Performance Evaluation of Travel-Time Estimation Methods for Real-Time Traffic Applications

Xuegang Jeff Ban, Yuwei Li, Alexander Skabardonis, and J. D. Margulici

Abstract

This paper presents a method to compare the performance of various travel-time estimation methods for real-time traffic applications. The methods are evaluated based on their accuracy, reliability, and computational efficiency. The results show that the proposed method outperforms the current methods in terms of accuracy and reliability.

Keywords

Travel-time estimation, real-time traffic, comparison, accuracy, reliability, computational efficiency.
Performance Evaluation of Travel-Time Estimation Methods for Real-Time Traffic Applications

XUEGANG JEFF BAN,1 YUWEI LI,2 ALEXANDER SKABARDONIS,2 and J. D. MARGULICI3

1Department of Civil and Environmental Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA
2Partners for Advanced Transit and Highways (PATH), Institute of Transportation Studies (ITS), University of California—Berkeley, Richmond, CA, USA
3California Center for Innovative Transportation (CCIT), Institute of Transportation Studies (ITS), University of California—Berkeley, Berkeley, CA, USA

Travel time for an itinerary constitutes one of the most relevant roadway traffic metrics. Numerous studies have been conducted to estimate travel times on the basis of data from loop detectors. This article focuses on evaluating the performance of a set of estimation methods—algorithms that estimate route travel times using specific speed data from dual loop detectors for real-time applications, such as displaying travel times on changeable message signs. The travel-time estimates are evaluated against probe vehicle data obtained from FasTrak in the San Francisco Bay Area to determine the accuracy of the estimates. The study reveals that (a) the dispersion of travel time is often larger in off-peak periods than in peak periods for heavily congested freeways; (b) assessed against the same set of speed data, the evaluated algorithms have similar performance; (c) the accuracy of travel-time estimates on the basis of loop-detector data is better in off-peak periods than in peak periods; (d) lane-by-lane speeds from loop detectors may be used to improve travel time estimation; and (e) travel time estimation quality is more sensitive to actual sensor locations if sensors are sparse.

Keywords median absolute deviation; probe vehicle; sensor location; sensor spacing; travel-time estimation; travel-time evaluation; travel-time representation

INTRODUCTION

The travel time for an itinerary constitutes one of the most relevant roadway traffic metrics. This is because (a) travel time is a crucial measure of traffic conditions and system performance; (b) travel time represents information that is easy for the driving public to understand and process; and (c) travel-time information can enable travelers to make better choices about their itineraries, departure times, or even transportation modes, resulting in a form of system self-management. Numerous studies revealed that commuters appreciate and value travel-time information, which reduces their stress and feeling of uncertainty (Khattak et al., 1994; Lindveld et al., 2000; Peng et al., 2004). In the past, many researchers have contributed to developing algorithms for travel-time estimation (e.g., see Coifman, 2002; Hartley, 2003; Hinsbergen et al., 2007; Huisken and Maarseveen, 2000; Oda, 1990; Rice and Zwet, 2001; Smith and Demetsky, 1997).

Measuring the quality of travel-time estimates is important because it helps understand the performance of travel-time estimation and point to needed improvements in traffic data collection. Although many studies provided certain discussions on performance of the proposed methods, the discussions were limited because (a) ground-truth travel times (e.g., those from probe vehicles) were not widely available because of technical or resource limitations; (b) most studies used average speeds across multiple lanes without looking at lane-by-lane travel-time variations; and (c) loop-detector locations were assumed...
to be given, and data from all detectors were used to estimate travel times. Note that in this article, we define *lane-by-lane* travel times as the travel times for individual lanes of a freeway computed through lane-specific speeds. They can be considered as the travel time of an imaginary vehicle traveling on a specific lane.

Studies devoted specifically to evaluating the performance of travel-time estimation algorithms were rare. Lindveld et al. (2000) are some of the few researchers who focused on comprehensive performance evaluation of several travel-time estimation methods using loop-detector data. They found that for up to moderate congestion, the evaluated travel-time estimation methods could produce reasonably accurate results (with a 10–15% estimation error); however, for heavy congestion, the results may degrade significantly. The ground-truth travel times in Lindveld et al. were collected through license plate readers, floating car runs, and toll-ticket collection. However, the number of observed data points using floating cars was not sufficient; travel times from toll-ticket collection had problems as well (Lindveld et al., pp. 46). In addition, lane-by-lane travel times were not computed and studied. Zhang et al. (1999) studied travel-time estimation methods on the basis of data from single-loop detectors. Floating car runs were conducted to gather the ground-truth travel times. As Kwon et al. (2006) pointed out, however, travel times estimated from limited floating car runs may be biased. Kwon et al. and Fujito et al. (2006) studied the relation between detector spacing and travel-time estimation quality. However, they used travel times computed from the baseline detector spacing as the ground-truth travel times, which may be very different from the actually experienced travel times by individual drivers.

