Highway Travel Time Data Fusion

F. Soriguera, D. Abeijón, L. Thorson and F. Robuste

Submission Date: July 31st, 2007
4,756 Words + 6 Figures + 4 Tables = 7,256 words

F. Soriguera (corresponding author), CENIT – Center for Innovation in Transport, Technical University of Catalonia (UPC), Jordi Girona 29, 2-A, 08034 Barcelona (Spain), Tel:+34934137667, Fax:+34934137675, francesc.soriguera@upc.edu.

D. Abeijón, CENIT – Center for Innovation in Transport, Technical University of Catalonia (UPC), Jordi Girona 29, 2-A, 08034 Barcelona (Spain), Tel:+34934137667, Fax:+34934137675, davidabei@hotmail.com.

L. Thorson, CENIT – Center for Innovation in Transport, Technical University of Catalonia (UPC), Jordi Girona 29, 2-A, 08034 Barcelona (Spain), Tel:+34934137667, Fax:+34934137675, leif.thorson@upc.edu.

F. Robuste, School of Civil Engineering, Technical University of Catalonia (UPC), Jordi Girona 1-3, B1, 08034 Barcelona (Spain), Tel:+34934017104, Fax:+34934017264, f.robuste@upc.edu

ABSTRACT
The development of new traffic monitoring systems and the increasing interest of road operators and researchers in obtaining reliable travel time measurements has leaded to the development of multiple travel time data sources. This situation represents a perfect environment for the implementation of data fusion systems to obtain the maximum accuracy from the available data.

This paper presents a new approach to calculate highway travel times fusing different data sources: inductive loop detectors and toll ticket data. Although the data fusion algorithm presented herein is applied to these types of data, it could easily be generalized to other equivalent sources.

The data fusion algorithm is applied to different travel time estimations in order to obtain a fused value more reliable than any of the individual estimations. The proposed algorithm overcomes some of the limitations of other methods, improving the spatial and temporal coverage, and determining the flow state (congested or not).

The results obtained in the application of the methodology to the AP-7 highway, near Barcelona in Spain, are found to be reasonable and accurate.

1. INTRODUCTION

Most of developed countries, unable to carry on with the classical strategy of extending the transportation infrastructures once these become saturated are switching to the optimization of road usage by means of operational management improvements.

This new approach is the result of environmental, budget and land occupancy limitations, being the last one especially important in metropolitan areas where the high population density is associated with increasing mobility needs of society.

The availability of reliable travel time information appears to be the key factor in the improvement of road networks management, since it allows an effective estimation of the traffic state and provides the most valuable and understandable information for road users [1].

For all these reasons, some European countries (Spain, France, Denmark, Italy, Finland, United Kingdom, Sweden, the Netherlands, Norway and Germany) grouped under the Trans-European Road Network (TERN) are developing travel time estimation projects [2].

This interest showed by road operators added to the development of ITS (Intelligent Transportation Systems) has leaded to a new framework in the traffic data management and has increased the variety of reliable, precise and economically viable road surveillance technologies [3-6]. In addition the appearance of ATIS (Advanced Traveler Information Systems) has made possible a simple and efficient information dissemination addressed to the road user.

This environment has brought researchers to an increasing interest in data fusion travel time techniques since late 90s. In the USA Palacharla and Nelson [7] studied the application of fuzzy logic to travel time estimation, evaluating which hybrid system turned out to be more effective (i.e. the fuzzy based on a neural network or on an expert system). They conclude that the neural network hybrid system is more precise, increasing the quality of the results obtained with classical travel time estimation methods. A similar methodology has been adopted by the Austrian Department of Traffic, Innovation and Technology (*bm vit*) who in 2006 presented a pioneer project, (still a pilot test) for obtaining reliable travel times and the congestion level of the road network, using multiple data sources (e.g. inductive loops, laser sensors, and taxi floating cars). The methodology used in this case was known as ANFIS (Adaptable Neural Fuzzy Inference System) which implied a reduction of 50% in the number of estimation mistakes [8]. Later, [9] relied on ANFIS to obtain delay times in signalized intersections achieving better results than the HCM (Highway Capacity Manual) [10], mainly in heavy congested situations.

Researchers in Singapore and China have tried to obtain predictions of traffic stream state using Bayesian inference on a neural structure from a unique source of data. The results improve those using simple neural networks in 85% of the cases [11].

In France, researchers have developed conceptually simple data fusion techniques. The best examples are the works of El Faouzi [12-14] in evidential Dempster-Shafer inference, which could be considered a generalization of Bayesian theory, improving the results of classical Bayes theories in pilot test runs in a highway near Toulouse. Two sources o data were used: license plate matching and inductive loops detectors. These experiences are being used by the French highway operators AREA, ASF, ESCOTA and SAPN for the calculation of travel times in their corridors [15-17].

Swedish and Scottish road operators (SRA - Sweden Road Administration and Transport for Scotland) since 2001 are studying the implantation of data fusion systems to obtain road travel times in their networks. In Scotland pilot tests in the A1 motorway, in the surroundings of Edinburgh, use up to 4 data sources: tracking of cellular phones, inductive loops detectors, floating car data and license plate matching. Surprisingly the cellular phone tracking stands out for its reliability [18-19]. In Germany, a private managed company has developed software capable of fuzzy inferring traffic variables to obtain a traffic flow state estimation, using a dynamic definition of domains. This software could be used to obtain the congestion delays, using the information provided by inductive loops and floating cars [20].

In Holland, van Lint, Hoogendoorn and van Zuylen [21] use neural networks for the prediction of travel times with gaps in the data, obtaining satisfactory results in spite of this partial information.

Park and Lee, both Koreans, have obtained travel times estimations in urban areas implementing neural networks and Bayesian inference, both independently, and using data from inductive loops and floating car. In both cases the results are considered promising [22].

In this context, the present paper proposes a new data fusion approach for travel time calculation. A simple algorithm using different travel time estimations (ITT - Instantaneous Travel Time and RTT - Reconstructed Travel Time) are considered to obtain a PTT (Predicted Travel Time) for any road corridor. The results of a pilot test in the AP-7 highway in Spain are also outlined in the paper, and they show that the developed methodology is sound.

The paper is organized as follows: Section 2 describes the different natures of travel time measurements. Section 3 introduces the methodology used for the development of the data fusion system providing also notation and mathematical algorithms. Section 3 presents the results of the application of the model to the AP-7 highway in Spain. Finally, some general conclusions and issues for further research are discussed in Section 4.

