

Technical report 15-005

# **Estimation of the generalized average traffic speed based on microscopic measurements\***

A. Jamshidnejad and B. De Schutter

*If you want to cite this report, please use the following reference instead:*

A. Jamshidnejad and B. De Schutter, “Estimation of the generalized average traffic speed based on microscopic measurements,” *Transportmetrica A: Transport Science*, vol. 11, no. 6, pp. 525–546, 2015.

Delft Center for Systems and Control  
Delft University of Technology  
Mekelweg 2, 2628 CD Delft  
The Netherlands  
phone: +31-15-278.51.19 (secretary)  
fax: +31-15-278.66.79  
URL: <http://www.dsc.tudelft.nl>

---

\*This report can also be downloaded via [http://pub.deschutter.info/abs/15\\_005.html](http://pub.deschutter.info/abs/15_005.html)

# Estimation of the generalized average traffic speed based on microscopic measurements

A. Jamshidnejad<sup>a</sup> \* and B. De Schutter<sup>a</sup>

<sup>a</sup>*Delft Center for Systems and Control, Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands*

The average speed of vehicles plays an important role in traffic engineering. In almost any model-based traffic monitoring, analysis, or control application the average speed is required as a measure of performance or as an input for traffic models used to simulate fuel consumption, vehicle emissions, or traffic noise. The average speed is also used in algorithms that estimate the travel time. It also appears in the fundamental equation of traffic where density is calculated based on the average speed and flow. This article presents a new methodology for estimating the time-space-mean speed (TSMS), which is an equivalent for the *generalized speed* introduced by Edie (1963). To this aim, first tight upper and lower bounds are developed for the TSMS using individual vehicle speeds that are obtained via point measurements. To estimate the TSMS from these bounds, and to deal with the cases where the trajectories of the vehicles might not be straight lines, a convex combination of the upper and lower bounds is introduced. In order to assess the convex combination and to compare its performance with other formulas in literature, two real-life data sets corresponding to the NGSIM data for the I-880 highway in the San Francisco Bay Area, and the A13 data near Rotterdam-Delft are used. At the end, MATLAB simulations are presented to cover scenarios that are not included in the available real-life data sets. The results produced by the new formula, both for the real-life data sets and for MATLAB simulations, are found to be more accurate compared with other available formulas in literature.

**Keywords:** average traffic speed; time-space-mean speed; microscopic point measurements;

## 1. Introduction

The average speed of the vehicles plays an important role in both performance and control-related works on the traffic. The average speed is in fact required for almost any model-based traffic application. Most of the traffic simulation models use the average speed in accident analyses, in economic studies, as an indication of the service level on the road, or as an input for estimation of other performance indicators such as the fuel consumption, the vehicle emissions, the travel time, and the traffic noise (May 1990).

Moreover, the average speed is a fundamental variable in traffic. Indeed, the average speed together with the flow and the density are commonly called the *fundamental variables* of traffic, since a basic relationship exists between the flow and the density by means of the average speed which is known as the *fundamental relationship* or the *fundamental equation* of traffic (Daganzo 1997; Wardrop 1952).

In addition to the discussed importance of the average speed, the availability of a reliable value of the travel time is important for traffic engineers in applications such as

---

\*Corresponding author. Email: a.jamshidnejad@tudelft.nl

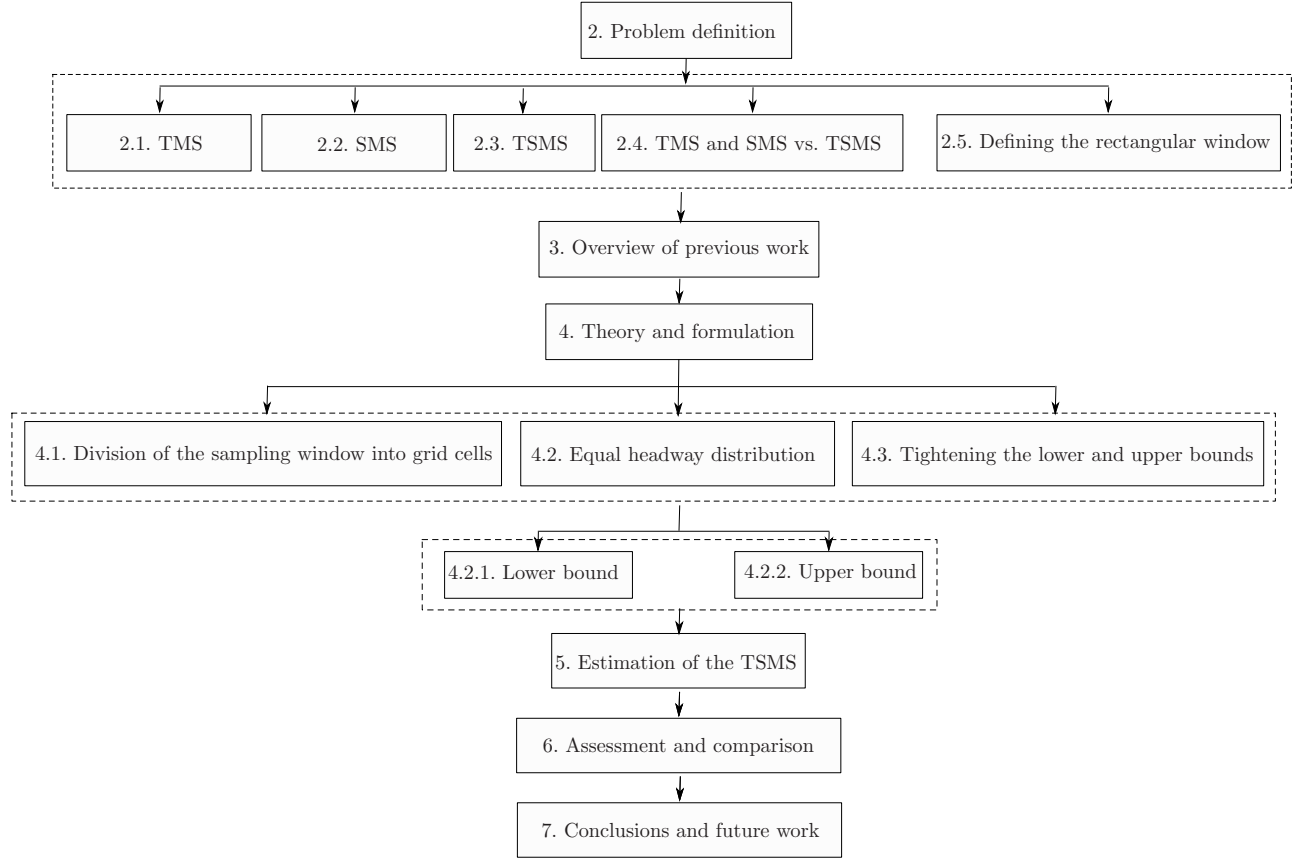


Figure 1. Road map of the paper

traffic signal coordination, in ‘before’ and ‘after’ studies of traffic, and also in estimation of other traffic states (Boyce and Xiong 2007; May 1990). Furthermore, travel time together with the average speed is used to identify and to assess operational problems along highways.

Two main methodologies are applied in order to measure or to estimate the travel time on a road: the direct measurement method and the indirect estimation method (Mori et al. 2015). In direct measurement, the total travel time between two predefined reference points is measured after all the vehicles have finished the distance between these two points. For the measurement, it is necessary to identify the time instant at which a particular vehicle appears at each of the two reference points. Therefore, the identifying technologies (being grouped under automatic vehicle identification systems) are implemented by means of license plates or toll tag IDs (Soriguera and Robusté 2011a).

The indirect travel time estimation is an alternative approach, where the fundamental traffic variables (i.e., flow, density, and average speed) are measured on a specific point of the road link and an algorithm is applied for travel time estimation. The travel time is considered to be the ratio of the length of the road’s section and the average speed (Soriguera and Robusté 2011a). Thus, for calculating the travel time, the average speed on the particular section of the road is needed (Soriguera and Robusté 2011a; Han et al. 2010; Wardrop 1952). Later on, in Section 2.3 we will argue that the time-space-mean speed (TSMS), which is an equivalent for the generalized speed proposed by Edie (1963), is the best averaging method for the speeds obtained from point measurements or any other data collecting method that represents the data as trajectories within the time-space plane.

Since their introduction in the early 1960's, inductive loop detectors have been widely used for the purpose of vehicle detection on roads and for point measurement of the speeds of vehicles. In fact, according to Klein et al. (2006) inductive loop detectors are the most popular and the most widely used forms of traffic detection systems. Both microscopic and macroscopic characteristics of the traffic flow can be obtained using inductive loop detectors.

Single-loop detectors have been deployed in order to provide information on the traffic flow (i.e., the number of vehicles passing a specific point on the road per unit time) and the lane occupancy (i.e., the fraction of the same observation time interval that the loop's detection zone is occupied by vehicles). Microscopic traffic flow characteristics including the time headway<sup>1</sup>, the vehicle occupancy time, and the spacing can be estimated using a single-loop detector (Ma et al. 2010). A double-loop detector consists of two single-loop detectors placed a few meters apart from one another. The main advantage of a double-loop detector over a single-loop detector is that it provides individual speed data (Wang and Nihan 2003).

The extensive use of inductive loop detectors in traffic systems all around the world and the relatively high costs of substituting them with modern detecting instruments have incited the development of efficient ways for estimating the fundamental traffic variables based on the information provided by these loop detectors. In this paper, we also propose a method for estimation of the time-space-mean speed using data from inductive loop detectors. Compared with the available formulas in the literature for estimation of the average traffic speed, the proposed method in this paper produces more accurate results, where the percentage of error produced by the proposed method in some cases could be up to 14% smaller than the least percentage of error produced by other formulas.

