BENCHMARKING TRAVEL TIME ESTIMATES

JD Margulici*
CCIT, Institute of Transportation Studies (ITS)
University of California - Berkeley
2105 Bancroft Way, Suite 300
Berkeley, CA 94720-383
Phone: 510-642-5929
Email: jd@calccit.org

Xuegang (Jeff) Ban
CCIT, Institute of Transportation Studies (ITS)
University of California - Berkeley
2105 Bancroft Way, Suite 300
Berkeley, CA 94720-383
Phone: 510-642-5112
Fax: 510-642-0910
Email: xban@berkeley.edu

Re-submittal to IET Journal of ITS

March, 2008

** Corresponding Author*

ABSTRACT

Travel time estimates are widely regarded as the most practical information about traffic conditions available to individual drivers. While there are numerous data collection and estimation methods in use today, few attempts have been made to evaluate them in a systematic manner. Even more fundamentally, there are no broadly accepted metrics to measure the quality of travel time estimates. This paper exposes the methodology and tools employed to conduct a benchmark of travel time estimates in the San Francisco Bay Area. The methodology and the proposed quality measures are intended to set a standard that can be universally applied. Their use is illustrated through a sample data set collected for 24 hours on one Bay Area freeway.

1 BACKGROUND AND PROBLEM

1.1 INTRODUCTION

Travel time estimates on selected itineraries represent information that is easy for the driving public to understand and process. Numerous studies reveal that commuters appreciate and value travel time information, which reduces their uncertainty and their stress (Peng et al., 2004; Lindveld et al., 2000; Khattak et al., 1994). Further, reliable information can arguably enable travelers to make educated choices about their itinerary, departure time or even transportation mode, with the result of bringing about a form of “system self-management.”

Travel time estimates have benefited from a flurry of innovations in traffic data collection, processing techniques, and information delivery modes over the past decade. Academic research has been very active in this area (Oda, 1990; Smith and Demetsky, 1997; Huisken and Maarseveen, 2000; Rice and Zwet, 2001; Hartley, 2003; Hinsbergen et al, 2007, to name just a few). On the front end, both government agencies and private media ventures across the world’s largest cities provide traffic information and travel time estimates through a variety of channels, including web browsing, traditional and satellite radio, mobile devices, navigation units and, increasingly, electronic signage on roadways.

1.2 WHY BENCHMARKING TRAVEL TIME ESTIMATES MATTER

Providers of traffic information, whether public or private, compete on two essential features: usability and information quality. In fact, it is often argued that it is because of shortcomings in both features that subscription-based information services have not yet established a substantial user base. In particular, measuring the quality of travel time estimates is important for the following reasons 1) the margins of errors of travel time estimates should be better understood and formulated so that drivers can develop adequate expectations; 2) robust validation and monitoring practices for travel time estimates can point to needed improvements in traffic data collection and they build up the confidence of network operators in the information that is delivered to the public; 3) in the context of public-private partnerships for data collection, aggregation and dissemination, quality metrics would help both government agencies and technology providers reach business agreements and develop a market.

In the literature, systematic benchmarks of travel time estimates have not been conducted in an authoritative manner, and debates over information quality are often anecdotal. Lindveld et al. (2000) is one of the few studies focusing on evaluating performances of several travel time estimation methods using loop detector data. The ground-truth travel times in this study were collected via license plate readers, floating car runs, and toll ticket collection. However, the number of observed data points using floating cars is not sufficient; travel times from toll ticket collection have problems as well (Lindveld et al., 2000, pp. 46). Zhang et al. (1999) studied travel time estimation methods based on single loop detector data. Floating car runs were conducted to gather the ground truth travel times. As pointed out in Kwon et al. (2006), however, limited floating car runs may be biased. Kown et al (2006) and Fujito et al. (2006) studied the relationship 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. As shown in Ban et al. (2007), this may be very different from actually experienced travel times by individual drivers. Recent studies by Dance (2007) are based on speed contour maps, but no quantitative measures were developed. Therefore, previous discussions on travel time quality evaluations were limited at least in the sense that 1) ground truth travel times from probe vehicles were not widely available, and 2) no widely accepted quality measures were developed.

Note also that evaluating travel time estimation quality is essentially different from studying travel time reliability, an issue that has recently gained much attention in the transportation research community (Chen et al., 1999; Chen et al., 2003; AL-DEEK and Emam, 2006; Liu et al., 2007). The former focuses on the differences of estimated and actual travel times, while the latter is supposed to study the features (such as distribution) of the actual travel times. Therefore, performance measures developed for travel time reliability (Fisher et al., 2003) may not be used directly for evaluating the performance of travel time estimations.

1.3 BENCHMARKING REQUIREMENTS

Benchmarking travel time estimates requires three conditions. First, significant volumes of ground truth data have to be collected. Ground truth data means observed trip times of vehicles traveling on the corridors along which the benchmark is conducted. This may seem like a daunting requirement and has certainly been the main obstacles to more systematic studies in the past. Most validation programs carried out by either government or private enterprises employ a limited fleet of so-called probe vehicles to record sparse observations. However, several technology options are now available to collect actual trip times much more massively: toll tag readers, license plate readers, cell phones, GPS-equipped professional fleets provide avenues not only for better estimates in the first place, but also for validation data.