2. TRAVEL TIME DEFINITIONS

There are two main methodologies to measure travel time in a road link: the direct measure and the indirect estimation. The direct travel time measure is based in measuring the time interval that a particular vehicle takes to travel from one point to another. The data collection techniques used in this case are the floating car data, the license plate matching, the AVI (Automated Vehicle Identification) from toll infrastructure or other advanced techniques like the GPS automatic vehicle location or the cellular phone tracking. In the direct measurement travel time data is directly obtained from measures, taking into account that the trip must be finished in order to obtain the measurement. So the obtained data are a kind of measurements of past situations. This nature of travel times are defined in the present paper as RTT - Reconstructed Travel Times.

The alternative is the indirect travel time estimation from traffic flow characteristics (density, flow and speed), obtained from magnetic loop detectors. To obtain travel time data from these measurements some type of algorithm must be applied. These algorithms can be based on the speed estimation in detection points or on a cumulative flow balance in a particular stretch [23]. Both rely on an instant spot estimation of a traffic variable. This type of travel times are defined as ITT - Instantaneous Travel Time in the present paper.

3. DATA FUSION SYSTEM

The proposed algorithm uses two different estimations of ITT. First one is calculated using a simple spot speed algorithm, while the second uses a flow balance algorithm. Details of these algorithms can be found at [24]. In addition a RTT from toll ticket data is also used [25] to obtain a PTT (Predicted Travel Time) that estimates with accuracy and reliability the real travel time in the corridor. The data fusion structure can be seen in figure 1.

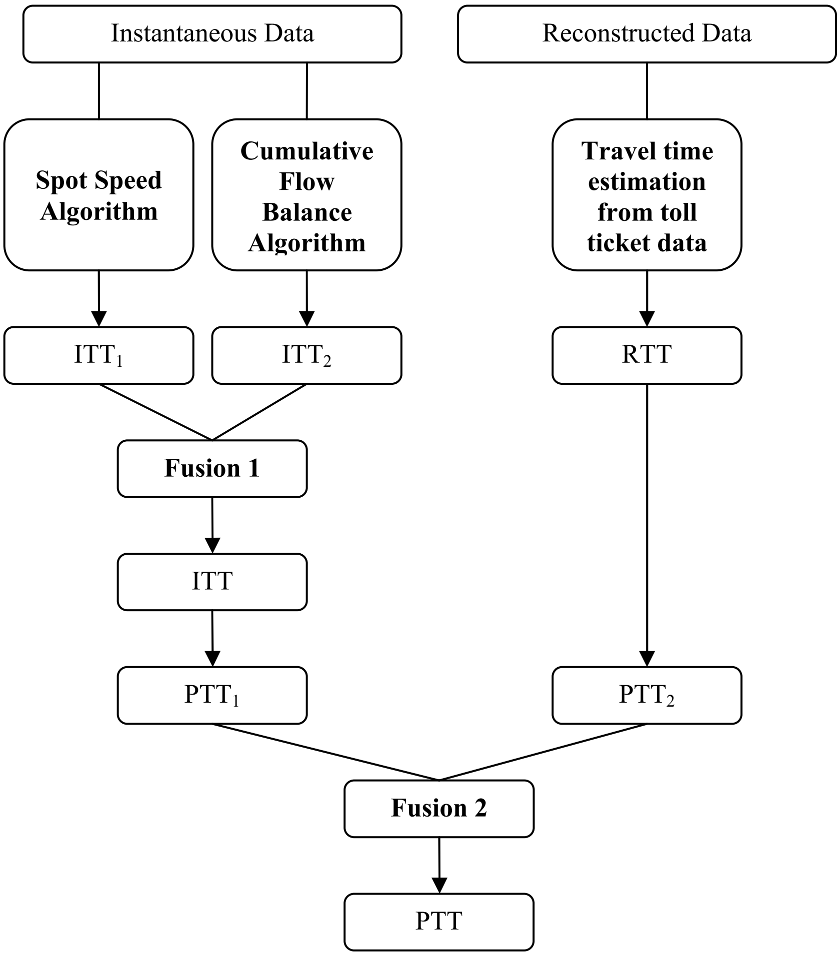


FIGURE 1 Structure of the Data Fusion System.

3.1 Fusion 1

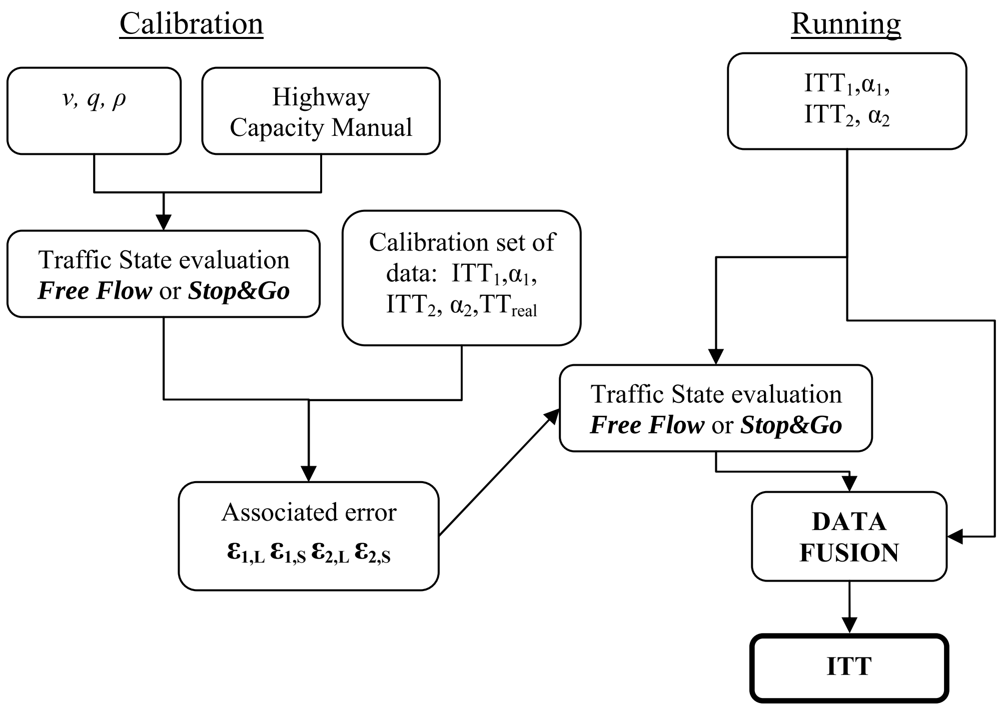
The first data fusion process fuses the two instantaneous travel times ITT_1 (form a spot speed algorithm) and ITT_2 (from a cumulative flow balance algorithm) to obtain a uniform ITT in terms of spatial and temporal coverage.

