

# Improving Travel Time Estimates from Inductive Loop and Toll Collection Data with Dempster–Shafer Data Fusion

Nour-Eddin El Faouzi, Lawrence A. Klein, and Olivier De Mouzon

**Dempster–Shafer data fusion can enhance travel time estimation for motorists and traffic managers. In this paper, travel time data from inductive loop road sensors and toll collection stations are merged through Dempster–Shafer inference to generate an improved estimate of travel time. The technique captures travel time data from the two sources and combines them by using Dempster’s rule and belief values (also called probability mass) calculated from a confusion matrix. The most probable travel time over the monitored road section is selected as that with the largest belief. A case study is provided to illustrate application of the fusion technique with data gathered on winter Saturdays for 2 years: 2003 data are used to compute the confusion matrices and belief values, and 2004 data are used for validation.**

Availability, accuracy, and reliability of systemwide travel time information contribute to effective decision making in support of safe and efficient operation of traffic management systems. Such information is made available in modern traffic management systems through sensor and data-processing technologies that gather real-time traffic and weather data that are descriptive of roadway conditions. Typical data sources include roadway sensor or transit fleet vehicle driver and operator reports of intersection and highway gridlock, 911 dispatch or traveler cellular telephone reports of traffic accidents, police radio or taxi communication channel reports of local congestion events, and roadway weather sensors.

Only a small amount of reported information is generally used at a traffic management center—partly due to automatic accumulation of multisource and multimedia data and the associated differences in formats, arrival times, and reliability of incident occurrence information and variations in the reported incident locations. Underutilization of available information is exacerbated by a lack of decision-support systems to process imprecise, incomplete, and conflicting information.

This paper describes and reports on application of Dempster–Shafer inference to assist traffic engineers and system managers at traffic operation centers in calculating travel time over a given section of roadway. Dempster–Shafer inference, a statistical-based classification technique, can be used when data sources contributing information

cannot associate a 100% probability of certainty to the events the data represent. The technique captures and combines whatever certainty or knowledge exists in the event classification capability of the information sources. The knowledge contributed by the information sources is combined, by using Dempster’s rule, to find the conjunction of the events and the associated probability or belief that the decision is correct (1–4).

The U.S. Department of Defense (DoD) data fusion model and Dempster–Shafer inference, in particular, as a tool for combining multisource data are discussed in the following sections. Field test data from inductive loop detector (ILD) pairs and toll collection stations are then used to demonstrate application of the technique to travel time analysis and estimation. This process involves computing a confusion matrix that describes the likelihood the sensors or toll collection devices are reporting correct travel times. Information derived from electronic toll collection (ETC) with tags, real-time credit card payments from the credit-card-only lane, and cash payments are used to find the “true” travel time values (used as the reference value) against which travel time estimates gathered from ILD data, toll tag data only, and toll tag plus real-time credit card payment data are compared. Travel times computed from toll tag data alone are an important travel time information source as those data can be supplied in real time to traffic management centers and used to supplement travel times derived from loops.

## DATA FUSION MODEL

Many data-processing techniques originally developed by the DoD to support identification and tracking of military objects, including Dempster–Shafer inference, can be used today to aid traffic management on surface streets and highways (5–7). The DoD data fusion model consists of a hierarchy of five processing levels. Level 0 deals with preprocessing of data from the contributing source. It may normalize, format, order, batch, and compress input data (7, 8). It may even identify subobjects or features in the data that are used later in Level 1 processing.

For highway and arterial traffic management, Level 1 processing concerns data gathering from all appropriate sources, including real-time point and wide-area traffic flow sensors, transit system operators, toll data, cellular telephone calls, emergency call box reports, probe vehicle and roving tow truck messages, commercial vehicle transmissions, and roadway-based weather sensors (3).

Level 2 processing identifies the probable situation causing the observed data and events by combining the results of Level 1 processing with information from other sources and databases. These

---

N.-E. El Faouzi and O. De Mouzon, Laboratoire d'Ingénierie Circulation Transports, INRETS-ENTPE, 25 Avenue François Mitterrand Case 24, 69675 Bron Cedex, France. L. A. Klein, Klein & Associates, 562 Giotto, Irvine, CA 92614. Corresponding author: N.-E. El Faouzi, [elfaouzi@inrets.fr](mailto:elfaouzi@inrets.fr).

*Transportation Research Record: Journal of the Transportation Research Board*, No. 2129, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 73–80.  
DOI: 10.3141/2129-09

sources may include highway patrol reports and databases, roadway configuration drawings, local and national weather reports, anticipated traffic mix, time-of-day traffic patterns, construction schedules, and special event schedules.

Level 3 processing assesses traffic flow patterns and other data with respect to the likely occurrence of a traffic event (e.g., traffic congestion, incident, construction or other preplanned special event, fire, or police action) that affects traffic flow.

