Transportation Research Part B 75 (2015) 19–35

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

A simple algorithm for the estimation of road traffic space mean speeds from data available to most management centres

Margarita Martínez-Díaz *, Ignacio Pérez [1]

Universidade da Coruña, E.T.S.I. Caminos, Canales y Puertos, Campus de Elviña, 15071 A Coruña, Spain

ARTICLE INFO

Article history:
Received 2 September 2014
Received in revised form 3 February 2015
Accepted 4 February 2015
Available online 2 March 2015

Keywords:
Space mean speed
Time mean speed
Travel time
Traffic loop detector
Stationarity
Log-normality

ABSTRACT

The control of the evolution of road traffic streams is highly related to productivity, safety, sustainability and, even, comfort. Although, nowadays, the findings from research efforts and the development of new technologies enable accurate traffic forecasts in almost any conditions, these calculations are usually limited by the data and the equipment available. Most traffic management centres depend on the data provided, at best, by double-loop detectors. These loops supply time means over different aggregation periods, which are indiscriminately used as the bases for subsequent estimations. Since space mean speeds are those needed in most applications (note the fundamental relationship between flow and density in traffic flow theory), most current practice begins with an error. This paper introduces a simple algorithm that the allows estimation of space mean speeds from the data provided by the loops without the need for any additional financial outlay, as long as the traffic in each time interval of aggregation is stationary and its speed distribution is log-normal. Specifically, it is focused on the calculation of the variance of the speeds with regard to the time mean, thus making possible to use the relationship between time mean speeds and space mean speeds defined by Rakha (2005). The results obtained with real data show that the algorithm behaves well if the calculation conditions help fulfil the initial hypotheses. The primary difficulties arise with transient traffic and, in this case, other specific methodologies should be used. Data fusion seems promising in this regard. Nevertheless, it cannot be denied that the improvement provided by the algorithm turns out to be highly beneficial both when used alone in the case of stationarity or as a part of a fusion.
© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

As society evolves, new requirements and needs may appear. With regard to road transport, researchers, administrations and private companies are aware that controlling the evolution of traffic results in an increase of productivity and safety, allows the exploitation of synergies among different means of transport and contributes to a more sustainable growth (SHRP2, 2013). Many different initiatives, such as real time calculation of travel times, active traffic management procedures or automated driving systems, emerge as examples of these key achievements.

Although these lines of research are very different, they share two commonalities: the need for appropriate data and well-founded calculations. The development of new technologies and computer software offers the possibility of collecting varied data and combining all this information in order to obtain accurate results (Yuan et al., 2014). Mobile phones, GPS (Global

* Corresponding author. Tel.: +34 981 167 000x6031.
E-mail addresses: margarita.martinez@udc.es (M. Martínez-Díaz), iperez@udc.es (I. Pérez).
[1] Tel.: +34 981 167 000x1451.

http://dx.doi.org/10.1016/j.trb.2015.02.003
0191-2615/© 2015 Elsevier Ltd. All rights reserved.

© <2015>. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
https://creativecommons.org/licenses/by-nc-nd/4.0/

Positioning System), Bluetooth, Optical Character Recognition (OCR) cameras and many other devices arise as invaluable sources of traffic data which can later be used for different calculations. As an example, GPS-enabled mobile phones have meant a new way of collecting traffic data as they are able to register vehicle trajectories (Hiribarren and Herrera, 2014). Unfortunately, neither totally accurate data nor the most complex programs are usually available (at least, in a sufficient amount) in less trafficked areas, such as, for example, secondary roads, rural areas or for small traffic management centres. In fact, the majority of these centres located in developed countries depend on pieces of equipment such as loop detectors and regular cameras (which are unable to identify vehicles). In these situations, loops are the main sources of data. Traffic researchers have clearly demonstrated the advisability of deploying double loops (in pairs in each section of each lane) rather than single loops in order to obtain a higher amount of data and, therefore, better results in subsequent calculations (Chen et al., 2003). Luckily, this trend is, at present, usually followed.

All inductive loop detectors are similar. They consist of a wire loop installed under the pavement of a lane which is able to detect the presence of a vehicle (in essence, a metallic object) thanks to the changes that it causes in the electromagnetic properties of the loop. The main differences are related to the software which manages and stores these data, as it can be programmed in several and various ways. When applying the double-loop configuration, the data usually available for previously determined time intervals of aggregation are:

- number of vehicles that pass over the detectors,
- lengths of these vehicles: the software that manages the information usually classifies them into groups and only stores the number of vehicles for each group. For example, in Spain the usual classification is as follows: groups of vehicles shorter than 6 m, between 6 and 10 m and larger than 10 m,
- spot speed measurements: again, and although individual spot speeds are initially detected, the software only calculates and registers their mean, i.e., the time mean speed, the average speed of all the vehicles passing over a particular spot,
- number of vehicles that pass over the detectors with a speed lower than a particular reference speed. In this case, it is usual to have two different references. Only the number of vehicles that meet this requirement is stored. It must be highlighted that the chance of obtaining these data directly from the software of loop detectors is not a standard in the USA, but it is quite common in Europe. As an example, all Spanish freeway traffic centres manage them.

The duration of the time intervals of aggregation ranges from 20 to 30 s in the USA up to 15 min in some European countries. Intervals between 3 and 5 min have proven to be the most suitable (Soriguera and Robusté, 2013). Nevertheless, both shorter and longer durations have some advantages and disadvantages, as discussed in Section 5.

Variation of traffic speeds at various places over the time turns out to be one of the basic inputs for subsequent studies, such as, for instance, the indirect estimation of travel times. However, the main problem is that most studies are based on the fundamental equation of traffic flow (Eq. (1)). This equation provides the relationship between flow q and density k through a specific type of speed, the so-called space mean speed $\overline{v_s}$, which is, in fact, a harmonic mean calculated under specific conditions (Wardrop, 1952). Further explanation about this point is included in Section 2.

$$q = \overline{v_s} k \tag{1}$$

The use of the data provided by loop detectors involves various difficulties when determining the evolution of speeds:

- individual speeds are measured at fixed points of a road and must be extrapolated to some extent, in order to achieve the spatial implication needed. This spatial generalisation is extremely complicated, particularly in the case of congestion,
- as mentioned, the software usually provides time mean speeds. The use of these time means as substitutes of the space means required for calculations can cause a considerable loss of accuracy in the results,
- although loops are simpler, more economical and more common than other devices commonly used to collect traffic data, their usefulness depends on their density on the road (Bachmann et al., 2013). Some invaluable research has resulted in the development of simple search algorithms which efficiently select the sensor locations in order to obtain suitable data when the number of available sensors is limited (Viti et al., 2014). Nevertheless, difficulties still remain in those roads already constructed.

The algorithm introduced in this paper is aimed at calculating spot space mean speeds exclusively from the data provided by double-loop detectors, thus avoiding extra expenses for governmental agencies and other related bodies. Specifically, it is focused on the calculation of the variance of the speeds with regard to the time mean, thus allowing, in the case of stationarity, the application of the relationship between time mean speeds and space mean speeds defined by Rakha (2005). Further improvements must be implemented in order to reach the final objectives and goals (for example, their extrapolation to entire links), but any lack of accuracy with regard to this first basic factor will already spread to any other result that may depend on it. The next sections of this paper cover the following aspects: Section 2 provides the background of different traffic speed definitions and summarises their relationships according to various researchers; Section 3 develops the proposed algorithm, whereas its implementation with artificial and real data is demonstrated in Section 4, as well as compared with other methodologies; Section 5 discusses the results, while conclusions and a proposal for new lines of research are presented in Section 6.

2. Background

Since 1952, when Wardrop stated his two principles concerning the idea of traffic equilibrium previously developed by Knight (1935), the differences between the time mean speed and the space mean speed have been widely demonstrated. The space mean speed $\overline{v_s}$ is the average speed of all the vehicles in a particular stretch of a road at a specific instant (Homburger et al., 1996). The time mean speed $\overline{v_t}$ is the average of the speeds of all the vehicles that pass over a section of a road during a certain time interval. It is easy to deduce that the time mean speed is greater than the space mean speed (Daganzo, 1997) since fast vehicles contribute more to the time-mean than the slow ones, whereas vehicles of all speeds equally contribute to the space-mean. Space averages equal time averages only in the case of space–time homogeneous traffic (Breiman, 1969).

As it has been explained before, loops in a road detect and average spot speeds in stipulated time intervals, thus also providing time mean speeds. However, if individual spot speeds were stored, $\overline{v_s}$ could be calculated by giving them a certain spatial nature and, also, by assuming stationary traffic in the section (Edie, 1965), as Eq. (2) shows:

$$\overline{v_s} = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} tt_i} = \frac{ndx}{\sum_{i=1}^{n} \frac{dx}{v_i}} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} \frac{1}{v_i}} \tag{2}$$

where

x_i = distance covered by vehicle i,
tt_i = time used by vehicle i to cover the distance x_i,
v_i = spot speed of vehicle i,
n = number of vehicles that pass over the detector during the time interval,
dx = differential length taken up by the detector.