To do so, the following inputs are needed:

- x_0: Start point of the considered section.
- x_1: Final point of the considered section.
- t: Finishing time instant of the capture of information.
- α: Statistical quality parameter of data samples used to estimate each ITT. It belongs to a range between 0 – worst and 1 – ideal.
- ITTs: Instantaneous Travel Times 1 and 2 to be fused.
- v: Loop detector measure of average vehicles speed.
- ρ: Loop detector measure of average road occupancy.
- q: Loop detector measure of vehicle flow.

- TT_{real}: Real Travel Time for the fusion calibration.

The first group of information, x_0, x_1 and t, provides the correct location in space and time of data. The second group represents the data to fuse, with its original reliability α. It is necessary to distinguish α, from the accuracy of each algorithm referred in this paper as ε. While α is a parameter which changes according to the number of vehicles computed in each data sample, ε depends on the particular travel time estimation algorithm used, and varies with the traffic stream state (congested or not). ε will be determined during the calibration process of the system. Finally, the third group of inputs, v, ρ and q, are used to determine the state of traffic stream, distinguishing between the free flow and the stop &go situation.



Note: $\varepsilon_{1,L}$ $\varepsilon_{1,S}$ $\varepsilon_{2,L}$ $\varepsilon_{2,S}$ represent the accuracy of instant travel times algorithms (1 and 2) in free flow conditions (L) or stop&go situations (S). It is determined in the calibration proces by computing the Quadratic Mean Relative Error [26] of the real measure (TT_{real}) in relation to the estimated value ($ITT_{1 \text{ or } 2}$).

FIGURE 2 Fusion 1 Structure.

3.1.1 Traffic Stream State Evaluation

The data fusion process needs to evaluate the traffic stream state, in order to determine the algorithm associated error (ε). To do so, a bivalent probabilistic system is chosen, which enables to determine the most probable traffic state (congested or not) from the traffic fundamental variables.

For every individual traffic fundamental variable, the probability of each traffic stream state is defined from a simplification of the fundamental traffic diagrams of the Highway Capacity Manual [10]. These probabilities are shown in Figure 3, and are mathematically represented by the following equations:

- Speed:
  - Free flow probability: for speeds lower than v_1 probability is null and for speeds higher than v_2 probability is maximum, defining between both values a transition zone.

$$p_v(Free) = \begin{cases} 0 \quad \text{if } v \le v_1 \\ \dfrac{(v - v_1)}{v_2 - v_1} \ \text{if } v_1 < v < v_2 \\ 1 \ \text{if } v \ge v_2 \end{cases} \tag{1}$$

  - Stop & Go probability is the opposite than free flow ones.

$$p_v(Stop \,\&\, Go) = \begin{cases} 1 \quad \text{if } v \le v_1 \\ 1 - \dfrac{(v - v_1)}{v_2 - v_1} \ \text{if } \ v_1 < v < v_2 \\ 0 \ \text{if } v \ge v_2 \end{cases} \tag{2}$$

- Flow:
  - Free flow probability: for flows lower than q_1 the flow variable could not determine the traffic state (probability of 0,5) for flows higher than q_2 there is the maximum likelihood that the highways is operating under capacity. In this situation and taking into account that the aggregation period of data is of 3 minutes, the free flow probabilities are maximum. A transition zone is defined between both situations.

$$p_q(Free) = \begin{cases} 0{,}5 \quad \text{if } q \le q_1 \\ 0{,}5 + 0{,}5 \dfrac{(q - q_1)}{q_2 - q_1} \quad \text{if } q_1 \le q \le q_2 \\ 1 \ \text{if } q \ge q_2 \end{cases} \tag{3}$$

  - Stop & Go probabilities is the opposite of free flow ones. Take into account the paradox that when flow is maximum, stop&go probability is minimal.

$$p_q(Stop \,\&\, Go) = \begin{cases} 0{,}5 \quad \text{if } q \le q_1 \\ 0{,}5 - 0{,}5 \dfrac{(q - q_1)}{q_2 - q_1} \ \text{if } q_1 < q < q_2 \\ 0 \qquad \text{if } q \ge q_2 \end{cases} \tag{4}$$

- Occupancy:
  - Free flow probability: for occupancies lower than ρ_1 and for the free flow probability is maximum while for occupancies higher than ρ_2 probability is minimal, defining between both values the transition zone.

$$p_\rho(Free) = \begin{cases} 1 \text{ if } \rho \le \rho_1 \\ 1 - \dfrac{(\rho - \rho_1)}{\rho_2 - \rho_1} \text{ if } \rho_1 < \rho < \rho_2 \\ 0 \text{ if } \rho \ge \rho_2 \end{cases} \tag{5}$$

– Stop & Go probability is the opposite of free flow probability.

$$p_\rho(Stop\ \&\ Go) = \begin{cases} 0 \quad \text{if } \rho \le \rho_1 \\ \dfrac{(\rho - \rho_1)}{\rho_2 - \rho_1} \text{ if } \rho_1 < \rho < \rho_2 \\ 1 \quad \text{if } \rho \ge \rho_2 \end{cases} \tag{6}$$

