STATE OF CALIFORNIA • DEPARTMENT OF TRANSPORTATION

TECHNICAL REPORT DOCUMENTATION PAGE

TR0003 (REV 10/98)

1. REPORT NUMBER	2. GOVERNMENT ASSOCIATION NUMBER	3. RECIPIENT'S CATALOG NUMBER
CA10-1127		

4. TITLE AND SUBTITLE

Statewide Travel Times - Business Case & Technology Deployment

5. REPORT DATE

February 2010

6. PERFORMING ORGANIZATION CODE

7. AUTHOR

Samer Madanat

8. PERFORMING ORGANIZATION REPORT NO.

N/A

9. PERFORMING ORGANIZATION NAME AND ADDRESS

California Center for Innovative Transportation
2105 Bancroft Way
Berkeley, CA 94720

10. WORK UNIT NUMBER

11. CONTRACT OR GRANT NUMBER

65A0212

12. SPONSORING AGENCY AND ADDRESS

California Department of Transportation
Division of Research & Innovation, MS 83
P.O. Box 942873
Sacramento, CA 94273-0001

13. TYPE OF REPORT AND PERIOD COVERED

Final Report:
6/30/2005 to 6/30/2008

14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

16. ABSTRACT

Travel times on preset highway itineraries are one of the most tangible outputs of Intelligent Transportation Systems (ITS). Travel times constitute an effective metric to describe level of service on roadways. It is a fine indicator of congestion and is well understood by the traveling public. Under this research the current capabilities of Caltrans to collect accurate travel time estimates on the highway network, and to identify the technologies and business models that could most effectively enhance these capabilities were assessed.

The results of this report, includes the following elements: A look at the value of travel time information, which is clearly associated with the quality of the underlying data; A survey of technologies available to collect traffic data; A series of systematic studies to characterize the quality of travel time information and evaluate its range;An industry level effort conducted through the North American Traffic Working Group (NATWG) to harmonize data quality benchmarking practices. This report investigates point detection, probe data, and segment detection. It concludes that a fusion of the data provided by these technologies would provide the best results in terms of travel time accuracy, reliability, and relevance. Data collected from cell phones and other connected device is limited to velocity information. Operators have traditionally relied on detection systems that provide measures of traffic flow and occupancy rate. It is unlikely and probably not even desirable that phone originating data will

17. KEY WORDS	18. DISTRIBUTION STATEMENT
Travel times, ITS, metric, congestion, highway network, accurate, NATWG, data, cell phones, connected device, travel time information	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161

19. SECURITY CLASSIFICATION *(of this report)*	20. NUMBER OF PAGES	21. COST OF REPORT CHARGED
Unclassified	153	

DISCLAIMER STATEMENT

FINAL REPORT

TASK ORDER 1016

STATEWIDE TRAVEL TIMES

FEBRUARY 2010

Prepared for:

CALTRANS DIVISION OF RESEARCH AND INNOVATION

CALTRANS DIVISION OF TRAFFIC OPERATIONS

Authors:

J.D. Margulici, *Associate Director*, CCIT

Jeff Ban, *Research Engineer*, CCIT

Kenneth Kuhn, *Post-Doctoral Researcher*, CCIT

Xiaohong Pan, *Development Engineer*, CCIT

Paul Supawanich, *Graduate Student Researcher*, CCIT

CALIFORNIA CENTER FOR INNOVATIVE TRANSPORTATION

UNIVERSITY OF CALIFORNIA BERKELEY · 2105 BANCROFT WAY, SUITE 300 · BERKELEY, CA 94720-3830

PHONE: (510) 642-4522 · FAX: (510) 642-0910 · HTTP://WWW.CALCCIT.ORG

PROJECT SUMMARY

Title: Statewide Travel Times

Sponsor: Caltrans Division of Traffic Operations

Project Organization: California Center for Innovative Transportation
2105 Bancroft Way
Berkeley, CA 94720
Phone: (510) 642-4522 – Fax: (510) 642-0910

Principal Investigator: Samer Madanat, Director, Institute of Transportation Studies

Project Manager: JD Margulici, Associate Director, CCIT
(510) 642-5929 – jd@calccit.org

Additional Investigators: Jeff Ban, *Research Engineer*, CCIT
Kenneth Kuhn, *Post-Doctoral Researcher*, CCIT
Xiaohong Pan, *Development Engineer*, CCIT
Paul Supawanich, *Graduate Student Researcher*, CCIT

Center Director: Thomas West

Administrative Officer: Coralie Claudel
(510) 642-5579 – coralie.claudel@bekeley.edu

Performance Period: June 2006 – September 2009

Project Cost: $205,000

CONTENTS

EXECUTIVE SUMMARY

CCIT task order 1016 was sponsored by Caltrans in the wake of a 2005 workshop on the collection and dissemination of roadway travel times held in San Diego, Cal. The goal of this task order was to assess the current capabilities of Caltrans to collect accurate travel time estimates on the highway network, and to identify the technologies and business models that could most effectively enhance these capabilities.

Travel times on preset highway itineraries are one of the most tangible outputs of Intelligent Transportation Systems (ITS). Travel times constitute an effective metric to describe level of service on roadways. It is a fine indicator of congestion and is well understood by the traveling public. Traditionally, traffic data is collected by sensors installed at fixed locations. While this method yields great results to estimate volume and occupancy, it does not allow deriving accurate travel time information unless the sensor coverage is very dense. Traditional traffic sensors are expensive to install and maintain, which prevents denser coverage. Over the past several years, a number of private industry vendors have approached Caltrans with solutions to collect travel time data on highways and city arterials. In this respect, cell-phone based technology is clearly the most promising avenue.

During the three-year performance period of this task order, there have been significant technological and competitive shifts in the traffic information industry. As a result, questions posed by practitioners today are different from those they asked at the time that task order 1016 was funded. Back in 2006, there was a lot of interest in companies such as AirSage that processed information collected from cell phone networks to derive traffic information. Today, this approach seems too onerous and there is a lot more focus on aggregating data from cell phones equipped with GPS receivers that can transmit their own position and speed. The project team followed the evolutions in the technology and business models, and adjusted the scope of work accordingly. Because these evolutions are ongoing, it has appeared more useful to focus on a general decision framework than to attempt to formulate specific recommendations.

The scope of the task order, and thus of this report, includes the following elements:

- A look at the value of travel time information, which is clearly associated with the quality of the underlying data;
- A survey of technologies available to collect traffic data;
- The project team undertook a series of systematic studies to characterize the quality of travel time information and evaluate its range;
- CCIT engaged in an industry-level effort conducted through the North American Traffic Working Group (NATWG) to harmonize data quality benchmarking practices.

This report investigates point detection, probe data, and segment detection. It concludes that a fusion of the data provided by these technologies would provide the best results in terms of travel time accuracy, reliability, and relevance. This type of fusion would be a planned combination of multiple layers of coverage (segment with radar, probe data with loop, etc). Sensor spacing between the various means of detection technology remains somewhat of an open question though it has received more attention of late. However, it is clear that the main source of new information is a massive influx of cost-effective probe data from cell phones equipped with GPS.

The massive availability of traffic data at a lower cost will have important consequences for both the traveling public and roadway operators. The dissemination of traffic information will enable a form of system 'self-management', in which individual commuters can make informed travel decisions. Not only will each user benefit personally, but the entire driving community will enjoy more balanced loads across the road network. Roadway operators will also benefit tremendously from collecting rich performance data that can be turned into a wide range of strategies, from demand management to dynamic traffic control.

However, the data collected from cell phones and other connected devices is limited to velocity information. Operators have traditionally relied on detection systems that provide measures of traffic flow and occupancy rate. It is unlikely and probably not even desirable that phone-originating data will supplant existing detection systems at any point in the foreseeable future. On the other hand, there is a definite possibility to use mobile probe data as a complement to existing detectors, resulting in a so-called 'hybrid' traffic data collection system. In such a setup offering ubiquitous availability of speed information, traditional traffic monitoring stations (TMS) will only be needed to maintain accurate flow counts. Larger spacing intervals could be allowed in-between stations, and therefore equipment could be deployed much more sparingly than it is today. In sum, operators such as Caltrans could be getting a lot more data for a lot less money.

A recently approved technical agreement named 'Hybrid Traffic Data Collection Roadmap' intends to investigate the feasibility and the business case for such a hybrid traffic data collection system. One very important feature of such a system is that it involves not just new technology but also a completely different business process. In today's environment, Caltrans purchases detection technology than it then installs, operates and maintains internally, even if contractors may be used. In the case of mobile probe data from GPS-equipped cell phones, Caltrans would have to procure data collected and processed by a third-party, be it a cellular network operator or a service provider. This is a fundamental shift that poses its own set of challenges, irrespective of the technical merits of a hybrid system.

As a result, the project team has been active within The North American Traffic Working Group (NATWG). Hosted by the Intelligent Transportation Society of America, NATWG works collaboratively to define, accept and advocate for the unique needs of North America traffic information services. Specifically, we have been leading a NATWG task force that aims to establish a unified methodology for traffic information benchmarking. Benchmarking traffic information with clear and consistent metrics is an essential requirement for the development of data services that may complement the data collected by states through their own infrastructure. The primary product of this effort is tentative guidelines on how to benchmark traffic data quality.

In parallel, the project team has pursued its own efforts to better understand the drivers of traffic estimation quality. This has resulted in three research papers that are part of this task order's deliverable. The papers advance the conversation on adequate metrics and sampling, the most two pressing issues regarding traffic information benchmarking.

INTRODUCTION

BACKGROUND

Travel times on preset highway itineraries are one of the most tangible outputs of Intelligent Transportation Systems (ITS). Travel times constitute an effective metric to describe level of service on roadways. It is a fine indicator of congestion and is well understood by the traveling public.

Traditionally, traffic data is collected by sensors installed at fixed locations. While this method yields great results to estimate volume and occupancy, it does not allow deriving accurate travel time information unless the sensor coverage is very dense. Traditional traffic sensors are expensive to install and maintain, which prevents denser coverage.

Over the past several years, a number of private industry vendors have approached Caltrans with solutions to collect travel time data on highways and city arterials. Cell-phone based technology is presented as the most promising avenue, although not all previous research and field tests have been conclusive.

The emergence of new technology that can provide accurate travel time data a lot cheaper than what was possible only a few years back is receiving due attention. Although travel times alone may not cover the full extent of the Caltrans' traffic data needs, accurate and reliable travel times could be used for both planning and operations purposes. Additionally, as the state's Department of Transportation, Caltrans is looking after the interests of the traveling public, the transportation industry, and the economy as a whole in California. Therefore, Caltrans is and will remain instrumental in facilitating the collection and diffusion of travel time data in California.

In December, 2005, the first statewide forum on travel time techniques and economics hosted by CCIT in San Diego was an opportunity to reflect on what the industry has to offer. The forum also provided the occasion for a call to action by Caltrans management with two messages. First, more ubiquitous and reliable data is urgently needed to manage the transportation system, and, second, Caltrans should intensify partnerships with the private sector in order to leverage travel information demand and benefit from the latest technological advances.

Around the same time, the GoCalifornia initiative provided the appropriate platform to enact the changes that were envisioned by Caltrans management. The following items were identified as critical to these changes:

- Collaboration with regional agencies
- Modalities of Public-Private Partnerships
- Technology decisions
- Coverage expansion plans

The initiative also provided the impetus for funding CCIT Task Order 1016. The intent of the task order was to assess the current capabilities of Caltrans to collect accurate travel time estimates on the highway network, and to identify the technologies and business models that could most effectively enhance these capabilities. In parallel, Caltrans received approval to conduct a pilot procurement of information from “virtual traffic monitoring stations (VTMS)” that would be derived from cell phones. Several factors contributed to slight changes in the scope of the project, which is outlined in the next section.

PROJECT SCOPE

Over the past three years, factors ranging from administrative issues both within Caltrans and between the University of California and Caltrans, to technological and competitive shifts in the industry have influenced the focus of this task order:

a. The project start was pushed back by administrative delays;
b. Technology has kept evolving since 2006, to the point where it has appeared more useful to focus on a general decision framework than to attempt to formulate specific recommendations. In particular, the first evaluations coming from pilot projects conducted in other states with network-based (i.e. non-GPS) cell phone data showed some clear shortcomings and Caltrans itself redirected the objectives of its VTMS project;
c. To echo the previous point, CCIT initiated a relationship with Nokia in late 2007 (who later acquired NAVTEQ) to study the collection of traffic data from GPS-equipped cell phones, first through the Mobile Century field test, and then with the Mobile Millennium pilot project;
d. District-based initiatives to broadcast travel time messages on Caltrans’ Changeable Message Signs (CMS) in urban areas were more or less complete by the end of 2006. Although coverage gaps remained, these are gradually being filled with traditional sensor technologies;
e. Caltrans did experiment with a few forms of public-private partnerships for data collection, in particular by taking advantage of the Federal ITIP program awarded to Traffic.com. As a result, a portion of traffic detectors in the main urban areas are owned and operated by this company (now part of NAVTEQ, and thus a subsidiary of Nokia);
f. The institutional map also evolved during the period of performance of the project. Caltrans, for one, adopted a more formal position with regards to traveler information, considering that it ought to be a facilitator but not an operator in that space. This somewhat diminished the interest for travel times, though the information is used by Caltrans practitioners as a performance measure. Besides, there were shifts in the industry, with winners and losers, and thus a new set of companies to engage with. Providers of traffic information, most prominently NAVTEQ and Inrix, now offer integrated data feeds that complement publicly available data with fleet data and, increasingly, cell-phone-originating data;

An important consequence of some of these points is that Caltrans can now examine a much more radical vision when it comes to collecting traffic information, both on the technology front and the business model front. As of this writing, the division of Traffic Operations is sponsoring CCIT to outline the modalities of a “hybrid data” business plan, which will consider the introduction of probe-originating traffic information delivered by third-party vendors into Caltrans’ traffic management systems. This represents a significantly larger effort than was first envisioned with the execution of task order 1016.

The actual scope of the task order, and thus of this report, includes the following elements:

A. A look at the value of travel time information, which is clearly associated with the quality of the underlying data;
B. A survey of technologies available to collect traffic data;

These two activities are covered in the main body of this final report. A consequence of these investigations has been to direct a very large portion of further work to the question of data quality. Data quality is central to both technology decisions and business model decisions, especially if procurement is involved. Thus, the additional two activities were conducted:

C. The project team undertook a series of systematic studies to characterize the quality of travel time information and evaluate its range;
D. CCIT engaged in an industry-level effort conducted through the North American Traffic Working Group (NATWG) to harmonize data quality benchmarking practices.

Activity C resulted primarily in three research papers that are included in this report as Appendix III. Activity D is ongoing and its current state is presented in Appendix IV. It includes a slide presentation that describes the nature and composition of NATWG, as well as the scope of the task that the project team has been leading as a part of the group. The primary product of this effort is tentative guidelines on how to benchmark traffic data quality. The strength of these guidelines is that they are being developed through a committee that is highly representative of the main industry players in the traffic information industry. The most recent draft guidelines are presented in Appendix IV.

Note that the main body of the report (that is, excluding Appendices III and IV) was primarily written in 2006 and 2007. As a result, some of its content appears outdated. Where the accuracy of the writing has been compromised by time, corrections have been made. In the case of gaps and omissions resulting from new developments that have taken place since then, references to more recent work have been provided.

1 THE VALUE OF TRAVEL TIME INFORMATION

1.1 OVERVIEW

Accurate, ubiquitous and timely traffic data is the foundation of system-level management for roadway networks. System management is a paradigm that considers the components of the transportation system as a whole rather than as isolated pieces. In other words, system management regards efficient mobility of people and goods as the outcome of interactions between transportation demand, the physical infrastructure, traffic regulations and control, and the availability of public transportation, among other factors. In the long-run, the system is also shaped by land use policies and investments in network capacity. Over the past decade, Caltrans has introduced system management thinking in some of its planning and operational decisions. Figure 1 illustrates the system management paradigm and highlights the foundational role played by traffic information.

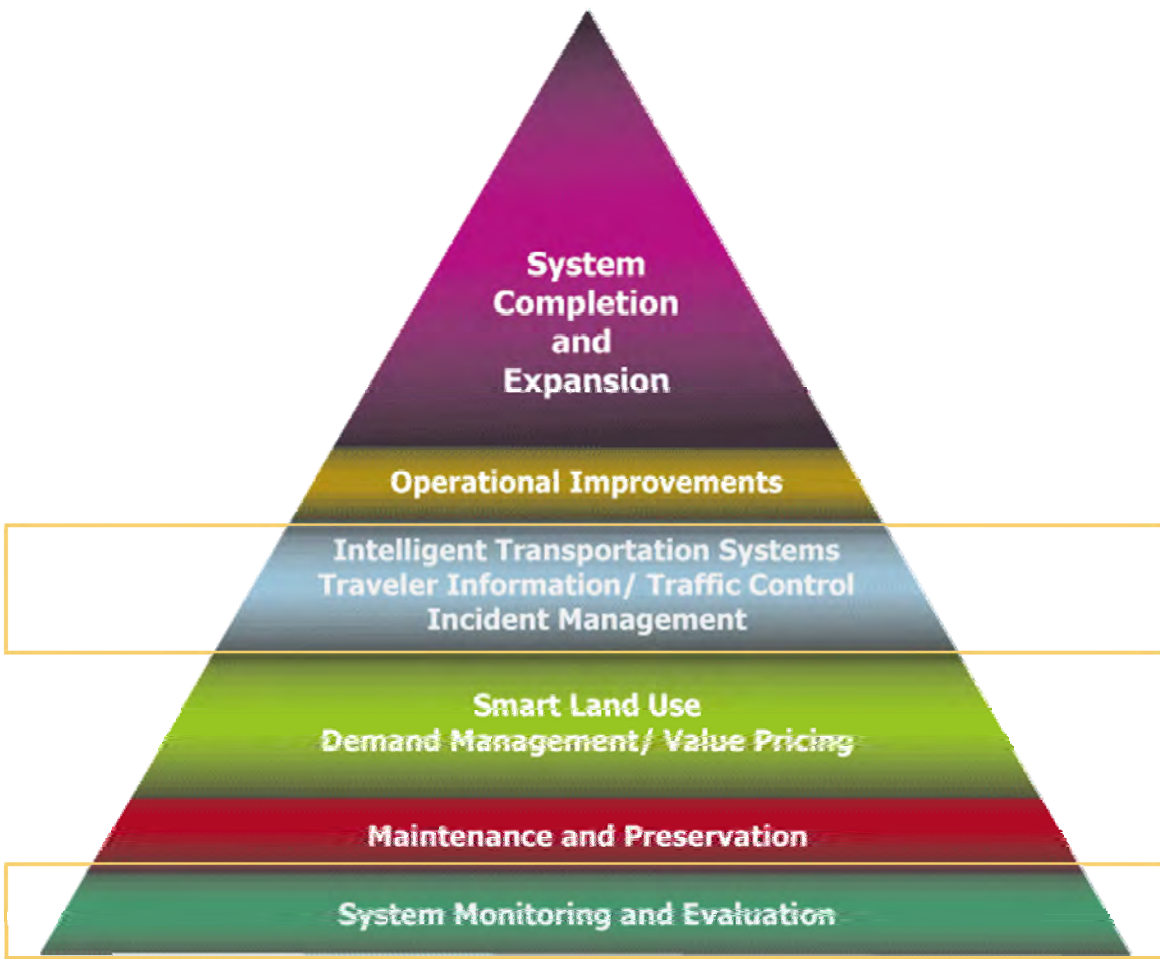


Figure 1 - Caltrans' pyramid illustrating system management thinking, with highlights on the critical role of traffic information

Travel times are only one piece of information regarding both historical and real-time traffic, but they are a very essential one because they capture congestion and the ensuing level of service to motorists, and are easy for drivers to interpret. In effect, travel times are used towards the following purposes:

- TRAVELER INFORMATION: travel times can be posted on Changeable Message Signs. They can also be disseminated by 511 service, over the phone or on the internet. With electronic mobile devices becoming ubiquitous, commuters are able to receive travel time information at any time in a variety of formats. Travel time information empowers drivers to make decisions that are right for them, both in the short term and the long term. This results in a form of system self-management that can be very effective to reduce congestion. Information also alleviates commuters' stress and demonstrates Caltrans' commitment to providing a high level of service to state residents.
- PERFORMANCE MEASUREMENT: as Caltrans describes itself as a mobility agency, travel times become the de facto #1 performance measure. Obtaining ubiquitous and accurate travel time data will bring practitioners the information they need to effectively manage the highway system and assess operational improvements.
- TRAFFIC CONTROL: on short section of freeways, travel time information is roughly equivalent to speed information. If provided with reliable data, operators can use travel times to monitor traffic and make control decisions.
- SYSTEM PLANNING: planners need large amounts of reliable data in order to effectively identify priorities. Reliable historical travel times can provide the bulk of this data by clearly showing which routes or areas are the most impacted by congestion and when.

Travel times are collected primarily in dense urban areas. Providing accurate travel time estimates in rural areas has been so far difficult because the sensors coverage is not dense enough. While travel times do not vary on the vast majority of rural roads, there are some exceptions such as routes leading to ski resorts on weekends. In Task Order 1023, CCIT also argues in favor of portable systems than can be used for construction zones, or in case of natural disasters or very severe accidents.

1.2 USER BENEFITS

The collection of travel times require both hardware and software technologies that come at a cost to Caltrans and other public agencies. In order to justify that cost and the need for further investments, it would be ideal if the resulting value could be measured in dollar terms. As one can well imagine, this is not an easy proposition, and every attempt at doing so requires assumptions that become easy targets for naysayers. The truth of the matter is that much of the value is intangible, and the tangible outcomes are extremely hard to estimate. In the following sections, we review the benefits of travel times to society. Of course, the value of information is tightly coupled with its overall quality, in this particular case accuracy and timeliness. Data quality gets addressed in the following section.

1.2.1 THE VALUE OF TRAVEL TIME INFORMATION FOR PRIVATE HOUSEHOLDS

One way to estimate the value of travel time information is to study how much people are willing to pay for this information. Estimate a demand curve for travel time information, indicating how many consumers would subscribe to a service at different price levels. It is here assumed that travel time information would be a monthly subscription service and that the consumers would be private households. The integral of the demand curve provides the value of travel time information, as in the diagram below.

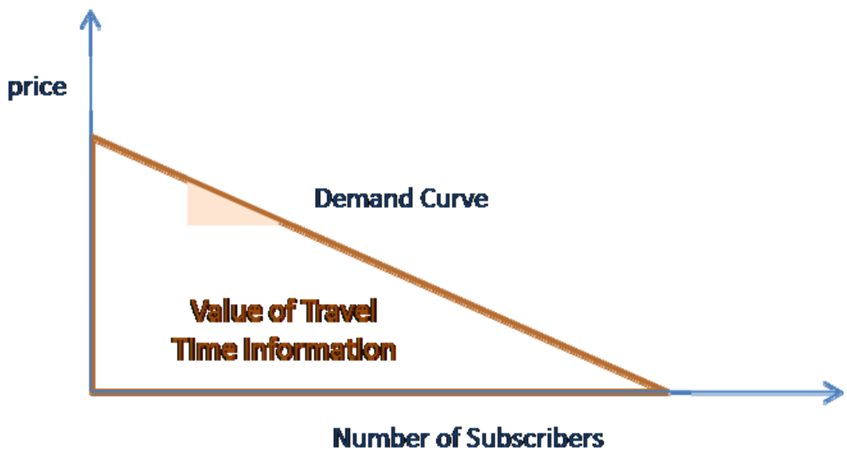


Figure 2 - Demand curve and consumer value of travel time information

Estimating the demand curve for travel time information is a tricky proposition for a number of reasons. First of all, much of the travel time data available today is distributed without charge. This makes it very difficult to perform a "revealed preference" study that would show how much people are currently paying for travel time information. Worse still, when asked, people may be unable or unwilling to estimate how much they are willing to pay for information they now get for free. This is known as cognitive dissonance bias (Polydoropoulou et al., 1996). When asked about the value of travel time information, people may also have a tendency to overestimate either in the hopes of encouraging public policy decision-making to their benefit (policy response bias) and/or because they do not have any real money at stake (non-commitment bias) (ibid.).

Finally and perhaps most importantly, travel time information is not a clearly defined market good. Different services that provide such information present this information very differently. Some services maintain websites users can navigate to find current travel time estimates on different roadways. Other services provide displays users put in their cars and update the maps shown on these displays, with highway network links color coded according to current driving conditions (free-flow, congested, or something in-between). Personal preferences are important in determining how different consumers value different presentations of travel time information.

Notwithstanding the problems listed above, an effort is made here to very roughly estimate the market value of travel time information. Let's begin by noting that in the year 2000, Census figures reported

California had around 11.5 million households[1]. Let us assume that only urban households are interested in obtaining travel time information on a regular basis. 94 percent of California's population lives in urban areas[2], and the proportion of households is roughly the same. That implies a potential market of 10.8 million households.

A 1998 Broad Area study asked 1,000 representatives of San Francisco Bay Area households to define their willingness-to-pay for traveler information systems that included travel time information, notifications of unexpected congestion, estimated delay times, and alternate route planning (Wlinetz et al., 2001). 185 respondents were willing to pay $7 a month for such a system, an additional 61 would pay $5, and an additional 81 would pay $3. Assuming these responses were typical of urban California households, that translates into a value for a similar statewide system of $20 million / month. (10.8 million * 185/1000 * $7 / month + 10.8 million * 61/1000 * $5 / month + 10.8 million * 81/1000 * $3 / month = $20 million / month)

This is likely a large overestimate of the amount of revenue traveler information services would generate. Such services would set a few fixed prices for information provided, but here we are counting every consumer's willingness to pay. This can be interpreted as measuring the value, to society, of providing travel time information to all free of charge. It can also be interpreted as measuring the maximum revenue commercial services could collect, if all consumers were differentiated and asked to pay their exact willingness to pay. Although $20 million / month is a large quantity, it may actually be a slight underestimate of this maximum possible revenue since those willing to pay $3, $5, or $7 may be willing to pay a bit more but this is not captured in the analysis above.

A 1994 study asked 220 respondents in the Boston area to estimate their willingness-to-pay for a traveler information system, mainly based on travel time information, known as SmarTraveler they did not currently have access to. 7% were "very likely" or "somewhat likely" to be willing-to-pay $15 a month for the service, an additional 7% would likely pay $10, an additional 13% would pay $5, and an additional 14% would pay $2.50. Again, assuming these responses are typical of urban households in California, this translates into a potential market of $30 million / month here.

Based on the studies listed above, a very rough approximation of the value, to society, of a travel time information system for California would be $20 to $30 million / month. This figure only considers private household users of such a system, although clearly such a service would also be beneficial for commercial vehicle fleet managers.

---

[1] http://www.census.gov/prod/2001pubs/c2kbr01-8.pdf

[2] http://www.ccp.ucdavis.edu/pubs/pdf/QuickFacts.pdf

1.2.2 NON-MARKETABLE BENEFITS OF TRAVEL TIME INFORMATION

Much of the value of travel time information cannot be captured by the willingness of private households to pay for this data. Some of this additional value stems from the notion that a more informed driving public will more efficiently use the roadways. Reducing congestion will result in reduced vehicle emissions and reduced environmental impacts. The impacts to society, in terms of improved air quality, reduced numbers of asthma cases, etc., may be quite substantial. More efficient use of roadways may also reduce the number of new roads that need to be constructed. Finally, government planners will derive significant benefits from travel time information. Roadway system decision-makers will be able to study the performance of the roadway network more closely. This will help them identify critical components of the system and design improvements. Disaster mitigation and relief efforts will also be more successful given accurate, real-time data about travel times.

1.3 MEASURING INFORMATION QUALITY

It is now commonplace for government agencies to provide travel time estimates to the driving public. These agencies frequently post these estimates on roadside changeable message signs or provide these estimates to firms that repackage the information before offering it to the public. Appraising various methods of collecting traffic data and specifically travel times is a definite need for Caltrans. Besides providing the bulk of the content required for ATIS, travel times also represent precious data to Caltrans as a network operator. While travel times alone may not cover the full extent of the department's traffic data needs, accurate and reliable travel times can be used for both planning and operations purposes.

Travel time estimates are provided for particular links of a highway network, although individual estimates are often published in groups as a part of network-wide traveler information systems. Travel time estimates on individual links are typically built using data from individual fixed-location sensors. In order to maximize the benefits of travel time information systems, it is important to assess the quality of different travel time estimation schemes. Doing so requires the consideration of network-level effects, since travel time estimates are often used for comparing alternate routes or for planning trips across a roadway network. It is also important to consider the quality of data provided by individual sensors, since this will impact the quality of travel time estimates.

It is here assumed that the quality of travel time estimates is based on providing accurate and reliable estimates for a number of particular routes within a roadway network. Improving network coverage, or the accuracy or reliability of estimates, increases the value of a travel time information system. When individual traffic sensors are used to calculate estimates, the accuracy and reliability of these estimates are a function of the accuracy, reliability, and granularity of data provided by individual sensors. The terms listed above are described in greater detail in the following subsections. For each factor that is identified here, a metric is given that provides a way to judge given travel time estimation schemes. The metrics are chosen to be clear, with intuitive meanings, and to distinguish the differences between alternate travel time estimation schemes. Example requirements, using the metrics given, are also provided. These are examples of requirements government agencies could demand of travel time information systems under development.

The main section developed in this report was later completed by three research papers that are reproduced in Appendix III.

1.3.1 TRUE TRAVEL TIMES

At the outset, let's consider four freeway links, all in the San Francisco Bay Area. These links are:

- Westbound on I-580 across the Richmond – San Rafael Bridge
- Westbound on Route 92 across the San Mateo Bridge
- Southbound on Route 101 across the Golden Gate Bridge
- Northbound on I-880 from Milpitas to Fremont

These links were chosen because toll tag readings are available that give 'ground truth' measurements of true travel times across the links. The true travel times of cars equipped with toll tags that traversed these links on Tuesday September 5, 2006, Wednesday September 6, 2006, and Thursday September 7, 2006 were studied. Data regarding true travel times on the links and days specified is attached as Figures 1.1-1.12 in Appendix I[3]. Summary statistics are presented below in Table 1.

Table 1 - Summary statistics of true travel times used here

Route	Median Travel Time (minutes)	Low (25th percentile) Travel Time	High (75th percentile) Travel Time	Data points per day
Richmond – San Rafael Bridge	11.2	10.7	11.9	2,611
San Mateo Bridge	9.9	9.5	10.2	2,708
Golden Gate Bridge	3.3	3.2	3.5	6,401
Milpitas to Fremont	9.4	7.9	12.2	1,276

It is worth noting that the summary statistics shown above and the plots shown in Appendix I come from processing raw data provided by FasTrak. The raw data was flawed in the sense that it contained certain highly irregular values for travel times (in some cases, negative values). These outlying data points may have corresponded to errors in the data processing done by FasTrak, or may have reflected the fact that certain cars traveled in a highly irregular way (traveling in what was assumed to be the wrong direction, stopping for lunch, etc.). It seems logical that cars that stop or travel in an otherwise highly irregular manner should not be considered when coming up with travel time estimates for cars traveling in a more typical manner. Here, outlying data points were removed using two sweeps of the MAD algorithm.

[3] Appendix I contains figures 1.1 through 1.24, all referenced in the present subsection.

1.3.2 RELIABILITY AND ESTIMATION

Different drivers traveling the same route at the same time will arrive at their destination at different times. This inherent variability in travel times makes it difficult to identify which data points are so ‘irregular’ that they constitute ‘outlying data points’ that should not be considered during travel time estimation.

Consistency in driving times of different cars is known as travel time reliability and the degree to which it exists on a link depends upon the length of the link in question and a host of other factors related to the specifics of the link in question and the conditions associated with its observation. Figures 1.13 and 1.14 in Appendix I examine travel time reliability on a given route across given off-peak and peak hour times when (based on Figure 1.1) driving conditions appeared more or less stable (i.e. travel times appeared to be uncorrelated to time of day). The range of travel times that covers the middle 50, 75, and 90% of drivers’ actual times are identified. Notice that in Figure 1.13, the middle 50% of drivers take between 16 to 19 minutes to traverse the link in question. Any single estimate of travel time will thus be off at least a good 8% for a driver whose driving time is in the middle 50% of travel times. Similar difficulties are also unavoidable during off-peak times.

Figures 1.13 and 1.14 show the middle 50%, 75% and 90% of driving times on a particular route. A question arises as to which of these ranges of travel times are relevant to a discussion of travel time estimation. This is really a question that has to do with the ambiguity of the term travel time estimate. Is it an estimate of the time that a ‘normal’ driver traveling on will take to travel directly between two points? Or is it more of a model of the time that all drivers take to travel that same route? The first definition is more attractive from a practical point of view. We can never hope to model the driving times of all drivers, and, given a single estimate, individual drivers can scale up or down based on their own knowledge of whether they travel slower or faster than normal. The second definition of a travel time estimate is more attractive from a theoretical point of view. Travel time estimates are available to all drivers and we should strive to make the information reasonably relevant to all.

Given the second point of view, and to a lesser degree even for the first point of view, it is impossible to divorce the notions of travel time reliability and the quality of travel time estimates. On a route where driving times are less reliable, it will be a bit more difficult to come with a ‘high quality’ travel time estimate that models the true driving time of a ‘normal’ driver and it will be much more difficult to come up with an estimate that is meaningful to a range of different drivers.

Let’s make a distinction here between accuracy and reliability. Let accuracy refer to the ability of a travel time estimate to predict typical travel time, while reliability measures how well an estimate reflects the range of travel times experienced by a variety of drivers. These definitions may sound contradictory; judging travel time estimates by how well they model two separate things. However, the two things being judged are closely related. The driving time of a ‘typical’ driver often is the best single estimate for the range of driving times of the masses. Figures 1.15, 1.16, and 1.17 show the maximum relative errors associated with different travel time estimates assuming they are used to model the driving times of 50%, 75%, and 90% of all drivers, respectively. (Relative errors and this methodology will be explained further below.) Note that for all graphs, the maximum relative error is minimized given a travel time estimate of around 18 minutes. 18 minutes is the 'typical' travel time for this particular link, and is the best single estimate of the ranges of travel times experienced by different travelers. The flip side to the above argument is that the confidence with which we can identify the driving time of a typical driver (the statistical reliability of an estimated typical travel time) can be measured by looking at the range of times observed.

It seems that ambiguity in the meaning of the term travel time estimate is not especially important in practice. The definitions proposed in this paper aim to reflect the ambiguity in the term travel time estimate, by at once using such estimates as models of normal drive times and as representative of the ranges of drive times of all drivers. Using the definitions described above, the reliability of a travel time estimate is very much related to the underlying reliability of true travel times while accuracy is less so.

1.3.3 LINK TRAVEL TIME ESTIMATE ACCURACY

Numerous alternate definitions of typical and methodologies for measuring accuracy exist. Here we look at the relative error associated with the median travel time. Say that a travel time estimate τ exists to describe trips on a particular link during a particular time-frame. There are a number of probe vehicle trips on this link during this time-frame for which we have data. The actual travel time of vehicle i is recorded as ti. The analysis done here first selects the median of these travel times, and then computes relative error: $| (\text{median}(ti) - \tau) / \text{median}(ti) |$.

Emphasis is placed on the median travel time because this is the best estimate of what a typical or normal travel time is. Using any alternate baseline travel time, for example the mean travel time, would make the analysis more sensitive to variations in the travel times of drivers that drive slower or faster than most other drivers. Calculating relative error is common in analyses estimating accuracy. Relative error has a clear intuitive meaning, for instance a relative error of 0.12 would mean a travel time estimate is 12% off in its estimation of median drive time.

EXAMPLE REQUIREMENT: Over the course of any 24 hour period, the relative error associated with the median travel time must be less than 0.05.

Figures 1.18 and 1.19 demonstrate the sensitivity of the accuracy metric to changing travel time estimates, using the same true travel times as Figures 1.13 and 1.14. Driving conditions are more or less stable across the time frames being surveyed in each graph, with Figure 1.18 showing travel times during non-rush conditions at mid-day and Figure 1.19 considering morning rush conditions. Note that in each graph a most accurate travel time estimate exists for which our metric equals 0. There is a true median travel time associated with the data and a travel time estimate equal to this value is considered to be perfectly accurate. As the travel time estimates stray from median travel time, the relative error increases linearly.

1.3.4 LINK TRAVEL TIME ESTIMATE RELIABILITY

The accuracy of a travel time estimate, as defined above, considers only a single value for travel time when we know real world drivers experience a range of travel times. It is also important to calculate the range of error values associated with a particular travel time estimation scheme. For instance, if travel time estimates are way off or 'unreliable' for 60% of drivers, this will not show up in a measure of accuracy but will greatly reduce the amount of faith the traveling public places in travel time estimates. Here we propose using the 75th percentile of relative errors as a measure of reliability. This shows how large errors can be for the bulk of drivers.

Note that unlike accuracy as defined above, this definition of reliability places great weight on travel times that are especially fast or slow. This is unavoidable since we are seeking to measure how well a travel time estimate reflects a range of true travel times. However, this also makes it is especially important to have high quality data that truly reflects travel times of ordinary vehicles traveling directly between two points of interest.

EXAMPLE REQUIREMENT: Over the course of any 24 hour period, the 75th percentile of relative errors must be below 0.20.