Level 4 processing seeks to improve the entire data fusion process by continuously refining predictions and assessments and evaluating the need for additional sources of information. Sometimes a fifth level is added to address issues concerned with enabling a human to interpret and apply the results of the fusion process.

Dempster–Shafer inference is an example of a Level 1 data fusion technique that can be used to support traffic management decision making when the available input data do not allow a definitive event or condition to be identified (9, 10). Once received, these data are combined in a meaningful way to address traffic management needs. The advantages of Dempster–Shafer data fusion lie in its ability to combine whatever information data sources are able to contribute and to find the conjunction of the events with the highest probability of occurrence.

## DEMPSTER–SHAFER INFERENCE

Assume a traffic-related event is of type  $h_1, h_2, \dots, h_p$ , representing the set of all propositions making up the hypothesis space, called the frame of discernment denoted by  $\theta$ . A probability mass  $m(h_i)$  is assigned to any of the original propositions or to a union of propositions based on available information from the data sources. For example, the union or disjunction that the event is of type  $h_1$  or  $h_2$  (denoted  $h_1 \cup h_2$ ) is assigned probability mass  $m(h_1 \cup h_2)$  as derived from information from one of the sources. Probability mass may also be assigned by the information source to the negation of a proposition such as  $m(\bar{h}_1) = m(h_2 \cup h_3 \cup \dots \cup h_p)$ .

When all probability mass cannot be directly assigned by the information source to any of the propositions, their unions, or the negations of propositions, the remaining mass is assigned to the frame of discernment  $\theta$  (representing uncertainty about further definitive assignment) as  $m(\theta) = m(h_1 \cup h_2 \cup \dots \cup h_p)$ . The mass assigned to  $\theta$  represents the uncertainty that remains concerning the ability of the information source to gather and interpret the data. The sum of probability masses over all propositions, uncertainty, and negation equals unity.

Each proposition or event is further characterized by its degree of support and degree of plausibility. Support for a given proposition is defined as the sum of all probability mass assigned directly by the sensor to that proposition (1). Thus, support for an event type as  $h_1, h_2, \text{ or } h_3$ , denoted by  $S(h_1 \cup h_2 \cup h_3)$ , contributed by a sensor or other information source is equal to

$$S(h_1 \cup h_2 \cup h_3) = m(h_1) + m(h_2) + m(h_3) + m(h_1 \cup h_2) + m(h_1 \cup h_3) + m(h_2 \cup h_3) + m(h_1 \cup h_2 \cup h_3) \quad (1)$$

Plausibility of a given proposition is defined as the sum of all mass not assigned to its negation; in other words, plausibility defines the mass free to move to the support of a proposition. The plausibility of  $h_i$ , denoted by  $Pl(h_i)$ , is written as

$$Pl(h_i) = 1 - S(\bar{h}_i) \quad (2)$$

where  $S(\bar{h}_i)$  represents the degree to which the evidence impugns a proposition (i.e., supports negation of the proposition).

An uncertainty interval is defined to capture the information contained in the support and plausibility for a proposition. The uncertainty interval is bounded from below by the support value, which is equal to the minimal commitment for the proposition based on direct sensor evidence. The upper bound for the uncertainty interval is given by the plausibility and represents the support plus any potential commitment. This interpretation of the uncertainty interval shows what proportion of evidence is truly in support of a proposition and what proportion results merely from ignorance or the requirement to normalize the sum of the probability masses to unity. Accordingly, the uncertainty interval is written as  $[S(h_i), Pl(h_i)]$ , where

$$S(h_i) \leq Pl(h_i) \quad (3)$$

Support and probability mass obtained from an information source represent different concepts. Support is calculated from the sum of the probability masses that directly support the proposition and its unions. Probability mass is determined from the ability of the information source to assign some certainty to a proposition based on the evidence.

Dempster's rule supplies the formalism to combine probability masses provided by multiple information sources for compatible propositions. The conjunction of the propositions with the largest probability mass is usually selected as the output of the process. Propositions are compatible when their conjunction exists. If the conjunctions form the empty set, the probability masses associated with the conjunctions are set equal to zero and the probability masses of the nonempty set intersections are increased by a proportionality factor  $K$  so that their sum is unity.

The general form of Dempster's rule for calculating the total probability mass committed to an event  $c$ , known as the orthogonal sum, is given by

$$m(c) = K \sum_{a_i \cap b_j = c} [m_A(a_i) m_B(b_j)] \quad (4)$$

where

$$K^{-1} = 1 - \sum_{a_i \cap b_j = \emptyset} [m_A(a_i) m_B(b_j)] \quad (5)$$

where  $a_i$  and  $b_j$  are the hypotheses or unions of hypotheses assigned probability masses by information sources  $A$  and  $B$ , and  $\emptyset$  is defined as the empty set. If  $K^{-1}$  is 0, then  $m_A$  and  $m_B$  are contradictory and the sum defined by Dempster's rule does not exist.