The main contribution of this paper is a new approach based on a microscopic point-of-view that produces a tight upper and a tight lower bound for the time-space-mean speed. We use a convex combination of these bounds to find an estimate of the time-space-mean speed. We then assess and compare the new formula with available formulas in literature using two real-life data sets, NGSIM for the I-880 highway in the San Francisco Bay Area, and the A13 data near Rotterdam-Delft. The rest of the paper is organized as follows: Section 2 describes the problem that is going to be discussed in this paper. In Section 3, an overview of the previous research on the problem is presented. Section 4 proposes a new method for finding a tight upper and a tight lower bound. The convex combination for estimating the time-space-mean speed from the bounds is given in Section 5. Finally, the results from real-life data sets and also from MATLAB simulations are presented and discussed in Section 6. Section 7 is allocated to the conclusions and topics for further research. A road map of the paper is given in Figure 1.

## 2. Problem definition

We will solve the problem of finding the appropriate average speed (which we will argue to be the time-space-mean speed (TSMS) in Section 2.3) giving a formula that includes microscopic data of point measurement type. Most previous research focuses on discovering a reliable way to approximate the average speed using a macroscopic point-of-view (where the aggregated data is used). An overview of the previous work in this area will be discussed in more detail in Section 3.

---

<sup>1</sup>The time between consecutive vehicle observations at a fixed location is usually called the time headway, and the distance separations between consecutive vehicles at a given instant is called the distance headway (May 1990) or alternatively the spacing if we follow the terminology used by Daganzo (1997).

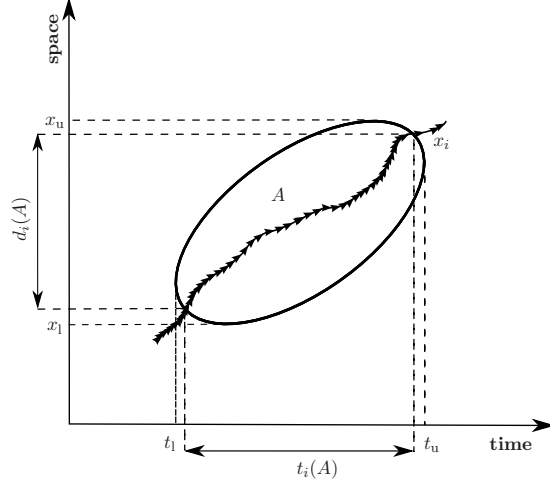


Figure 2. Representation of the section of the trajectory of vehicle  $i$  that is enclosed by the area  $A$  and is applied for the generalized definitions of the fundamental traffic variables

In this paper, we will propose a new solution for the problem which originates from a microscopic point-of-view by dividing the sampling window into smaller cells (see Section 4.1), and hence its results are expected to provide a higher accuracy and more detailed information for the aim of traffic control.

Many of the macroscopic and microscopic features of traffic including the average speed are defined from three different perspectives, i.e., across the time axis (called the *time variables*), across the space axis (the *space variables*), and in the time-space plane (the *time-space variables*). Next, we will present different definitions given for the average speed in literature.

### 2.1 Time-mean speed (TMS)

The *time-mean speed* (TMS) involves averaging across a time interval at a fixed location. A stationary observer such as a loop detector that has a sampling period of  $T_A$  and is placed at a fixed position  $x_{\text{loop},A}$  observes the vehicles during its sampling period,  $T_A$ , at the position  $x = x_{\text{loop},A}$ . Hence, from the given definition, the arithmetic mean of the speeds observed by the loop detector is the TMS.

### 2.2 Space-mean speed (SMS)

The SMS involves averaging across a stretch of the road at a specific instant of time. Therefore, for a traffic detection system such as a traffic camera that provides at a specific time instant a data set that covers a stretch of road of length  $L_A$ , the arithmetic mean of the speeds being observed in the photograph is the SMS.

### 2.3 Time-space-mean speed (TSMS)

In this section we will discuss the third average, which is called the time-space-mean speed or briefly the TSMS. The fundamental traffic equation relates two fundamental concepts (i.e., the flow and the density), one of which is defined across a time interval and the other one is defined across a length interval. Considering the ratio of these two quantities, we need an intermediary variable (with the same unit as the velocity) which is neither a local nor an instant variable, but a variable that is defined simultaneously through the time and the space axes. We call this variable the TSMS, which is equivalent

with what Edie (1963) calls the generalized speed.

According to the definitions given by Newell (1995) for the flow and density, which originate from a mathematical and physical perspective, the formulation given by Edie (1963) is the most appropriate average for traffic speed. Newell (1995) considers a mathematical interpretation for the traffic density. First a joint probability function  $F(x_1, x_2, \dots, x_N; t)$  is defined, which is the probability that at time instant  $t$  vehicle  $j$ , for all  $j \in \{1, 2, \dots, N\}$ , has a position less than  $x_j$ . Based on the joint probability, the marginal probability that a specific vehicle, say vehicle  $i$ , has a position less than  $x_i$  is given by

$$F_i(x_i; t) = F(L_A + \Delta, L_A + \Delta, \dots, x_i, L_A + \Delta, \dots; t) \quad (1)$$

where  $L_A$  is the length of the given segment  $A$  of the road, and  $\Delta > 0$  (where  $\max_{j=1, \dots, N, j \neq i} x_j \leq L_A + \Delta$ ). If this function is differentiable, then the probability density for the position of vehicle  $i$  would be:

$$f_i(x; t) = \frac{\partial}{\partial x} F_i(x; t) \quad (2)$$

where  $f_i(x; t)dx$  for the small value of  $dx$  is the probability for vehicle  $i$  to be between  $x$  and  $x + dx$  at time  $t$ . The total density at coordinate  $x$  at time instant  $t$  (which could also be considered as the *instant* density apart from the *generalized* density which will be defined afterwards) for the  $n_A$  observed vehicles on the road segment is then:

$$\rho(x, t) = \sum_{i=1}^{n_A} f_i(x; t) \quad (3)$$

In this way we can successfully define density for a road segment of any desired length. The flow can similarly be defined through the time axis.

A practical way of implementing the mathematical interpretation of the density, the flow, and correspondingly the average speed for a traffic network is given by Edie (1963) where the *generalized* definitions of the fundamental variables are represented. Consider an arbitrary-shaped area  $A$  within the time-space plane for which the covered time intervals at different positions are sub-intervals of  $[t_l \ t_u]$  and the covered position intervals at different time instants are sub-intervals of  $[x_l \ x_u]$  (see Figure 2). Suppose that a part of the trajectory of vehicle  $i$  ( $i = 1, \dots, n_A$ ) is enclosed by the area  $A$  such that the projection of this trajectory onto the time and space axes are  $t_i(A)$  and  $d_i(A)$  respectively. Then the generalized density  $\rho(A)$  and flow  $q(A)$  for the area  $A$  are given by:

$$\rho(A) = \frac{1}{|A|} \sum_{i=1}^{n_A} t_i(A), \quad (4)$$

$$q(A) = \frac{1}{|A|} \sum_{i=1}^{n_A} d_i(A) \quad (5)$$

where  $d_i(A)$  and  $t_i(A)$  are indeed the distance traveled and the time spent in the area by vehicle  $i$ , and  $n_A$  is the number of vehicles that have been observed in the area  $A$ .

Newell (1995) shows that Edie's definitions of the generalized fundamental variables are equivalent with the mathematical definitions of density given by (3) and flow (which is obtained following the same reasoning as for the density). Finally, inspired by the

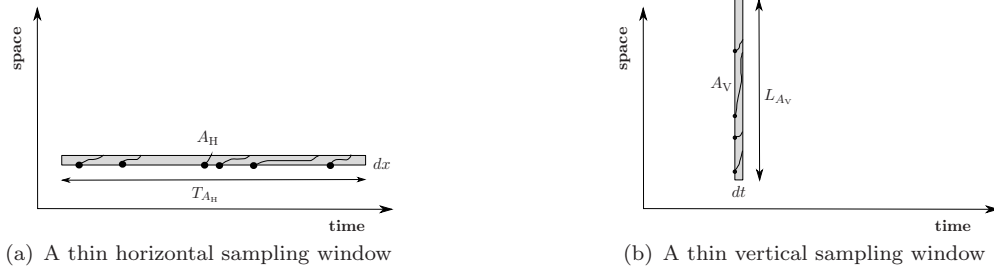


Figure 3. The TMS and SMS versus the TSMS

fundamental traffic equation the generalized average speed, is defined as the ratio of the generalized flow and density, i.e.,

$$v(A) = \frac{q(A)}{\rho(A)} = \frac{\sum_{i=1}^{n_A} d_i(A)}{\sum_{i=1}^{n_A} t_i(A)} \quad (6)$$

#### 2.4 TMS and SMS vs. TSMS

In the following discussion we consider the relationships between the TMS and the SMS with the TSMS. We will apply definition of the generalized average speed to the following two cases:

- (1) A thin horizontal sampling window with length  $dx$  and width  $T_A$  (see Figure 3(a))
- (2) A thin vertical sampling window with length  $L_A$  and width  $dt$  (see Figure 3(b))

The first case could represent the detection zone  $A_H$  of a loop detector. Due to the fact that  $dx$  is very small, the possibility that a trajectory enters or leaves the window through its left or right edge is negligible. Hence,

$$\text{TSMS}(A_H) = \frac{\sum_{i=1}^{n_{A_H}} d_i(A_H)}{\sum_{i=1}^{n_{A_H}} t_i(A_H)} = \frac{dx \cdot n_{A_H}}{\sum_{i=1}^{n_{A_H}} \frac{dx}{v_{A_H,i}}} = \frac{1}{\frac{1}{n_{A_H}} \sum_{i=1}^{n_{A_H}} \frac{1}{v_{A_H,i}}} \quad (7)$$

and

$$\text{TMS}(A_H) = \frac{1}{n_{A_H}} \sum_{i=1}^{n_{A_H}} v_{A_H,i} \quad (8)$$

where  $n_{A_H}$  is the number of vehicles detected by the loop detector during  $T_{A_H}$ . Thus, for a thin horizontal sampling window the TSMS corresponds the harmonic mean of the detected speeds, while the TMS is the arithmetic mean of the speeds. This is also discussed extensively by Treiber and Kesting (2013).