Figures 1.20 and 1.21 demonstrate the sensitivity of the reliability metric to changing travel time estimates, using the same true travel times as Figures 1.18 and 1.19. Again non-rush (Figure 1.20) and morning rush (Figure 1.21) conditions are studied. Note that it is impossible to have perfectly reliable travel time estimate; our metric always exceeds 0. It is impossible for any single travel time estimate to capture the range of travel times experienced by different drivers. The minimum level of the metric possible, as well as the sensitivity of the metric to changing travel time estimates, is determined by the specific conditions of the link being investigated, the traffic during the time-frame of interest, and possibly the data collection schemes used during metric estimation. Note that although Figures 1.20 and 1.21 both make use of FakTrak data from the same segment of freeway, they reflect different time-frames of interest and are quite different.

1.3.5 LINK TRAVEL TIME ESTIMATE CONSISTENCY

Travel time estimates are useful because driving conditions vary widely, even for a single route observed on a single day. The timing of "rush" periods and the intensity of the congestion associated with such periods is difficult to predict. In order to be valuable travel time estimation schemes must be relatively accurate throughout rush and non-rush periods; they must be consistently accurate.

A metric is constructed here that aims to capture how consistently accurate different travel time estimation schemes are. Different routes will experience congestion at different times of day. Sporting events and major holidays can dramatically increase traffic at times when otherwise there would be no congestion. Travel time estimates benefit motorists by warning them at times when unexpected congestion appears. For these reasons, the analysis here does not pre-specify 'rush' and 'non-rush' times of day.

Instead, we split up the day into time windows and examine the accuracy, using the metric defined previously, of travel time estimates in the time windows. What results is a time-series of time estimate accuracy over the course of a day (or multiple days.) A metric for consistency could then be applied that looks at the higher values for our accuracy metric, corresponding to the times when travel time estimates were most inaccurate.

The discussion of accuracy and reliability above avoided specifying time windows of analysis, but this is impossible here. We are seeking to capture the performance of travel time estimation schemes across different conditions, conditions which are fleeting. It would be ideal if we could estimate the accuracy of travel time estimates at all points in time throughout a day to judge consistency. However, in order to make measures of accuracy meaningful, it is necessary to have observed several cars' true travel times and to compare these travel times to estimates. There is a tradeoff to be made in choosing the size of time windows to look at. Too large time windows make it difficult to judge how well travel time estimation schemes adapt to rapidly changing traffic circumstances. Too small time windows make individual measurements of accuracy less meaningful.

Whatever the choice of time window, witnessing traffic over an extended period of time, we are likely to run into windows of time where very few cars are on the road. Thus we are likely to run into periods of time where our accuracy metric is less meaningful in some sense. For this reason, it is not wise to use a consistency metric that looks at something like the maximum value of our accuracy metric across an entire day.

The metric proposed here to estimate consistency is 90th percentile of the relative errors associated with the median travel times of 5 minute windows. Figures 1.22 through 1.24 show the relative errors associated with the median travel times of 5 minute windows for 3 different travel time estimation schemes that were applied to cars traveling between Albany and Carquinez, California. Horizontal lines indicate the 90th percentile of the errors across the day, for each travel time estimation scheme.

Notice the high degree of noise in the plots. This is most likely a result of relatively few cars being observed in certain 5 minute windows, which yields noisy estimates of accuracy in these time windows. There are even a number of 'holes' in the plots shown, corresponding to time windows where no cars were observed and no data on accuracy exist. Despite these difficulties, the plots and the 90th percentile of the accuracy metric do draw attention to the times of day when the travel time estimation schemes were poorest: evening rush.

There are no universal answers to the questions of setting the length of time window to investigate or the percentile of the accuracy metric to study. The amount of data available should influence both decisions. With more data, it becomes possible to more precisely characterize the accuracy of a travel time estimation scheme over the course of a day, and how this accuracy is related to time of day. What is important is to realize that travel time estimates must be consistently accurate. Consistency, over time, is worthy of consideration alongside more traditional notions of overall accuracy and reliability.

1.3.6 NETWORK COVERAGE

The value of a travel time information system depends upon the degree to which the system in question covers routes drivers use. For instance, travel time information is often used for route planning. In order to compare multiple routes, it is necessary to have travel time information for all of these routes. The clearest possible metric for network coverage would be the number of links for which travel time estimates are available.

EXAMPLE REQUIREMENT: Travel time estimates must be available for 21 of the 25 links in the network.

1.3.7 SINGLE DETECTOR DATA ACCURACY

Accuracy is a concern for measuring the performance of individual traffic sensors, as it is for measuring travel time estimates. Unlike travel time estimates, it is expected that the error associated with most measurements of individual traffic sensors approach zero. In this case, mean relative error is the most logical way to measure accuracy. Detectors may report data related to individual vehicle trips, or may aggregate data associated with multiple vehicle trips. Either way, mean relative error can be used and has a simple meaning; readings from the sensor are, on average, x% off.

EXAMPLE REQUIREMENT: Over the course of any 24 hour period, the mean relative error of time-mean speeds reported by the sensor must be less than 0.05.

1.3.8 SINGLE DETECTOR DATA RELIABILITY

Again it is important to consider estimate reliability as well as accuracy. If data from a particular detector is, even occasionally, way off, it may reduce the accuracy or reliability of link travel time estimates. As for such estimates, let's examine the 95th percentile of relative errors as a metric for reliability.

EXAMPLE REQUIREMENT: Over the course of any 24 hour period, the 95th percentile of relative errors of time-mean speeds reported by the sensor must be below 0.20.

1.3.9 SINGLE DETECTOR DATA GRANULARITY

The accuracy of travel time estimates is judged by looking at individual vehicle trips, but traffic sensors often aggregate measurements of multiple vehicles. Even the most accurate and reliable measurements of time-mean speeds may be not be that useful if the data aggregates vehicles over too large time periods. As a metric of single detector data granularity, examine the frequency with which the sensor reports data.

EXAMPLE REQUIREMENT: The speed detector must report time-mean speed estimates every minute.

2 DEDICATED DATA COLLECTION TECHNOLOGIES

Over the past several years, a number of private industry vendors have approached Caltrans with solutions to collect travel time data on highways and city arterials. Solutions revolve around two basic concepts and trends. The first trend suggests leveraging new technologies that significantly lower the cost of fixed detection. Both in-pavement technologies such as wireless magnetometers from Sensys Networks, Inc. and off-pavement technologies such as radar-based sensors by Speedinfo, Inc. offer much more attractive price points than inductive loops and make it conceivable to augment detection to a level that would yield accurate traffic maps and travel time estimates. An alternative concept is to use so-called mobile traffic probes to measure travel times from actual trips. Mobile traffic probes are essentially vehicles that are tagged and tracked along a corridor. This concept can be implemented by toll collection tags and readers, or by automated license plate readers. In either of those two cases, travel times are collected for preset segments of roadways in-between readers. For instance, the San Francisco Bay Area 511 system relies for a large part on data collected from FasTrak readers.

There is yet another way to collect traffic information, by leveraging the flow of cell phones and other connected devices. Cell phone-based methods have the immense advantage that for the most part, they do not require any additional infrastructure. Collection of travel time information from cell phones is the object of its own section, following the present one which focuses on 'dedicated' technologies.

Various technologies available have different capabilities and costs. The following subsections summarize the technology options and describe some of their results in the field. Point detection technologies, which include intrusive and non-intrusive technologies, are discussed first. This is followed by segment reidentification technologies.

It is also reasonable to assume that multiple technologies can be embedded together in a particular ratio to achieve high quality data at a lesser price. There are some areas (e.g. San Francisco Bay Area) that are currently using multiple, overlapped sources of data to produce travel times. However, data fusion remains practically in its infancy. There are no State DOTs that are using a planned and researched combination of multiple technologies in order to produce travel times.

2.1 METHODOLOGY

In late 2005, CCIT organized a Travel Times Forum in San Diego, Cal. The purpose of this gathering was to discuss "the needs and the availability of ubiquitous, accurate travel time information." This conference was a valuable tool in creating a network of contacts within the field of study surrounding travel times. For this section of the report, we have surveyed this network by focusing on three types of organizations: state agencies, suppliers, and solution integrators or content providers. These organizations were chosen based on their ability to help us investigate five particular questions and their implications in real-world applications:

1. Rationale for Highway Sensor Spacing (distances between traffic detectors)
2. Algorithms for Travel Time Calculation

3. Traffic Data/Travel Time Performance Metrics
4. Public/Vendor Feedback
5. Pros/Cons of New/Existing Sensor Technology

From our list of possible contacts in the three categories (state agencies, suppliers, integrators), research was conducted on each organization to determine its role in finding answers to the five questions posed above. Based on their company's role and fit to our goal, they were ranked on a priority system of 1-4 (1 = lowest, 4 = highest). State agencies have a major role in this study as they are the primary client for travel time technologies. Often, they are the final judge in determining sensor spacing, installation techniques, and detection system types. Suppliers provide the detection and software technology to the governmental organization, act as implementation consultants, and often have a greater breadth of understanding in this area of study. Finally, solution integrators serve as third-party vendors of detection equipment, while content providers aggregate and process traffic information for media distribution. These organizations implement traffic detection and process data, thus having an understanding of the entire value chain. All three categories described above were included in our investigation while providing valuable information for our study. Interview notes are provided in Appendix II.

2.2 POINT DETECTION

Point detection sensors have been used in highways for many years and make up the majority of traffic detection on US highways. In the context of highways, these sensors are placed at varying distances and determine a number of particular measures from the traffic that passes by. These measures typically include: presence, occupancy (density), speed, vehicle classification. These sensors in conjunction with additional hardware and software can produce valuable information about a highway segment, including travel times. Their main advantage is to produce a complete set of measurements, based on exhaustive sample, at a particular location. A disadvantage of the use of point detection is maintenance of sensors sites and the high costs of implementation of an effective sensor infrastructure. Some of these issues will be discussed in the following summary about various point detection technologies used in the calculation of travel times. The summary will be categorized first by embedded (intrusive) followed by non-intrusive technologies.

2.2.1 INDUCTIVE LOOP DETECTION

Inductive loop detectors are among the oldest technology in traffic detection that is available today. Inductive loop detectors (ILD) make up a vast majority of the point detection found in our road. As a brief background, here is a description of the workings of an ILD provided by the Vehicle Detector Clearinghouse.

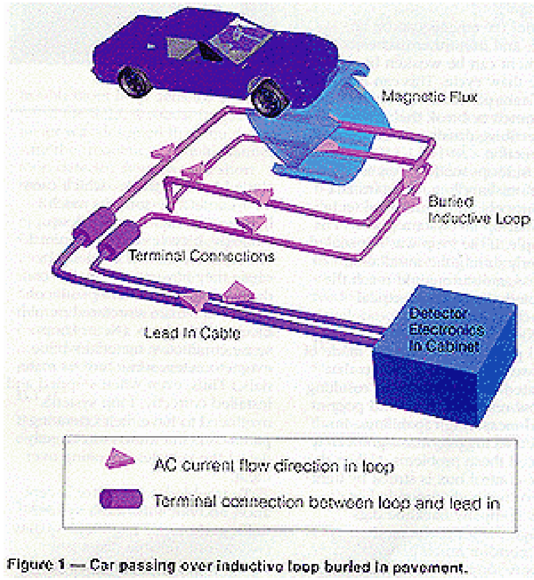


Figure 3 - ILD Schematic[4]

Passive magnetic devices [ILD] measure the change in the earth's magnetic flux created when a vehicle passes through a detection zone. Active magnetic devices, such as inductive loops, apply a small electric current to a coil of wires and detect the change in inductance caused by the passage of a vehicle.

Some governmental agencies have steadily attempted to move away from this intrusive technology. However, the sheer number of ILD installations currently in place has been a deterrent in transforming towards different technologies (infrastructure is already place to support loops and loop data). Loops are needed per lane, and have a high maintenance cost (the loop itself as well as lane closures). Thus, the costs of many miles of loop detectors on a 4-lane highway quickly escalate. Additionally, loop detectors and their supporting infrastructure are prone to breakage or failure based on general road maintenance (one study concludes 20% failure rate/year). Advantages come in the sense that the technology has a very high accuracy rate and requires little training to implement. The technology is well proven in the field, and can be an effective solution in certain situations.

2.2.2 WIRELESS MAGNETOMETER

Sensys Networks Inc. of Berkeley, Cal., manufactures a unique vehicle detection system that employs small wireless sensors embedded in pavement. The technical concepts that underlie the establishment of Sensys Networks were developed to fulfill a research contract issued by Caltrans Division of Research and Innovation.

Functionally similar to inductive loop detectors, Sensys Wireless sensors are used for traffic flow monitoring and signal control. The data they collect enables freeway applications including include

[4] Federal Highway Administration http://www.tfhrc.gov/pubrds/septoct98/loop.htm

system performance monitoring, ramp metering, incident detection, travel time estimations, vehicle classification, and bottleneck analysis. The key value proposition of the Sensys Wireless Vehicle Detection System is that it offers a substantially lower lifecycle cost than traditional loop detectors without sacrificing performance.

Figure 4 - Sensys VSN240-f sensor node

A trial of Sensys technology at the Berkeley Highway Lab established that initial installation and setup times are reduced by almost one order of magnitude. Data collected over a 2-week period was of similar quality to that of the incumbent loop technology, proving to be timely, complete, valid and accurate. Yet the lifecycle cost of Sensys Vehicle Detection Station is estimated to be $22,500 over a 15-year period, less than half of the $49,500 cost for the same station using inductive loops. Over that same period, the Sensys station would require closing 15 lane.hours while loops would require 56 lane.hours. A Road User Cost model shows that such closures translate in an additional lifecycle cost to the traveling public ranging from $3,000 to $7,000 per site where loops are installed, as opposed to approximately $400 where Sensys sensors are employed.

As of this report, Sensys Networks supplies several DOTs and dozens of cities across the United States and internationally. Operational deployment at Caltrans is also underway in several Caltrans districts.

2.2.3 DOPPLER MICROWAVE

Doppler microwave is non-intrusive technology that has been used for gathering traffic speeds with low power consumption (ability to be solar powered). These units work by transmitting low-energy microwave radiation at a directed point on the highway pavement. This reflection is then used to detect moving vehicles and determine their speed.

Figure 5 - DVSS-100 by SpeedInfo

These types of systems hold their strength in calculating speed, but are found to be less accurate at determining vehicle counts and other pertinent information. Doppler microwave units have been implemented in a few states on the eastern seaboard New Jersey, Maryland, and North Carolina. California's Bay Area has used SpeedInfo Doppler sensors in conjunction with loop detectors to calculate travel times. North Carolina has used SpeedInfo sensors alone to calculate travel times. Studies report that Doppler microwave units cannot detect speed under a certain minimum. This minimum speed varies between sensors.

Some companies (ASIM Technologies) have bundled this technology together with others in a single product in order to provide a broader wealth of traffic information for users. However, Doppler microwave sensors may be a good solution for applications where speed is the primary required measurement.

2.2.4 PASSIVE ACOUSTIC

Passive acoustic sensors use methods of "acoustic imaging" to determine characteristics of highway traffic data. The graphic above depicts what an acoustic sensor detects. These sensors are capable of capturing a wide variety of traffic data characteristics and have utilized in a small segment of transportation.

Figure 6 - Multi-lane highway monitoring using SAS-1

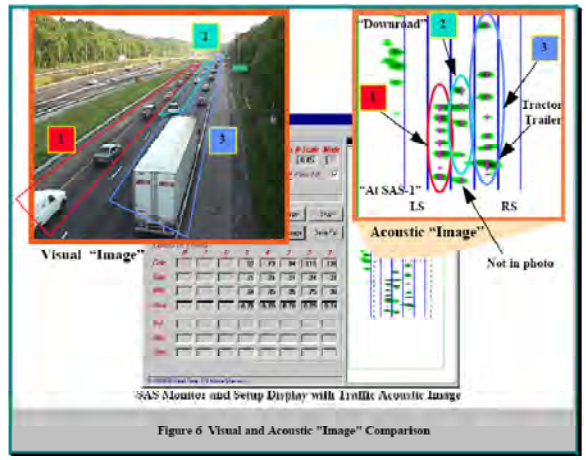


Figure 7 - Smartek Systems SAS-1 Passive Acoustic Sensor

When installed, these non-intrusive devices have a very low power consumption (1/4 – 1/5 the amount versus radar, which will be discussed shortly) and come at a relatively inexpensive price. Their low power consumption gives passive acoustic devices the ability to be completely solar powered. Upon installation, this technology requires a higher installation point compared to radar and video. This is to assist with occlusion issues. Acoustic units have an advantage of being resistant to adverse weather conditions with the exception of temperature. Colder weather has been known to disturb the accuracy of these sensors. Additionally, these units have not performed well in stop-and-go traffic conditions. Thus far, there have been mixed reviews about the accuracy of acoustic detection systems. A 1999 Texas Transportation Institute study states that passive acoustic detection devices do a relatively poor job of determining speed (55% accuracy), therefore most are limited to observing vehicle counts. On the contrary, a more recent report produced by the Minnesota DOT reports that speed accuracy was within 8% of baseline under their test of SmarTek SAS-1 passive acoustic sensors . New York State's Long Island is currently using SAS-1 detectors and will begin using them for the purpose of determining travel times within the next year. Traffic.com is also a know user of passive acoustic sensors. This type of sensors may be best suited where an overhead sensor is needed, but right-of-way is limited and overhead utilities are not of issue.

Table 2 - Models of passive acoustic sensors

Manufacturer	Model	Approx. Costs	Locations Used
International Road Dynamics	Smartsonic	$6000/4 lanes	AZ, TX, VA, MA
Smart-Tek Systems	SAS-1	$3500/4 lanes	VA, OH, NY

2.2.5 RADAR MICROWAVE

Radar technology has made significant advances within the last decade and has made a substantial impact. Currently, NAVTEQ (traffic.com), one of the nation's largest providers of travel time data uses radar devices for 90% of their traffic detection nationwide. Radar detection devices in side-mounted configuration are effective in reading across multiple lanes (8-10 lanes) and can determine all necessary traffic characteristics (volume, occupancy, speed).

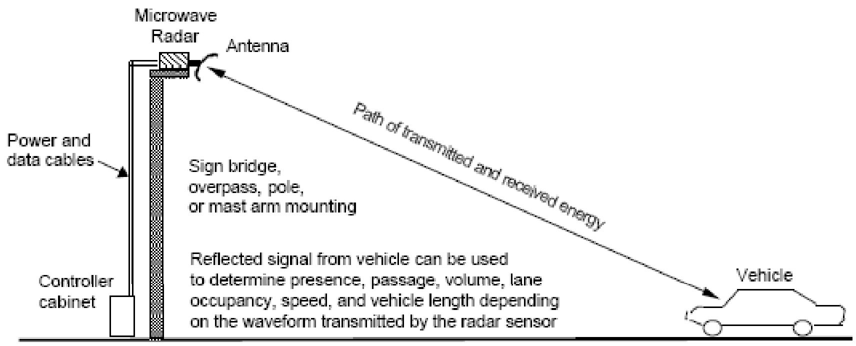


Figure 8 - Microwave radar setup

Radar microwave's advantages come from its relatively low initial and operational cost. Since it is non-intrusive, it can be easily added along current roadway configurations in a side-fire or overhead position. It is important to note that the actual detector placement is critical in the unit's accuracy and effectiveness. Depending on its application, radar may prove to be an effective means of calculating travel times. Drawbacks include its susceptibility to occlusion (vehicles being lost behind other vehicle profiles), and inability to detect over barriers. The major players in our study that are producing radar-type detectors are EIS and Wavetronix. A relatively recent study completed by the state of Pennsylvania concluded that the Wavetronix Smartsensor had a higher accuracy rate than the EIS RTMS units. More information can be found in the Traffic Data Collection Methodologies Final Report (Feb 2006).

Table 3 - Models of microwave radars

Manufacturer	Model	Approx. Costs	Locations Used
EIS	RTMS (and variants)	$4000/8 lanes	WI, OH, CA, NY, NJ, TX
Wavetronix	Smartsensor		UT, CA, IA, TX , FL, NY, NV

2.2.6 VIDEO DETECTION

The Georgia Department of Transportation (GDOT) is the country and world's largest employer of video-detection for use in their travel time system (Georgia Navigator). The technology works by mounting a video camera high above the highway segment so it can "read" segments up to ½ mile long and multiple lanes (up to eight). The feed data is then processed where computer software can generate vehicle speed, lengths, and counts.

Figure 9 - Autoscope Video Sensor (VDC 2002)

An advantage of this technology is that it can effectively detect traffic over multiple lanes and give operators a visual image of traffic flow for incident detection and speed validation purposes. Drawbacks to this technology come from its high costs, susceptibility to poor weather conditions (including wind and any precipitation) and occlusion[5]. A large amount of effort and maintenance is also needed to keep the video cameras clean and free from natural obstructions. However, based on studies from Georgia and other states, they have found that environmental conditions have had little impact on video-detection's ability to predict accurate travel times.

At the time of implementation (early 1990s), GDOT made the decision to use video technology, as it was the best non-intrusive detection technology available. However, some critics believe that the technology is now best suited for intersection control and incident detection. Nonetheless, the Georgia Department of Transportation transformed its traffic detection system from not only detecting incidents, but also to predicting travel times which are then displayed on changeable message signs as well as their Georgia Navigator website. In general, video detection is an effective means of collecting traffic data where a rich and diverse amount of data is required over a wide field of view (e.g. multi-lane highway applications).

Table 4 has been provided to give a brief summary of the technologies listed above. The data is categorized by technology type and should be used in general comparison. It should be noted that these statistics do not necessarily represent the performance of all types of sensors within a particular technology category.

[5] Occlusion: occurs when vehicles are blocked from camera view from obstructions (e.g. trucks, larger vehicles)

Table 4 - Models of video sensors

Manufacturer	Model	Approx. Costs	Locations Used
Computer Recognition Services	TAS-2	$17,000/site	CA, MA
Image Sensing Solutions	Autoscope	$18,200/site	N/A
Iteris	Vantage One/Edge/Plus	$6000/approach	CA, TX, VA
Traficon	VIP 3.1	$5,000/approach	FL, GA

2.2.7 SUMMARY OF POINT DETECTION CHARACTERISTICS

The following table sums up essential characteristics of selected technologies covered in the previous section.

Table 5 - Summary of point detector characteristics

	Count	Presence	Speed	Occupancy	Classification	Multi-Lane Detection	Communication Bandwidth6	General Accuracy	Average Lifespan	Manufacturers
Inductive Loop	X	X	X	X	X		Low to moderate	Volume = 99.6% Occupancy = 99% Speed	5-15 years, depending on installation conditions	Reno A&E Neverfail Truvelo Peek Traffic 3M
Magnetometer[7]	X	X	X	X	X			Volume = 96% Occupancy = 98% Speed = 99%	60,000 hours (MTBF)[8]	Sensys Networks
Doppler Microwave	X		X			X	N/A	Volume = Occupancy = Speed = 99%	60,000 hours (MTBF)[9]	SpeedInfo

[6] Bandwidth is defined as the speed in which data travels between the sensor and the central traffic system: Low: 2.4kbps-19.2kbps, Moderate: 19.2kbps – 115kbps, High: 115+kbps

[7] Sensys Networks Wireless Sensor, CCIT Report TO 5

[8] Manufacturer Spec Sheet (Sensys Networks)

[9] Manufacturer Spec Sheet (Speedinfo)
Westchester County Advanced Traveler Information Systems (ATIS) Traffic Flow Sensor Technology Design Alternatives Analysis

Passive Acoustic	X	X	X	X	X	X	Low to moderate	Volume = 94% Occupancy = N/A Speed = 95%	7 years (MTBF) 61,320 hours	SmartTek Systems Smartsonic
Doppler Radar	X	X	X	X	X	X	Moderate	Volume = 95% Occupancy = 95% Speed = 90%	10 years (MTBF), 87,600 hours	EIS Wavetronix
Video Image	X	X	X	X	X	X	Low to High	Volume = 34% Occupancy = 38% Speed = 35%	10 years (MTBF) 87,600 hours	Computer Recognition Systems Iteris Image Sensing Systems Traficon

2.3 SEGMENT-BASED TECHNOLOGIES

Segment reading types of technologies have only been recently used in the collection of traffic data for the purpose of estimating travel times. Systems using segment read types of technologies have the ability to provide a “ground truth” of travel times through a particular corridor. When implemented, these systems determine the true travel time between two points by tagging (or reading) a vehicle as it passes an “entry gate”, then again at an “exit gate” downstream. Technology such as toll tag reading and license plate recognition provide the truest form of travel time data. However, the use of this data for travel time estimation is often in question because of the lag time involved for a vehicle to cross the segment distance. Thus, the travel time of a vehicle leaving the segment may be very different than the travel time for a vehicle nearing the entrance of the segment. This lag time is a detriment to this type of technology for real-time applications.

Segment detection technology has a strong potential for application in regions with typical free-flow speeds, but may be hampered by weather conditions, scheduled static delays, or other types of incidents. Additionally, segment reading currently has the ability to be deployed in portable units, notably including Bluetooth readers, opening up the possibility for temporary, seasonal or emergency deployment.

2.3.1 ELECTRONIC TOLL TAG

Electronic toll tags and similar transponder devices are currently being utilized in a wide number of US states. These small RFID transponders provide a means of automatic vehicle identification (AVI) as a vehicle passes by a data collection site (typically a toll plaza or a portable collection unit). To date, only a few locations are using these toll tags to aid in the calculation of travel times. The city of San Antonio, TX is one metropolitan area that is utilizing AVI as a component in determining travel times. In San Antonio, a rolling average travel time algorithm is used to determine the travel time, which has proved to be of a higher accuracy than those reported by loop detectors. The state of California has also been using this type of data in the Bay Area in conjunction with point detection sites for the calculation of travel times.

Figure 10 - Electronic toll transponder

2.3.2 LICENSE PLATE MATCHING

The concept of license plate matching and vehicle identification for the calculation of travel times has been used as early as the 1950s. However, the more common use of this technology is for origin-destination research. Today, automatic license plate readers conduct this process of plate-matching (more commonly referred to as automatic vehicle identification (AVI)) and provide the ability to calculate actual travel times through corridor segments. In our study, we have found New York State and San Antonio, TX to be the only areas that are utilized license plate readers as permanent means of travel time prediction. There have been a few applications where license plate readers were utilized for short-term projects (construction, workzones, etc). A CCIT study conducted as part of Task Order 1009 provided an extensive test bench for this technology.

Figure 11 - Licence plate matching technology by Pips

Advantages of using license plate matching include the ability to provide travel times for small time intervals (incidents, peak hours) and its ability to provide a representative estimate of the travel time by use of random sampling. However, disadvantages come in its moderate costs, impracticality for low-traffic corridors, and ability to collect only travel times (which may breakdown in stop/go conditions).

Table 6 - Selected LPR Technologies and use for Travel Time estimations

Manufacturer	Model	Approx. Costs	Locations Used
Computer Recognition Systems	NRS	N/A	N/A

PIPS Technologies	P372	AZ, VA

2.3.3 ADDENDUM TO SEGMENT-MATCHING TECHNOLOGIES

As of the writing of this report, new technologies enable segment detection. These include wireless detectors by Sensys Networks, already covered in the section on point detection. By deploying an array of 3-5 sensors across a lane, Sensys can assemble a relatively unique magnetic signature for each passing vehicle, allowing reidentification downstream at a similarly equipped location. Current tests indicate a reidentification capability of about 70%, which is in line with the capabilities of license plate readers.

Another solution that has generated much interest in the past couple of years consists in setting up Bluetooth readers by the side of the roadway and to capture the MAC addresses[10] of networked personal electronics such as cell phones, navigation systems and laptops. Known manufacturers include Traffax and TrafficCast, whose product is nicknamed BlueToad.

2.4 CURRENT FINDINGS IN SENSOR SPACING

The determination of optimal distances between data sensors is an important yet relatively unresearched component in the travel time estimation puzzle. A primary goal of organizations that produce travel times is to do so accurately and reliably. This is driven first and foremost by the amount of traffic data produced, which correlates with the amount of sensors placed in the roadway. As sensor density increase, the associated costs also rise. Based on these factors, it was our goal to involve different agencies and academics on the study of sensor spacing for the means of determining travel time estimates. Note that since the investigations reported in this section, CCIT has conducted and delivered PATH Task Order 6328 (Optimal Sensor Requirements). That task order's final report provides a much more complete and in-depth assessment of the sensor spacing problem and makes some noted contributions.

This investigation in spacing can be further broken down into two categories. These categories include sensor spacing for point detectors, and sensor spacing of segment detectors. Point detector spacing is the primary objective for this study. Previous history has suggested that equidistant spacing configuration of these sensors (e.g. ½, ¼ mile) has been the typical means of implementation. However, there has not been many academic research to validate that this equal spacing provides good data for the lowest capital expenditure.

In our study of a few states that are currently predicting travel times, sensor spacing ranges approximately between ¼ - ½ mile. Unfortunately, these distances were not calculated nor assumed by scientific means. It seems as though these approximate distances were inferred as standard based on

[10] MAC addresses are unique identifier assigned to computer networking gear and nodes

systems built prior to those mentioned directly in this study. It is also of importance to note that these distances typically did not form from the need to predict travel times, but come originally from the need of incident detection and vehicle counting. Therefore, the distances provided by the few states in our survey have no academic nor quantitative reasoning, yet have been claimed by their respective states to hold the capability to predict accurate travel times. In most cases, this validation of travel time prediction has been done through simplistic and fragmented methods. Then to help improve and correct for better accuracy, changes are then made to the algorithms that actually predict the highway's travel time. On the contrary, Traffic.com, a nationwide traffic information provider, has been installing their RTMS Radar cameras at distances of approximately 1 – 2 miles, which is fairly large change in traditional spacing distances. However, these standards share their lack of solid support or justification for their spacing distances.

The second category involves segments detection such as toll tag readers or license plate readers. These carry the stipulation that the segment can be infinitely long and calculate a correct travel time. However, as segment length increases, the lag time between the calculated travel time, and the start point of the segment increase in a direct relationship. Thus, the accuracy of a travel time using segment technology can potentially be very inaccurate in longer segments. Segment detection technology has not been overly cited in the determination of travel times. In terms of our study, New York State was the only group to use their segment detection technology as the sole means of calculating travel times. The figure below shows some of our findings in terms of sensor distances in a variety of applications.

Table 7 - Sensor spacing practices

Organization	Current Spacing	Justification	Future Plans
Arizona DOT	¼ - ½ mile (loop)	N/A	Move towards 1-mile spacing due to rising costs
California DOT	1 mile (loop)		N/A
Georgia DOT	1/3 mile (VIDS)	Spacing adequate for incident detection	Plan on staying at 1/3 mile
Traffic.com (CA)	1-2 mile (Radar)		
Minnesota DOT	½ mile	Implemented in conjunction with ramp meters and finding volume counts	N/A
San Antonio TransGuide	½ mile		
New York DOT	N/A	Using Segment Reads to Determine Travel Times	N/A

Through the research of this report, it has been found that most standards for sensor spacing have not been thoroughly researched. Most standards that are currently in practice ¼ - ½ + mile spacing, has truly been a product of a follow-the-leader mentality, with little regard to the reliability of the traffic data that comes from a segment. Now that a secondary role for traffic detection sensors has arrived in the form of travel time detection, we have seen a shift in spacing standards (example being radar). Yet, once again, these standards have little quantitative support and justification.

3 THE USE OF CELL PHONES AS TRAFFIC PROBES

Floating point technology is one of the newest technologies to date in the determination of travel times. Studies have been conducted to investigate the movement of cell phones and/or GPS units (sometimes embedded in cell phones) as an indication of vehicle movement. Floating point technologies have most recently taken the form of either cell phone signals and/or GPS. Both technologies have gone through several trial evaluations in the United States, but have not been fully implemented by any state. Both of these technologies, along with their strengths and limitations, will be addressed in the following sections.

Note that the investigations reported in this section were conducted in late 2006 to early 2007. Much has happened in the field since then, though technical explanations remain valid. In-depth analysis of probe data originating from GPS-equipped cell phones was conducted as part of the Mobile Century and Mobile Millennium experiments. The technical notes corresponding to these experiments can be found in the final report to CCIT Task Orders 1021, 1029, and the Mobile Millennium Research Technical Agreement. The following subsections do not incorporate these findings but have been updated wherever they have become inaccurate as a result of the passage of time.

3.1 OVERVIEW

To this point in the paper, the discussion has focused on estimating travel times using fixed-location sensors of traffic conditions. Nascent technologies use data collected from cell phones treated as traffic probes to estimate travel times. The movements of cell phones indicate the movements of vehicles, and analyses of cell phone location data can reveal real-time traffic speeds and travel times throughout a roadway network. Cell phone location data can be gathered relatively cheaply (possibly by passively monitoring cell phone service provider data streams), without having to install or maintain roadway infrastructure. Such data could include data relating to traffic on arterial roads, costly to obtain via alternate methods. It could even provide estimates of origin - destination demands and travel times. For these reasons, the use of cell phones as traffic probes presents an exciting opportunity.

A PATH simulation of travel time estimation shows the promise of using cell phones as traffic probes. Simulation results showed that if 5% of highway vehicles are equipped with cell-phones then it becomes possible to predict link travel times with 95% accuracy (California PATH, 2000). As a follow-up to their simulation, PATH also surveyed San Francisco Bay Area residents to roughly estimate what proportion of vehicles has functioning cell phones. This survey concluded that, in November 1998, “37.7% of the Bay Area commuters are expected to travel with a cell phone” and that “approximately 50% of freeways [sic] users are traveling with a cell phone” (California PATH, 2000). The PATH simulation and survey, taken together, indicate that if even a fraction of the cell phones in use on highways today could be used as traffic probes, reliable travel time information could be obtained.

The primary methods for using cell phones as traffic probes rely on being able to precisely track the movements of cell phones. Important technical hurdles involve geographically locating cell phones and then connecting the movements of cell phones to particular vehicle paths on a roadway network. The geographical spacing between roadway segments and the precision with which cell phones can be located become important here. Once a sufficiently large number of vehicle paths have been collected, it is possible to say something meaningful about travel conditions on roadways. Traffic lights, stop-and-go traffic, and vehicles traveling in an unusual manner (motorcycles passing queued cars, delivery vehicles stopping in the roadway, etc.) complicate analyses.

Alternate methods for using cell phone data involve monitoring the movements of large numbers of phones. This type of analysis avoids the privacy concerns raised when tracking individual phones / vehicles. Less precise geographical data is necessary. However, the results of this type of analysis are limited. It may be possible to say something about traffic on major highways, especially in rural areas, but it is impossible to track traffic on local roads.

Past projects involving using cell phones as traffic probes have produced mixed results. According to one summary report, initial tests have "not produce[d] data of sufficient quality or quantity to provide reliable traffic condition estimates" (University of Virginia, 2005). Difficulties have been especially pronounced on urban arterial roads and during off-peak hours. On a more positive note, a few different technologies have geographically located cell phones with some precision. Let's examine the different technologies used to locate cell phones.

3.1.1 CELL TOWER METHODS

Time Difference of Arrival (TDOA) technology can calculate the position of a cell phone by tracking the time signals take to travel from the phone to the towers that provide service to the phone (Cell-Loc Inc, 2002). Once at least three times have been recorded, it is possible to use triangulation to "make a fix" (determine the position of the phone) to an accuracy of around 100m. In remote locations or when physical objects like buildings obscure signals, it may be difficult to use TDOA to make a fix. Cell-Loc of Calgary, Canada has used TDOA technology to measure traffic speeds as part of a proof-of-concept for Transport Canada. Decell Technologies has used measurements of the strengths of the signals cell phones receive from various towers in place of TDOA data to track traffic in Israel[11]. The technology is similar to TDOA and it can be inferred that the accuracy and reliability of fixes obtained using signal strength data are roughly equivalent to those associated with TDOA data.

It is worth noting that the precision with which TDOA and signal strength techniques identify the location of cell phones can be increased by considering a history of past measurements alongside current data. Scottish firm Applied Generics uses such a technique and claims to have a higher "data

[11] http://www.decell.com/index.htm

quality" than the methods described above[12]. Many other firms, including IntelliOne and AirSage, are combining similar technologies with cell handoff data[13].

According to a 2005 review of the use of cell phones as traffic probes, "the majority of recent deployments use cell handoffs to define vehicle paths and speeds" (University of Virginia, 2005). A cellular phone service provider has a number of base stations that transmit to and receive signals from phones in the adjacent area, i.e. within a base station's 'cell.' A 'cell handoff' occurs when a phone moves from one cell to another and begins communicating with a new base station. Service providers currently collect data on cell handoffs, making it easy to obtain this data. Cells vary in size from less than a kilometer in central business districts to around 60 km across in rural areas (California PATH, 2000). Because of the large size of cells, it is difficult to map cell handoff data to vehicle paths on a roadway network. However, cell handoff data can still be useful in determining a rough aggregate measure of how well traffic is flowing on major highways. This data is available now and collecting such data does not place any additional load on cellular communication networks.