Therefore, in these conditions the space mean speed could be calculated as the harmonic mean of the individual spot speeds. It must be highlighted, however, that, in the origin of this formulation, neither a time mean nor a space mean was established, but a generalised definition of the average speed. The fact of labelling this generalised definition of the average speed as space mean speed $\overline{v_s}$ is an abuse of notation. Actually, $\overline{v_s}$ does not share the spatial implications of the original space mean speed definition unless traffic is stationary. Some limitations have been imposed for that reason, considering that this identification is only performed when the average speed is computed over a narrow rectangular strip in the x–t plane with a spatial width dx and a time length T, which corresponds to the measurement region of a loop detector on a highway. Taking this definition into account, space mean speed appears, for example, in the mathematical definition of the average travel time $\overline{TT}$ of n vehicles that cover a specific distance of a road L at a constant speed v_i (Eq. (3)):

$$\overline{TT} = \frac{\sum_{i=1}^{n} TT_i}{n} = \frac{\sum_{i=1}^{n} \frac{L}{v_i}}{n} = L * \frac{1}{n}\sum_{i=1}^{n} \frac{1}{v_i} = \frac{L}{\overline{v_s}} \tag{3}$$

In consequence, travel times would be underestimated if $\overline{v_t}$ were used instead of $\overline{v_s}$ (Soriguera and Robusté, 2011). This substitution could lead to other inaccuracies such as wrong estimates of jam densities or shock wave speeds (Knoop et al., 2007). The data aggregation process is, in fact, an important source of noise and errors normally present in conventional measures of the traffic state (Coifman, 2014). Many authors have already stated the importance of using time-based or space-based data accurately and correctly, regardless of their source. For example, the inverse of the harmonic mean of instantaneous speeds from probe vehicles results in an unbiased and consistent estimator of the mean segment travel time when sampling by space, but biased upwards when sampling by time (Jenelius and Koutsopoulos, 2015).

Clearly, upgrades in the loop software would allow these devices to store individual data or even to directly calculate space mean speeds. However, the large number of loops deployed world-wide and human inertia have so far precluded those modifications. Therefore, many researchers have tried to calculate space mean speeds from the time mean speeds provided by the loops, especially in the case of stationarity, which is the common hypothesis of all the following methodologies.

The first of these relationships, shown in Eq. (4), is due to Wardrop (1952):

$$\overline{v_t} = \overline{v_s} + \frac{\sigma_s^2}{\overline{v_s}} \tag{4}$$

where σ_s^2 is the variance of the speed with regard to the space mean for the specific time interval of aggregation which has been chosen. The accuracy of the formula has been experimentally verified, but most traffic management centres cannot use it because individual speeds are needed in order to calculate the variance with regard to the space mean. This formula was actually devised to calculate time means from space means, what is not usually necessary in real life.

Another postulated formula to relate both means is that of Garber (2002), which is shown in Eq. (5):

$$\overline{v_t} = 0.966\overline{v_s} + 3.541 \tag{5}$$

The main problem of this relationship is that it was established based only on experimental data. Thus, it cannot be extrapolated to many situations in which the boundary conditions differ from the original ones. Therefore, it must be continuously calibrated and, ultimately, it is not really worth using.

Eq. (6) has been used in several traffic studies. It was first derived by Khisty (2003), but, in fact, it was Rakha (2005) who proved it analytically:

$$\overline{v_s} = \overline{v_t} - \frac{\sigma_t^2}{\overline{v_t}} \tag{6}$$

In this equation σ_t^2 is the variance of the speed with regard to the time mean for the specific time interval of aggregation which has been chosen. However, the impossibility of calculating the variance arises again. Nevertheless, and taking into account the usefulness of this formula, Soriguera and Robusté (2011) was able to estimate this variance by applying the common hypothesis of stationary traffic to each time interval of aggregation and, additionally, by assuming normality of the speed distribution. The variance with regard to the time mean speed is obtained through Eq. (7) below:

$$\sigma_t = \frac{v^a - \overline{v_t}}{F^{-1}\left[\frac{n_{v^a}}{n}\right]} \tag{7}$$

σ_t = standard deviation of the speed with regard to the time mean,
v^a = value of the speed chosen by traffic management centres,
F^{-1} = inverse of the cumulative standard normal distribution,
n_{v^a} = number of vehicles that pass over the detectors with a speed lower than v^a in each time interval of aggregation,
n = number of vehicles that pass over the detectors in each time interval of aggregation.

Although this methodology shows a good performance in specific conditions, El Faouzi and Maurin (2007) has already pointed out that it is inappropriate to use it indiscriminately, especially in the cases of shock wave onsets and offsets or with "stop and go situations". As Cassidy (1998) has already stated, stationarity ensures some otherwise senseless relationships. However, the relationship established by Rakha (2005) has been proved to be useful under certain conditions, even with non-spot data such as those from GPS (Poomrittigul et al., 2008).

Another fact that must be taken into account in order to establish these relationships between speeds is that they approximately fit common statistical distributions. Thus, an as a general rule, normal, log-normal, gamma and bimodal distributions tend to appear in the majority of traffic studies.

Normal distribution is, undoubtedly, the most widely used because of its simplicity, and it performs well when traffic conditions are homogeneous. Consequently, the assumption of multivariate normal distributions for link travel times tends to be quite common as well (Jenelius and Koutsopoulos, 2013). However, the log-normal and gamma distributions are usually more suitable, as they have additional advantages (Haight, 1962):

- they help to avoid the appearance of negative speeds,
- they keep their shape when either time speeds or space speeds are fitted.

In the case of the log-normal distribution, another important advantage is the fact that the distribution of travel times based on speeds that fit this distribution maintains the same shape (El Faouzi and Maurin, 2007). If the log-normal speed distribution has a mean μ and a standard deviation σ, the distribution of travel times will follow Eq. (8):

$$f_t(t) = \frac{1}{\sqrt{2\pi}\sigma t} e^{\left[-\frac{(Lnt+\mu)^2}{2\sigma^2}\right]} \tag{8}$$

In the cases where the traffic is too heterogeneous (for example, because there are many different vehicle types that may behave differently or because of the existence of phases of free flow after congestion periods), unimodal distributions should be avoided (Dey et al., 2006). In this situation, bimodal or even multimodal distributions might be used instead. Each of their components would often be a normal or a log-normal distribution (May, 1990).

Many other complex distributions have already been applied in different pieces of research, but their high level of complexity prevents them from being put into practice (Zou and Zhang, 2011). Even for log-normal distributions, some improvements can be expected if the distributions are truncated because only a range of speeds makes sense. In addition, the variances of these truncated distributions are always smaller than those of the original ones (Wang, 2012).

3. Simple algorithm for the estimation of space mean speeds from the data provided by double-loop detectors

Having analysed previous investigations, and taking into account the data available, the authors decided to use the equation of Rakha (2005) in order to solve the problem of not having an explicit value of the variance. The reason underlying this motivation is based on the fact that the validity of this formula has been widely demonstrated in experimental studies.

To be able to estimate the variance, two important hypotheses are assumed. In each time interval of aggregation T:

- traffic is stationary,
- the speed distribution is log-normal.

The validity of these hypotheses will be discussed in Section 5. The first one has also been taken for granted in the other methodologies discussed in this paper. With regard to the second, the authors exploit the advantages of the log-normal distribution mentioned in Section 2. Assuming that the distribution of individual speeds v_i in each time interval of aggregation T is log-normal, the distribution of the logarithms of these speeds $x = Ln\,v$ is a normal distribution $N(\mu_x, \sigma_x)$. Therefore, the probability density function of the speeds, their mean and their variance are given by Eqs. (9)–(11), respectively:

$$f_v(v) = \frac{1}{\sqrt{2\pi}\sigma_x v} e^{\left[-\frac{(Ln\,v-\mu_x)^2}{2\sigma_x^2}\right]} \text{ con } v > 0 \tag{9}$$

$$\mu_v = \overline{v_t} = e^{\mu_x} + \frac{\sigma_x^2}{2} \tag{10}$$

$$\sigma_v^2 = \sigma_t^2 = \left(e^{\sigma_x^2} - 1\right) * e^{2\mu_x + \sigma_x^2} \tag{11}$$

where
v = individual speed,
μ_x = arithmetic mean of the logarithms of the speeds,
σ_x^2 = variance of the logarithms of the speeds with regard to the mean.

