

Chapter 3

A Simple Algorithm for the Estimation of Road Traffic Space Mean Speeds from Data Available to Most Management Centers



Margarita Martínez-Díaz

Abstract In traffic engineering, a lot of valuable information is obtained after appropriate processing of data collected by certain sensors. However good the data may be, the information extracted can be completely wrong if the processing is inadequate. One of the most common simplifications in the field, which, for example, affects some travel time estimation methodologies, is the use of temporal average speeds as equivalent to spatial averages. This chapter explains the causes of this bad practice, which is linked to the most traditional (and most extended) road equipment and procedures. To correct this trend, a highly applicable solution in the form of an algorithm is proposed. Although the results of the algorithm are not fully robust, they are favorable in a wide variety of cases, with the added bonus that no additional investment would be required.

3.1 Introduction

As society progresses, 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 exploiting synergies among different means of transport and contributes to a more sustainable growth (SHRP 2 2013). Many different initiatives such as real-time calculation of travel times, active traffic management or automated driving emerge as examples of key achievements.

Although these lines of research are very different, they have two commonalities, namely the need for (i) appropriate data and (ii) well-founded calculations. The

M. Martínez-Díaz (✉)

UPC-BarcelonaTech, Department of Civil and Environmental Engineering, Area of Transport and Territorial Infrastructures, Barcelona Innovative Transportation (BIT) Research Group, Polytechnic University of Catalonia, Barcelona, Spain
e-mail: margarita.martinez.diaz@upc.edu

development of new technologies and computer software offers the possibility of collecting varied data and combining them to obtain accurate results (Yuan et al. 2014). Mobile phones, GPS (Global Positioning System), Bluetooth, Optical Character Recognition (OCR) cameras, and many other devices are sources of traffic data usable for calculations. As seen, GPS-enabled cell phones, RFID technologies, etc. have opened a new way of collecting traffic data, as they are able to register individual vehicle trajectories (Hiribarren and Herrera 2014). Furthermore, vehicles themselves will also act as high-level “sensors” in future cooperative scenarios. With these lagrangian sensors moving within the traffic stream, increasing amounts of data will be available. Therefore, it will be possible to design much more precise methodologies, either for real-time travel time estimation or for any other purpose aimed at the dynamic management of traffic.

However, currently, neither totally accurate data nor the most complex programs are usually available, at least in a sufficient amount, in less trafficked areas. This is the case for example on secondary roads, in rural areas or for small traffic management centers. In fact, the majority of these centers in developed countries depend on equipment such as loop detectors and common cameras (unable to identify vehicles). That is, loops are the main sources of data. Traffic researchers have demonstrated the advisability of deploying double loops (in pairs in each section of each lane) rather than single loops to obtain more data and thus better results in later calculations (Chen et al. 2003). Fortunately, at present this trend is usually fulfilled. Moreover, the fact that there is a single data source on any road is expected to gradually disappear. Anyway, until today’s scenarios evolve, some modifications can be performed in the procedures currently implemented in traffic centers so that they better manage traffic. In the case of this chapter, travel time estimation by means of spot speed methods will be improved only with the application of traffic flow theory, and maintaining loops as the unique data source. First, a remainder of the basics of these detectors is included next.

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) because of the change that it causes in the electromagnetic properties of the loop. The main differences among loops are related to the software that manages and stores these data, which can be programed in several ways. As explained in Chap. 1, the data usually available in previously determined time intervals of aggregation, Δt , with the double-loop configuration are follows:

- Number of vehicles that pass over the detectors.
- Lengths of these vehicles: the software that manages the information usually classifies them into groups and keeps only the number of vehicles in each group. For example, in Spain the usual classification is as groups of vehicles shorter than 6 m, between 6 and 10 m and larger than 10 m.
- Spot speed measurements: again, although at first individual spot speeds are detected, the software calculates and registers only their mean, i.e., the time mean speed, \bar{v}_t , the average speed of all vehicles passing over a particular spot.

- Number of vehicles that pass over the detectors with a speed lower than a particular reference speed. It is common 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 centers manage them.

The duration of the time intervals of aggregation ranges from 20–30 s in the USA up to 15 min in some European countries. Intervals between 3–5 min have proven to be the most suitable (Soriguera and Robusté 2013): both shorter and longer durations have some advantages but also some disadvantages, as it will be discussed in Sect. 3.5.

Variation of traffic speeds at various places over time turns out to be one of the basic inputs for subsequent studies, for example, the indirect estimation of travel times. However, the problem is that most studies are based on the fundamental equation of traffic flow (Eq. 3.1, introduced as Eq. 1.8 in Chap. 1); it provides the relationship between flow, q , and density, k , by means of a specific type of speed, the so-called space mean speed, \bar{v}_s , which is really a harmonic mean calculated under particular conditions (Wardrop 1952). Further explanation about this point is included in Sect. 3.2.

$$q = \bar{v}_s * k \quad (3.1)$$

The use of 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 to achieve the spatial implication needed. This spatial generalization is extremely complicated, particularly in case of congestion.
- As mentioned, the software delivers 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 final results.
- Although loops are simpler, more economical and more common than other devices used to collect traffic data, their utility depends on their density on the road (Bachmann et al. 2013). Some research has resulted in the development of simple search algorithms that efficiently select sensor locations in order to obtain suitable data when the number of available sensors is limited (Viti et al. 2014). Nevertheless, difficulties remain on those roads already constructed.

The goal of the algorithm introduced in this chapter is to calculate spot space mean speeds exclusively from the data provided by double-loop detectors, avoiding extra expenses for the administrations. Specifically, it is focused on the calculation of the variance of the speeds with respect to the time mean, which allows using the relationship between time mean speeds and space mean speeds in the event of stationarity defined by Rakha and Zhang (2005). As explained in Chap. 2, further improvements must be implemented to obtain more accuracy in the final objectives, in this case, in the travel time estimations. Once working with space mean speeds,

a procedure for the generalization of these speeds over the links between detectors based on traffic dynamics and queue evolution would be the next challenge to face. Anyway, improvements in this first basic input have, as it is next demonstrated, satisfactory consequences.

The remaining sections of this chapter are as follows. Section 3.2 gives the background of different traffic speed definitions and summarizes their relationships according to various researchers. Section 3.3 develops the proposed algorithm, whose implementation is demonstrated in Sects. 3.4 and 3.5 with artificial and real data, and also compared with other methodologies. After the discussion of the results, attempts to find new relationships between mean speeds are performed in Sect. 3.6. Finally, Sect. 3.7 includes the conclusions and a proposal for new lines of research.

3.2 Background

Since 1952, when Wardrop (1952) 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, \bar{v}_s , is the average speed of all vehicles in a particular stretch of a road at a specific instant (Homburger et al. 1996). The time mean speed, \bar{v}_t , is the average of the speeds of all 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) because vehicles that are faster contribute more to the time-mean than the slow ones. On the contrary, vehicles of all speeds contribute equally to the space-mean. Space averages equal time averages only in case of space–time homogeneous traffic (Breiman 1969).

As it has been explained before, loops on a road detect and average spot speeds in stipulated time intervals, thus providing time mean speeds. However, if the individual spot speeds were stored, \bar{v}_s could be calculated by giving them certain spatial nature and by considering stationary traffic in the section (Edie 1965) as Eq. 3.2 shows:

$$\bar{v}_s = \frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n tt_i} = \frac{n * dx}{\sum_{i=1}^n \frac{dx}{v_i}} = \frac{1}{\frac{1}{n} \sum_{i=1}^n \frac{1}{v_i}}, \quad (3.2)$$

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 generalized definition of the average speed. The fact of labelling this generalized definition of the average speed as space mean speed \bar{v}_s is an abuse of notation. Actually, \bar{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, the space mean speed appears for example in the mathematical formulation of the average travel time \bar{tt} of n vehicles that cover a specific distance of a road L at a constant speed v_i (Eq. 3.3, already introduced in Chap. 2 as Eq. 2.1):

$$\bar{tt} = \frac{\sum_{i=1}^n tt_i}{n} = \frac{\sum_{i=1}^n \frac{L}{v_i}}{n} = L * \frac{1}{n} \sum_{i=1}^n \frac{1}{v_i} = \frac{L}{\bar{v}_s}. \quad (3.3)$$

In consequence, travel times would be underestimated if \bar{v}_t were used instead of \bar{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. 2009). The data aggregation process is in fact an influential source of noise and errors present in conventional measures of the traffic state (Coifman 2014). Many authors have stated the importance of correctly using time-based or space-based data, no matter their source. For example, the inverse of the harmonic mean of instantaneous speeds from probe vehicles is an unbiased and consistent estimator of the mean segment travel time when sampling by space, whereas it is biased upward when sampling by time (Jenelius et al. 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 worldwide 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 case of stationarity, which is the common hypothesis of all the following methodologies.