3.1.2 GPS METHODS

Technologies based on the Global Positioning System (GPS) can determine locations to an accuracy of 5 - 15m (Cell-Loc, 2002), are improving every year, and are less sensitive to local environmental conditions than the systems described above (Globis, 2005). GPS technology relies on a network of 27 satellites that circle the earth. These satellites beam down very accurate time information to terrestrial devices equipped with GPS receivers. Devices receiving signals from at least three satellites can then use trilateration to calculate their own latitude and longitude. GPS devices are so precise that they have been used as true descriptions of position in studies of alternate cell phone locating technologies (California PATH, 2000; Globis, 2005). GPS data from cell phones have not been widely used for travel time estimation because (1) cell phones with GPS chips were rare and (2) the cost of obtaining GPS data was thought to be significantly higher than alternate methods[14]. However, GPS phones have now become commonplace and the cost of transmitting data has gone down a great deal over time.

Singapore, a country where electronic tolls and transponders are common in vehicles, has had the opportunity to conduct studies on the quality of traffic data collected via GPS units. In a 2002 study conducted by the National University of Singapore, it was concluded that speed accuracy begins to diminish when the probe vehicle drops below 15% on arterial roadways. In order to maintain accuracy

---

[12] http://www.appliedgenerics.com/downloads/RoDIN24-Brochure.pdf

[13] IntelliOne Technical Report Version 4.13G (February 2006)

[14] http://www.appliedgenerics.com/downloads/RoDIN24-Brochure.pdf

within 5.0 km/hr 95% of the time, a 4-5% percentage of GPS-equipped vehicles need to be active on the roadway[15].

Qualcomm has pioneered a system known as Advanced GPS (A-GPS) that increases the efficiency of GPS-based cell phone location algorithms (Globis, 2005). In A-GPS, a dedicated network server identifies which GPS satellite signals individual cell phones should look for to define their position. The cell phones receive the requested signals and pass the information back to the server. The server then performs the calculations necessary to make a fix. It is hoped that this system will shorten the time it takes to make a fix, bring down the costs incurred, and perform better than traditional GPS in environments (building interiors, etc.) where it is difficult to receive signals.

3.2 EXPERIMENTS

3.2.1 CAPITAL – ITS PROJECT

The first major project involving using cell phones as traffic probes was known as CAPITAL - ITS and was conducted over a 27-month period that concluded in November 1995 (Transportation Studies Center, 1997). The Federal Highway Administration and Virginia Department of Transportation contracted Raytheon E-Systems division to monitor traffic on major highways in the Washington, DC area. TDOA techniques were used to make fixes, which were then used to assign vehicles to roadway links and directions of travel. The University of Maryland independently monitored and evaluated data including that related to positioning accuracy and travel time estimation.

Cell phones in this project were geolocated with an accuracy of just over 100m. This was typically good enough to locate vehicles traveling on a highway on the correct link and direction of travel. It was estimated that 4 to 5 fixes were required to estimate the speed of any one vehicle. Link speed estimates could not be reliably computed during the project. The major obstacles were the limited accuracy of the fixes provided and the limited number of fixes received per vehicle. According to the concluding report of this project, "if the geolocation accuracy can be reduced to 5 to 25 meters and the signal can be consistently received, the system shows promise if the costs of doing this are not overwhelming" (Transportation Studies Center, 1997).

3.2.2 CALIFORNIA PATH PROJECT

One interesting project conducted by the California PATH organization in 2000 and 2001 compared travel information gleamed from GPS units to data from a cell phone locating technology developed by US Wireless (California PATH, 2001). Although the US Wireless locating technology is never identified, it appears to involve triangulation of data from cell phone towers like TDOA or signal strength

[15] Note that the Mobile Century experiment finds that the required penetration rate is much lower than this, probably in the order of ½% on a relatively dense urban freeway.

methodologies. The cell phone data was imprecise and, of the vehicle tracks studied, “66% had one or more points outside of a 200-meter accuracy range” (California PATH, 2001). The GPS data was found to be significantly more accurate.

The high accuracy and precision of the GPS data made it relatively easy to determine the roadways the GPS units traveled on. According to the report, software that analyzed GPS data “identified 93% of the path followed by [a] probe vehicle” traversing a known route (California PATH, 2001). Tracking probe vehicles, especially on non-highway roads proved to be “a minor problem for a GPS system but a major issue that must be addressed for cell phone tracking where the accuracy is poor” (California PATH, 2001). The inability to match the US Wireless data to vehicle paths resulted in “GPS data produc[ing] almost seven times as much travel information per unit tracking time” as the non-GPS data (California PATH, 2001). The conclusions are clear: (1) the precision of geographic locating technology has a direct impact on the amount of traffic information data that can be collected, and (2) GPS technology is precise and can be used to generate useful traffic information data.

3.2.3 HAMPTON ROADS, VA – 2003

In 2003, a joint venture between the Virginia DOT (VDOT), FHWA, and AirSage, a private vendor began in the Hampton Roads region of Virginia. This project was conducted to test the feasibility of cellular phone data collection over an approximate distance of 90 centerline miles. AirSage, a company from Atlanta, GA produces traffic data by mining “handoff” data that collected by wireless service providers. This data is then process and converted into a speed. In December of 2005, the system was finally tested with less than desirable results. In the evaluation, it was found that speeds determined by AirSage were in error of +20 miles/hour 68% of the time. Additionally, the performance tended to be worse in congested conditions. At the time, AirSage claimed that major improvements would be implemented in their system within 2006. The final report concludes that the AirSage system was not yet capable of producing accurate travel time estimates in its current form.

3.2.4 TEL AVIV, ISRAEL – 2005

In early 2005, ITIS was involved in a test of its WLT technology. Similar to AirSage, ITIS used handoff technology in the determination of its traffic data. The system was tested against a baseline provided by loop detectors spaced at a distance of 500m. Day testing under non-congestion conditions provided good results that closely followed that of the loop detectors. However congestion or low-volume conditions provided inconsistencies ranging from 10-30% from the loop detector baselines. In a report produced from Ben-Gurion University in Israel, it was concluded that the overall test results were “good”, but additional studies are need to quantify the results for various applications.

3.2.5 SUMMARY

The following table gives a brief summary of some of the field results of cellular network-based traffic studies conducted in the five years preceding this investigation (i.e. 2007).

Location	Year	Vendor	Technology	Results
Noord-Brabant, Netherlands	2003	Applied Generics/LogicaCMG		Reliable at speeds < 20 km/h Low sample sizes caused high variation
Hampton Roads, VA	2003	AirSage	Handoff-Based	68% of speed estimates with errors of >20 mph No reliability measures could be generated
Munich, Germany	2003	Vodafone		Call volume not well correlated to traffic volume Limited accuracy results
Tel Aviv, Israel	2005	IT IS	Handoff-Based	Limited data during off peak or night hours WLT estimates varied from loop data by 10-30% during congested conditions

Additional tests underway at the time of these investigations included Missouri (2006), Maryland (2006), and Georgia (2007). As of this writing (2010), these investigations confirmed that methods based on data collected from cell phone networks absent GPS localization are relatively ineffective. While they provide relevant data on dense urban freeways, they do not do so at a cost considerably lower than traditional means of detection. On signalized arterial, the data is of poor quality.

CONCLUSIONS

This report investigated point detection, probe data, and segment detection. It is noted by some experts, that a fusion of the data provided by these technologies would provide the best results in terms of travel time accuracy, reliability, and relevance. This type of fusion would be a planned combination of multiple layers of coverage (segment with radar, probe data with loop, etc). However, up to this point, few agencies or institutions that are conducting research in this field.

Table 8 - Technology summary

	General Guidelines
Point Detection – Non Intrusive	Adequate infrastructure is necessary (power, data communication, mounting hardware) Accordance to strict height/mounting requirements Consider overhead vs. sidefire mounting techniques Electromagnetic interference may disrupt microwave radar detectors
Point Detection - Intrusive	Cannot be used in poor pavement Cannot be used in bridge decks Adequate infrastructure is necessary (power, data communication) Consider lane closure impacts VIP sites require adequate night lighting Acoustic sensors cannot be placed in areas of already high noise
Probe Data	Frontage roadways or other busy streets in close proximity may disrupt travel readings Motorcycles cause alterations in data readings Low-traffic volumes produce less accurate data results Cellular/Handoff technologies perform best at high densities of cellular coverage
Segment Technology	Low-traffic volumes produce less accurate data results Ineffective in stop-and-go conditions

In closing, it is likely that travel time prediction will be implemented on an increasing number of highway corridors in the future. Therefore, it is important to investigate and understand the current infrastructure solutions. In terms of travel time prediction, this study was not prepared to evaluate detection technologies, but instead to objectively report on state-level decisions to choose certain technologies and determine those solution's effect in practice.

In the realm of detector technology, despite recent advancements in sensor technology, older, traditional technology continues to reign in many locations. This is due to the high amount of infrastructure, and operational and maintenance training that has been invested in implementing these systems. However, non-intrusive technologies have begun to show up in a variety of states and in use by third-party traffic companies. Technologies such as radar have the ability to provide equivalent data as traditional loop detectors, at a fraction of the cost, and without any in-ground maintenance.

Sensor spacing between the various means of detection technology remains somewhat of an open question. A fairly narrow range of distances is typical in a state-by-state comparison. However, despite this low variance in numbers, none of the states seem to have a strong understanding of the origin of these distances. At current, states are finding ways to keep predicted travel time accuracy high, but are not necessarily as confident in providing reliability measures. Some believe that high marks in both metrics can be accomplished, but only by meshing segment reading technologies as well as point detection technologies.

Theoretical simulations and previous studies have shown that it is possible to use GPS technology to track vehicles and determine travel times, even on urban arterial roads. Other technologies for tracking cell phones have demonstrated an ability to locate vehicles traveling on major highways. However, scholarly research has yet to demonstrate the potential of cell tower based technologies to provide useful travel time data, or to locate vehicles traveling on local roads. Yet most private firms interested in using cell phone data to track traffic have focused on cell tower based technologies. Moving forward, the first priority should be to look for a way to cheaply obtain data feeds of GPS fixes for cell phones.

The massive availability of traffic data at a lower cost will have important consequences for both the traveling public and roadway operators. The dissemination of traffic information will enable a form of system 'self-management', in which individual commuters can make informed travel decisions. Not only will each user benefit personally, but the entire driving community will enjoy more balanced loads across the road network. Roadway operators will also benefit tremendously from collecting rich performance data that can be turned into a wide range of strategies, from demand management to dynamic traffic control.

However, the data collected from cell phones and other connected devices is limited to velocity information. Operators have traditionally relied on detection systems that provide measures of traffic flow and occupancy rate. It is unlikely and probably not even desirable that phone-originating data will supplant existing detection systems at any point in the foreseeable future. On the other hand, there is a definite possibility to use mobile probe data as a complement to existing detectors, resulting in a so-called 'hybrid' traffic data collection system. In such a setup offering ubiquitous availability of speed information, traditional traffic monitoring stations (TMS) will only be needed to maintain accurate flow counts. Larger spacing intervals could be allowed in-between stations, and therefore equipment could be deployed much more sparingly than it is today. In sum, operators such as Caltrans could be getting a lot more data for a lot less money.

A recently approved technical agreement named 'Hybrid Traffic Data Collection Roadmap' intends to investigate the feasibility and the business case for such a hybrid traffic data collection system. One very important feature of such a system is that it involves not just new technology but also a completely different business process. In today's environment, Caltrans purchases detection technology than it then installs, operates and maintains internally, even if contractors may be used. In the case of mobile probe data from GPS-equipped cell phones, Caltrans would have to procure data collected and processed by a third-party, be it a cellular network operator or a service provider. This is a fundamental shift that poses its own set of challenges, irrespective of the technical merits of a hybrid system. Thus, in addition to engineering work on data integration optimal trade-offs between mobile probe information and traditional detection, the new agreement dedicates substantial resources to investigate the business aspects of traffic data procurement, be it from cell phones or other technological means.

By focusing specifically on data quality assessment, and in partnering with industry providers through the North American Traffic Working Group (NATWG), the project team has laid the ground for the upcoming work on determining the business case and the adequate procurement mechanism. The product of these activities are presented in Appendices III and IV, respectively.

BIBLIOGRAPHY

Bar-Gera, Hillel (2006). "Evaluation of a Cellular Phone-Based System for Measurements of Traffic Speeds and Travel Times: A Case Study from Israel". Department of Industrial Engineering and Management. Ben-Gurion University of the Negev, Israel.

Bartin, Bekir, Kaan Ozbay, Cem Iyigun (2006). "A Clustering Based Methodology for Determining the Optimal Roadway Configuration of Detectors for Travel Time Estimation. Rutgers University.

California PATH (2000). Travel Time Estimation on the San Francisco Bay Area Network Using Cellular Phones as Probes. *Working Paper UCB-ITS-PWP-2000-18*.

California PATH (2001). Investigation of Vehicles as Probes Using Global Positioning System and Cellular Phone Tracking: Field Operational Test. *Working Paper UCB-ITS-PWP-2001-09*.

Cell-Loc Inc (2002). Traffic Monitoring Application of Cellular Positioning Technology: Proof of Concept. *Transportation Development Centre of Transport Canada Report TP 13936E*.

Demidovich, Mark. Georgia Department of Transportation. Personal Interview. 9 November 2006.

Fairburn, Bob. EIS. Personal Interview. 8 November 2006.

Fariello, Brian. "White Paper Response: ITS America RFI Travel Time Projects in North America" San Antonio District, Texas Department of Transportation.

Fontaine, Michael, Brian L. Smith, Aron R. Hendricks, William T. Scherer (2007). "Wireless Location Technology-Based Traffic Monitoring: Preliminary Recommendations to Transportation Agencies Based on a Synthesis of Experiences and Simulation Results." University of Virginia.

Globis Data Inc (2005). Development and Demonstration of a System for Using Cell Phones as Traffic Probes. *Transportation Development Centre of Transport Canada Report TP 14359E*.

Halvorsen, Donald. MSI Sensors. Personal Interview. 4 December 2006.

Jones, Peter. PIPS Technology. Personal Interview. 17 November 2006.

Kary, Brian. Minnesota Department of Transportation. Personal Interview. 22 November 2006.

Kothuri, Sirisha M., Kristin Tufte, Soyoung Ahn, Robert L. Bertini (2006). "Using Archived ITS Data to Generate Improved Freeway Travel Time Estimates. Portland State University.

Kreth, William. Transport Data Systems. Personal Interview. 3 December 2006.

Kuhn, Kenneth (2006). "Literature Review: The Use of Cell Phones as Traffic Probes." California Center for Innovative Transportation.

Leach, Roger. Computer Recognition Systems. Personal Interview. 17 November 2006.

Manor, Dan. EIS. Personal Interview. 20 November 2006.

Martin, Peter T. and Yuqi Feng (2003). "Traffic Detector Selection Procedure," *Report No. UT-03.30. Utah Department of Transportation Research*.

Middleton D., Jasek D., and Parker R (1999). "Evaluation of Some Existing Technologies for Vehicle Detection." *Texas Transportation Institute Project Summary Report 1715-S*.

Pan, Changxuan, Jiangang Lu, Dawei Wang, Bin Ran (2006). "A GPS-based Arterial Roadway Network Travel Time/Delay Data Collection Method." Presented at TRB 2007.

Pieper, Greg. SmarTek. Personal Interview. 13 November 2006.

Polydoropoulou, A., Gopinath, D., and Ben-Akiva, M. (1996). Willingness to Pay for Advanced Traveler Information Systems: SmartTraveler Case Study. *Transportation Research Record 1588*.

Seeling, Lee-Ann S. Trichord Inc. Personal Interview. 13 November 2006.

Semmahi, Farid. Traficon. Personal Interview 12 December 2006.

Shupp Donald R. Western Pacific Signal LLC. Personal Interview. 5 November 2006.

Sutula, Tom. Reno A&E. Personal Interview. 20 November 2006.

Transportation Studies Center, University of Maryland (1997). Final Evaluation Report for the CAPITAL-ITS Operational Test and Demonstration Program.

University of Virginia Center for Transportation Studies, Virginia Transportation Research Council (2005). Probe-based Traffic Monitoring State-of-the-Practice Report. *NCHRP Project 70-01 Report*.

Versavel, Jo (2007). "Traffic Data Collection: Quality Aspects of Video Detection". Traficon n.v.

Wolfson, Dave. Arizona Department of Transportation. Personal Interview. 6 November 2006.

Wolinetz, L., Khattak, A., and Yim, Y. (2001). TravInfo Evaluation: Traveler Response Element Willingness to Pay for Traveler Information: Analysis of Wave 2 Broad Area Survey. *Path Working Paper UCB-ITS-PWP-2001-05*

Zibinski, Rick. New York Department of Transportation. Personal Interview. 17 November 2006.

APPENDIX I: FIGURES FOR SECTION 1

LIST OF FIGURES:

FIGURE 1.1

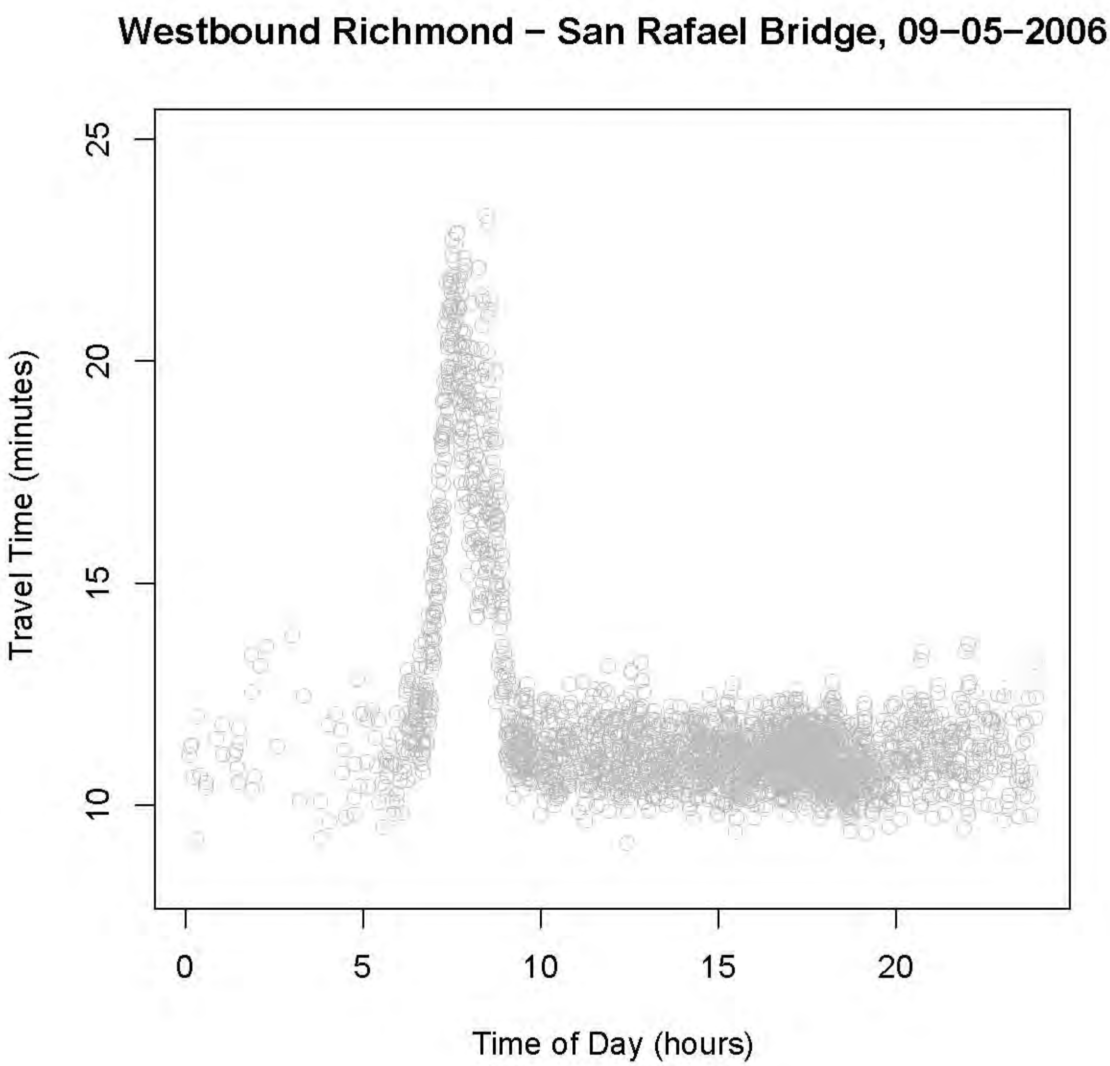

Westbound Richmond – San Rafael Bridge, 09–05–2006
Travel Time (minutes)
25
20
15
10
0
5
10
15
20
Time of Day (hours)

FIGURE 1.2

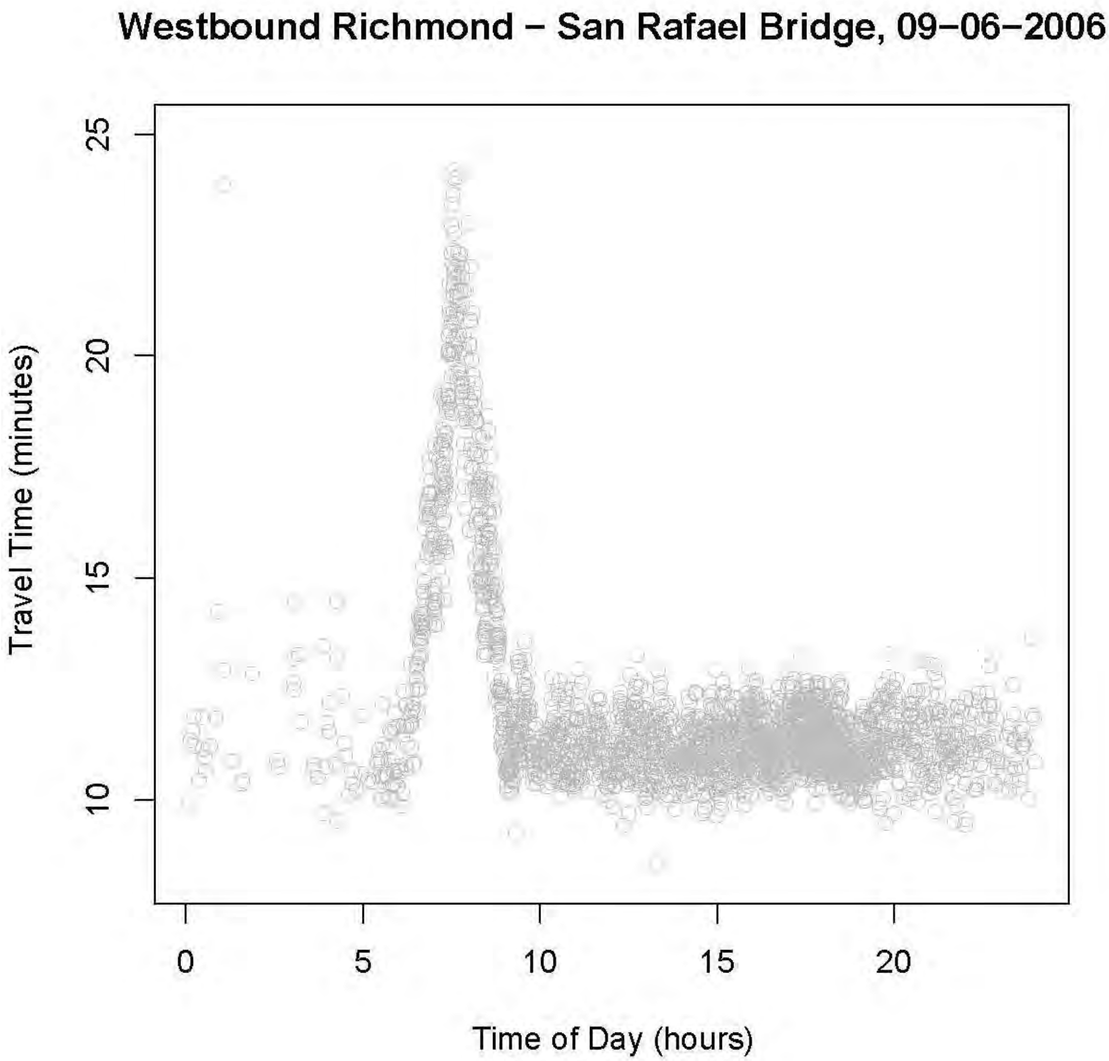

Westbound Richmond – San Rafael Bridge, 09–06–2006
Travel Time (minutes)
25
20
15
10
0
5
10
15
20
Time of Day (hours)

FIGURE 1.3

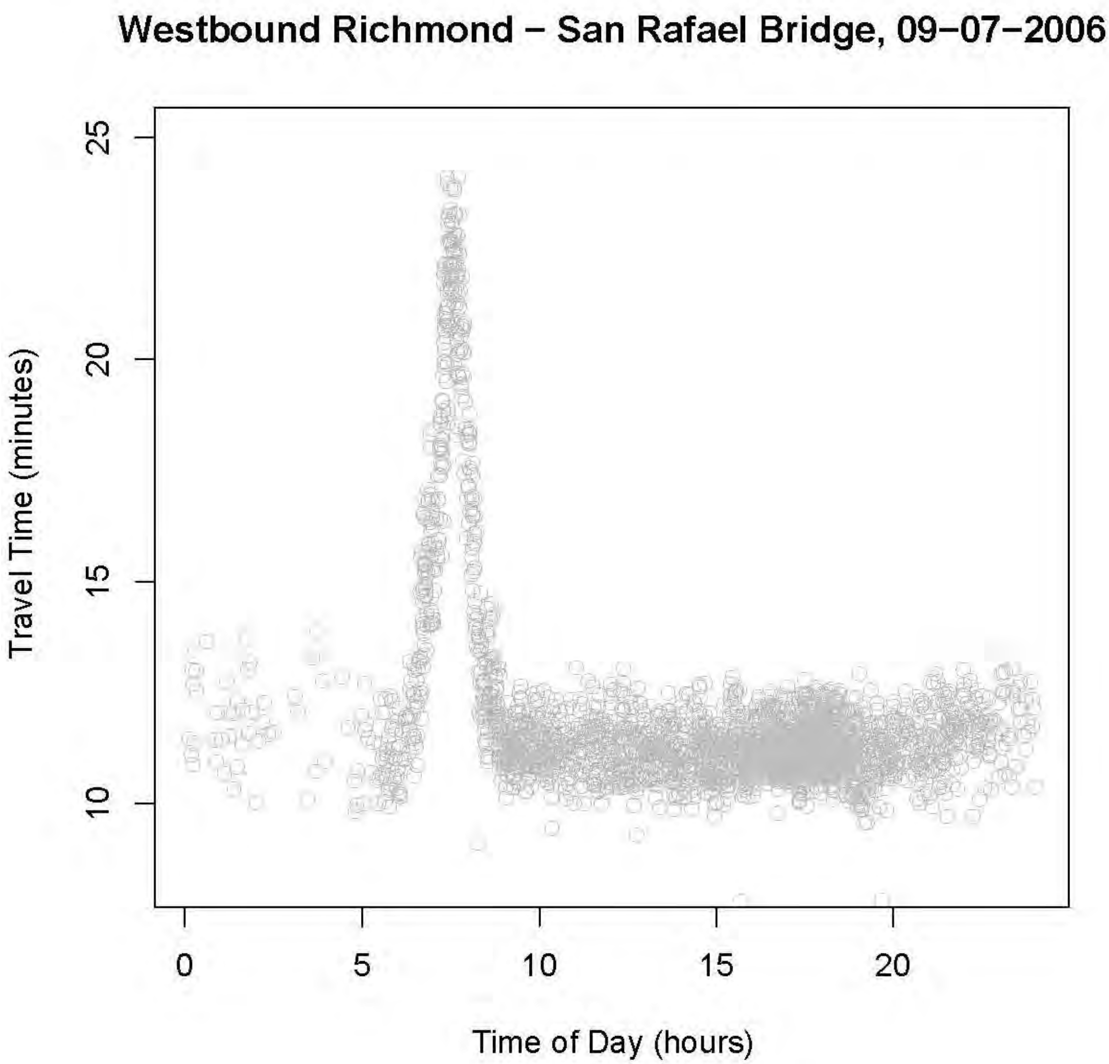

Westbound Richmond – San Rafael Bridge, 09–07–2006
Travel Time (minutes)
25
20
15
10
0
5
10
15
20
Time of Day (hours)

FIGURE 1.4

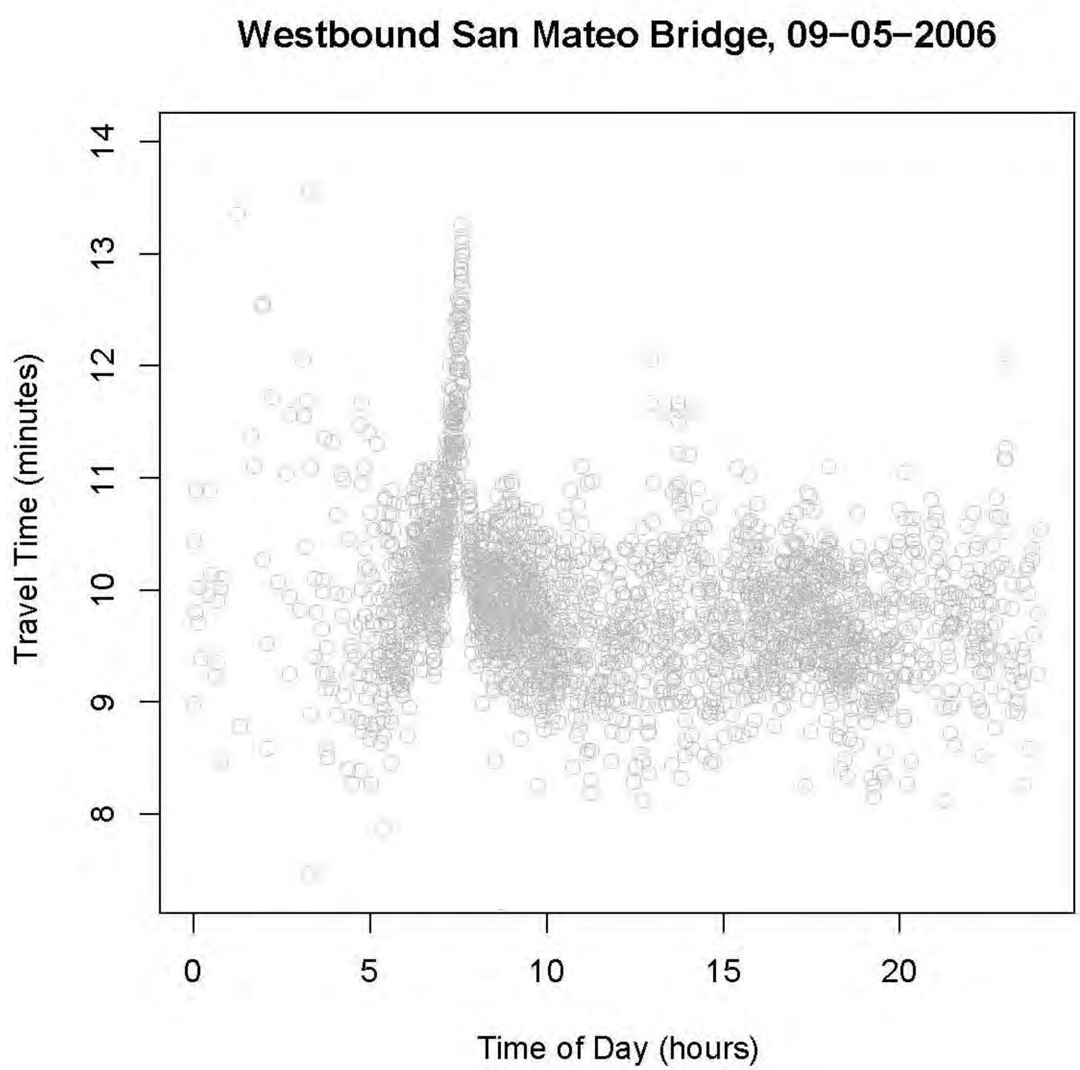
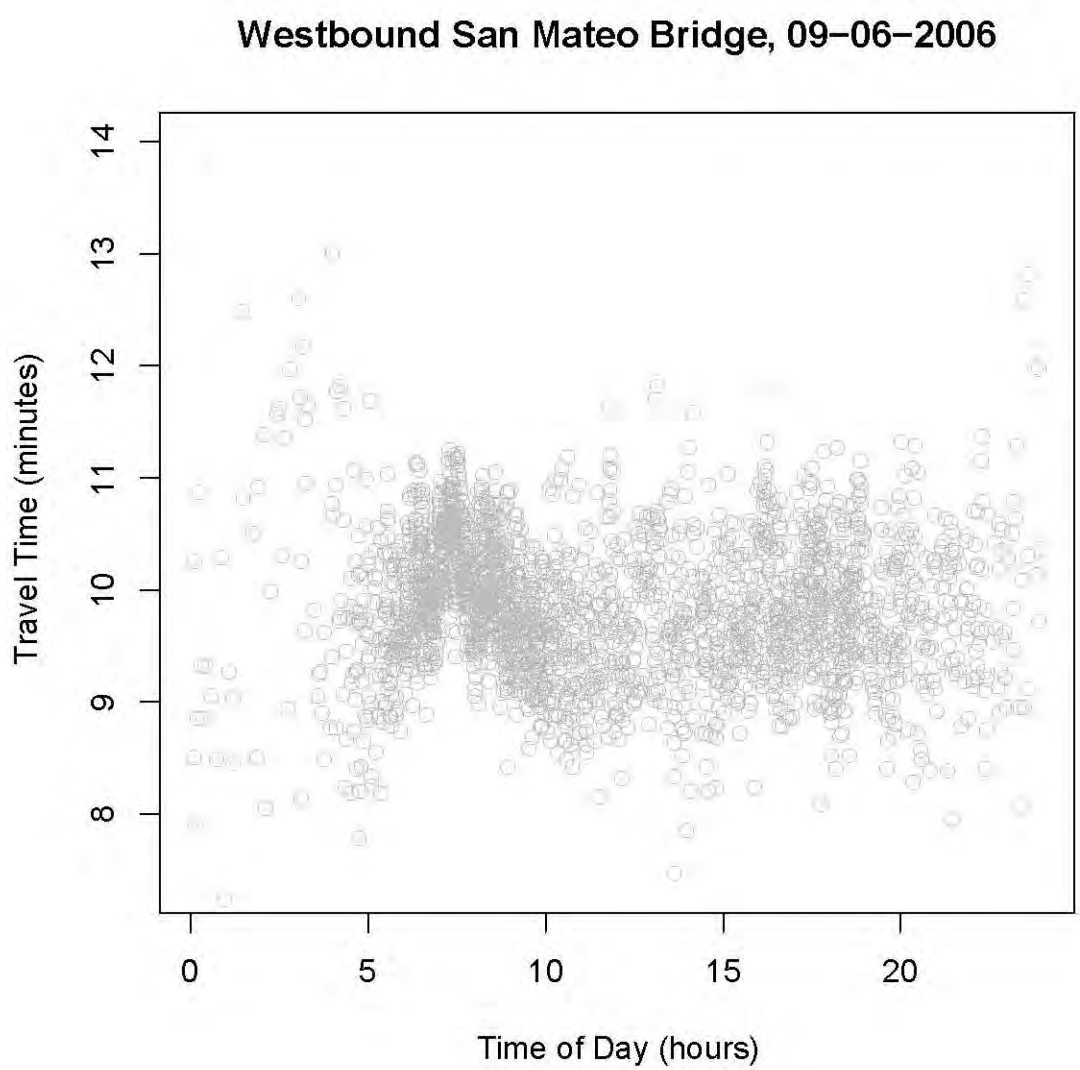

Westbound San Mateo Bridge, 09–05–2006
Travel Time (minutes)
8
9
10
11
12
13
14
0
5
10
15
20
Time of Day (hours)

FIGURE 1.5



Westbound San Mateo Bridge, 09−06−2006
Travel Time (minutes)
8
9
10
11
12
13
14
0
5
10
15
20
Time of Day (hours)