We note that evaluating travel-time estimation quality is essentially different from studying the variability of travel times (e.g., Chen et al., 2003; Noland and Polak, 2002; Van Lint et al., 2004). The former focuses on the differences between estimated and actual travel times, whereas the latter focuses on the features (e.g., the statistical distribution) of the actual travel times. For similar reasons, our study also differs from the travel-time reliability studies that have gained much attention recently in the transportation-research communities (e.g., Al-Deek and Eman, 2006; Chen et al., 2003; Chen et al., 1999; Liu et al., 2007; Schmöcker and Lo, 2009).

In this article, we evaluate a set of estimation methods (i.e., algorithms that give point estimates of route travel times using specific speed data from dual-loop detectors). The three evaluated algorithms in this article are popularly used in real-time traffic applications such as displaying travel times on changeable message signs: the instantaneous, dynamic, and linear-regression travel times (Rice and Zwet, 2001). Both the average and lane-by-lane speeds from double-loop detectors are used to compute the average and lane-specific travel times. Travel-time estimates are compared against probe vehicle travel times, obtained from FasTrak in the San Francisco Bay Area. Because of the large amount of data samples, FasTrak travel times are expected to provide a more accurate representation of the ground-truth travel times than can limited floating car runs, although the data contain outliers that need to be filtered.

We start with developing methods to remove outliers in FasTrak data. For this purpose, we develop an improved scheme for the median absolute deviation (MAD) method (Barnett and Lewis, 1984; Hoaglin et al., 1983). We call this scheme the *local MAD method* (LMM) because it uses variable window length to remove outliers. The LMM captures properly the traffic characteristics under different traffic conditions and is thus appropriate for processing probe-vehicle data.

Two quality measures are defined in this article for evaluating the performance of the selected travel-time estimation methods. One is based on the absolute differences between the loop-based travel-time estimates and the estimated median travel times via FasTrak. The second one is based on whether the loop-based travel-time estimates fall within certain range of the travel times estimated from the dispersion of FasTrak data.

Data for a particular route in the San Francisco Bay Area are subsequently used to conduct the performance evaluation. The evaluation is repeated for different traffic conditions associated with different times of the day and various detector-spacing scenarios. The results show that (a) the accuracy of travel-time estimates on the basis of loop-detector data is better in off-peak periods than in peak periods; (b) lane-by-lane speed data from loop detectors may be used to improve travel-time estimation; and (c) the quality of travel-time estimates is more sensitive to actual sensor locations if sensors are sparse.

The extent to which the findings in this article can be applied to other sites may vary. We thus provide discussions at the end of the article on how these findings can be further verified and applied to improve travel-time estimation.

**Ground-Truth Travel Times from Probe Vehicles**

The travel time $T_r(t)$ for an arbitrary vehicle traveling from a given freeway location (termed *origin*) to a certain destination via route $r$ and starting at time $t$ can be treated as a random variable. The population of concern is the set of different people driving different cars who potentially may take route $r$ at time $t$. In Noland and Polak (2002), this randomness is termed *vehicle-to-vehicle travel-time variation*. Note that in certain cases, it may make sense to impose further restrictions such as "not using the carpool lane," "with FasTrak onboard," or "not stopping over at a point along the route." A sufficiently large number of drivers who take route $r$ at time $t$ is desirable because they constitute a sample of $T_r(t)$. However, if steps of time and distance are arbitrarily fine, at any time $t$ there is only at most one vehicle passing through the exact starting location. Thus, strictly speaking, we can have at most one sample for $T_r(t)$ in reality. To obtain estimates of not only the median but dispersion of the travel-time variable, we can define $T_r(t)$ as the (random) travel time for a vehicle to travel route $r$, starting from the origin at some time $t' \in [t - \Delta/2, t + \Delta/2]$. This is feasible when we are only interested in (or limited to) discrete points on the time scale. Under this alternative definition, we
may have multiple observations for $T_r(t)$. However, now the difference of the observed travel times may be due not only to driver/vehicle differences but also to traffic condition change from time $t - \Delta/2$ to time $t + \Delta/2$ if $\Delta$ is large.

Real observations of $T_r(t)$ (e.g., those obtained from probe vehicles) are called *ground-truth travel times*. The ground-truth travel times in this article are obtained from FasTrak in the San Francisco Bay Area. FasTrak is used statewide in California to automatically collect road and bridge tolls. FasTrak readers are currently installed at each toll booth and at every 5 to 10 miles on selected highways. It has a large market penetration in the Bay Area: nearly 50% of drivers used FasTrak to pay their bridge tolls in 2007 (Golden Gate District, 2007).

The ground-truth travel times obtained from FasTrak can be interpreted as observations of $T_r(t)$ under the alternative definition (i.e., for the period of $[t - \Delta/2, t + \Delta/2]$), with $\Delta$ being reasonably small (e.g., 5–15 min). In this case, $T_r(t)$ may contain multiple observations and the differences between observed values are largely due to driver/vehicle differences. In practice, $\Delta$ cannot be too small—otherwise there will be too few observations in $[t - \Delta/2, t + \Delta/2]$ to characterize the random variable $T_r(t)$. Thus, choosing an appropriate $\Delta$ is important in using FasTrak data to estimate the median and the dispersion of travel times. This issue will be further discussed later in this section. We first propose ways to remove outliers from raw FasTrak data.

