "connecting" areas within **S** as, given $\tau(N,\rho)$, there are simply less vehicles to "connect".

As we shall see below in the two scenarios in section 4, the densities are around 15 to 20 vehicles per kilometre in normal, i.e. uncongested traffic states and up to 70 vehicles per kilometre during congestion.

Weighting

From Figure 2a, but mainly after convolution, i.e. from Figure 2b, we see that some of the diagonals contain substantially more high scores than others. This is due to the fact, that some vehicles at the upstream station have an individual length that has more possible counterparts at the downstream station. Thus there are more possible matches and therefore more red to light blue cells on the diagonal. The same holds for vehicles at the downstream station (see vertical lines in Figure 2b): Some vehicles at the downstream station have an individual length, which has more possible counterparts at the upstream station.

Since especially the diagonals with high scores can disturb the subsequent filtering step, we weight them leading to matrix **T'** as follows (see equation (5)): First, the mean of the diagonal elements is computed, i.e. the sum of all diagonal elements is calculated and divided by the number of elements on the diagonal. Since for some diagonals this can lead to very low sums (or even zero), a constant γ is added to prevent the weights from becoming infinite. A value of γ= 0.05 has led to good results. However, the estimation results are not very sensitive on γ. The second term on the r.h.s. of equation (5) is the weight for the diagonal that includes element $T(x,y)$. Subsequently each element $T(x,y)$ is multiplied by its appropriate weight, which leads to $T'(x,y)$. Formally this can be written as

$$T'(x,y) = T(x,y) \cdot \left(\gamma + \frac{\sum_{x'=x_{\min}}^{x_{\max}} T\left(x', x' - x_{\min} + y_{\min}\right)}{x_{\max} - x_{\min} + 1} \right)^{-1} \tag{5}$$

where $x_{\min} = x_{\min}(x,y) = \max[1, x-(y-y_{\min})]$ and $x_{\max} = x_{\max}(x,y) = \min[X, x+(Y-y)]$ denote the minimal and maximal value considered in x-direction, respectively, X and Y represent the number of cells in x- and y-direction, respectively (see also explanations on the computation of score s_{ij}), $y_{\min} = \text{ceil}(\tau_{\min} / \Delta t)$ denotes the number of cells in y-direction that represents the minimum travel time with $\tau_{\min}$ being the minimal travel time for the considered origin-destination pair. The resulting image, i.e. matrix **T'** (again of size $X \times Y$) is shown in Figure 2c.

Resizing

Normally matrix **T'** comprehends a large number of columns that corresponds to *X* time bins. With the aim to enable an efficient computation of the subsequent steps, we need to reduce the number of columns. This is done as follows: We consider matrix **T'** as an image and resize it by bicubic interpolation to image **G** with dimension *M* (columns) times *N* (rows), *while preserving the image information*. For further details on the implementation of the bicubic interpolation we refer to Press et al. (2007).

2.2.3 Filtering

The goal of the filtering step is to extract structures of relevant size from image **G** (see Figure 2c). Furthermore, it is also an important step as it provides a better basis, i.e. a slightly smoothed image, for the subsequent step that determines the path with maximum costs, which is key in the estimation of the travel times. The filtering procedure includes three steps and is implemented as follows:

First, we apply a two-dimensional Discrete Fourier Transform (2D-DFT) on image **G**, which can be written as

$$D(m,n) = \frac{1}{\sqrt{MN}} \sum_{p=0}^{M-1} \sum_{q=0}^{N-1} G(p,q) \exp\left[-i2\pi \left(\frac{mp}{M} + \frac{nq}{N} \right) \right] \tag{6}$$

where *M* (width) and *N* (height) are the dimensions of the image after the resizing (see above) and **D** represents the image in the frequency domain, with the image coordinates $p \in \{1,\ldots,M\}$ and $q \in \{1,\ldots,N\}$ and the spectral coordinates $m \in \{1,\ldots,M\}$ and $n \in \{1,\ldots,N\}$.
Once we have transformed the image into the frequency domain we perform a pointwise multiplication with the filter function H_{BP}, which leads to the filtered image $D'(m,n)$ in the frequency domain. This is done as follows:

$$D'(m,n) = D(m,n) \cdot H_{\mathrm{BP}}(m,n) \tag{7}$$

The filter function H_{BP} that we use is a second-order two-dimensional Butterworth band-pass, composed of the difference between the two first-order low-pass filters. It is defined as:

$$\begin{aligned} H_{\mathrm{BP}}(m,n) &= H_{\mathrm{BP}}\left(\omega(m,n); p, \omega_1, \omega_2\right) \\ &= H_{\mathrm{LP}}^{(2)}\left(\omega(m,n); p, \omega_2\right) - H_{\mathrm{LP}}^{(1)}\left(\omega(m,n); p, \omega_1\right) \\ &= \frac{1}{1+\left(\omega(m,n)/\omega_2\right)^{2p}} - \frac{1}{1+\left(\omega(m,n)/\omega_1\right)^{2p}} \end{aligned} \tag{8}$$

where the frequency $\omega(m,n)$ is defined by

$$\omega(m,n) = \sqrt{\left(\frac{2m}{M}\right)^2 + \left(\frac{2n}{N}\right)^2} \tag{9}$$

with $-\lfloor M/2 \rfloor \le m \le \lfloor (M-1)/2 \rfloor$ and $-\lfloor N/2 \rfloor \le n \le \lfloor (N-1)/2 \rfloor$.

The three parameters used in equation (8), p, ω_1 and ω_2 have the following meaning (see also Figure 3 below): $2 \cdot p$ is an exponent, e.g. $(\omega(m,n)/\omega_2)^{2p}$, and in this case determines the slope of the transfer function in the "non-passing range", i.e. above around three times the cut-off frequency. In our application, we have set $p = 1$, accounting for the constraint $p \ge 1$. ω_1 and ω_2 are, as already mentioned, the cut-off frequencies of the two low-pass filters, i.e. the frequencies at which the amplitude of

the transfer function is half of its maximum. We determined ω_1 = 1e-4 and ω_2, = 1.5e-2 as optimal for our application. For the cut-off frequencies the following constraints must always be met: $0 < \omega_1 < \omega_2 < 0.5$.