FIGURE 1.6

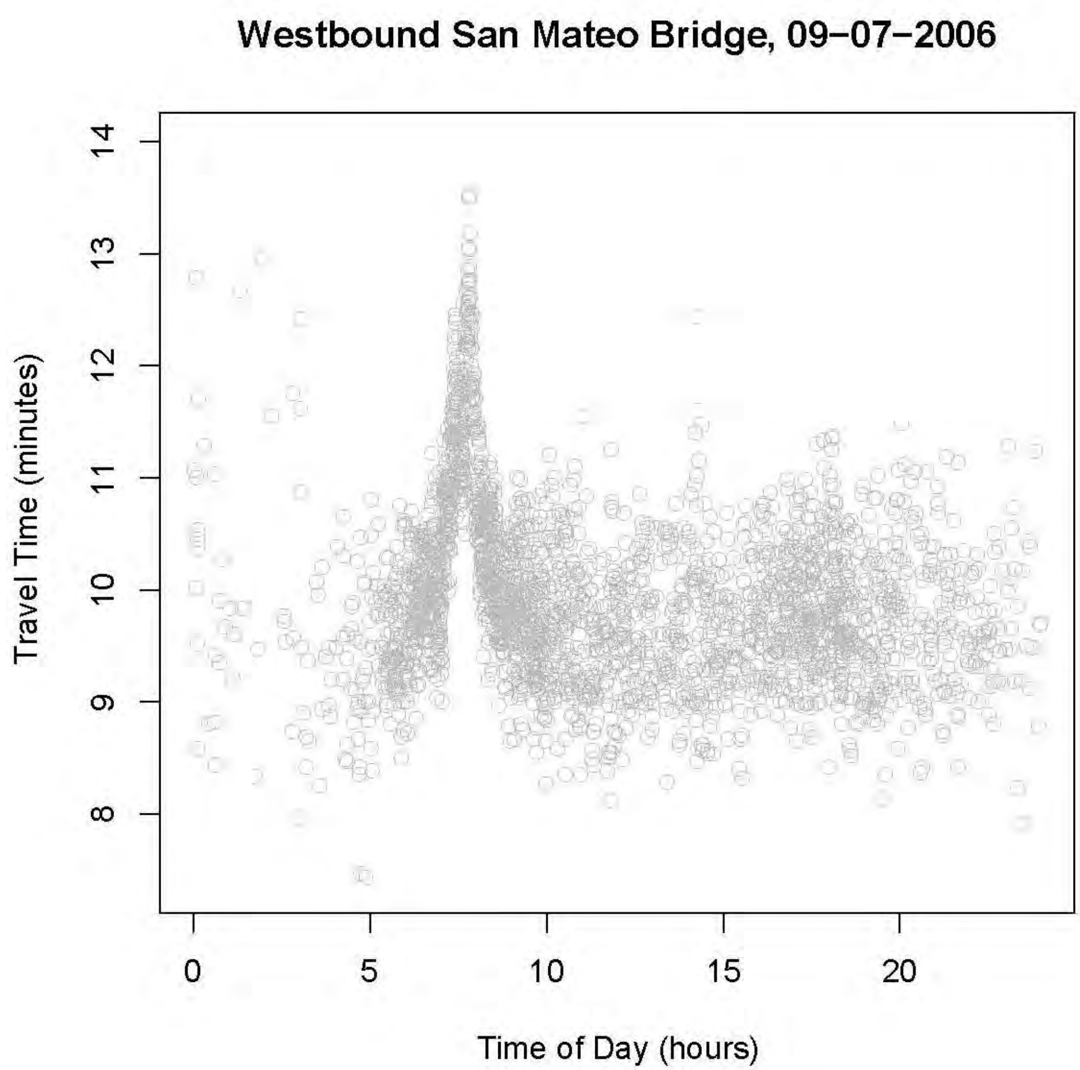

Westbound San Mateo Bridge, 09-07-2006
Travel Time (minutes)
8
9
10
11
12
13
14
0
5
10
15
20
Time of Day (hours)

FIGURE 1.7

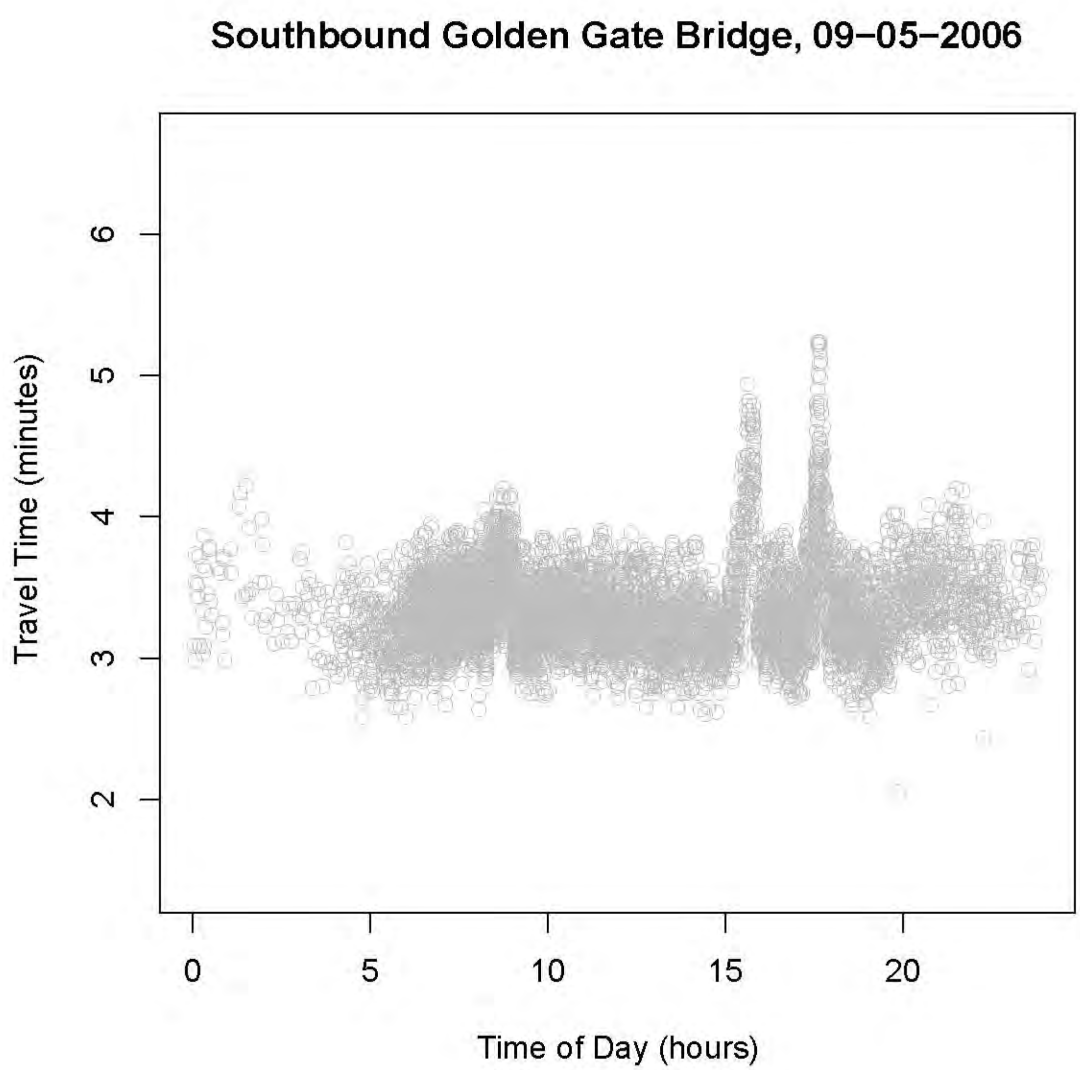
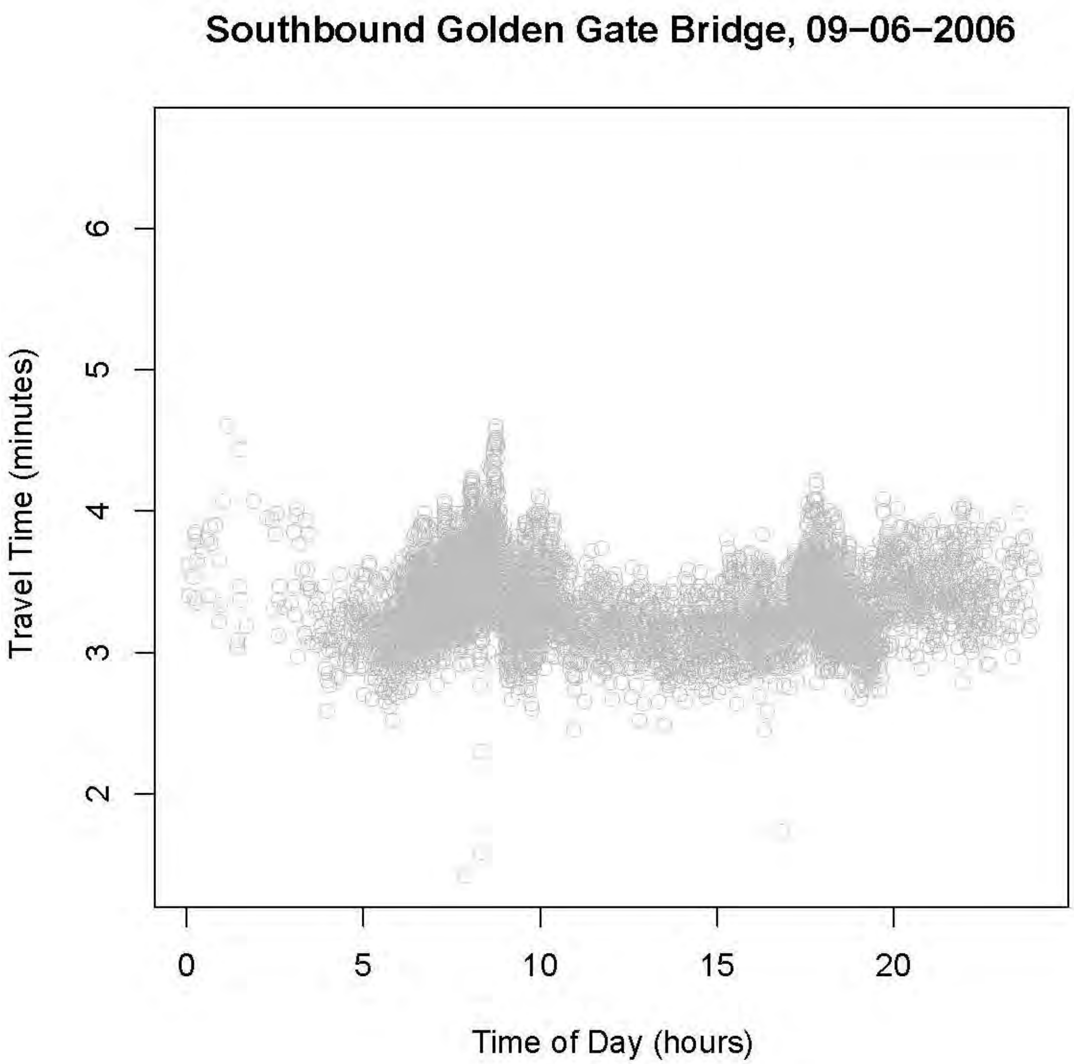

Southbound Golden Gate Bridge, 09–05–2006
Travel Time (minutes)
2
3
4
5
6
0
5
10
15
20
Time of Day (hours)

FIGURE 1.8



Southbound Golden Gate Bridge, 09−06−2006
Travel Time (minutes)
2
3
4
5
6
0
5
10
15
20
Time of Day (hours)

FIGURE 1.9

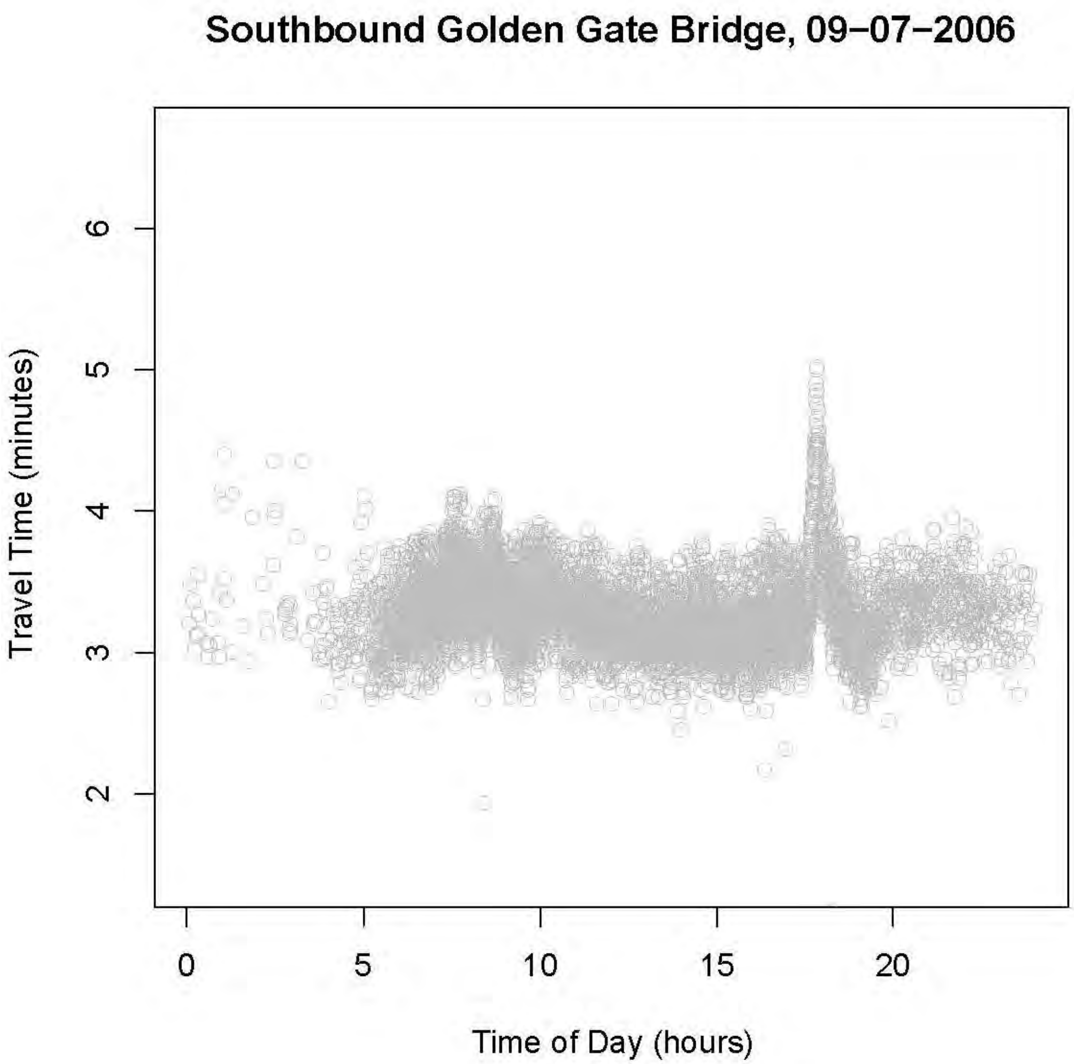

Southbound Golden Gate Bridge, 09−07−2006
Travel Time (minutes)
2
3
4
5
6
0
5
10
15
20
Time of Day (hours)

FIGURE 1.10

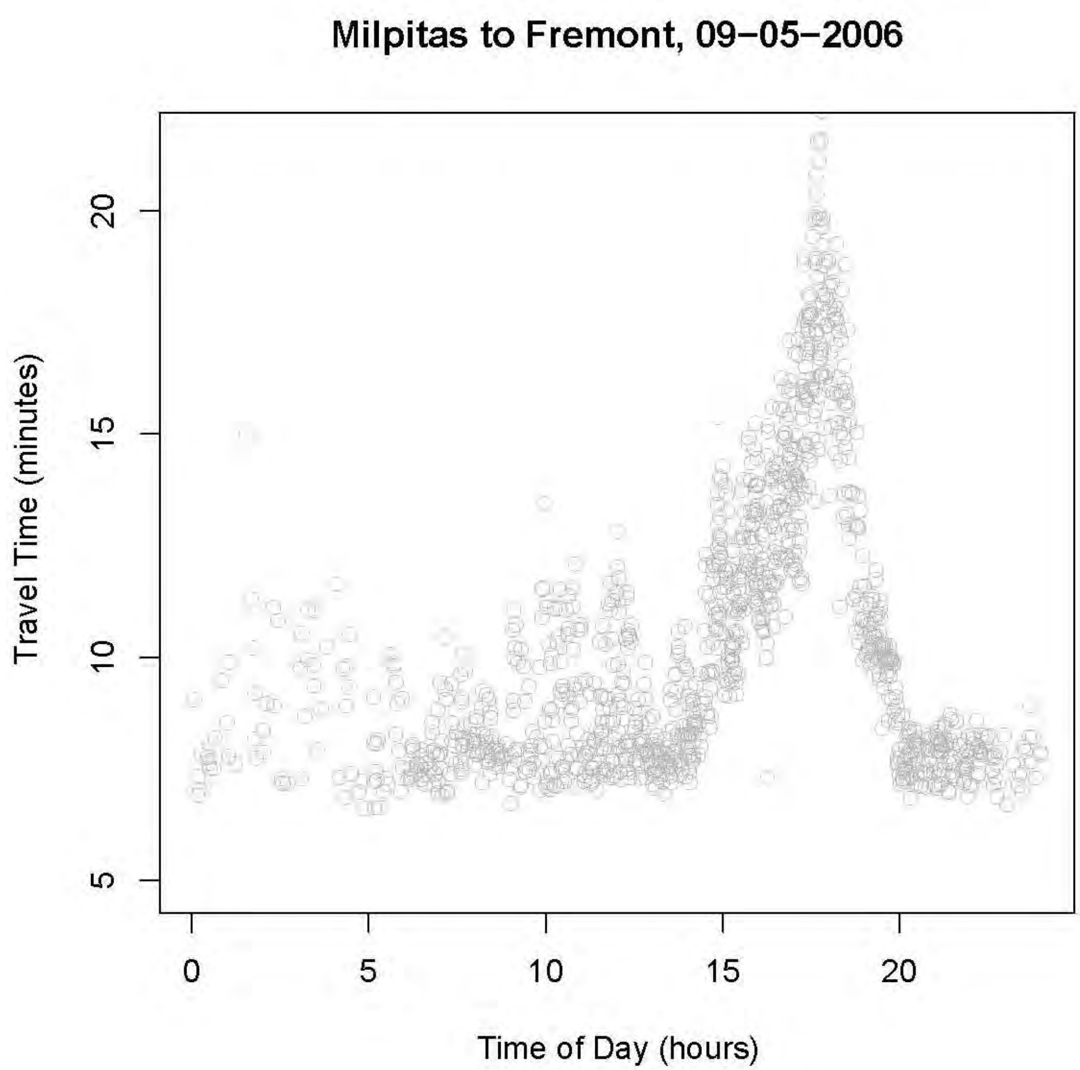

Milpitas to Fremont, 09−05−2006
Travel Time (minutes)
5
10
15
20
0
5
10
15
20
Time of Day (hours)

FIGURE 1.11

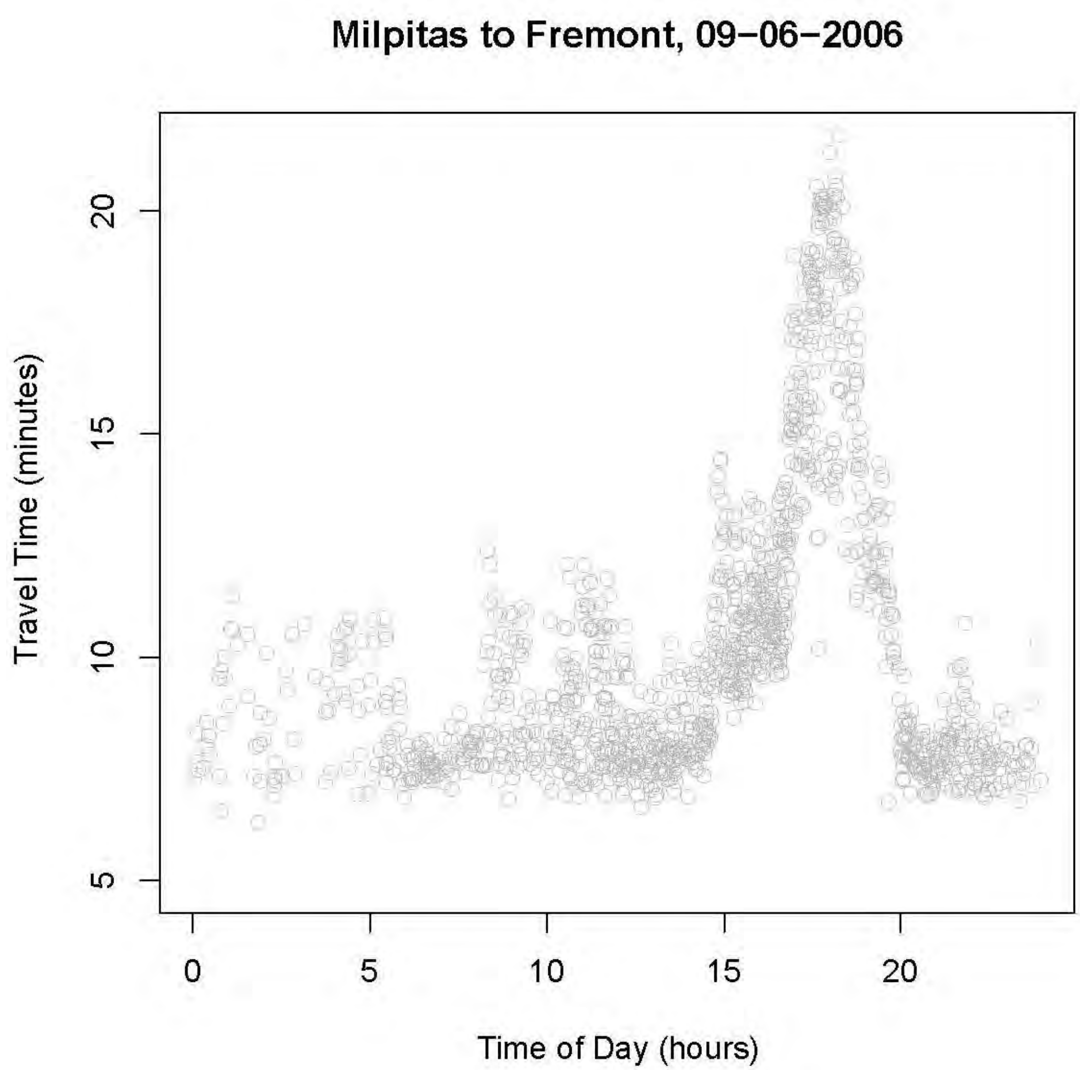

Milpitas to Fremont, 09–06–2006
Travel Time (minutes)
5
10
15
20
Time of Day (hours)
0
5
10
15
20

FIGURE 1.12

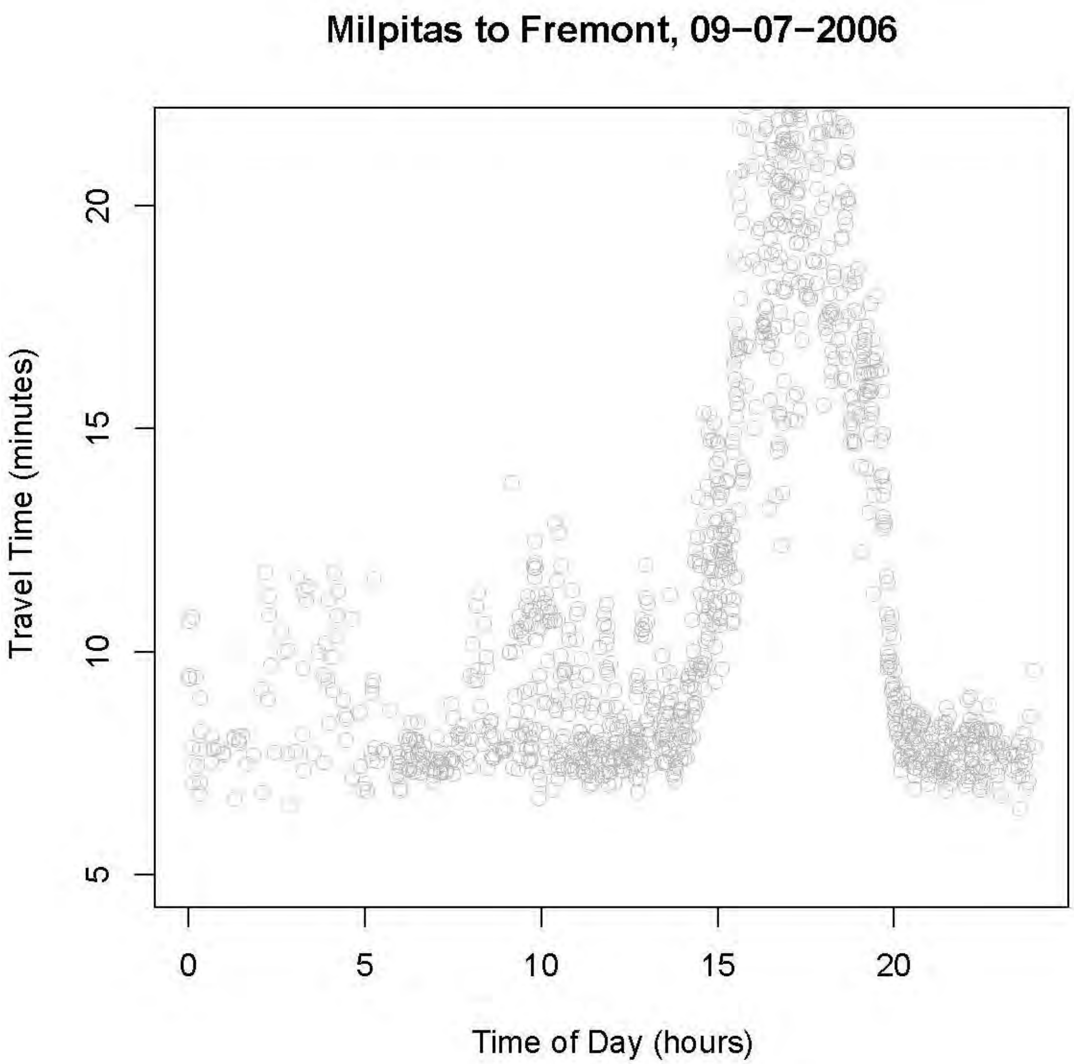

Milpitas to Fremont, 09−07−2006
Travel Time (minutes)
5
10
15
20
0
5
10
15
20
Time of Day (hours)

FIGURE 1.13

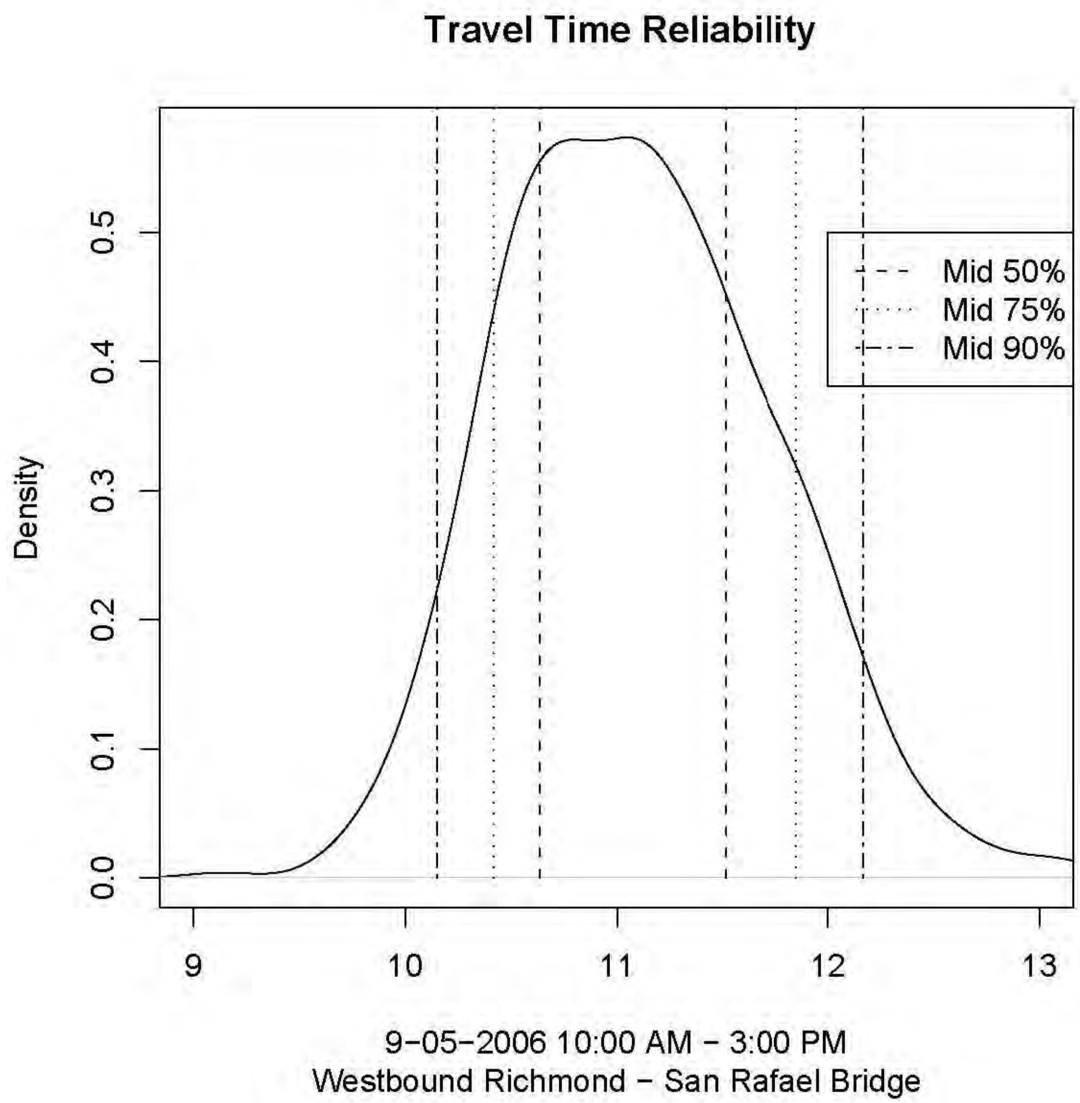

Travel Time Reliability
Density
0.0
0.1
0.2
0.3
0.4
0.5
9
10
11
12
13
Mid 50%
Mid 75%
Mid 90%
9−05−2006 10:00 AM − 3:00 PM
Westbound Richmond − San Rafael Bridge

FIGURE 1.14

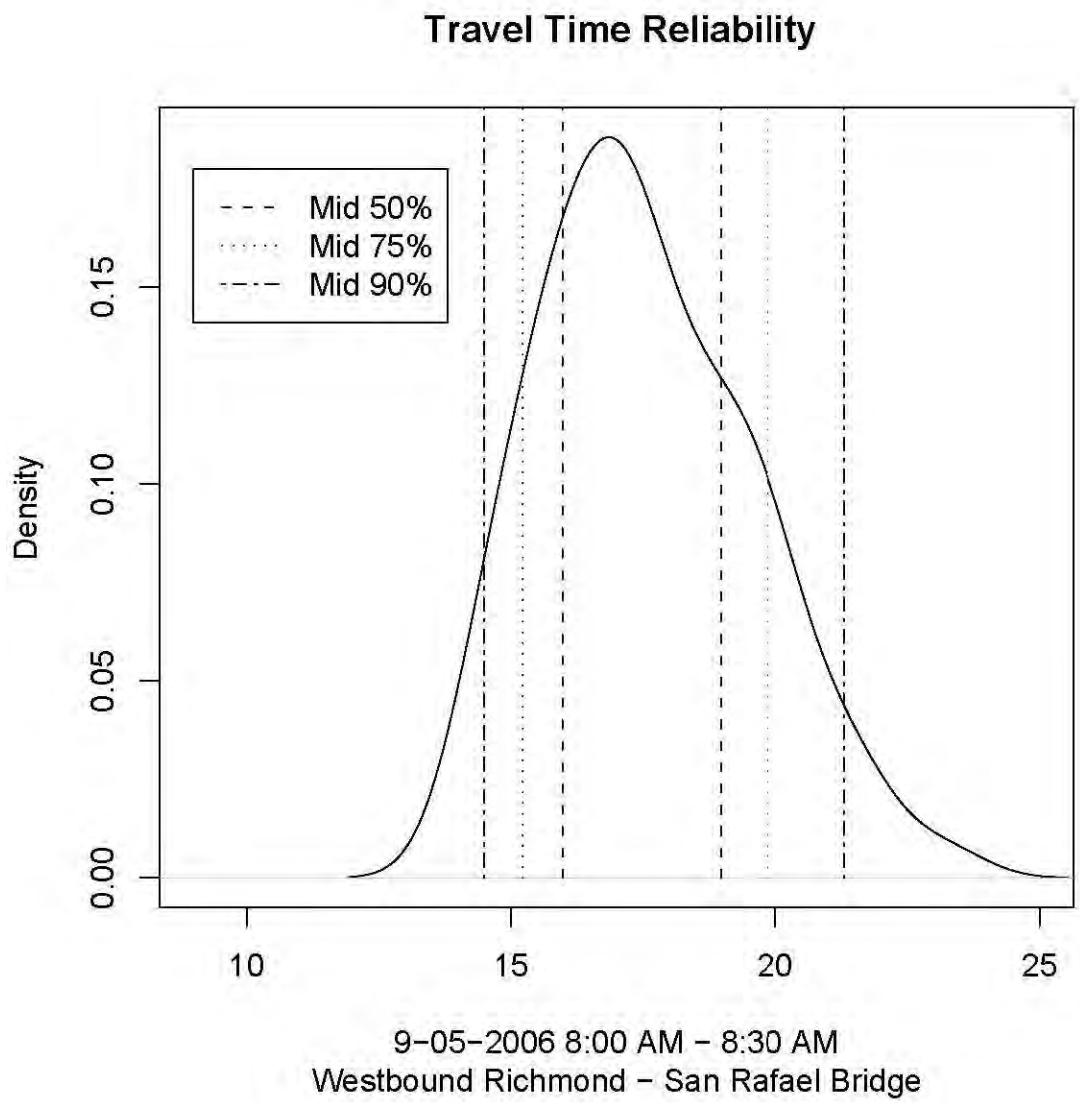

Travel Time Reliability
Mid 50%
Mid 75%
Mid 90%
Density
0.00
0.05
0.10
0.15
10
15
20
25
9-05-2006 8:00 AM - 8:30 AM
Westbound Richmond - San Rafael Bridge

FIGURE 1.15

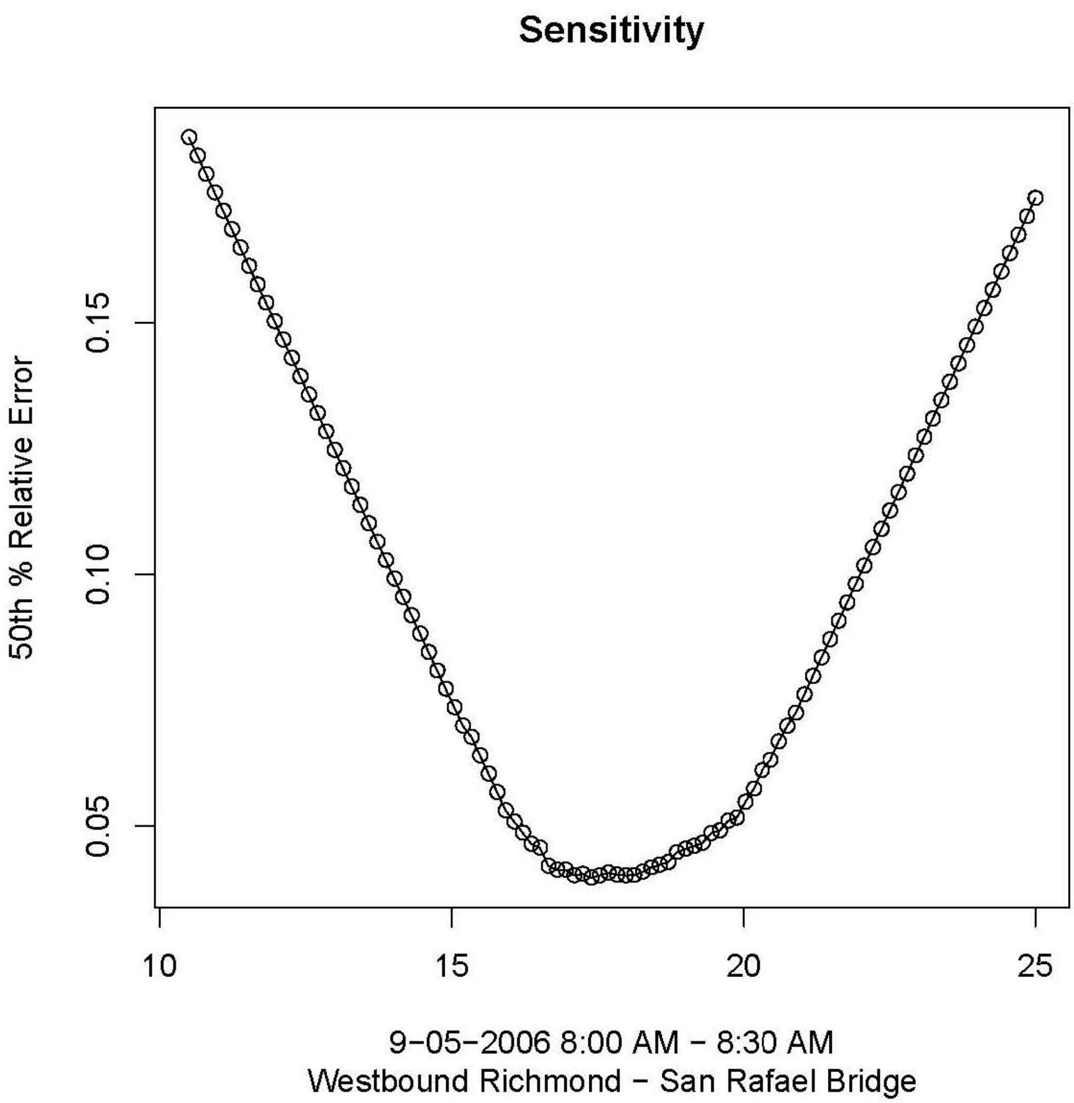

Sensitivity
50th % Relative Error
0.05
0.10
0.15
10
15
20
25
9−05−2006 8:00 AM − 8:30 AM
Westbound Richmond − San Rafael Bridge