## ESTIMATION OF MOTORWAY TRAVEL TIME

### Test Site Description

Travel time data from 7 days in 2003 and 5 days in 2004 were available from a 7-km section of the Autoroutes Rhône-Alpes (AREA) motorway in France. This section, illustrated in Figure 1, shows the location of toll stations, ILD pairs in each lane, exit and entry ramps, and rest area.

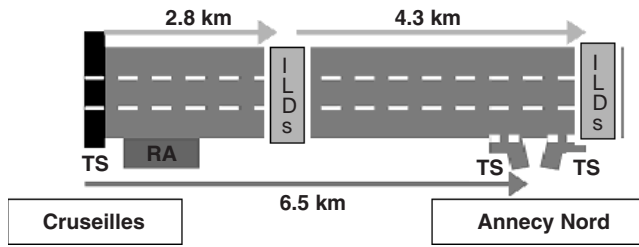


FIGURE 1 Motorway section over which travel time data were collected and analyzed (TS = toll station, RA = rest area).

**Data Sources**

The inductive loop detector pairs give 6-min aggregated volume, occupancy, and speed data. Toll collection data provide entry and exit times at toll gates, identification of entry and exit toll gates, class of vehicle (car, heavy vehicle, truck, motorcycle, bus), and means of payment (e.g., electronic toll tag, real-time credit card payment, or cash).

Implementation of the Dempster–Shafer fusion process requires the available data to be subjected to some preliminary processing to convert them into travel times and travel time intervals, where each travel time interval is assigned a likelihood by using a probability mass value. This procedure is the primary factor in determining the success of the fusion operation.

**Travel Time Derivation from Loop and Toll Collection Data**

Information conveyed by each source is converted to travel time before fusion occurs. ILD data are transformed to travel time data by a speed extrapolation method (11). This simple strategy estimates speed between loop detector stations from upstream and downstream measurements by assuming that measurements made by a detector located at a fixed point are still representative of the traffic state prevailing in its neighborhood  $N$ . The associated speed field is given by

$$\hat{V}_x(t_i) = \begin{cases} V_u(t_i) & \text{if } x \in N_{x_u} \\ V_d(t_i) & \text{if } x \in N_{x_d} \end{cases} \quad (6)$$

where

- $x_u$  and  $x_d$  = upstream and downstream locations, respectively, of the loop detectors,
- $D$  = length of the section between the two detectors, and
- $N_x$  = neighborhood of position  $x$  defined by  $N_x = \{x; |x - x_s| \leq D/2\}$ .

Toll collection data are filtered with a statistical-based filter (12) to remove extremely long and short travel times (outliers or whiskers). These extreme values are due to stops for resting or entering service areas located within the test section shown in Figure 1 and to motorcycles that often travel between lanes and do not experience prevailing congestion. The filtered data are then aggregated to calculate the 6-min travel time intervals. Electronic toll tag data are extracted from toll collection data and used as a supplementary source to enhance ILD-computed travel times.

Once these steps are completed, travel time is separated into four intervals to form the frame of discernment. They are defined according to prevailing traffic conditions

$$h_1 = \{TT(t)$$

so that

$$TT(t) \leq 1.1 \times TT_{ff} \} \quad (7)$$

$$h_2 = \{TT(t)$$

so that

$$1.1 < \frac{TT(t)}{TT_{ff}} \leq 1.3 \} \quad (8)$$

$$h_3 = \{TT(t)$$

so that

$$1.3 < \frac{TT(t)}{TT_{ff}} \leq 1.5 \} \quad (9)$$

$$h_4 = \{TT(t)$$

so that

$$TT(t) > 1.5 \times TT_{ff} \} \quad (10)$$

where  $TT_{ff}$  is free-flow travel time when vehicle speed is set equal to the speed limit of 130 km/h (80 mph).

Figure 2 illustrates the travel time hypotheses with respect to traffic-flow conditions before, during, and after a midday incident. Thus,  $h_1$  represents free flow at the prevailing speed limit,  $h_2$  and  $h_3$  represent two stages of travel time during the onset or conclusion of congestion, and  $h_4$  represents travel time at peak congestion.

The travel times are then further segregated by time of day to determine whether probability mass values are sensitive to time of day. Similar operations are performed for the rest of the travel time data collected on the remaining 6 days of 2003.

**IMPLEMENTATION OF DEMPSTER–SHAFER INFERENCE DATA FUSION**

**Generation of Probability Mass Functions**

As there is no standard method to generate probability mass functions, it was decided to generate and combine them with a statistical training procedure using confusion matrices. Confusion matrices, one for each data source under consideration, were created from the 24-h and time-of-day travel time data as follows. The first confusion matrix compared the “true” travel times computed using all the toll collection data (electronic toll tag + real-time credit card payments + cash) with estimated travel times computed from the inductive loop sensor pairs over a 24-h data collection period. The second confusion matrix replaced the estimated loop travel times with travel times computed using only ETC data. The true travel times were the same as before.