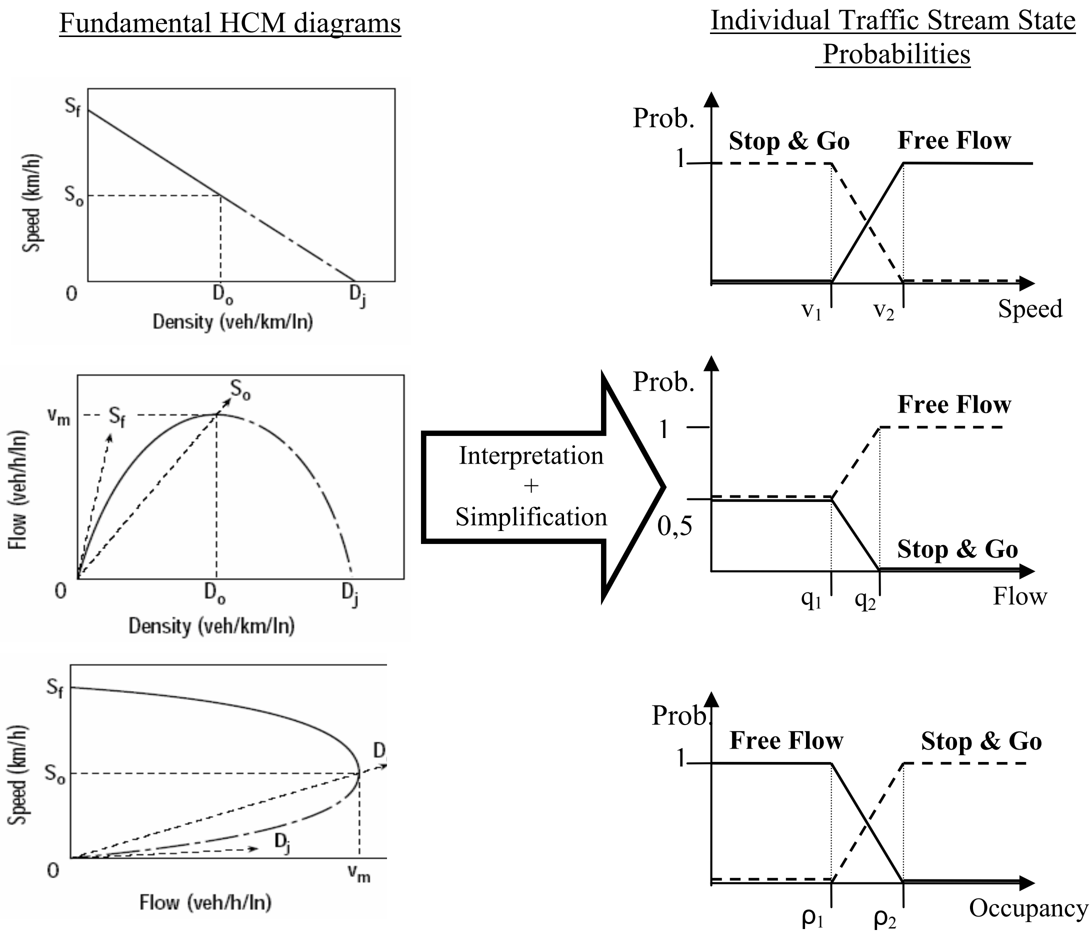


FIGURE 3 Individual Traffic Stream State Probabilities.

The limits of the transition zones should be determined in each particular highway section by using field data. However if this is not possible due to the unavailability of data, the values that define the different LoS in the Highway Capacity Manual [10] could be used.

In the present paper, the transition zone in the probability diagrams corresponds to the zone between LoS C and F, being the LoS A to C those that define the free flow zone. Stop&go

situation is assumed to be represented by LoS F of the Highway Capacity Manual. The Manual states that these values should not be taken for a direct application, so if possible it is recommended to calibrate them with some real information of the road. Note that a great precision is not necessary in the determination of these values, due to the intrinsic variability in the changes of traffic stream state. In addition, inaccuracies will be laminated in the fusion process of the three individual probabilities.

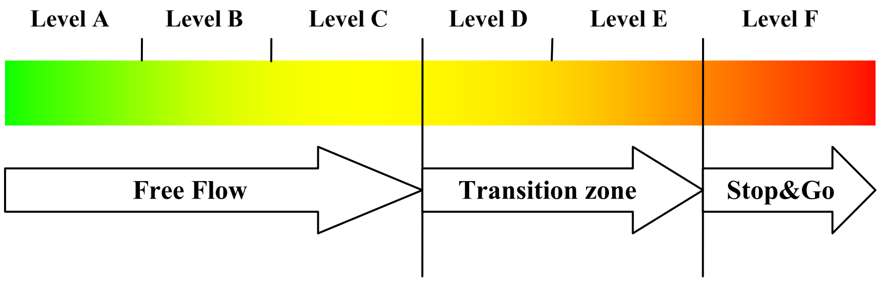


FIGURE 4 Relation between Level of Service and Flow State.

The fusion of the three individual probabilities guarantees consistency in the determination of the flow state, as any of the fundamental parameters is capable of characterizing itself the traffic stream state. It could be stated that any of the probability functions are able to represent the reality in critical moments by itself.

The simple arithmetic mean operator is chosen for the fusion, as the easiest example of a fuzzy logic with a context independent and constant mean behavior [27]:

$$p(Free) = \frac{p_v(Free) + p_q(Free) + p_\rho(Free)}{3} \tag{7}$$

$$p(Stop\ \&\ Go) = \frac{p_v(Stop\ \&\ Go) + p_q(Stop\ \&\ Go) + p_\rho(Stop\ \&\ Go)}{3} \tag{8}$$

Then the maximum likely state is determined as:

$$\text{if} \quad p(Free) \geq p(Stop\ \&\ Go) \rightarrow Free\ Flow \tag{9}$$

$$\text{if} \quad p(Stop\ \&\ Go) > p(Free) \rightarrow Stop\ \&\ Go \tag{10}$$

As it will be shown in the next section the results obtained in this flow state evaluation are very accurate.

3.1.2 Fusion 1 operator

For this travel time fusion two different algorithms were carefully analyzed. The first one was a weighted quadratic mean. The fused ITT was expressed in this case as:

$$ITT = \sqrt{\frac{((1-\varepsilon_{1,k}+\alpha_1)\cdot ITT_1^2) + ((1-\varepsilon_{2,k}+\alpha_2)\cdot ITT_2^2)}{2+\alpha_1+\alpha_2-\varepsilon_{1,k}-\varepsilon_{2,k}}} \tag{11}$$

Where:

ITT: Instantaneous travel time resultant of the fusion process

k: Traffic stream state in same time and space of ITT_1 and ITT_2.
α_i: Statistical quality parameter of every sample

If a new ε_3 error related to the fused ITT was defined, then it would be described by the following equation:

$$ITT = TTreal + (TTreal \cdot \varepsilon_3) = \\ = \sqrt{\frac{(1-|\varepsilon_1|)\cdot(TTreal + (\varepsilon_1 \cdot TTreal))^2 + (1-|\varepsilon_2|)\cdot(TTreal + (\varepsilon_2 \cdot TTreal))^2}{(2-|\varepsilon_1|-|\varepsilon_2|)}} \qquad (12)$$

Solving equation 12:

$$\varepsilon_3 = -1 + \sqrt{\frac{2\sum_{i=1}^{2}\varepsilon_i(1-|\varepsilon_i|)\cdot\left(1+\frac{\varepsilon_i}{2}\right)}{(2-|\varepsilon_1|-|\varepsilon_2|)} + 1} \qquad (13)$$

Analyzing this result, it was found that the accuracy of the fused value had an inconsistent behavior depending on the values of ε_1 and ε_2, being in occasions grater in absolute value than both of the original errors.

