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 ITT₁ (form a spot speed algorithm) and ITT₂ (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:
- \( x₀ \): Start point of the considered section.
- \( x₁ \): 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.
- \( TT_{\text{real}} \): Real Travel Time for the fusion calibration.

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

![FIGURE 2 Fusion 1 Structure.](image)

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 process by computing the Quadratic Mean Relative Error [26] of the real measure (\( TT_{\text{real}} \)) in relation to the estimated value (\( \text{ITT}_1 \) or \( \text{ITT}_2 \)).

**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 (\( \varepsilon \)). 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.

\[
p_{v}(\text{Free}) = \begin{cases} 
0 & \text{if } v \leq v_1 \\
\frac{(v-v_1)}{v_2-v_1} & \text{if } v_1 < v < v_2 \\
1 & \text{if } v \geq v_2
\end{cases}
\]  

(1)

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

\[
p_{v}(\text{Stop & Go}) = \begin{cases} 
1 & \text{if } v \leq v_1 \\
1-\frac{(v-v_1)}{v_2-v_1} & \text{if } v_1 < v < v_2 \\
0 & \text{if } v \geq v_2
\end{cases}
\]  

(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.

\[
p_{q}(\text{Free}) = \begin{cases} 
0,5 & \text{if } q \leq q_1 \\
0,5 + 0,5 \frac{(q-q_1)}{q_2-q_1} & \text{if } q_1 \leq q \leq q_2 \\
1 & \text{if } q \geq q_2
\end{cases}
\]  

(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.

\[
p_{q}(\text{Stop & Go}) = \begin{cases} 
0,5 & \text{if } q \leq q_1 \\
0,5 - 0,5 \frac{(q-q_1)}{q_2-q_1} & \text{if } q_1 < q \leq q_2 \\
0 & \text{if } q \geq q_2
\end{cases}
\]  

(4)

• Occupancy:
  – Free flow probability: for occupancies lower than \( \rho_1 \) and for the free flow probability is maximum while for occupancies higher than \( \rho_2 \) probability is minimal, defining between both values the transition zone.
\[
p_{\rho}(\text{Free}) = \begin{cases} 
1 & \text{if } \rho \leq \rho_1 \\
1 - \frac{(\rho - \rho_1)}{\rho_2 - \rho_1} & \text{if } \rho_1 < \rho < \rho_2 \\
0 & \text{if } \rho \geq \rho_2 
\end{cases}
\]
(5)

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

\[
p_{\rho}(\text{Stop & Go}) = \begin{cases} 
0 & \text{if } \rho \leq \rho_1 \\
\frac{(\rho - \rho_1)}{\rho_2 - \rho_1} & \text{if } \rho_1 < \rho < \rho_2 \\
1 & \text{if } \rho \geq \rho_2 
\end{cases}
\]
(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](image)

**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]:

\[
p(\text{Free}) = \frac{p,F(\text{Free}) + p,F(\text{Free}) + p,F(\text{Free})}{3} \tag{7}
\]

\[
p(\text{Stop} & \text{ Go}) = \frac{p,F(\text{Stop} & \text{ Go}) + p,F(\text{Stop} & \text{ Go}) + p,F(\text{Stop} & \text{ Go})}{3} \tag{8}
\]

Then the maximum likely state is determined as:

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

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

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:

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

Where:

- *ITT*: Instantaneous travel time resultant of the fusion process
Traffic stream state in same time and space of \( ITT_1 \) and \( ITT_2 \).

\( \alpha_i \): Statistical quality parameter of every sample

If a new \( \varepsilon_3 \) error related to the fused ITT was defined, then it would be described by the following equation:

\[
ITT = TT_{real} + (TT_{real} \cdot \varepsilon_3) = \\
\sqrt{\left(1 - |\varepsilon_1|\right) \cdot (TT_{real} + (1 - |\varepsilon_2|) \cdot TT_{real} + \varepsilon_1 \cdot TT_{real})^2 + \left(1 - |\varepsilon_2|\right) \cdot (TT_{real} + \varepsilon_2 \cdot TT_{real})^2} \\
(2 - |\varepsilon_1| - |\varepsilon_2|)}
(12)
\]