It is important to remember that this algorithm is aimed at estimating σ_v^2, which corresponds to the variance with regard to the time mean speed that Rakha (2005) named σ_t^2. Therefore, μ_x and σ_x are needed. μ_v is provided by the loops (the time mean speed, named $\overline{v_t}$ by Rakha (2005)).

Let n_v^a be the number of vehicles that pass over the detectors in a section with a speed lower than v^a in one time interval of aggregation T. The probability that a vehicle passes over the detector with such a speed is shown in Eq. (12):

$$P[V \leqslant v^a] \approx \frac{n_v^a}{n} \approx P\left[e^X \leqslant e^{x^a}\right] \approx P\left[Lne^X \leqslant Lne^{x^a}\right] \approx P[X \leqslant x^a] = F[Z(x^a)] = F\left[Z(Ln\,v^a)\right] = F\left[\frac{Ln\,v^a - \mu_x}{\sigma_x}\right] \tag{12}$$

where:
v^a = speed chosen as a reference,
n = number of vehicles that pass over the detectors in each time interval of aggregation,
x^a = logarithm of the speed v^a,
F = cumulative standard normal distribution,
$Z_{(\ )}$ = standardised value.
By rearranging Eqs. (10) and 12, a system with two equations (Eqs. (13) and (14)) and two unknowns is obtained:

$$2\mu_x + \sigma_x^2 = Ln\overline{v_t}^2 \tag{13}$$

$$\mu_x + F^{-1}\left[\frac{n_v^a}{n}\right]\sigma_x = Ln\,v^a \tag{14}$$

where
F^{-1} = inverse of the cumulative standard normal distribution.
Finally, Eq. (15) is obtained:

$$\sigma_x^2 - 2F^{-1}\left[\frac{n_v^a}{n}\right]\sigma_x + Ln\left(\frac{v^a}{\overline{v_t}}\right)^2 = 0 \tag{15}$$

When solving Eq. (15), two possible values of σ_x arise. For two reference values of the speed (v^{a1} and v^{a2}), four values are provided. In practice, some of these are nullified during the calculations due to the fact that some mathematical limitations for the algorithm arise. In each time interval of aggregation T:

- n cannot be too small or, otherwise, the initial substitution of the theoretical probability for the accumulated frequency (Eq. (12)) may be problematic and the confidence interval of the estimations is too small,
- the conditions $n_v^a \neq 0$ and $n_v^a \neq n$ are necessary. This prevents the inverse of the cumulative standard distribution from tending to infinite,
- $\left(F^{-1}\left[\frac{n_v^a}{n}\right]\right)^2$ must be greater than $Ln\left(\frac{v^a}{\overline{v_t}}\right)^2$ to avoid square roots of negative numbers when solving Eq. (15),
- the condition $\frac{v^a}{\overline{v_t}} \neq 0$ is necessary in order to avoid natural logarithms of zero.

In those cases where more than one value of σ_x is obtained, an action protocol must be established in order to help to choose the most suitable. One possibility is to keep the value with the smallest confidence interval for a specific level of confidence.

When a value of σ_x is found and introduced into Eq. (13), the corresponding μ_x can be calculated. By using both values in Eq. (11), σ_t^2 is finally obtained and can be introduced into Eq. (6) in order to estimate $\overline{v_s}$. The flow chart in Fig. 1 summarises the main steps of the algorithm.

As it can be deduced, it is not easy in practice to choose the best estimate of σ_t^2 from more than one possible value. There are no simple methods to calculate the confidence intervals for the variance of log-normal distributions. Bayesian procedures seem to be the most suitable (Harvey and van der Merwe, 2012), although quite difficult to implement.

A naive solution could be to apply the confidence intervals of a parameter calculated in a previous step of the method, for example σ_x. If the best value of σ_x is chosen, the best σ_t^2 and, thus, a more accurate $\overline{v_s}$ will be obtained. Therefore, and thanks to the fact that the variable x is normally distributed, the solution for the confidence interval limits of σ_x proposed by Soriguera and Robusté (2011) and developed in Eqs. (16) and (17) can be used:

$$\begin{aligned} \varepsilon_{\sigma_{x(1)}} &= -\frac{(v^a-\mu_x)\varepsilon_{z(1)}}{z\left(z+\varepsilon_{z(1)}\right)} \\ \varepsilon_{\sigma_{x(2)}} &= -\frac{(v^a-\mu_x)\varepsilon_{z(2)}}{z\left(z+\varepsilon_{z(2)}\right)} \end{aligned} \tag{16}$$

where

$$\begin{aligned} \varepsilon_{Z(1)} &= F^{-1}\left(p+\varepsilon_p\right)-F^{-1}(p) \\ \varepsilon_{Z(2)} &= F^{-1}\left(p-\varepsilon_p\right)-F^{-1}(p) \end{aligned} \tag{17}$$

Variable p corresponds to the probability of a vehicle with a speed lower than v^a passing over the detector in the time interval of aggregation. The circulation of vehicles over the detectors can be considered as a Bernouilli process; the possibilities are their driving slower than a reference speed or not, these trials being independent. Thus, the estimator of p ($\hat{p}$) matches Eq. (18):

$$\hat{p} = \frac{n_v^a}{n} \tag{18}$$

The proposed methodology heavily relies on the availability of n_v^a. If n_v^a is not reported to the traffic management centre during the normal operation of the system, the method cannot be applied. Obviously, carrying out modifications in the controllers in order to obtain these data lacks any sense, as it would be simpler, in this case, to introduce other modifications in

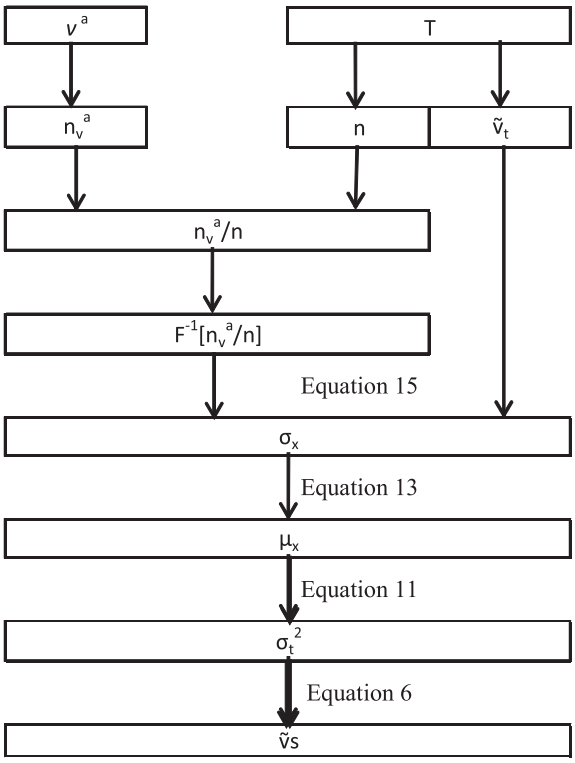

Fig. 1. Steps of the algorithm.

order to directly obtain $\overline{v_s}$. Nevertheless, in the substantial number of countries where n_v^a is available, the fact of using the estimated $\overline{v_s}$ instead of working with $\overline{v_t}$ (the current procedure) when making subsequent calculations would imply a higher level of accuracy without the need of any re-coding.

4. Implementation of the algorithm with artificial and real data

4.1. Implementation with artificial data

In order to first verify the proper performance of the algorithm, it was successfully tested with data generated with Matlab and readjusted to fulfil both the main hypotheses of the method (i.e., the stationarity of the traffic and the log-normality of the speed distribution in each time interval of aggregation T) and the mathematical requirements detailed in Section 3. For this very last reason, the reference values were set at 101 km/h and 110 km/h (90% and 98% of the total time mean speed), thus ensuring the participation of enough vehicles in the calculations. The steps followed and the results are shown in Table 1, whereas Fig. 2 shows them in comparison with time means and real space mean speeds.

The estimated space mean speeds are much closer to the real space mean speeds than the time mean speeds the loops provide. The error introduced by the latter is 2.17%, compared to 0.65% for the estimations of the algorithm. The validity of the algorithm has been therefore demonstrated in these ideal conditions. The mean relative error was calculated taking into account the absolute values of the differences. In addition, and with regard to the estimated space means, only values with differences smaller than the maximum difference reported by the loops were admitted. This procedure was also followed in Section 4.2 with real data.

4.2. Implementation with real data

The validity of the algorithm has been demonstrated in an ideal situation where all the initial conditions that were assumed when defining the method were met. However, it is also necessary to test it with different combinations of real data for which one or more of these conditions probably will not apply.