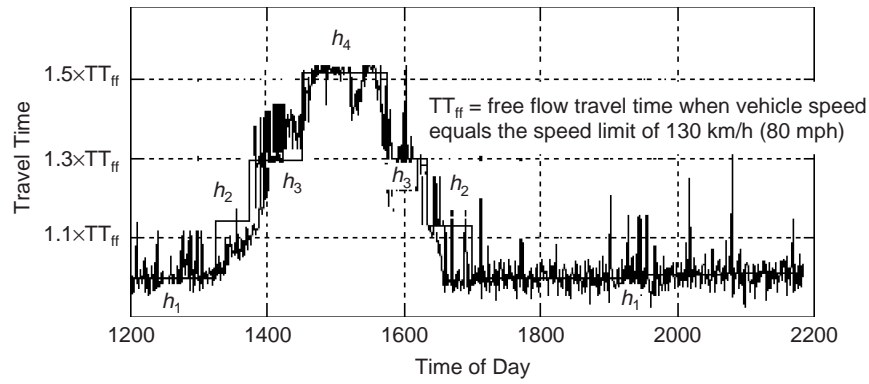


FIGURE 2 Separation of travel time into four hypotheses about traffic conditions.

The entries in the confusion matrix are the numbers of instances  $n$  a travel time hypothesis estimated by a source agrees with the true travel time over the data collection period. Accordingly, the confusion matrix ( $CM_j$ ) for each source  $j$ , with  $j \in \{ILD, ETC\}$  is specified as shown in Equation 11 as a  $p \times p$  table that shows  $n_{ik}^{(j)}$ , where  $p$  designates the number of travel time hypotheses and  $i$  and  $k$  are the row and column indices, respectively.

$$CM_j = \left. \begin{array}{c} \text{columns represent travel time} \\ \text{intervals estimated by a} \\ \text{selected source} \\ \left[ \begin{array}{ccc} n_{11}^{(j)} & \cdots & n_{1p}^{(j)} \\ \vdots & \ddots & \vdots \\ n_{p1}^{(j)} & \cdots & n_{pp}^{(j)} \end{array} \right] \\ \text{rows represent true} \\ \text{travel time intervals} \end{array} \right\} \quad (11)$$

The  $CM_j$  displays the similarity between the travel time hypothesis decision vector estimated by each source and the vector representing the true hypothesis. The diagonal elements are the number of correctly classified travel time intervals from each data source, while the off-diagonal elements are the number of misclassified travel time intervals. Thus,  $n_{ii}^{(j)}$  is the number of instances the travel time interval  $h_i$  estimated by data source  $j \in \{ILD, ETC\}$  matches the true travel time interval  $h_i$  derived from all toll collection data (electronic toll tag + real-time credit card payments + cash), and  $n_{ik}^{(j)}$ ,  $i \neq k$ , is the number of instances data source  $j \in \{ILD, ETC\}$  estimated travel time interval  $h_i$  when the true one was  $h_k$ . The matrix is updated each time a travel time estimate is processed during the data collection period.

As an example of how the matrix is populated, consider the four-hypothesis problem. At the first 6-min time step, the travel time interval estimated by the inductive loops is  $h_2$  and the true travel time is also  $h_2$ . Thus, the confusion matrix appears as

$$CM_{ILD} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad (12)$$

after the first time step data are incorporated. Inductive loop data from the second time step estimate the travel time interval as  $h_2$ , while the true travel time is  $h_3$ . Accordingly, the matrix appears as

$$CM_{ILD} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad (13)$$

after the second time step. If the third sample contains the same information as the first, then the value of  $n_{22}^{ILD}$  is updated to 2. Column two is continually updated and the other columns are populated with inductive loop travel time estimates as data collection proceeds over various traffic flow conditions that occur during the 24-h period.

The probability mass functions are found by normalizing the confusion matrix of Equation 11 by either of two strategies. For simplicity of notation, the  $j$  superscript that appeared with  $n$  is dropped.

Strategy 1. The frame of discernment  $\theta$  is included as a potential travel time decision to model ignorance about travel time on the part of the data source. Normalization of the confusion matrix occurs by dividing each matrix element by the total of all the matrix elements. Probability masses  $m$  are assigned to each travel time hypothesis as follows: if source  $j$  gives  $h_k$  as an output, then select the  $k$ th column of confusion matrix  $CM_j$ —for example,  $\{\tilde{n}_{1k}, \dots, \tilde{n}_{pk}\}$ , where  $\tilde{n}_{ik} = n_{ik} / \sum_i n_{ik}$  and  $p =$  number of travel time hypotheses—and define

$$\left. \begin{array}{l} m^{(j)}(h_i) = \tilde{n}_{ik} \\ m^{(j)}(\theta) = 1 - \sum_i \tilde{n}_{ik} \end{array} \right\} \quad (14)$$