The sampling window of case two could be the observation area of a camera. Due to the fact that  $dt$  is very small, the possibility that a trajectory enters or leaves the window

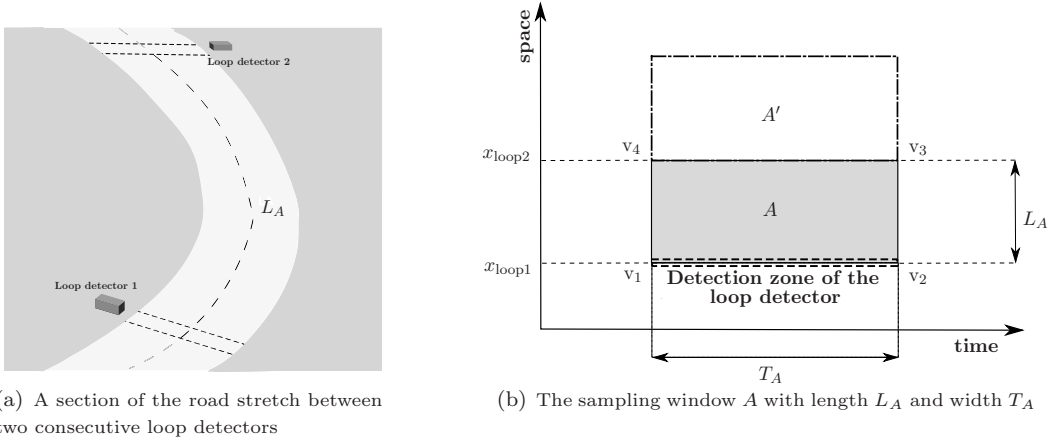


Figure 4. Specifying a sampling window for a sampling road section in the time-space plane to estimate the TSMS

through its bottom or top edge is negligible. Thus we have,

$$\text{TSMS}(A_V) = \frac{\sum_{i=1}^{n_{A_V}} d_i(A)}{\sum_{i=1}^{n_{A_V}} t_i(A_V)} = \frac{\sum_{i=1}^{n_{A_V}} dt \cdot v_{A_V,i}}{dt \cdot n_{A_V}} = \frac{1}{n_{A_V}} \sum_{i=1}^{n_{A_V}} v_{A_V,i} \quad (9)$$

and

$$\text{SMS}(A_V) = \frac{1}{n_{A_V}} \sum_{i=1}^{n_{A_V}} v_{A_V,i} \quad (10)$$

where  $n_{A_V}$  is the number of vehicles observed in the photograph captured by the camera. Then the TSMS at a specific instant of time corresponds the arithmetic mean of the observed vehicles, which by definition is equal to the SMS. Therefore, for a thin vertical rectangle the TSMS and the SMS are equivalent.

## 2.5 Defining the sampling windows for estimation of the TSMS

From the discussions in Section 2.3, the TSMS is the required average speed for the traffic applications addressed in Section 1. Since for the TSMS traffic information is averaged on an *area* over two dimensions (i.e., space and time axes), and since we will use the trajectories of the vehicles to consider features of the traffic stream, we should work in the time-space plane. We will consider a rectangular window in the time-space plane that represents the road segment under consideration during one sampling period. From now, we will simply call this rectangular time-space window the *sampling window*.

Consider the road section illustrated in Figure 4(a), where two consecutive loop detectors are positioned at the downstream and upstream of the section. Suppose that the sampling period of the loop detectors is  $T_A$ , and the distance between the two loop detectors, i.e., the length of the black dashed string, is  $L_A$ . Then the length and the width of the sampling window will respectively be  $L_A$  and  $T_A$  (see Figure 4(b)).

The downstream loop detector collects data in its detection zone, which is illustrated by the thin dotted rectangle at  $x = x_{\text{loop1}}$  in Figure 4(b). Under such a configuration where point measurements are available at discrete points on the road where the loop detectors



are installed, the challenge is to find an approach to estimate the average speed during each sampling period through the length  $L_A$ . In the next section, we discuss previous work on estimating the average speed.

### 3. Overview of previous work

For stationary traffic if we define the flow locally (i.e., at a fixed location), and the same way define the density instantly (i.e., at a specific time instant), then the SMS and the TSMS are equivalent, and the ratio of the local flow and instant density would be the SMS (Daganzo 1997). Most literature considers a way to estimate the SMS, while in this paper we consider the general case where traffic is not necessarily stationary. Therefore, the generalized definition of the speed, i.e., the TSMS, should be applied.

Wardrop (1952) develops a relationship between the SMS and the TMS using a macroscopic point-of-view:

$$\text{TMS} = \text{SMS} + \frac{\sigma_{\text{SMS}}^2}{\text{SMS}} \quad (11)$$

where  $\sigma_{\text{SMS}}^2$  is the standard deviation of the space distribution of speeds. However, since in the case of a loop detector the only available information is the TMS, it is not straightforward to estimate  $\sigma_{\text{SMS}}^2$ .

Han et al. (2010) discuss a theoretical approach in combination with an empirical method to solve (11) for the SMS using the definition of  $\sigma_{\text{SMS}}^2$ . The applied procedure is as follows:

$$\begin{aligned} \sigma_{\text{SMS}}^2 &= \text{E}[(v_i - \text{SMS})^2] \\ &= \text{E}[v_i^2] + \text{SMS}^2 - 2\text{SMS} \cdot \text{TMS} \end{aligned} \quad (12)$$

where  $v_i$  is the speed of the  $i^{\text{th}}$  vehicle passing through the loop detector, and  $\text{E}[\cdot]$  denotes the expected value operator. The above relation is applied to (11) and the resulting quadratic equation is solved to eventually result in:

$$\text{SMS} = \frac{3\text{TMS} \pm \sqrt{9\text{TMS}^2 - 8\text{E}[v_i^2]}}{4} \quad (13)$$

In (13) the value of  $\text{E}[v_i^2]$  is unknown. Han et al. (2010) propose a quadratic relationship between  $\text{E}[v_i^2]$  and  $\text{E}[v_i]$ :

$$\text{E}[v_i^2] = a \cdot \text{TMS}^2 + b \cdot \text{TMS} + c \quad (14)$$

where a previously collected set of data from the loop detectors should be used to estimate the constant coefficients  $a$ ,  $b$ , and  $c$  empirically.

Dailey (1999) presents an algorithm to estimate the mean speed using data from a single-loop detector. Dailey (1999) considers the statistical nature of the measurements made by single-loop detectors and presents an algorithm which estimates the speed. The two typical measurements, i.e., the flow  $q$  and the occupancy  $o$ , are taken into account:

$$o = \frac{1}{T} \sum_{i=1}^{q \cdot T} \frac{l_i}{v_i} \quad (15)$$

where  $l_i$  is the effective length (i.e., the length of the vehicle + the length of the detection zone) of the  $i^{\text{th}}$  vehicle, and  $T$  is the duration of the measurement.

Furthermore, the speed and the length of the vehicles are random variables that could be rewritten as the summation of their mean value (respectively  $\bar{v}$  and  $\bar{l}$ ) and the deviation (respectively  $\Delta v_i$  and  $\Delta l_i$ ). Using these expressions, Dailey (1999) obtains the following equation for the mean speed,  $\bar{v}$ :

$$o \frac{T}{\bar{l}} \bar{v}^3 - q \bar{v}^2 - q \sigma_v^2 = 0 \quad (16)$$

Note that Dailey (1999) does not distinguish between the TMS, the SMS, and the TSMS and only uses the notation  $\bar{v}$  in general. Two estimation methods are introduced by Dailey (1999). The first one that solves (16) to find the real root  $\bar{v}$  is called the ‘*root finding*’ method, which is based on the deterministic values and which yields an unbiased estimator for  $\bar{v}$  assuming idealized noiseless measurements. The second method is the ‘*filtering*’ method, which considers real measurements. This method is represented as an algorithm that gives an estimate of  $\bar{v}$ , and also provides a reliability test for the computed average speed.

Rakha and Zhang (2005) also propose a relationship between the TMS and the SMS using an approach similar to Dailey (1999) for the statistics of the measurements. Contrary to the formula given by Wardrop (1952) where the TMS is solved from the SMS, for the formula proposed by Rakha and Zhang (2005), the SMS is estimated from the TMS:

$$\text{SMS} \simeq \text{TMS} - \frac{\sigma_{\text{TMS}}^2}{\text{TMS}} \quad (17)$$

In (17) for estimation of the SMS, we need to know the value of the TMS and also  $\sigma_{\text{TMS}}^2$ . However, regular loop detectors only report the value of the TMS to the traffic management center. Therefore, Soriguera and Robusté (2011b) propose to assume a normal distribution for the vehicle speeds in order to find an estimate for  $\sigma_{\text{TMS}}$  that can be used in (17). A confidence interval is also derived by Soriguera and Robusté (2011b) for the estimated value of the SMS.

#### 4. New formulas for a tight upper and a tight lower bound for the TSMS based on a microscopic point-of-view: theory and formulation

Since traffic monitoring and control decisions are made based on the information available from point measurements at the location of the loop detectors, from trajectory data, or from floating car data (FCD), it is critical to make some approximations. The main objective is to make these assumptions more accurate. Not only is a good estimate of the TSMS desired for traffic control and performance evaluation, but we may also need to predict this value for a short upcoming time interval such that the control center can take the appropriate actions in real time. To this aim, a microscopic approach is selected in this article. Next, we present a new solution for the problem described in Section 2.