The data used for this study were collected during two days, on March 31th and April 1st, 2014 in a section with double loops (P.K. 86 + 211, with two lanes in the direction of A Coruña) of the AP-9 freeway, which runs north and south along the west coast of Galicia in Spain. The data were provided per lane and for aggregation time intervals T of 15 min. During the normal management of this freeway, the common data available are:

- number of vehicles that pass over the loops (n),
- number of vehicles with lengths L shorter than 6 m, between 6 and 10 m or longer than 10 m,
- time mean speeds $\overline{v_t}$: at an initial stage these speeds are averaged every 5 min, but then they are smoothed for time intervals of 15 min,
- number of vehicles (n_v^a) that pass over the loops with speeds lower than 50 km/h (V^{a1}) and 100 km/h (V^{a2}), respectively.

Specifically for investigation purposes, however, on this occasion the individual speeds and lengths were also provided, thus allowing an analysis of the algorithm with a wide range of different boundary conditions, as well as a comparison of the estimated space mean speeds with the real ones. The algorithm was executed with data obtained on different days, in different lanes (the left, for the fastest vehicles, and the right, for medium–low speed vehicles) and for all vehicles or only those whose lengths L were within a specified range. In addition, different time intervals of aggregation (T, in minutes) and reference speeds (V^{a1} and V^{a2}) were used. N corresponds to the number of vehicles detected during the entire data acquisition period. Table 2 below shows the cases which have been analysed:

In addition, it is also important to point out that neither the stationarity of the traffic flow nor the log-normality of the speeds is guaranteed. This issue will be discussed in Section 5.

4.2.1. Results obtained with real data

Table 3 shows the difference between using the time mean speeds provided by the loop detectors or the space mean speeds estimated with the algorithm as the substitute for real space mean speeds. This difference is shown here as it has been previously done in Section 4.1, i.e., by determining the mean relative error in each case.

In 8 out of the 11 cases analysed (and taking into account that case V has been subdivided) the application of the algorithm implies an improvement, but there are 2 cases where the results have been worse and another in which no reasonable value has been obtained. This behaviour, which was also analysed and understood, is discussed in Section 5.

Besides, it is worth pointing out that, in most cases, it is not possible to determine the validity of the algorithm by focusing only on one of the boundary conditions; thus, attention to the combination of all of them is required. Nevertheless, once all the conditions for the calculation have been established, its performance can be improved by changing only one of these conditions. As an example, between cases VI (Fig. 3) and VII (Fig. 4) only the reference speeds are different. However, the algorithm only shows a good performance in the latter case.

Table 1
Estimation of space mean speeds and comparison of the results obtained with the data provided by the loops and with real values.

Time period T	Number of vehicles n	Classification of vehicles according to the speeds			n_{va}/n		Time mean speed (km/h)	Estimated space mean speed (km/h)	Real space mean speed	Difference time mean – real space mean	Error (%)	Difference estimated space mean – real space mean	Error (%)
		$V < 101$	$101 \leqslant V < 110$	$V \geqslant 110$	$V^a = 101$	$V^a = 110$							
7:30:00	28	0	4	24.00	0.0000	0.1429	123.56	122.22	122.47	1.09	0.89	−0.25	0.20
7:45:00	25	3	9	13.00	0.1200	0.4800	111.10	110.41	110.30	0.79	0.72	0.11	0.10
8:00:00	33	11	8	14.00	0.3333	0.5758	106.51	105.23	105.62	0.89	0.84	−0.39	0.37
8:15:00	32	1	11	20.00	0.0313	0.3750	114.48	113.98	113.58	0.89	0.79	0.39	0.35
8:30:00	36	1	5	30.00	0.0278	0.1667	119.43	118.55	118.41	1.02	0.86	0.14	0.12
8:45:00	45	2	4	39.00	0.0444	0.1333	124.11	122.77	122.67	1.45	1.18	0.10	0.08
9:00:00	36	9	10	17.00	0.2500	0.5278	110.18	108.62	108.55	1.63	1.50	0.07	0.06
9:15:00	51	14	18	19.00	0.2745	0.6275	107.80	106.71	106.87	0.93	0.87	−0.16	0.15
9:30:00	43	7	14	22.00	0.1628	0.4884	111.38	110.37	110.05	1.32	1.20	0.32	0.29
9:45:00	39	11	9	19.00	0.2821	0.5128	108.55	107.13	106.93	1.62	1.52	0.21	0.19
10:00:00	31	8	6	17.00	0.2581	0.4516	112.33	109.87	108.34	3.99	3.68	1.53	1.41
10:15:00	22	10	2	10.00	0.4545	0.5455	104.23	101.48	100.36	3.87	3.86	1.11	1.11
10:30:00	32	18	3	11.00	0.5625	0.6563	103.71	99.80	95.59	8.12	8.49	4.21	4.40
10:45:00	29	4	11	14.00	0.1379	0.5172	110.39	109.70	109.61	0.78	0.71	0.09	0.08
11:00:00	16	5	1	10.00	0.3125	0.3750	114.11	112.97	111.29	2.82	2.54	1.69	1.52
11:15:00	24	2	6	16.00	0.0833	0.3333	114.12	113.28	112.84	1.28	1.14	0.44	0.39
11:30:00	26	8	3	15.00	0.3077	0.4231	111.88	108.45	108.36	3.52	3.25	0.10	0.09
11:45:00	29	7	6	16.00	0.2414	0.4483	116.54	109.99	110.89	5.65	5.09	−0.90	0.81
Total vehicles	577.0	121.0	130.0							Mean error	2.05		0.59

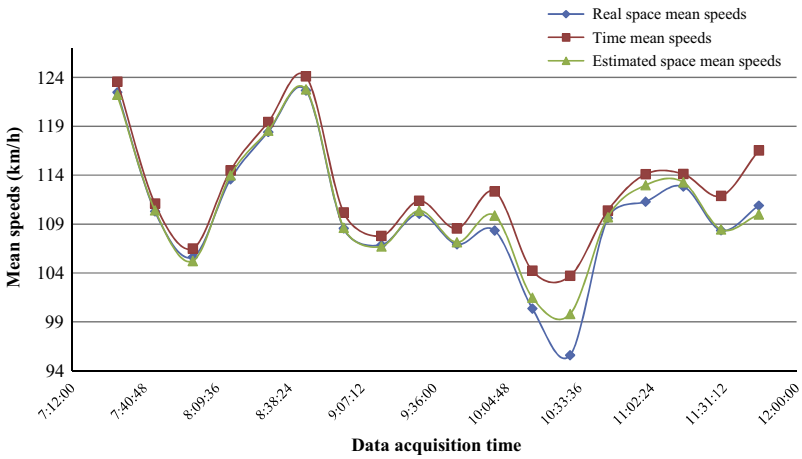

Fig. 2. Comparison of the real space mean speeds, the time mean speeds and the space mean speeds estimated with the algorithm from data that completely fulfil the initial conditions of the method.

Table 2
Cases analysed to test the algorithm.

Case	Day	Lane	T (')	L	N	V^{a1}	V^{a2}
I	31 March	right	15	all	4662	50	100
II	01 April	right	15	all	2841	50	100
III	01 April	right	15	all	2841	98	107
IV	01 April	right	5	all	2841	50	100
V	01 April	right	5	$L < 10$ m	2489	50	100
				$L \geqslant 10$ m	352	50	100
VI	31 March	left	15	all	769	50	100
VII	31 March	left	15	all	769	110	120
VIII	01 April	left	15	all	596	50	100
IX	01 April	left	5	all	595	50	100
X	01 April	left	5	all	595	50	115

Table 3
Comparison between the errors derived from the use of time means and those of the algorithm.

Case	I	II	III	IV	Va	Vb	VI	VII	VIII	IX	X
Weighted error derived from the use of time means (%)	1.35	1.19	1.21	2.04	1.68	0.27	0.56	0.47	0.59	1.48	0.93
Weighted error of the algor. (%)	0.79	0.87	0.99	0.59	0.46	–	0.86	0.44	0.78	0.58	0.50

The reason underlying this fact is that, in case VI, the sample includes fewer vehicles, since most of them were driving at speeds higher than 50 km/h. Another example is based on cases IV (Fig. 5) and V (Fig. 6). The segregation of the sample in accordance with the vehicle length usually improves the performance for light vehicles, as the hypothesis of log-normality is met in a more complete way. As for heavy vehicles, the algorithm, in this specific example, does not even run due to the small size of the sample. The influence of the length of the time interval of aggregation can be observed, for example, in cases II and IV (Figs. 7 and 5). The results of case IV, where $T = 5$ min, are, undoubtedly, much better.