### A local MAD Method for Probe-Vehicle Data Processing

Figure 1 depicts the raw FasTrak travel times for a 15-mile section of the EB I-880 freeway (for detailed descriptions of the route, see the “Study Site” section later in this article). The time-dependent pattern of the travel time is obvious in Figure 1, but the raw data contain a significant amount of outliers. Outliers include those vehicles that took excessively long time to travel the route, possibly because they left and reentered the freeway at some intermediate points. Outliers may also come from detection errors that result in those negative travel times in Figure 1. Vehicles that used the highway occupancy vehicle (HOV) lane during afternoon peak hours are also treated as outliers because their travel times are much lower than those using general-purpose lanes.

To remove outliers, we apply the MAD method. MAD is a statistical measure to capture the variation of a given set of data points (Barnett and Lewis, 1984; Hoaglin et al., 1983). Assume that $\{x_i, i = 1, \ldots, K\}$ is the set of data points that are samples of a random variable. Then, MAD can be defined as follows:

$$\text{MAD} = \text{median}(|x_i - \bar{x}|).$$

(1)

Here, $\bar{x}$ is the median value of $\{x_i, i = 1, \ldots, K\}$. To detect whether $x_i$ is an outlier, a $z$ score needs to be computed for each data point:

$$z_i = 0.6745 \cdot \frac{|x_i - \bar{x}|}{MAD}.$$  

(2)

Then, if $z_i \geq \bar{z}$ for a given threshold, $\bar{z}$, $x_i$ can be regarded as an outlier. A $z$ score 3 or 3.5 is often used in practice such as to detect and remove outliers for normally distributed samples (Fallon and Spada, 1997). Here, we use $\bar{z} = 3.0$, which is commonly used in practice.

### LMM Algorithm

Because the raw FasTrak data in Figure 1 have a clear time-dependent pattern, we apply the MAD method locally to data points in a time window (band) with a proper bandwidth. We call this method the LMM. Choosing the bandwidth (i.e., $\Delta$ as we discussed previously), however, is not trivial. First, to
adequately capture the time-dependent trend, it is desirable to use a small bandwidth, but this may result in only a few data points within a band for which the MAD method may not be properly applied (also the basic statistics such as the mean and standard deviation of the travel-time variable cannot be properly estimated either). Alternatively, a large bandwidth can certainly result in sufficient data points, but the time-dependent pattern may be smoothed out.

To address the aforementioned two issues, we adopted a method of variable bandwidth in this article. We first set the default bandwidth as \( h_0 \) min, which is expansible so that each band contains at least \( M \) data points. This is to make the bandwidth as small as possible, while still keeping the analysis statistically meaningful. Such a bandwidth will be constructed for each of the data points in the set \( \{ x_i, i = 1, \ldots, N \} \), as illustrated in the subsequent algorithm.

Step 1. Initialization. Set \( i = 1 \), \( h_0 = 10 \) min, \( M = 25 \), \( \Delta h = 2 \) min, and \( \bar{z} = 3.0 \).

Step 2. Major iteration.
Step 2.1. Determine the bandwidth. Assume the timestamp of the current data point \( x_i \) is \( t_i \).
Set \( h = h_0 \). Assume the number of data points in the time window \([ t_i - h/2, t_i + h/2 ]\) is \( m \). If \( m \geq M \), go to Step 2.2; otherwise, set \( h = h + \Delta h \), update the bandwidth, recalculate \( m \), and check again.

Step 2.2. LMM method. Compute the \( z \) statistics using Equations (1) and (2) for the current data point \( x_i \) using all points in the band \([ t_i - h/2, t_i + h/2 ]\). If \( z \geq \bar{z} \), record \( x_i \) as an outlier.

Step 3. If \( i = N \), go to Step 4; otherwise, set \( i = i + 1 \) and go to Step 2.

Step 4. Remove all recorded outliers from the set \( \{ x_i, i = 1, \ldots, N \} \).

The aforementioned algorithm shows that we initially set the bandwidth as 10 min, which can be expanded if needed symmetrically to both sides of the band (using 1 min as the increment) until the band contains at least 25 data points. Figure 2 illustrates how the bandwidths will change over time if we apply this method to the FasTrak data for a particular day. We considered 11% of the data points to be outliers and thus removed them for this particular data set. Figure 2 clearly shows that to have enough data samples, the bandwidth should be much larger in off-peak hours than in peak hours. For example, nearly 60% of the bandwidths in afternoon peak hours (3:00 p.m. to 7:00 p.m.) are less than or equal to 15 min and the minimum bandwidth is 10 min, whereas for other periods of the day, the minimum bandwidth is much larger. This is consistent with the trip characteristics of commuter traffic: there are usually more travelers, especially those using FasTrak toll tags, in peak hours (the afternoon peak in this case) than in off-peak hours. As a result, there are more FasTrak data samples per unit time during peak hours than off-peak hours. As discussed later in the Characterization of Travel Times section, the bandwidth generated this way is suitable for processing FasTrak data.