Second, there needs to be clearly defined metrics to measure the quality of estimates over time. No such metrics have been developed and promoted in the industry to date. Part of the problem lies in defining what a good travel time estimate is: in effect, every driver actually experiences a slightly different trip time. Then again, the absence of adequate data, as described in the previous paragraph, has limited the practical importance of systematic metrics for quality measurements.

The third requirement for a systematic benchmark is the accumulation of sufficient data to allow meaningful comparisons. Clearly, an evaluation of travel time estimates that is based on several weeks of collected data will carry for more weight than an anecdotal study of the peak hour on a randomly selected day. In turn, this data requirement points to setting up an adequate computing infrastructure that insures proper collection, storage and processing of large volumes of data.

2 METHODOLOGY

In this paper, we propose a methodology and tools to conduct a systematic benchmark of travel time estimates on selected corridors. One of the main innovations in this methodology is the development of adequate metrics to track the quality of travel time estimates. Another key component of the proposed methodology is a database that enables the collection of data from multiple sources over extended periods of time. Functions have been developed to process the data that is hosted in the database, resulting in automated and systematic evaluation procedures.

On the flip side, the proposed methodology assumes that significant numbers of observed trip times can be collected. As already noted, this requirement is becoming much less stringent. For example, an agency interested in this methodology could affordably deploy a pair of mobile license plate readers and rotate it among the various routes that it wishes to benchmark. Recently, a team of researchers in Minnesota has used time-stamped voice recorders to register the last four digits cars' license plates passing by selected locations, thereby accomplishing a similar feat with very little equipment costs!

As an example, the methodology is applied to a benchmark of travel times estimates along four corridors in the San Francisco Bay Area. In that area, massive amounts of anonymized trip times are available from toll tag readers that have been installed as part of the local 511 program.

2.1 METRICS DEVELOPMENT

Measuring the quality of travel time estimates is complicated by the fact that there is no single trip time value on a given road segment at a particular time. Not all vehicles travel at the same speed, and drivers experience different trip times as a result. However, for practical purposes, traveler information systems provide an estimate that captures the likely trip time of most vehicles, whether it is on the web, a phone vocal server, or a changeable message signs. Note that some services or public agencies provide a range instead of a single value. This may arguably be a better way to communicate expectations, but the range is still derived from a baseline estimate. We can therefore assume that a traveler information system tries to provide, for each route and each refresh interval, a single travel time estimate. In effect, the system produces a time series $\hat{\tau}(t)$. It is the quality of this estimate over time that is in question.

Each individual driver observes two values: a travel time estimate, and his or her actual trip time. Therefore, it is possible to calculate an individual relative error between those two values, defined as the ratio of their difference to the actual trip time. Individual relative errors should ideally be as close as possible to zero. They deviate because 1) the estimate may be biased and 2) certain individuals travel slower or faster than most other drivers. The second factor is not controllable but will nonetheless affect the perceived usefulness of the estimate to those individuals.

Assume we can observe *M* drivers in total over a period of time *T*. During that period, the travel time estimate $\hat{\tau}(t)$ may change many times. The m-th driver's actual trip time is τ_m and its estimated travel time is $\hat{\tau}_\mathrm{m} = \hat{\tau}(t)$. We define the *relative error* for that driver, denoted e_m, as:

$$e_m = \frac{\hat{\tau}_\mathrm{m} - \tau_\mathrm{m}}{\tau_\mathrm{m}} \tag{1}$$

Note that e_m has convenient properties. First, it is an algebraic number, either negative or positive depending on whether $\hat{\tau}_\mathrm{m}$ under- or overestimates τ_m. Second, it is a relative measure, which eliminates the need to account for route length. Relative errors for different trips can be readily compared.

From there, we assemble two metrics: aggregate error and relevance. The *aggregate error* captures the overall inaccuracy of the estimates over the time interval T. It is defined as the mean relative error within this period, i.e.

$$E_T = \frac{\sum_{m=1}^{M} e_m}{M}. \tag{2}$$

The aggregated error is an algebraic percentage value that can be assimilated to the systemic bias of the estimate. Note that it is a more informative measure than, say, averaging the absolute values of individual relative errors. Absolute values would always add up to a non-negligible error term, which depends more on the variability of travel times between individual drivers than on how representative the travel time estimates are. As an example, consider two drivers traveling on the same route, for which they are given a trip time estimate of 20 minutes. If one driver experiences a travel time of 18 minutes and the other driver experience a travel time of 22 minutes, the aggregate error would be null, as it should be. On the other hand, the average absolute error would be 10%. It reflects variability, but not accuracy.