4.2.2. Comparison between the proposed methodology, the algorithm based on the normal distribution and previous methods.

Because the proposed algorithm is somewhat more complicated than that introduced by Soriguera and Robusté, 2011, a comparative analysis was performed to verify if it is worth using. In case I, for example, the proposed algorithm demonstrated good behaviour, thus diminishing by 0.58% the error derived from the use of time mean speeds. Table 4 and Fig. 8 compare these results with that obtained following the methodology of Soriguera, which, as mentioned before, assumes normality and stationarity in each time interval of aggregation T.

Only for comparison purposes and being conscious of the dependence of the formula developed by Garber on the boundary conditions, Table 4 also includes the results that would be obtained from its application. The equation of Wardrop, as it has previously been stated, is useful only to calculate $\overline{v_t}$ from $\overline{v_s}$, what is not necessary in practical uses.

It can be observed that the use of the algorithm proposed in this paper is more advisable, mainly because, as explained in Section 2, log-normal distributions represent the character of the speeds more accurately. It can also be observed that, in

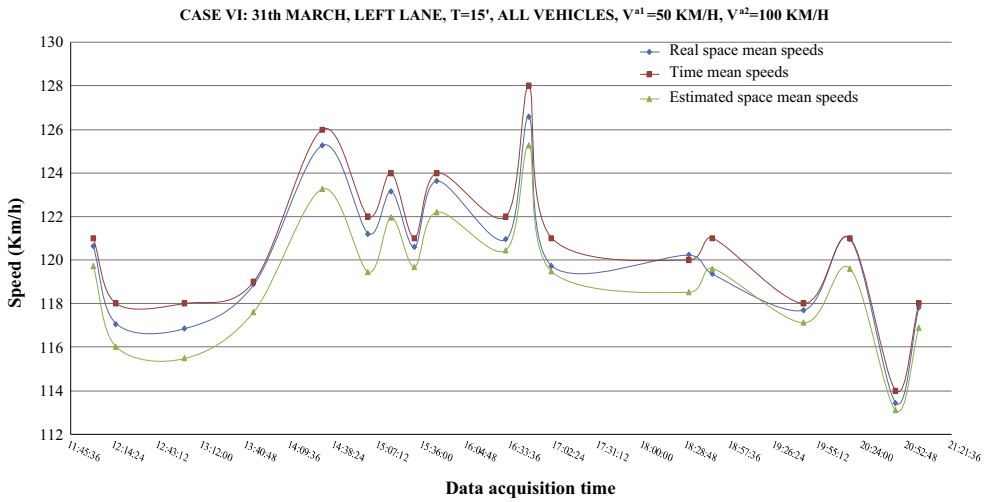

Fig. 3. Comparison of the real space mean speeds, the time mean speeds and the space mean speeds estimated with the algorithm in case VI.

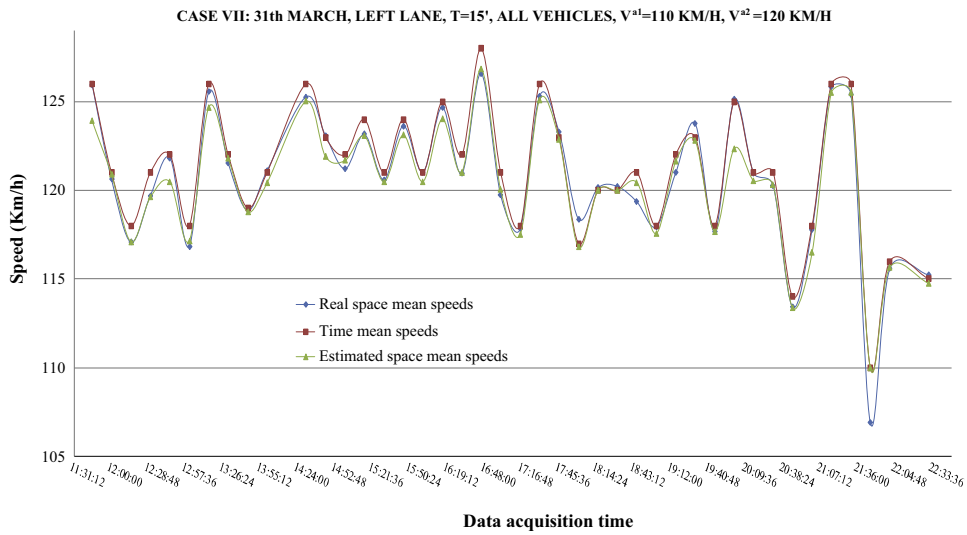

Fig. 4. Comparison of the real space mean speeds, the time mean speeds and the space mean speeds estimated with the algorithm in case VII.

spite of the mathematical limitations of the proposed methodology (discussed in Section 3), the final number of vehicles involved in the calculations is greater when assuming log-normality. Besides, and as it had been expected, it would be even better to substitute space mean speeds for the time mean speeds directly provided by the loops, rather than using the results obtained from Eq. (5).

5. Discussion

Based on the accuracy of the estimates obtained in each case, some conclusions have been drawn. It seems that the algorithm is worth using in numerous situations, as the results are usually more accurate than the currently accepted values of time mean speeds. However, while it clearly shows a better performance in some of these cases, it does not work very well in others.

The analysis was carried out taking into account the following boundary conditions:

- sample size,
- log-normality of the speed distribution,
- speeds chosen as references,
- length of the time interval of aggregation,

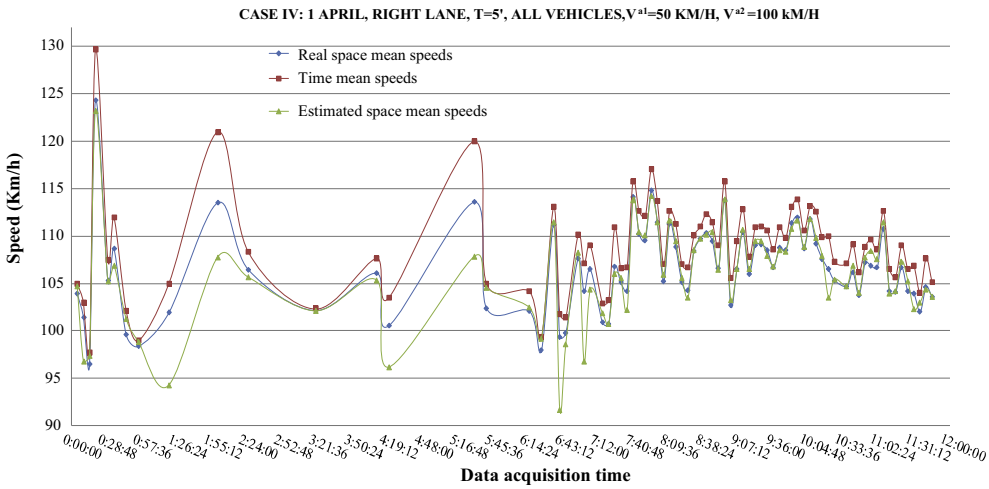

Fig. 5. Comparison of the real space mean speeds, the time mean speeds and the space mean speeds estimated with the algorithm in case IV.

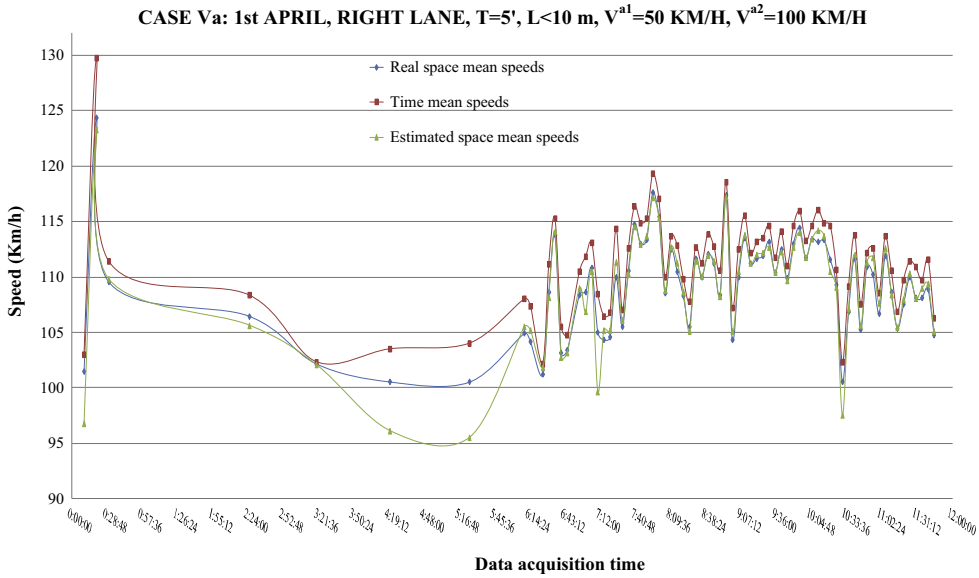

Fig. 6. Comparison of the real space mean speeds, the time mean speeds and the space mean speeds estimated with the algorithm in case Va.