Given the inconsistency of the proposed operator it was decided to analyze a context dependent operator with constant mean behavior [27]. Since it is context dependent, it is necessary to define three of contexts A, B and C. In each context the data fusion algorithm will follow a different expression. Their definitions are:

- **A context**:

$$ITT_i + \varepsilon_i + (1-\alpha_i) \geq ITT_j + \varepsilon_j + (1-\alpha_j) \;\cup\; ITT_i - \varepsilon_i - (1-\alpha_i) \leq ITT_j - \varepsilon_j - (1-\alpha_j) \qquad (14)$$

where $i, j = 1,2 \quad i \neq j$

- **B context**:

$$ITT_i + \varepsilon_i + (1-\alpha_i) \geq ITT_j + \varepsilon_j + (1-\alpha_j) \;\cup\; ITT_i - \varepsilon_i - (1-\alpha_i) \geq ITT_j - \varepsilon_j - (1-\alpha_j) \qquad (15)$$

where $i, j = 1,2 \quad i \neq j$

- **C context**:

$$ITT_i + \varepsilon_i + (1-\alpha_i) \geq ITT_j + \varepsilon_j + (1-\alpha_j) \;\cup\; ITT_i - \varepsilon_i - (1-\alpha_i) \geq ITT_j + \varepsilon_j + (1-\alpha_j) \qquad (16)$$

where $i, j = 1,2 \quad i \neq j$

Note that **ε** error parameters depend on the estimation algorithm (1 or 2) but also on the traffic state, having minimized notation to simplify the comprehension of the equations.

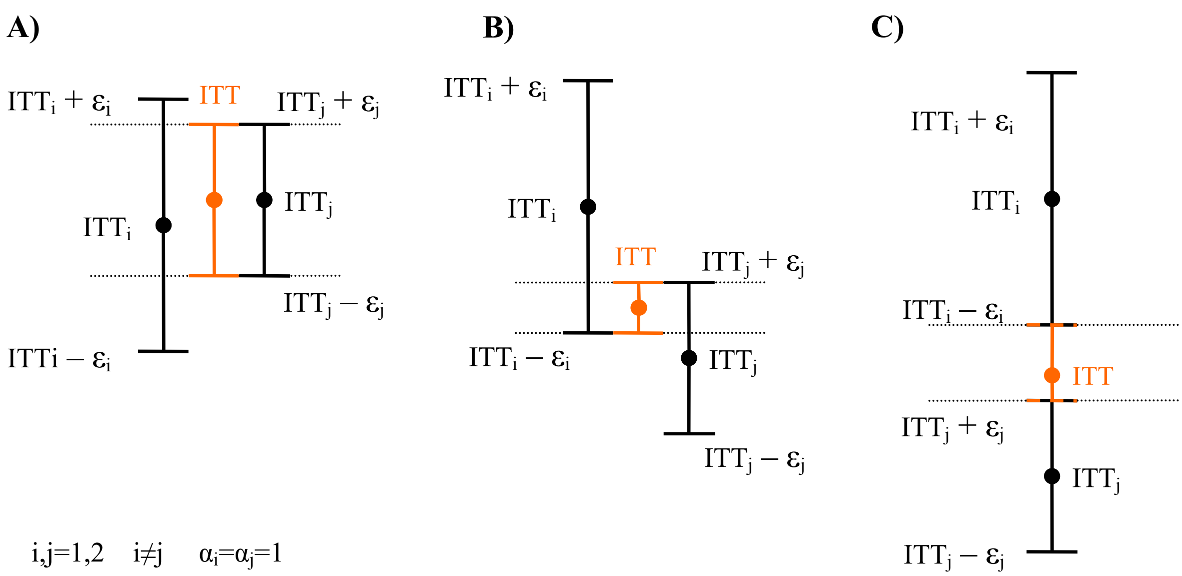


FIGURE 5: ITT Fusion Contexts.

So, given:

$$l = \left(ITT_i - \varepsilon_i - (1-\alpha_i)\right) - \left(ITT_j + \varepsilon_j + (1-\alpha_j)\right) \tag{17}$$

The fusion operators are:

$$ITT = \begin{cases} \dfrac{\left(ITT_j + \varepsilon_j + (1-\alpha_j)\right) + \left(ITT_j - \varepsilon_j - (1-\alpha_j)\right)}{2} & \text{if A} \\ \dfrac{\left(ITT_j + \varepsilon_j + (1-\alpha_j)\right) + \left(ITT_i - \varepsilon_i - (1-\alpha_i)\right)}{2} & \text{if B} \\ \left(ITT_j + \varepsilon_j + (1-\alpha_j)\right) + \left(l \cdot \dfrac{\left(\varepsilon_j + (1-\alpha_j)\right)}{\left(\varepsilon_i + (1-\alpha_i)\right) + \left(\varepsilon_j + (1-\alpha_j)\right)}\right) & \text{if C} \end{cases} \tag{18}$$

The analytic expression for the error in this case is:

$$\varepsilon_3 = \begin{cases} \min\left(\varepsilon_i ; \varepsilon_j\right) & \text{if A} \\ \left(ITT_j + \varepsilon_j + (1-\alpha_j)\right) - \left(ITT_i - \varepsilon_i - (1-\alpha_i)\right) \rightarrow \varepsilon_3 \le \min\left(\varepsilon_i ; \varepsilon_j\right) & \text{if B} \\ l = \left(ITT_i - \varepsilon_i - (1-\alpha_i)\right) - \left(ITT_j + \varepsilon_j + (1-\alpha_j)\right) & \text{if C} \end{cases} \tag{19}$$

In Figure 5 it is possible to observe that the operator is more consistent, being its behavior clearly determined by the context. In A and B contexts the resultant error is smaller or equal to the smallest of the errors ε_1 and ε_2 while in context C this could not be true. Context C must be always avoided since it represents a mistake in the calibration of ε_1 and ε_2. This means that a correct calibration of these parameters is basic for the correct functioning of the algorithm.