Figure 3 shows the processed FasTrak data, which are the ground-truth travel times. It illustrates that it is inadequate to characterize the travel time at a time instant as a single value as many previous studies have done. It is also important to differentiate the travel-time variations at any given time in Figure 3 with those studied in travel-time reliability. In Figure 3, travel-time variations largely come from vehicle-to-vehicle variations (i.e., even for the same time instant, travel times from different vehicles are different because of varied driving behaviors; see Noland and Polak, 2002). In travel-time reliability studies, however, travel-time variations across different days (i.e., day to day) are considered; for a particular day, travel times are treated as single valued. Hence, for a given time instant, travel-time
variations in travel-time reliability studies mainly reflect traffic condition changes across different days; in Figure 3, however, the variations are mainly due to varied driving behaviors (i.e., aggressiveness) of individual drivers rather than changes of traffic conditions. We also note in Figure 3 that the dispersion of travel times is much smaller during afternoon peak than other periods of the day for the study route. We will show in the next section why this is the case.

Characterization of Travel Times

The travel-time dispersion in Figure 3 imposes difficulty on how to properly characterize travel times. One may use a single representative value such as the average or the median without capturing the dispersion. To capture the dispersion, however, we propose to use the interval of the 15th- and 85th-percentile travel times. Figure 4 illustrates the estimated 15th-, the median (50th-), and the 85th-percentile travel times on the basis of the processed FasTrak data in Figure 3. The use of the 15th and the 85th percentiles is somewhat arbitrary. However, this gives us an interval that encloses the travel times of the middle 70% of drivers (in terms of driving aggressiveness). The percentile travel times in Figure 4 were obtained through the method of local linear fit, originally proposed by Koenker and Bassett (1978). This method was later applied by Small et al. (2005) and Liu et al. (2007) to process travel times computed from loop data. Details of the method are omitted here, and one can refer to Small (2005) for detailed descriptions.

Several observations follow Figures 3 and 4. First, the most congested period for this route is the afternoon peak hours (3:00 p.m. to 7:00 p.m.) during which the time-dependent trend of travel times is evident. The median travel time increases from about 15 min to almost 35 min from 3:00 p.m. to 5:00 p.m., then decreases to about 15 min after 7:00 p.m. For other periods of the day, travel times are fairly stable. This pattern can also be observed on other days. Second, for the congested period (i.e., afternoon peak hours), the travel-time dispersion at a given instant is small, whereas it is much larger during noncongested periods. The second observation may be explained by the difference in driving behavior in congested and noncongested periods. That is, under noncongested periods, drivers have more freedom to maneuver and drive at their desired speeds, resulting in higher variation of individual travel times. During heavily congested periods (which is the case for the afternoon peak hours of the studied route), however, drivers cannot do too much other than going with the flow. Hence, their individual driving preferences may not be reflected at all, resulting in the nearly homogeneous travel times during the congested period.

This pattern has been previously suggested by Daganzo (1997, p. 142) who stated that in a queue, delays (and thus travel times) can be predicted “independent of drivers’ shenanigans.” Our results here provide empirical evidences of the pattern. Note that the travel-time dispersion in off-peak hours ranges, as expected, from about 800 s to 1,000 s, corresponding to speeds from 55 to nearly 70 miles per hour.

Our observation is not inconsistent with previous findings in Chen et al. (2003) who reported that travel-time variability is proportional to the mean travel time. The reason is that in Chen et al., the variability is calculated on the basis of travel times from multiple days, which is mainly day-to-day variability. In our case, variations come from different driving behaviors. One should also note that our observation is based on a heavily congested route; for lightly congested routes, large dispersions may still be observed during peak hours.

The aforementioned discussions also imply that LMM is appropriate for processing the FasTrak data (or other types of probe vehicle data). This is because although LMM generates large bandwidths during off-peak hours as shown in Figure 2, the ground-truth travel times do not change much either. Therefore, the large time window in off-peak hours will not smooth out the trend of travel times. Although the trend changes fairly rapidly during peak hours, LMM generates rather small bandwidth (mostly less than 15 min), which should be adequate to capture the time-dependent trend of travel times during peak hours.

Methodology for Performance Evaluation of Travel-Time Estimation Methods

This section describes how the evaluation is conducted. The travel-time estimation methods include algorithms applied to the speed data for individual lanes and the average of all lanes. Two quality measures are defined and applied. Following a description of the study site (route), two scenarios used to examine the effect of different detector spacing are described.

Travel-Time Estimation Algorithms

We evaluate three benchmark travel-time estimation algorithms: the instantaneous, the dynamic, and the linear-regression
(LR) algorithms. The instantaneous travel time assumes that traffic conditions remain unchanged from the time a vehicle enters a route until it leaves the route. Therefore, it is a snapshot travel time computed by considering only the current prevailing traffic information (speeds or travel times). The travel time of a route can be computed by simply summing the travel times of the constituent links at the time the vehicle enters the route. The dynamic route travel time is also a summation of travel times of its constituent links; however, the link travel time is computed using the traffic condition at the future time when a vehicle would enter a particular link. In other words, we compute the dynamic travel time by walking through the traffic data both temporally and spatially. Therefore, the calculation of dynamic travel times requires future traffic information.