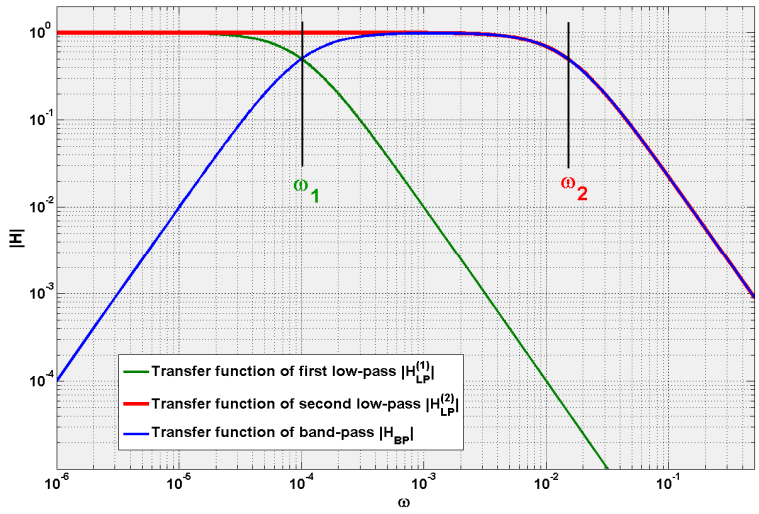


Figure 3. Transfer functions of the two low-pass filters and the band-pass filter.
The cut-off frequencies ω_1 and ω_2 define the frequency range of the band-pass.

In the last step, we transform the filtered image $D'(m,n)$ back to the spatial domain by applying the two-dimensional Inverse Discrete Fourier Transform (2D-IDFT), which leads to image **G'**:

$$G'(p,q) = \frac{1}{\sqrt{MN}} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} D'(m,n) \exp\left[i2\pi \left(\frac{mp}{M} + \frac{nq}{N} \right) \right] \tag{10}$$

Good general introductions on image processing and specifically on filtering procedures can be found in Burger and Burge (2006) or in Pitas (2000). Detailed information on the Discrete Fourier Transform and its applications can be found in Burger and Burge (2006), pp. 331 to 354.

2.2.4 Determination of the path with maximum costs

Although the travel time development is already well visible in Figure 2d (see red to light blue areas), we require an additional step to *automatically* find the path that connects the relevant areas (indicated by red to light blue colours) and forms an estimate for the travel time. Hence, we implemented an algorithm that searches for the path with the maximum costs, within a given image. The associated pseudo code is shown in Figure 4 below. Therein ϕ_1 and ϕ_2 determine the maximum meaningful change of travel time in positive and negative direction, respectively. M and N denote the width and height of the image and matrix **P** comprehends the travel time estimation.

Initialisation
Read image **G'** (with dimension $M \times N$)
Declare matrices **C** (cost) and **P** (path) with dimension $M \times N$ and set all cells to zero

Forward processing (left to right) → Determine cost matrix
Copy first column of **G'** to first column of **C**
for columns $p = 2$ with step size 1 to M
  for rows $q = 1$ with step size 1 to N
    1. Determine the additional costs **d** from the feasible positions (determined by ϕ_1 and ϕ_2) in the image at column p–1: **d** = **C** [p–1, max(1, $q - \phi_1$) to min(N, $q + \phi_2$)]
    2. Add value to cost matrix at position: $C(p, q) = \max(\mathbf{d}) + G'(p, q)$
  end
end

Backward processing (right to left) → Find path
Determine starting cell at the right boundary of the image
for columns $p = M$–1 with step size –1 to 1
  1. Determine feasible cells for step from position p+1 back to p, again within the range (determined again by ϕ_1 and ϕ_2) at column p–1: **f** = **C** [p, max(1, $q - \phi_1$) to min(N, $q + \phi_2$)]
  2. Determine position q with maximum value for **f**, i.e., the cell in **f** with the maximum costs for getting from p+1 to p
  3. Given p and q, set $P(p, q) = 1$
end

*Figure 4. Pseudo code of the procedure to find the path with maximum costs in image **G'**.*

3. Application

The examined network corresponds to an existing part of the Swiss National freeway network, from which single car data are available. It was implemented with the micro simulation tool AIMSUN NG (v5.1.7) and has dimensions according to Figures 6a/7a. To feed the simulation with a vehicle mixture that is close to reality, distributions of the lengths at real detector stations were analysed (see Figure 5). From this, we derived vehicle classes together with appropriate properties (e.g. maximum speed, vehicle length, driving characteristics of the driver (reaction time) and the vehicle (maximum acceleration)) as part of the simulation input.

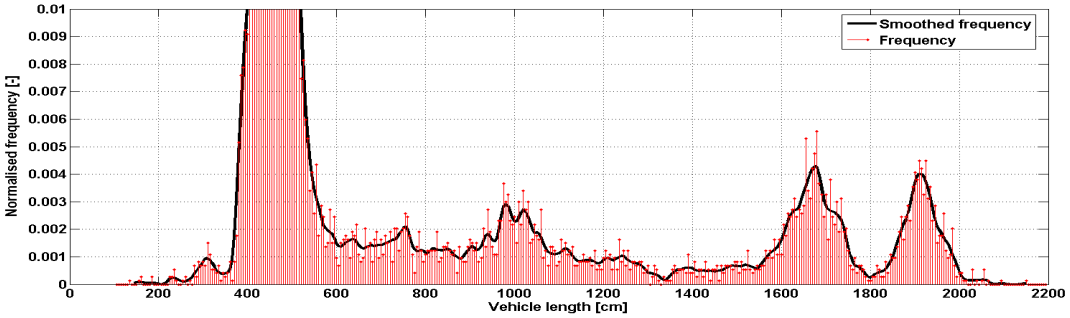


Figure 5. Normalised frequency of real vehicle lengths detected at a station on the Swiss National freeway network at a typical weekday.
This formed the basis for the definition of the implemented vehicle types and numbers. The maximum frequency is 0.03 at a length of around 450 cm.

Many freeway stretches are equipped with double inductive loop detectors. This holds also for the Swiss National freeway network: The detectors used are mainly 'Marksman M660', manufactured by Golden River Ltd. According to Taxomex (1999) and Rubin (2007), the accuracy of the length measurements is $\Delta L < \pm$ 20 cm. To investigate the influence of the random

measurement error on the travel time estimation, we performed simulations with measurement errors of ± 20 cm, ± 30 cm, and ± 40 cm, respectively. In addition, simulations were performed for five different values of structural errors (bias) in the length measurement: -20 cm, -10 cm, 0 cm, +10 cm, and +20 cm.