3.2 Fusion 2

This second data fusion process starts with two different predicted travel times PTT_1 and PTT_2, with the objective of obtaining a unique PTT more precise, reliable and with more temporal coverage then both previous ones. Similar than fusion 1, the following inputs are needed:

- x_0: Start point of the considered section.
- x_1: Final point of the considered section.
- t: Finishing time instant of the capture of information.

- f_i: Temporal frequency (t^{-1}) of PTT_1 and PTT_2 data actualization.
- PTTs: Predicted Travel Times 1 and 2 to be fused.
- TT_{real}: Real Travel Time for the fusion calibration.

3.2.1 Spatial and temporal alignment

Unlike the fusion 1, in the fusion 2 process the information it is not provided by the same data source and therefore the data are not equally located in space and time. So, a spatial and temporal alignment is needed before the data can be fused.

In the present paper the chosen dimension for the alignment has been the PTT_1 temporal dimension and PTT_2 spatial dimension. It has been supposed that the spatial dimension of PTT_2 is the larger of both spatial dimensions, and that the temporal one of PTT_1 is the smallest of temporal dimensions. That agrees with the real case analyzed in the section 4 of this paper.

3.2.1.1 Spatial Alignment. PTT_1 spatial dimension will be aligned to PTT_2. It is necessary to define:

- $x_1,\ldots,x_n$: Section limits for PTT_1
- x_I, x_{II}: Section limits for PTT_2
- t_0: Temporal instant of PTT_1 update

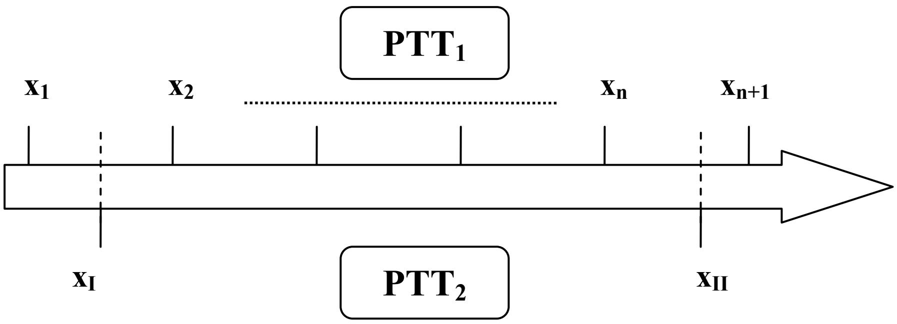


FIGURE 6 Spatial Alignment.

The resulting spatial alignment equation can be expressed as:

$$PTT_{1,x_I \to x_{II}}(t_0) = \left(PTT_{1,x_1 \to x_2}(t_0) \cdot \left(\frac{x_2 - x_I}{x_2 - x_1} \right) \right) + \sum_{i=2}^{n-1} PTT_{1,x_i \to x_{i+1}}(t_i) + \left(PTT_{1,x_n \to x_{n+1}}(t_n) \cdot \left(\frac{x_{n+1} - x_{II}}{x_{n+1} - x_n} \right) \right) \quad (20)$$

Where:

$$t_i = t_0 + \left(PTT_{1,x_1 \to x_2}(t_0) \cdot \left(\frac{x_2 - x_I}{x_2 - x_1} \right) \right) + \sum_{j=3}^{i} PTT_{1,x_{j-1} \to x_j}(t_{j-1}) \qquad \forall i = 1, n \tag{21}$$

Since in the data fusion system has not advanced forecasting capabilities:

$$PTT_1(t_i) = PTT_1(t_0) \qquad t_i > t_0 \tag{22}$$

Simplifying from *(21)* and using *(22)* results:

$$PTT_{1,x_I \to x_{II}}(t) = \left(PTT_{1,x_1 \to x_2}(t) \cdot \left(\frac{x_2 - x_I}{x_2 - x_1} \right) \right) + \sum_{i=2}^{n-1} PTT_{1,x_i \to x_{i+1}}(t) + \left(PTT_{1,x_n \to x_{n+1}}(t) \cdot \left(\frac{x_{n+1} - x_{II}}{x_{n+1} - x_n} \right) \right) \tag{23}$$

3.2.1.2 Temporal Alignment The temporal alignment of PTT_2 has to be done increasing artificially its updating frequency. It is necessary to define:

- f_1: Temporal frequency (t^{-1}) of PTT_1 data actualization.
- f_2: Temporal frequency (t^{-1}) of PTT_2 data actualization. $f_2 < f_1$
- t_0: Temporal instant of PTT_2 update

Not considering forecasting capabilities, the best temporal alignment is:

$$PTT_2\left(t_0 + \left(\frac{m}{f_1} \right) \right) = PTT_2(t_0) \qquad \text{where} \qquad m = 0, \ldots, \left(\frac{f_1}{f_2} - 1 \right) \tag{24}$$

3.2.2 PTT's fusion operator

Once PTT_1 and PTT_2 are spatially and temporally aligned it is possible to proceed to the calibration of the fusion algorithm. This operator uses the probabilistic logic [28], based on Bayes' Theory:

$$p(E \mid x_1, x_2) = p(E) \frac{p(x_2 \mid E, x_1) p(E \mid x_1)}{p(x_2 \mid x_1)} \tag{25}$$

If x_1 and x_2 are independent variables, then results:

$$p(E \mid x_1, x_2) = \frac{p(x_2 \mid E) p(x_1 \mid E) p^2(E)}{p(x_1) p(x_2)} \tag{26}$$

Applying this rule to travel time variables, TT_{real}, PTT_1 and PTT_2:

$$p(TT_{real} \mid PTT_1, PTT_2) = \frac{p(PTT_2 \mid TT_{real}) p(PTT_1 \mid TT_{real}) p^2(TT_{real})}{p(PTT_1) p(PTT_2)} \tag{27}$$

The probabilities $p(PTT_1 \mid TT_{real})$ and $p(PTT_2 \mid TT_{real})$ are obtained by a statistic analysis of the calibration samples. In the determination of these probabilities, PTT_1 and PTT_2 are rounded up to the next whole number, in order to obtain more representative relations. This does not affect the quality of the results, because the user perception of travel time is never lower than this minute unit.