##### 4.1 Division of the sampling window into grid cells

Consider the sampling window of dimensions  $T_A$  and  $L_A$  shown in Figure 5 that contains the trajectories of the vehicles that have been observed by the loop detector installed at the downstream of the road section. The number of the observed vehicles is  $N_A$ , and the

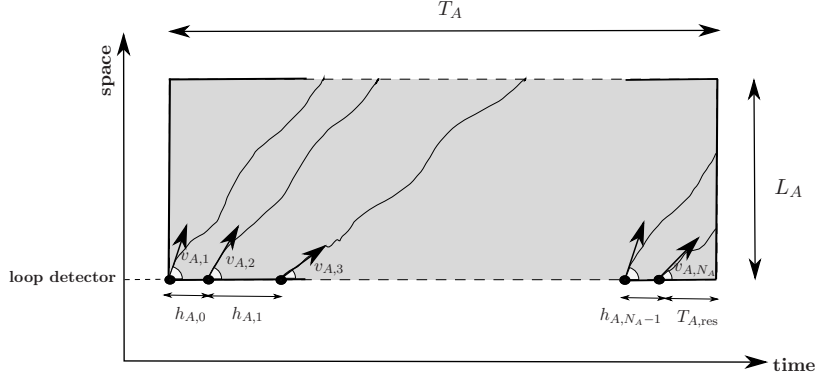


Figure 5. Trajectories of the observed vehicles within the sampling window  $A$

corresponding microscopic data that can be extracted from the loop detector during one sampling period are:

$$\begin{aligned}\mathbf{V}_A &= \{v_{A,i} \mid i = 1, 2, \dots, N_A\}, \\ \mathbf{H}_A &= \{h_{A,i-1} \mid i = 1, 2, \dots, N_A\}\end{aligned}\tag{18}$$

where  $v_{A,i}$  and  $h_{A,i-1}$  are respectively the velocity of the  $i^{\text{th}}$  observed vehicle, and the time headway between the  $(i-1)^{\text{th}}$  and the  $i^{\text{th}}$  vehicle (Note that for  $i = 1$ ,  $h_{A,0}$  indicates the time duration from the beginning of the observation cycle until the first vehicle is observed).

From the above measurements, we obtain the following information:

$$\begin{aligned}v_{A,\min} &= \min_{i=1,\dots,N_A} (v_{A,i}), \\ v_{A,\max} &= \max_{i=1,\dots,N_A} (v_{A,i}), \\ h_A &= \frac{1}{N_A} \sum_{i=1}^{N_A} h_{A,i-1}\end{aligned}\tag{19}$$

where  $v_{A,\min}$ ,  $v_{A,\max}$ , and  $h_A$  are the minimum and maximum velocities of the observed vehicles and the mean time headway.

The sampling window is first divided into  $n_A$  grid cells of length  $L_A$  and width  $h_A$  (see Figure 6), where:

$$n_A := \left\lceil \frac{T_A}{h_A} \right\rceil\tag{20}$$

where  $\lceil \cdot \rceil$  denotes the ceiling function. Note that the two parameters  $n_A$  and  $N_A$  are not necessarily equal. However, if we assume to have  $T_{A,\text{res}} < h_A$ , where  $T_{A,\text{res}}$  is the time duration from the last observation until the end of the observation period (see Figure 5), then we will have:

$$N_A = \frac{1}{h_A} \sum_{i=1}^{N_A} h_{A,i-1} = \frac{1}{h_A} (T_A - T_{A,\text{res}}) \xrightarrow[\text{since } N_A \in \mathbb{N}]{\text{if } T_{A,\text{res}} < h_A} N_A = \left\lceil \frac{T_A}{h_A} \right\rceil := n_A\tag{21}$$

that is  $n_A$  and  $N_A$  are equal. So from now, we assume to have  $n_A = N_A$ .

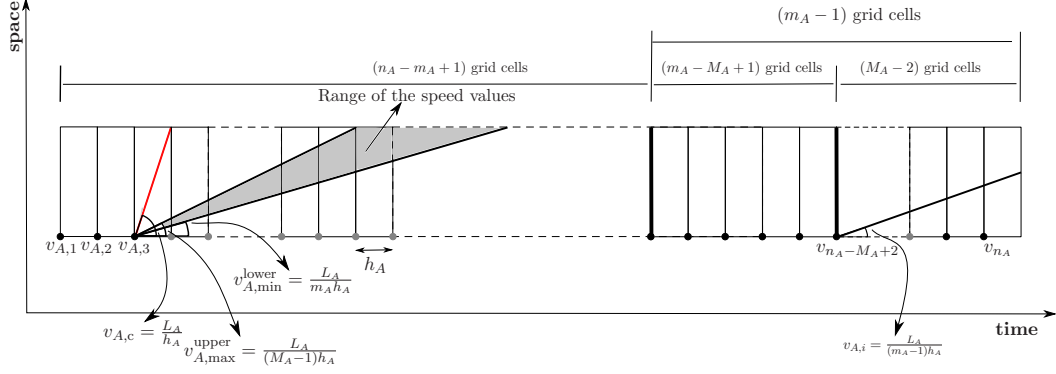


Figure 6. Dividing the grid cells into three parts in order to find a lower and an upper bound of the TSMS

At this point we introduce a new parameter called the *cell speed*, which depends on the dimensions of the sampling window and which is defined by:

$$v_{A,c} = \frac{L_A}{h_A} \quad (22)$$

Note that the cell speed is the least required speed for traveling the complete length  $L_A$  of the road within one grid cell. Now we determine the integer value  $m_A$  such that:

$$m_A - 1 \leq \frac{v_{A,c}}{v_{A,min}} < m_A \Rightarrow m_A = \left\lceil \frac{L_A}{h_A v_{A,min}} \right\rceil + 1 \quad (23)$$

Similarly, the integer value  $M_A$  is determined such that:

$$M_A - 1 \leq \frac{v_{A,c}}{v_{A,max}} < M_A \Rightarrow M_A = \left\lceil \frac{L_A}{h_A v_{A,max}} \right\rceil + 1 \quad (24)$$

## 4.2 Equal time headway distribution

In Figure 6, a sampling window is shown that is divided into  $n_A$  grid cells, such that at the left bottom corner of each grid cell one and only one vehicle is located. Here, we assume the case for which the time headways of the vehicles are all the same and equal to the mean time headway value<sup>1</sup>.

We divide the vehicles into two sets based on the grid cell that they are at. We call these two sets the ‘*first*’ and the ‘*second*’ set. By the *first set* we refer to the vehicles in the first  $n_A - m_A + 1$  grid cells, and by the *second set* we mean the vehicles within the last  $m_A - 1$  grid cells.

Furthermore, the second set will also be divided into two subsets called the *first* and the *second* subsets, where the first subset includes the first  $m_A - M_A + 1$  vehicles of the second set and the second subset includes the remaining  $M_A - 2$  vehicles (see Figure 6).

From (23) we have:

$$v_{A,min} > \frac{v_{A,c}}{m_A} \quad (25)$$

<sup>1</sup>Note that in (Jamshidnejad and De Schutter 2014) it is shown that the results obtained in Sections 4 and 5 of this paper are still valid for the case of non-equal time headway distribution.

which indicates that the vehicles that face  $m_A$  or more number of grid cells in front of them will travel a distance greater than or at least equal to  $L_A$  during the observation time interval. From Figure 6, the vehicle with index  $n_A - m_A + 1$  and all its predecessors in  $A$  (i.e., the *first set*) fulfill this condition. Consequently, those vehicles that enter the detection zone within the first  $(n_A - m_A + 1)h_A$  seconds of the observation time interval will always leave the sampling window  $A$  through its upper edge, while for the vehicles that enter the detection zone in the last  $(m_A - 1)h_A$  seconds of the observation time interval, the trajectories *might* intersect the right edge of the sampling window. Therefore, we need to consider the vehicles in the *second set* more carefully. We expand (6) as

$$v(A) = \frac{\sum_{i=1}^{n_A-m_A+1} L_A + \sum_{i=1}^{m_A-1} t_{n_A-i+1} \cdot v_{A,n_A-i+1}}{\sum_{i=1}^{n_A-m_A+1} \frac{L_A}{v_{A,i}} + \sum_{i=1}^{m_A-1} t_{n_A-i+1}} \quad (26)$$

Next, we will determine a lower bound and an upper bound for (26).

#### 4.2.1 Lower bound

To find a lower bound for  $v(A)$  in (26), we consider the case for which all vehicles within the second group will move with  $v_{A,\min}$ . The TSMS for such a case is definitely a lower bound for any other possible scenario in which the first set of the vehicles move with their real speed values, while we change the speeds of the vehicles in the second set. For this situation, all vehicles in the *second set* will stay within  $A$  until the end of the observation period. Therefore,

$$t_{n_A-j+1} = jh_A, \quad j = 1, 2, 3, \dots, m_A - 1 \quad (27)$$

For the generalized mean speed given by (26) the following holds:

$$v(A) \geq \frac{\sum_{i=1}^{n_A-m_A+1} L_A + h_A \sum_{i=1}^{m_A-1} i \cdot v_{A,n_A-i+1}}{\sum_{i=1}^{n_A-m_A+1} \frac{L_A}{v_{A,i}} + h_A \sum_{i=1}^{m_A-1} i} \quad (28)$$

To continue, we find an upper bound for the denominator using the definition of  $M_A$  in (24). For the first  $n_A - m_A + 1$  vehicles entering the area we can write:

$$h_A \leq \frac{L_A}{(M_A - 1)v_{A,i}} \Rightarrow (n_A - m_A + 1)h_A \leq \frac{1}{(M_A - 1)} \sum_{i=1}^{n_A-m_A+1} \frac{L_A}{v_{A,i}} \quad (29)$$

Substituting  $h_A$  for  $i = 1, 2, \dots, n_A - m_A + 1$  in the denominator of (28) with its upper bound from (29), and  $v_{A,i}$ ,  $i = n_A - m_A + 2, \dots, n_A$  in the numerator with the lower

bound of  $v_{A,\min}$  from (23) a lower bound for  $v(A)$  is obtained:

$$v(A) > \frac{\left(\sum_{i=1}^{n_A-m_A+1} L_A\right) + h_A \left(\sum_{i=1}^{m_A-1} i\right) \frac{L_A}{m_A h_A}}{\left(\sum_{i=1}^{n_A-m_A+1} \frac{L_A}{v_{A,i}}\right) + \frac{1}{(M_A-1)(n_A-m_A+1)} \left(\sum_{i=1}^{n_A-m_A+1} \frac{L_A}{v_{A,i}}\right) \sum_{i=1}^{m_A-1} i} \quad (30)$$

Finally, after simplification we get the following lower bound for  $v(A)$ :

$$v^{\text{lower}}(A) = \frac{n_A - \frac{m_A - 1}{2}}{(n_A - m_A + 1) + \frac{m_A - 1}{2} \left(\frac{m_A}{M_A - 1}\right)} H_{A,1 \rightarrow n_A - m_A + 1} \quad (31)$$

in which  $H_{A,1 \rightarrow n_A - m_A + 1}$  stands for the harmonic mean of the speeds of the first  $n_A - m_A + 1$  vehicles.