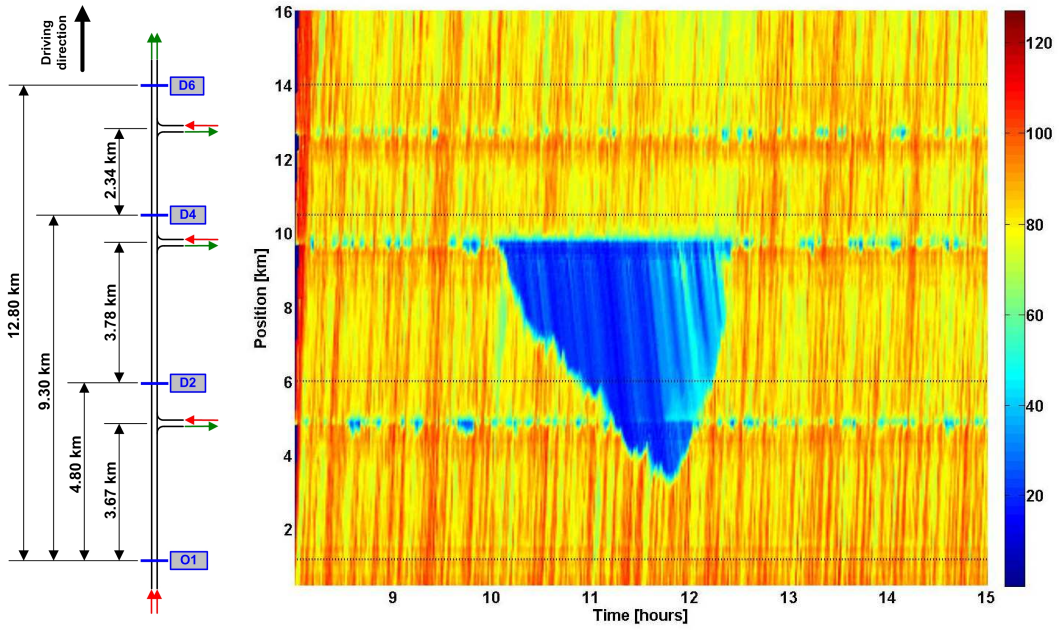
To simulate the random length measurement errors that are experienced in practice, for each vehicle generated by the AIMSUN simulation environment, a random error ΔL was added to its true length, independently at the upstream and downstream detector station, respectively. Based on the assumptions made above regarding the accuracy, the simulated errors are i.i.d., uniformly distributed between ± ΔL. To additionally consider the structural errors, values for the bias within the range defined above were added to each downstream vehicle length L_j.

4. Investigated scenarios

To assess the presented method, we investigated two different scenarios with both of them having congestion situations, caused either by demand overload or by an accident.
A third scenario, which we considered, was a test scenario, which shows that the algorithm also performs well under normal conditions, i.e. with low demand and neither congestion nor incident. This is important as it represents a standard operating case. However, due to the limited space, we do not elucidate this here.

4.1 Scenario 1 – Congestion due to overload

The first scenario has high volumes flowing into the system at origin O1. The mean flow over both lanes was around 1500 veh/hour/lane together with densities around 15 to 20 veh/km/lane. At 10:00 a.m., the flow at the second on-ramp (position 8.60 km; see Figure 6a) was increased during 60 minutes (see Figure 6b). This had the effect, that the flow exceeded the maximum capacity of the network around this position, which immediately led to congestion. Although the higher inflow lasted only for one hour, it has taken the system almost 2½ hours to operate normally again. Figure 6b shows the time-space view of the average speed over both lanes. We see, that the average speed is around 80 to 90 kilometres per hour in normal condition and breaks down to around 20 kilometres per hour during congestion. We also see, that around the on- and off-ramps, velocities are slightly lower than before and after (horizontal patterns around the positions of the on- and off-ramps). This is caused by the fact, that even in normal operation the inflows are relatively high, which in some cases leads to small temporal bottlenecks.

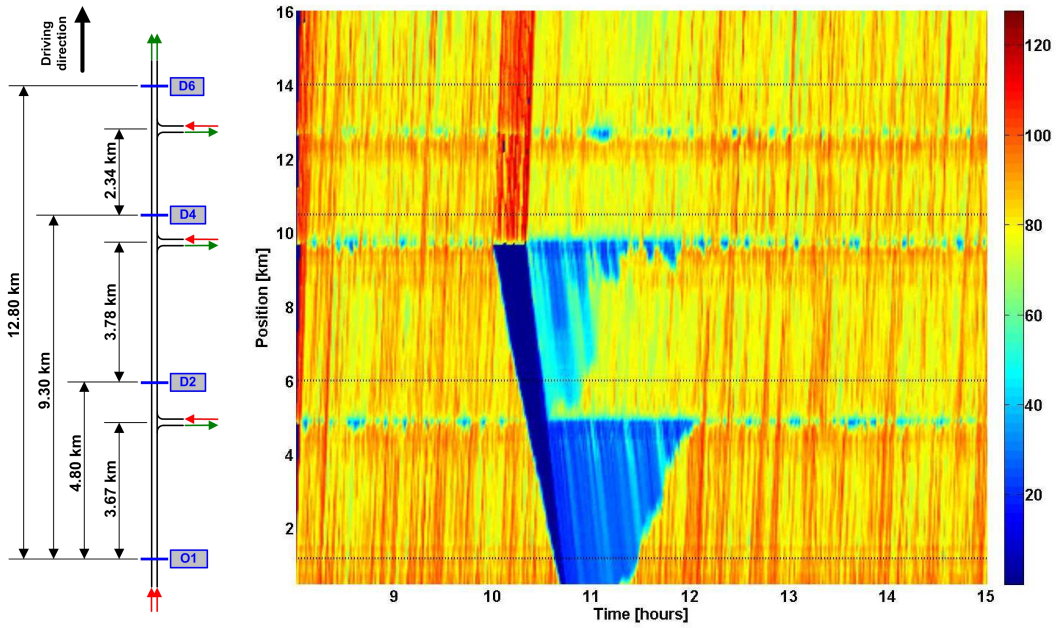


(a) *(b)*

Figure 6. (a) The examined network displays the detector positions (origin O1; destinations: D2, D4 and D6) together with the relevant distances. (b) Time-space view of the average speed (over both lanes) for scenario 1 and the whole simulation period.
The spatial resolution is 50 meters, the temporal resolution is 1 minute. The colour bar on the right side indicates the scale for the average speed in kilometres per hour.