Yet, variability between drivers is also of interest. Even if an estimate accurately tracks the variations of the mean driving times, we should still ask what the perception from the public will be. Each individual driver will judge for themselves how accurate *they* think the estimates are. To that end, the *relevance measure* sets an acceptable error threshold and captures the proportion of drivers whose actual driving time differs from the estimate given to them by less than that threshold. In other words, if ε is the acceptable threshold, say a 15% error, then the relevance measure for time period of T is formulated as:

$$R_T^{\varepsilon} = \Pr(|e| \le \varepsilon). \tag{3}$$

where e represents the relative error of an arbitrary vehicle traveling during period T, and Pr refers to the empirical probability defined by the distribution of errors observed during that same period.

To better illustrate the concepts of aggregate error and relevance measures proposed in this report, we provide an example below. Table 1 lists the actual and estimated travel times for individual drivers for a given time period (in total 15 drivers). The relative error (calculated by equation (1)) and the absolute value of the relative value for every driver are also shown in the table. For the presented sample, the aggregated error is the mean value of the forth column: -7%. The 10%-relevance (R10) measure is 80%: 12 out of 15 vehicles have an absolute relative error 10% or less. Similarly, the 15%-relevance (R15) measure is 87%.

Driver #	Actual Travel Time	Estimated Travel Time	Relative Error (%)	Absolute Value of Relative Error (%)
1	18'27"	16'57"	-8	8
2	18'58"	16'57"	-11	11
3	18'56"	16'57"	-10	10
4	18'30"	16'57"	-8	8

5	20'12"	16'57"	-16	16
6	20'13"	16'57"	-16	16
7	20'47"	19'45"	-5	5
8	20'23"	19'45"	-3	3
9	20'45"	19'45"	-5	5
10	21'25"	19'45"	-8	8
11	21'41"	19'45"	-9	9
12	21'24"	20'59"	-2	2
13	20'48"	20'59"	1	1
14	20'59"	20'59"	0	0
15	21'13"	20'59"	-1	1

Table 1 - Performance Measures for Sample Travel Time Estimation

One of the most remarkable features of the two performance measures we have defined is that they can accommodate any route length or duration of observation. Their definition enables them to flexibly handle scaling the time-space area considered up or down because it is based on aggregating individual observations. It is also insensitive to the refresh rate of travel time estimates: it doesn't matter whether estimates are provided every minute or every five minutes. These properties are important in order to establish those measures as true *benchmarks*, in that they can be applied to any setting and produce directly comparable numbers.

2.2 TOOLS

The Messaging Infrastructure for Travel Time Estimates to a Network of Signs (MITTENS) was initially developed at CCIT to provide travel time estimates on freeway signs in the San Francisco Bay Area (CCIT, 2006). The data is provided by the Bay Area's 511 program operated by the Metropolitan Transportation Commission (MTC) and relies on a combination of toll tag readers, inductive loops and microwave radars.

MITTENS has been reconfigured to add an experimental component to its operational function. A data repository and a computing platform allowing travel time calculations from segment-level information, it is well positioned to collect data from additional sources and compare different estimates over long periods of time.

Figure 1 below shows the system architecture of MITTENS. It is basically a data archiving, processing, and dissemination system, with the MITTENS DB as the core component. In the figure, dashed lines indicate the components that are related to travel time evaluation. The "DataTranslator" retrieves traffic data from three data sources and archives the data into MITTENS DB. The "TTCalculator" module operates on the traffic data in MITTENS DB and computes travel times on pre-defined routes. Three travel time methods have been implemented in the system: *instantaneous*, *dynamic* and *Linear Regression* (*LR*) travel times. More descriptions regarding the three data sources and three travel time methods are provided later in this paper. The "TTEvaluation" module compares the performances of the three travel time methods over the data sources with the "ground-truth" travel times from probe vehicles.