#### 4.2.2 Upper bound

To find an upper bound for  $v(A)$  in (26), the case is considered in which all vehicles within the second set move with  $v_{A,\max}$ . The TSMS for any other scenario (that has different speed values in the second set of the vehicles, but the same speeds for vehicles in the first set) will not exceed the TSMS for the scenario considered here. From (24) we know:

$$v_{A,\max} < \frac{v_{A,c}}{M_A - 1} \quad (32)$$

From (32) and Figure 6 for those vehicles that are located in the first subset (i.e., in the first  $m_A - M_A + 1$  grid cells of the second set) a distance larger than or at least equal to  $L_A$  will be traveled during the sampling period  $T_A$ , whereas for vehicles in the second subset (i.e., in the last  $M_A - 2$  grid cells of the second set) the traveled distance will be less than  $L_A$ .

Accordingly, to calculate an upper bound for  $v(A)$ , the second term of the numerator of (26) is split into two terms:

$$v(A) < \frac{\left(\sum_{i=1}^{n_A-m_A+1} L_A\right) + (m_A - M_A + 1)L_A + \frac{1}{M_A - 1} \left(\sum_{i=1}^{M_A-2} i\right) L_A}{\left(\sum_{i=1}^{n_A-m_A+1} \frac{L_A}{v_{A,i}}\right) + (m_A - M_A + 1)(M_A - 1)h_A + \left(\sum_{i=1}^{M_A-2} i\right) h_A} \quad (33)$$

where the following expressions, which are obtained from Figure 6, for the traveled time and the traveled distance of vehicles in the last  $M_A - 2$  grid cells have been substituted in (26):

$$\begin{aligned} t_{n_A-j+1} &= j h_A \\ d_{n_A-j+1} &= v_{A,n_A-j+1} \cdot t_{n_A-j+1} = \frac{j}{M_A - 1} L_A \end{aligned} \quad \text{for } j = 1, 2, \dots, M_A - 2$$

Applying (23) to find a lower bound for  $h_A$  in the denominator, we obtain an upper

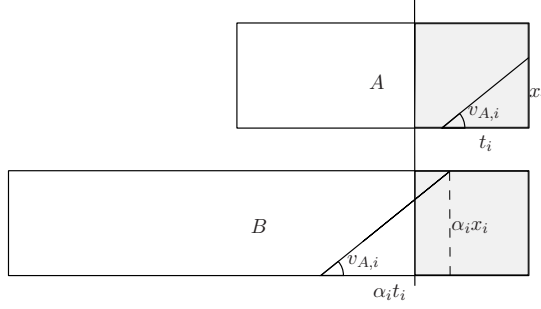


Figure 7. Sampling windows  $A$  and  $B$  with the same data sets (i.e., speed values, number of vehicles, and length of the window), but different time headways (we have  $\alpha_i > 1$  and  $\alpha_i x_i = L_A$ )

bound for  $v(A)$ :

$$v^{\text{upper}}(A) = \frac{n_A - \frac{M_A}{2} + 1}{(n_A - m_A + 1) + \frac{(M_A - 1)}{2m_A}(2m_A - M_A)} H_{A,1 \rightarrow n_A - m_A + 1} \quad (34)$$

### 4.3 Tightening the lower and upper bounds

In this section we use (31) and (34) with the aim of finding bounds that are very tight. Suppose that we have the data sets given by (18) from which we can derive the maximum and minimum speeds,  $v_{A,\max}$  and  $v_{A,\min}$ , and the mean time headway,  $h_A$ , given by (19). Equations (31) and (34) yield a lower and an upper bound for the TSMS for sampling window  $A$ . Therefore, we can write:

$$v^{\text{lower}}(A) \leq \text{TSMS}(A) \leq v^{\text{upper}}(A) \quad (35)$$

Now we construct a new sampling window  $B$  (see Figure 7) with all speed data the same as that of  $A$ , but with the headway equal to a different constant, i.e.,

$$h_B = h_A + \Delta h \quad (36)$$

and

$$L_B = L_A, \quad T_B = n_A h_B \quad (37)$$

and we want to find  $\Delta h$  such that:

$$v^{\text{lower}}(B) \leq \text{TSMS}(A) \leq v^{\text{upper}}(B) \quad (38)$$

Note that  $n_A = n_B$ . From Figure 7 we see that for  $\Delta h > 0$  some of the vehicles that are located in the second set for  $A$ , might be located in the first set for  $B$ . Let  $W$  be the number of vehicles from the second set of  $A$  that are in the first set for  $B$ . Therefore, we have:

$$m_B = m_A - W \quad (39)$$

Now we consider the upper and lower bounds separately, so we denote the extended sampling window for the upper bound by  $B_u$  and for the lower bound by  $B_l$ . In addition,

let  $W_u$  and  $W_l$  be the number of vehicles from the second set of  $A$  that are respectively in the first set for  $B_u$  and  $B_l$ .

First we consider the conditions under which the right-hand inequality in (38) is satisfied for  $B_u$ . By determining where the maximum of the difference between the harmonic mean for the first set of the vehicles in  $B_u$  and  $\text{TSMS}(A)$  is reached, and after some elaborate calculus (see (Jamshidnejad and De Schutter 2014) for the detailed proof), it can be verified that the upper bound calculated for the sampling window  $B_u$ , is also an upper bound for  $\text{TSMS}(A)$  if:

$$W_u \leq \min \left\{ m_A \frac{M_A - 2}{M_A - 1}, n_A - \frac{M_A}{2} + 1 \right\} \quad (40)$$

Likewise, based on (38) we can show that if

$$W_l \leq m_A - M_A + 1 \quad (41)$$

then the lower bound calculated for the sampling window  $B$  will also be a lower bound for  $\text{TSMS}(A)$ .

Finally, if we seek for a very tight upper bound for the TSMS,  $W_u$  found by the right-hand side of (40) could be used in (39) to determine  $m_{B_u}$ . Then we can compute  $h_{B_u}$  from (23), and correspondingly we can calculate  $M_{B_u}$  from (24). Substituting  $m_{B_u}$  and  $M_{B_u}$  in (34), we obtain a new upper bound,  $v^{\text{upper}}(B_u)$ , which we already know is an upper bound for  $\text{TSMS}(A)$  (see Jamshidnejad and De Schutter (2014)). We use the following notation:

$$\text{TSMS}^{\text{upper}}(A) := v^{\text{upper}}(B_u) \quad (42)$$

Similarly, a tight lower bound,  $v^{\text{lower}}(B_l)$ , can be found from (31). We use the following representation:

$$\text{TSMS}^{\text{lower}}(A) := v^{\text{lower}}(B_l) \quad (43)$$

## 5. Estimation of the TSMS from the upper and lower bounds

In this section, we introduce a formula for estimating the TSMS using a convex combination of the lower and upper bounds found in Sections 4.2 and 4.3:

$$\text{TSMS}_{\text{est}}(A) = \frac{\text{TSMS}^{\text{lower}}(A) + \gamma_A \text{TSMS}^{\text{upper}}(A)}{1 + \gamma_A} \quad \text{with } \gamma_A \geq 0 \quad (44)$$

In the above expression,  $\gamma_A$  could be identified based on a large data set including individual speeds of the vehicles for different traffic scenarios for the soar section under consideration.

However, such an extensive data set might not always be available or there might be a need for immediate application of the formula for a new area (without a priori knowledge about the possible traffic scenarios or without any available data). For such cases, a parametric expression for  $\gamma_A$  is more applicable. We propose a weighted average for the estimated TSMS with the weight being a function of the speed range, since in the given expression when the ratio of  $v_{A,\text{max}}$  and  $v_{A,\text{min}}$  increases, the share of the upper



bound in the estimated TSMS should be higher compared with the share of the lower bound. The proposed formula for the estimated TSMS is:

$$\text{TSMS}_{\text{est}}(A) = \frac{\text{TSMS}^{\text{lower}}(A) + \frac{v_{A,\text{max}}}{v_{A,\text{min}}} \cdot \text{TSMS}^{\text{upper}}(A)}{1 + \frac{v_{A,\text{max}}}{v_{A,\text{min}}}} \quad (45)$$

Now we explain how to estimate the TSMS from the convex combination given by (45) using the bounds for the TSMS from Section 4.3. We should use the bounds  $W_u$  and  $W_l$  (see (40) and (41)) for the same sampling window to apply (45). Therefore, for the estimation of TSMS we should substitute  $W$  with the following:

$$W = \min\{W_u, W_l\} \quad (46)$$

and correspondingly, we determine  $m_B$  and  $M_B$  and substitute them in (31) and (34). Finally, we substitute  $\text{TSMS}^{\text{lower}}(B)$  and  $\text{TSMS}^{\text{upper}}(B)$  in (45) to find an estimate for  $\text{TSMS}(A)$ .