- prevailing type of vehicles,
- general traffic conditions,
- place, day and moment of data acquisition.

Regarding the sample size, the larger the sample, the better the performance of the algorithm. The main reasons underlying this behaviour are the increase in the probability of having a log-normal distribution of speeds in each time interval of aggregation and the reduction in the emergence of mathematical inconsistencies during the calculations.

The log-normality of the speed distribution in each time interval of aggregation is one of the main hypotheses of the method, and, therefore, it must hold. Depending on the conditions established for the calculations, it may be more or less difficult to meet this hypothesis. For example, with low traffic densities, the behaviours of fast (cars) and slow (trucks, buses, vans) vehicles can be very different (Dey et al., 2006). If the estimation is made with samples from all lanes, however, bimodal or even multimodal distributions will probably appear. Therefore, the analysis must be carried out by lane (Soriguera and Robusté, 2011). Nevertheless, with high-medium densities, log-normality could appear even in the whole section because faster vehicles will not be able to reach their usual speeds. As previously mentioned, log-normality is more suitable when working with large samples.

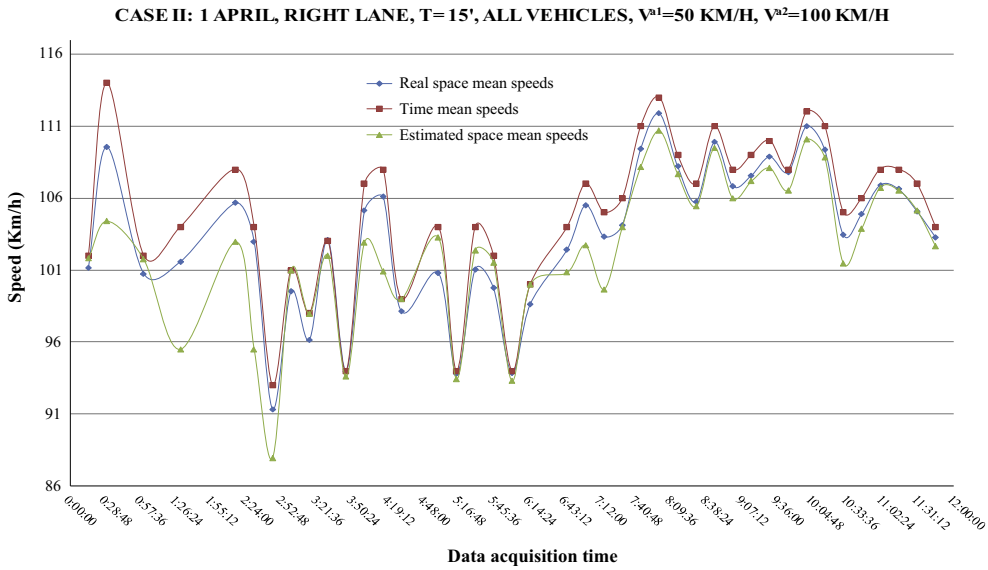

Fig. 7. Comparison of the real space mean speeds, the time mean speeds and the space mean speeds estimated with the algorithm in case II.

Table 4
Comparison of the errors introduced by different methodologies in case I.

Methodology	Vehicles suitable for calculations		Weighted mean error (%)
	Number	% of total vehicles	
Use of time mean speeds directly delivered by loop detectors	4662	100.00	1.35
Use of the equation (Eq. (5)) proposed by Garber (2002)	4662	100.00	1.56
Use of the algorithm (Eq. (7)) proposed by Soriguera and Robusté (2011)	4547	97.53	1.05
Use of the algorithm (Eq. (6) plus Eq. (15) and precedent) proposed in this paper	4628	99.27	0.79

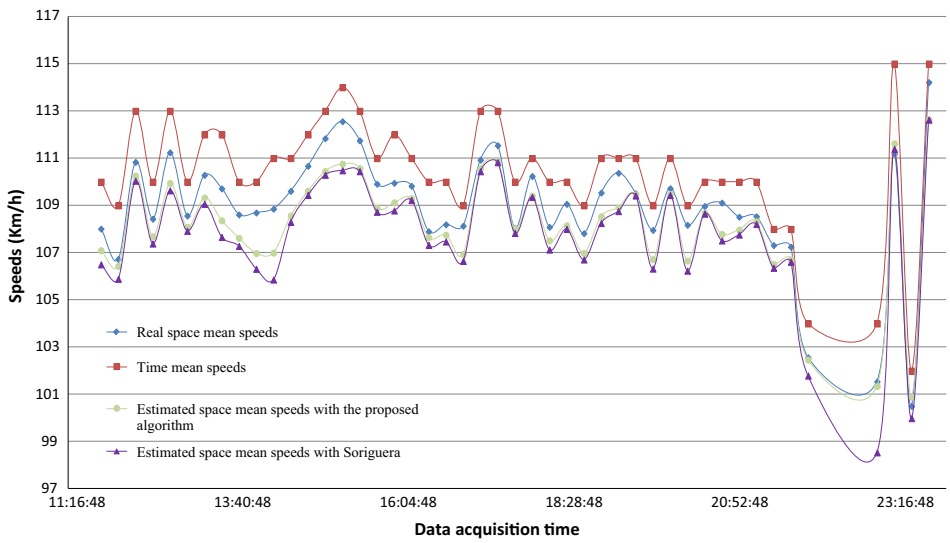

Fig. 8. Comparison of the real space mean speeds, the time mean speeds and the space mean speeds estimated with the proposed algorithm and the algorithm of Soriguera and Robusté (2011) in case I.

To illustrate the importance of fulfilling this hypothesis, two time intervals of $T = 5$ min obtained from case Va were chosen (time intervals between 7.40 and 7.45 a.m. and between 11.10 and 11.15 a.m.). The errors of estimation in these intervals were among the smallest (0.04% and 0.03%, respectively). The logarithms of the speeds were tested with the

Kolmogorov–Smirnov Test. Table 5 shows the results, where the p-value in both cases was greater than 0.05, indicating normality of the logarithms and thus log-normality of the speeds. Figs. 9 and 10 also roughly represent this trend.

The election of the speeds chosen as a reference must be made in a logical way with the specific purpose of having a sufficient number of vehicles in the sample. In the particular case of the AP-9 freeway, the values used were 50 and 100 km/h. Since it is obviously uncommon for a vehicle to drive slower than 50 km/h on a freeway, some data will still be missed. And since the individual speeds were also available, other values were chosen for some of the analyses, what consequently led to better results. In this research, values of 90% and 98% of the average speed were chosen. In practice, these values could be based on (recent) historical data.

As for the lengths of the time intervals of aggregation, both long and short intervals show advantages and disadvantages. Short durations are more likely to meet the other main hypothesis of this method (i.e. the stationarity of the traffic flow), as well as to yield more accuracy in subsequent calculations in real time (for example, in travel time calculations).

On the contrary, longer periods involve a greater sample size and a lower need for calculation capacity, since only a smaller number of iterations will be run each day.

Again, the prevailing type of vehicle is related to the convenience of making the estimations per lane or in a whole section in order to ensure the appearance of log-normal distributions. If possible, it is always advisable to work per lane and, even, to divide the vehicles into groups by their usual speeds, although this last step obviously adds some extra effort. In the case of working per lane, later estimates for the section can be obtained with equations such as Eq. (19), where the superscript i corresponds to the lanes of the section (Soriguera and Robusté, 2011):

$$\overline{v_s^{section}} = \frac{1}{\left[\frac{1}{\sum_i \overline{n^i}}\right] \cdot \sum_i \left(n^i / \overline{v_s^i}\right)} \tag{19}$$

A preliminary analysis of the behaviour of each type of vehicle should be carried out in order to avoid wasting time and effort. In this study, the classification of the vehicles into the three different sizes established by the Galician traffic management centre provided the same results as their classification into only two different sizes (presumably the fastest and the slowest ones), or even worse, because of the lack of individuals in some groups.

It is important to point out that the hypothesis of stationarity for the traffic flow has conditioned most of the steps followed when deriving the algorithm, and, therefore, its observance is essential to achieve a good performance. This stationarity is assumed for each (short) time interval of aggregation, and, consequently, it is quite likely to occur. Nevertheless, it should also be taken into account that there will be frequent occasions in which transients (shock waves, stop and go

Table 5
KS test results for two time intervals with accurate estimates.