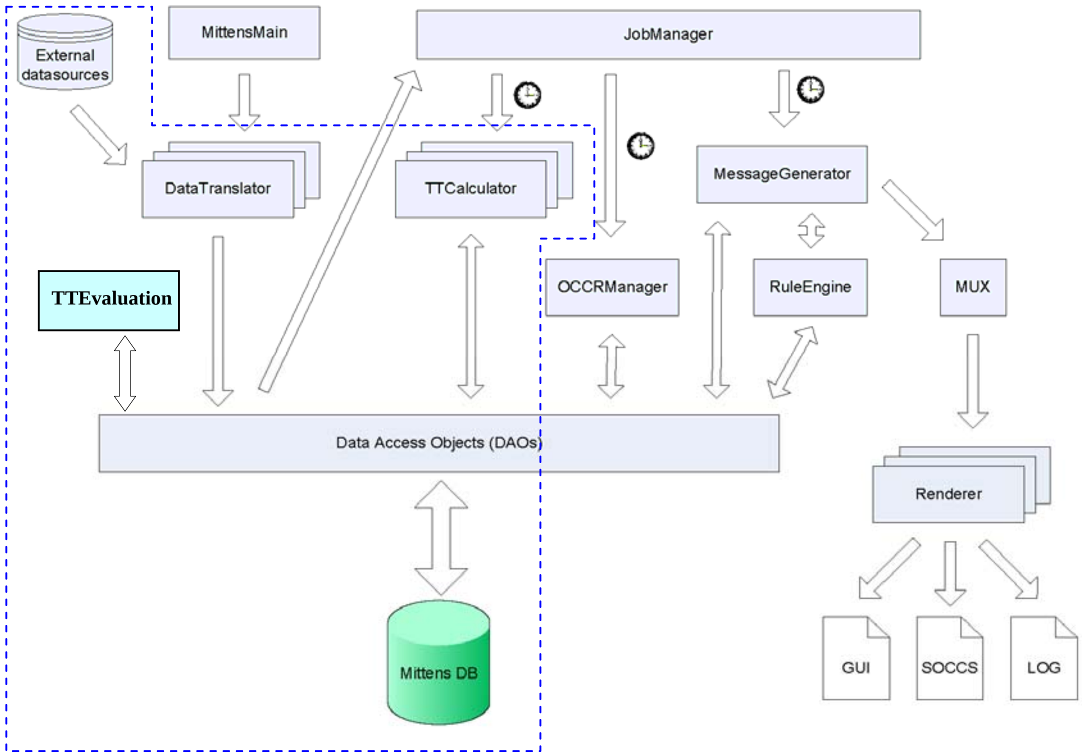

Figure 1 - System Architecture of MITTENS

The basic premise of the benchmarking methodology is to assemble anonymous individual trip times. In the case of the San Francisco Bay Area benchmark, those trip times are provided by toll tag readers (i.e., FasTrak readers). While such observations do not result in perfect estimates when processed in real time, they do provide large, continuous volumes of ground truth data. MITTENS performs both on-line and off-line calculations to calculate travel times estimates. It then uses the ground truth data to assemble the aggregated error and relevance metrics for each route of interest and estimation method over arbitrary time periods.

3 DESIGN OF EXPERIMENT

For the San Francisco Bay Area benchmark, four routes were selected. Data fed to MITTENS includes toll tag reader data, which, ex-post, constitutes ground truth data, as well as fixed detector data from inductive loops, microwave speed radars, and 511 link-based travel times. Contacts have also been made with private providers of real-time traffic information to obtain their own estimates that will be stored and benchmarked over time.

3.1 CORRIDOR SELECTION

The primary selection criterion for the four corridors was the coverage with respect to three data sources: SpeedInfo radars, Loop detectors, FasTrak, and 511 link travel times. All routes are defined between two FasTrak toll tag readers, which provide the ground-truth travel times. The first route is about 14 miles (22 KM) along Interstate 80 EB from the City of Albany to the Carquinez Bridge Toll Plaza. The second route is on WB I-80 from the Cutting Blvd Exit to the I-580 E Split. It is a little more than 3 miles (5 KM) in length. The third route is from North of Stevenson to South of SR-237 on I-880 SB, with a length of about 9 miles (14.5 KM). Finally, the forth route is on I-580 WB from West of the I-238 Split to East of SR-13, a length of about 7.7 miles (12 KM). Figure 2(a) – 2(d) depicts the locations of these four routes.

Table 2 shows the data coverage for the four routes. Loop detector data includes both 30-sec raw data and 5-min aggregated data. SpeedInfo radar sensors record the average speed across all lanes of a freeway location in 1-min intervals. FasTrak provides the ground-truth travel time of individual drivers between two toll tag readers. 511 travel time estimates are updated every minute for pre-defined freeway links by fusing different data sources.

	SpeedInfo	Loops	FastTrak	511 estimates
Route 1		✓	✓	✓
Route 2		✓	✓	✓
Route 3		✓	✓	✓
Route 4	✓		✓	✓