## 6. Assessment and comparison

In this section we present the results for the formulas given by Wardrop (1952), i.e., (11), by Rakha and Zhang (2005), i.e., (17), and the new formula given in this paper, i.e., (45), for estimating the TSMS. We determine the relative errors of these formulas with respect to the real TSMS found by (6). As a comparison, we also present the relative error of the harmonic mean of the speeds with respect to the real TSMS, since the TSMS equals the harmonic mean of the speeds if the two averages are computed for the detection zone of a loop detector, i.e., a thin horizontal sampling window as shown in Figure 3(a) (see Section 2.4). This way we can determine how it affects the results when the detection zone of the loop detector is considered instead of the sampling window  $A$  (see Figure 4(b)). We first use real-life data from the NGSIM data set on the I-880 highway in the San Francisco Bay Area (<http://ngsim-community.org/>) and a Rotterdam-Delft data set for the highway A13 ([http://data.3tu.nl/repository/collection:traffic\\_flow\\_obs](http://data.3tu.nl/repository/collection:traffic_flow_obs)) that has been extracted from movies captured by a helicopter. Figure 8 shows the map of the A13 including the data set location.

Furthermore, to cover the possible scenarios that are not included in the real-life data sets, we also include simulations in MATLAB at the end of this section.

### 6.1 Real-life data (NGSIM, I-880 and Rotterdam-Delft, A13)

For multi-lane roads where lane-changing is also permitted (like the I-880 and the A13), there are two possible ways of estimating the average speed: the first considers the average speed separately for each lane, and the other considers multiple lanes together and estimates their overall average speed. Due to the lane changes, a measurement on one lane might affect the speed on the other lane(s), which is why we consider this joined case for the real-life data in this section.

Figure 9(a) shows the results of calculating the TSMS for the NGSIM data. We have extracted four traffic data sets from the NGSIM I-880 data set, where the corresponding values for  $L_A$ ,  $h_A$ ,  $T_A$ ,  $v_{A,\text{min}}$ , and  $v_{A,\text{max}}$  are given in Table 1. From the relative errors shown in Figure 9(a) we see that the convex combination (45) yields the best performance.

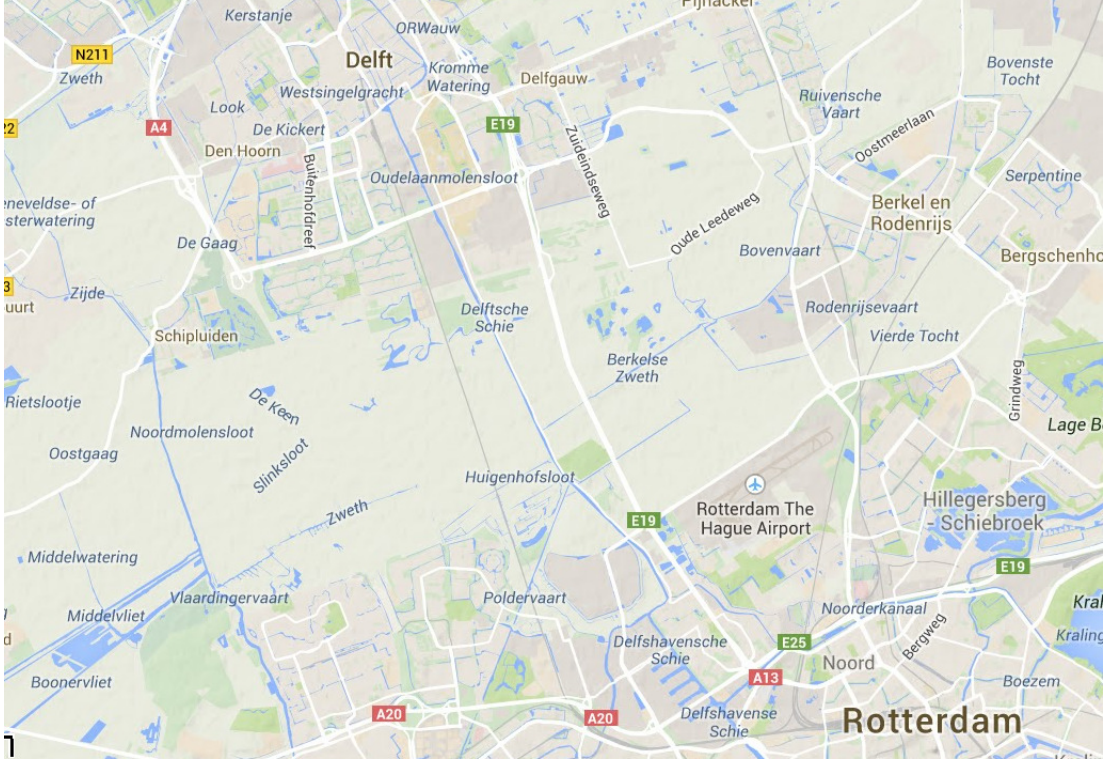


Figure 8. Map displaying the highway A13, Rotterdam-Delft for which the real-life traffic data is available

Table 1. The values of the parameters for the NGSIM I-880 data sets

Data set	$L_A$ (m)	$h_A$ (s)	$T_A$ (min)	$v_{A,\min}$ ( $\frac{m}{s}$ )	$v_{A,\max}$ ( $\frac{m}{s}$ )	$n_A(-)$
DS 1	500	0.34	102.00	10.00	34.50	300
DS 2	500	0.34	102.56	18.25	33.09	300
DS 3	500	0.34	102.65	11.33	32.87	300
DS 4	500	0.34	102.36	17.96	30.85	300

Table 2. The values according to the Rotterdam-Delft A13 data sets

Data set	$L_A$ (m)	$h_A$ (s)	$T_A$ (s)	$v_{A,\min}$ ( $\frac{m}{s}$ )	$v_{A,\max}$ ( $\frac{m}{s}$ )	$n_A(-)$
DS 5	300	0.81	183.00	6.82	38.94	227
DS 6	300	1.07	204.00	12.24	35.30	191
DS 7	300	1.01	220.00	13.33	34.24	217
DS 8	300	0.87	192.00	10.30	35.15	221
DS 9	300	0.92	213.00	12.42	39.55	231
DS 10	300	0.99	192.00	12.88	35.15	214
DS 11	300	0.93	387.00	6.81	38.94	418
DS 12	300	1.94	400.00	13.63	40.60	206
DS 13	300	1.05	27.00	8.50	27.50	24
DS 14	300	1.30	40.00	7.80	31.20	32
DS 15	300	1.50	41.00	6.90	31.03	30
DS 16	300	1.50	25.00	8.90	27.32	18

Next we consider 12 data sets that are extracted from the Rotterdam-Delft A13 real-life data. The values for  $L_A$ ,  $h_A$ ,  $T_A$ ,  $v_{A,\min}$ , and  $v_{A,\max}$  are given in Table 2. Note that we have illustrated the results for the first 8 data sets (DS 5 to DS 12) and the last 4 data sets (DS 13 to DS 16) in two separate figures (see Figures 9(b) and 9(c)), since the ranges of the errors for these two cases were very different.

Figure 9(b) represents the relative errors of the formulas by Rakha and Zhang (2005), by Wardrop (1952), the harmonic mean of the speeds, and the convex combination with respect to the real TSMS (calculated by (6)) for the Rotterdam Delft A13 real-life data

Table 3. Classification of speed ranges for flow scenarios

Traffic scenario	Range of speeds ( $\frac{m}{s}$ )
Over-saturated (Breakdown)	$< 15.5$
Queue discharge	$15.5 - 25.0$
Under-saturated (Free flow)	$\geq 25.0$

Table 4. The values for traffic parameters used in different scenarios simulated in MATLAB

Scenario	$L_A(m)$	$h_A(s)$	$T_A (min)$	$v_{A,min} (\frac{m}{s})$	$v_{A,max} (\frac{m}{s})$	$n_A(-)$
Over-saturated (1 <sup>st</sup> )	500	2.0	5	2.0	15.5	150
Over-saturated (2 <sup>nd</sup> )	500	2.0	1	2.0	15.5	30
Queue discharge (1 <sup>st</sup> )	500	1.0	5	15.5	25.0	300
Queue discharge (2 <sup>nd</sup> )	500	1.0	1	15.5	25.0	60
Under-saturated (1 <sup>st</sup> )	500	0.5	5	25.0	40.0	600
Under-saturated (2 <sup>nd</sup> )	500	0.5	1	25.0	40.0	120
Under-saturated (3 <sup>rd</sup> )	500	2.0	5	25.0	40.0	150
Under-saturated (4 <sup>th</sup> )	500	2.0	1	25.0	40.0	30

sets DS 5 to DS 12 for which the sampling period, i.e., the time during which the TSMS is calculated, is more than 3 min. From Figure 9(b) for most data sets the convex combination yields the best performance. Among the data sets, only for DS 9 and DS 10 the harmonic mean and the formula by Rakha and Zhang (2005) respectively, produce smaller errors.

Figure 9(c) shows the results produced for the Rotterdam-Delft real-life data sets DS 13 to DS 16 where the sampling period is less than 1 min. It is again observed in Figure 9(c) that the convex combination has the best performance.

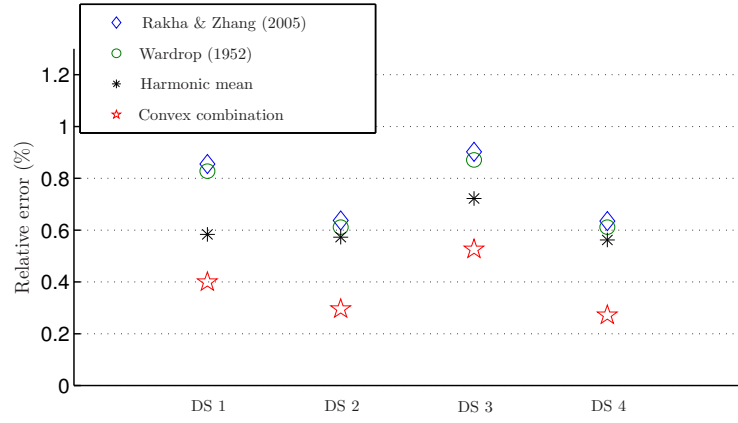
Comparing Figures 9(a), 9(b), and 9(c) when the minimum observed speed,  $v_{A,min}$ , decreases and the sampling period  $T_A$  along which the TSMS is computed decreases, the percentage of errors for all formulas increases. However, this percentage is still less than 1% for the convex combination, while for the other formulas the percentage of errors grows dramatically. In Figure 9(c), for DS 13 for example, there is a difference of almost 7% between the least percentage of error (which corresponds to the convex combination) and the percentage of error produced by the harmonic mean (which performs better than the formulas by Rakha and Zhang (2005) and by Wardrop (1952)).