The first of these relationships, shown in Eq. 3.4, is due to Wardrop (1952):

$$\bar{v}_t = \bar{v}_s + \frac{\sigma_s^2}{\bar{v}_s}, \quad (3.4)$$

where σ_s^2 is the variance of the speed with regard to the space mean for the specific time interval of aggregation chosen. The accuracy of the formula has been experimentally verified, but most traffic management centers 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 formula postulated to relate both means is that of Garber (2002) shown in Eq. 3.5:

$$\bar{v}_t = 0.966 * \bar{v}_s + 3.541. \quad (3.5)$$

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. It must be continuously calibrated and, ultimately, it is not worth using.

Equation 3.6 has been used in several traffic studies. It was first derived by Khisty (2003), but they were Rakha and Zhang (2005) who proved it analytically:

$$\bar{v}_s = \bar{v}_t - \frac{\sigma_t^2}{\bar{v}_t}. \quad (3.6)$$

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. However, the impossibility of calculating the variance arises again. Nevertheless, and taking into account the utility of the formula, Soriguera and Robusté (2011) were able to estimate this variance by imposing the common hypothesis of stationary traffic in each time interval of aggregation and additionally assuming normality of the speed distribution. Then, the variance with regard to the time mean speed is given by Eq. 3.7:

$$\sigma_t = \frac{v^a - \bar{v}_t}{F^{-1}\left[\frac{n_v^a}{n}\right]}, \quad (3.7)$$

where.

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 performs well in specific conditions, Soriguera and Robusté (2011) warned that it is inappropriate to use it indiscriminately, especially in cases of shock wave onsets or offsets or with “stop and go” situations. As Cassidy (1998) declared, stationarity ensures some otherwise senseless relationships. However, the relationship established by Rakha and Zhang (2005) has been proven 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 to establish relationships between speeds is that they more or less fit common statistical distributions. The normal, log-normal, gamma and bimodal distributions appear in the majority of the traffic studies. The normal distribution is undoubtedly the most used because of its simplicity, and it performs well when traffic conditions are homogeneous. Consequently, it is also the common assumption of multivariate normal distributions for link travel times

(Jenelius et al. 2013). However, the log-normal and gamma distributions are usually more suitable because they have additional advantages (Haight 1962):

- They avoid the appearance of negative speeds.
- They keep their shape if 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 et al. 2007). If the log-normal speed distribution has a mean μ and a standard deviation σ , the distribution of travel times will follow Eq. 3.8:

$$f_t(t) = \frac{1}{\sqrt{2 * \pi * \sigma * t}} * e^{\left[-\frac{(Lnt + \mu)^2}{2 * \sigma^2} \right]}. \quad (3.8)$$

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

Many other complex distributions have been used in research, but their complexity prevents them from being put into practice (Zou and Zhang 2012). 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).

3.3 Simple Algorithm for the Estimation of Space Mean Speeds from the Data Provided by Double-Loop Detectors

Having analyzed previous investigations and taking into account the data available, the author decides to use the equation of Rakha and Zhang (2005) to solve the problem of not having an explicit value of the variance. The motivation is that the validity of this formula has been widely demonstrated in experimental studies. However, a particular analysis has been performed in order to compare it with other possible relationships. Section 3.6 contains the results of this comparison, which effectively verifies the goodness of this formula against the others.

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 Sect. 3.5.4. The first one has also been taken for granted in the other methodologies discussed in the chapter. With regard to the second, the author exploits the advantages of the log-normal distribution mentioned in Sect. 3.2. 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 $x = Lnv$ 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. 3.9 to 3.11, respectively,

$$f_v(v) = \frac{1}{\sqrt{2 * \pi * \sigma_x * v}} * e^{\left[-\frac{(Lnv - \mu_x)^2}{2 * \sigma_x^2} \right]} \quad \text{with } v > 0, \quad (3.9)$$

$$\mu_v = \bar{v}_t = e^{\mu_x + \frac{\sigma_x^2}{2}}, \quad (3.10)$$

$$\sigma_v^2 = \sigma_t^2 = \left(e^{\sigma_x^2} - 1 \right) * e^{2 * \mu_x + \sigma_x^2}, \quad (3.11)$$

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.

Note that the goal of the algorithm is to estimate σ_v^2 , which corresponds to the variance with regard to the time mean speed, termed σ_t^2 by Rakha and Zhang (2005). Therefore, μ_x and σ_x are needed. μ_v is supplied by the loops (the time mean speed, termed \bar{v}_t by Rakha and Zhang (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. 3.12:

$$\begin{aligned} P[V \leq v^a] &\approx \frac{n_v^a}{n} \approx P[e^X \leq e^{x^a}] \approx P[Lne^X \leq Lne^{x^a}] \approx P[X \leq x^a] = F[Z(x^a)] \\ &= F[Z(Lnv^a)] = F\left[\frac{Lnv^a - \mu_x}{\sigma_x} \right], \end{aligned} \quad (3.12)$$

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) = standardized value.

Rearranging Eqs. 3.10 and 3.12 yields a system with two equations (Eqs. 3.13 and 3.14) and two unknowns:

$$2\mu_x + \sigma_x^2 = Ln\bar{v}_t^2, \quad (3.13)$$

$$\mu_x + F^{-1}\left[\frac{n_v^a}{n}\right] * \sigma_x = Ln v^a, \quad (3.14)$$

where

F^{-1} = inverse of the cumulative standard normal distribution.

Finally, Eq. 3.15 is obtained

$$\sigma_x^2 - 2 * F^{-1}\left[\frac{n_v^a}{n}\right] * \sigma_x + Ln\left(\frac{v^a}{\bar{v}_t}\right)^2 = 0. \quad (3.15)$$

Solving Eq. 3.15, two possible values of σ_x arise. For two reference values of speed (v^{a1} and v^{a2}), four values are provided. In practice, some of these are nullified during the calculations because there are some mathematical limitations for the algorithm. In each time interval of aggregation T :

- n cannot be too small or the initial substitution of the theoretical probability by the accumulated frequency (Eq. 3.12) is problematic and the confidence interval of the estimations is too small.
- It is necessary that $n_v^a \neq 0$ and $n_v^a \neq n$. This keeps the inverse of the cumulated standard distribution from tending to infinite.
- $\left(F^{-1}\left[\frac{n_v^a}{n}\right]\right)^2$ must be greater than $Ln\left(\frac{v^a}{\bar{v}_t}\right)^2$ to avoid square roots of negative numbers when solving Eq. 3.15.
- It is necessary that $\frac{v^a}{\bar{v}_t} \neq 0$ to avoid natural logarithms of zero.

In those cases when more than one value of σ_x results, an action protocol must be established that helps to choose the most suitable. One possibility is to keep the value with the smallest confidence interval for a specific level of confidence. Once a value of σ_x is found and introduced into Eq. 3.13, the corresponding μ_x can be calculated. By using both values in Eq. 3.11, σ_t^2 is finally obtained and can be introduced into Eq. 3.6 to estimate \bar{v}_s . The flow chart in Fig. 3.1 summarizes the main steps of the algorithm.

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

A naïve solution could be to consider the confidence intervals of a parameter calculated in a previous step of the method, for example σ_x . If the best σ_x is chosen, the best σ_t^2 and thus a more accurate \bar{v}_s will be obtained. Because 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. 3.16 and 3.17 can be used:

$$\varepsilon_{\sigma_{x(1)}} = -\frac{(v^a - \mu_x) * \varepsilon_{z(1)}}{Z(Z + \varepsilon_{z(1)})}, \quad (3.16a)$$

$$\varepsilon_{\sigma_{x(2)}} = -\frac{(v^a - \mu_x) * \varepsilon_{z(2)}}{Z(Z + \varepsilon_{z(2)})}, \quad (3.16b)$$

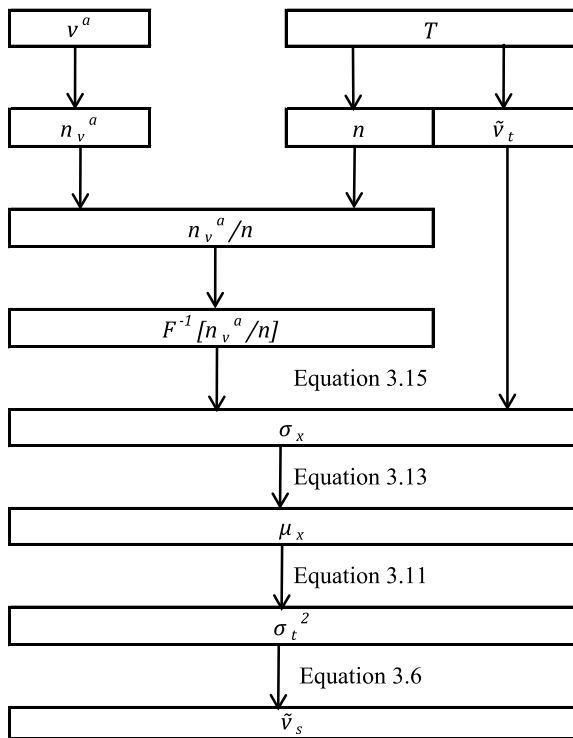
where

$$\varepsilon_{Z(1)} = F^{-1}(p + \varepsilon_p) - F^{-1}(p), \quad (3.17a)$$

$$\varepsilon_{Z(2)} = F^{-1}(p - \varepsilon_p) - F^{-1}(p). \quad (3.17b)$$

The variable p is the probability of a vehicle with a speed smaller than v^a passing over the detector in the time interval of aggregation. The circulation of vehicles over the detectors can be observed as a Bernoulli 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. 3.18:

Fig. 3.1 Steps of the algorithm to obtain space mean speeds from loop detector data



$$\hat{p} = \frac{n_v^a}{n}. \quad (3.18)$$

The proposed methodology relies heavily on the availability of n_v^a . If n_v^a is not reported to the traffic management center in the normal functioning of the system, the method cannot be applied. Obviously, carrying out modifications in the controllers in order to achieve these data lacks any sense, as it would be simpler, in this case, to introduce other modifications in order to directly obtain \bar{v}_s . Nevertheless, in those countries where n_v^a is available (a substantial number), the fact of using the estimated \bar{v}_s instead of working with \bar{v}_t (the current procedure) for later calculations would imply a higher level of accuracy without the need of any re-coding.