Test KS	7_45	11_15
Z Kolmogorov–Smirnov	0.481	0.764
P value (bilateral)	0.975	0.604

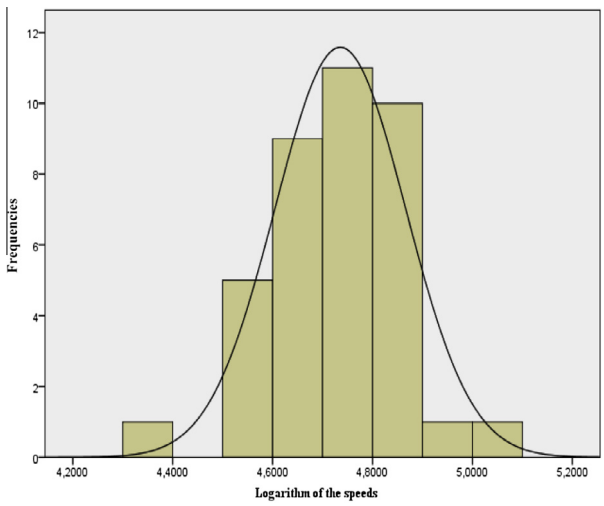

Fig. 9. Log-normal trend for time interval between 7.45 and 7.50 a.m.

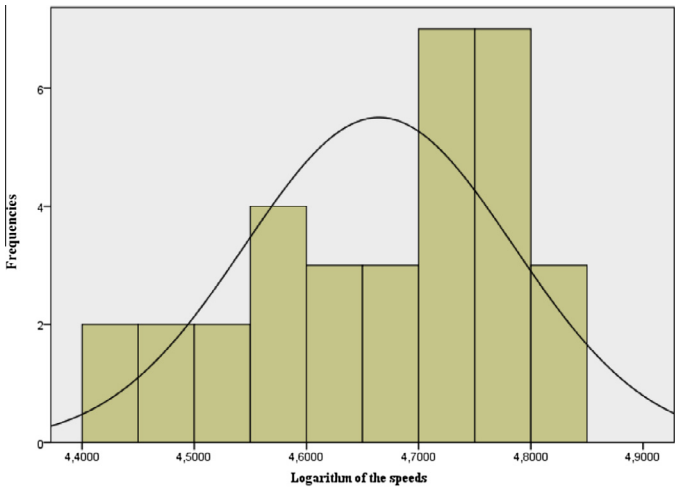

Fig. 10. Log-normal trend for time interval between 11.15 and 11.20 a.m.

behaviour, etc.) will be present. In these cases, the algorithm, as it stands, will not provide accurate estimates and would need some complex changes.

To detect these situations, some simple measures can be taken. One parameter that can help detect the presence of transients is the coefficient of variation (CV) (Eq. (20)):

$$CV_v = \frac{\sigma_v}{\bar{v}} \tag{20}$$

where

CV_v = speed coefficient of variation,
σ_v = speed standard deviation,
$\bar{v}$ = mean speed.

Theoretically, and assuming the existence of stationary traffic, this parameter tends to increase as the mean speed does. Although it is in the denominator, the more the mean increases, the more the deviation does. The coefficient of variation also indicates the importance of distinguishing time mean speeds and space mean speeds based on the relationships established by Wardrop (1952) or Rakha (2005), as Eq. (21) shows:

$$\overline{v_t} - \overline{v_s} = \frac{\sigma_t^2}{\overline{v_t}} = \frac{\sigma_s^2}{\overline{v_s}} = CV \cdot \sigma = CV^2 \cdot \bar{v} \tag{21}$$

The formula indicates that greater differences will occur with high CVs and high mean speeds. However, and based on empirical data, the greatest differences have been commonly found with high CVs and low mean speeds, which are a supposedly incompatible pairing. This fact clearly indicates that the traffic is not stationary (May, 1990;Rakha, 2005; Soriguera and Robusté, 2011). Fig. 11 shows the relationship between the mean speed and the CV in case VI, in which the algorithm did not perform well. In this case the CV diminishes with the mean, thus indicating the presence of transients and explaining this poor operation of the algorithm. In case IX (Fig. 12), the trend matches the assumption of stationarity and the algorithm provides good results.

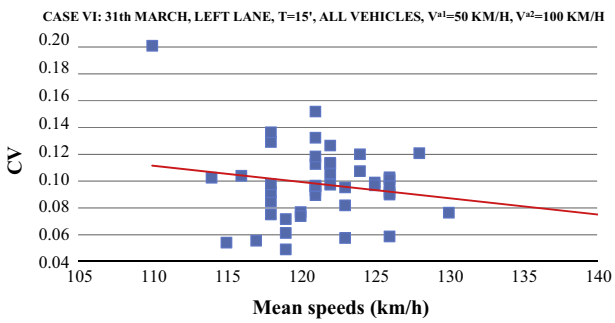

Fig. 11. Mean speeds vs. the coefficient of variation in case VI.

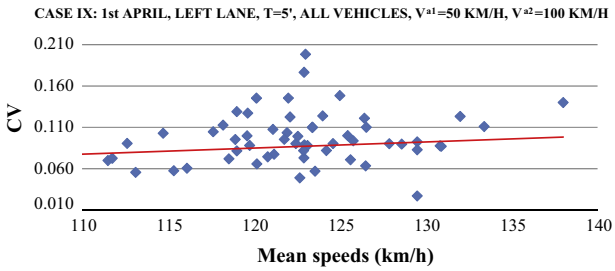

Fig. 12. Mean speeds vs. the coefficient of variation in case IX.

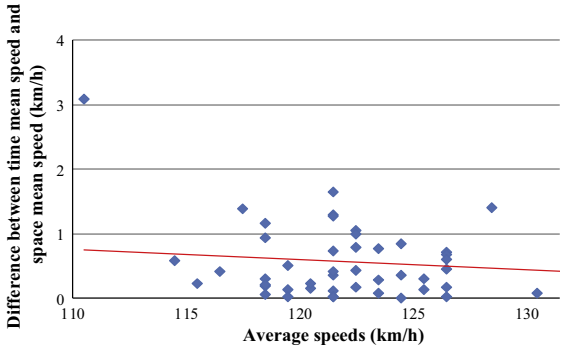

Fig. 13. Average speeds vs. the difference between time mean speed and space mean speed in case VI.

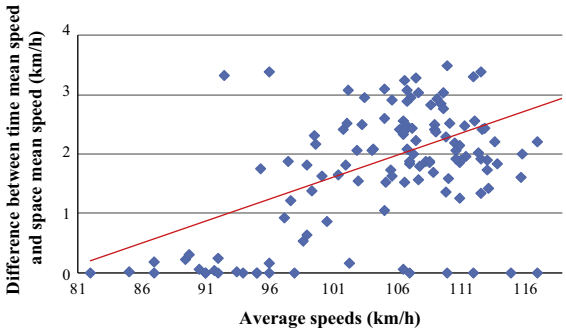

Fig. 14. Average speeds vs. the difference between time mean speed and space mean speed in case IX.

Although similar trends are usually observed by directly comparing average speeds with the difference between time and space means (Figs. 13 and 14), the fact of not taking into account the variance of the speeds could result in an exaggerated impression of the magnitude of the relationship. The use of CV is therefore strongly advised.

Finally, the exact place, day and moment when the data are collected are related to some of the issues previously mentioned. For example, the number and type of vehicles that use a freeway toward a capital on a workday morning in March will be very different from that on a Sunday in August on a secondary road surrounding a small town. Therefore the speeds and traffic conditions will also be very different.

6. Conclusions and further research

The development of road networks and new technologies has proven to be a useful tool to respond to the increasing demands of society regarding the total control of traffic evolution. Nevertheless, fundamental traffic theory must be correctly incorporated in modern methodologies in order to obtain accurate results.

This paper introduces an algorithm that estimates space mean speeds in a specific time interval of aggregation. These data can later be used as a first step, for example, for the calculation of occupancies or travel times. Thus, and after analysing the results obtained, three main conclusions can be extracted:

- it is possible to improve the current procedure followed by the majority of traffic management centres, i.e., considering time means equal to space means. It can be done inexpensively by exploiting all the data delivered by loop detectors. Specifically, the proposed algorithm allows an estimation of space mean speeds values that are accurate in most cases, or, at least, much closer to the real values than time mean speeds. The use of these data also improves the results of subsequent calculations,
- the good performance of the algorithm depends on the fulfilment of its initial hypotheses, i.e., stationarity of the traffic stream and log-normality of speeds in each time interval of aggregation. The boundary conditions for data acquisition and for the calculations can be established to a certain extent in order to achieve these characteristics,
- in the case of transients, for example the formation or dissipation of shock waves, most of the steps followed to design the algorithm, such as the extrapolation of the spot speeds to a section at the very beginning of the process, are not valid, and, thus, other specific methodologies should be used. In this regard, data fusion seems promising, as well as other completely different approaches which try to explain the propagation of traffic oscillation by means of car-following models (Li and Cui, 2014).