4.2 Scenario 2 – Congestion due to an incident

Scenario 2 has the same high volumes flowing into the system at origin O1 as in scenario 1. The average flow over both lanes was again around 1500 veh/hour/lane together with densities around 15 to 20 veh/km/lane. Unlike scenario 1, we do not have a congestion due to a network overload but one caused by an incident. At 10:00 a.m., an accident happened at position 8.62 km. i.e. right after the second on-ramp (see Figure 7a). The accident caused a full blocking of both lanes for 20 minutes. In Figure 7b this is represented by the dark blue area starting its propagation at 10:00 a.m. in the opposite direction of the vehicle flow. Within this area the speed is zero, i.e. full congestion.
Although the two lanes are reopened after 20 minutes, due to the high volumes even in normal mode, the back-propagation of the congestion queue causes a secondary congestion around the first on-ramp (at position 3.70 km). In Figure 7b we see, that this "second congestion" has also a very high impact. Although the incident lasted for twenty minutes only, it has taken the system almost 2 hours to return to normal conditions. We see again the horizontal patterns around the positions of the on- and off-ramps, caused in some cases by small temporal bottlenecks (see comments on Figure 6).



(a) *(b)*

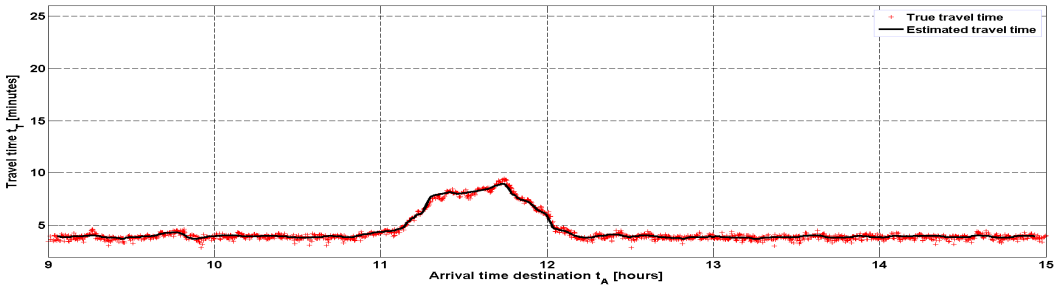
Figure 7. (a) The network under investigation displays the detector positions (origin O1; destinations: D2, D4 and D6) together with the relevant distances. (b) Time-space view of the average speed (over both lanes) for scenario 2 and the whole simulation period.
The spatial resolution is 50 meters, the temporal resolution is 1 minute. The colour bar on the right side indicates the scale for the average speed in kilometres per hour.

5. Estimation results

5.1 Qualitative assessment

In this section, a first assessment of the quality of the estimation results is done by comparing the estimated graphs visually with the true travel times for the two scenarios and for three OD-combinations considered. The measurement noise is assumed to be i.i.d. and uniformly distributed within $\Delta L = \pm 20$ cm (see also section 3). For these investigations it was assumed that no additional structural errors exist. The travel time estimates were performed here for the right lane and for vehicle lengths ≥ 6 m only.

(a)



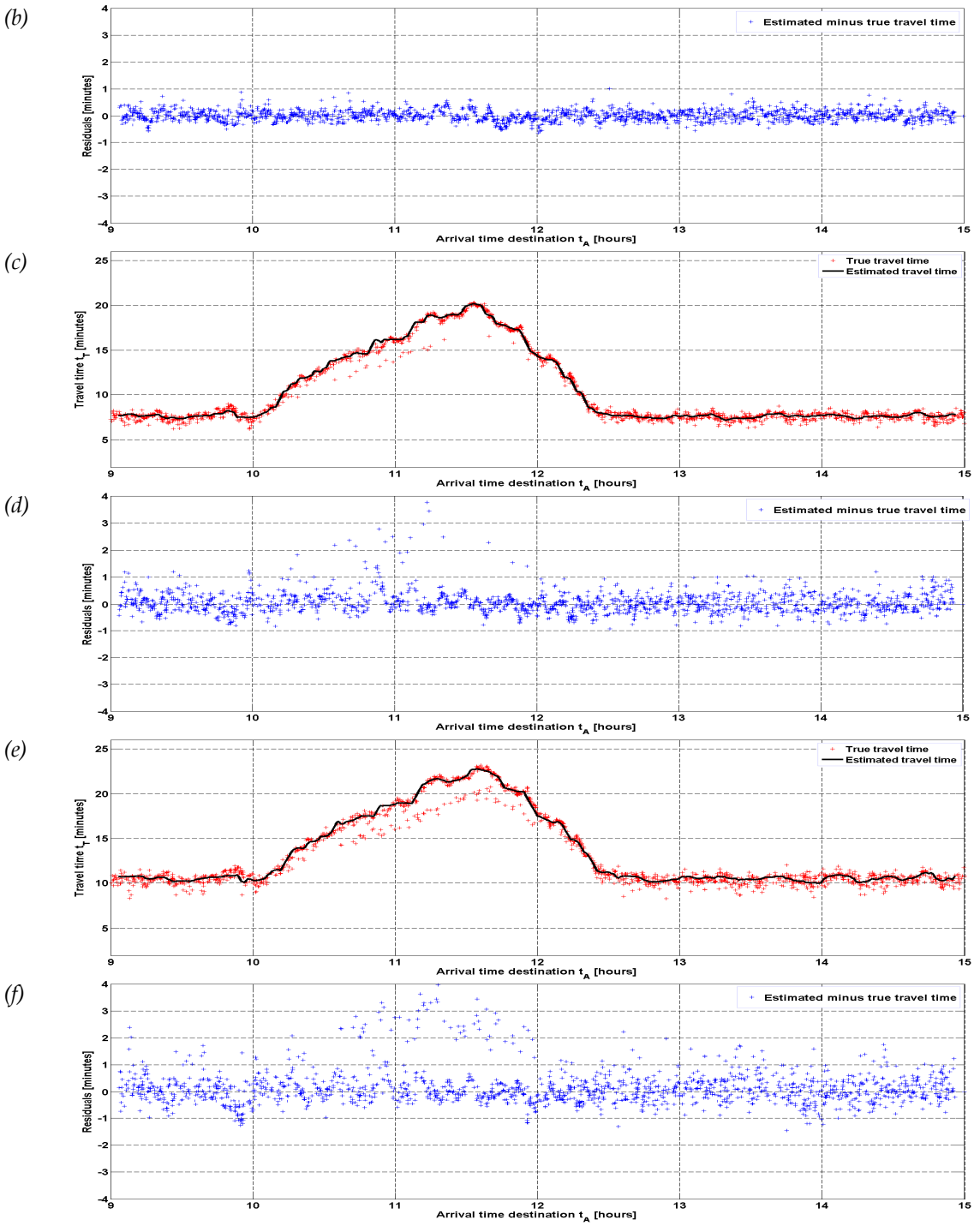
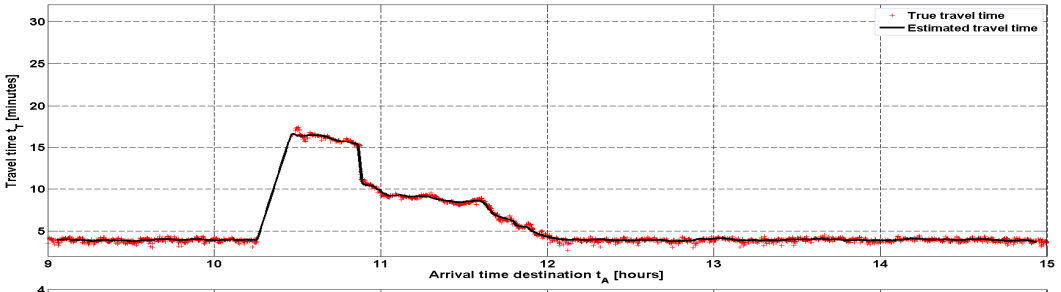