Table 2 - Data Coverage of the Four Evaluation Routes

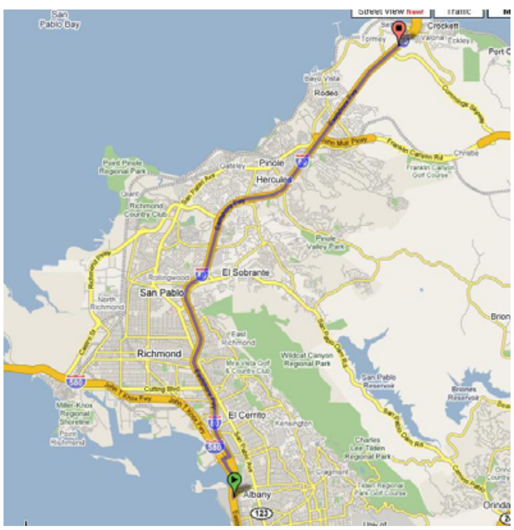

Figure 2 - Route 1 - EB I-80

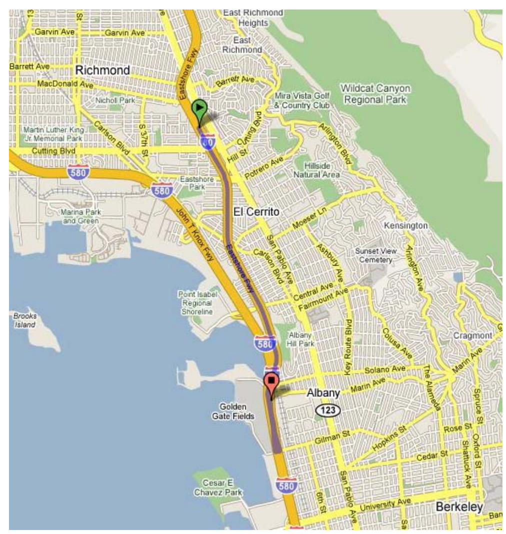

Figure 3 - Route 2 – WB I-80

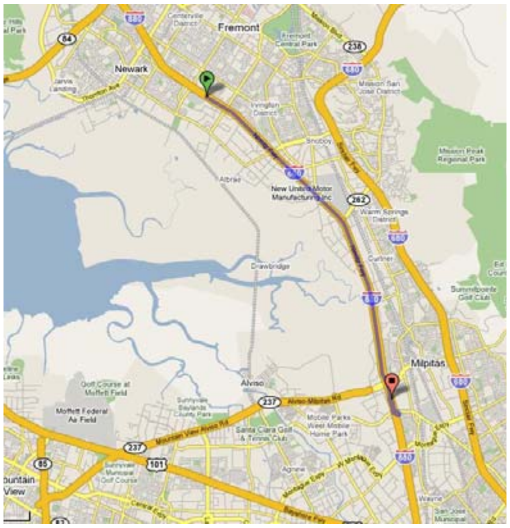

Figure 4 - Route 3 – SB I-880

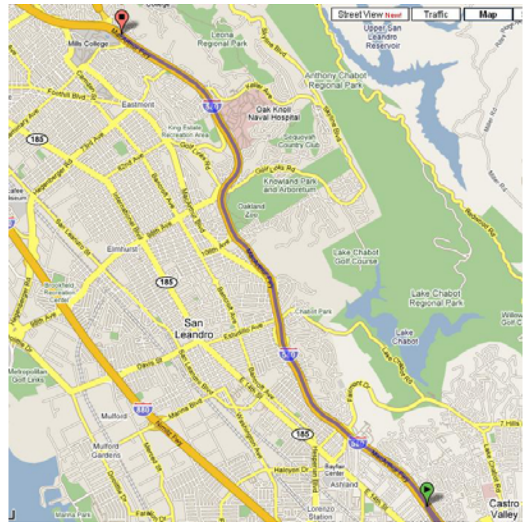

Figure 5 - Route 4 – WB I-580

3.2 ESTIMATION ALGORITHMS

We are benchmarking three travel time estimation algorithms, the so-called *instantaneous*, *dynamic*, and *linear-regression* (*LR*) travel time methods. All three use fixed detectors data, which is representative of the practice in most jurisdictions where trip times are estimated for traveler information purposes. The *instantaneous* travel time assumes traffic conditions remain unchanged from the time a vehicle enters a route until it leaves the route. Therefore, the route travel time can be computed by simply summing travel times of the constituent links at the time a 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 will be computed using the latest traffic condition at the time a vehicle enters a particular link. Note that this method can not be applied in real time, but it is useful for freeway performance monitoring and calibration. The *LR* method linearly combines the instantaneous and historical dynamic travel times so that the historical variations of travel times for a given route can be considered to certain extent. Detailed descriptions of the LR method can be found in Rice and Zwet (2001) and Chen et al. (2004).

4 RESULTS

Preliminary results were assembled for a 24-hour period on December 17th, 2007, and are presented for Route 3. As shown in Table 2, three data sources are available for this route: loops, Fastrak, and 511. Figure 6 shows the travel times collected from toll tag readers during that day.