Further research can be carried out to improve the accuracy of the results or to broad the scope of application of the proposed algorithm. Some lines could be:

- including a smoothing process to remove erroneous data derived from the tendency for traffic loops to drift,
- including the necessary steps in the algorithm to calculate the confidence interval for the means in order to be able to choose the most accurate when more than one value is obtained,
- designing other algorithms adapted to other common speed distributions in addition to that introduced in this paper and that of Soriguera and Robusté (2011). Thus, after the application of a prior step which may help to find the most suitable distribution for the speeds, the appropriate algorithm could be chosen in each case.

As it has been observed, it is necessary to develop different and more evolved methodologies to estimate space mean speeds in the case of transients, as loop data may probably be insufficient in these situations. Other researchers have already achieved good results with various techniques of data fusion (Soriguera and Robusté, 2011; Bachmann et al., 2013;Yuan et al., 2014). However, there is still much work to do, since it is difficult to put most of them into practice due to their degree of complexity and/or high costs.

Acknowledgments

The authors would like to express their gratitude to the staff of the Centro de Gestión de Tráfico del Noroeste of the DGT (Dirección General de Tráfico), which belongs to the Ministerio del Interior (Government of Spain), and, particularly, to its Manager, Mr. Ramiro Martínez Rodríguez, for giving us access to the data managed by the centre. We would also like to give thanks to David Cota Mascuñana, Innovation and Development Manager of Iceacsa S.L.U., for supplying these data. Furthermore, comments provided by anonymous reviewers have greatly contributed to the improvement of the final version of the paper.

References

Bachmann, C., Abdulhai, B., Roorda, M., Moshiri, B., 2013. A comparative assessment of multi-sensor data fusion techniques for freeway traffic speed estimation using microsimulation modeling. Transportation Research Part C 26, 33–48. http://dx.doi.org/10.1016/j.trc.2012.07.003.

Breiman, L., 1969. Space-time relationships in one-way traffic flow. Transportation Research 3, 365–376. http://dx.doi.org/10.1016/S0041-1647(69)80020-1.

Cassidy, M., 1998. Bivariate relations in nearly stationary highway traffic. Transportation Research Part B 32 (1), 49–59. http://dx.doi.org/10.1016/S0191-2615(97)00012-X.

Chen, C., Kwon, J., Skabardonis, A.V., 2003. Detecting errors and imputing missing data for single loop surveillance systems. Transportation Research Record: Journal of the Transportation Research Board 1855, 160–167. http://dx.doi.org/10.3141/1855-20.

Coifman, B., 2014. Revisiting the empirical fundamental relationship. Transportation Research Part B 68, 173–184. http://dx.doi.org/10.1016/j.trb.2014.06.005.

Daganzo, C., 1997. Fundamentals of Transportation and Traffic Operations. Pergamon, Oxford, ISBN: 0080427855.

Dey, P., Chandra, S., Gangopadhaya, S., 2006. Speed distribution curves under mixed traffic conditions. Journal of Transportation Engineering 132 (6), 475–481. http://dx.doi.org/10.1061/(ASCE)0733-947X(2006)132:6(475).

Edie, L., 1965. Discussion of traffic stream measurements and definitions. In: Proceedings 2nd International Symposium on the Theory of Traffic Flow. OECD, París, pp. 139–154.

El Faouzi, N., Maurin, M., 2007. Reliability of travel time under log-normal distribution: methodological issues and path travel time confidence derivation. Transportation Research Board 86th Annual Meeting (CD-ROM). Transportation Research Record, Washington, DC.

Garber, N.H., 2002. Traffic and Highway Engineering. Brooks/Cole, California, ISBN: 0-534-38743-8.

Haight, F.M., 1962. A practical method for improving the accuracy of vehicular vehicle speeds distribution measurements. Highway Research Board Bulletin 341, 92–116.

Harvey, J., van der Merwe, A., 2012. Bayesian confidence intervals for means and variances of log-normal and bivariate log-normal distributions. Journal of Statistical Planning and Inference, 1294–1309. http://dx.doi.org/10.1016/j.jspi.2011.12.006.

Hiribarren, G., Herrera, J.C., 2014. Real time traffic states estimation on arterials based on trajectory data. Transportation Research Part B 69, 19–30. http://dx.doi.org/10.1016/j.trb.2014.07.003.

Homburger, W., Hall, J., Loutzenheiser, W., Reilly, W., 1996. Fundamentals of Traffic Engineering. Institute of Transportation Studies, University of California, Berkeley, California. ISSN: 0192-5911.

Jenelius, E., Koutsopoulos, H.N., 2013. Travel time estimation for urban road networks using low frequency probe vehicle data. Transportation Research Part B 53, 64–81. http://dx.doi.org/10.1016/j.trb.2013.03.008.
Jenelius, E., Koutsopoulos, H.N., 2015. Probe vehicle data sampled by time or space: consistent travel time allocation and estimation. Transportation Research Part B 71, 120–137. http://dx.doi.org/10.1016/j.trb.2014.10.008.
Khisty, C.L., 2003. Transportation Engineering: An Introduction. Prentice Hall, New Jersey, ISBN: 0-13-033560-6.
Knight, F., 1935. The Ethics of Competition and other Essays. Harper & Bros and George Allen & Unwin, New York and London, ISBN-10: 0836910885.
Knoop, V., Hoogendoorn, S., Zuylen, H., 2007. Empirical Differences Between time Mean Speed and Space Mean Speed. Traffic and Granular Flow'07. Springer, pp. 351–356. ISBN: 978-3-540-77074-9.
Li, X., Cui, J., An, S., Parsafard, M., 2014. Stop-and-go traffic analysis: theoretical properties, environmental impacts and oscillation mitigation. Transportation Research Part B 70, 319–339. http://dx.doi.org/10.1016/j.trb.2014.09.014.
May, A., 1990. Traffic Flow Fundamentals. Prentice Hall, New Jersey, ISBN: 0-13-926072-2.
Poomrittigul, S., Pan-ngum, S., Phiu-Nual, K., 2008. Mean Travel Speed Estimation using GPS Data without ID number on Unner City Road. ITS Telecommunications. IEEE, Phuket, ISBN 978-1-4244-2857-1, pp. 56–61.
Rakha, H.Z., 2005. Estimating traffic stream space-mean speed and reliability from dual and single loop detectors. Transportation Research Record: Journal of de Transportation Research Board 1925, 38–47. http://dx.doi.org/10.3141/1925-05.
SHRP2, 2013. Travel Time Reliability 2030: Innovations and Strategies for Today and Tomorrow. Transportation Research Board, Washington.
Soriguera, F., Robusté, F., 2011. Estimation of traffic stream space-mean speed from time aggregations of double loop detector data. Transportation Research Part C 19 (1), 115–129. http://dx.doi.org/10.1016/j.trc.2010.04.004.
Soriguera, F., Robusté, F., 2011. Highway travel time accurate measurement and short-term prediction using multiple data sources. Transportmetrica 7 (1), 85–109. http://dx.doi.org/10.1080/18128600903244651.
Soriguera, F., Robusté, F., 2013. Freeway travel-time information: design and real-time performance using spot-speed methods. IEEE Transactions on Intelligent Transportation Systems 14 (2), 731–742. http://dx.doi.org/10.1109/TITS.2012.2234454.
Viti, F., Rinaldi, M., Corman, F., Tampere, C., 2014. Assessing partial observability in network sensor location problems. Transportation Research Part B 70, 65–89. http://dx.doi.org/10.1016/j.trb.2014.08.002.
Wang, Y., 2012. A note on speed and travel time estimation based on truncated normal and log-normal distributions. Transportation Research Record: Journal of the Transportation Research Board 2315, 66–72. http://dx.doi.org/10.3141/2315-07.
Wardrop, J., 1952. Some theoretical aspects of road traffic research. Proceedings of the Institute of Civil Engineers 1 (2), 325–378. http://dx.doi.org/10.1680/ipeds.1952.11259.
Yuan, Y., van Lint, H., Hoogendoorn, S., van Wageningen-Kessels, F., 2014. Network-wide traffic state estimation using loop detector and floating car data. Journal of Intelligent Transportation Systems 18 (1), 41–50. http://dx.doi.org/10.1080/15472450.2013.773225.
Zou, Y., Zhang, Y., 2011. Use of skew-normal and skew-t distributions for mixture modeling of freeway speed data. Transportation Research Record: Journal of the Transportation Research Board, 65–75. http://dx.doi.org/10.3141/2260-08.

Glossary

AVI: AutomaticVehicle Identification
CV: Coefficient of variation
DGT: Dirección General de Tráfico (Spanish Traffic Management Center)
GPS: Global Positioning System
KS: Kolmogorov–Smirnov
OCR: Optical Character Recognition
TT: Travel time