Figure 8. Travel time estimates and residuals for scenario 1: (a) origin O1 to destination D2 (4.8 km), (b) residual of estimates in Figure 8a, (c) origin O1 to destination D4 (9.3 km), (d) residuals of estimates in Figure 8c, (e) origin O1 to destination D6 (12.8 km), (f) residuals of estimates in Figure 8e.
In Figures 8a, 8c, and 8e the red crosses indicate true travel times of individual vehicles and the black line represents the estimated travel times as a function of the downstream arrival time. In Figures 8b, 8d, and 8f the blue crosses indicate the residuals of the travel times, i.e. estimated minus true travel times.

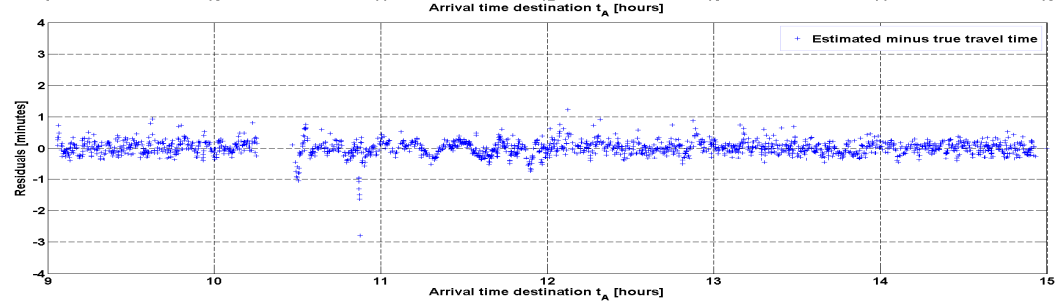
In Figure 8 we see, that for all three OD-combinations the estimated travel time (black line) matches the true travel times (red crosses) well, almost independently of the distance between

origin and destination. As can be seen, especially in Figures 8c/d and 8e/f, there are some travel times below/above the others. These are travel times of vehicles, which travelled for a substantial part of their journey between origin and destination on the left lane and therefore were a little faster than the vehicles travelling mainly on the right lane. Besides that we see from the residual plots (8b, 8d and 8f) that there is no systematic structural error (bias).

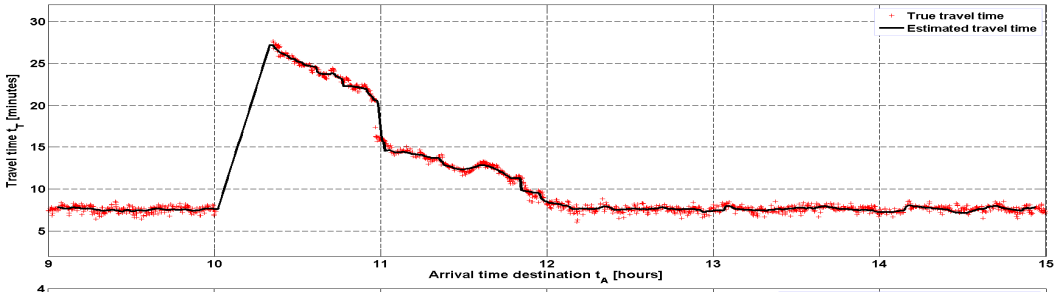
(a)



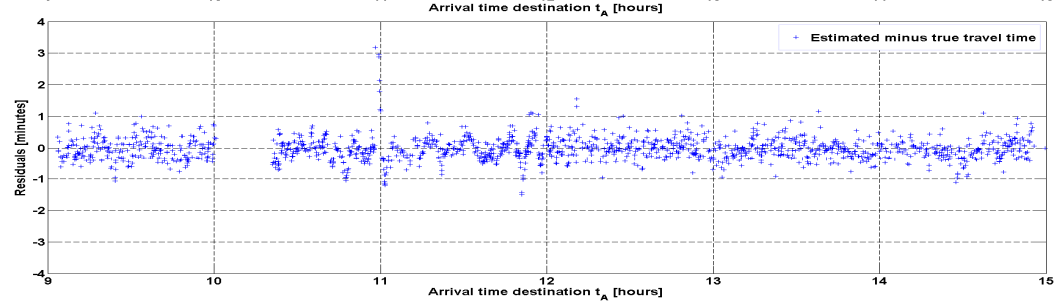
(b)