Figure 5 demonstrates these two types of travel times. Here, we only provide some conceptual descriptions of the two methods; more detailed discussions can be found in Lindveld et al. (2000). The figure depicts a time-space diagram, where station \( i \) denotes the \( i \)th loop station. The route is comprised of \( n \) stations, and one is interested in calculating the route travel time at time \( t_0 \). The thin line represents the instantaneous trajectory of an imaginary vehicle traversing the route, approximated by instantaneous loop speeds at time \( t_0 \). For a link between two stations, the link speed is assumed to be the average speed of the two stations (shown as blue dots in Figure 5). The approximate trajectory of a vehicle on a link can then be constructed as a straight line. Connecting all link trajectories results in the route trajectory, which is piecewise linear. Notice that all dots are for time \( t_0 \) when the route travel time is calculated. The dynamic trajectory, as indicated by the bold line in Figure 5, is also constructed by link trajectories. However, the speed of a link is the average speeds of its two neighboring stations at the time when the imaginary vehicle enters the link. Because loop detectors usually provide speeds at discrete time instants (such as every 5 min), interpolation is needed when the entrance time does not coincide with the exact time grid. The loop speeds that are needed for calculating dynamic travel times at \( t_0 \) are shown using squares in Figure 5, which span to time \( t_m \) in the time domain. In Figure 5, it is obvious that the instantaneous travel time can be calculated in real time, whereas future speeds are needed to compute the dynamic travel time. For real-time applications, future traffic information may be available through traffic prediction models. In this article, the dynamic method is applied through postprocessing: the dynamic travel time at time \( t_0 \) can be calculated after time \( t_m \) as shown in Figure 5 when the imaginary vehicle finishes its trip.

The LR algorithm combines (linearly) the instantaneous and historical dynamic travel times so that both the real-time and historical trends of travel times for a given route can be considered to certain extent (Chen et al., 2004; Rice and Zwet, 2001). The LR algorithm can be expressed using the following equation:

\[
d_0^l(t) = \bar{d}^l(t) + (\bar{d}_r^d(t) - \bar{d}_r^l(t)) \cdot \lambda(t). \tag{3}
\]

Here, \( d_0^l(t) \) : the LR travel time for route \( r \) at departure time \( t \), \( \bar{d}^l(t) \) : the average dynamic travel time at time \( t \) computed from historical data, \( \bar{d}_r^l(t) \) : the instantaneous travel time computed at time \( t \), \( \bar{d}_r^d(t) \) : the average instantaneous travel time computed at time \( t \) from historical data, \( \lambda(t) \) : parameter which needs to be estimated.

Parameter \( \lambda(t) \) can be estimated using least square estimation using historical data. (For details, see Chen et al., 2004.) In practice, \( t \) is discretized into 5-min intervals; that is, we will have 288 parameters for a given route for an entire day. Because \( \bar{d}_r^l(t) \) and \( \bar{d}_r^d(t) \) can be computed with historical data in advance, we can calculate the LR travel time in real time using Equation (3).

In summary, computing the instantaneous and the LR travel times only require real-time speeds. Therefore, they are suitable for real-time traffic applications, such as posting travel times on CMS. The instantaneous method is the simplest travel-time algorithm and is used widely in practice. For example, the CMS travel-time system in the San Francisco Bay Area is using this method for computing travel times. Alternatively, the LR method represents a class of more sophisticated algorithms trying to combine real-time information and the historical data, which are numerous in the literature. The dynamic travel time needs speeds in the future, and their application requires future traffic information that may be obtained through traffic-prediction models. The dynamic travel-time algorithm also provides benchmark travel times that the other two algorithms can compare with. We note that the evaluation methods and scenarios proposed in this article are general and can be directly applied to other algorithms if needed.

**Quality Measures**

The travel-time estimates given by the three algorithms are point estimates. In this section, we define two quality measures: the relative error and the accuracy index. The former is based on the absolute difference between the loop-based travel-time estimate and the median travel time from FastTrak. The latter is based on whether the loop-based travel-time estimate falls
within a certain range of travel times estimated from the dispersion of the FasTrak data.

The relative error. From the processed FasTrak data, we obtained \( \hat{T}_r(t) \), the median travel time for route \( r \) for vehicles entering the first link of \( r \) at time \( t \). Here, \( t \) is the discrete time instant (e.g., in every 5 min). Similarly, \( \overline{T}_r(t) \) is the loop-based estimated travel time for the same route at time \( t \). Then, the relative error can be defined as follows:

\[
E_r(t) = \left| \frac{\hat{T}_r(t) - \overline{T}_r(t)}{\overline{T}_r(t)} \right|.
\]  

Equation (4) defines the accuracy measure for a particular time instant, which is referred as *disaggregated* measure. Sometimes aggregating quality measures over a certain time period may be of more interest, especially from practitioners’ point of view. The typical periods of a day may include morning off peak, morning peak, midday, afternoon peak, and afternoon off peak (Fujito et al., 2006). For a given period \( p \), the aggregated measure can be computed using the following equation:

\[
E_r^p = \frac{\sum_{t=1}^{n} E_r(t)}{n}.
\]  

Here, \( n \) is the total number of estimates within the period \( p \).