## 6.2 MATLAB simulations

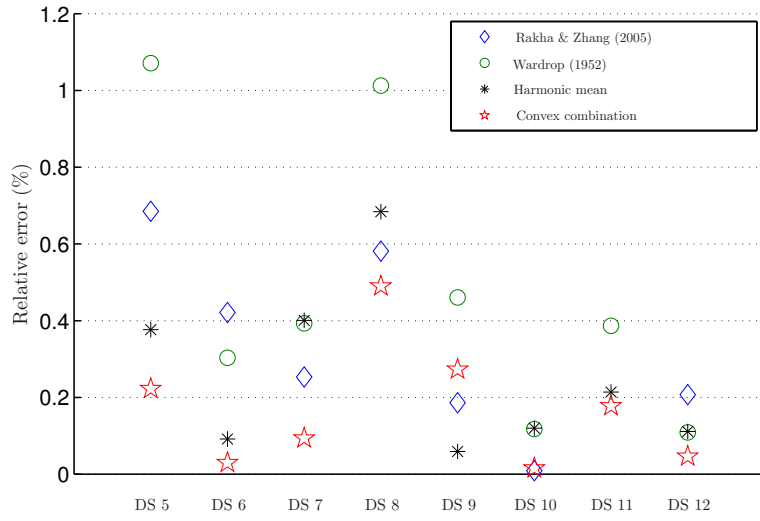
In order to consider the scenarios that are not covered by the real-life data given in Section 6.1, we now present simulations using MATLAB. In this setting we can repeat the experiments for different ranges of  $h_A$ ,  $v_{A,min}$ , and  $v_{A,max}$  as many times as is desired, and construct the boxplots of the errors produced by the different averaging formulas. Such a boxplot shows the shape of the distribution of the errors, its central value, and variability. The produced picture indicates the most extreme values in the data set, the lower and upper quartiles, and the median.

We generate a set of individual speeds using a normal<sup>1</sup> distribution with the mean value equal to the average of  $v_{A,min}$  and  $v_{A,max}$ , which are chosen for different scenarios using Table 3. In order to produce the individual speeds based on a normal distribution, we also need the standard deviations for which we use the values given by Huey et al. (2012) for different traffic scenarios. To find the real TSMS using (6), the trajectories of the vehicles are assumed to be straight lines, an assumption that is confirmed by

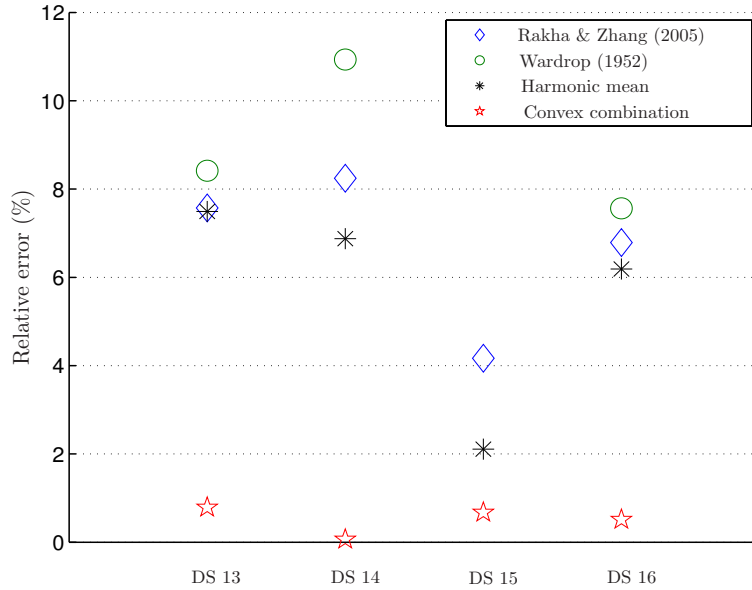
<sup>1</sup>According to May (1990) one of the most commonly used mathematical distributions for representing the measured speed values is the normal distribution.



(a) The relative errors for the formulas by Rakha and Zhang (2005), and by Wardrop (1952), the harmonic mean of the individual speeds, and the convex combination developed in this paper for the **NGSIM I-880** data sets DS 1 to DS 4 given in Table 1



(b) The relative errors for the formulas by Rakha and Zhang (2005), and by Wardrop (1952), the harmonic mean of the individual speeds, and the convex combination for the **Rotterdam-Delft A13** data sets DS 5 to DS 12 given in Table 2



(c) The relative errors for the formulas by Rakha and Zhang (2005), and by Wardrop (1952), the harmonic mean of the individual speeds, and the convex combination for the **Rotterdam-Delft A13** data sets DS 13 to DS 16 given in Table 2

Figure 9. Percentage of the errors produced by different formulas using two real-life data sets **NGSIM I-880** and **Rotterdam-Delft A13**

the trajectories obtained from the NGSIM and the Rotterdam-Delft data. Since we consider loop detectors, we use  $L_A = 500$  m, which is the prevalent distance between two consecutive loop detectors in most parts of Europe, and also in many areas of the US.

We will consider the scenarios given in Table 3. The speed ranges for different traffic scenarios in Table 3 are based on the information given by HCM (2000). In Section 6.1 we observed that by decreasing the sampling period,  $T_A$ , the error produced by the formulas by Rakha and Zhang (2005), by Wardrop (1952), and by the harmonic mean increases, while the convex combination still yields great results. Therefore, here for each of the over-saturated and queue discharge scenarios, MATLAB simulations are repeated two times for  $T_A = 5$  min and  $T_A = 1$  min. For the mean time headway of the vehicles,  $h_A$ , we use the results provided by Zou et al. (2014) and Brackstone et al. (2009), where the correlation between the speed and the time headway is studied. Based on the results of these papers, we have selected  $h_A$  for each scenario as shown in Table 4.

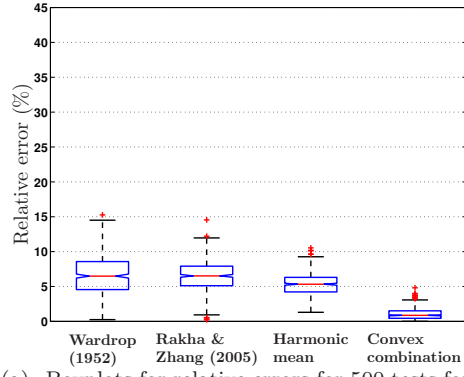
We should note that in (Zou et al. 2014) and (Brackstone et al. 2009), a range of values is given for the time headway for each traffic scenario. In the fundamental diagram, the under-saturated scenario covers small to large flows. Therefore, it depends on the number of the vehicles traveling on the road that which time headway is observed. If there is a large number of vehicles moving freely on the road, then  $h_A$  decreases. However, if the number of the vehicles reduces, then  $h_A$  increases. Therefore, for the under-saturated scenario, MATLAB simulations are repeated four times where in the first and second simulations  $h_A = 0.5$  s is considered and in the third and fourth simulations  $h_A = 2.0$  s is used (see Table 4 for the parameter values and Figures 10(e), 10(f), 10(g), and 10(h) for the results of the simulations).

The results for MATLAB simulations are given in Figure 10. From this figure we see that in general the errors by all formulas decrease when the flow scenario changes from the over-saturated to the under-saturated case, where it becomes negligible for example in Figure 10(e). In all cases the convex combination gives the best performance. The differences in performance are more highlighted for the over-saturated case, where the best formula is clearly the convex combination. Here for  $T_A = 5$  min (see Figure 10(a)) an improvement of almost 4% is obtained by the proposed convex combination. For a shorter sampling period, i.e.,  $T_A = 1$  min (see Figure 10(b)) the improvement resulted by the convex combination increases to 14%. Comparing the results in the first and second columns of Figure 10, the same result as in Section 6.1 is obtained, i.e., in general by decreasing the value of  $T_A$ , the error values produced by the formulas by Rakha and Zhang (2005), by Wardrop (1952), and by the harmonic mean increase dramatically, while the convex combination produces considerably smaller errors compared with the other formulas.

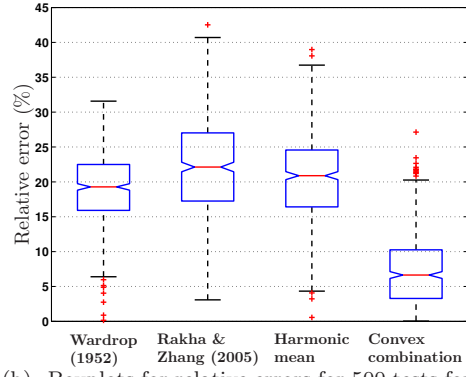
## 7. Conclusions and future work

We have first developed tight upper and lower bounds for the time-space-mean speed (TSMS), which is an equivalent for the generalized speed introduced by Edie (1963). In our proposed method, we have used microscopic traffic point measurements. Afterwards, we have introduced a convex combination of the upper and lower bounds such that an appropriate estimate of the TSMS is obtained.