3.4 Implementation of the Algorithm with Artificial Data

To first verify the proper functioning of the algorithm, it was tested successfully with data generated with Matlab and readjusted to fulfil 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 as well as the mathematical requirements detailed in Sect. 3.3. For this last reason, the reference values were set at 101 km/h and 110 km/h (90 and 98% of the total time mean speed), ensuring enough vehicles participating in the calculations. The steps followed and the results are shown in Table 3.1, whereas Fig. 3.2 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 that 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 absolute values of the differences. In addition, regarding the estimated space means, only values with differences smaller than the maximum difference incurred by the loops were admitted. This procedure was followed also in Sect. 3.5 with real data.

3.5 Implementation of the Algorithm 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.

Table 3.1 Estimation of the space mean speeds and comparison of the results obtained with the data provided by the loops and with the real values

Time period T	Number of vehicles n	Classification of vehicles according to speeds			n _{est} /h	Time mean speed (Km/h)	Estimated space mean speed (Km/h)	Real space mean speed	Difference time mean - real space mean	Error (%)	Difference space mean - real space mean	Error (%)
		V < 101	101 < = V < 110	V > = 110								
07:30:00	28	0	4	24.00	0.0000	123.56	122.22	122.47	1.09	0.89	-0.25	0.20
07:45:00	25	3	9	13.00	0.1200	111.10	110.41	110.30	0.79	0.72	0.11	0.10
08:00:00	33	11	8	14.00	0.3333	106.51	105.23	105.62	0.89	0.84	-0.39	0.37
08:15:00	32	1	11	20.00	0.0313	114.48	113.98	113.58	0.89	0.79	0.39	0.35
08:30:00	36	1	5	30.00	0.0278	119.43	118.55	118.41	1.02	0.86	0.14	0.12
08:45:00	45	2	4	39.00	0.0444	124.11	122.77	122.67	1.45	1.18	0.10	0.08
09:00:00	36	9	10	17.00	0.2500	110.18	108.62	108.55	1.63	1.50	0.07	0.06
09:15:00	51	14	18	19.00	0.2745	107.80	106.71	106.87	0.93	0.87	-0.16	0.15
09:30:00	43	7	14	22.00	0.1628	111.38	110.37	110.05	1.32	1.20	0.32	0.29
09:45:00	39	11	9	19.00	0.2821	108.55	107.13	106.93	1.62	1.52	0.21	0.19
10:00:00	31	8	6	17.00	0.2581	112.33	109.87	108.34	3.99	3.68	1.53	1.41
10:15:00	22	10	2	10.00	0.4545	104.23	101.48	100.36	3.87	3.86	1.11	1.11
10:30:00	32	18	3	11.00	0.5625	103.71	99.80	95.59	8.12	8.49	4.21	4.40
10:45:00	29	4	11	14.00	0.1379	110.39	109.70	109.61	0.78	0.71	0.09	0.08
11:00:00	16	5	1	10.00	0.3125	114.11	112.97	111.29	2.82	2.54	1.69	1.52
11:15:00	24	2	6	16.00	0.0833	114.12	113.28	112.84	1.28	1.14	0.44	0.39
11:30:00	26	8	3	15.00	0.3077	111.88	108.45	108.36	3.52	3.25	0.10	0.09
11:45:00	29	7	6	16.00	0.2414	116.54	109.99	110.89	5.65	5.09	-0.90	0.81
TOTAL VEHICLES	577.0	121.0	130.0						MEAN	2.05		0.59
									ERROR			

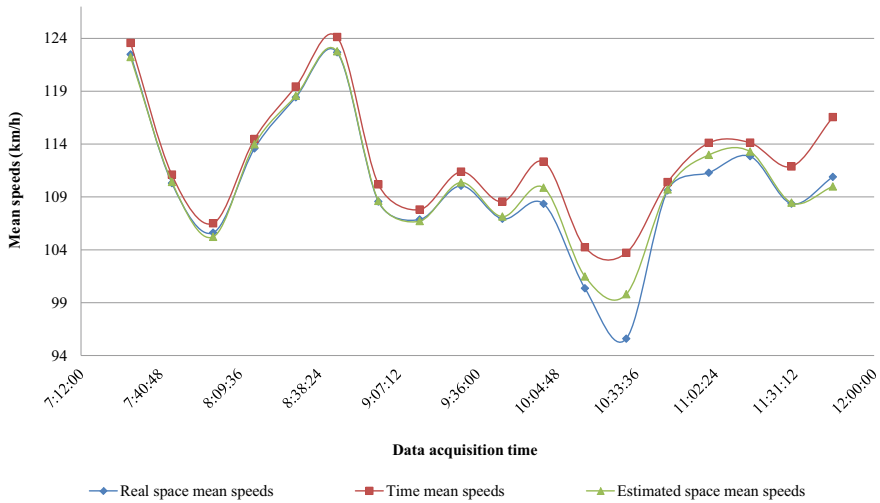


Fig. 3.2 Comparison of the real space mean speeds, the time mean speeds and the space mean speeds estimated with the algorithm from data that completely fulfil the initial conditions of the method

3.5.1 The Data

The data used for this study were collected during two days, on March 31th, 2014 and April 1st, 2014 in a section with double loops (P.K. 86 + 211, with two lanes in the direction toward 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. It must be noted that the fact that the data is a few years old has no special implication. In fact, the traffic control center in charge of this freeway still generates this type of information on a daily basis.

During the normal management of this freeway, the common data available were and are as follows:

- 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 \bar{v}_t : in 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 the 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

Table 3.2 Cases analyzed to test the algorithm

Case	Day	Lane	T(')	L	N	V^{a1}	V^{a2}
I	31 March	Right	15	all	4,662	50	100
II	01 April	Right	15	all	2,841	50	100
III	01 April	Right	15	all	2,841	98	107
IV	01 April	Right	5	all	2,841	50	100
V	01 April	Right	5	L < 10 m	2,489	50	100
				L > = 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

Note that neither the stationarity of the traffic flow nor the log-normality of the speeds is guaranteed. This issue is discussed in Sect. 3.5.4

of aggregation (T , in minutes) and reference speeds (V^{a1} and V^{a2}) were used. N is the number of vehicles detected during the entire data acquisition period. Table 3.2 shows the cases that have been analyzed:

3.5.2 The Results

Table 3.3 shows the difference between using the time mean speeds provided by the loop detectors or the space mean speeds estimated with the algorithm as substitutes for real space mean speeds. This difference is shown as in Sect. 3.4, i.e., by determining the mean relative error in each case.

In 8 out of the 11 cases analyzed (and taking into account that case V has been subdivided) 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 behavior was analyzed and understood; it is discussed in Sect. 3.5.4.