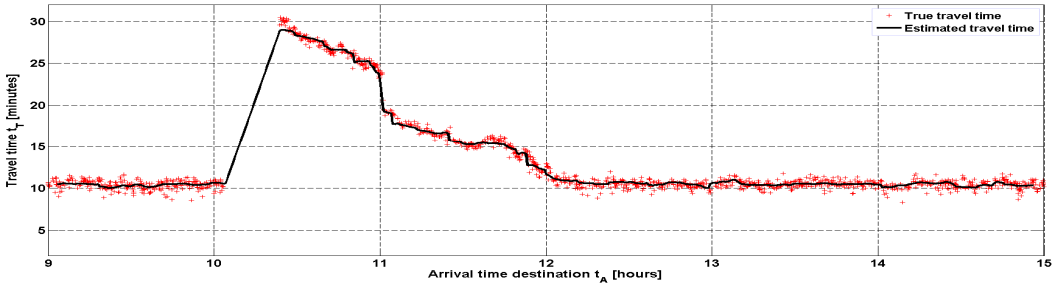
(c)



(d)



(e)



(f)

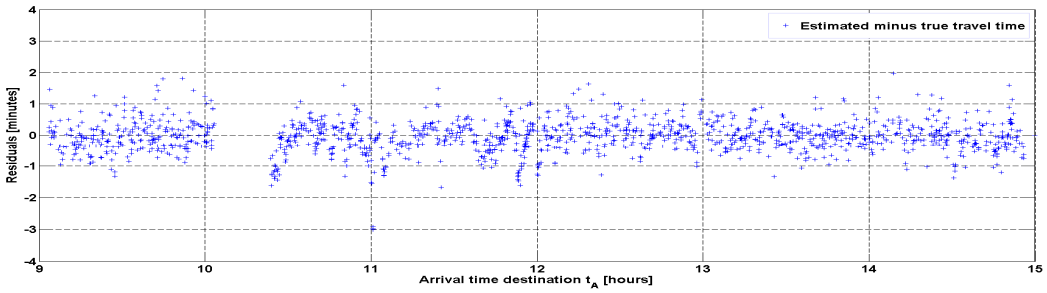


Figure 9. Travel time estimates and residuals for scenario 2: (a) origin O1 to destination D2 (4.8 km), (b) residual of estimates in Figure 9a, (c) origin O1 to destination D4 (9.3 km), (d) residuals of estimates in Figure 9c, (e) origin O1 to destination D6 (12.8 km), (f) residuals of estimates in Figure 9e.
In Figures 9a, 9c, and 9e the red crosses indicate true travel times of individual vehicles and the black line represents the estimated travel times as a function of the downstream arrival time. In Figures 9b, 9d, and 9f the blue crosses indicate the residuals of the travel times, i.e. estimated minus true travel times.

In Figure 9 we see again, that for the three considered OD-combinations the estimated travel time (black line) matches the true travel times (red crosses) quite well, almost independently of the distance between origin and destination. Due to some fast travel time changes (at arrival times between 10 and around 11 a.m.), we observe some slightly larger differences as in the other areas. Besides that we see from the residual plots (9b, 9d and 9f) that there is no systematic structural error (bias).

We can conclude that the above assessments of both scenarios lead to very satisfying results.

5.2 Quantitative assessment

In addition to the investigations of the plots discussed in the previous section, we now carry out some quantitative assessments of the scenarios. For this we considered again the same three OD pairs as in the previous section. In a first step, we analysed the dependency of the 'Mean Absolute Percentage Error' (MAPE) on random measurement errors. In a second step, we investigated the effects of some structural errors (bias) of the lengths. For each case we performed R = 30 simulation runs to get statistically meaningful values for the mean of the MAPE. The MAPE is defined as follows:

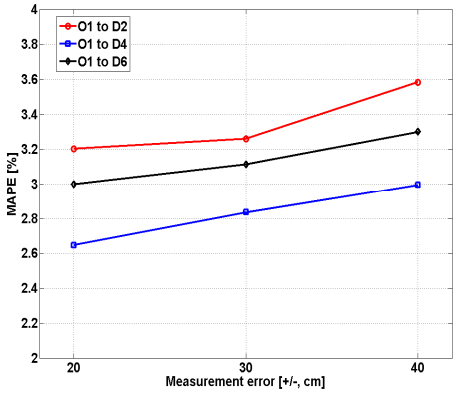
$$MAPE_{od}^{(r)} = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|\tau_{od}^{(r,i)} - \hat{\tau}_{od}^{(r,i)}\right|}{\tau_{od}^{(r,i)}} \tag{11}$$

where N is the number of individual travel times available for vehicles $i = 1,\ldots,N$, $\hat{\tau}_{od}^{(r,i)}$ and $\tau_{od}^{(r,i)}$ denote the estimated and true travel time, respectively, for origin o to destination d, simulation run $r \in \{1,\ldots,R\}$, and vehicle i. The computation of the mean of the MAPE from the R simulation runs is straightforward and hence is not explained in detail here.

5.2.1 Scenario 1

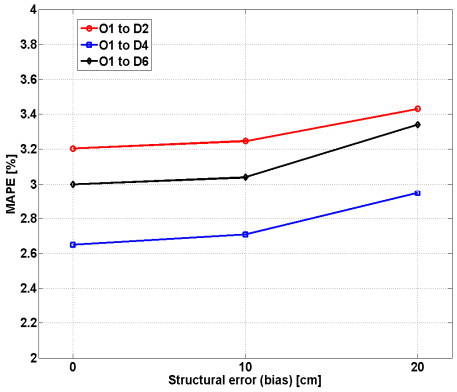
The results (MAPE) for random measurement errors (with no bias) for scenario 1 are graphically represented by Figure 10a. The MAPE for different structural errors (with a random error of ΔL = ± 20 cm) is shown in figure 10b.

Random error influences



(a)

Structural error influences



(b)

Figure 10. (a) Mean of the MAPE as a function of the measurement error for the three investigated OD-combinations, (b) Mean of the MAPE as a function of the structural error for the three investigated OD-combinations.

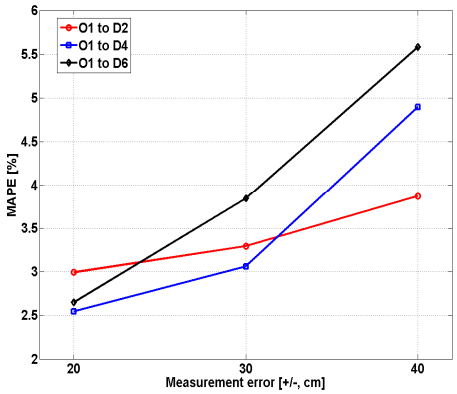
From Figures 10a and 10b we can conclude that the travel time estimates for scenario 1 are for all investigated cases very satisfying (MAPE well below 4%). This holds for the dependency on the random measurement errors as well as for the dependency on the structural errors.