FIGURE 1.16

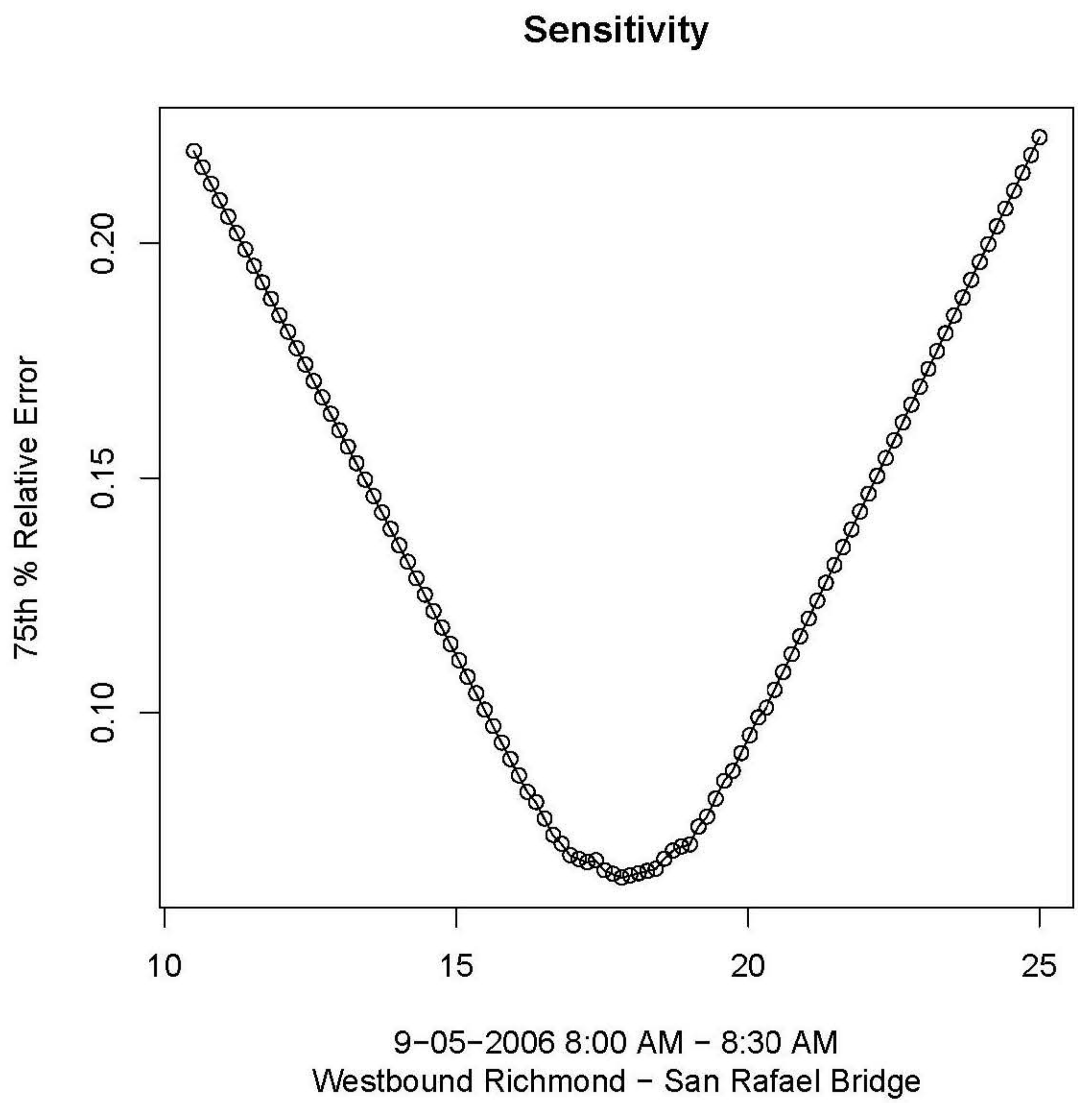

Sensitivity
75th % Relative Error
0.10
0.15
0.20
10
15
20
25
9−05−2006 8:00 AM − 8:30 AM
Westbound Richmond − San Rafael Bridge

FIGURE 1.17

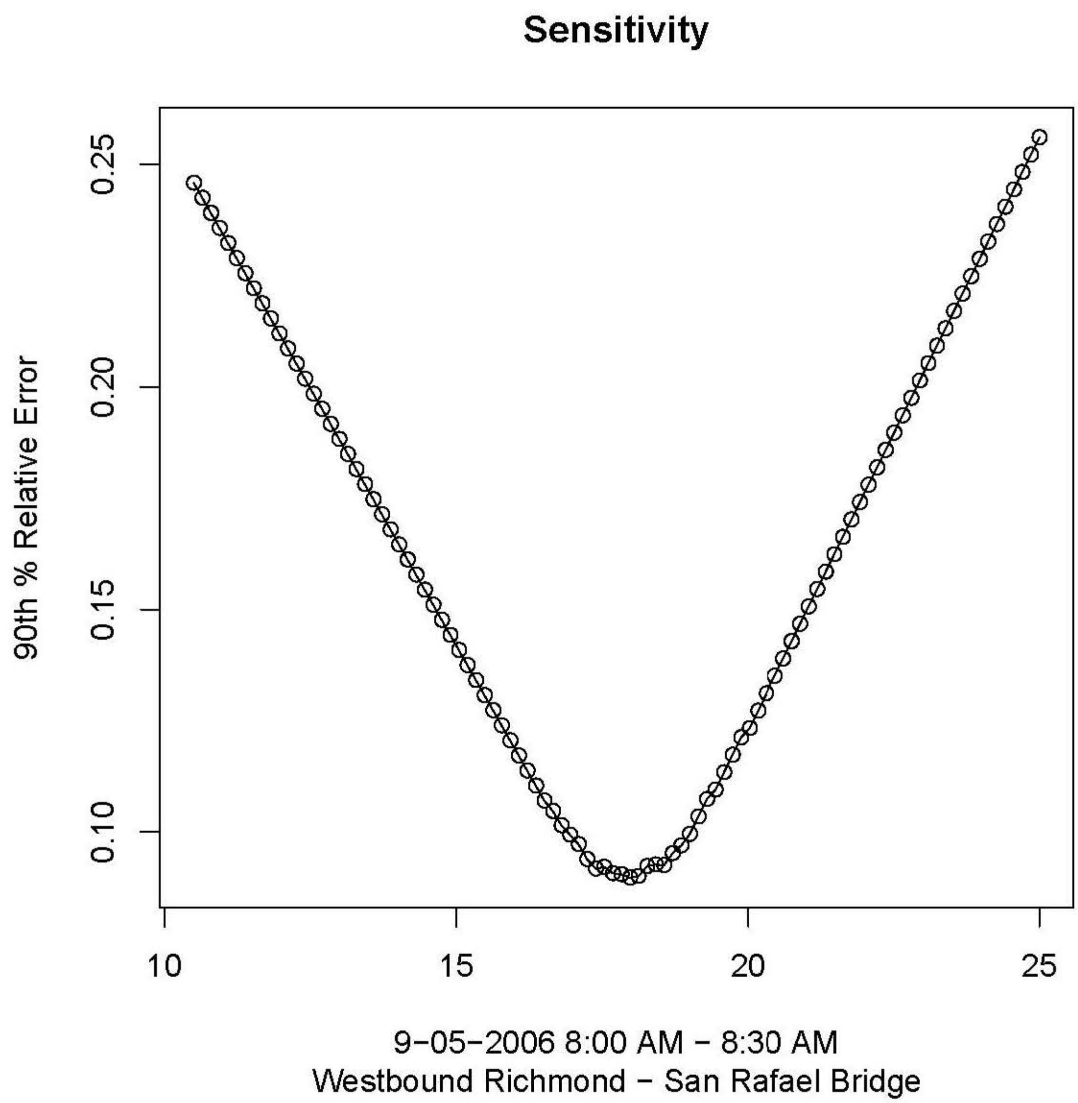

Sensitivity
90th % Relative Error
0.10
0.15
0.20
0.25
10
15
20
25
9−05−2006 8:00 AM − 8:30 AM
Westbound Richmond − San Rafael Bridge

FIGURE 1.18

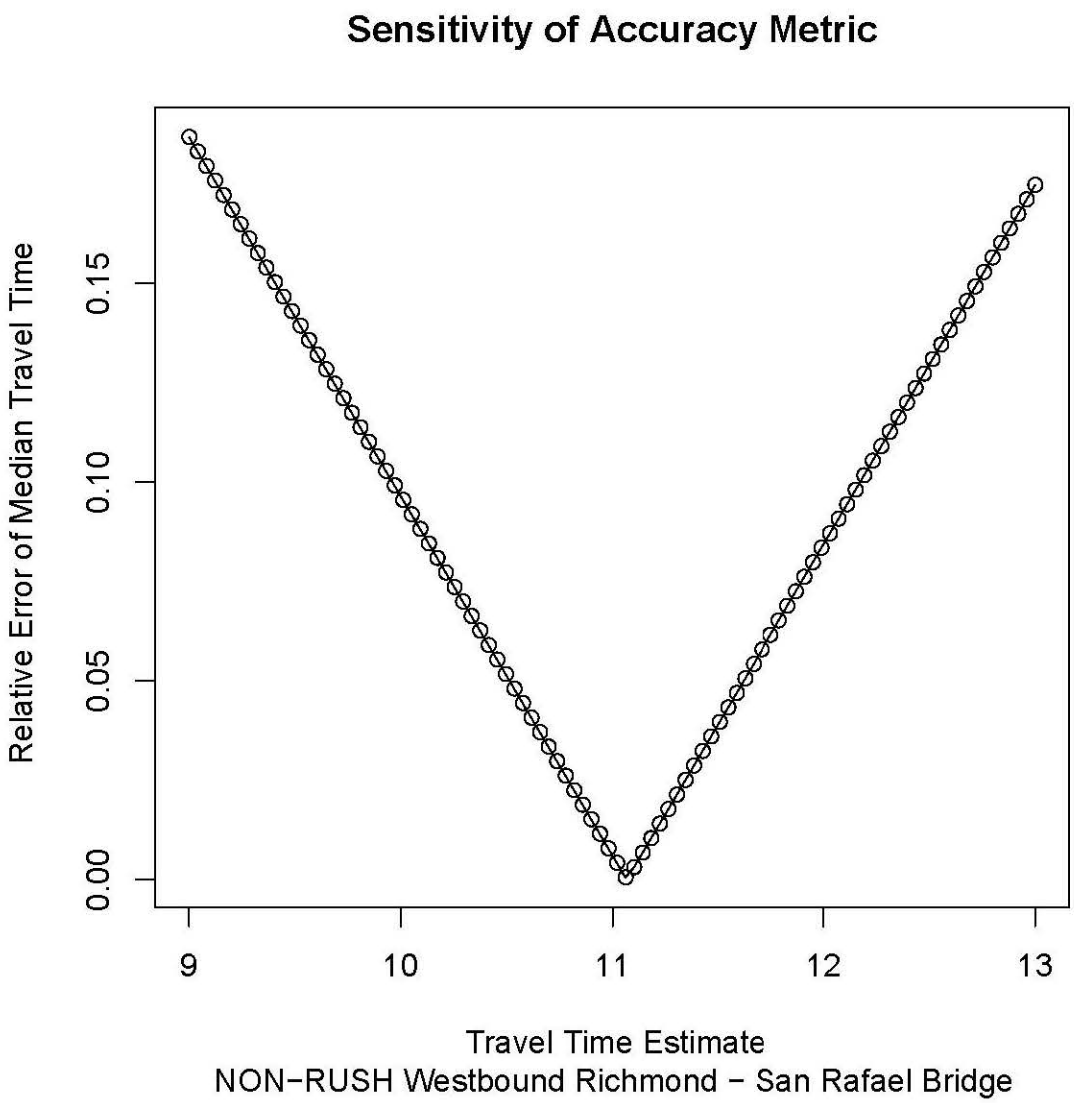

Sensitivity of Accuracy Metric
Relative Error of Median Travel Time
0.00
0.05
0.10
0.15
9
10
11
12
13
Travel Time Estimate
NON−RUSH Westbound Richmond − San Rafael Bridge

FIGURE 1.19

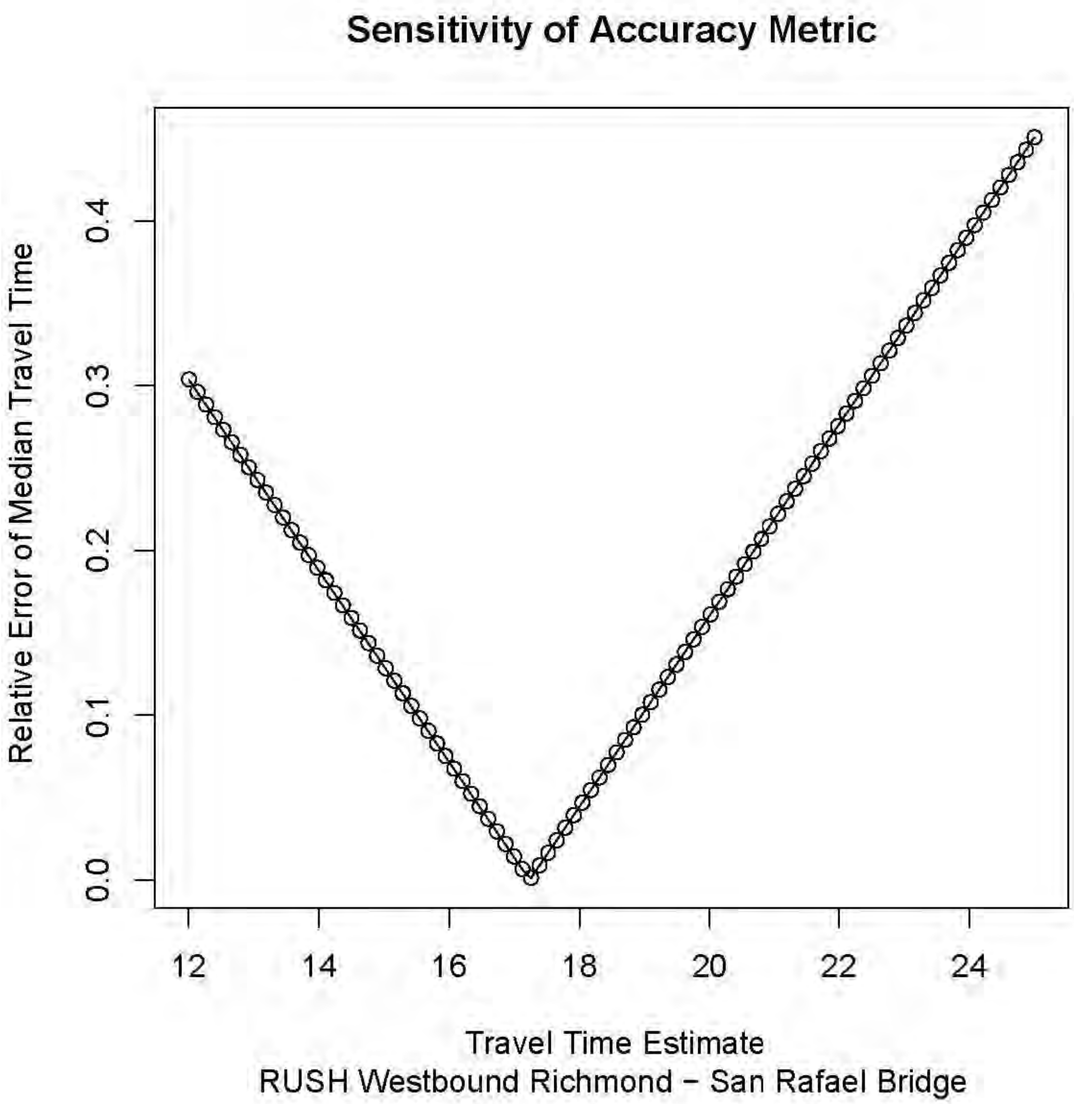

Sensitivity of Accuracy Metric
Relative Error of Median Travel Time
0.0
0.1
0.2
0.3
0.4
12
14
16
18
20
22
24
Travel Time Estimate
RUSH Westbound Richmond – San Rafael Bridge

FIGURE 1.20

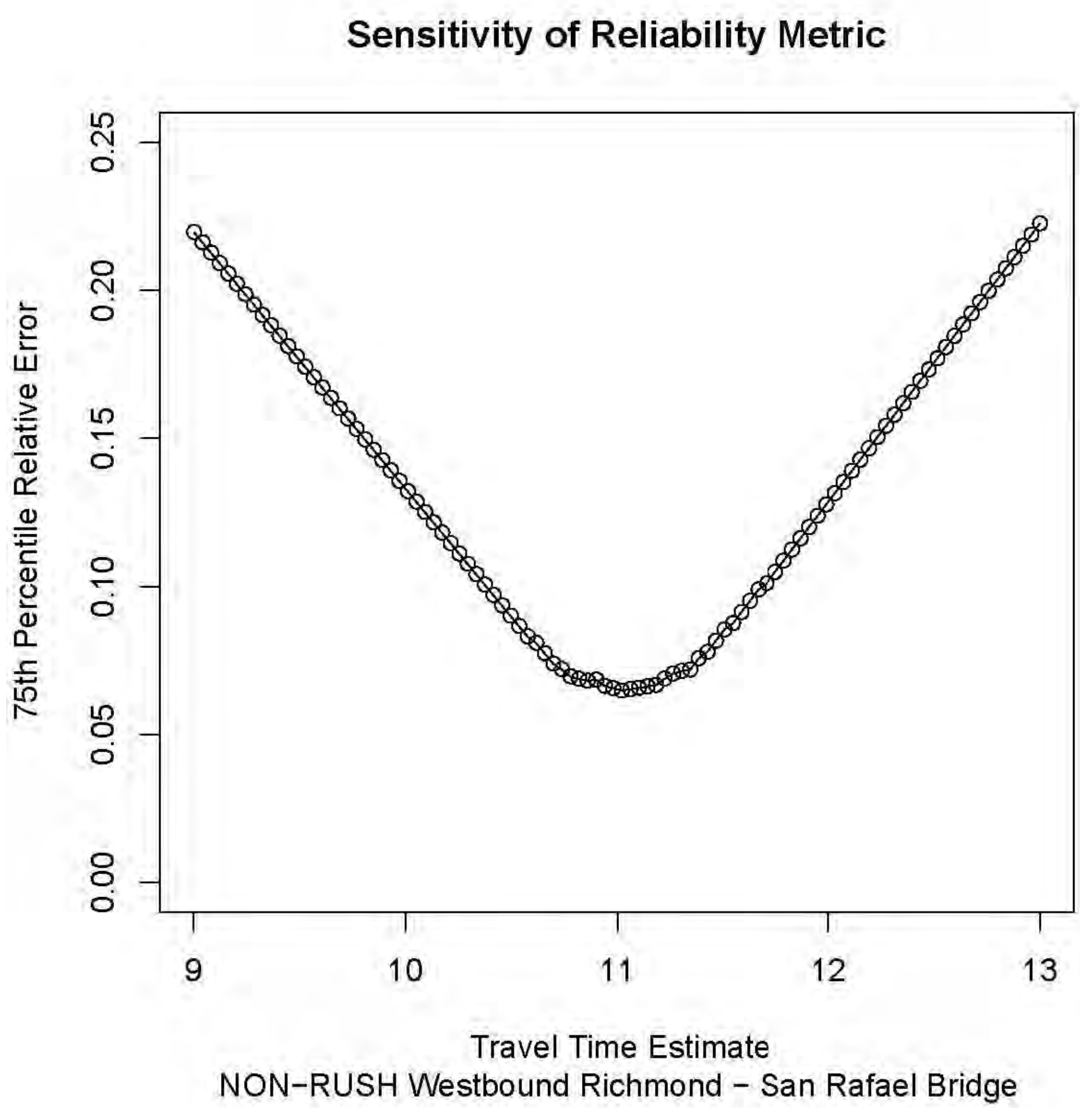

Sensitivity of Reliability Metric
75th Percentile Relative Error
0.00
0.05
0.10
0.15
0.20
0.25
9
10
11
12
13
Travel Time Estimate
NON−RUSH Westbound Richmond − San Rafael Bridge

FIGURE 1.21

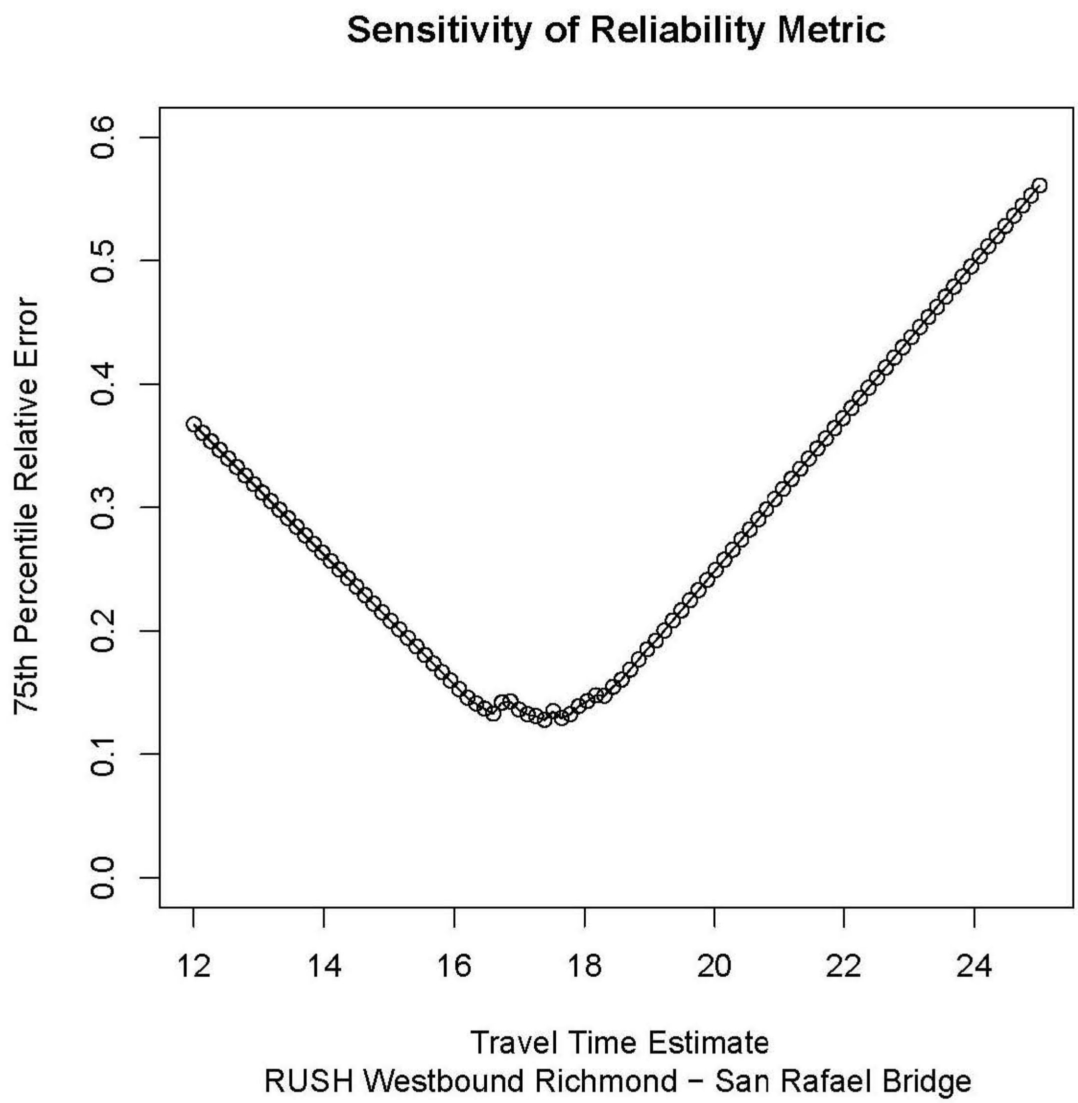

Sensitivity of Reliability Metric
75th Percentile Relative Error
0.0
0.1
0.2
0.3
0.4
0.5
0.6
12
14
16
18
20
22
24
Travel Time Estimate
RUSH Westbound Richmond − San Rafael Bridge

FIGURE 1.22

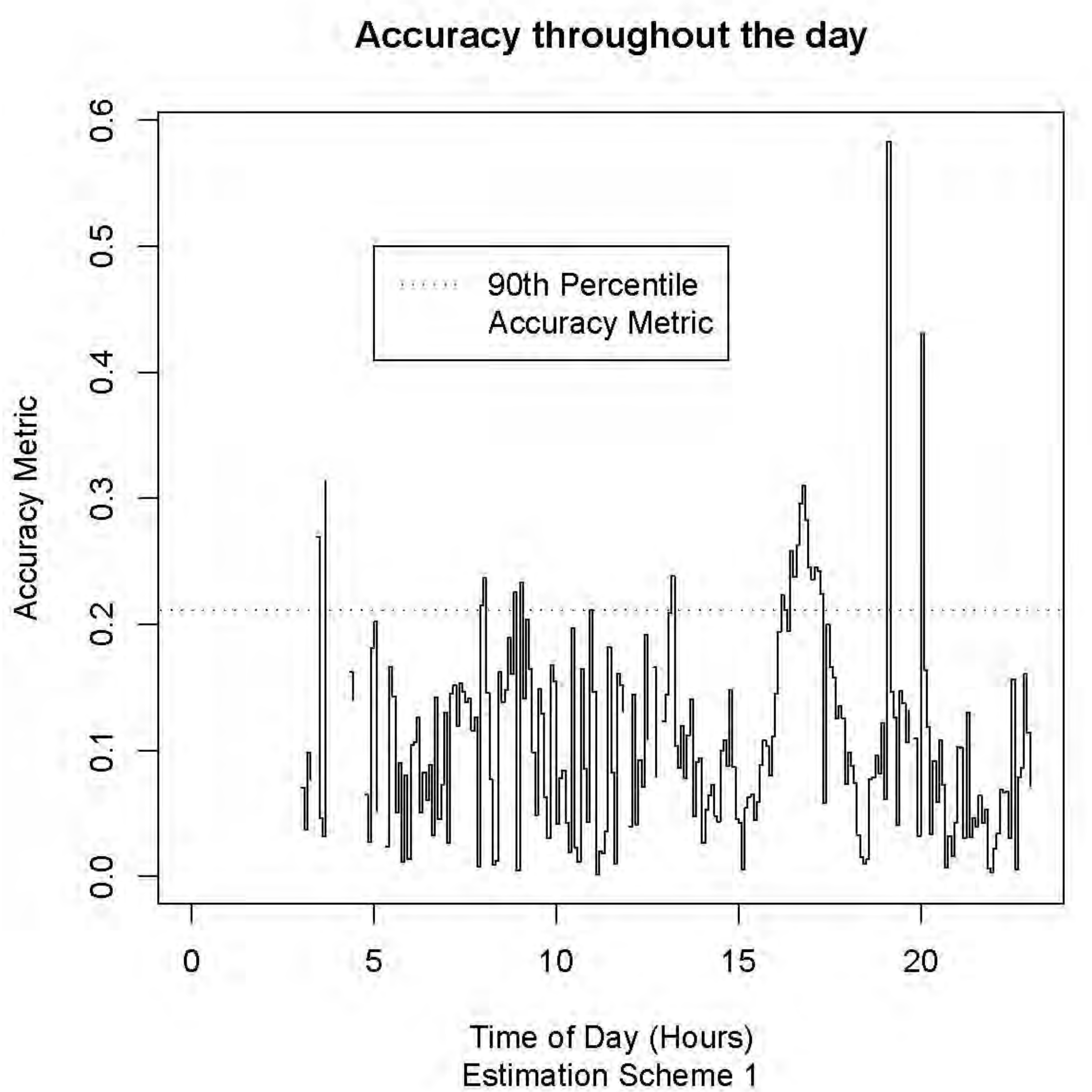

Accuracy throughout the day
Accuracy Metric
0.0
0.1
0.2
0.3
0.4
0.5
0.6
90th Percentile
Accuracy Metric
0
5
10
15
20
Time of Day (Hours)
Estimation Scheme 1

FIGURE 1.23

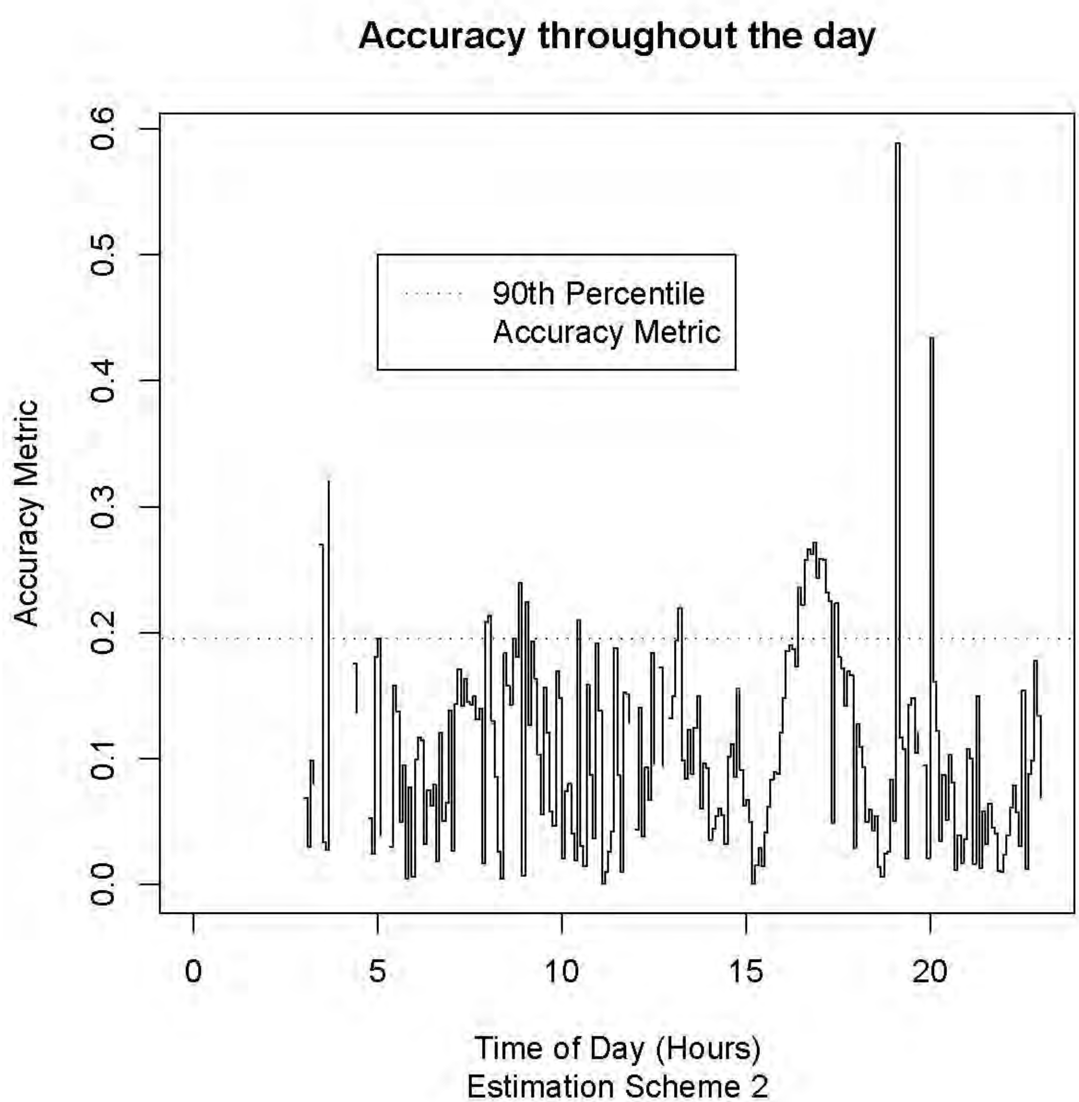
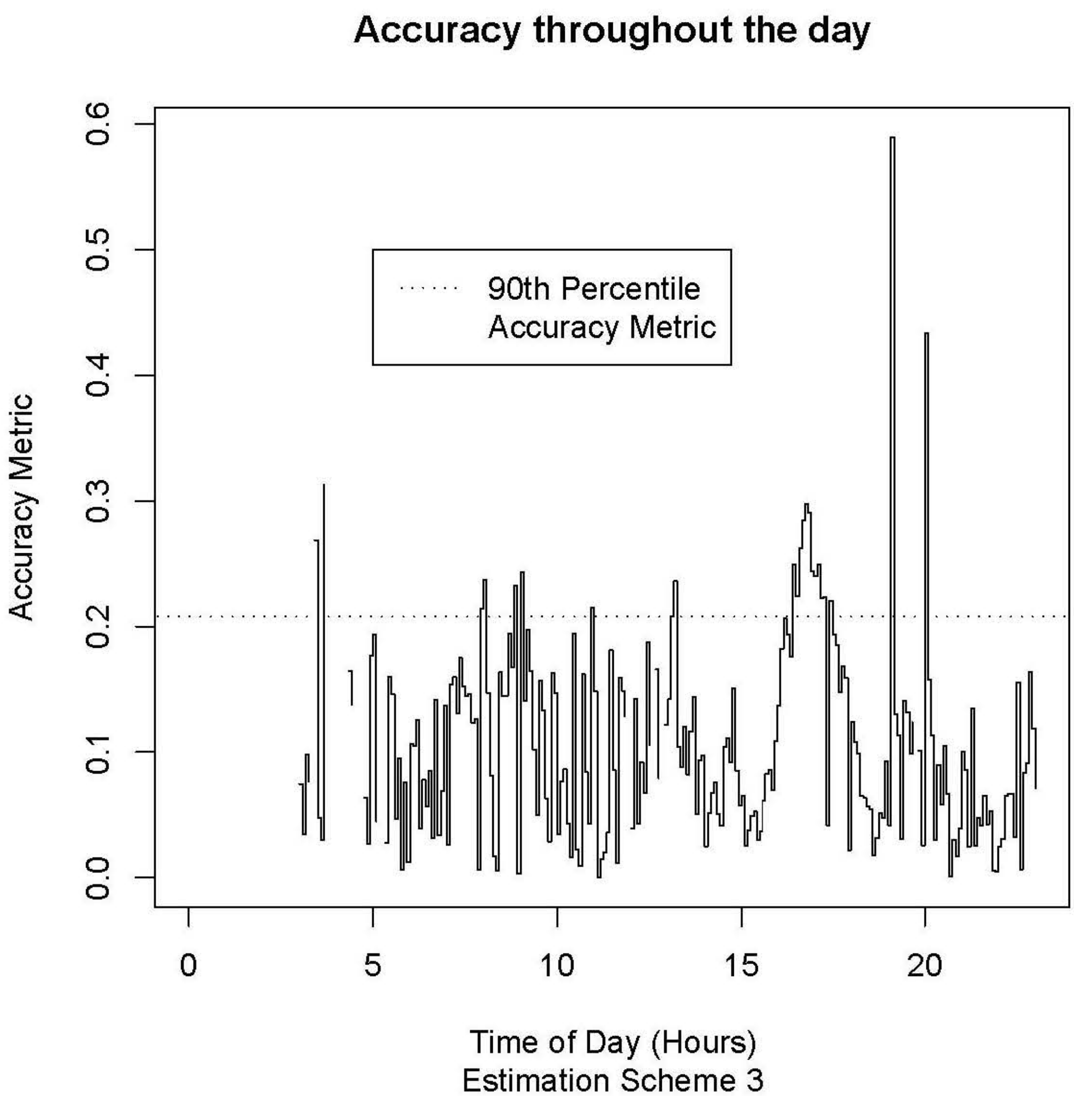

Accuracy throughout the day
90th Percentile
Accuracy Metric
Accuracy Metric
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
5
10
15
20
Time of Day (Hours)
Estimation Scheme 2

FIGURE 1.24


Accuracy throughout the day
90th Percentile
Accuracy Metric
Accuracy Metric
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
5
10
15
20
Time of Day (Hours)
Estimation Scheme 3

APPENDIX II: INTERVIEWS

DAVE WOLFSON – MARICOPA COUNTY – DAVEWOLFSON@MAIL.MARICOPA.GOV – 11/16/06

Dave Wolfson works for Maricopa Cnty, currently the ADOT spaces their sensors at 1/4 and 1/2 mile spacing, but is moving that distance up to 1 mile primarily due to rising costs, loop detectors are the primary method being used, there is not overlapping of detection technology, OZ Engineering handles much of their data processing and algorithms, Dave refered me to a gentleman there who can help answer additional questions.

Currently travel times are only accessable through the ADOT website, current message signs are not being used to display calculated travel times, and travel times are only being calculated in Phoenix and Tuscon.

DAN MANOR – EIS – 11/20/06

EIS: primary solution in the US in travel time prediction… traffic.com : 90% of their sensing requirements are provided by point detectors which are RTMS. Make use of all information available - data fusion is not that easy, but is being done:

Made the point: travel time is really travel time prediction, be able to predict… algorithms need to rely on volume. Changes in variablity is dependant on volume, and recurrent congestion.