The accuracy index. As previously mentioned, the relative error measure does not capture the dispersion of travel times. In this article, we (a) construct an interval by the estimated 15th and 85th percentile travel times on the basis of the FasTrak data and (b) define an accuracy index of the loop-based estimated travel time at time \( t \), denoted as \( A_r(t) \). The accuracy index is 1 if the loop-based estimated travel time lies in the interval; otherwise, it is 0:

\[
A_r(t) = \begin{cases} 
1, & \hat{T}_{15}(t) \leq \hat{T}_r(t) \leq \hat{T}_{85}(t) \\
0, & \text{otherwise}
\end{cases}
\]  

Here \( \hat{T}_{15}(t) \) and \( \hat{T}_{85}(t) \) denote the estimated 15th and 85th percentile travel times at \( t \), respectively, on the basis of FasTrak data. For simplicity, we will subsequently refer the estimated 15th/50th/85th percentile travel time at \( r \) on the basis of FasTrak data as the 15%/50%/85% *ground truth*. The accuracy index at a single time instant is a binary value (0 or 1). This definition can be extended to a time period, such as morning or afternoon peak hours, as follows:

\[
A_r^p = \frac{\sum_{t=1}^{n} A_r(t)}{n},
\]  

where \( n \) is the number of time instants in the time period \( p \). The accuracy index over a time period defined in Equation (7) may be more practical than what is defined in Equation (6) for a single time instant.

The Study Site

We selected a 15-mile section of the I-80 freeway from the city of Albany to the Carquinez Bridge in the San Francisco Bay Area for the evaluation of the methods (see Figure 6). In this figure, the dark arrow and the asterisk indicate, respectively, the origin and destination of the route. Most of the evaluation route has four lanes, but the last six miles of the route has only three lanes. The leftmost lane is designated as the HOV lane during the morning (5:00 a.m. to 10:00 a.m.) and afternoon (3:00 p.m. to 7:00 p.m.) peak hours. We further selected 4 weekdays in mid-September 2006 for the evaluation. There are 33 double-loop detectors deployed almost evenly in this route, and most of them worked properly during the 4 evaluation days. We use 5-min loop speeds to compute the estimated travel times. The data can be downloaded from PeMS (PeMS, 2009), where the data quality is also reported.

Scenarios for Examining the Effect of Different Detector-Spacing Scenarios

We evaluated two scenarios of detector spacing. First, we tested the baseline (existing) configuration of detector spacing, approximately half mile. We were interested in assessing the performance of the three travel-time estimation algorithms at different times of day, which we used as a proxy for recurrent congestion. In this article, different periods of a day are defined as morning off peak (12:00 a.m. to 7:00 a.m.), morning peak (7:00 a.m. to 10:00 a.m.), midday (10:00 a.m. to 3:00 p.m.), afternoon peak (3:00 p.m. to 7:00 p.m.), and afternoon off peak (7:00 p.m. to 12:00 a.m.). Furthermore, because there are loop detectors in each lane, we computed the lane-by-lane travel times and compared the resulting performance.

In the second scenario, we varied the detector spacing and investigated how it would affect travel-time estimation. For this purpose, we randomly took out detectors. This scheme was previously used in Kwon et al. (2006) and Fujito et al. (2006). However, detectors were removed in Kwon et al. purely randomly so that the remaining detectors may be distributed highly unevenly. Thus, for the same average spacing, the estimated travel times may be very different depending on whether sensors are located. In Fujito et al., however, detectors were taken out in such a way that the remaining detectors were distributed almost evenly. Hence, the method in Fujito et al. resulted in only a few detector-deployment settings for a given number of remaining detectors, and the variation of performance may not be easily captured.

In this article, we generalize the random selection process by randomly removing detectors in such a way that the remaining detector spacing satisfies the following condition:

\[
\max_{i \in N} s_i - \min_{j \in N} s_j \leq \rho \bar{s}.
\]  

Here \( s_i \) is the \( i \)th spacing, \( \bar{s} \) is the average spacing, and \( \rho \) is a constant. By using different values of \( \rho \), one can control the variation of individual detector spacing for a given number of detectors. For example, if \( \rho = 0 \), we required sensors to be deployed absolutely evenly (close to what was done in Fujito et al.); if \( \rho = +\infty \), sensors can be selected completely randomly.
In our study, $\rho = 2$ was used.