Solving equation 12:

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

Analyzing this result, it was found that the accuracy of the fused value had an inconsistent behavior depending on the values of \( \varepsilon_1 \) and \( \varepsilon_2 \), being in occasions greater 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:**
  \[
  ITT_i + \varepsilon_i + (1 - \alpha_i) \geq ITT_j + \varepsilon_j + (1 - \alpha_j) \quad \cup \quad ITT_i - \varepsilon_i - (1 - \alpha_i) \leq ITT_j - \varepsilon_j - (1 - \alpha_j) \quad \text{where} \quad i, j = 1, 2 \quad i \neq j
(14)
\]

- **B context:**
  \[
  ITT_i + \varepsilon_i + (1 - \alpha_i) \geq ITT_j + \varepsilon_j + (1 - \alpha_j) \quad \cup \quad ITT_i - \varepsilon_i - (1 - \alpha_i) \geq ITT_j - \varepsilon_j - (1 - \alpha_j) \quad \text{where} \quad i, j = 1, 2 \quad i \neq j
(15)
\]

- **C context:**
  \[
  ITT_i + \varepsilon_i + (1 - \alpha_i) \geq ITT_j + \varepsilon_j + (1 - \alpha_j) \quad \cup \quad ITT_i - \varepsilon_i - (1 - \alpha_i) \geq ITT_j + \varepsilon_j + (1 - \alpha_j) \quad \text{where} \quad i, j = 1, 2 \quad i \neq j
(16)
\]

Note that \( \varepsilon \) 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:

\[ I = (ITT_i - \varepsilon_i - (1 - \alpha_i)) - (ITT_j + \varepsilon_j + (1 - \alpha_j)) \]  

(17)

The fusion operators are:

\[
ITT = \begin{cases} 
\frac{(ITT_j + \varepsilon_j + (1 - \alpha_j)) + (ITT_i - \varepsilon_i - (1 - \alpha_i))}{2} & \text{if A} \\
\frac{(ITT_j + \varepsilon_j + (1 - \alpha_j)) + (ITT_i - \varepsilon_i - (1 - \alpha_i))}{2} & \text{if B} \\
(ITT_j + \varepsilon_j + (1 - \alpha_j)) + \frac{(\varepsilon_j + (1 - \alpha_j))}{l + (\varepsilon_j + (1 - \alpha_j)) + (\varepsilon_j + (1 - \alpha_j)))} & \text{if C}
\end{cases}
\]  

(18)

The analytic expression for the error in this case is:

\[
\varepsilon_3 = \begin{cases} 
\min(\varepsilon_i; \varepsilon_j) & \text{if A} \\
(ITT_j + \varepsilon_j + (1 - \alpha_j)) - (ITT_i - \varepsilon_i - (1 - \alpha_i)) \rightarrow \varepsilon_3 \leq \min(\varepsilon_i; \varepsilon_j) & \text{if B} \\
l = (ITT_i - \varepsilon_i - (1 - \alpha_i)) - (ITT_j + \varepsilon_j + (1 - \alpha_j)) & \text{if C}
\end{cases}
\]  

(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 \( \varepsilon_i \) and \( \varepsilon_j \) while in context C this could not be true. Context C must be always avoided since it represents a mistake in the calibration of \( \varepsilon_i \) and \( \varepsilon_j \). 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 PTT1 and PTT2, with the objective of obtaining a unique PTT more precise, reliable and with more temporal coverage than both previous ones. Similar than fusion 1, the following inputs are needed:
- \( x_0 \): Start point of the considered section.
- \( x_f \): Final point of the considered section.
- \( t_f \): Finishing time instant of the capture of information.
- \( f_i \): Temporal frequency \((t^i)\) of PTT1 and PTT2 data actualization.
- PTTs: Predicted Travel Times 1 and 2 to be fused.
- TTreal: 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 PTT1 temporal dimension and PTT2 spatial dimension. It has been supposed that the spatial dimension of PTT2 is the larger of both spatial dimensions, and that the temporal one of PTT1 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. PTT1 spatial dimension will be aligned to PTT2. It is necessary to define:
- \( x_1, \ldots, x_n \): Section limits for PTT1
- \( x_I, x_{II} \): Section limits for PTT2
- \( t_0 \): Temporal instant of PTT1 update