Strategy 2. Here one is always able to select one of the travel time hypotheses as the output of the data source. Normalization is performed by column (i.e., in each column, the entries are divided by the total of the column entries) so that each column vector representing probability mass values sums to unity. Probability masses  $m$  are assigned to each travel time hypothesis as follows: if source  $j$  gives  $h_k$  as an output, then select the  $k$ th column of confusion matrix  $CM_j$ —for example,  $\{\tilde{n}_{1k}, \dots, \tilde{n}_{pk}\}$ , where  $\tilde{n}_{ik} = n_{ik} / \sum_i n_{ik}$ —and define

$$\left. \begin{array}{l} m^{(j)}(h_i) = \tilde{n}_{ik} \\ m^{(j)}(\theta) = 0 \end{array} \right\} \quad (15)$$

To investigate the sensitivity of probability mass values to time of day, other confusion matrices were derived by the same procedures

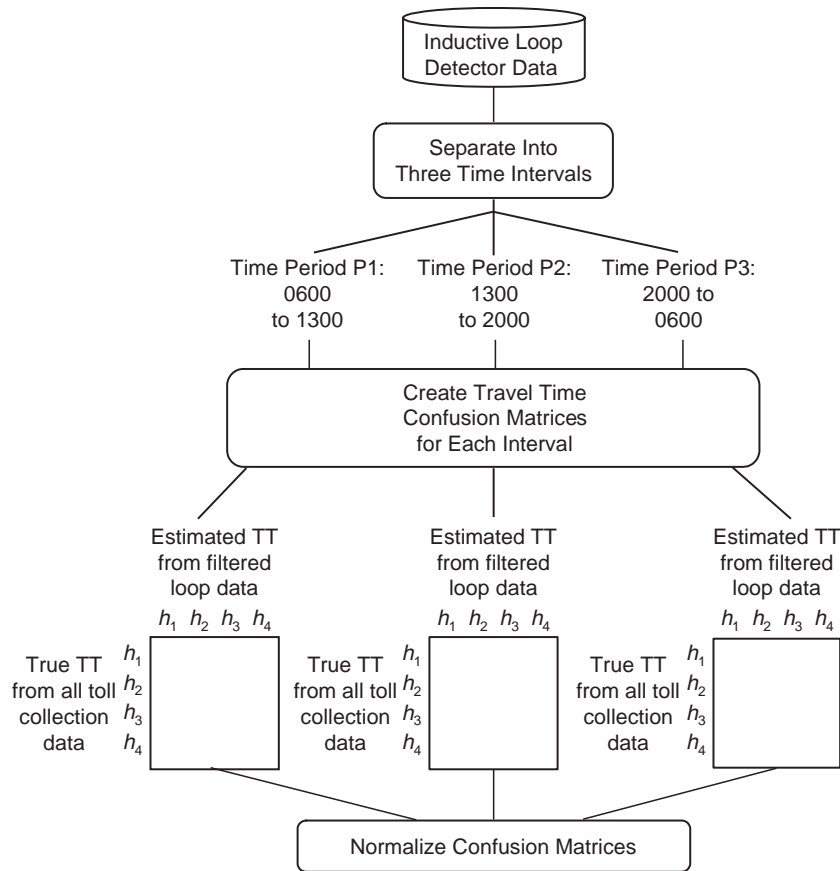


FIGURE 3 Confusion matrix derivation using travel times from ILD data.

but with travel time data segregated by time of day for the morning (time period P1: 0600 to 1300), afternoon (time period P2: 1300 to 2000), and night (time period P3: 2000 to 0600). The time-of-day formation of the confusion matrices corresponding to travel time hypotheses is shown in Figure 3 for ILD data. The same method applies to ETC data. The process of using probability masses derived from time-of-day considerations is referred to as dynamic fusion, while the process that uses probability masses derived from the entire ensemble of 24-h data is referred to as static fusion.

### Implementation of Dempster’s Rule

The application of Dempster’s rule is demonstrated by using the probability mass values found from the confusion matrices representing travel times calculated from ILD data and ETC + real-time credit card payment data. The corresponding probability masses are shown in Tables 1 and 2 for application of Strategy 2. The hypotheses are defined in Equations 7 through 10.

Dempster’s rule is implemented by forming a matrix with the probability masses that are to be combined entered along the first column and last row as shown in Table 3. Inner matrix (row, column) elements are computed as the product of the probability mass in the same row of the first column and the same column of the last row. The proposition corresponding to a matrix element is equal to the conjunction of the propositions that are multiplied.

For example, consider the fusion of the column probability mass vectors for  $h_2$ . The  $h_2$  column vector from Table 1 is entered

along the first column and the  $h_2$  column vector from Table 2 is entered in the last row of Table 3. Superscripts 1 and 2 correspond to the ILD and ETC + real-time credit card payments source, respectively.