As we see from Figure 10a, the MAPE varies with the distance between origin and destination, e.g. it is higher for origin O1 to destination D2 than for origin O1 to destination D4. This is due to the fact that the different origin-destination pairs face quite different traffic conditions during the time period considered (see figure 6b), which has a strong influence on the travel time and thus on the MAPE (computed according to equation (11)).

5.2.2 Scenario 2

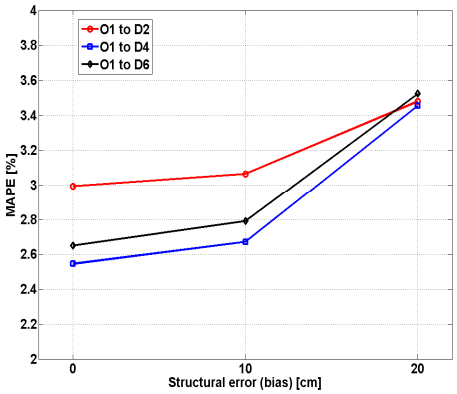
The results (MAPE) for random measurement errors (with no bias) for scenario 2 are graphically represented by Figure 11a. The MAPE for different structural errors (with a random error of ΔL = ± 20 cm) is shown in Figure 11b.

Random error influences



(a)

Structural error influences



(b)

Figure 11. (a) Mean of the MAPE as a function of the measurement error for the three investigated OD-combinations, (b) Mean of the MAPE as a function of the structural error for the three investigated OD-combinations.

From Figures 11a and 11b we see that the travel time estimates for scenario 1 are for most cases satisfying. However, in Figure 11a we see, that there are in general larger MAPE for measurement errors ≥ ± 30 cm. In addition, the dependency of the MAPE on the structural errors is slightly higher than for scenario 1 but still well acceptable.

As we see from Figure 11a, the MAPE also varies with the distance between origin and destination as it was the case for scenario 1 (see Figure 10a). However, due to the fact that for scenario 2 the traffic conditions during the time period considered (see Figure 7b) differ substantially from those in scenario 1, the dependencies of the MAPE on the OD-distance are also different from those for scenario 1.

5.2.3 Summary

As one could expect, larger random or structural errors lead to larger travel time estimation errors (MAPE). At the current stage for random measurement errors up to ± 20 cm together with structural errors up to ± 10 cm one can expect very good estimation results for both investigated scenarios, independent of the OD-combination. If either the random measurement error or the structural error exceeds the above values, the MAPE usually deviates from its optimal values (see Figures 10 and 11).

6. Conclusions and future work

In this paper we presented a novel travel time estimation method, which relies only on vehicle lengths (e.g. from double inductance loops) and the corresponding time stamps at the detector stations. We performed tests with simulated single vehicle data, for origin-destination distances between 4.8 and 12.8 kilometres. Furthermore, depending on the OD-combination, up to three unobserved on- and off-ramps each were part of network under investigation. Only vehicles on the right most lane and with lengths ≥ 6 m were considered for the procedure.

We could show that for the investigated origin-destination combinations all relevant characteristics of the travel times were captured. For practically meaningful values of random measurement errors and structural errors, the travel time estimates were well within satisfying error boundaries.

Besides these results, some practical benefits of the method are: (i) Provided that single car data are available at the detector stations with sufficient accuracy, no additional infrastructure investment is required; (ii) The method works fully anonymously, i.e. no privacy issues arise; (iii) extensions to use more sophisticated detection technologies that provide additional vehicle features, are straightforward; and (iv) the computational speed (the travel time estimation for 7 hours takes about 5 seconds) allows for an efficient processing of historical data and as such, can provide valuable additional input data for subsequent algorithms.

As mentioned already in section 1, we focussed in this paper on the estimation of mean arrival travel times. Thus, one of the next steps will be the extension of the procedure such that it allows for the estimation of mean departure travel times, which forms a basis for travel time prediction algorithms.

Besides that, the next steps will also include the application of the method on real data sets and various distances together with some adjustments/optimisations of the algorithm.

In addition, we will test the applicability of the method on other than the right most lane. First attempts for the left lane at distances of 4.8 km were very promising, even though there are significantly less longer vehicles available. Furthermore, we plan to perform tests with larger random errors for short distances with the intention, that even detectors with lower sampling

rates could be used (down to 60 Hz). As a final step it is intended to integrate the method to field applications.

Acknowledgements

We would like to thank the IDP for financial support of this project as well as the CTI (Swiss Innovation Promotion Agency) and Siemens Schweiz AG for financial support of a preceding project (Grant No. KTI 6949.2 FHS-ES). Finally, we would like to thank the two anonymous referees for providing very useful suggestions, which helped to improve the quality of this paper.

References

Abdulhai, B. and Tabib, S.M. (2003). Spatio-temporal inductance-pattern recognition for vehicle re-identification. *Transportation Research Part C*, vol. 11, no. 3-4, pp. 223-239.

Anagnostopoulos, C.N., Anagnostopoulos, I.E., Loumos, V. and Kayfas, E. (2006). A license plate-recognition algorithm for intelligent transportation systems. *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 3, pp. 377-392.

Bar-Gera, H. (2007). Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: A case study from Israel. *Transportation Research Part C*, vol. 15, no. 6, pp. 380-391.

Bovy, P.H.L. and Thijs, R. (2000). *Estimators of travel time for road networks*. Delft University Press, Delft, The Netherlands.

Brockfeld, E., Lorkowski, S., Mieth, P. and Wagner, P. (2007). Benefits and limits of recent floating car technology – An evaluation study. Paper presented at the *11th WCTR Conference*, June 24-28, 2007, Berkeley, USA.

Burger, W. and Burge, M.J. (2006). *Digitale Bildverarbeitung – Eine Einführung mit Java und ImageJ*. Springer, Berlin und Heidelberg.

Cayford, R. and Yim, Y.B. (2006). A field operation test using anonymous cell phone tracking for generating traffic information. *Transportation Research Board 2006 annual meeting*, CD-ROM, paper number 06-2511.

c.c.com (2006). BLIDS - Bluetooth-based traffic data collection system. Information from project. available at: http://www.cccom.at/blids/ (assessed February 2008).

Coifman, B. (1998a). A new algorithm for vehicle reidentification and travel time measurement on freeways. *Proceedings of the 5th International Applications of Advanced Technologies in Transportation Engineering*, 1998, ASCE, pp. 167-174.