Note that in most cases it is not possible to determine the validity of the algorithm by focusing only on one of the boundary conditions; 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 them. As an example, between cases VI (Fig. 3.3) and VII (Fig. 3.4) only the reference speeds are different. However, the algorithm only shows a good performance in the latter case. The reason underlying this fact is that, in case VI, the sample includes fewer vehicles because most of them were driving at speeds higher than 50 km/h. Another example is based on cases IV (Fig. 3.5) and V (Fig. 3.6). Segregating the sample according to the vehicle length improves the performance for light vehicles

Table 3.3 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 algorithm (%)	0.79	0.87	0.99	0.59	0.46	-	0.86	0.44	0.78	0.58	0.50

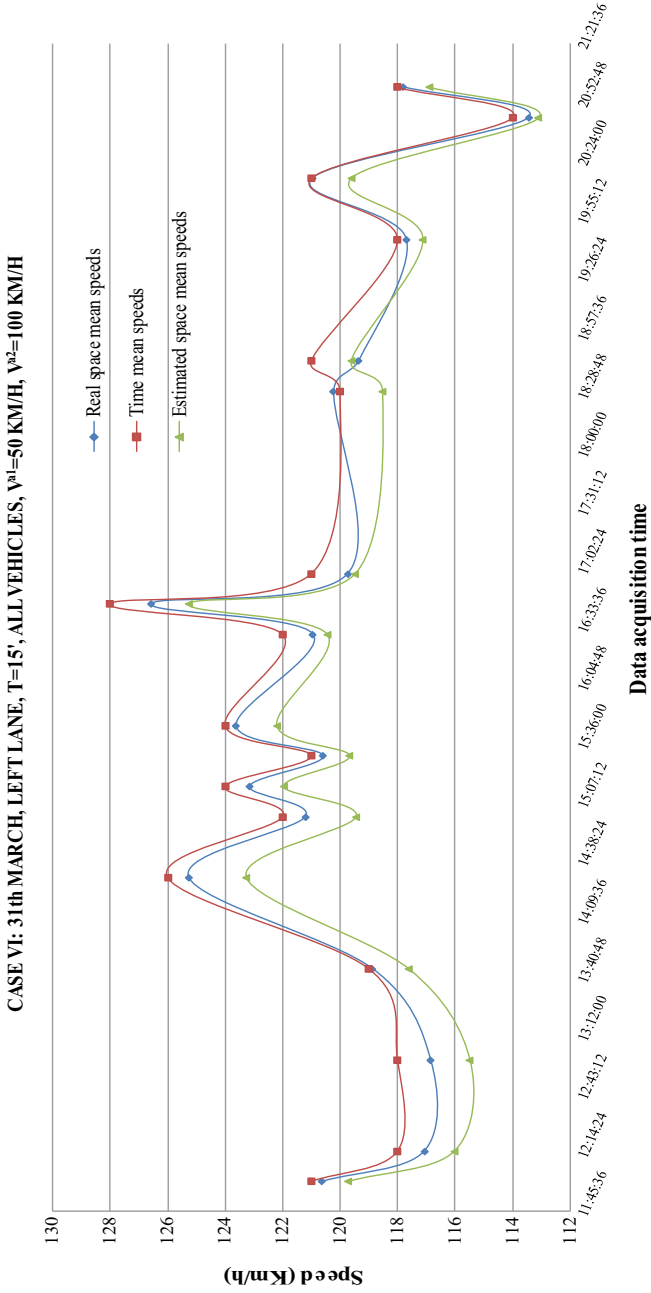


Fig. 3.3 Comparison of the real space mean speeds, the time mean speeds and the space mean speeds estimated with the algorithm in case VI

CASE VII: 31th MARCH, LEFT LANE, T=15', ALL VEHICLES, $v^{a1}=110$ KM/H, $v^{a2}=120$ KM/H

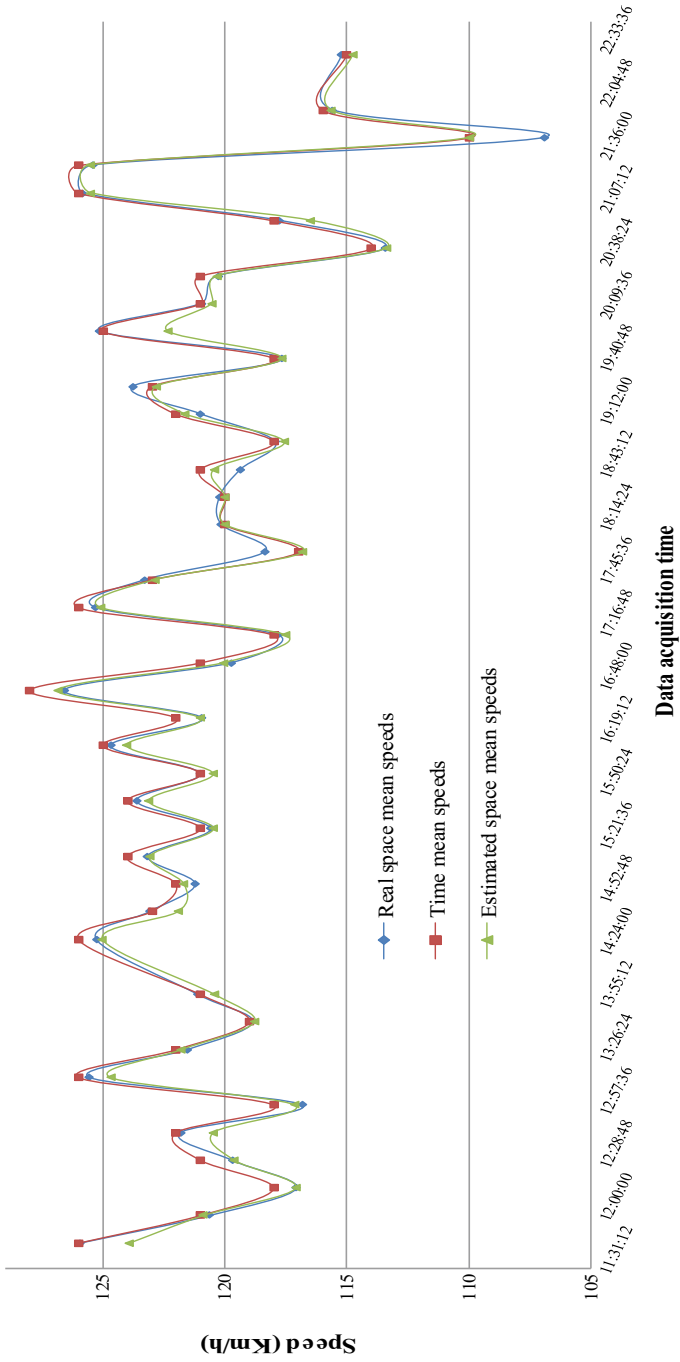


Fig. 3.4 Comparison of the real space mean speeds, the time mean speeds and the space mean speeds estimated with the algorithm in case VII

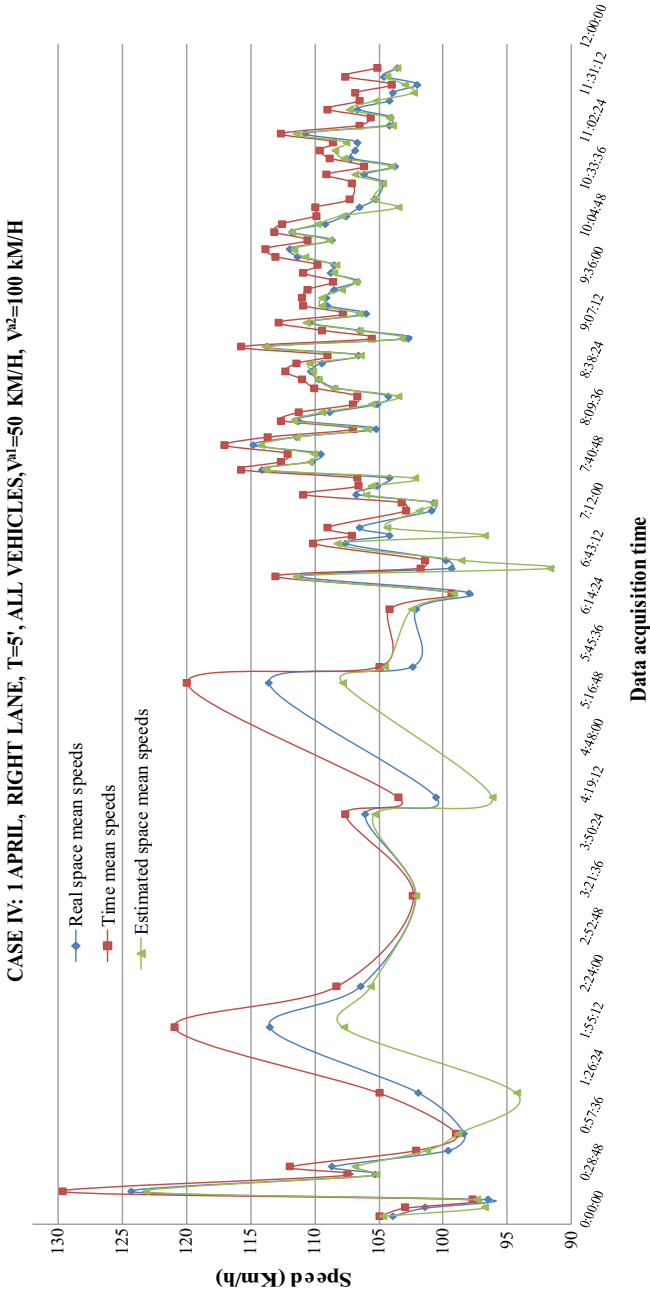


Fig. 3.5 Comparison of the real space mean speeds, the time mean speeds and the space mean speeds estimated with the algorithm in case IV