In this example, matrix element (1, 2) represents the proposition formed by the conjunction of  $m^{(1)}(h_1)$  from ILD data and  $m^{(2)}(h_1)$  from ETC data. The un-normalized probability mass  $m(h_1)$  associated with the intersection of the propositions (i.e., travel time less than  $1.1 \times$  free-flow travel time) is

$$m(h_1) = m^{(1)}(h_1) \times m^{(2)}(h_1) = (0.20)(0.03) = 0.0060 \quad (16)$$

The off-diagonal elements in Table 3 are members of the empty set  $\phi$ . Therefore, the mass assigned to  $\phi$  must be redistributed to

TABLE 1 Probability Masses for Travel Time Hypotheses from ILDs and True Values from All Toll Collection Data over 24 h

	$h_1$	$h_2$	$h_3$	$h_4$
$h_1$	1	0.20	0.00	0.00
$h_2$	0.00	0.61	0.08	0.00
$h_3$	0.00	0.16	0.69	0.05
$h_4$	0.00	0.03	0.23	0.95

**TABLE 2 Probability Masses for Travel Time Hypotheses from ETC and True Values from All Toll Collection Data over 24 h**

	$h_1$	$h_2$	$h_3$	$h_4$
$h_1$	0.36	0.03	0.00	0.00
$h_2$	0.60	0.35	0.01	0.00
$h_3$	0.04	0.51	0.35	0.28
$h_4$	0.00	0.11	0.64	0.72

the nonempty set elements by using the value  $K$  from Equation 5, where

$$K^{-1} = 1 - (0.0183 + 0.0048 + 0.0009 + 0.0700 + 0.0560 + 0.0105 + 0.1020 + 0.3111 + 0.0153 + 0.0220 + 0.0671 + 0.0176) = 0.3044 \tag{17}$$

and

$$K = \frac{1}{0.3044} = 3.285 \tag{18}$$

As shown in Table 4, the probability masses corresponding to the null set elements are set to 0 and the probability masses of the nonempty set elements are multiplied by  $K$  so that their sum is unity. This procedure results in hypothesis  $h_2$  having the largest probability mass, equal to

$$m(h_2) = 0.70 \tag{19}$$

Therefore, it is selected as the most probable travel time hypothesis and the output of the Dempster–Shafer fusion process (i.e., travel times are between  $1.1 \times TT_{ff}$  and  $1.3 \times TT_{ff}$ ).

When three or more information sources contribute data, the application of Dempster’s rule is repeated recursively by using the results of the first application of the rule as the probability mass elements of the first column and the probability mass values from the third information source as the elements of the last row (or vice versa). The final output is obtained after the probability masses from all the information sources have been incorporated.

**RESULTS AND DISCUSSION**

Dempster–Shafer inference was applied to the 2003 data to compute probability mass values. The same values were used with the 2004 data for validation purposes. The combined travel times from ILD estimates and ETC estimates were calculated for each data set.

Outputs of the fusion process, using the two strategies defined by Equations 14 and 15, were compared with the actual travel time intervals derived from all toll collection data. Performance of the applied fusion strategies in terms of correct classification percentage is presented in Table 5. Two scenarios that reflect different automated toll payment market penetration rates are listed for each strategy. Scenario A uses only ETC data collected during the selected study days to compute travel time, whereas Scenario B uses both ETC data and real-time credit card payment transactions to compute travel time.

Table 5 shows that the ILD source almost always outperforms the ETC travel time data. The percentage of correct classification obtained from fusion is always better than that of individual sources. Furthermore, Strategy 2 exhibits more improvement than Strategy 1.

The sensitivity of fusion performance to a varying ETC market penetration rate was explored for each of the three time-of-day periods using the market penetration data in Figure 4.

Examination of the evolution of the correct classification percentage by ETC market penetration rates in Table 5 shows that an increase in the ETC market penetration (from Scenario A to Scenario B) results in better performance for the individual source and the fused data sets. This situation occurs because the true or reference travel time is computed from all toll data. Therefore, as more electronic toll data are used (here through the addition of the real-time credit card transactions to

**TABLE 3 Application of Dempster’s Rule for Combining Probability Masses for Travel Time Hypothesis  $h_2$  from ILD and ETC Data**

	$m^{(2)}(h_1) = 0.03$	$m^{(2)}(h_2) = 0.35$	$m^{(2)}(h_3) = 0.51$	$m^{(2)}(h_4) = 0.11$
$m^{(1)}(h_1) = 0.20$	$m(h_1) = 0.0060$	$m(\phi) = 0.0700$	$m(\phi) = 0.1020$	$m(\phi) = 0.0220$
$m^{(1)}(h_2) = 0.61$	$m(\phi) = 0.0183$	$m(h_2) = 0.2135$	$m(\phi) = 0.3111$	$m(\phi) = 0.0671$
$m^{(1)}(h_3) = 0.16$	$m(\phi) = 0.0048$	$m(\phi) = 0.0560$	$m(h_3) = 0.0816$	$m(\phi) = 0.0176$
$m^{(1)}(h_4) = 0.03$	$m(\phi) = 0.0009$	$m(\phi) = 0.0105$	$m(\phi) = 0.0153$	$m(h_4) = 0.0033$