Once the conditional probabilities $p(TT_{real} \mid PTT_1, PTT_2)$ are determined, a modified maximum posteriori probability decision rule is chosen. This decision rule is a modification of the rule of maximum likelihood, which results in a more stable behavior. It consists in taking into account the occurrence probabilities of the two TT_{real} values adjacent to the most likely one. This decision rule can be formulated as follows:

$$
\begin{gathered}
TT_{real_k} = \arg\max_{1 \le i \le r} \{a + b + c\} \\
a = 0{,}5 \cdot p\left(TT_{real_i} \mid PTT_1, PTT_2\right) \qquad b = 0{,}25 \cdot p\left(TT_{real_{i-1}} \mid PTT_1, PTT_2\right) \\
c = 0{,}25 \cdot p\left(TT_{real_{i+1}} \mid PTT_1, PTT_2\right) \qquad r: \text{Subset number of } \mathrm{TT}_{\mathrm{real}}
\end{gathered}
\tag{28}
$$

The decision to leave result void is taken if the probability value does not overcome a threshold defined by the user of the system. This situation denotes little probability that PTT_1 and PTT_2 values coincide in the same section and time interval (e.g. it is slightly probable that PTT_1 = 1 min. and PTT_2=15 min.). A great number of voids in the running phase of the fusion shows a great weakness of the travel time estimation algorithms.

From the Bayes' Theory it is also possible to obtain the accuracy of the result. Since multiplying conditional probabilities, part of the sample information gets lost, the uncertainty of the result related with a pair of PTT_1 and PTT_2, could be defined as:

$$
I(PTT_1, PTT_2) = 1 - \sum_{i=1}^{r} p\left(\mathrm{TT}_{\mathrm{real_i}} \middle| PTT_1, PTT_2\right) \tag{29}
$$

The goal of any travel time estimation system should be the reduction of this uncertainty, as this parameter is a good reliability indicator of the final result.

Once calibrated the running of the fusion algorithm is very simple because after the spatial and temporal alignment of PTT_1 and PTT_2, only it is needed to check the table of probabilities and obtain the corresponding fused PTT.

4. APPLICATION TO THE AP-7 HIGHWAY IN SPAIN

The data fusion technique proposed in this paper was tested in the AP-7 toll highway in Spain. The AP-7 highway runs along the Mediterranean cost corridor, from the French border to the Gibraltar strait. Nevertheless, the pilot test was restricted to the north east stretch of the highway between "La Roca del Vallès" and "Lloret de Mar" toll plazas, near Barcelona. This stretch is approximately 30 km long.

The pilot test was performed with the July 10th 2005 afternoon data in southbound direction. This was a very conflictive period in terms of traffic, as it was a very sunny Sunday of July when a lot of people use this stretch of the AP-7 highway for returning to the Barcelona after a weekend at the coast.

The surveillance equipment installed on the highway consists of 10 loop detectors (i.e. an average of 1 detector every 3 km). Moreover, the highway operates in a closed tolling system, where each vehicle entering the highway receives a ticket (real -usually a card with magnetic band- or virtual –using an electronic toll collection ETC device-), which is collected at the exit. Since the ticket includes the entry point, and the exact time of entry, by cross-checking entry and exit data, the precise time taken to travel along the itinerary (route) can be determined.

On one hand, from loop detector data, ITT_1 is obtained using a simple spot speed algorithm [24], and ITT_2 using Nam's cumulative flow balance algorithm [29]. The updating time interval of these ITTs is 3 minutes. On the other hand RTT is obtained from toll ticket data using the algorithm proposed in [25] and being updated every 15 minutes.

4.1 Fusion 1

4.1.1 Traffic Stream State Evaluation

The limits of the transition zones in the probability diagrams (see Fig. 3) were determined using the Highway Capacity Manual's [10] parameters for a three lane basic freeway segment in a metropolitan environment (see table 1). Note that the HCM's density parameters must be traduced into occupancy to be applied in the data fusion process.

TABLE 1 Traffic Stream State Transition Limits

	Transition start	Transition End
Speed	V_1 = 60 km/h	V_2 = 70 km/h
Flow	q_1 = 4,340 veh/h	q_2 = 6,200 veh/h
Density	k_1 = 20 veh/km	k_2 = 27 veh/km

The accuracy of the traffic state evaluation is validated comparing the real measured travel times in the different classified states. The results are found to be relatively accurate (see table 2 results).

TABLE 2 Accuracy of Traffic Stream State Evaluation

	Free Flow Intervals	Stop & Go Flow Intervals	No Data Intervals	Percentage $TTmax_{free}$>$TTmin_{stop}$
Loop 1	28	13	4	35%
Loop 2	17	27	1	33%
Loop 3	13	31	1	15%
Loop 4	2	28	15	0%
Loop 5	0	31	14	0%

Once the traffic stream state is determined, the different values of the accuracy of the ITT estimation algorithms (ε) could be obtained. Results are shown in table 3:

TABLE 3 Original Average Accuracy of the Instantaneous Travel Time Algorithms (ε)

ε	Free Flow	Stop & Go
ITT_1	$\varepsilon_{1,L}$ = 0,327	$\varepsilon_{1,S}$ = 0,376
ITT_2	$\varepsilon_{2,L}$ = 0,285	$\varepsilon_{2,S}$ = 0,388

It is stated that both ITT algorithms have a similar behavior, being more accurate in free flow conditions and resulting in similar errors in all the situations.

4.1.2 Calculation of fused ITT

After calibrating the ε parameters, ITT is easily obtained applying equation 11. It is interesting to analyze the improvement achieved with this first fusion. This improvement is monitored by the first fusion accuracy parameter ε_3 defined in equation 19 and comparing it to the original algorithms accuracy (ε_1 and ε_2). Average results shown in table 4a are found to be positive, in spite of the lack of information for the accurate calibration of the algorithm.

4.2 Fusion 2

The last step consists in the second fusion process taking into account RTT information. Equations 20 to 29 are applied to obtain PTT resulting from the fusion of PTT_1 and PTT_2. The average improvement achieved in terms of the reduction of the average quadratic mean relative error [26] is shown in table 4b.

TABLE 4 Data Fusion Technique Average improvement

	Free Flow		Stop & Go Flow		Global	
	ITT_1	ITT_2	ITT_1	ITT_2	Free Flow	Stop & Go
Improvement in relation to original ITT values in each traffic state	20,7%	9,5%	6,0%	9,2%	8,4%	5,4%

(a) ITT fusion 1 improvement

	Quadratic Mean Relative Error
PTT_1	0,19
PTT_2	0,47
PTT	0,11

(b) PTT fusion 2 improvement

5. CONCLUSIONS AND FURTHER RESEARCH

The paper presents a simple approach for reliable road travel time estimation, using data fusion techniques. The system can be easily put into practice with the existing infrastructure, and is able to use data obtained from any kind of sensor in any type of road link.