Traffic.com - they operate 1000 RTMS units, don't have them every 1/2 mile, usually have them at 1 or 2 mile segments. Data is collected is every min (vol/occ/speed), based on "my" calculations, and use all three parameters in their algorithm, (Virginia Tech Professor) - Algorithms Knowledge - Antwoin Hobeika,

Other technologies:

floating vehicle/probe data - cell phone data. (tracking cell phones, toll tag readers)

Speedinfo: fundamentally not different from Wheelen: - all the rage 10 years ago, people figured out it doesn't work, Maryland, I-80 New Jersey: Motion Detector (doppler based microwave detector), - BUT, doesn't' do what radar does. Only thing they claim is average speed, not volume, occupancy, or per lane information), to achieve doppler, parallel to road. 20-30degree parallel from road.

Problem: motion detector: doppler signal doesn't work if speeds less than 10 mph. (if at congestion, it doesn't see any vehicles, parking lot looks like empty road), other problem is that beam cross multiple lanes, even in moderate traffic, multiple vehicles inside the beam, all in same footprint. In that circumstance, no way that motion detector can decompose doppler signial from multiple vehicles, unless you take a) stronger signal (truck will win against others) b), fastest moving vehicle (e.g. HOV lane), the speed you get is not the average.

Speed by itself is not sufficient information about prediction.

Atlanta: took autoscope (3 lane) 1/2 mile spacing, all they do is provide average speeds, speeds per lane and averages property,

Chicago with loops, developed segment travel times by average speed…. Summing up travel time. (this is both and chicago and atlanta)

Loops - give occupancy.

Video - not as good at getting occpancy, effected by shadows, footprints can be extended by shadows, night also has problems.

Efficiency of Video: found niche in stop bar detection in control, no new highway projects with video. Specify with radar now.

Loops have good weight, staying power. Various algorithms of re-identification.

Problem with toll readers, license plate readers: measure exactly and directly travel times through a segment. However, there is a lag…no volume.

Hybrid Technology is needed: toll tag readers, used sparingly, not for full blown system. Close the loop on algorithm development. Should not be the main thing.

BOB FAIRBURN – EIS – 11/9/06

Work directly with customer, represented by two organizations through distributors. Rely on distributors locally for support/ presale and postsale. (Distributors are: Bill McDonald (McDonald Engineering in Livermore), Western Pacific Siginal (San Leandro). Distance is location specific, not necessarily. Transportation authorities specify distances. More interested in specific locations versus distances.

Detection, Communication of Data, Intrepretation: FTMS - Does the Intrepretation, (Mcmaster Algorithm) - designed to determine incidents versus false detection. NEWS - some interpretative elements of travel time prediction. Predictive Travel Times are often calculated using different systems but only using EIS data.

9000 hr MTBF (mean time between failure)

Within EIS: contact Dan Manor (Company President)

EIS radar technology is much more robust, climate, etc… video (video has more mainteance), Detection is much more challenging. (low power consumption, 4.5 watts of power consumption, rugged and durable unit platforms.

MARK DEMIDOVICH – GEORGIA DOT – 11/9/06 (EMAIL)

1. Rationale for Highway Sensor Spacing: We space our sensors at 1/3 mile intervals. This spacing was selected back in 1996 when we first deployed the Navigator system. At that time, the rationale for 1/3 mile was mainly for incident detection, not trip time calculation. It was felt that 1/3 mile spacing was best for quick detection of interruptions to traffic flow. We still use this spacing today. I also find it very good for trip time calculations. I think going up to ½ mile spacing would be ok too, but much more than that, I think you are risking too much space between sensors that could “hide” areas of slow traffic. Of course, this applies to point sensors only. If you are using “probe” type sensors, you can probably use much longer segments, because you are measuring the actual trip times, not point speeds.

2. Algorithms for Travel Time Calculation (With One or More Technologies): All of our sensor data, which come from a variety of sources, are placed into a database of “stations.” A station is basically a location along the roadway (in one direction), that is an aggregate of all lanes at that location. Those speed data, which are basically averages of all the lanes over the past x seconds (x usually = 20 seconds, but can depend on the source), are fed into our trip time algorithms. Algorithms are plural in our case because we have one for our website and another for our dynamic message signs. The website algorithm sums the trip times through each 1/3 mile segment over a longer corridor. So… we take the speed that we get at each station (and assume it is constant for the whole 1/3 mi segment), divide it over the length of that segment, and get a trip time for that 1/3 mi piece. Then we add up all the 1/3 mile segment trip times and come up with the corridor trip time. Our “corridors” are actually selected by the user – they pick a start and end point.

For our message signs (which display messages like “TRAVEL TIME TO MAPLE ST IS 14-16 MINS”), we use a slightly different algorithm. We break the corridor of interest into two “zones.” Then we calculate the average speed of all the sensor stations in Zone 1, and all the sensor stations in Zone 2. Remember that stations are the averages of all lanes in one direction. Then we go to a pre-established look-up table (which is unique to each sign and is 4 rows by 4 columns or 16 cells/records) and find where the average speeds of Zone 1 and Zone 2 fit into the table. Behind each cell is a corresponding message, which is then displayed on the sign. We use 2 minute ranges to give ourselves a little leeway. We update our sign messages once a minute.

3. Traffic Data/Travel Time Performance Metrics As with most things DOT, we use public feedback as one of our strongest indicators of accuracy. If we get few complaints, we assume the system is working well. If we start to get complaints, we will investigate and “tweak” the system as necessary. Additionally, when we first start displaying trip times on a new message sign, we test them rigorously using our cameras. We pick a brightly colored car on camera and begin a stopwatch as the car passes under the sign in question. Then we follow that car through the corridor, from one camera to the next, and measure the total amount of time it took to traverse that segment. We cross-check that number with the number our sign was displaying.

Trip times are also becoming a major performance metric for GDOT in the area of congestion. Our Board has adopted a policy that “success” is having a peak-hour trip time index of no greater than 1.35. Trip time index (or TTI) is simply the peak hour trip time divided by the free-flow trip time. A peak hour trip time of 40 minutes that should normally take 20 minutes (in the middle of the night for instance) would have a TTI of 2.0. I think trip times were selected as a performance measure because it is a simple concept for everyone to grasp and relate to (as opposed to trying to tell the public that the volume/capacity ratio is x or Level of Service is y).

4. Public/Vendor Feedback The public has been very supportive (if not demanding!) of trip time information. We were one of the first metro areas to display trip times on message signs back in 1998. Trip times seem to be the one thing most people can relate to, because there is little subjectivity (as opposed to a message that might say “Traffic Moving Well Next 10 miles” or “Traffic Normal” – what does that mean? “normal” for rush hour?) When for some reason our signs or website are not giving trip times, we get lots of feedback.

5. Pros/Cons of New/Existing Sensor Technology In Georgia, we use primarily video-detection based sensor systems. These are made up of black and white cameras mounted 70-80 feet above the road, whose video is fed into processors. These processors can “read” the cars as they pass through the video image. The output from these processors is vehicle speed, length and counts. In addition to video detection, we use about 15% radar detection, specifically the EIS brand (RTMS). These units supply basically the same output, but without using video. Video can be subject to errors with bad weather like heavy rain or fog, and shadows, but we haven’t had too much of a problem. Radar systems also have their own challenges, such as seeing over barrier walls and getting false returns from stationary objects nearby. Lately, we have been exploring using probe technology, specifically cell phones. Two companies (Airsage and CellInt) have contracts with us to provide trip time data on various roadways that they have gathered from cell phones. We are just now beginning to receive this data, so it is too early to deem whether it meets our accuracy requirements (we generally like errors no greater than +/- 5-10% from actual). Probe data is also missing a key element that other sensors provide – and that is counts. However, we don’t feel it is necessary to get counts AND speeds every 1/3 mile. We can probably live with much fewer count stations, as long as we are getting good speed/trip time data along a corridor.

BRIAN KARY – MINNESOTA DOT – 11/22/06 (EMAIL)

Your question on Travel Times was passed on to me. Give me a call if you have any more questions.

We started using travel times in fall 2004. We were able to utilize our existing dynamic message signs and our existing loop detectors to calculate and deploy the travel times. The system was developed in house.

Generally speaking the loop detector stations are in place along the mainline at about ½ mile spacing. Speed data from the detectors is calculated every 30 seconds. Using the speed data from the detectors and then the distance between detectors we are able to calculate the travel time and deploy that time on our existing DMS.

To improve the accuracy of the system we gathered GPS coordinate data so that we could accurately determine the distance between detector stations. We also had to calibrate our detectors to make sure the speed data collected was accurate. As the travel times were deployed along a corridor, we had staff drive the corridor and test the accuracy of the travel time calculations.

We also capped the travel time during freeflow conditions so that only the travel time for vehicles traveling the posted speed limit was displayed. During extreme congestion when speeds are so slow they are difficult to calculate we will default to a maximum travel time and display a message such as " over 30 minutes." To minimize errors we only give travel times to destinations 12 miles away or less.

The feedback we have gotten from the public is very positive. The public likes the additional information and finds the signs to be very accurate.

1/2 distance: original system, (1972) - ramp meters, mainly for accurate volume counts. (300 miles of system), (calibration of detector)… (avg vehicle length), in house system: (IRIS) : software system that operates all field equipment (module within IRIS), fairly accurate. To keep error rate, (6-12 mile destination points)

RICK ZIBINSKI – NY STATE DOT – 12/6/06 – 518-457-2516 – RZIBINSKI@DOT.NY.STATE.US

Planning on doing some travel time estimation on long island (inform traffic management system), (track a vehicle between two points), have been using transit system, have been using toll-pass, (downstate), easy pass readers along highways.

No forecasting, reporting current travel times, taking the average travel time from the toll tag readers, plan on giving range (4-6 minutes). Toll tag readers, loop detectors, radar, acoustic, all available: not using these other technologies to calculate travel times.

A few names:

- TMC Long Island - Emilio Sosa 631.952.6781 - esosa@dot.state.ny.us
- Transcom: (serves NY area, acts as a clearinghouse 16+ transportation system operators in NY Metro) - allows shared data: processes travel data, Tom Bats: 201.963.4033 - batz@xcm.org
- Hudson Valley TMC - Bob Rella : 914.742.6010 - brella@dot.state.ny.us

PETER JONES – PIPS TECHNOLOGIES – 11/7/06

Citilog- Philidelphia - incident management software (tunnels and bridges),

PIPS: area of travel time : focus on travel times in work zones/temp work zones. Two permanent areas (were work zones), fairly expensive technology (10-11K per camera), I-81 Corridor (18 mi segment) - Virginia, Tampa-FL Area.

Idea for work zones, trailer mountable, speed calculation, enforcement capture 33/35% percent of license plates for accuracy, as low as 10% of plates +/- 2 mph. Primary use : law enforcement and toll enforcement.

Max Corridor Length: - Not too much distance between each node, (lose plates, typically 1 mile, two miles), very situational, closed system, one beginning, one exit, If there are nodes that exit, there would have more.

Example: 24 miles bridge, lag calculation is included, maybe camera is included in between if links are two long, no traffic, system intergrators: 3rd party or jointly: deciding factor on Lag Time: transmit wireless several 100 yards from message sign, server may be vicinity.

Portable System (Quixote Transportation Technologies): ALPR cameras mounted on portable solar powered camera, immediately deployable.

TOM SUTULA – RENO A&E

Primary customers - sell through distribution to different municipal state agencies through distribution. If there is a need, we try to provide a solution: Orincon (defense contractor) - District 5, studies of being able to track trucks, types of vehicles coming up from border. Loop detectors that can provide signatures of vehicles, identification of vehicle by inductive signature.

GREG PIEPER – SMARTEK SYSTEMS

In area number 5, I can give you the pro's and con's of existing technology as I know it.

Non Embedded sensors have all the advantages of not being in the road bed that comes with this attribute for installation and maintenance. Among them are sensors based on video, radar and passive acoustic technology. Active acoustic and infrared are available as well. They share many of the attributes of the radar and video image processing technologies, but they have not made a serious foray into the US market. Most non-embedded sensors, if not tied to a legacy relay interface, are remotely re-configurable and do not require an expensive cabinet controller, and may have wireless home runs.

Video sensing -- pro: High zone density, video zones easily correlated by user, especially during the day because the sense of sight is most developed, usually mounted high enough that occlusion is mitigated to a large extent, per vehicle speed calculation usually based on speed trap like zone to zone calculations independent of vehicle length. Re-configurable, and depending on height, may detect 8 or more lanes. Limited vehicle classification available. Volume, lane occupancy or headway, speed are standard data stream items if a relay interface is not used. -- con keeping lenses clean, objective windows clear (winter months require heaters and blowers in north climates which use a lot of power. Must be an AC site most of the time.) Day/night transition, light bloom on wet pavement, fog, dust storms, usually must have a more stable mounting platform than other technologies, must be mounted relatively high for most accurate results. If the processing is done in the camera housing, a lightening strike may result in a more costly replacement than if the camera is separated from the video processor in a road side cabinet. If a data stream is all that is required, then it may be cheaper to do video processing in the camera housing and risk the lightening hit.

Radar -- usually cheaper than video and less prone to weather effects, large established base and acceptance within the industry, very accurate speed when used to face a low number of lanes from an overhead position (much more accurate than the side fire location), relatively maintenance free compared to cameras, more lanes than passive acoustic. When used as a side fire device, gets most lanes and relatively good speed info. Newer models have longer MTBF and not maintenance once installed. Re-configurable, may detect up to 8 or 10 lanes. Limited vehicle classification available. Volume, lane occupancy or headway, speed are standard data stream items if a relay interface is not used. Cabinet controller not required. -- must get a good return from the sides of vehicles for lane differentiation, so low mounting is better than mounting high where lane resolution is lost, but this results in occlusion issues. (Low mounting height lends to far lane occlusion issues in areas with high truck volume in lanes nearest the sensor.) Analog versions have difficulty with barriers in the middle of the road because range gate issue and more difficulty in separating vehicles in far lanes because of zone spreading away from sensor. Speed calculations based largely on an average vehicle length. An active, radiating sensor, it uses more power than some video, loops or passive acoustic, so larger solar panels are required if AC is not available. There are relatively cheap radar detection systems that only provide an average road bed speed based on samples taken across all lanes. (not lane by lane speeds based on individual vehicle speeds being sampled.) These are deployed using using relatively small solar panels, but these units do not continuously sample the road bed to conserve power. Tests in Northern VA were not seen as completely satisfactory. I do not know the test results in other areas, including California.)

Passive Acoustic Detectors -- Like video PADs are mounted higher than radars to mitigate occlusion. Mounting structures do not have to be as stable as for video. Like video, per accurate vehicle speed calculation independent of vehicle length. 1/4 to 1/5 power usage of radars lends itself to solar applications when continuously monitoring all lanes. Side fire device. Long MTBF and no maintenance once installed. Re-configurable. Usually lower priced than the digital radars and video systems for ITS use. Limited vehicle classification available. Volume, lane occupancy or headway, speed are standard data stream items if a relay interface is not used. Cabinet controller not required. -- con as a passive device, it is limited in the number of lanes it can reliably resolve beyond 5 or 6. Requires a mounting height about twice that of a radar for comparable coverage. Must carefully follow manufactures recommendations when mounting near bridges, tunnels and other road side objects of this nature to mitigate effects of reverberation. Users not always able to conceive of an acoustic detection zone easily like a video image.

Embedded Systems / Standard Loops. Pro -- ubiquitous, so relatively cheaply available hard ware. Low training requirement. Closest sensor to object being sensed, it should be the most accurate. Can be made different sizes at installation for different application (ITS vs. stop line sensing). Usually have an industry standard relay interface. Con -- While hardware is cheap, lane closures and traffic control are very expensive, typically more than the cost of side fire systems above. Embedded sensors typically have high failure rates due to milling, lane re-stripping, and other issues other than cables being cut. (One study put loop failure at over 20% per year) Must close lanes to install and maintain. Home run requires digging and filling. Break up the road surface, lowering roadway life expectancy. Very critical if bridge decking to be considered, especially in north climates. Not Re-configurable. Volume, lane occupancy or headway, speed are available only if an expensive cabinet controller are available.

Micro loops. Pro- getting to be more reliable, sensors less costly. Closed sensor to the object being sensed, it should be the most accurate. Some are re-configurable by movement inside the tube they are in under the road. Those in tubes in the road are not subject to fatigue failure like the loops above. Con-- installation requires a bore from the side of the road with some clearance and a close tolerance for depth. May be very expensive, depending on soil type and location. Non bore types which are cored into the road have limited battery life (3 to 5 years) before then must be replaced, requiring lane closure like normal loops. Also, wireless transmission for real time operations from the sensor may be occluded in heavy traffic volume. The transmitters are relatively low power to preserve battery life and must have a receiver places strategically near the road bed to receive their signals. Those that are not put physically into the road, but are meant to be "stuck" to the road with adhesive get removed with the passing of the first snow plow or dragging muffler.

Combinations of sensor technologies using those above lend to very accurate results at the penalty of higher cost. No sensor type is perfect by itself, so the life cycle cost and accuracy issues are part of the trades, as are training and maintenance, geometry constraints, legacy interfaces, etc.

FARID SEMMAHI – TRAFFICON – 12/12/06

Company in Belgium: In traffic business for 20 years/video detection. Main applications depends on region, Europe: incident detection/data, US: presesnce detection at interesection, data incidents. Atlanta, Traficon gathers raw data for units brought into navigator server. Algorithms are done by state DOTs,

Our job gathers the raw data: advantages: we do a better job in multiple lanes, RTMS might not deal with occurancy well. Visual aspect: a wider area of dectection. We can validate problems, whereas radar, you cannot, (monitoring tools,) Occurancy is better than RTMS. Compensation algorithm exists. Their core business is the detector unit, distributors provide the backend.

LEE-ANNE SEELING – TRICHORD – 11/13/06

Mr. Pieper is correct that we at Trichord can shed some insight into your areas of exploration. I've recently by talking with JD Margulici regarding a potential CalTrans project but would be would be delighted to chat with you on any of your topics at the ITS-CA meeting tomorrow thru Wednesday in Sacramento – is there a time that would be convenient for you?

If not, perhaps we can arrange a time to get together either in Berkley or down in Southern California where my offices are. Look forward to speaking to you,

RYAN LINDSEY – WAVETRONIX – 2/1/07

UT, CA, IA, TX (largest implementation), FL, NY, NV

Both Radar: Digital Wave Radar, lock in on target and doesn't drift, better detection, new technology, bette technology, ease of installation, automatic. Most algorithms are provided the customer.

Up to 10 lanes of detection automatically, (volume count, veh speed, avg length speed, 85% speed, (approach)>info from those different lanes.

DONALD SHUPP – WESTERN PACIFIC – REPLIED TO EMAIL DIRECTED TO NAZTEC

Paul,

I wanted to introduce myself as the exclusive local factory representative for Naztec, Inc. I believe you had made an inquiry last week with their system developer, Bryan Beyer. Bryan asked me to find out if I could provide you with local service and support relating to your desire to learn more about how traffic data is accurately recorded through highway sensors on our system. Our central Advanced Traffic Management System (ATMS) software uses incoming vehicle detectors to predict travel times in either highway corridors, or local surface street links (intersection-intersection nodes). This is performed by periodic sampling (1 min – 15 min typical reads), and using a lookup table by location to measure against either volumes, or occupancy percentages. Since our local controller software can “gate” the incoming information to only report detector occupancy when the associated phase movement is green or yellow, the field data is highly accurate and effective.

If this system is of interest to you, I would be glad to invite you to our local San Leandro office for a tour of our facility, or meet with you at your office. Myself, and Jeff McMullen (Transp. Manager, District 6, retired) participate in the UCB ITS extension course on the Model 2070 traffic controller. The ATMS software that we work with communicates with both 170E controllers (running Calltrans C8 local software), and 2070 controllers (running NTCIP 1202-based software) via a GIS and Web-based map selection utility. In addition to selecting controller objects from the map, control of CMS / DMS, CCTV, and field modem elements may be launched from the GIS.

In terms to your interest with multiple detection technologies, our system collects detector information with loops, video, and microwave (RTMS, Accuwave, or MS Sedco products). We have teamed with several qualified ITS partners to integrate fully scalable ATMS central systems, such as Naztec, Cisco Systems, Actelis Networks, EIS (RTMS), and Eberle Design (EDI). I look forward to discussing more details with you soon. I’ll be returning from the Sacramento annual ITS-CA (CITE / CAATS) conference on Thursday.

Donald R. Shupp, VP
Western Pacific Signal LLC
(510) 483-6400 Ofc
(510) 754-6058 Cell

APPENDIX III: RESEARCH PAPERS

This Appendix contains three research papers, one of which was published and is reproduced here for the exclusive consumption of Caltrans as the sponsoring organization. The three papers are as follows:

- Margulici, J.D., and Ban, X. **Benchmarking travel time estimates**. IET Journal of Intelligent Transportation Systems, 2008. – 15 pages.
- Margulici, J.D., Pan, X., and Mellet A. **An Extended Benchmark of Travel Time Estimates on Urban Freeway Corridors**. California Center for Innovative Transportation, 2009. – 19 pages.
- Pan, X., Margulici, J.D., and Yazbek, R. **Treatment and Analysis of Ground Truth Travel Time Data on Urban Freeway**. California Center for Innovative Transportation, 2009. – 17 pages.

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

Submitted 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.

BACKGROUND AND PROBLEM

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.

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.

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.

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.

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:

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

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 - Example of Performance Measures for 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.

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.

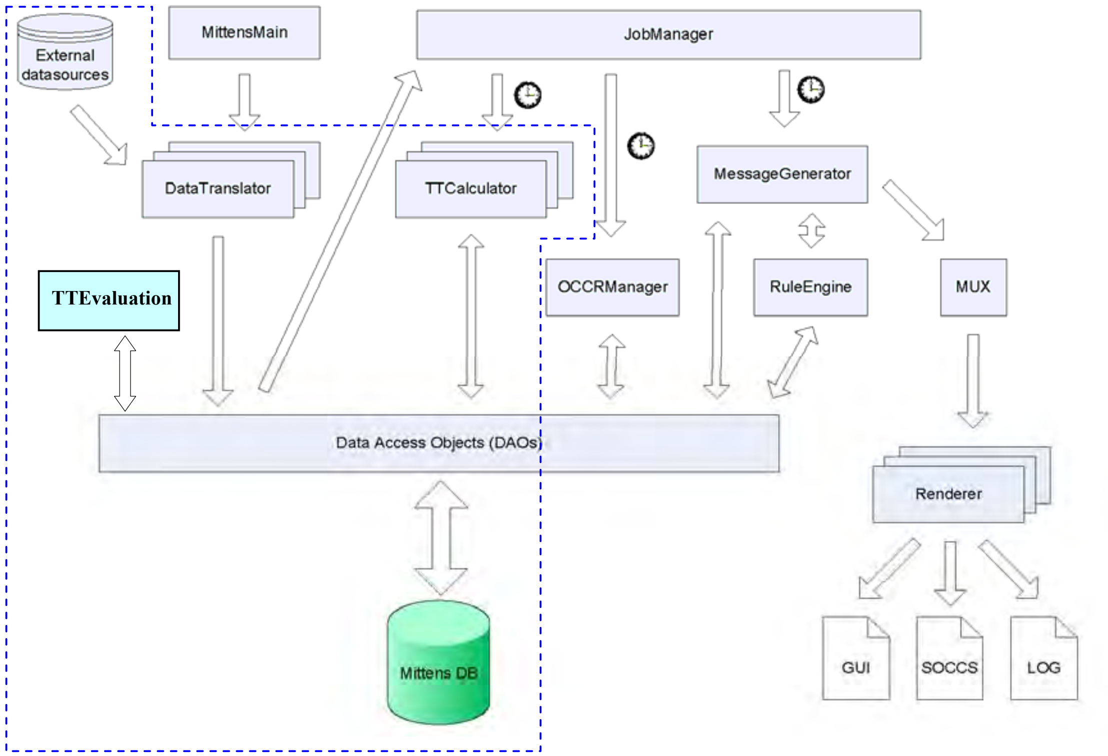


Figure 1 - System Architecture of MITTENS

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.

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.

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	FasTrak	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

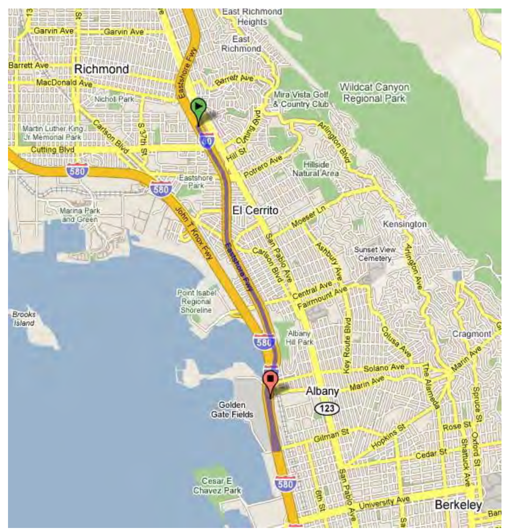


Figure 3 - Route 2 – WB I-80

Figure 4 - Route 3 – SB I-880

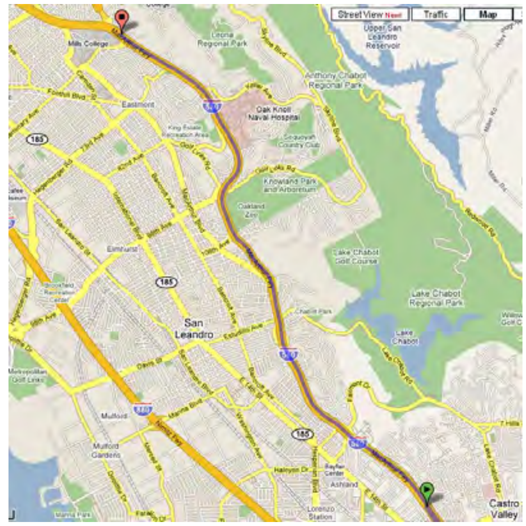


Figure 5 - Route 4 – WB I-580

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).

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.

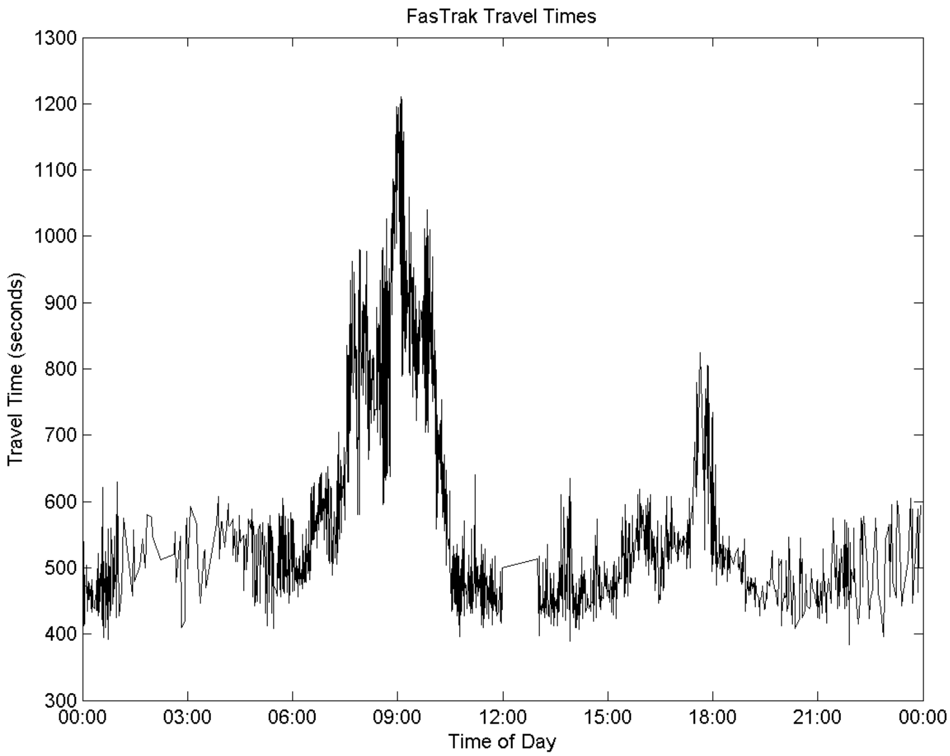


Figure 6 - Individual travel times

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.

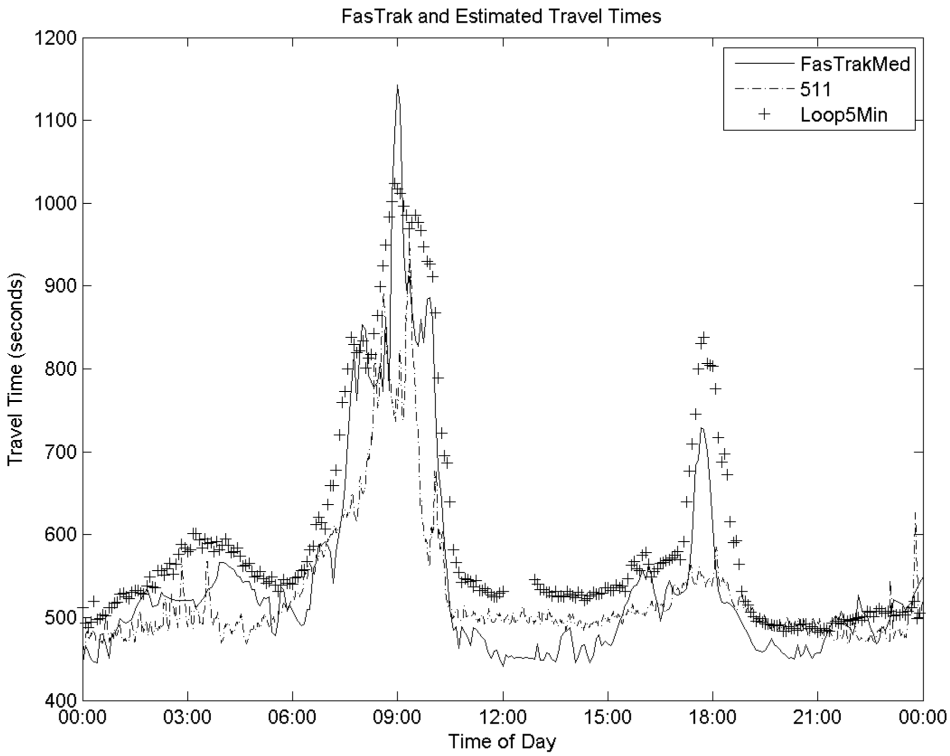


Figure 7 - Comparison of ground truth travel times with estimates from local 511 and loop detectors

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.

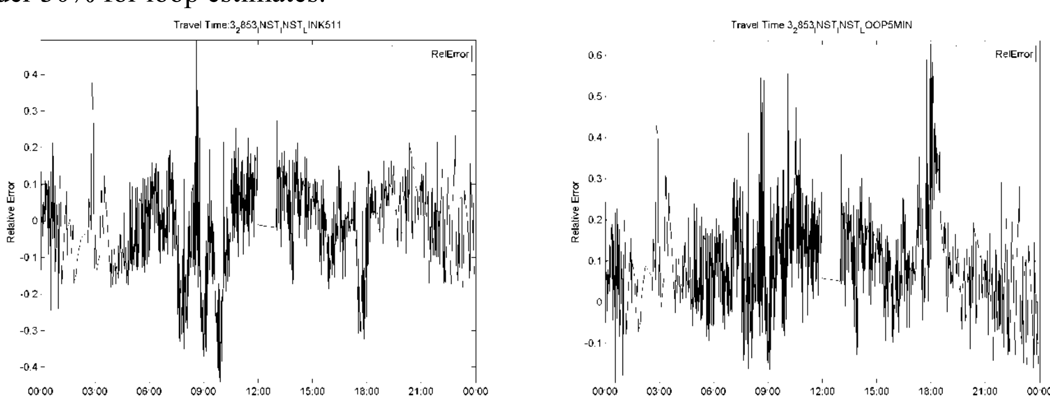


Figure 8 - Individual relative errors for estimates from 511 and loop detectors

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 5 selected time periods

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.

An Extended Benchmark of Travel Time Estimates

on Urban Freeway Corridors

J.D. Margulici*
California Center for Innovative Transportation
University of California, Berkeley
2105 Bancroft Way, Suite 300
Berkeley, CA 94720-3830
Email: jd@calccit.org

Xiaohong Pan
California Center for Innovative Transportation
University of California, Berkeley
2105 Bancroft Way, Suite 300
Berkeley, CA 94720-3830
Email: xpan@calccit.org

Aymeric Mellet
École Polytechnique
Route de Saclay
91128 PALAISEAU Cedex, France
Email: aymeric.mellet@polytechnique.edu

July, 2009

Word count 4658 + 2750 (6 figures + 5 tables) = 7408

Submitted for Presentation and Publication to:
Transportation Research Board 89th Annual Meeting
Washington, D.C.

* Corresponding Author

An Extended Benchmark of Travel Time Estimates on Urban Freeway Corridors

ABSTRACT

This paper further refines a proposed methodology and metrics to benchmark the quality of traffic information. The methodology hinges on the collection of ground truth travel time data, which we do here from electronic toll tags. Compared to earlier studies, we consider a much larger set of routes over an extended period of time. This enables us to test our benchmark methodology and examine its features in practice. We conclude that the benchmark metrics we proposed earlier seem wholly adequate for our purpose, which is to introduce a standard practice for the evaluation of traffic information.

An Extended Benchmark of Travel Time Estimates on Urban Freeway Corridors

1. INTRODUCTION

1.1. Background and Problem Statement

Travel time estimates on selected itineraries are a natural indicator of traffic congestion and are well understood by the traveling public. Travel time information can be disseminated through various channels, including the internet, mobile devices, navigation units, and dynamic message signs. Numerous studies reveal that travelers appreciate and value travel time information, especially because it reduces travel uncertainty and stress (*1-3*). In theory at least, travelers can also respond to real time information by adjusting their travel schedules, travel routes, or even travel modes (*4*). From an operational standpoint, travel times constitute an effective performance metric to describe the level of service on a given roadway segment.

The quality of travel time estimates generally relies on sufficient traffic data collection, as well as adequate processing techniques and information delivery modes. Each of these steps has benefited from a flurry of innovations over the past decade (*5-10*). Nevertheless, literature on systematic evaluations of the quality of travel time estimates remains limited. Our research team has conducted research on this issue and highlighted its importance (*11*). First, real-time traveler information is becoming ubiquitous and demand is driven by mobile devices and location-based services. Where users were previously content to simply get any data, quality starts to matter a lot more. Second, the emergence of private aggregators of real-time traffic information that collect data from mobile probes means that Departments of Transportation now have the option to purchase third-party feeds instead of deploying their own detectors. In this context, public agencies need a methodology to properly assess and continuously audit data quality. Our previous publication (*11*) suggested such a methodology and highlighted a set of two standardized metrics aimed at benchmarking travel time estimates.

1.2. Research Objectives/Questions

The present paper was written in conjunction with a second, related paper (*12*). The two papers split results obtained from further work done around the methodology and the metrics that we have already exposed. Our focus here is on practically implementing the methodology as well as validating the adequacy of the metrics to benchmark estimates in a variety of situations.

With respect to practical implementation, we primarily integrate an outlier removal methodology previously developed by one of us (*13*) and further investigate its applicability. Regarding the two metrics used to benchmark travel times estimates, we were interested in testing them on a much larger number of routes than we had done prior. The metrics need to achieve two objectives: a) synthetically capture as much information as possible about the quality of various estimates over time and along different itineraries; and b) provide a graded scale that offers good discriminating power between mediocre estimates and accurate estimates – in other words, the metrics must be calibrated such that they immediately and reliably distinguish good from bad. In order to assess how well those two objectives are attained, empirical verification was required. In the process, we also realized that we needed a finer comprehension of how collected ground truth travel times vary, both within a given sample and over time. Obviously, such a comprehension is instrumental to developing proper estimates in the first place, and we therefore think that our work has inherent value for traffic modelers. Systematic analysis of ground truth travel time data along a large set of itineraries is hard to come by in the literature, probably because such data has not been widely available until fairly recently.

In sum, we considered the following research questions in this expanded travel time benchmark study:

- What could be a sufficient/reasonable sample size of ground truth travel time for the purpose of benchmarking estimates?
- How do ground truth travel times vary in the first place? In particular, what metrics can be designed to feature different variance components due to individual drivers and the effect of congestions?
- How do the previously proposed benchmarking metrics perform in our comprehensive experiment?

This paper addresses the third question, while the companion paper (*12*) responds to the first two. These results are intended to help researchers and practitioners with theoretical and practical knowledge about travel time variations and estimates benchmarking. A broader goal is to feed into a conversation about the proper assessment of traffic information quality and foster the adoption of standards that can serve the industry by accelerating the dissemination and commercialization of data.

A short literature review section is followed by a methodology section in which we review the development of benchmark metrics and describe the data sets used for this study. We present and analyze our results in section 4.

2. LITERATURE REVIEW

2.1. Importance of Benchmarking Travel Time Estimates

Roadway travel time estimates have been studied extensively in the past couple of decades (*8, 14, 15-18*). This has been driven by the development of technologies and the gradual deployment of Intelligent Transportation Systems (ITS), as well as data-hungry transportation professionals and commuters. Further, a number of private concerns have developed methodology to collect, aggregate, process and disseminate traffic data. Today, a majority of Departments of Transportation utilize at least some traffic information supplied by the private sector, either in the form of raw data or speed estimates, by opposition to collecting it from detectors that they own, operate and maintain. Yet in terms of volume, this remains somewhat limited. Basic economic considerations would suggest that this trend can only grow, as third-party traffic data providers can implement new technologies at a much faster rate than most public agencies, make money by providing services to consumers, and benefit from economies of scale.