**TABLE 4 Normalized Probability Masses for Travel Time Hypotheses**

	$m^{(2)}(h_1) = 0.03$	$m^{(2)}(h_2) = 0.35$	$m^{(2)}(h_3) = 0.51$	$m^{(2)}(h_4) = 0.11$
$m^{(1)}(h_1) = 0.20$	$m(h_1) = 0.02$	$m(\phi) = 0$	$m(\phi) = 0$	$m(\phi) = 0$
$m^{(1)}(h_2) = 0.61$	$m(\phi) = 0$	$m(h_2) = 0.70$	$m(\phi) = 0$	$m(\phi) = 0$
$m^{(1)}(h_3) = 0.16$	$m(\phi) = 0$	$m(\phi) = 0$	$m(h_3) = 0.27$	$m(\phi) = 0$
$m^{(1)}(h_4) = 0.03$	$m(\phi) = 0$	$m(\phi) = 0$	$m(\phi) = 0$	$m(h_4) = 0.01$

TABLE 5 Individual Sources and Fusion Schemes in Terms of Correct Classification (%), 2003 Data

Time Period	Type of Fusion	ILD Data Only	ETC Data Only in SA (% of total tolls)	ETC Data Only in SB (% of total tolls)	Fusion of ILD + ETC Data (SA) Using Strategy 1	Fusion of ILD + ETC + CC Data (SB) Using Strategy 1	Fusion of ILD + ETC Data (SA) Using Strategy 2	Fusion of ILD + ETC + CC Data (SB) Using Strategy 2
P1	Static	26	3 (32)	17 (42)	29	31	29	32
	Dynamic				29	32	29	34
P2	Static	29	12 (19)	23 (28)	32	34	37	40
	Dynamic				32	34	34	40
P3	Static	31	1 (34)	10 (46)	31	33	31	35
	Dynamic				32	37	32	37

NOTE: SA = Scenario A, SB = Scenario B, CC = credit card.

the ETC data), the more the electronic toll data set resembles the reference value data set. Strategy 1 is less sensitive to variation in ETC rates than Strategy 2, where the percentage of correct classification is strongly dependent on the increase in ETC market penetration. However, Strategy 1 performs worse overall than Strategy 2.

Dynamic fusion, which incorporates the temporal dimension of the data sources, is generally as good as or better than static fusion (in 11 of 12 cases). According to these results, the best strategy appears to be dynamic fusion using Strategy 2. Thus, Strategy 2 is selected for fusing travel times from the 2004 ILD and ETC data. The probability mass values were kept the same as those computed from the 2003 data and applied to the 2004 travel time data. Results of this validation process are shown in Table 6.

Analysis of the 2004 data reveals two principal findings:

- Contrary to 2003, the ETC travel time data source outperforms the ILD data source except for Scenario A in time period P3. The percent of correct classification increases with an increase in market penetration of ETC data when changing from Scenario A to Scenario B. This trend is more pronounced for individual data sources.
- Dynamic fusion is generally as good as or better than static fusion (in five of six cases). However, fusion does not always outperform

individual sources. This situation is especially apparent in Strategy 1 (not shown), which almost never improves in overall performance (in one of six cases). The results with Strategy 2 and Scenario B are better only for dynamic fusion and only in P1 and P3 (i.e., two of six cases).

This disappointing result for the fusion process can be explained by the structural change in the motorway ETC deployment policy that resulted in increased market penetration of ETC tags. Both the Scenario A and Scenario B ETC market penetration percentages are similar in 2003 and 2004, except for P2 in 2004 (see 2004 data in Figure 4 that show almost constant market penetration for Scenario A and Scenario B during P2), which explains why Strategy 2–Scenario B fusion performs poorly in P2 of 2004.

CONCLUSIONS

Application of Dempster–Shafer inference to estimate travel times by combining data from two sources was presented. This technique is conceptually clear, easy to apply, and effective for processing information from different data sources. The output of the process can offer