The proposed system is capable of determining traffic stream state using a probabilistic approach, combining flow, speed and occupancy data. This state evaluation is an intermediate result for obtaining travel time estimations with better spatial and temporal coverage. This fused travel times are also more reliable than the initial ones and more accurate if the calibration process is carefully analyzed.

The results of the pilot test carried out on the AP-7 highway in Spain indicate the importance of the calibration in the performance of this process and the suitability of the data fusion system for a better usage of the different surveillance equipment already installed in the roads.

Further developments are possible with the model, such as an initial data control system to verify the original quality of data, a flow traffic stream evaluation in multiple states or the introduction of forecasting capabilities in the algorithm. This would lead to a more complex and probably more accurate data fusion system.

ACKNOWLEDGMENTS
This research was performed within a contract managed by Mr. Josep A. Grau and in cooperation with Abertis, the Spanish leading private transportation and communications infrastructures management corporation. The authors acknowledge the collaboration and data providing of Joan Altarriba and José V. Solano from Abertis. The contents of this report reflect the views of the authors who are responsible for the facts and accuracy of the data presented herein.

REFERENCES
1. Palen, J. The Need for Surveillance in Intelligent Transportation Systems. *Intellimotion*, 6:1, pp. 1-3, 10. University of California PATH, Berkeley, CA, 1997.
2. Hopkin, J., D. Crawford, and I. Catling. *Travel Time Estimation*. Summary of the European Workshop organized by the SERTI project, Avignon, November 2001.
3. Klein, L. A. 2001. *Sensor Technologies and Data Requirements for ITS*. Artech House Publishers.
4. Martin, P. T., Y. Feng and X. Wang. 2003. *Detector Technology Evaluation*. Department of Civil and Environmental Engineering, University of Utah sponsored by US Department of Transportation.
5. Skesz, S.L. 2001, *"State-of-the-Art" Report on Non-Traditional Traffic Counting Methods*. Arizona Department of Transportation.
6. US Department of Transportation. 2006. *T-REX Project: ITS Unit Costs Database*. http://www.itscosts.its.dot.gov.
7. Palacharla, P.V. and P. C. Nelson. 1999. Application of fuzzy logic and neural networks for dynamic travel time estimation. *International Transactions in Operational Research 6*.
8. Bm vit, FFG and Trust consult. 2006. Intelligente Infrastuktur. *3.I2-Vernetzungsworkshop.*
9. Sazi Murat,Y. 2006. Comparison of fuzzy logic and artificial neural networks approaches in vehicle delay modeling. *Transportation Research Part C* 14.
10. TRB (Transportation Research Board). 2000. *Highway Capacity Manual.* National Academy of Sciences.
11. Weizhong Z., D.H. Lee and Q. Shi. Short-Term Freeway Traffic Flow Prediction: Bayesian Combined Neural Network Approach. *Journal of Transportation Engineering, ASCE.*
12. El Faouzi, N.E. 2005. Bayesian and evidential approaches for traffic data fusion: Mehodological issues and case study. *85 Transportation Research Board Annual Meeting.*
13. El Faouzi, N.E.. 2005. Fusion de données en I.T. Estimation multisource du temps de parcours. *1er Séminaire TISIC,Paris*. INRETS.
14. El Faouzi, N.E. and C. Simon. 2000. Fusion de données pour l'estimation des temps de parcours via la théorie de l'evidence. *Recherche transports sécurité* nº 68:15-27.
15. AREA. 2006. Système d'information Temps de Parcours sur le réseau autoroutier AREA. *Asecap PULA 2006* pag. 105-115.
16. Ferré, J. 2005. Quality Control and Fusion of Road and Weather Data at SAPN. *Centrico Briefing Note.*

17. Guiol, R. (ESCOTA) and N. Schwab (ASF). 2006. Vers une information «temps de parcours» toujours plus performante et adaptée aux besoins de la clientèle: L'expérience ASF-ESCOTA. *Asecap PULA 2006* pag. 97-104.
18. Peterson, A. (SRA). 2006. Travel Times – A Multi-Purpose Issue. *4th Conference of the Euro-regional Projects, I2Tern*.
19. Scott, G. (Transport Scotland). 2006. Scottish Nacional Journey System and Development of Data Fusion. *4th Conference of the Euro-regional Projects, I2Tern*.
20. ddg (2005) Gesellschaft für Verkehrdaten mbH. *Von der Meßwerterfassung bis zur automatisch generierten Verkehrsmeldung*.
21. van Lint, J.W.C., S.P. Hoogendoorn and H.J van Zuylen. 2005. Accurate freeway travel time prediction with state-space neural networks under missing data. *Transportation Research Part C* 13.
22. Park, T. and S. Lee. 2004. *A Bayesian Approach for Estimating Link Travel Time on Urban Arterial Road Network*. A. Laganà et al. (Eds.): ICCSA 2004, LNCS 3043. Springer-Verlag Berlin Heidelberg 2004.
23. Turner, S. M., W. L. Eisele, R. J. Benz, and D.J. Holdener. *Travel Time Data Collection HandBook*. Research Report FHWA-PL-98-035. Texas Transportation Institute, Texas A&M University System, College Park, 1998.
24. Soriguera, F., D. Abeijón, D. Rosas, L. Thorson and F. Robusté. (2007). Highway Travel Time Estimation from Multiple Data Sources. Proceedings of the *11th World Conference on Transport Research*. 24-28 June 2007. Berkeley, California.
25. Soriguera F., L. Thorson and F. Robuste. 2007. Travel Time Measurement Using Toll Infrastructure. *Transportation Research Board*. In edition. TRB, National Research Council, Washington, D.C.
26. Bloch, I. 1996. Information Combination Operators for Data Comparative Review with Classification. IEEE Transactions on Systems, Man and Cybernetics – Part A: *Systems and Humans* Vol. 26 nº 1
27. Cohen, S., J. Bomier, N. Schwab. 2002. Estimation et prediction des temps de parcours sur un resau autoroutier. *Modelisation du Trafic, Actes INRETS* n° 83 pag. 25-36
28. Hall, David L. y Sarah A. H. McMullen. 2004. *Mathematical Techniques in Multisensor Data Fusion*. Artech House Publishers.
29. Nam D. H. y D. R. Drew. 1996. Traffic Dynamics: Method for estimating freeway travel times in real time from flow measurements. *Journal of Transportation Engineering*, vol. 122, pp. 185-191.