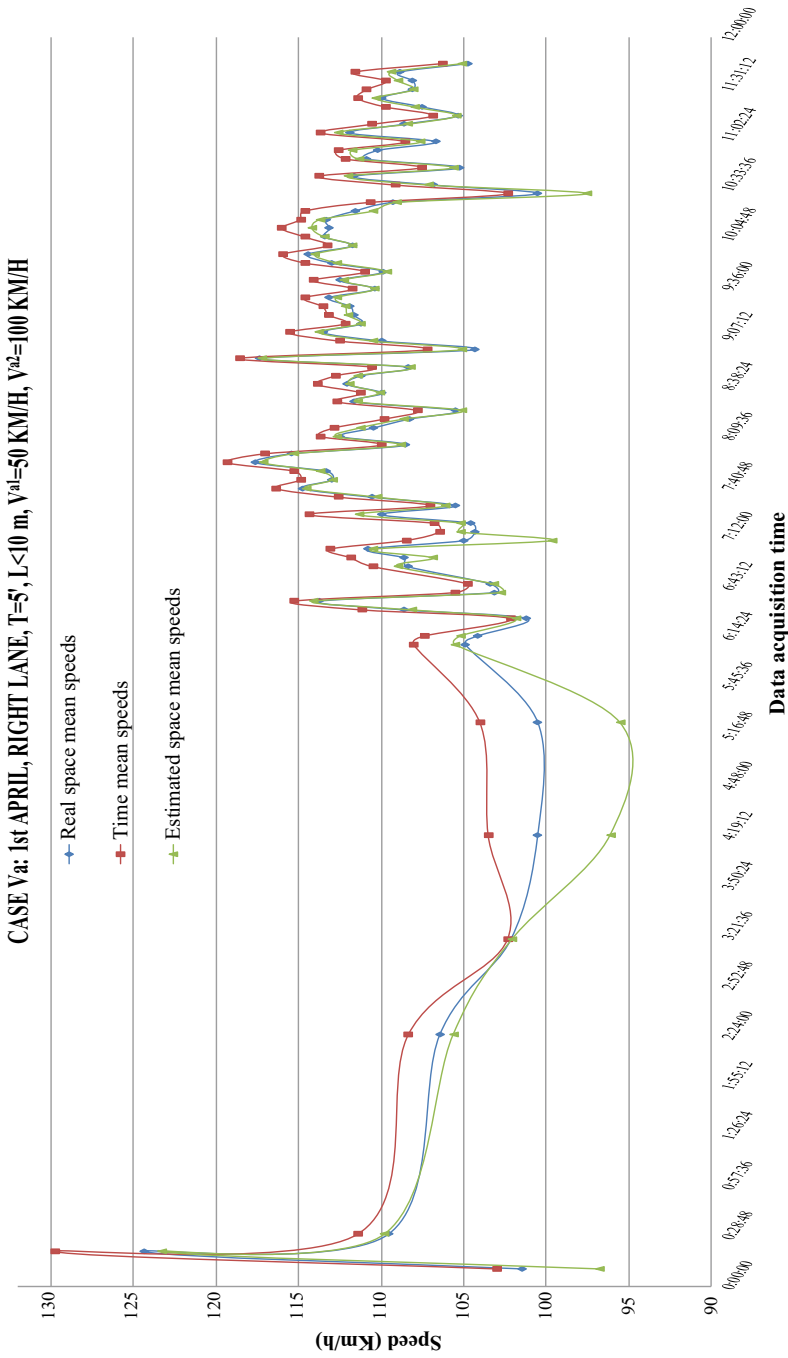


Fig. 3.6 Comparison of the real space mean speeds, the time mean speeds and the space mean speeds estimated with the algorithm in case Va

because the hypothesis of log-normality is better achieved. As for heavy vehicles, the algorithm in this specific example does not even run due to the small sample size of these vehicles. The influence of the length of the time interval of aggregation can be observed for example between cases II and IV (Figs. 3.7 and 3.5). The results of case IV, where $T = 5$ minutes, are much better.

3.5.3 *Comparison Between the Proposed Algorithm and Other Methods*

Because the proposed algorithm is somewhat more complicated than that introduced by Soriguera and Robusté (2011), a comparative analysis was performed to verify that it is worth using. In case I for example, the proposed algorithm demonstrated good behavior, diminishing the error incurred by the use of time mean speeds by 0.58%. Figure 3.8 and Table 3.4 Comparison of the errors introduced by different methodologies in case I. compare these results with that obtained with the methodology of Soriguera and Robusté (2011), which, as mentioned before, assume normality and stationarity in each time interval of aggregation T .

In spite of being conscious of the dependence of the formula of Garber on the boundary conditions, Table 3.4 also includes the results that would be obtained from its application, only for comparison purposes. The equation of Wardrop, as it has been previously stated, is clearly useful only to calculate \bar{v}_t from \bar{v}_s , what is not necessary in practical uses.

3.5.4 *Discussion*

Given the accuracy of the estimates achieved in each case, some conclusions must be drawn. It seems that the algorithm is worth using in numerous situations because results are usually more accurate than the currently accepted time mean speeds. However, while it clearly performs better in some of these cases, it does not do so 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.
- Prevailing type of vehicles.
- General traffic conditions.
- Place, day and moment of data acquisition.

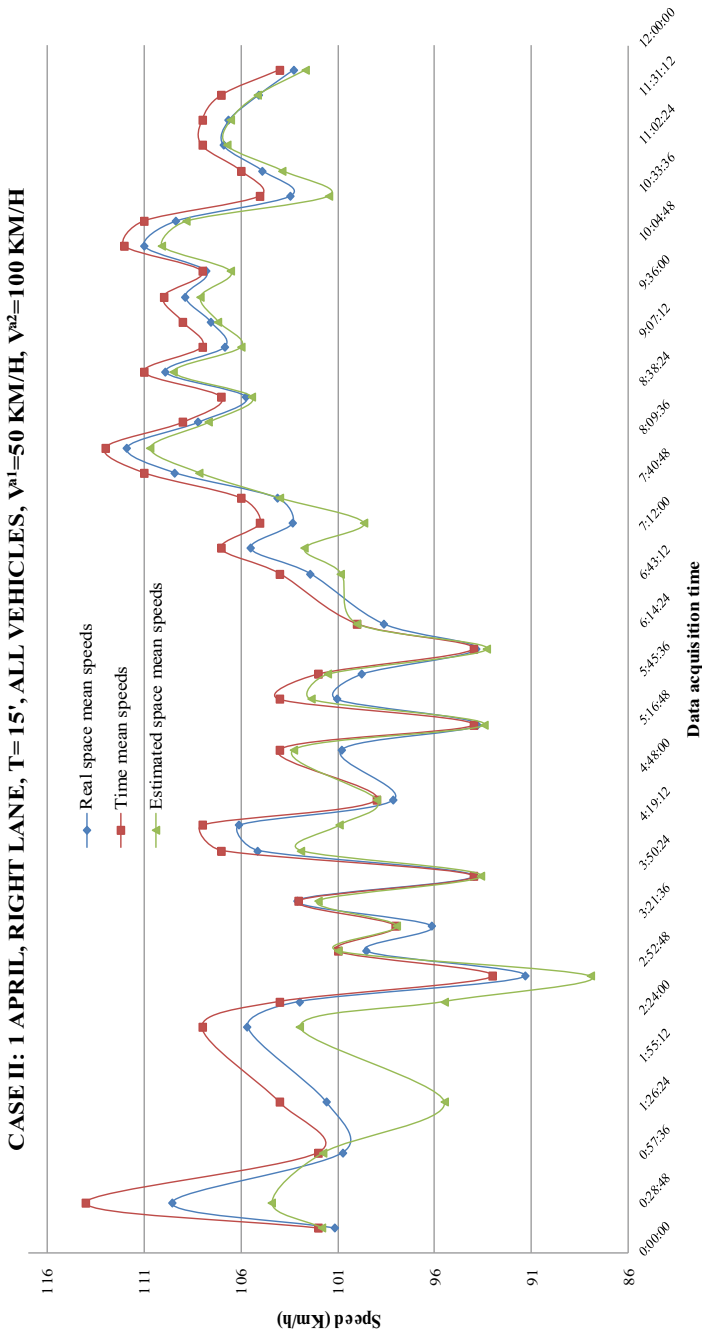


Fig. 3.7 Comparison of the real space mean speeds and the space mean speeds estimated with the algorithm in case II