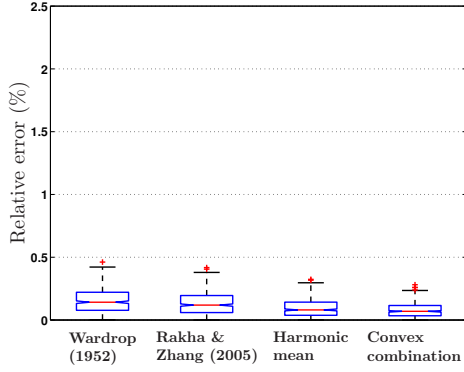
In order to assess and to compare the performance of the different formulas, we have applied them to two real-life traffic data sets corresponding to the NGSIM data for the I-880 highway in the San Francisco Bay Area, and the A13 data near Rotterdam-Delft. Moreover, we have also included a number of MATLAB simulations to consider some of the traffic scenarios that are not covered by the real-life data sets. In this way different traffic scenarios including the under-saturated, the queue discharge, and the



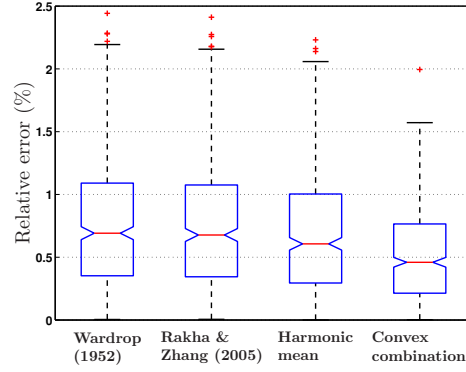
(a) Boxplots for relative errors for 500 tests for the 1<sup>st</sup> over-saturated scenario



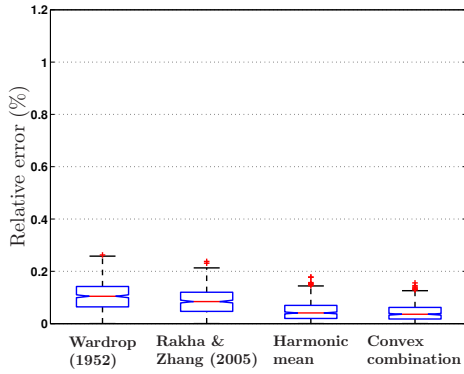
(b) Boxplots for relative errors for 500 tests for the 2<sup>nd</sup> over-saturated scenario



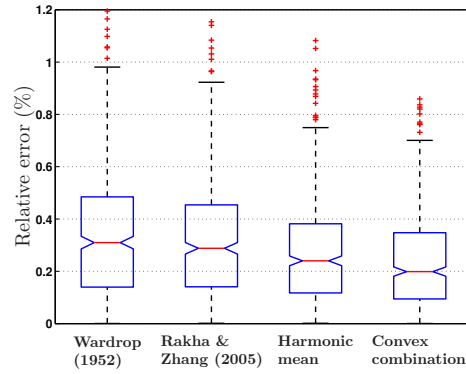
(c) Boxplots for relative errors for 500 tests for the 1<sup>st</sup> queue discharge scenario



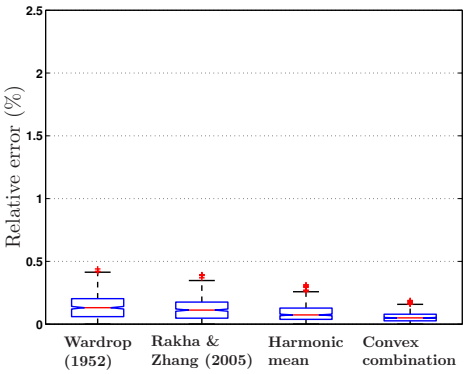
(d) Boxplots for relative errors for 500 tests for the 2<sup>nd</sup> queue discharge scenario



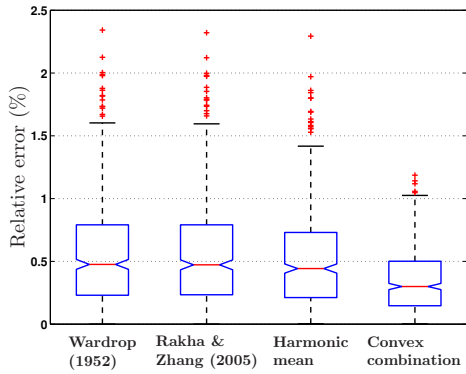
(e) Boxplots for relative errors for 500 tests for the 1<sup>st</sup> under-saturated scenario



(f) Boxplots for relative errors for 500 tests for the 2<sup>nd</sup> under-saturated scenario



(g) Boxplots for relative errors for 500 tests for the 3<sup>rd</sup> under-saturated scenario



(h) Boxplots for relative errors for 500 tests for the 4<sup>th</sup> under-saturated scenario

Figure 10. The percentage of errors for the formulas by Wardrop (1952), and by Rakha and Zhang (2005), the harmonic mean of the individual speeds, and the convex combination developed in this paper for **MATLAB** simulations



over-saturated flow have been covered.

The results from both the real-life data and from the MATLAB simulations show that the values produced by the new formula give better estimates of the TSMS compared with the formulas by Wardrop (1952), and by Rakha and Zhang (2005), and the pure harmonic mean of the individual speeds. An improvement of 7% has been obtained using the convex combination proposed in this paper for the real-life data sets. The improvement resulted by the convex combination for the MATLAB simulations has been even higher, i.e., an improvement of 14% has been obtained.

For future work, we will consider an important subject that is missing in the literature considering estimation of the average speed, and that is how to handle the vehicles that remain on a given road section for more than one sampling cycle. These vehicles are detected by the downstream loop detector once they enter the road section. However, in future cycles the data of these vehicles will not be considered by the loop detector.

## Acknowledgments

This research has been supported by the NWO-NFSC project “Multi-level predictive traffic control for large-scale urban networks” (629.001.011), which is partly financed by the Netherlands Organisation for Scientific Research (NWO).

We thank dr.ir. Andreas Hegyi for providing the real-life traffic data set for the highway A13, Rotterdam-Delft, and for his valuable comments.

## References

- Boyce, D., and Xiong, C. 2007. “Forecasting Travel for Very large Cities: Challenges and Opportunities for China.” *Transportmetrica* 3 (1): 1–19.
- Brackstone, M., Waterson, B., and McDonald, M. 2009. “Determinants of Following Headway in Congested Traffic.” *Transportation Research Part F: Traffic Psychology and Behaviour* 12 (2): 131–142.
- Daganzo, C. F. 1997. *Fundamentals of Transportation and Traffic Operations*. Oxford, U.K.: Pergamon-Elsevier.
- Dailey, D. J. 1999. “A Statistical Algorithm for Estimating Speed from Single Loop Volume and Occupancy Measurements.” *Transportation Research Part B: Methodological* 33 (5): 313–322.
- Edie, L. C. 1963. “Discussion of Traffic Stream Measurements and Definitions.” In *Proceedings of the 2nd International Symposium on the Theory of Traffic Flow*, Paris, France. 139–154.
- Han, J., Polak, J. W., Barria, J., and Krishnan, R. 2010. “On the Estimation of Space-Mean Speed from Inductive Loop Detector Data.” *Transportation Planning and Technology* 33 (1): 91–104.
- HCM. 2000. *Highway Capacity Manual*. TRB, Pub No: LP-608.
- Huey, R., De Leonardis, D., and Freedman, M. 2012. *National Traffic Speeds Survey II: 2009*. Alexandria, VA 22312 USA: National Technical Information Service.
- Jamshidnejad, A., and De Schutter, B. 2014. *Estimation of the Generalized Traffic Average Speed based on Microscopic Measurements: Addendum. Tech. Rep. 14-013*. Delft Center for Systems and Control, Delft University of Technology. Delft, The Netherlands. [http://www.dsc.tudelft.nl/~bdeschutter/pub/rep/14\\_013.pdf](http://www.dsc.tudelft.nl/~bdeschutter/pub/rep/14_013.pdf).
- Klein, L. A., Mills, M. K., and Gibson, D. R. 2006. *Traffic Detector Handbook*. 3rd ed. U.S. Department of Transportation, Federal Highway Administration.
- Ma, Y., van Dalen, J., de Blois, C., and van Nuen, J. 2010. “Estimating Dynamic Transport Population for Official Statistics Based on GPS/GSM.” In *Proceedings of the 7th International Conference on Traffic and Transportation Studies (ICTTS)*, Kunming, China. 1076–1088. Aug. 3–5.
- May, A. D. 1990. *Traffic Flow Fundamentals*. New Jersey, U.S.: Prentice Hall.

- Mori, U., Mendiburu, A., Álvarez, M., and Lozano, J. A. 2015. “A Review of Travel Time Estimation and Forecasting for Advanced Traveller Information Systems.” *Transportmetrica A: Transport Science*, 11 (2): 119–157.
- Newell, G. F. 1995. *Theory of Highway Traffic Flow 1945-1965*. University of California at Berkeley: Transportation Research Board.
- Rakha, H., and Zhang, W. 2005. “Estimating Traffic Stream Space-Mean Speed and Reliability From Dual and Single Loop Detectors.” *Transportation Research Record* (1925): 38–47.
- Soriguera, F., and Robusté, F. 2011a. “Highway Travel Time Accurate Measurement and Short-Term Prediction Using Multiple Data Sources.” *Transportmetrica* 7 (1): 85–109.
- Soriguera, F., and Robusté, F. 2011b. “Estimation of Traffic Stream Space Mean Speed from Time Aggregations of Double Loop Detector Data.” *Transportation Research Part C* 19 (1): 115–129.
- Treiber, M., and Kesting, A. 2013. *Traffic Flow Dynamics*. Berlin Heidelberg: Springer.
- Wang, Y., and Nihan, N. L. 2003. “Can Single-Loop Detectors Do the Work of Double-Loop Detectors.” *Journal of Transportation Engineering* 129 (2): 169–176.
- Wardrop, J. G. 1952. “Some Theoretical Aspects of Road Traffic Research.” *Proceedings of the Institute of Civil Engineers* 1: 325–378.
- Zou, Y., Zhang, Y., and Zhu, X. 2014. “Constructing a Bivariate Distribution for Freeway Speed and Headway Data.” *Transportmetrica A: Transport Science* 10 (3): 255–272.