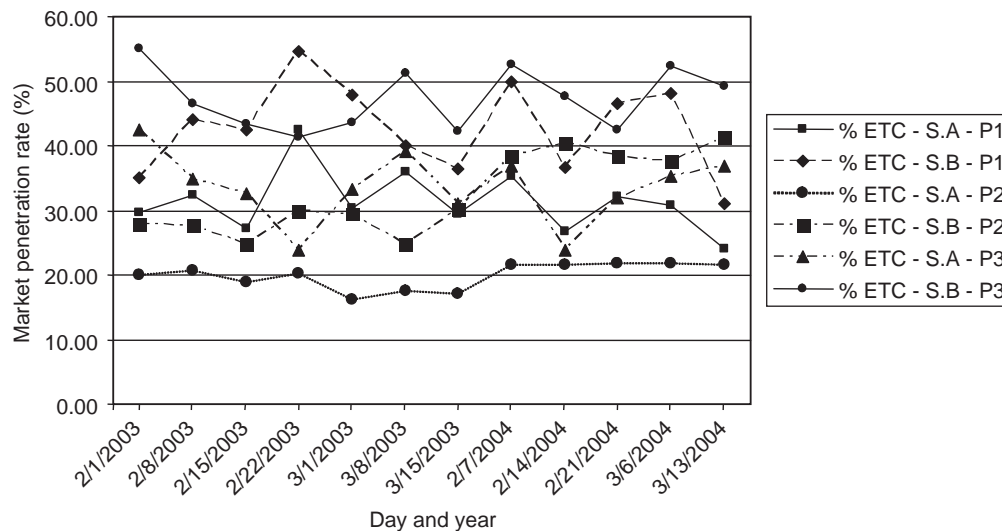


FIGURE 4 ETC market penetration rate as a function of time of day, scenario, day, and year (S.A = Scenario A, S.B = Scenario B).

TABLE 6 Individual Sources and Fusion Schemes in Terms of Correct Classification (%), 2004 Data

Time Period	Type of Fusion	ILD Data Only	ETC Data Only in SA (% of total tolls)	ETC + CC Data in SB (% of total tolls)	Fusion of ILD + ETC Data (SA) Using Strategy 2	Fusion of ILD + ETC + CC Data (SB) Using Strategy 2
P1	Static	4	12 (29)	50 (42)	15	04
	Dynamic				04	58
P2	Static	9	19 (22)	48 (39)	21	13
	Dynamic				21	13
P3	Static	11	5 (33)	18 (48)	11	11
	Dynamic				11	27

timely updates to assist in traffic management decision making as it is designed to extract whatever information exists in an otherwise incomplete database derived from a variety of data sources. The benefits gained from implementing the system on the AREA motorway in the Rhône-Alpes region of France are highly transferable to other locations. Dempster–Shafer inference supports fast identification of travel time changes induced by congestion and thus can be used to support the timely display of traffic advisory messages and removal of incidents that improve the quality of transportation system services to motorists and transportation system managers.

Implementation of the Dempster–Shafer data fusion technique indicates that the most critical item influencing its effectiveness is accurate probability mass values for each data source that contributes information. One technique for generating the probability mass values was described, although others exist (11). Further evaluation of the technique is important for its validation before it can be applied operationally. For example, the data fusion method should be assessed in other regions, over longer sections of highway, and with other data sources such as mobile telephone data. Issues such as the effect of the fusion technique on the time interval over which the travel time estimate is valid also need to be explored.

## REFERENCES

1. Shafer, G. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, N.J., 1976.
2. Waltz, E., and J. Llinas. *Multisensor Data Fusion*. Artech House, Norwood, Mass., 1990.
3. Klein, L. A. *Sensor Technologies and Data Requirements for ITS*. Artech House, Norwood, Mass., 2001.
4. Klein, L. A. *Sensor and Data Fusion: A Tool for Information Assessment and Decision Making*. Monograph 138. SPIE Press, Bellingham, Wash., 2007.
5. White, F. E., Jr. Joint Directors of Laboratories Data Fusion Subpanel Report: SIGINT session. *Technical Proceedings of Joint Service Data Fusion Symposium*, Vol. I, DFS-90, 1990, pp. 469–484.
6. Data Fusion Development Strategy Panel. *Functional Description of the Data Fusion Process*. Office of Naval Technology, Washington, D.C., 1991.
7. Steinberg, A. N., C. L. Bowman, and F. E. White. Revisions to the JDL Data Fusion Model. *Proceedings of SPIE*, Vol. 3719, 1999, pp. 430–441.
8. Hall, D. L. *Mathematical Techniques in Multisensor Data Fusion*. Artech House, Norwood, Mass., 1992.
9. Byun, S. C., D. B. Choi, B. H. Ahn, and H. Ko. Traffic Incident Detection Using Evidential Reasoning Based Data Fusion. *Proc., 6th World Congress on Intelligent Transportation Systems*, Toronto, Canada, Nov. 1999.
10. Klein, L. A., P. Yi, and H. Teng. Decision Support System for Advanced Traffic Management Through Data Fusion. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1804, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 173–178.
11. El Faouzi, N.-E., A. B. Aissa, O. De Mouzon, and J. Sau. *Highway Travel Time Estimation and Prediction from Toll Collection System*. Transport and Traffic Engineering Laboratory, INRETS, Bron, France, 2006.
12. De Mouzon, O., and N.-E. El Faouzi. *Innovative Processing of Toll Collection Data: Filtering, Enrichment and Real-Time Fusion with Loop Detectors for Travel Time Estimation*. Technical Report 0604. Institut National de Recherche sur les Transports et Leurs Sécurité, Bron, France, 2006.

The Intelligent Transportation Systems Committee sponsored publication of this paper.