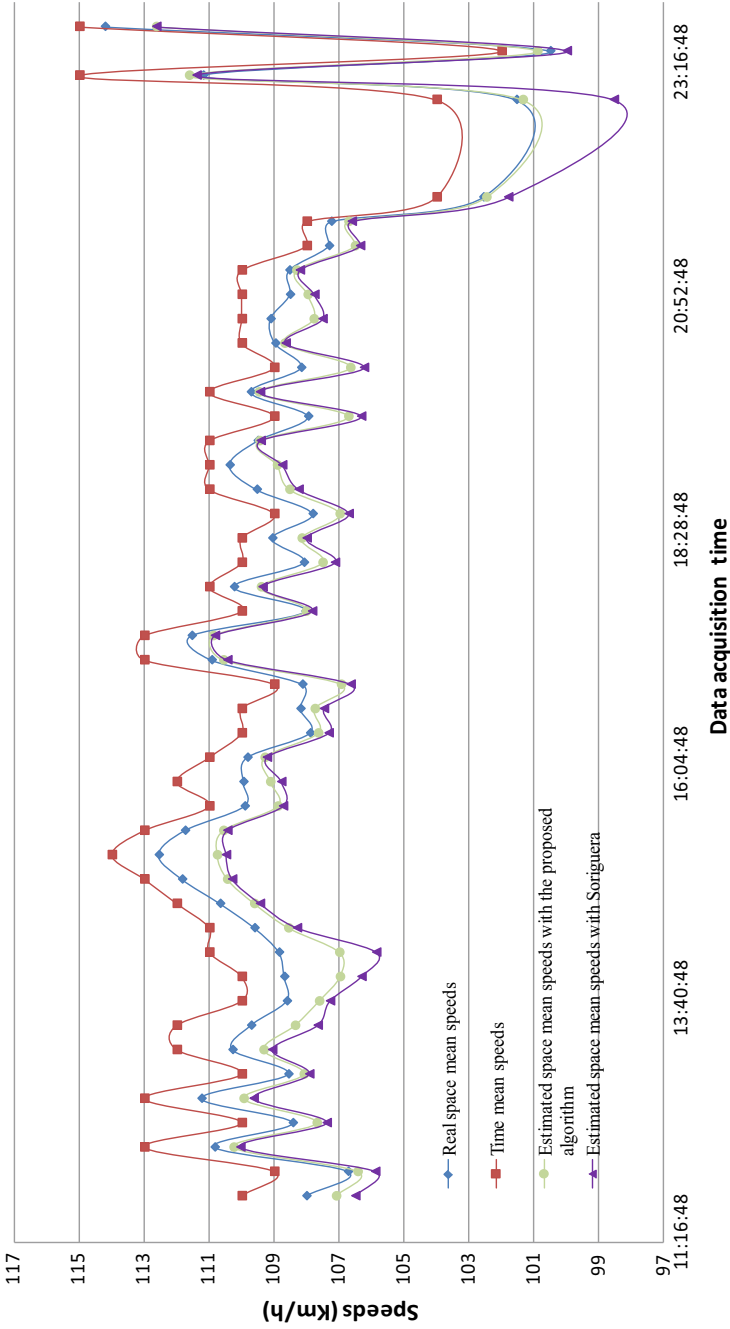


Fig. 3.8 Comparison of the real space mean speeds, the time mean speeds and the space mean speeds estimated with the proposed algorithm and the algorithm of Soriguera and Robusté (2011) in case I

Table 3.4 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	4,662	100	1.35
Use of the equation (Eq. 3.5) proposed by Garber (2002)	4,662	100	1.56
Use of the algorithm (Eq. 3.7 proposed by Soriguera (2011))	4,547	97.53	1.05
Use of the algorithm (Eq. 3.6 plus Eq. 3.15 and precedent) proposed in this paper	4,628	99.27	0.79

Regarding the sample size, the larger the sample, the better the algorithm performs. The main reasons are that the probability of having a log-normal distribution of speeds in each time interval of aggregation increases and because fewer mathematical inconsistencies appear 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 be met. This can be more or less difficult depending on the conditions established for the calculations. For example, with low traffic densities, the behaviors of fast (e.g. cars) and slow (e.g. trucks, buses, vans) vehicles can be very different (Dey et al. 2006). If the estimation is made with samples from all lanes, bimodal or even multimodal distributions will probably appear. Therefore, the analysis must be made by lane (Soriguera and Robusté 2011). However, with high-medium densities, log-normality could appear even in the whole section because the faster vehicles will not be able to reach their usual speeds. As previously mentioned, log-normality is more suitable with large samples. To illustrate the importance of fulfilling this hypothesis, two time intervals of $T = 5$ minutes of 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 (KS) Test. Table 3.5 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. Figures 3.9 and 3.10 also roughly represent this trend.

The election of the speeds chosen as a reference must be made in a logical way with the only purpose of having a sufficient number of vehicles in the sample. In the specific case of the AP-9 freeway, the values used were 50 and 100 km/h. As it is

Table 3.5 KS test results for two time intervals with accurate estimates

Test KS	7:45	11:15
Z Kolmogorov-smimov	0.481	0.764
P value (bilateral)	0.975	0.604

Fig. 3.9 Log-normal trend for time interval between 7.45 and 7.50 a.m

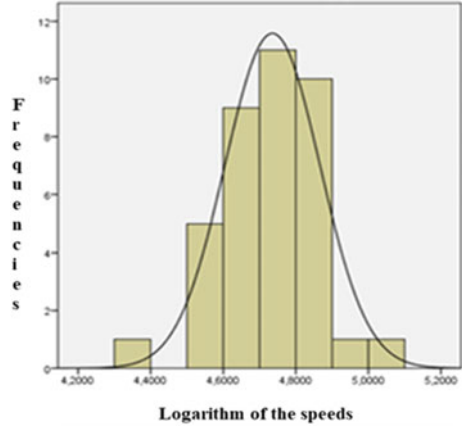
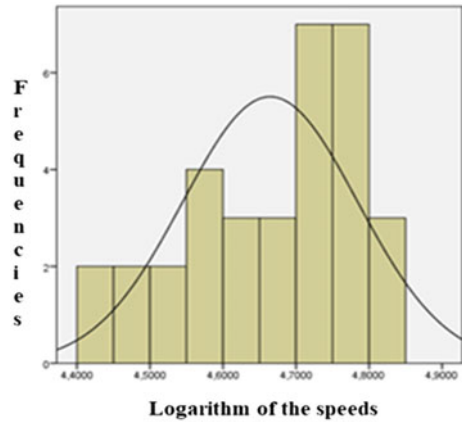


Fig. 3.10 Log-normal trend for time interval between 11.15 and 11.20 a.m



obviously uncommon for a vehicle to drive slower than 50 km/h on a freeway, some data will still be missed. Since the individual speeds were available, other values have been chosen for some of the analyses, what has 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 comply with the other main hypothesis of the method, i.e., the stationarity of the traffic flow, and 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 because 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 to help 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 adds some extra effort. In case of working per lane, later estimates for the section can be obtained with equations such as Eq. 3.19, where the superscript i labels the lanes of the section (Soriguera and Robusté 2011):

$$\overline{v_s^{section}} = \frac{1}{\left[\frac{1}{\sum_i n^i} \right] * \sum_i (n^i / \overline{v_s^i})}. \tag{3.19}$$

A preliminary analysis of the behavior of each type of vehicle should be done to avoid useless work. In this study, dividing the vehicles into the three sizes established by the Galician traffic management center generally provided the same results as classifying them into only two sizes (presumably the fast and slow ones), or even worse ones in some time intervals of aggregation lacking of vehicles of specific groups in the sample.

Note that the hypothesis of stationarity for the traffic flow has conditioned most of the steps followed when deriving the algorithm and, thus, is essential to achieve a good performance. This stationarity is assumed for each time interval of aggregation, and it is quite likely to occur. Nevertheless, there will also be frequent occasions in which transients (shock waves, stop and go behavior, etc.) will be present, and, thus, in which the algorithm as it is 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 to detect the presence of transients is the coefficient of variation (CV) (Eq. 3.20):

$$CV_v = \frac{\sigma_v}{\bar{v}}, \tag{3.20}$$

where

CV_v = speed coefficient of variation,

σ_v = speed standard deviation,

\bar{v} = mean speed.

Theoretically, if stationary traffic is assumed, 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. Besides, the coefficient of variation indicates the importance of distinguishing time mean speeds from space mean speeds based on the relationships established by Wardrop (1952) or Rakha and Zhang (2005), as Eq. 3.21 shows:

$$\overline{v_t} - \overline{v_s} = \frac{\sigma_t^2}{\overline{v_t}} = \frac{\sigma_s^2}{\overline{v_s}} = CV * \sigma = CV^2 * \bar{v}. \tag{3.21}$$

The formula indicates that greater differences will occur with high CV_S and high mean speeds. However, empirically, it is common that the greatest differences appear with high CV_S and low mean speeds, a supposedly incompatible pairing. This fact indicates that the traffic is not stationary (May 1990; Rakha and Zhang 2005; Soriguera and Robusté 2011). Figure 3.11 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, indicating the presence of transients and thus explaining the poor functioning of the method. In case IX (Fig. 3.12), the trend agrees with the assumption (stationarity) and the algorithm provides good results.

Although similar trends are usually obtained by directly comparing average speeds with the difference between time and space means (Figs. 3.13 and 3.14), 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 strongly advised.