Coifman, B. (1998b). Vehicle reidentification and travel time measurement in real-time on freeways using the existing loop detector infrastructure. *Transportation Research Record*, no. 1643, pp. 181-191.

Coifman, B. (2002). Estimating travel times and vehicle trajectories on freeways using dual loop detectors. *Transportation Research Part A*, vol. 36, no. 4, pp. 351-364.

Coifman, B. and Cassidy, M. (2002). Vehicle reidentification and travel time measurement on congested freeways. *Transportation Research Part A*, vol. 36, no. 10, pp. 899-917.

Coifman, B. and Ergueta, E. (2003). Improved vehicle reidentification and travel time measurement on congested freeways. *Journal of Transportation Engineering*, vol. 129, no. 5, pp. 475-483.

Coifman, B., Krishnamurthy, S. (2007). Vehicle reidentification and travel time measurement across freeway junctions using the existing detector infrastructure. *Transportation Research Part C*, vol. 15, no. 3, pp. 135-153.

Coifman, B., McCord, M., Mishalani, R.G., Iswalt, M. and Ji, Y. (2006). Roadway traffic monitoring from an unmanned aerial vehicle. *IEE Proceedings, Intelligent Transport Systems*, vol. 153, no. 1, pp. 11-20.

Dion, F. and Rakha, H. (2006). Estimating dynamic roadway travel times using automatic vehicle identification data for low sampling rates. *Transportation Research Part B*, vol. 40, no. 9, pp. 745-766.

Haoui, A., Kavaler, R. and Varaiya, P. (2008). Wireless magnetic sensors for traffic surveillance. *Transportation Research Part C*, vol. 16, no. 3, pp. 294-306.

Helbing, D. (1997). *Verkehrsdynamik – Neue physikalische Modellierungskonzepte*. Springer, Berlin.

Helbing, D. (2001). Traffic and related self-driven many-particle systems. *Reviews of Modern Physics*, vol. 73, pp. 1067-1141.

Helbing, D., Hennecke, A., Shvetsov, V. and Treiber, M. (2002). Micro- and macro-simulation of freeway traffic. *Mathematical and Computer Modelling*, vol. 35, no. 5-6, pp. 517-547.

Kühne, R.D. and Immes, S. (1993). Freeway control systems for using section-related traffic variable detection. *Proceedings of the Pacific Rim TransTech Conference*, July 25-28, 1993, Seattle, Washington, USA, pp. 56-62.

Kwon, T.M. (2006). Blind deconvolution processing of loop inductance signals for vehicle reidentification. *Proceedings of the 85th Annual Meeting of the Transportation Research Board*, January 22-26, 2006.

Li, Y. and McDonald, M. (2002). Link travel time estimation using single GPS equipped probe vehicle. *IEEE 5th Int. Conf. On Intelligent Transport Systems*, Singapore, Sept. 2002.

Lindveld, C.D.R., Thijs, R., Bovy, P.H.L. and Van der Zijpp, N.J. (2000). Evaluation of online travel time estimators and predictors. *Transportation Research Record*, no. 1719, pp. 45-53.

Michalopoulos, P. and Hourdakis, J. (2001). Review of non-intrusive advanced sensor devices for advanced traffic management systems and recent advances in video detection" *Proceedings of the Institution of Mechanical Engineers. Part I, Journal of Systems and Control Engineering*. vol. 215, no. 4. pp. 345-355.

Nam, D.N. and Drew, D.R. (1996). Traffic dynamics: Method for estimating freeway travel times in real time from flow measurements. *Journal of Transportation Engineering*, vol. 122, no. 3, pp. 185-191.

Neubert, L., Santen, L., Schadschneider, A. and Schreckenberg, M. (1999). Single-vehicle data of highway traffic - A statistical analysis. *Physical Review E*, vol. 60, pp. 6480.

Oh, J.-S., Jayakrishnan, R. and Recker, W. (2002). Section travel time estimation from point detector data. *Working Paper UCI-ITS-TS-WP-02-14*, Institute of Transportation Studies, University of California, Irvine.

Pathirana, P.N., Savkin, A.V., Bulusu, N. and Plunkett, T. (2006). Speed control and policing in a cellular mobile network: SpeedNet. *Computer Communications*, vol. 29, no. 17, pp. 3633-3646.

Pfannerstill, E. (1984). A pattern recognition system for the re-identification of motor vehicles. *Proceedings of the 7th IEEE International Conference on Pattern Recognition*, Montreal, New Jersey, pp. 553–555.

Pitas, I. (2000). *Digital Image Processing Algorithms and Applications*. John Wiley & Sons, New York.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (2007). *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press.

Quiroga, C.A. and Bullock, D. (1998). Travel time studies with global positioning and geographic information systems: an integrated methodology. *Transportation Research Part C*, vol. 6, no. 1-2, pp. 101-127.

Reinartz, P., Lachaise, M., Schmeer, E., Krauss, T. and Runge, H. (2006). Traffic monitoring with serial images from airborne cameras. *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 61, no. 3-4, pp. 149-158.

Rubin, M. (2007). Swiss Federal Roads Office FEDRO. Personal communication.

Taxomex (1999). *Datenblatt Schlaufenkarte zu Marksman 660*. Taxomex S.A., Nyon, Switzerland.

Toth, C.K. and Grejner-Brzezinska, D. (2006). Extracting dynamic spatial data from airborne imaging sensors to support traffic flow estimation. *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 61, no. 3-4, pp. 137-148.

van Lint, J.W.C. (2004). *Reliable Travel Time Prediction for Freeways*. PhD Thesis. Delft University of Technology, Delft, The Netherlands.

Westerman, M. and Immers, L.H. (1992). A method for determining real-time travel times on motorways. *Proceedings of the 25th ISATA Conference*, Florence, Italy, pp. 221-228.

Yim, Y. (2003). The state of cellular probes. *Research Report UCB-ITS-PRR-2003-25*, Institute of Transportation Studies, University of California, Berkeley.

Yim, Y. and Cayford, R. (2001). Investigation of vehicles as probes using global positioning system and cellular phone tracking: Field operational test. *Report UCB-ITS-PWP-2001-9. California PATH Program*, Institute of Transportation Studies, University of California, Berkeley.

Yim, Y. and Cayford, R. (2006). Field operational test using anonymous cell phone tracking for generating traffic information. Proceedings of the *85th Annual Meeting of the Transportation Research Board*.