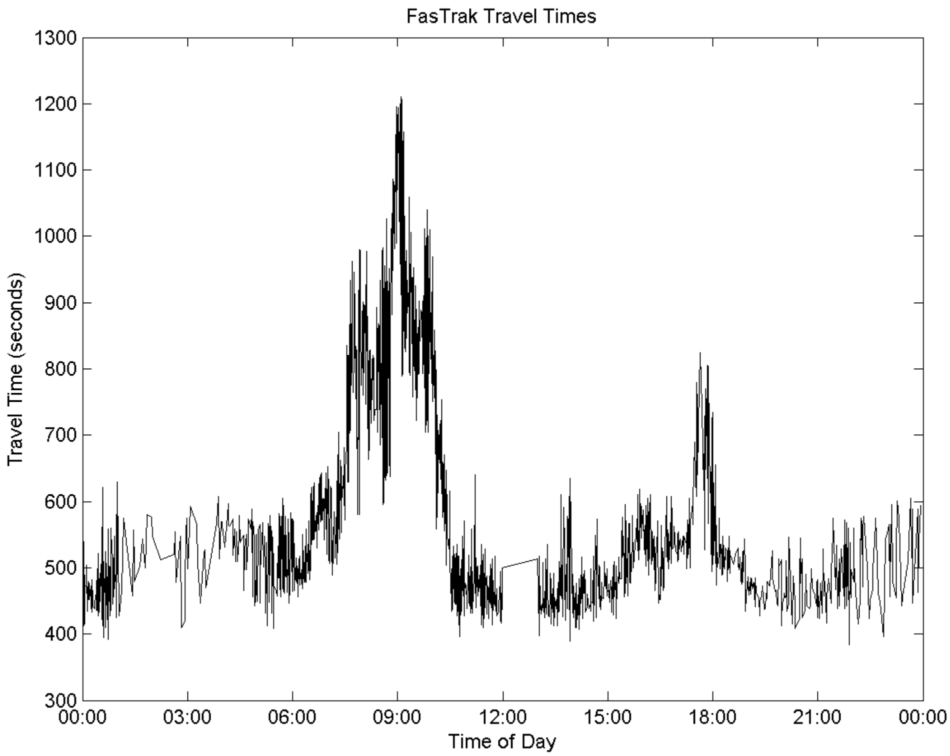

Figure 6 - Individual Travel Times

On that same day, travel time estimates produced by the San Francisco metropolitan area 511 service as well as from 5-min-interval loop detector data were assembled. The local 511 service estimates travel times from the data sources mentioned above, including toll tags. However, toll tag data is used in real time, and even though it provides a form of ground truth, there is a delay between the time of collection and the origination time, which translates into estimation latency. With loop data, estimates were assembled using the instantaneous method. The estimates produced by employing the dynamic of LR methods did not differ significantly from those derived by the instantaneous method and were therefore discarded. Possible explanations for the similarity are the facts that the study route is relatively short in length and that traffic conditions change relatively slowly. More fundamentally, this also seems to underline the importance of data sources, which, in the authors' opinion, trumps algorithmic refinements. However, this claim would need to be backed up by considerably more data to be asserted with certainty.

Figure 7 compares estimates from 511 as well as from loop detectors with median toll tag travel times computed in 10-min intervals. It appears that both sets of estimates do a reasonably good job of tracking the overall trend of measured travel times, especially the morning peak. However, it is also clear that estimates fall short of accurately measuring some of the deviations. For instance, the 511 estimate pretty much misses the afternoon peak, while the loop estimates overestimate its magnitude.

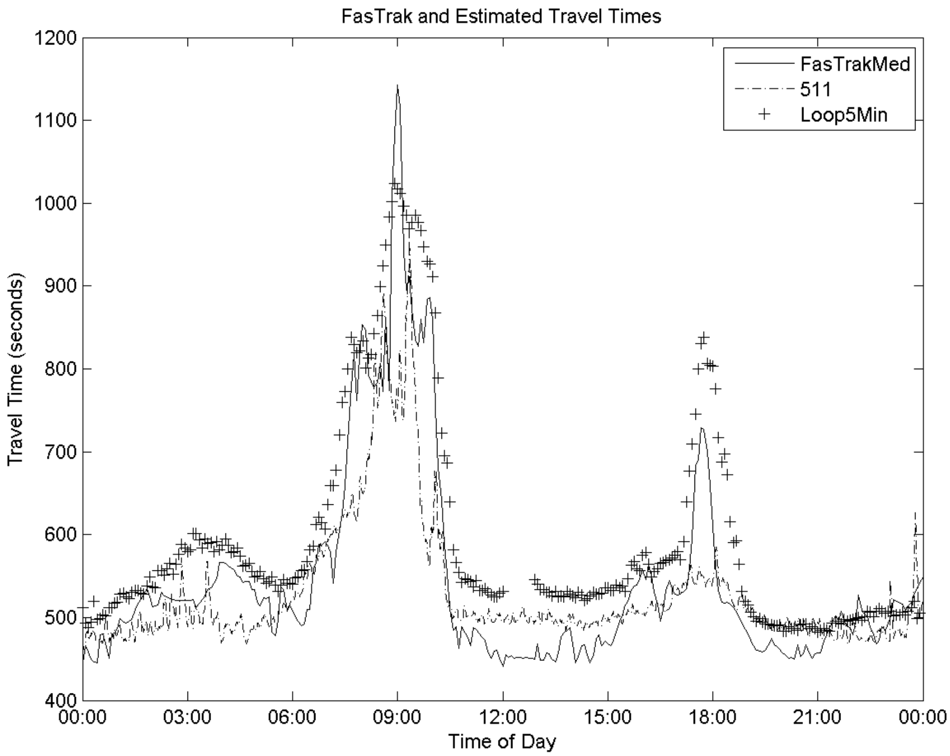

Figure 7 - Comparison of Ground Truth Travel Times with Estimates From Local 511 and Loop Detectors

In order to calculate the benchmark metrics, we first compute (via equation (1)) the individual relative errors induced by each estimate for every toll tag measurement. Scatter plots of those errors are presented on Figure 8. Individual errors establish themselves in a range spanning up to 50%, with most values remaining under 20% for 511 estimates, and under 30% for loop estimates.