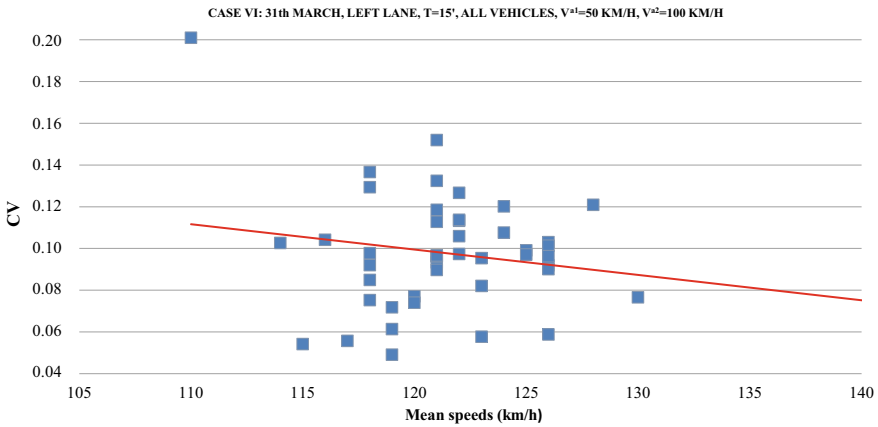


Fig. 3.11 Mean speeds versus the coefficient of variation in case VI

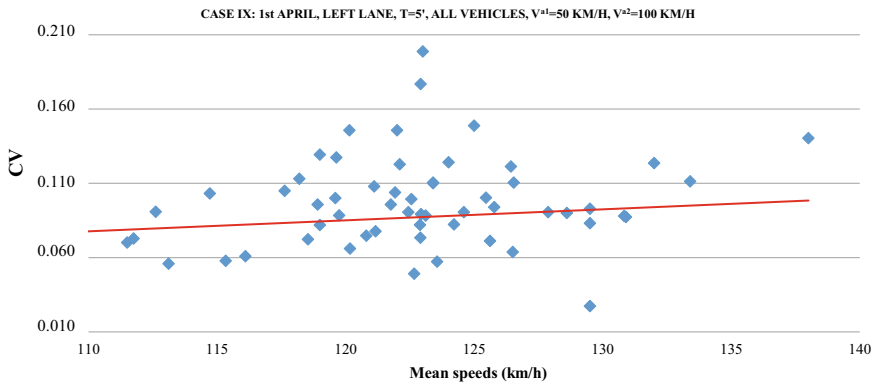


Fig. 3.12 Mean speeds versus the coefficient of variation in case IX

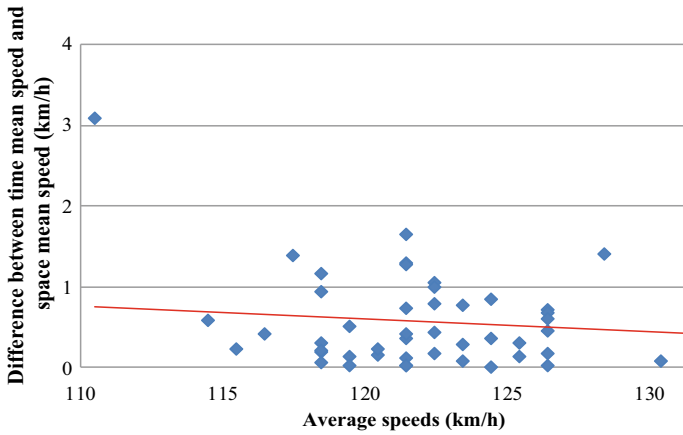


Fig. 3.13 Average speeds versus the difference between time mean speed and space mean speed in case VI

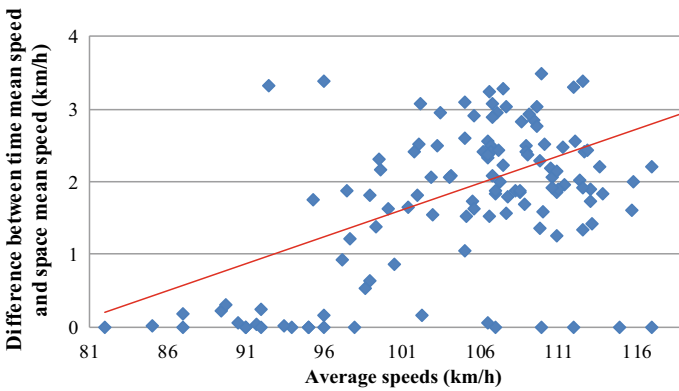


Fig. 3.14 Average speeds versus the difference between time mean speed and space mean speed in case IX

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

3.6 In Search of Other Relationships Between Mean Speeds

As explained, the algorithm proposed in Sect. 3.3 draws from the premise that the formula derived by Rakha and Zhang (2005) is the one that best defines a relationship between time mean speeds and space mean speeds, under different boundary conditions. Several researchers (e.g. Soriguera and Robusté 2011) reached the same conclusion, and this chapter has demonstrated the goodness of this formula, for example, compared to that of Garber’s. However, the author wanted to check whether it would be possible to find a formula that would yield better results for the real case study analyzed in Sect. 3.5, even assuming a priori the impossibility of extrapolation. As the reference speeds played no role in this analysis, a different and more concise nomenclature has been defined (Table 3.6).

As explained in Sect. 3.5.1, individual spot speed data were in this case available. This allowed the calculation of the exact time mean speeds (arithmetic means) and space mean speeds (harmonic means). Then, space mean speeds were estimated from time means by using Garber’s and Rakha and Zhang’s relations. The mean absolute and mean relative errors in relation to the real space mean for each case were also calculated.

In addition, an attempt was made to find another kind of correlation between both means. More in particular, the possibility of a linear, quadratic, cubic, logarithmic, inverse, exponential or power-type relationship was analyzed (Table 3.7).

Both the corrected coefficient of determination, R_c^2 , and the p-value were determined for this purpose. As it is already known, R_c^2 is a downward correction of R^2 based on the sample size n and on the number of independent variables k' , as shown in Eq. 3.22 below:

$$R_c^2 = R^2 - \left[\frac{k' * (1 - R^2)}{(n - k' - 1)} \right]. \tag{3.22}$$

Table 3.6 Cases analyzed to verify the best relationship between the time mean speeds and the space mean speeds

Case	Day	Lane	T(°)	L	N
1	31 March	right	15	all	4,662
2	31 March	left	15	all	769
3	01 April	right	15	all	2,841
4	01 April	left	15	all	596
5	31 March	left	5	all	769
6	01 April	right	5	all	2,841
7	01 April	right	5	L < 10 m	2,489
8	01 April	right	5	L > = 10 m	352

Table 3.7 Tested correlations between space and time mean speeds

Correlation	Outline
Lineal	$v_{ms} = a*v_{mt} + b$
Logarithmic	$v_{ms} = a*\ln(v_{mt}) + b$
Inverse	$v_{ms} = a*(1/v_{mt}) + b$
Quadratic	$v_{ms} = a*v_{mt}^2 + b*v_{mt} + c$
Cubic	$v_{ms} = a*v_{mt}^3 + b*v_{mt}^2 + c*v_{mt} + d$
Power	$v_{ms} = b*v_{mt}^a$
Exponential	$v_{ms} = b*\exp(a*v_{mt})$

The p-value is related to the contrast of the regression (ANOVA). In this case, the null hypothesis stands for a value of R^2 that equals zero. If the significance (p-value) in the statistical F-test is lower than 5% (for a confidence level of 95%), the null hypothesis can be rejected and, therefore, the existence of a correlation is proved.

In each of the cases studied, the estimated space mean speeds and the errors for the most suitable correlation were calculated. In this way, the best relationship both in general and for each particular case was determined. In order to remove the possible outliers, a slight smoothness was also made.

It should be highlighted that the variance with regard to the time mean for each specific time interval of aggregation introduced in Rakha and Zhang’s equation was again calculated from individual spot speeds, which, as said, are not usually available. It is also important to notice that the data used in this study fit different types of distributions depending on the time interval of aggregation, being lognormal and normal distributions the most commonly found, as expected.