![FIGURE 6 Spatial Alignment.](image)

The resulting spatial alignment equation can be expressed as:

\[
PTT_{x_I \rightarrow x_{II}} (t_0) = \left( PTT_{x_I \rightarrow x_2} \left( t_0 \right) \cdot \left( \frac{x_2 - x_I}{x_2 - x_1} \right) \right) + \sum_{i=2}^{n-1} PTT_{x_i \rightarrow x_{i+1}} \left( t_f \right) + \left( PTT_{x_{n-II} \rightarrow x_{n}} \left( t_f \right) \cdot \left( \frac{x_{n+1} - x_{n-II}}{x_{n+1} - x_n} \right) \right) \quad (20)
\]

Where:
\[ t_i = t_0 + \left( \frac{\sum_{j=1}^{i} PTT_{L_{x_i \rightarrow x_j}}(t_j) \cdot \left( x_{j} - x_{j-1} \right)}{x_{j} - x_{j-1}} \right) + \sum_{j=1}^{i} PTT_{L_{x_i \rightarrow x_j}}(t_{j-1}) \quad \forall i = 1, n \]  \hspace{1cm} (21) 

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

\[ PTT_i(t_i) = PTT_i(t_0) \quad t_i > t_0 \]  \hspace{1cm} (22) 

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

\[ PTT_{L_{x_i \rightarrow x_n}}(t) = \left( PTT_{L_{x_i \rightarrow x_n}}(t) \cdot \left( x_{n} - x_{n-1} \right) \right) + \sum_{i=2}^{n-1} PTT_{L_{x_i \rightarrow x_n}}(t) + \left( PTT_{L_{x_i \rightarrow x_n}}(t) \cdot \left( x_{n} - x_{n-1} \right) \right) \]  \hspace{1cm} (23) 

3.2.1.2 **Temporal Alignment**  
The temporal alignment of PTT2 has to be done increasing artificially its updating frequency. It is necessary to define:

- \( f_1 \): Temporal frequency (t^{-1}) of PTT1 data actualization.
- \( f_2 \): Temporal frequency (t^{-1}) of PTT2 data actualization.  \( f_2 < f_1 \)
- \( t_0 \): Temporal instant of PTT2 update

Not considering forecasting capabilities, the best temporal alignment is:

\[ PTT_{2}(t_0 + \left( \frac{m}{f_1} \right)) = PTT_{1}(t_0) \quad \text{where} \quad m = 0, \ldots, \left( \frac{f_1}{f_2} - 1 \right) \]  \hspace{1cm} (24) 

3.2.2 **PTT’s fusion operator**  
Once PTT1 and PTT2 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:

\[ p(E \mid x_1, x_2) = p(E) \frac{p(x_2 \mid E, x_1)p(E \mid x_1)}{p(x_2 \mid x_1)} \]  \hspace{1cm} (25) 

If \( x_1 \) and \( x_2 \) are independent variables, then results:

\[ p(E \mid x_1, x_2) = \frac{p(x_2 \mid E)p(x_1 \mid E)p^2(E)}{p(x_1)p(x_2)} \]  \hspace{1cm} (26) 

Applying this rule to travel time variables, TT\(_{\text{real}}\), PTT1 and PTT2:

\[ p(TT_{\text{real}} \mid PTT_1, PTT_2) = \frac{p(PP_{TT_2} \mid TT_{\text{real}})p(PP_{TT_1} \mid TT_{\text{real}})p^2(TT_{\text{real}})}{p(PP_{TT_1})p(PP_{TT_2})} \]  \hspace{1cm} (27)
The probabilities \( p(PTT_1 \mid TT_{\text{real}}) \) and \( p(PTT_2 \mid TT_{\text{real}}) \) are obtained by a statistic analysis of the calibration samples. In the determination of these probabilities, PTT1 and PTT2 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 \( p(TT_{\text{real}} \mid PTT_1, PTT_2) \) 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 \( TT_{\text{real}} \) values adjacent to the most likely one. This decision rule can be formulated as follows:

\[
TT_{\text{real},r} = \arg \max_{1 \leq r \leq 5} \{a + b + c\}
\]

\[
a = 0.5 \cdot p(TT_{\text{real}} \mid PTT_1, PTT_2) \quad b = 0.25 \cdot p(TT_{\text{real},-1} \mid PTT_1, PTT_2) \quad c = 0.25 \cdot p(TT_{\text{real},+1} \mid PTT_1, PTT_2) \quad r : \text{Subset number of } TT_{\text{real}}
\]

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 PTT1 and PTT2 values coincide in the same section and time interval (e.g. it is slightly probable that PTT1 = 1 min. and PTT2 = 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 PTT1 and PTT2, could be defined as:

\[
I(PTT_1, PTT_2) = 1 - \sum_{i=1}^{r} p(TT_{\text{real}} \mid PTT_1, PTT_2)
\]

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 PTT1 and PTT2, 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, ITT₁ is obtained using a simple spot speed algorithm [24], and ITT₂ 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**

<table>
<thead>
<tr>
<th></th>
<th>Transition start</th>
<th>Transition End</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed</td>
<td>$V_1 = 60$ km/h</td>
<td>$V_2 = 70$ km/h</td>
</tr>
<tr>
<td>Flow</td>
<td>$q_1 = 4,340$ veh/h</td>
<td>$q_2 = 6,200$ veh/h</td>
</tr>
<tr>
<td>Density</td>
<td>$k_1 = 20$ veh/km</td>
<td>$k_2 = 27$ veh/km</td>
</tr>
</tbody>
</table>

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**

<table>
<thead>
<tr>
<th></th>
<th>Free Flow Intervals</th>
<th>Stop &amp; Go Flow Intervals</th>
<th>No Data Intervals</th>
<th>Percentage TTmaxfree&gt;TTminstop</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loop 1</td>
<td>28</td>
<td>13</td>
<td>4</td>
<td>35%</td>
</tr>
<tr>
<td>Loop 2</td>
<td>17</td>
<td>27</td>
<td>1</td>
<td>33%</td>
</tr>
<tr>
<td>Loop 3</td>
<td>13</td>
<td>31</td>
<td>1</td>
<td>15%</td>
</tr>
<tr>
<td>Loop 4</td>
<td>2</td>
<td>28</td>
<td>15</td>
<td>0%</td>
</tr>
<tr>
<td>Loop 5</td>
<td>0</td>
<td>31</td>
<td>14</td>
<td>0%</td>
</tr>
</tbody>
</table>

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

**TABLE 3 Original Average Accuracy of the Instantaneous Travel Time Algorithms ($\varepsilon$)**

<table>
<thead>
<tr>
<th></th>
<th>Free Flow</th>
<th>Stop &amp; Go</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITT₁</td>
<td>$\varepsilon_{1,L} = 0.327$</td>
<td>$\varepsilon_{1,S} = 0.376$</td>
</tr>
<tr>
<td>ITT₂</td>
<td>$\varepsilon_{2,L} = 0.285$</td>
<td>$\varepsilon_{2,S} = 0.388$</td>
</tr>
</tbody>
</table>
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 $\varepsilon$ 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 $\varepsilon_3$ defined in equation 19 and comparing it to the original algorithms accuracy ($\varepsilon_1$ and $\varepsilon_2$). 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>
<thead>
<tr>
<th>TABLE 4 Data Fusion Technique Average improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Free Flow</td>
</tr>
<tr>
<td>ITT$_1$</td>
</tr>
<tr>
<td>Improvement in relation to original ITT values in each traffic state</td>
</tr>
</tbody>
</table>

(a) ITT fusion 1 improvement

<table>
<thead>
<tr>
<th>Quadratic Mean Relative Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTT$_1$</td>
</tr>
<tr>
<td>PTT$_2$</td>
</tr>
<tr>
<td>PTT</td>
</tr>
</tbody>
</table>

(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