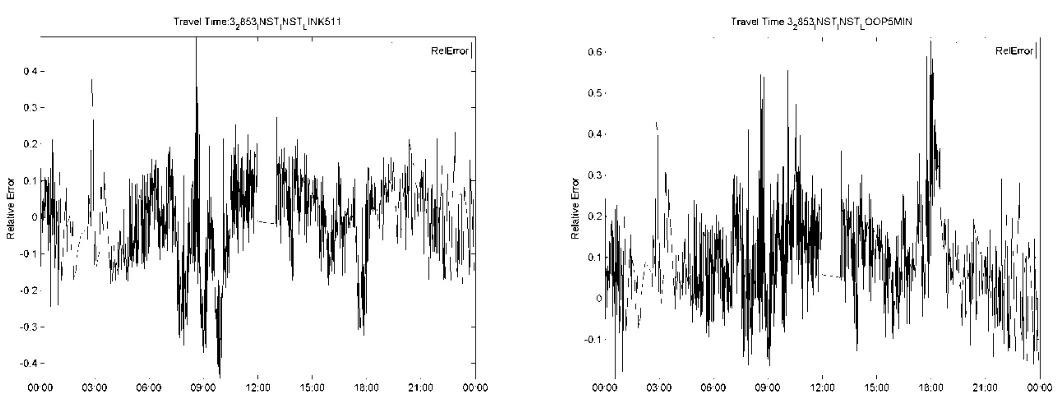

Figure 8 - Individual Relative Errors for Estimates from 511 and Loop Detectors

Benchmark metrics are then calculated for five distinct time periods spanning the entire day and reflective of the daily travel time variations. These are as follows:

1. Early morning, off-peak, from 12am to 7am;
2. Morning peak period, from 7am to 10am;
3. Mid-day period, from 10am to 3pm;
4. Afternoon peak from 3pm to 7pm;
5. Evening, from 7pm to 12am.

Table 3 displays the aggregated error and the 15%-relevance (R15) metrics for both the 511 estimates and the estimates from loops over each time period. This reveals that the 511 estimates are relatively unbiased: the aggregated error is less than 5% for all time periods but the morning peak, during which it underestimates actual travel times by an average of 11%. The loop data doesn't perform as well and overestimates travel times by more than 10% for most of the day. Yet the relevance scores are all above 90% for both set of estimates during the entire day. Only at times when the error increases does the relevance drop below 95%, as is the case for the loop data during the mid-day period (92%). It is interesting to note two facts. First, while the aggregated error and the relevance are correlated, they are not identical: a greater error can still result in a higher relevance. This is essentially indicative of travel time dispersion: at times when the spread is greater, even an unbiased estimate will miss the mark for a greater number of drivers; at times when traffic is homogeneous, a slightly off estimate may still be good enough for most people. Second, and as a consequence of the first point, the relevance measure appears less discriminating. While the 511 estimates are clearly better than loops in terms of accuracy, their superiority is only marginal according to the relevance measure. It will be interesting to accumulate more data for more days and routes and observe how this plays out in general. In particular, we will determine which threshold usually strikes the best balance between discriminating power (lower values) and sensibility (no one expects travel time estimates to be within one minute all the time).

Data Source	Period	AggError	Relevance – R15
511	AM off-peak (12am – 7am)	0%	98%
	AM peak (7am – 10am)	-11%	96%
	Mid-day (10am – 3pm)	4%	98%
	PM peak (3pm – 7pm)	-2%	98%
	PM off-peak (7pm – 12am)	2%	97%
Loop5Min	AM off-peak (12am – 7am)	8%	98%
	AM peak (7am – 10am)	10%	97%
	Mid-day (10am – 3pm)	16%	92%
	PM peak (3pm – 7pm)	14%	96%
	PM off-peak (7pm – 12am)	4%	97%

Table 3 - Benchmark Metrics for Five Selected Time Periods

5 CONCLUSION

At a time when the business of collecting precise traffic information is becoming more normalized and professionalized due to the growing success of in-car navigation systems and the availability of new data collection techniques, benchmarking the quality of travel time estimates is becoming a market necessity. This paper offers a methodology and associated metrics to conduct benchmarks of travel time estimation schemes. The methodology assumes that individual travel times can be collected continuously on a fairly large scale. While this is a strong assumption, it is also becoming more and more realistic with the development of various means to anonymously track vehicles, such as toll tag readers, license plate cameras and GPS receivers. Of the two proposed metrics, aggregated error mostly refers to bias and accuracy, while relevance examines the extent to which a given estimate captures the driving time of a majority of vehicles. It is understood that the latter depends strongly on the dispersion of travel times on a given route, regardless of how accurate the estimate may be.

Applying the methodology and metrics on one route in the San Francisco Bay Area during a 24-hour period shows that estimates produced by the local 511 program are more accurate than those produced by data from loop detectors. They also are slightly more “relevant” as a result, though the effect is less clear. Overall, the 511 estimates show little or no bias when considered as a whole over 5 selected time periods that cover the entire day. 15%-relevance is typically above 95% except for one time period during which it is 92%. Those figures indicate which proportion of drivers traveling the studied route during the various time periods considered are provided with travel time estimates that end up matching their actual driving time within a 15% margin.