As pointed by Margulici and Ban (*11*), measuring the quality of travel time estimates matters for several reasons: a) the margins of errors of travel time estimates should be better understood and formulated so that drivers can develop adequate expectations; b) 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; c) 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 accelerate the market.

2.2. Existing Researches/Practices Related to Benchmarking Travel Time Estimates

Although many innovative estimation techniques have been proposed, there have not been many publications to discuss the topic of travel time estimates benchmarking systematically. Guo et al. (*19*) is one of a few studies that examined the quality of link speed and travel time estimations. Their study developed and tested a structured evaluation procedure, focusing on system design and information technology resources (i.e., Geographic Information Systems (GIS) and Global Positioning Systems (GPS)). The methodology relies on floating cars to collect baseline travel time and link speed. However, the comparison procedure merely calculates the difference between the estimates and baseline traffic data, and no attempt is made to propose generalized metrics. Kothuri et al (*18*) also pointed to the importance of measuring the accuracy of travel time estimates. Their study compared ground truth data collected from probe vehicle runs with real-time travel time estimates in Portland, Oregon. Again, the focus of the study was on the sources of error and corresponding solutions rather than on evaluation methods/metrics.

Another study (*2*) focused on evaluating the performances of several travel time estimation methods using loop detector data. Although it used multiple data sources for ground truth travel times, i.e., license plate readers, floating car runs, and toll ticket collection, the validity of this baseline data did not appear to be very solid.

Many studies related to innovative data processing schemes also tried to validate the accuracy of their methods. For instance, Zhang et al. (*15*) utilized floating car runs to gather ground truth travel times and test an estimation method based on single loop detector data. Kown et al. (*20*) and Fujito et al. (*21*) used travel times computed from “baseline” detector spacing as ground truth to evaluate their research systems and methods. Recent studies by Dance et al. (*22*) utilized speed contour maps to conduct evaluation, but no quantitative measures were developed.

Most of these existing studies on evaluating the quality of travel time estimates tend to be limited in the sense that sufficient and reliable ground truth travel times were not widely available, and that the comparison metrics were not designed to be comprehensive and generalized.

3. METHODOLOGY

3.1. Conceptual Framework

In our earlier study, we proposed benchmark metrics and tested them using a small sample of traffic data from the San Francisco Bay Area. For more information regarding the methodology and results of the earlier study, please see the paper prepared by Margulici and Ban (*11*). As the study suggested, we enhanced the development of quality assessment metrics and expanded the scope of evaluation experiments, both in time and coverage. The conceptual framework is illustrated in Figure 1, including two major interactive tracks – metric development and experiment design. Detailed discussion for these two tracks is presented in the following sections.

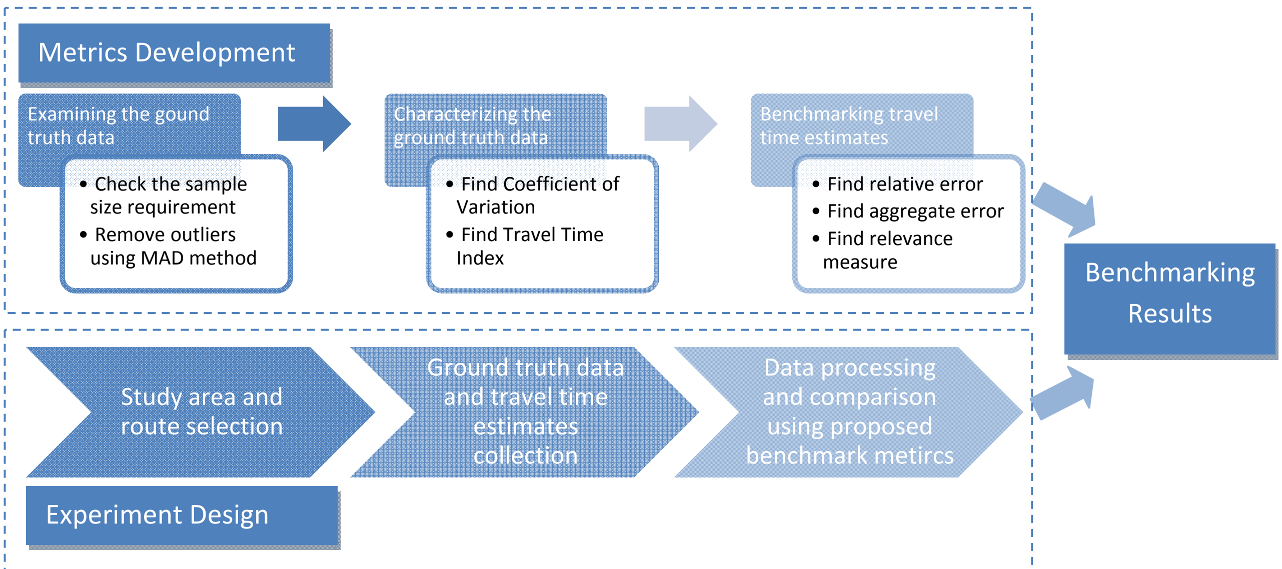


Figure 1. Conceptual framework

Before discussing the development metrics for travel time estimates, we find it necessary to point out a few considerations and assumptions we used in our research:

- Our objective is to come up with a set of clearly defined, simple, and straightforward benchmark metrics. Additionally, those metrics should be as standardized as possible, so that they can easily be applied in various settings and with different sources of data. We do not pretend to have reached a final answer and rather hope to open a debate about what constitutes good metrics for the industry to move forward.
- The proposed benchmarking methodology assumes the existence of significant volumes of ground truth travel times with sufficient reliability. As mentioned in the previous section, this requirement is becoming much less stringent thanks to the increasing deployment of ITS and the constant development of innovative and lower cost information technology solutions. In the present work, travel times collected from the San Francisco Bay Area's FasTrak toll tag readers were used as ground truth travel time.

- For the purpose of analysis, we divide time periods into elementary time slots. A time slot *i* may be set to 5 minutes, 10 minutes, 1 hour etc. We mostly stick to symbolic notations in this section and shift to numerical examples in the presentation of results.

3.2. Benchmarking travel time estimates

In this subsection, we restate the definition of the benchmark metrics developed by team members in a previous publication (*11*). Those metrics stem from a user-centric approach: in effect, each individual driver who has access to real-time traveler information observes two values: a travel time estimate, and his or her actual trip time. We calculate an *individual relative error* between those two values, defined as the ratio of their difference to the actual trip time. This number is affected by both the quality of travel time estimates and individual differences between drivers. Individual relative errors are then bunched together into two metrics: *aggregate error* and *relevance*. One of the advantages of this approach is that it is indifferent to the refresh rate of travel time estimates and doesn't need to be mapped to set time intervals. In particular, comparison between data feeds that have different refresh rates is transparent. Further, both the aggregate error and the relevance are dimensionless and normalized with respect to route length and study period duration.

3.2.1. Individual Relative Error

For the m^{th} driver traveling along certain route, his/her *relative error* is defined as $e_m = \frac{\widehat{tt_m} - tt_m}{tt_m}$, where $\widehat{tt_m}$ refers to the estimated travel time provided for the route at the start of the driver's trip and tt_m refers to the actual journey time experienced by this driver.

3.2.2. Aggregate Error

The *aggregate error* captures the overall inaccuracy of the estimates over an arbitrary time interval T. It is defined as the mean relative error within this period, $E_T = \frac{\sum_{m=1}^{M} e_m}{M}$, where M refers to total number of drivers recorded on the route segment studied over the time interval T.

3.2.3. Relevance Measure

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: $R_T^{\varepsilon} = \Pr\left(|e| \le \varepsilon\right)$, where e represents the individual 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.

3.2.4. Discussion

Both the aggregate error and the relevance measure can be expressed as percentages and provide fairly intuitive results. The aggregate error essentially measures estimate bias. It is an algebraic number which can be either positive (systemic overestimation) or negative (systemic underestimation). Some of the criticism that we have received pointed out that summing up individual errors algebraically cancels out the deviations from the mean and results in a number that is not reflective of the error that a typical driver experiences. The mean absolute error or the root mean squared error would accomplish that, but the downside is that some information is lost in the process. By keeping two separate metrics, we can draw comparisons between estimates in terms of overall bias, and still provide an assessment of the driver experience with the relevance metric. The relevance measure schematically states: the estimates was within 15% (or 10, or 20…) of actual driving time for N% of drivers. We feel that this piece of information is actually more interesting than to state that the mean absolute error is, say, 12%, because it addresses a more comprehensive performance objective. Part of our objective in this paper is to examine the sensitivity of the relevance measure to different thresholds and determine which is most appropriate.

3.3. Experiment Design

Compared to our previous study, we propose a more comprehensive experiment to examine the variability factors that affect travel time estimates and to assess the validity of the travel time estimates benchmark methodology and technique.

3.3.1. Study Area and Route Selection

The study area was set to the San Francisco Bay Area due to the remarkable advantage of its traffic data availability. The experiment was hosted by the MITTENS[1] system, which generates travel time estimates using a variety of sources, such as loop detectors and radars. Another important resource is the "ground truth" traffic data collected from FasTrak toll tag readers, which can be retrieved in the MITTENS system in disaggregate form or in the PeMS[2] system at an aggregated level.

The experiment routes form a representative set of the various features that distinguish them from one another. These include sample size of FasTrak data (small/moderate/large), type of area (urban/suburb/rural), geographic location, whether comprising a bridge or not, as well as magnitude of the variations of ground truth travel times (small/moderate/large). It is made up of 20 highway segments around the Bay Area. We encoded these 20 routes into Google Maps for reference[3]. Additionally, we also include those 4 experiment routes used in our earlier study. An overview of the total 24 study routes is presented in Table 1.

Table 1. Overview of the Study Routes

Route ID	Length (mile)	Freeway	Features
1	9.14	SR-24E	-
2	6.48	I80-E	including bay bridge

[1] MITTENS stands for "Messaging Infrastructure for Travel Time Estimates to a Network of Signs", which is an automated system developed by CCIT for Caltrans District 4 to display travel times on CMS in the Bay Area.
[2] PeMS stands for Freeway "Performance Measurement System", a web-based tool originally designed at UC Berkeley to host, process, retrieve, and analyze road traffic condition information.
[3]http://maps.google.com/maps/ms?ie=UTF8&hl=en&oe=UTF8&msa=0&msid=110448774267712153593.0004693 1135a9cc4d5c52&start=0&num=200&z=9

3	1.75	I80-E	relatively short distance
4	7.75	I80-E	-
5	3.00	I80-E	relatively short distance
6	3.25	I80-E	suburban area; relatively short distance
7	26.50	I80-E	suburban area; relatively long distance
8	6.48	I80-W	including bay bridge toll plaza
9	7.84	SR84-W	-
10	1.75	SR92-E	suburban area; relatively short distance
11	11.03	SR92-W	-
12	25.00	I880-N	-
13	8.00	I680-N	-
14	41.33	I680-N	suburban area; relatively long distance
15	7.75	I580-W	-
16	10.75	SR101-N	-
17	4.95	SR101-N	urban area
18	16.67	SR101-N	suburban area
19	11.75	SR101-N	suburban area
20	2.49	SR101-S	-
21	13.90	I80-E	-
22	3.15	I80-W	-
23	9.00	I880-S	-
24	7.75	I580-W	-

3.3.2. Ground Truth Travel Time and Travel Time Estimates Collection

Ground truth travel times for the 24 experiment routes selected were systematically retrieved within the MITTENS system from early April to early June, 2009. Travel time estimates generated from 511 data were also collected for the same routes and the same time period.

Our data feeds for ground truth travel times and estimations did not perform reliably over the entire collection period. Because of the wealth of data available to us, we took a very conservative approach in rejecting samples that were partial or looked suspicious. The net result is that we ended up with 170 day.routes for which both ground truth travel times and estimations were deemed satisfactory for analysis purposes. We also conducted a PM peak analysis with 98 data samples (i.e. pairs of routes and PM periods). The reason for the lower number is that we imposed more stringent completeness criteria for the data in the latter analysis.

3.3.3. Data Processing and Comparison Using Benchmark Metrics

The first data processing step consists in rejecting observations for which the minimum sample size requirement is not met. We remove the outliers among those for which it is. We then analyze the features of the ground truth travel times, including individual differences and congestion-driven variations. Finally, we compare those ground truth travel times with the travel time estimates using the proposed benchmark metrics. We conclude by examining the comprehensiveness and discriminating power of the metrics across the entire route set.

4. RESULTS AND FINDINGS

4.1. A First Look at the Data

We arbitrarily present results for route 1 in order to illustrate the methodology presented in the previous section. Route 1 is just over 9 miles in length on CA State Route 24. Ground truth travel times for that route collected on May 5, 2009, are presented, in seconds, on Figure 2. The plot gives a sense of the very substantial amount of data collected during that period, and how it provides an unequivocally adequate sampling of travel times for that route.

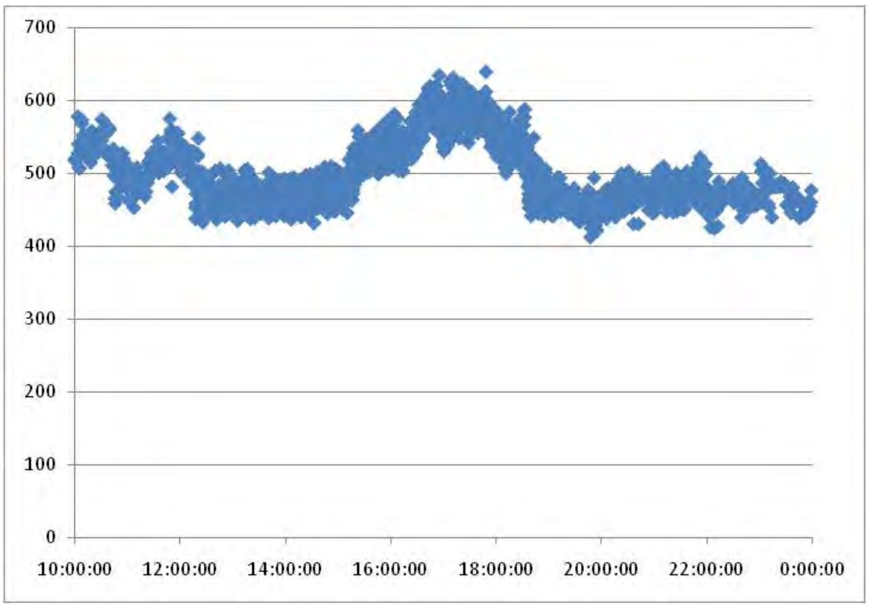


Figure 2. Ground truth travel time profile for route 1 on May 5, 2009

Figure 3 displays individual relative errors between these ground truth observations and the 511 estimates. As can be seen, most individual errors are 10% or less in absolute value, establishing that the 511 estimate is very accurate for this route.

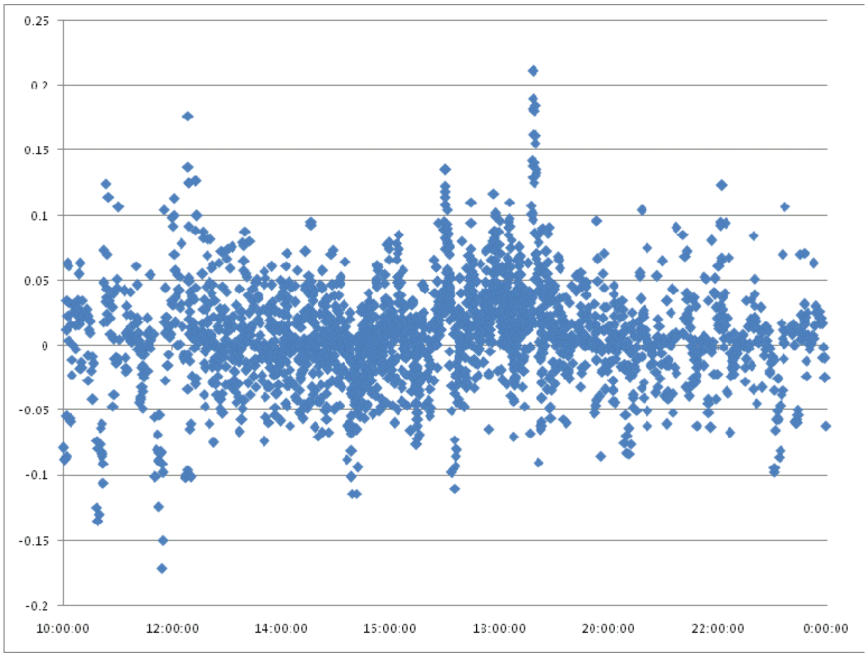


Figure 3. Example of individual relative errors for route 1 on May 5, 2009

Table 2. Benchmark Metrics for 511 Estimates on Route 1, Selected Days

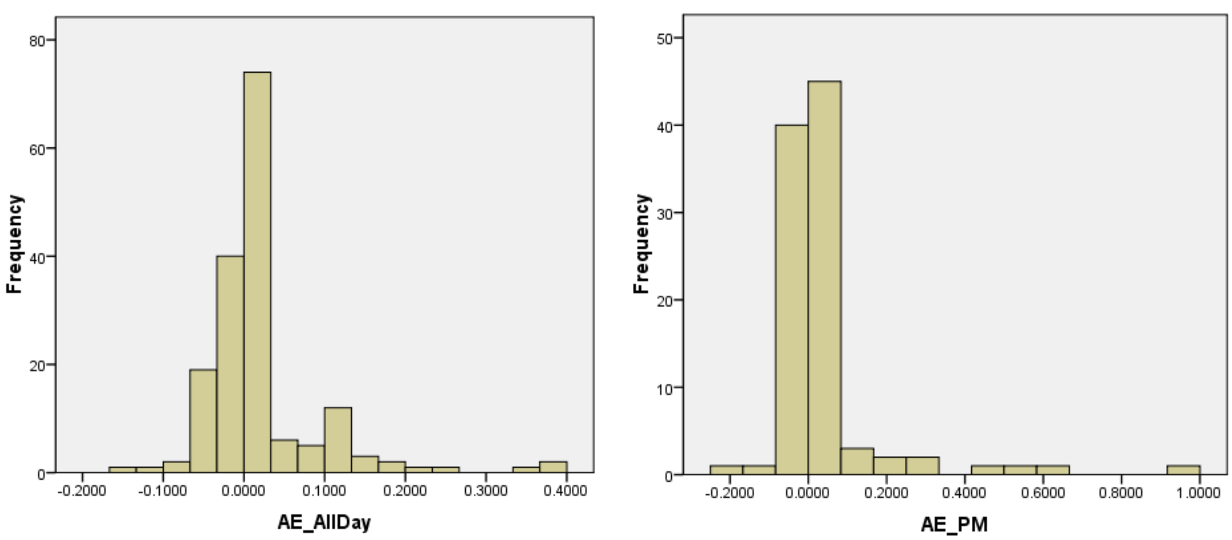
Date	AE	R10	R15	R20
5/5/2009	0.8%	97.4%	99.5%	100.0%
5/6/2009	-0.2%	98.7%	100.0%	100.0%
5/7/2009	0.2%	99.4%	100.0%	100.0%
5/26/2009	-0.1%	94.7%	99.8%	100.0%
6/1/2009	-0.7%	92.9%	100.0%	100.0%
6/2/2009	1.1%	97.5%	99.8%	100.0%
6/3/2009	1.6%	92.7%	98.8%	99.8%
6/4/2009	2.1%	91.8%	98.9%	99.7%
6/5/2009	2.0%	91.2%	98.6%	99.8%
6/6/2009	1.9%	92.9%	98.5%	100.0%
6/9/2009	1.5%	97.6%	99.5%	100.0%
6/10/2009	0.6%	100.0%	100.0%	100.0%

Table 2 summarizes benchmarking results for route 1 on a dozen different days. The aggregated error stays less than 2% in absolute value, suggesting that the travel time estimates from 511 for this route are unbiased overall. We also indicate the relevance measures with three different thresholds at 10%, 15% and 20%. Based on our ground truth sample, virtually 100% of all drivers on route 1 for the 12 selected days would have received travel time estimates from 511 that correctly predicted their trip within 20%. With a threshold of 10%, the proportion oscillates between 90% and 100%, with a median of 96%.

4.2. Cross-Sample Analysis of Aggregated Errors

We now turn to the consolidated data from all 24 routes. In total, we have 170 day.route pairs, and 98 such pairs when we restrict the study to the PM peak hour. For each pair, we calculated the aggregated error and the relevance measure at different thresholds.

Figure 4 summarizes those results for the aggregated error measure. It shows histograms of the measure's distribution across both the all-day set and the PM-peak-period set. As can be seen, the first result is that the 511 travel time estimates tend to be very accurate on most routes and most days. The vast majority of the measurements fall within 10% in absolute value. At the same time, we observe a limited number of higher values. Those higher values indicate that estimates have been consistently wrong for extended periods of time on the days and routes to which they correspond. In this sense, the aggregated error measure has the good taste to immediately point to problems in the estimation system. As explained in the methodology section, it is also possible that a good aggregated error measure hides significant individual errors that cancel out over the study period (i.e. a day or the entire PM peak). If that is the case, the relevance measure would reveal this deficiency, and it could be further concluded that the estimation errors are not consistent (in other words, there is a combination of overestimates and underestimates).

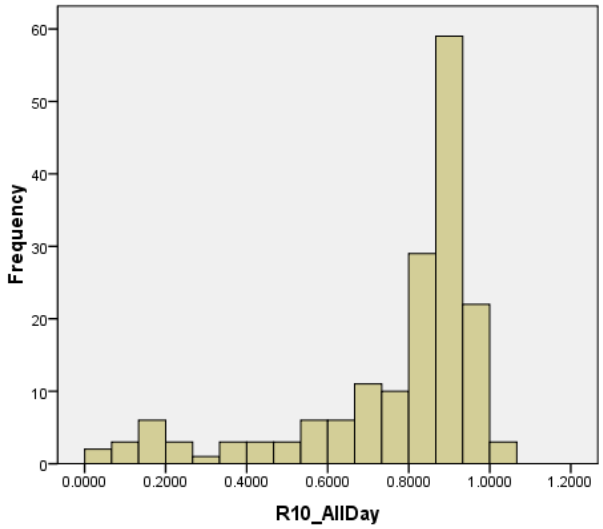


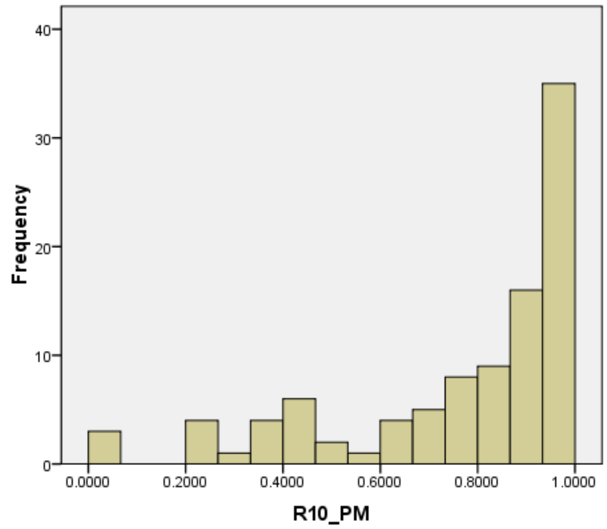
(a) (b)

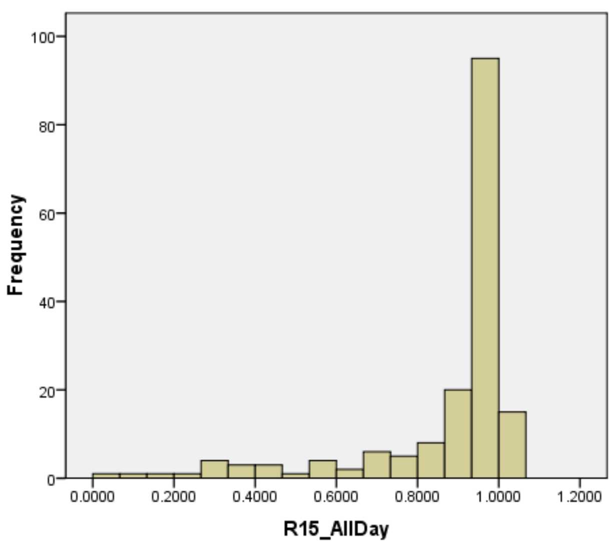
Figure 4. (a) Distribution of Aggregate Error for all days and all routes (n=170)
(b) Distribution of Aggregate Error for all PM peaks and all routes (n=98)

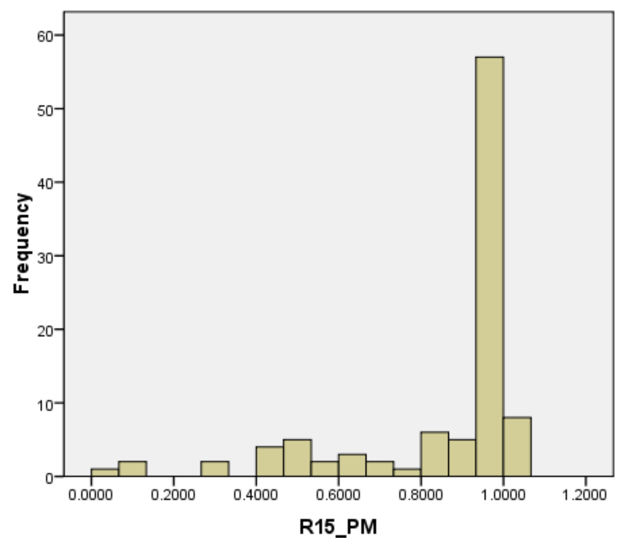
4.3. Analysis of Relevance Measures

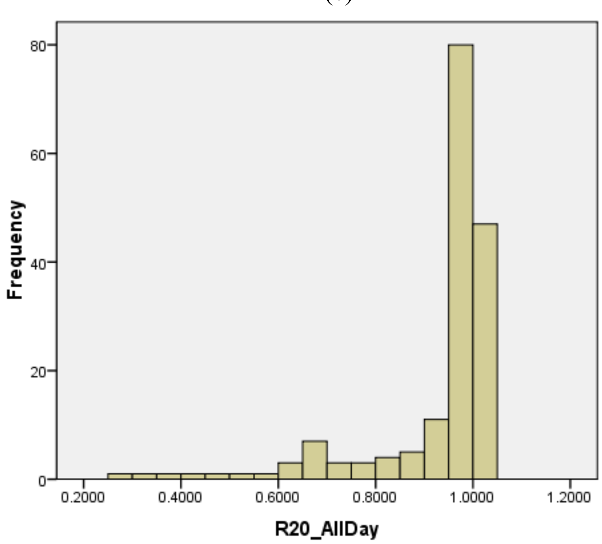
We conducted the same distribution analysis for the relevance measures, with three different thresholds set at 10, 15 and 20%, respectively. Again, the first observation is that the 511 estimates perform very well overall. For about ¾ of all day.route pairs, the R15 benchmark metric is 90% or more (Figure 5 (c) & (d)). In other words, for most routes and on most days, 90% or more of the drivers can get a trip time estimate that is within 15% of they actually experience.


Frequency
R10_AllDay
(a)


Frequency
R10_PM
(b)


Frequency
R15_AllDay
(c)


Frequency
R15_PM
(d)


Frequency
R20_AllDay
(e)

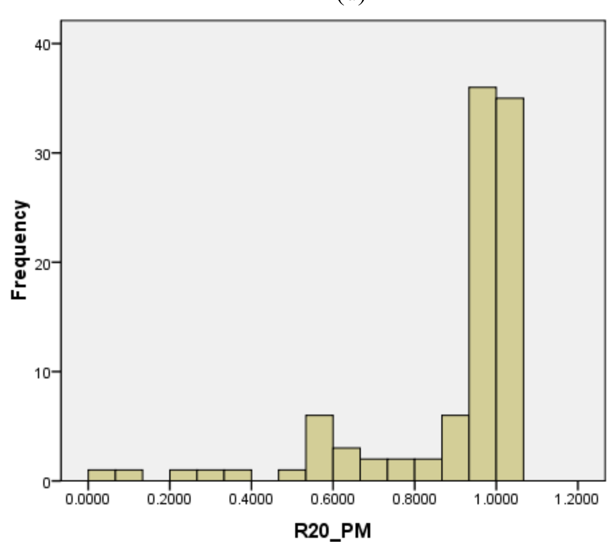

Frequency
R20_PM
(f)

Figure 5. (a) Distribution of Relevance (10% threshold) for all days and all routes (n=170)
(b) Distribution of Relevance (10% threshold) for all PM peaks and all routes (n=98)
(c) Distribution of Relevance (15% threshold) for all days and all routes (n=170)
(d) Distribution of Relevance (15% threshold) for all PM peaks and all routes (n=98)
(e) Distribution of Relevance (20% threshold) for all days and all routes (n=170)
(f) Distribution of Relevance (20% threshold) for all PM peaks and all routes (n=98)

As we may expect, the results look even quite more favorable at the 20% threshold (Figure 5 (e) & (f)). At that level, nearly 50 day.route pairs score 100% -that is, after removing outliers, no ground truth observation differ from the 511 estimate by more than 20%. On the other hand, the picture is more contrasted at the 10% level. On the one hand, 10% appears very stringent. This especially true of shorter routes, where such accuracy becomes unnecessary: estimates are typically rounded to the next minute, and thus the rounding itself may approach or even exceed 10%. On the other hand, using the 10% threshold results in a more even distribution, which is a desirable feature in the context of a benchmark. This threshold level holds a greater ability to discriminate between 'good' and 'bad' estimates, whereas the distinctions may be more difficult to draw when estimates are compared based on R20 scores.

4.4 Further Analysis

In this subsection, we single out a few routes to complement the general analysis presented above. Table 3 shows results for Route 12, which at 25 miles is one of the longer routes in the set.

Table 3. Example of route 12 (I880-N, 25-mile, from North of Stevenson to I-980 Split)

Date	AE	R10	R15	R20
5/5/2009	-2.4%	71.4%	87.6%	94.2%
5/6/2009	5.8%	55.6%	81.1%	85.6%
5/7/2009	7.2%	44.1%	71.6%	85.5%
5/26/2009	7.6%	60.4%	85.3%	92.2%
6/1/2009	-2.4%	68.6%	97.1%	100.0%
6/2/2009	-1.9%	76.5%	93.1%	96.2%
6/3/2009	-2.2%	70.6%	89.9%	95.3%
6/4/2009	1.8%	65.7%	82.7%	90.5%
6/5/2009	-0.9%	80.0%	92.5%	97.2%
6/6/2009	-2.5%	93.8%	100.0%	100.0%
6/9/2009	-0.6%	59.8%	75.1%	84.1%

At a glance, we can observe that these results are not as good as the prevalent benchmarking numbers obtained for other routes. The aggregated error measure stays within reasonable bounds, though on a few days (5/6, 5/7, 5/26) the 511 estimates were systematically too high. Differences from other routes are more pronounced in the relevance values. The R15 values are less than 90% on 7 out of 11 days, and the R10 values drop as low as 44%. Obviously, route length affects those values. Because the estimates provided are considered predictive, the longer routes create more room for errors as traffic conditions evolve along the way.

Another interesting aspect of this data is that while higher values of AE tend to correspond to lower relevance values, this relationship is not systematic. For instance, some of the lowest relevance values are found on 6/9, when the absolute aggregated error is only 0.6%. This finding comforts our choice of benchmark metrics by showing real complementarities between the two sets of values.

Route 18 is another relatively long route, nearly 17 miles in length. Yet the results of the benchmark, show on Table 4, happen to be quite good. The lowest value for R15 is 96.6%. Again, this happens to be on a day when the aggregated error is almost null (0.4%, on 6/6), confirming that the two metrics carry different information.

Table 4. Example of Route 18 (SR101-N, 16.7-mile, from Novato to Petaluma)

Date	AE	R10	R15	R20
5/5/2009	-0.3%	96.0%	99.8%	100.0%
5/6/2009	1.5%	92.7%	99.0%	99.5%
5/7/2009	-4.2%	85.2%	99.2%	99.9%
5/26/2009	-4.2%	93.3%	99.6%	100.0%
6/1/2009	-0.1%	89.2%	98.6%	100.0%
6/2/2009	-0.3%	96.7%	99.9%	100.0%
6/3/2009	0.5%	92.7%	98.5%	99.6%
6/4/2009	-1.8%	87.1%	97.9%	99.1%
6/5/2009	-1.5%	90.0%	98.9%	100.0%
6/6/2009	0.4%	88.9%	96.6%	100.0%
6/9/2009	-0.7%	89.0%	98.1%	100.0%
6/10/2009	0.6%	94.1%	99.0%	100.0%

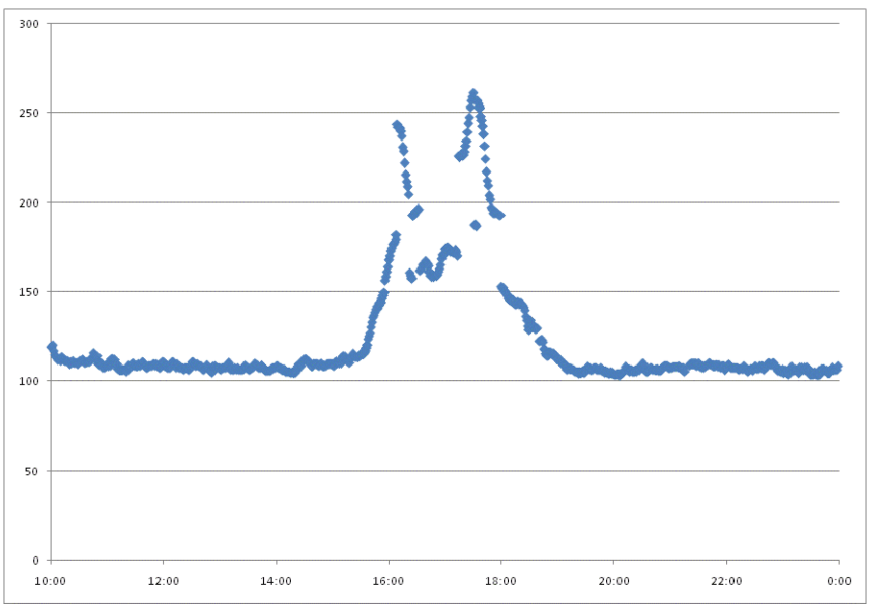
Conversely, we also examined Route 3 in more detail. Route 3 is extremely short, only 1.75 miles and thus almost too much so to be truly considered a "route". However it is one of the East Bay approaches to the San Francisco Bay Bridge, and thus a highly congested segment. The results for this route are all over the place. Overall, it looks at first glance as if the 511 estimates for this short stretch are consistently wrong. Relevance measures below 50% can be interpreted to mean that it is more likely for a driver to receive a poor estimate than a correct estimate.

Table 5. Example of Route 3 (I80-E, 1.75-mile, from I-580 Split to Powell St. Exit)

Date	AE	R10	R15	R20
5/5/2009	12.3%	15.3%	34.5%	69.3%
5/6/2009	18.6%	8.6%	26.9%	58.4%
5/7/2009	-8.7%	6.9%	17.6%	36.0%
5/26/2009	0.5%	17.8%	34.5%	47.2%
6/1/2009	12.2%	28.8%	70.4%	88.8%
6/2/2009	14.4%	15.3%	34.9%	67.4%
6/3/2009	15.7%	13.3%	41.2%	74.8%

6/4/2009	9.8%	25.9%	48.3%	69.8%
6/5/2009	13.1%	15.2%	33.3%	63.8%
6/6/2009	19.4%	0.0%	12.2%	63.5%
6/9/2009	12.5%	13.8%	31.3%	62.7%
6/10/2009	22.8%	0.0%	3.7%	29.3%

A look at the raw data seems in order. It is presented on Figure 6 for one day on May 5, 2009. It shows two features of interest. First, the ground truth travel times become extremely dispersed during peak time. This means that there are large individual differences between drivers, and thus no one estimate can do a proper job at capturing the range of actual travel times. Thus the benchmark relevance measures are necessarily low during the peak. The second feature is that we can tell graphically that the 511 estimates track the overall trend remarkably well. This calls for a couple of key comments. The first one is to recognize that the benchmark metrics we developed are based on relative errors. For shorter routes with more volatile variations, this becomes penalizing and is less meaningful anyway because large percentage errors can still correspond to small absolute errors, which is what drivers care about. However, as explained in the methodology section, relative errors offer the advantage to be normalized, which is essential for a benchmark methodology. A second comment requires the understanding that the 511 estimates on route 3 are derived from the same FasTrak data that we used as ground truth. However, because estimates are run in real-time, before the corresponding FasTrak traces have been observed. In other words, past traces are used to generate current estimates, and there is latency between the 511 information and the ground truth data. The net result is that the estimates closely follow the ground truth, which is apparent on the plots. Yet because the travel times vary so rapidly, the estimates can still be significantly off at any given time.