We implemented the aforementioned method as a random selection algorithm. For the route shown in Figure 6, we removed detectors in such a way that the resulting average detector spacing is 0.75 mile, 1 mile, 1.5 miles, 2 miles, 2.5 miles, 3 miles, and 5 miles, respectively (i.e., the remaining number of detectors is correspondingly 20, 15, 10, 8, 6, 5, and 3). For each of the aforementioned seven scenarios, we ran the random selection algorithm for multiple times so that 100 distinct detector settings were obtained. These detector settings are used in the Effect of Different Detector-Spacing Scenarios section of this article to evaluate the performance of travel-time algorithms under different detector spacing scenarios.
Evaluation Results

Baseline Scenario

For space consideration, we only show results on the basis of the September 6, 2006 data, as shown in Figure 7 for the evaluation route. Similar results can also be found for the other 3 evaluation days. Figure 7(a) is for instantaneous travel times, Figure 7(b) is for dynamic travel times, and Figure 7(c) is for LR travel times. In each figure, the estimated travel times using speeds of individual lanes are shown in different broken lines. Lane 1 is the leftmost lane, and during peak hours only high-occupancy vehicles are allowed in this lane. Lane 2 is the second from the left, and Lane 3 is the third from the left. The line labeled all lanes results from the average speeds across all lanes except for the afternoon peak hours (from 3:00 p.m. to 7:00 p.m.). During the afternoon peak, Lane 1 speeds are excluded from the calculation of the average speeds. This is because, as mentioned in the LMM Algorithm section, the LMM method in this article effectively filtered travel times from Lane 1 during the afternoon peak. For a fair comparison, therefore, Lane 1 speeds are also excluded when calculating average speeds for all lanes during the afternoon peak period. Furthermore, the ground-truth travel times are plotted with solid lines, each representing a different percentile.

It turns out that for the evaluation route, when the same speed data are used, different estimation algorithms do not make much difference, under both the free-flow and the recurring-congestion conditions. Theoretically, the dynamic travel times should be superior to instantaneous travel times when congestion forms or dissipates rapidly. However, because the route travel time is relatively short (15 min under free-flow conditions and 35 min when most congested), and the transition from free-flow condition to maximum-congestion condition is slow (taking almost 2 hr), the travel times from the tested estimation algorithms are not significantly different. This suggests that under similar circumstances, using the instantaneous travel-time algorithm is sufficient.

Our second observation is that data from different lanes greatly affect the estimated travel times. Because all three estimation algorithms yield similar results, we focus on Figure 7(b) for simplicity. During the free-flow period, travel times calculated using Lane 1 data are the closest to the 15% ground truth line; travel times estimated with Lane 2 data are the closest to the 50% ground truth line; and travel times using Lane 3 data are often longer than the ground truth (except during 12:00 a.m. to 7:00 a.m.). This is plausible because the speed in the right lane is lower than that in other lanes, and travelers going through the whole route (who are the most relevant users of the travel-time display) tend to stay in left lanes, except during 12:00 a.m. to 7:00 a.m., when a large percentage of through traffic are trucks, and they stay in right lanes more often than do cars. Travel time estimated with data from all lanes (i.e., the average speed) is closest to the 85% ground truth, in general. To display the travel time estimated for an average driver who is interested in the median travel time, using data from Lane 2 is a better choice than using data for other lanes or for all lanes.

During the recurring-congestion period (i.e., afternoon peak), the percentile ground-truth lines should be interpreted a little differently. As explained in the LMM Algorithm section, we processed FasTrak data to eliminate outliers, which, in this study, effectively removed travel times from HOV vehicles during afternoon peak hours. When we turn to the loop-detector data, it is not surprising that estimated Lane 1 (the leftmost lane) travel times are significantly less than the percentile ground-truth lines because the former is for the HOV lane and the latter is for regular lanes. Meanwhile, travel times from Lane 2 underestimated the ground-truth travel times during this period, whereas Lane 3 travel times generally overestimated the ground-truth travel times. Travel times estimated with all-lanes data (note that Lane 1 speeds are excluded), however, are the closest to the median ground truth lines. The aforementioned findings can also be observed for other evaluation days.

The second observation can also be easily seen from Table 1, which shows the aggregated relative errors and accuracy indexes of dynamic travel times from the average and lane-by-lane speeds computed using Equations (5) and (7). In Table 1, we show in bold text the best performance for each time period. It is clear, especially through the relative errors, that all-lanes travel times have the best performance during afternoon peak period, whereas Lane 2 travel times have the best performance for the other periods of the day (except the morning off-peak period during which the all-lanes error is slightly smaller than that for Lane 2). Note that although one prefers smaller relative errors, larger values of the accuracy index represent higher possibilities that the loop-based travel times lie within the interval of FasTrak-based 15th and 85th percentile travel times, and they are thus more desirable.