While this paper has the ambition to push for a widely accepted standard in how travel time estimation accuracy is measured, it by no means pretends to have achieved that feat. In other words, the authors wish not only to accumulate far greater amounts of data to determine how well the proposed metrics capture the differences between various types of estimates, but also to engage the ITS research community and the industry in becoming interested in the question they pose. Our hope is to receive numerous suggestions and to generate a fruitful discussion in order to elevate the quality and objectivity of the debates that revolve around traffic data quality. Clarity and transparency are crucially needed if we want to fully develop traveler information services.

ACKNOWLEDGMENTS

The authors thank, first and foremost, the CCIT staff members and graduate students who have contributed to this project: Kossi Adjaka, Bensen Chiou, Kenneth Kuhn, Florent Robineau, Paul Supawanich and Samuel Yang; the California Department of Transportation (Caltrans), who is sponsoring the effort and providing local support; the 511 team at the Metropolitan Transportation Commission, without whom this benchmark would remain a thought experiment; and Professor Pravin Varaiya of the EECS department at UC Berkeley, who first advocated the use of anonymous toll tag data for research and development purposes. We also extend our gratitude to the organizers of the 14th World Congress on Intelligent Transportation Systems, which took place in Beijing, China, in November, 2007, for selecting a preliminary version of this paper for presentation.

REFERENCES

1. Al-Deek, H., and Emam, E. (2006). A new methodology for estimating transportation network reliability. *Journal of Intelligent Transportation System*, 10(3), 117-129.

2. Ban, X., Li, Y., Skabardonis, A. (2007) Performance evaluation for travel time estimation methods, In *Proceedings of the 11th World Conference on Transport Research (WCTR)*.

3. California Center for Innovative Transportation (CCIT). *Travel Times on Changeable Message Sing in District 4*. Final Report of CCIT Task Order 13, 2006. Avaiable http://www.calccit.org/resources/2006%20PDF/CCIT%20TO%2013%20-%20Final%20Report.pdf

4. Chen, A., Yang, H., Lo, H.K., and Tang, W. (1999) A capacity related reliability for transportation networks. *Journal of Advanced Transportation*, 33, 183-200.

5. Chen, C., Skabardonis, A., and Varaiya, P. (2004) A system for displaying travel times on changeable message signs. In *Proceedings the 83rd Transportation Research Board Annual Meeting (CD-ROM),* Washington, DC.

6. Dance, F., Gawley, D., Hein, R., and Kates, R. (2007) Enhancing navigation systems with quality controlled traffic data. SAE International.

7. Fisher, K. et al. (2003) Performance Measures of Operational Effectiveness for Highway Segments and Systems. NCHRP Synthesis 311.

8. Fujito, I., Margiotta, R., Huang, W., and Perez, W.A. (2006). The Effect of Sensor Spacing on Performance Measure Calculations. *In Proceedings of the 85th Annual Meeting of Transportation Research Board (CD-ROM)*, Washington, DC.

9. Hartley, J.K. (2003) Prediction of link travel times in the context of Nottingham's urban road network. In Proceedings of the 17th European Simulation Multi-Conference, 423-428.

10. Hinsbergen, C.P.IJ, Lint, J.W.C., and Sanders, F.M. (2007). Short-term prediction models. In *Proceedings of the 14th world congress on ITS (CD-ROM)*.

11. Huisken, G., and M.v. Maarseveen (2000). Congestion prediction on motorways: a comparative analysis. In *Proceedings of the 7th world congress on ITS (CD-ROM)*.

12. Khattak, A., Kanafani, A., and Colletter, E.L. (1994) Stated and reported route diversion behavior: implications of benefits of advanced traveler information system. *Journal of Transportation Research Board* 1464, 28-35.

13. Kwon, J., McCullough, B., Petty, K. and Varaiya, P. (2006) Evaluation of PeMS to Improve the Congestion Monitoring Program. Final Report for PATH TO 5319.

14. Lindveld, C.D.R., Thijs, R., Bovy, P.H.L, and Van der Zijpp, N.J. (2000) Evaluation of online travel time estimators and predictors. *Journal of Transportation Research Board* 1719, 45-53.

15. Liu, X.H., He, X.Z., Recker, W. (2007) Estimation of the time-dependency of values of travel time and its reliability from loop detector data. *Transportation Research, Part B*, 41(4), 448-461.

16. Oda T. (1990). An algorithm for prediction of travel time using vehicle sensor data. *In Proceedings of Third International Conference on Road Traffic Control*, 40 - 44.

17. Peng, Z.R., Guequierre, N., and Blakeman, J.C. (2004) Motorist response to arterial variable message signs. *Journal of Transportation Research Board*, 1899, 55-63.

18. Rice, J., and Zwet, E.V. (2001) A simple and effective method for predicting travel times on freeways. In *Proceedings of IEEE Intelligent Transportation Systems*, 227-232, Oakland, CA.

19. Smith, B.L., and M.J. Demetsky (1997). Traffic flow forecasting: comparison of modeling approaches. *Journal of Transportation Engineering*, 123(4), 261-266.

20. Zhang, X.Y., Rice, J., and Bickel, P. (1999) *Empirical Comparison of Travel Time Estimation Methods*. California Path Research Report, UCB-ITS-PRR-99-43.