Table 3.8 shows the results of the curvilinear estimation. The corrected coefficient of determination indicates that the quadratic correlation is the most suitable in most cases. The coefficients of the quadratic correlation for each analysis are included in Table 3.9. New estimates of space mean speeds were calculated with these values. A level of significance was given to each coefficient, being the value of the null

Table 3.8 Coefficients and their significance for quadratic relationships

Case	Non-standardized coefficients and p-value					
	a	p	b	p	c	P
1	-0.011	0.058	3.487	0.010	-137.497	0.057
2	0.000	0.623	1.083	0.000	-8.046	0.380
3	-0.006	0.008	2.095	0.000	-51.618	0.028
4	0.002	0.000	0.485	0.000	29.919	0.000
5	0.001	0.083	0.861	0.000	8.133	0.124
6	0.003	0.036	0.410	0.126	33.357	0.034
7	0.002	0.000	0.574	0.000	23.899	0.000
8	0.002	0.008	0.689	0.000	13.764	0.009

Table 3.9 Corrected coefficient of determination and significance for each correlation

Case	Corrected R2 and p-value													
	Lineal	Logarithm	Inverse	Cuadratic	Cubic	Power	Exponent	Lineal	Logarithm	Inverse	Cuadratic	Cubic	Power	Exponent
1	0.971	0.972	0.973	0.973	0.000	0.973	0.970	0.971	0.972	0.973	0.000	0.973	0.973	0.000
2	0.996	0.990	0.968	0.996	0.000	0.994	0.985	0.996	0.996	0.000	0.994	0.985	0.994	0.000
3	0.964	0.968	0.966	0.969	0.000	0.969	0.961	0.964	0.969	0.000	0.969	0.961	0.969	0.000
4	0.992	0.984	0.967	0.994	0.000	0.994	0.994	0.992	0.994	0.000	0.993	0.994	0.993	0.000
5	0.986	0.965	0.915	0.986	0.000	0.986	0.973	0.986	0.986	0.000	0.984	0.973	0.984	0.000
6	0.977	0.960	0.930	0.979	0.000	0.980	0.978	0.977	0.979	0.000	0.979	0.978	0.979	0.000
7	0.988	0.970	0.936	0.989	0.000	0.989	0.985	0.988	0.989	0.000	0.987	0.985	0.987	0.000
8	0.997	0.994	0.987	0.997	0.000	0.997	0.997	0.997	0.997	0.000	0.997	0.997	0.997	0.000

hypothesis equal to zero. As shown in Table 3.9 most coefficients are significant (p -value < 0.05), that is, they are needed to establish a good correlation. A linear relationship could achieve the same results only in two cases (as coefficient a is non-significant).

Finally, the mean absolute and mean relative errors with respect to the real space mean speeds (i.e., those calculated from individual speeds) encountered with the formula of Rakha and Zhang, that of Garber and with the quadratic correlation were also compared. The results are included in Table 3.10. It can be observed that the relationship of Rakha, despite being the most complex practice because of the need of estimating the variance with regard to the time mean, is worth considering. Both the absolute and relative errors are at the lowest level in all the cases analyzed in this study. Therefore, it has been again demonstrated its appropriateness to be part of the algorithm presented in Sect. 3.3.

Table 3.10 Errors observed with the different estimations of space mean speeds from time mean speeds

Case		Rakha		Garber		Quadratic correlation	
		Mean abs error	Mean relative error (%)	Mean abs error	Mean relative error (%)	Mean abs error	Mean relative error (%)
1	March 31th-right lane-T = 15'	0.09	0.09	2.60	2.40	0.32	0.29
2	March 31th-left lane-T = 15'	0.13	0.11	1.76	1.44	0.45	0.38
3	April 1st-right lane-T = 15'	0.45	0.44	2.47	2.35	0.79	0.76
4	April 1st-left lane-T = 15'	0.10	0.08	1.70	1.38	0.42	0.35
5	March 31th-left lane-T = 5'	0.30	0.25	1.54	1.27	0.61	0.51
6	April 1st-left lane-T = 5'	0.67	0.63	2.16	2.04	0.87	0.82
7	April 1st-left lane-T = 5' light veh	0.56	0.52	1.93	1.75	0.84	0.77
8	April 1st-left lane-T = 5' heavy veh	0.06	0.07	0.40	0.46	0.16	0.18

3.7 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 chapter introduces an algorithm that estimates space mean speeds in a specific time interval of aggregation as a first step, for example, for the calculation of travel times or occupancies. After analyzing the results obtained, three main conclusions can be drawn:

- It is possible to improve the current procedure followed by most traffic management centers, 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 speed values that are accurate in most cases, or, at least, much closer to the real values than time mean speeds. Consequently, 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 case of transients, for example the formation or dissipation of shock waves, most of the steps followed to design the algorithm are not valid (starting from the extrapolation of the spot speeds to a section). Thus, other specific methodologies should be used. Data fusion appears promising in this respect, as well as other completely different approaches that try to explain the propagation of traffic oscillation by means of car-following models (Li et al. 2014).

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

- Including a smoothing process to remove erroneous data derived from the tendency of traffic loops to drift.
- Including in the algorithm the steps necessary 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 chapter and that in Soriguera and Robusté (2011). Thus, after the application of a prior step that may help to find the most suitable distribution for the speeds, the appropriate algorithm could be chosen in each case.

As noted, it is necessary to develop different and more evolved methodologies to estimate space mean speeds in case of transients. Loop data are probably insufficient in these situations. Other researchers have 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 because of their complexities and/or high costs. Of course, the same issues arise when thinking of data-driven approaches.

In view of the results, usual spot speed methods enhanced by the proposed algorithm would be satisfactory to estimate travel times in stable traffic conditions. Their combination with more elaborated methodologies that only partially rely on loop data would allow making the most of these widespread detectors on other occasions. For example at present when congestion exists or even in future driving environments.

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. *Transp Res Part C: Emerg Technol* 26:33–48
- Breiman L (1969) Space-time relationships in one-way traffic flow. *Transp Res* 3:365–376
- Cassidy M (1998) Bivariate relations in nearly stationary highway traffic. *Transp Res Part B: Methodol* 32(1):49–59
- Chen C, Kwon J, Skabardonis AV (2003) Detecting errors and imputing missing data for single loop surveillance systems. *Transp Res Record: J Transp Res Board* 1855:160–167
- Coifman B (2014) Revisiting the empirical fundamental relationship. *Transp Res Part B: Methodol* 68:173–184
- Daganzo CF (1997) Fundamentals of transportation and traffic operations. Pergamon, Oxford
- Dey P, Chandra S, Gangopadhaya S (2006) Speed distribution curves under mixed traffic conditions. *J Transp Eng* 132(6):475–481
- Edie LC (1965) Discussion of traffic stream measurements and definitions. *Proc. 2nd International Symposium on the Theory of Traffic Flow, OECD, Paris*, 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)*. Washington DC: Transportation Research Board
- Garber NH (2002) *Traffic and highway engineering*. California: Brooks/cole. ISBN 0-534-38743-8
- Haight FM (1962) A practical method for improving the accuracy of vehicular vehicle speeds distribution measurements. *Highw Res Board Bull* 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. *J Stat Plan Inf*, 1294–1309
- Hiribarren G, Herrera JC (2014) Real time traffic states estimation on arterials based on trajectory data. *Transp Res Part B: Methodol* 69:19–30
- Homburger W, Hall J, Loutzenheiser W, Reilly W (1996) *Fundamentals of traffic engineering*. California: Institute of Transportation Studies. University of California, Berkeley. ISSN 0192-5911
- Jenelius E, Koutsopoulos HN (2013) Travel time estimation for urban road networks using low frequency probe vehicle data. *Transp Res Part B: Methodol* 53:64–81
- Jenelius E, Koutsopoulos HN (2015) Probe vehicle data sampled by time or space: Consistent travel time allocation and estimation. *Transp Res Part B: Methodol* 71:120–137
- Khisty CL (2003) *Transportation engineering: an introduction*. Prentice Hall, New Jersey

- Knight F (1935) *The ethics of competition and other essays*. New York and London: Harper & Bros and George Allen & Unwin. ISBN-10: 0836910885
- Knoop, V., Hoogendoorn, S. and H. Zuylen (2007). *Empirical Differences Between time Mean Speed and Space Mean Speed*. In *Traffic and Granular Flow'07*, 351–356. Springer. 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. *Transp Res Part B: Methodol* 70:319–339
- May A (1990) *Traffic flow fundamentals*. Prentice Hall, New Jersey
- Poomrittigul S, Pan-ngum S, Phiu-Nual K (2008) Mean Travel Speed Estimation using GPS Data without ID number on Inner City Road. *ITS Telecommunications* 2008:56–61
- Rakha H, Zhang W (2005) Estimating traffic stream space-mean speed and reliability from dual and single loop detectors. *Transp Res Record: J De Transp Res Board* 1925:38–47
- SHRP2 (2013) *Travel time reliability 2030: innovations and strategies for today and tomorrow*. Washington, DC: Transportation Research Board
- Soriguera F, Robusté F (2011) Estimation of traffic stream space-mean speed from time aggregations of double loop detector data. *Transp Res Part C: Emerg Technol* 19(1):115–129
- Soriguera F, Robusté F (2013) Freeway travel-time information: design and real-time performance using spot-speed methods. *IEEE Trans Intell Transp Syst* 14(2):731–742
- Viti F, Rinaldi M, Corman F, Tampère CMJ (2014) Assessing partial observability in network sensor location problems. *Transp Res Part B: Methodol* 70:65–89
- Wang Y (2012) A note on speed and travel time estimation based on truncated normal and log-normal distributions. *Transp Res Record J Transp Res Board* 2315:66–72
- Wardrop J (1952) Some theoretical aspects of road traffic research. *Proceedings of the Institute of Civil Engineers* 1(2):325–378
- Yuan Y, van Lint H, Hoogendoorn S, van Wageningen-Kessels F (2014) Network-wide traffic state estimation using loop detector and floating car data. *J Intell Transp Syst* 18(1):41–50
- 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*, pp 65–75