<table>
<thead>
<tr>
<th>Period of Day</th>
<th>Relative Error</th>
<th>Accuracy Index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All Lanes</td>
<td>Lane 1</td>
</tr>
<tr>
<td>Morning off peak</td>
<td>0.02989</td>
<td>0.12954</td>
</tr>
<tr>
<td>Morning peak</td>
<td>0.13502</td>
<td>0.06701</td>
</tr>
<tr>
<td>Midday</td>
<td>0.10898</td>
<td>0.07101</td>
</tr>
<tr>
<td>Afternoon peak</td>
<td><strong>0.05088</strong></td>
<td>0.34182</td>
</tr>
<tr>
<td>Afternoon off peak</td>
<td>0.07710</td>
<td>0.05555</td>
</tr>
</tbody>
</table>

Bold values indicate the best performance for the given period of day.
Figure 7  (a) Instantaneous travel times versus ground-truth travel times. (b) Dynamic travel times versus ground-truth travel times. (c) Linear-regression travel times versus ground-truth travel times.
Effect of Different Detector-Spacing Scenarios

To evaluate the effect of different detector spacing scenarios, we only investigated the congested period (i.e., the afternoon peak hours). One may expect that detector spacing does not greatly change the performance for noncongested periods when vehicles are in (nearly) free flow. Our previous discussions showed that the average travel times estimated using all-lanes speeds are the closest to the ground-truth travel times during afternoon peak hours (note that Lane 1 speeds were excluded in this period). Therefore, we focus on average travel times in this section. The performance measure we use is the aggregated relative error defined in Equation (5). We first show, for each of the 4 weekdays, the aggregated relative error versus detector spacing, as depicted in Figure 8(a). For each detector-spacing scenario (except the baseline scenario), we randomly generated 100 detector-deployment configurations. We show in each figure the boxplot among these 100 configurations. Because there is only one baseline scenario (which is currently used in the field), the plot for the baseline (i.e., the 0.5-mile spacing) in Figure 8(a) reduces to a single point (i.e., the short horizontal bar).

Two observations thus follow. First, the median relative error increases, slowly and not monotonically, as detector spacing increases. This is intuitive because as the distance between detectors increases, less information can be collected regarding the traffic condition, which leads to less accurate estimation of travel times.

The second observation is that as detector spacing increases, so does the variation of the relative errors. This can be seen more clearly in Figure 8(b), which shows the difference of the 75th and 25th relative errors (i.e., the so-called interquantile of relative errors). From the Scenarios for Examining the Effect of Different Detector Spacing Scenarios section, the variation of relative errors for each detector-spacing scenario for a single day is obtained from 100 detector configurations, which were randomly generated. Therefore, the large variation of performance for larger detector spacing seems imply that if detectors are more sparse, the accuracy of travel-time estimation becomes more sensitive to the actual locations of the detectors. When the detectors are dense, nevertheless, the actual locations of detectors do not matter too much. This finding is consistent with our recent investigations on optimal sensor deployment for freeway travel-time estimation (Ban et al., 2009).

Concluding Remarks

In this study, we evaluated performance of a set of estimation methods (i.e., algorithms that estimate route travel times using specific speed data from dual loop detectors, for real-time applications. We first proposed the LMM method to process travel-time data from probe vehicles. The method captures the characteristics of commuter trips during both off-peak and peak hours, and allows us to estimate percentiles of travel times on
Figure 8  (a) Performance of detector spacing—relative error. (b) Performance of detector spacing—variation.
the basis of ground-truth data. We verified that during heavily congested peak hours, travel-time dispersion in a small interval is small.

We then compared the performance of the three travel-time estimation algorithms that use speed data from loop detectors. We found that when the route travel time is relatively short and the transition from free-flow to maximum congestion is slow, the differences using different estimation algorithms are not significant, and the instantaneous travel time can be adopted for its simplicity. Alternatively, using speed data from different lanes makes significant differences. For example, for the evaluation route, we recommend calculating travel times using speeds from the middle-lane (Lane 2) only during free-flow periods, but speeds from all lanes during the recurring-congestion periods. This finding may be site-specific, but it suggests that for other routes, both the average and lane-by-lane travel times should be computed from archived historical data from loop detectors, and compared to probe vehicle data. If patterns similar to those shown in this article are found, lane-by-lane loop data may then be used to improve travel-time estimation quality.

We also evaluated the performance of travel-time estimates with different detector-spacing scenarios. We found that both the median relative error and the variation of relative errors increase (not monotonically) as detector spacing increases, and travel-time estimation quality becomes more sensitive to actual sensor locations when sensors are sparse.

The concept of lane-by-lane travel time is hypothetical in the sense that the travel time may not correspond to any real vehicles; however, the lane-by-lane travel time is computable as long as lane-specific speeds are available. The lane-by-lane travel times may merit further investigations. For this, the differences in travel times for vehicles that keep traveling on specific lanes should be studied when trajectories of individual vehicles become widely available (e.g., through the IntelliDrive, 2009).

We only investigated in this article travel times computed using 5-min resolution data from loop detectors on a single route for 4 weekdays. We are currently evaluating performance of travel-time methods using other types of data sources (such as 30-s resolution loop data and speed radar sensors) on multiple routes and over an extended period of time. Finer resolution data are expected to improve the performance of travel-time estimation methods during peak periods, whereas data over multiple days can help determine whether revealed patterns for a route is consistent across a longer period of time. Results in this regard will be presented in subsequent articles.

REFERENCES