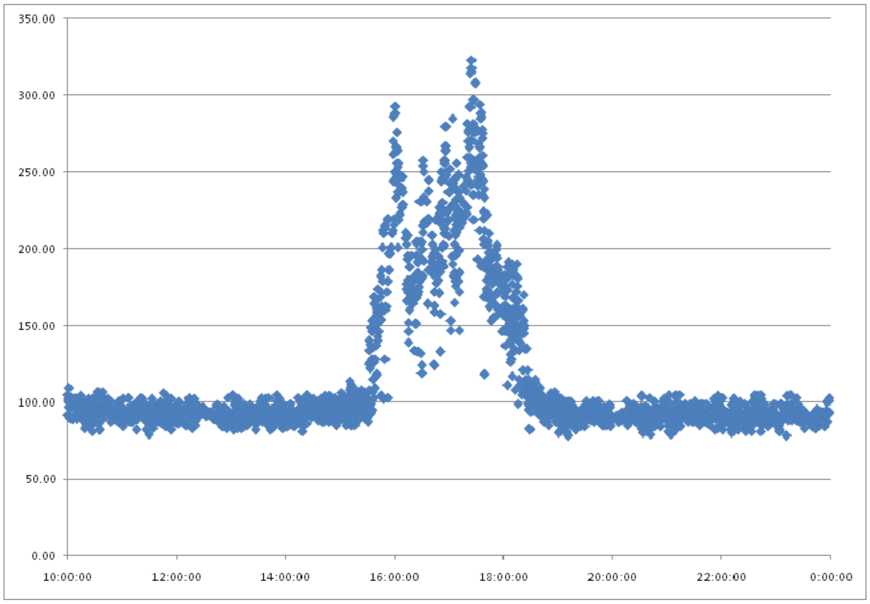


Figure 6. Comparison between travel time estimates and ground truth data (route 3, May 5, 2009)

This is a situation that the benchmark metrics we developed cannot capture. Regardless, it is debatable whether an estimation scheme that accurately tracks ground truth travel times with some latency must be seen as good or not. On the one hand, it does a good job of following the trend, but on the other hand stale information is useless to travelers. One of the premises of our methodology is to be user-centric, and it is therefore coherent that the benchmark measures indicate mediocre results for route 3, in spite of the aforementioned caveats.

CONCLUSION

This paper set out to examine how a previously developed benchmark methodology for travel time estimates could perform with a large sample of routes over an extended period of time. We found that the methodology worked quite well, combining practicality, flexibility and rigor. The two benchmark metrics capture comprehensive and complementary information about the differences between travel time estimates and ground truth data. Errors due to latency only are not flagged, which is possibly penalizing. Yet it is not clear whether or not this would be a desirable feature.

We wanted to determine a good threshold for the relevance measure. We find that 10% holds the most explanatory power, but it can also bring confusing results because it is more stringent than is truly realistic. Thus 15% remains the best all-around value.

We hope that this work can help with the development of a universally accepted methodology and metrics to benchmark traffic information quality. Note that one of the limitations of this study is that it only addresses freeways. Future work should be conducted on signalized arterials, which present a different environment with data a lot noisier.

ACKNOWLEDGMENTS

We wish to thank the 511 team at the Metropolitan Transportation Commission as well as Caltrans for making ground truth travel time data available for this study. We also thank our colleagues at the California Center for Innovative Transportation who have designed, implemented and maintained the MITTENS system without which the collection and processing of data could not have taken place.

REFERENCES

1. Khattak, A., Kanafani, A., and Colletter, E.L. Stated and reported route diversion behavior: implications of benefits of advanced traveler information system. Journal of Transportation Research Board 1464, 1994, pp28-35.
2. Lindveld, C.D.R., Thijs, R., Bovy, P.H.L, and Van der Zijpp, N.J. Evaluation of online travel time estimators and predictors. Journal of Transportation Research Board 1719, 2000, pp45-53.
3. Peng, Z.R., Guequierre, N., and Blakeman, J.C. Motorist response to arterial variable message signs. Journal of Transportation Research Board, 1899, 2004, pp55-63.
4. FHWA. Travel Time Reliability: Making It There On Time, All the Time. www.ops.fhwa.dot.gov/publications/tt_reliability/TTR_Report.htm. Accessed July, 2009.
5. Oda T. An algorithm for prediction of travel time using vehicle sensor data. In Proceedings of Third International Conference on Road Traffic Control, 1990, pp40-44.
6. Smith, B.L., and M.J. Demetsky. Traffic flow forecasting: comparison of modeling approaches. Journal of Transportation Engineering, 123(4), 1997, pp261-266.
7. Huisken, G., and M.v. Maarseveen. Congestion prediction on motorways: a comparative analysis. In Proceedings of the 7th world congress on ITS (CD-ROM), 2000.
8. Rice, J., and Zwet, E.V. A simple and effective method for predicting travel times on freeways. In Proceedings of IEEE Intelligent Transportation Systems, Oakland, CA, 2001, pp227-232.
9. Hartley, J.K. Prediction of link travel times in the context of Nottingham's urban road network. In Proceedings of the 17th European Simulation Multi-Conference, 2003, pp423-428.
10. Hinsbergen, C.P.IJ, Lint, J.W.C., and Sanders, F.M. Short-term prediction models. In Proceedings of the 14th world congress on ITS (CD-ROM), 2007.
11. Margulici, J.D., and Ban, X. Benchmarking travel time estimates. IET Journal of Intelligent Transportation Systems, 2008.
12. Pan, X., Margulici, J.D., and Yazbek, R. Treatment and Analysis of Ground Truth Travel Time Data on Urban Freeway Corridors. Submitted for presentation and publication to 89th Annual Meeting of the Transportation Research Board, Washington, D.C., 2010.
13. Ban, X., Li, Y., and Skabardonis, A. Local MAD Method for Probe Vehicle Data Processing. In Proceedings of the 14th ITS World Congress, 2007.
14. Liu, X.H., He, X.Z., Recker, W. Estimation of the time-dependency of values of travel time and its reliability from loop detector data. Transportation Research, Part B, 41(4), 2007, pp448-461.
15. Zhang, X.Y., Rice, J., and Bickel, P. Empirical Comparison of Travel Time Estimation Methods. California Path Research Report, UCB-ITS-PRR-99-43, 1999.

16. Van Lint, J. W. C., and H. J. van Zuylen. Monitoring and Predicting Freeway Travel Time Reliability, Using Width and Skew of Day-to-Day Travel Time Distribution. Journal of the Transportation Research Board, No. 1917, 2005, pp. 54-62.
17. Fei, X., Mahmassani, H. S., and Eisenman, S. M. Sensor Coverage and Location for Real-Time Traffic Prediction in Large-Scale Networks. Journal of the Transportation Research Board, No. 2039, 2007, pp. 1-15.
18. Kothuri, S. M., Tufte, K. A., Fayed, E., and Bertini, R. L. Toward Understanding and Reducing Errors in Real-Time Estimation of Travel Times. Journal of the Transportation Research Board, No. 2049, 2008, pp. 21-28.
19. Guo, J., Smith, B. L., and Zhang, G. Information Technology Used to Evaluate Nontraditional Traffic Monitoring Systems. Journal of the Transportation Research Board, No. 2024, 2007, pp. 1-7.
20. Kwon, J., McCullough, B., Petty, K. and Varaiya, P. Evaluation of PeMS to Improve the Congestion Monitoring Program. Final Report for PATH TO 5319, 2006.
21. Fujito, I., Margiotta, R., Huang, W., and Perez, W. A. Effect of Sensor Spacing on Performance Measure Calculations. Journal of the Transportation Research Board, No. 1945, 2006, pp. 1-11.
22. Dance, F., Gawley, D., Hein, R., and Kates, R. Enhancing navigation systems with quality controlled traffic data. SAE International, 2007.

Treatment and Analysis of Ground Truth Travel Time Data on Urban Freeway Corridors

Xiaohong Pan
California Center for Innovative Transportation
University of California, Berkeley
2105 Bancroft Way, Suite 300
Berkeley, CA 94720-3830
Email: xpan@calccit.org

J.D. Margulici*
California Center for Innovative Transportation
University of California, Berkeley
2105 Bancroft Way, Suite 300
Berkeley, CA 94720-3830
Email: jd@calccit.org

Rudy Yazbek
American University of Beirut
P.O.Box 11-0236
Riad El-Solh, Beirut 1107 2020, Lebanon
Email: rry02@aub.edu.lb

July, 2009

Word count 4505 + 2500 (5 figures + 5 tables) = 7005

Submitted for Presentation and Publication to:
Transportation Research Board 89th Annual Meeting
Washington, D.C.

* Corresponding Author

Treatment and Analysis of Ground Truth Travel Time Data on Urban Freeway Corridors

ABSTRACT

This paper examines a very large set of ground truth travel time data collected from toll tag data in the San Francisco Bay Area. Our purpose is to introduce a standard practice for evaluating the quality of traffic information. To that end, we propose benchmark metrics in a different paper based on the same study. The present paper focuses on the treatment and analysis of ground truth travel time data, including outlier removal, distribution shape, issues of required sample size, and variations, both within a sample and between samples. We draw conclusions in each of these areas, which will interest both researchers and practitioners.

Treatment and Analysis of Ground Truth Travel Time Data on Urban Freeway Corridors

1. INTRODUCTION

1.1. Background and Problem Statement

Travel time estimates on selected itineraries are a natural indicator of traffic congestion and are well understood by the traveling public. Travel time information can be disseminated through various channels, including the internet, mobile devices, navigation units, and dynamic message signs. Numerous studies reveal that travelers appreciate and value travel time information, especially because it reduces travel uncertainty and stress (*1-3*). In theory at least, travelers can also respond to real time information by adjusting their travel schedules, travel routes, or even travel modes (*4*). From an operational standpoint, travel times constitute an effective performance metric to describe the level of service on a given roadway segment.

The quality of travel time estimates generally relies on sufficient traffic data collection, as well as adequate processing techniques and information delivery modes. Each of these steps has benefited from a flurry of innovations over the past decade (*5-10*). Nevertheless, literature on systematic evaluations of the quality of travel time estimates remains limited. Our research team has conducted research on this issue and highlighted its importance (*11*). First, real-time traveler information is becoming ubiquitous and demand is driven by mobile devices and location-based services. Where users were previously content to simply get any data, quality starts to matter a lot more. Second, the emergence of private aggregators of real-time traffic information that collect data from mobile probes means that Departments of Transportation now have the option to purchase third-party feeds instead of deploying their own detectors. In this context, public agencies need a methodology to properly assess and continuously audit data quality. Our previous publication (*11*) suggested such a methodology and highlighted a set of two standardized metrics aimed at benchmarking travel time estimates.

1.2. Research Objectives/Questions

The present paper was written in conjunction with a second, related paper (*12*). The two papers split results obtained from further work done around the methodology and the metrics that we have already exposed. Our focus here is on practically implementing the methodology as well as validating the adequacy of the metrics to benchmark estimates in a variety of situations.

With respect to practical implementation, we primarily integrate an outlier removal methodology previously developed by one of us (*13*) and further investigate its applicability. Regarding the two metrics used to benchmark travel times estimates, we were interested in testing them on a much larger number of routes than we had done prior. The metrics need to achieve two objectives: a) synthetically capture as much information as possible about the quality of various estimates over time and along different itineraries; and b) provide a graded scale that offers good discriminating power between mediocre estimates and accurate estimates – in other words, the metrics must be calibrated such that they immediately and reliably distinguish good from bad. In order to assess how well those two objectives are attained, empirical verification was required. In the process, we also realized that we needed a finer comprehension of how collected ground truth travel times vary, both within a given sample and over time. Obviously, such a comprehension is instrumental to developing proper estimates in the first place, and we therefore think that our work has inherent value for traffic modelers. Systematic analysis of ground truth travel time data along a large set of itineraries is hard to come by in the literature, probably because such data has not been widely available until fairly recently.

In sum, we considered the following research questions in this expanded travel time benchmark study:

- What could be a sufficient/reasonable sample size of ground truth travel time for the purpose of benchmarking estimates?
- How do ground truth travel times vary in the first place? In particular, what metrics can be designed to feature different variance components due to individual drivers and the effect of congestions?
- How do the previously proposed benchmarking metrics perform in our comprehensive experiment?

This paper addresses the first two questions, while the companion paper (*12*) responds to the third question. These results are intended to help researchers and practitioners with theoretical and practical knowledge about travel time variations and estimates benchmarking. A broader goal is to feed into a conversation about the proper assessment of traffic information quality and foster the adoption of standards that can serve the industry by accelerating the dissemination and commercialization of data.

A short literature review section is followed by a methodology section in which we describe the data sets used for this study and the descriptive statistics that we employ to characterize ground truth travel time data. We present and analyze our results in section 4.

2. LITERATURE REVIEW

2.1. Difference between Quality of Travel Time Estimates and Travel Time Reliability

Because of terminology, it appears necessary to clearly distinguish between evaluating the quality of travel time estimates and studying travel time reliability (*11*). The topic of travel time reliability had recently received a lot of attention in the literature (*14-18*). The concept focuses on the distribution of actual travel times, so as to help travelers, shippers, and network managers better understand the performance and reliability of the roadway network (*19*). The benchmarking of travel time estimates requires an assessment of the differences between estimated and actual travel times, which tells travelers and transportation professionals about the accuracy and reliability of the estimations of travel times –rather than the travel times themselves! Yet clearly, variations in ground truth travel times, and thus travel time reliability, affect the quality of estimates. This paper calls heavily on the concept of travel time reliability in order to explicate natural travel time variations and their bearing on the estimation task.

2.2. Requirements for Benchmarking Travel Time Estimates

In order to systematically process the benchmark of travel time estimates, the research team has summarized several necessary conditions (*11*):

- Significant volumes of ground truth data with sufficient reliability: ground truth travel time, or baseline travel time, means observed trip times of vehicles traveling on the roadway segments where the benchmark is conducted. This requirement has been one of the main obstacles for systematic validation studies in the past, though various methods have been attempted (*20*). Nowadays, the increasing deployment of ITS technologies has enabled the provision of actual trip travel times to transportation agencies, whether from toll tag readers, license plate readers, Bluetooth readers, magnetic signatures, or GPS probes.
- Clearly defined, widely accepted, and standardized quality assessment measures: no such metrics have been developed in the literature. Part of the reasons may be that most researchers are more interested in the underlying phenomenon than in its measurement. Then again, the fact that actual travel times are distributed rather than uniform creates a bit of a hurdle. But mostly, the absence of large volume of available and reliable ground truth data may have occulted the need to develop more systematic quality evaluation metrics.
- Sophisticated data collection, processing, and storage platform: a systematic benchmark needs to accumulate sufficient data to allow meaningful comparisons. This requires an upfront investment in a software infrastructure than can effectively and flexibly store and process this data, which is not to be discounted. Guo et al. (*21*) published a paper that insisted on system design in order to carry out a structured evaluation procedure for link speed and travel time estimations.

3. METHODOLOGY

3.1. Conceptual Framework

In our earlier study, we proposed benchmark metrics and tested them using a small sample of traffic data from the San Francisco Bay Area. For more information regarding the methodology and results of the earlier study, please see the paper prepared by Margulici and Ban (*11*). As the study suggested, we enhanced the development of quality assessment metrics and expanded the scope of evaluation experiments, both in time and coverage. The conceptual framework is illustrated in Figure 1, including two major interactive tracks – metric development and experiment design. Detailed discussion for these two tracks is presented in the following sections.

Before discussing the development metrics for travel time estimates, we find it necessary to point out a few considerations and assumptions we used in our research:

- Our objective is to come up with a set of clearly defined, simple, and straightforward benchmark metrics. Additionally, those metrics should be as standardized as possible, so that they can easily be applied in various settings and with different sources of data. We do not pretend to have reached a final answer and rather hope to open a debate about what constitutes good metrics for the industry to move forward.

- The proposed benchmarking methodology assumes the existence of significant volumes of ground truth travel times with sufficient reliability. As mentioned in the previous section, this requirement is becoming much less stringent thanks to the increasing deployment of ITS and the constant development of innovative and lower cost information technology solutions. In the present work, travel times collected from the San Francisco Bay Area's FasTrak toll tag readers were used as ground truth travel time.
- For the purpose of analysis, we divide time periods into elementary time slots. A time slot i may be set to 5 minutes, 10 minutes, 1 hour etc. We mostly stick to symbolic notations in this section and shift to numerical examples in the presentation of results.

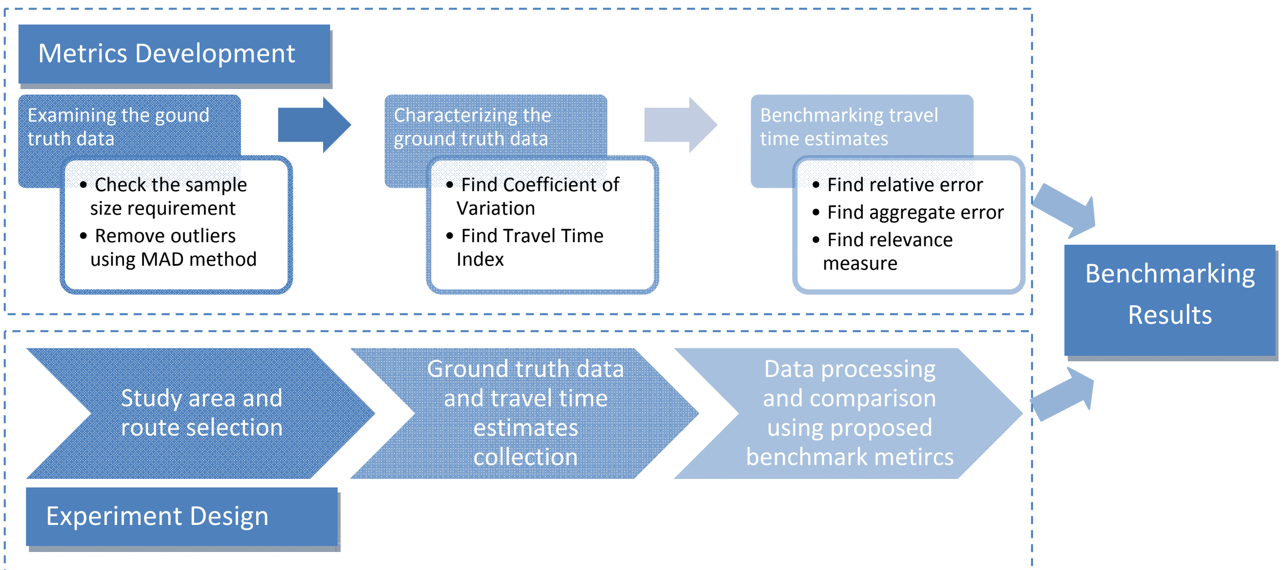


Figure 1. Conceptual framework

3.2. Examining the Ground Truth Data

3.2.1. Sample Size Adequacy

One of the great advantages of collecting ground truth travel times from toll tag readers is that the sample sizes end up fairly large on most routes and for most of the time. Nonetheless, sample size is a critical component of ground truth data collection and we set out to determine a threshold that would allow us to accept or reject the data collected during a given time slot based on the number of observed trip times.

Let tt_{i1}, tt_{i2} ... tt_{ij} be observed travel times on a selected roadway segment during time interval i, where tt_{ij} refers to the *j*th ground truth travel time observed during that time. One of our assumptions is that the length of time intervals is small enough that travel times can be considered to be homogeneously distributed within each interval (5-minute intervals constitute a typical standard in this circumstance, and 15-minute intervals still work as a reasonable approximation). Further, we assume the travel times within interval i to be normally distributed, a feature that we verify in the next section of this paper. Let μ_i be the true mean and σ_i be the true standard deviation of travel times within interval i.

Let $\overline{tt_i} = \frac{\sum_{j=1}^{J} tt_{ij}}{J}$ be the sample mean of travel times over time interval i, where J refers to the total number of ground truth travel time observations during time slot i.

Let $SD_i = \sqrt{\frac{\sum_{j=1}^{J} (tt_{ij} - \overline{tt_i})^2}{J-1}}$ be the sample standard deviation of travel times for time interval i.

We know $\frac{\overline{tt_i} - \mu_i}{SD_i/\sqrt{J}} \sim T_{J-1}$, where T_{J-1} is a random variable which follows a Student's t-distribution with $J-1$ degrees of freedom.

Thus for a fixed tolerance probability, say significance level α (or for a $1-\alpha$ confidence interval), and a fixed band width of travel time distribution at time interval i, say B_i, we have

$$\alpha \le P(|\overline{tt_i} - \mu_i| \le B_i) = P\left(\frac{|\overline{tt_i} - \mu_i|}{SD_i/\sqrt{J}} \le \frac{B_i\sqrt{J}}{SD_i}\right) = P\left(|T_{J-1}| \le \frac{B_i\sqrt{J}}{SD_i}\right).$$

We can then search the t table to find the smallest n for which the above inequality holds. In our analysis, we set $\alpha = 0.05$ and $B_i = 10\%\overline{tt_i}$. The minimum sample size requirements can then be determined for each route and for each time interval for which we have large enough samples. In turn, we can generalize these results to estimate sample size adequacy.

3.2.2. Outlier Removal Using the MAD Method

With large data samples, and irrespective of the kind of data source employed to collect it, a substantial number of outliers that must be filtered is inevitable. As suggested by Ban et al. (*13*), we apply the Median Absolute Deviation (MAD) method to remove outliers, where MAD is a statistical measure for capturing variations within a set of data points. For a given time interval i, MAD is calculated as follows:

$$MAD_i = median(|tt_{ij} - median(tt_{ij})|)$$

Outliers can be detected by comparing the z-score z_j with a given threshold $\bar{z}$, where z_j is defined as:

$$z_j = \frac{|tt_{ij} - median(tt_{ij})|}{MAD_i}$$

If $z_j \ge \bar{z}$, then tt_{ij} is an outlier. Ban et al. conducted extensive analysis based on ground truth driving times across the Golden Gate bridge and determined that the optimal value for $\bar{z}$ was 4.5 (*13*). Our own informal analysis confirms that 4.5 is a good value. A $\bar{z}$ value higher than

4.5 leaves obvious outliers after the MAD method is performed. On the other hand, a lower $\bar{z}$ value may mistakenly remove large but valid travel time observations due to non-recurring traffic congestions.

3.3. Characterizing the Ground Truth Data

Different features and variations of actual travel times influence the performance of travel time estimations. Here, we attempt to capture the travel time variations as comprehensively and synthetically as possible. Several types of variation for ground truth travel times are discussed, i.e., individual differences between drivers for the same roadway segment at the same time, intra-day travel time variations due to traffic congestion patterns, and day-to-day variations.

3.3.1. Capturing Individual Differences

In order to capture the magnitude of individual differences, we use the coefficient of variation (CV), a normalized measure of dispersion within a data set. The coefficient of variation is defined as the ratio of the standard deviation to the mean:

$$CV_i = {}^{SD_i}/_{\overline{tt_i}}$$

CV presents the advantages over other dispersion measures to be normalized and dimensionless. It is therefore well suited to establish comparisons across routes of various lengths and at various times.

In order to provide a measure of the overall effect of individual differences on a given route, we aggregate the time series of CV_i obtained over a longer period of time, such as an entire day, by extracting the median. The median CV of ground truth travel time in day t is defined as $CV^{t,50th}$, which holds all the same features as CV_i for an individual time slot. It can be easily applied for comparison of the contribution of individual trip time differences across multiple days and road segments. The Median CV can also be extracted for a set time period during to measure peak-hour individual travel time variations.

3.3.2. Capturing Intra-Day Travel Time Variations

To characterize intra-day travel time variations, we use the Travel Time Index (TTI). This metric was initially introduced to measure travel time reliability. It captures the average congestion level on a roadway segment over a set time period (*4*). For each time slot i, TTI is computed as the average ground truth travel time divided by the free flow travel time, as follows:

$$TTI_i = \frac{\overline{tt_i}}{tt_{ff}}$$

This formulation, which is also normalized and dimensionless, captures the ratio of additional time required to travel along a segment when it is congested during time slot i, as compared to a light traffic situation. Because of these properties, TTI can be aggregated for multiple time slots and is not dependent on travel distance. It is therefore suitable for comparison across different roadway segments.

In order to compare the intra-day variations of ground truth travel times between routes, we aggregate the time series TTI_i in two way, i.e., Aggregate Travel Time Index (ATTI) for day t, and 95th Percentile Travel Time Index for day t.

- ATTI is defined as the sum of absolute TTI values over a selected day divided by the total number of time slots in that day, $ATTI^t = \frac{\sum_{i=1}^{I}|TTI_i|}{I}$, where I refers to the total number of time slots in day t. ATTI captures the average excess travel time compared to the free flow situation.
- The 95th Percentile Travel Time Index is defined as the 95th percentile of all computed travel time indexes in day t, $TTI^{t,95th}$, which captures the near worse case travel times away from the free flow situation.

The 95th Percentile TTI Index complements ATTI by measuring the height of the peak, where ATTI measured its overall magnitude. A low ATTI value does not necessarily mean that intra-day travel time variations are less, because there may be severe traffic congestion happening for a short period of time, a feature that will be revealed by the 95th Percentile TTI Index. Both ATTI are normalized, dimensionless metrics that can directly be compared across routes and days.

3.3.3. Capturing Day-to-Day Travel Time Variations

Day-to-day travel time variations are driven by congestion patterns and are thus no different in nature from intra-day variations. We assume that if a particular travel time estimation mechanism properly captures intra-day travel time variations, then it should be able to handle day-to-day variations as well. Therefore, we don't provide specific metrics for day-to-day variations because they add no further information that we can bring to bear to explicate the performance of travel time estimates on given routes.

3.4. Experiment Design

Compared to our previous study, we propose a more comprehensive experiment to examine the variability factors that affect travel time estimates and to assess the validity of the travel time estimates benchmark methodology and technique.

3.4.1. Study Area and Route Selection

The study area was set to the San Francisco Bay Area due to the remarkable advantage of its traffic data availability. The experiment was hosted by the MITTENS[1] system, which generates travel time estimates using a variety of sources, such as loop detectors and radars. Another

[1] MITTENS stands for "Messaging Infrastructure for Travel Time Estimates to a Network of Signs", which is an automated system developed by CCIT for Caltrans District 4 to display travel times on CMS in the Bay Area.

important resource is the "ground truth" traffic data collected from FasTrak toll tag readers, which can be retrieved in the MITTENS system in disaggregate form or in the PeMS[2] system at an aggregated level.

The experiment routes form a representative set of the various features that distinguish them from one another. These include sample size of FasTrak data (small/moderate/large), type of area (urban/suburb/rural), geographic location, whether comprising a bridge or not, as well as magnitude of the variations of ground truth travel times (small/moderate/large). It is made up of 20 highway segments around the Bay Area. We encoded these 20 routes into Google Maps for reference[3]. Additionally, we also include those 4 experiment routes used in our earlier study. An overview of the total 24 study routes is presented in Table 1.

Table 1. Overview of the Study Routes

Route ID	Length	Freeway	Features
1	9.14	SR-24E	-
2	6.48	I80-E	including bay bridge
3	1.75	I80-E	relatively short distance
4	7.75	I80-E	-
5	3.00	I80-E	relatively short distance
6	3.25	I80-E	suburban area; relatively short distance
7	26.50	I80-E	suburban area; relatively long distance
8	6.48	I80-W	including bay bridge toll plaza
9	7.84	SR84-W	-
10	1.75	SR92-E	suburban area; relatively short distance
11	11.03	SR92-W	-
12	25.00	I880-N	-
13	8.00	I680-N	-
14	41.33	I680-N	suburban area; relatively long distance
15	7.75	I580-W	-
16	10.75	SR101-N	-
17	4.95	SR101-N	urban area
18	16.67	SR101-N	suburban area
19	11.75	SR101-N	suburban area
20	2.49	SR101-S	-
21	13.90	I80-E	-
22	3.15	I80-W	-
23	9.00	I880-S	-
24	7.75	I580-W	-

[2] PeMS stands for Freeway "Performance Measurement System", which is a web-based tool designed by UC Berkeley to host, process, retrieve, and analyze road traffic condition information.
[3] http://maps.google.com/maps/ms?ie=UTF8&hl=en&oe=UTF8&msa=0&msid=110448774267712153593.00046931135a9cc4d5c52&start=0&num=200&z=9

3.4.2. Ground Truth Travel Time and Travel Time Estimates Collection

Ground truth travel times for the 24 experiment routes selected were systematically retrieved within the MITTENS system from early April to early June, 2009. Travel time estimates generated from 511 data were also collected for the same routes and the same time period.

3.4.3. Data Processing and Comparison Using Benchmark Metrics

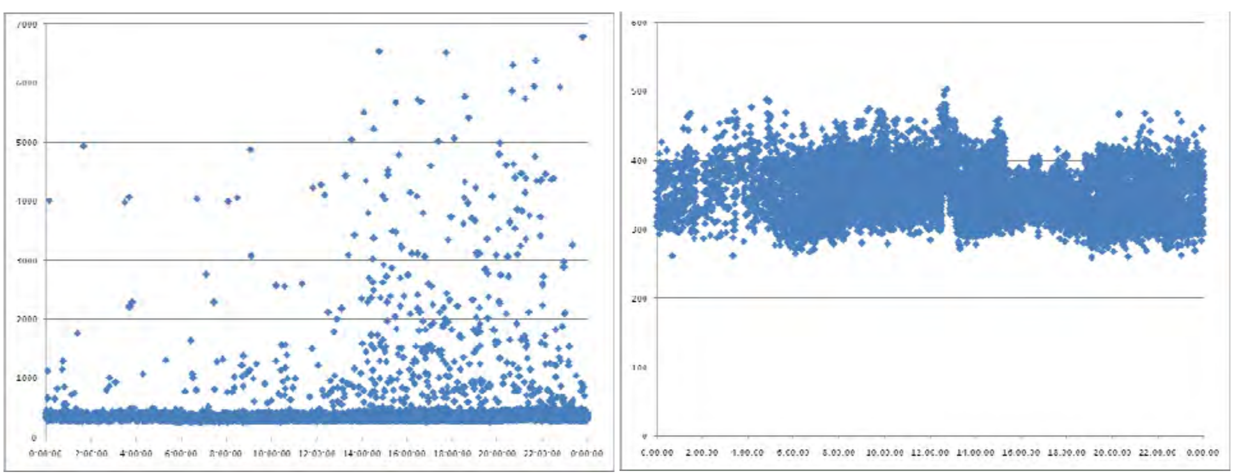
The first data processing step consists in rejecting observations for which the minimum sample size requirement is not met. We remove the outliers among those for which it is. We then analyze the features of the ground truth travel times, including individual differences and congestion-driven variations. Finally, we compare those ground truth travel times with the travel time estimates using the proposed benchmark metrics. We conclude by examining the comprehensiveness and discriminating power of the metrics across the entire route set.

4. RESULTS AND FINDINGS

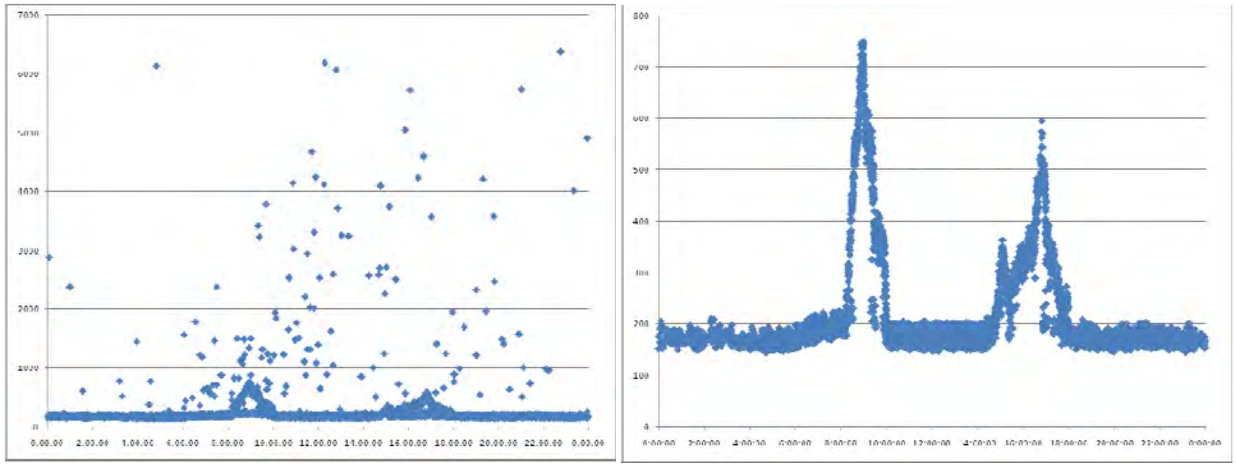
4.1. Outlier Removal Procedure

By conducting sensitivity analysis on toll tag data, Ban et al. (*13*) found that, "if data samples are fairly large, a band width of 15-30 minutes can be chosen in order to capture the trend of travel times and the statistical rigor of the local MAD method." We tested different band width between 15 and 30 minutes. The 15-minute time interval was found to work best for our research purpose.

Figure 2 illustrates the outlier removal procedure by showing before-and-after results for route 4 on May 18, 2009, and route 22 on June 5, 2009, respectively. It exemplifies the effectiveness of the procedure, both when travel times vary little and when they feature pronounced peaks.



(a) (b)



(c) (d)

Figure 2. Examples of travel time data before and after outlier removal

(a) Raw FasTrak travel time data for route 4 on May 18, 2009; (b) processed FasTrak travel time data for route 4 on May 18, 2009; (c) raw FasTrak travel time data for route 22 on June 5, 2009; (d) processed FasTrak travel time data for route 22 on June 5, 2009.

Table 2 shows the percentage of outliers removed for a subset of routes. It indicates that there exists a substantial amount of outliers within the FasTrak "ground truth" travel times, and confirms the importance of the removal procedure.

Table 2. Percentage of Outliers Removed

Route ID*	1	3	4	6	9	10	11	12	13	15	16	18	19	20	22	23
% of Outliers Removed	7.9	10.0	14.8	11.1	7.7	9.2	10.3	16.8	10.9	8.0	17.8	14.4	16.7	11.1	6.8	14.2

* Some routes are not reported due to the ground true data being unavailable.

4.2. Sample Size and Distribution

Consistent with the MAD outlier removal procedure, we use 15-minute time intervals as the base unit to examine issues related to sample size and distribution. Also, as mentioned in the methodology section, we used 95% confidence interval ($\alpha = 0.05$) and $\pm 10\%$ of the sample mean as the acceptable band width for mean estimation ($B_i = 10\%\overline{tt_i}$). Table 3 shows sample size requirement calculations and percentages of time intervals those requirements.

As our numerical results show, the median sample size requirement for most routes is 3-5. Further, a sample size of 5-6 is adequate for 95% of the collected samples on a majority of routes. A sample size of 8 nearly universally guarantees adequacy. If we roll back outliers in the sample size, then we can loosely state that a sample size of 5-10 at the initial data collection step is sufficiently large based on the criteria we selected. As the table shows, the calculations yield a higher number in a very few cases, but these can be attributed to statistical singularities given their rare occurrence. In our data processing, we used a conservative minimum of 15 samples for each 15-minute time interval as a filter.

Table 3. Requirement of Sample Size and Percentage of Time Intervals Qualified

Route ID*	1	3	4	6	9	10	11	12	13	15	16	18	19	20	22	23
% of Intervals Qualified**	100	99.2	99.7	99.2	100	99.9	100	98.3	100	100	99.9	100	99.6	100	100	100
Required Sample Size (Min)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Required Sample Size (Median)	3	4	6	4	4	4	4	4	4	3	5	3	4	5	4	4
Required Sample Size (95th Percentile)	4	15	21	8	5	6	5	8	6	4	11	5	6	6	8	6
Required Sample Size (Max)	20	89	99	79	41	60	34	68	13	21	91	13	74	54	59	16

* Some routes are not reported due to the ground true data being unavailable.
* *Percentage of time intervals that contain sufficiently large sample size.

We also tested the normality of the ground truth travel time data distribution. The normality of this data is a prerequisite to the method we used to determine the minimum sample size requirement within a time interval i. To that end, we selected time intervals with particularly large sample sizes and produced normal Q-Q plots to verify this feature. Q-Q plot, or Quantile-Quantile plot, is a graphical method for comparing two probability distributions by plotting their quantiles against one another. If the observed data matches the theoretical distribution (normal in our case), the plot aligns on a straight line. Figure 3 shows an example of normal Q-Q Plot for travel times collected from 12:00pm to 12:15pm, April 3, 2009 on route 1. The high degree of fit indicates a normal distribution of travel times in that time interval.

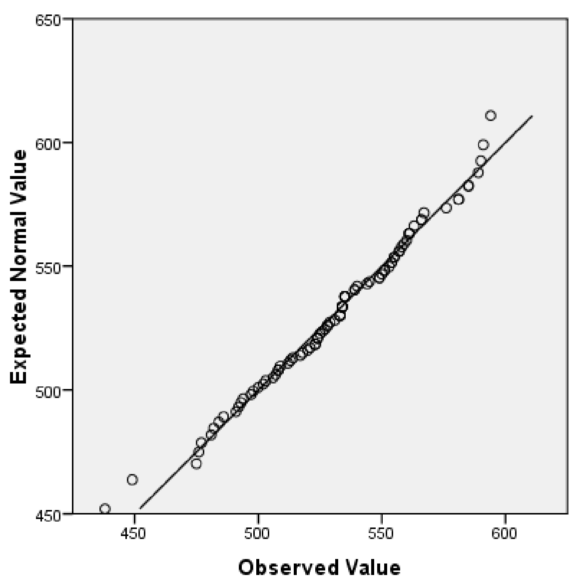


Figure 3. Example of normal Q-Q Plot for travel times (route 1, April 3, 2009, 12:00-12:15pm)

4.3. Intra-Sample Variations

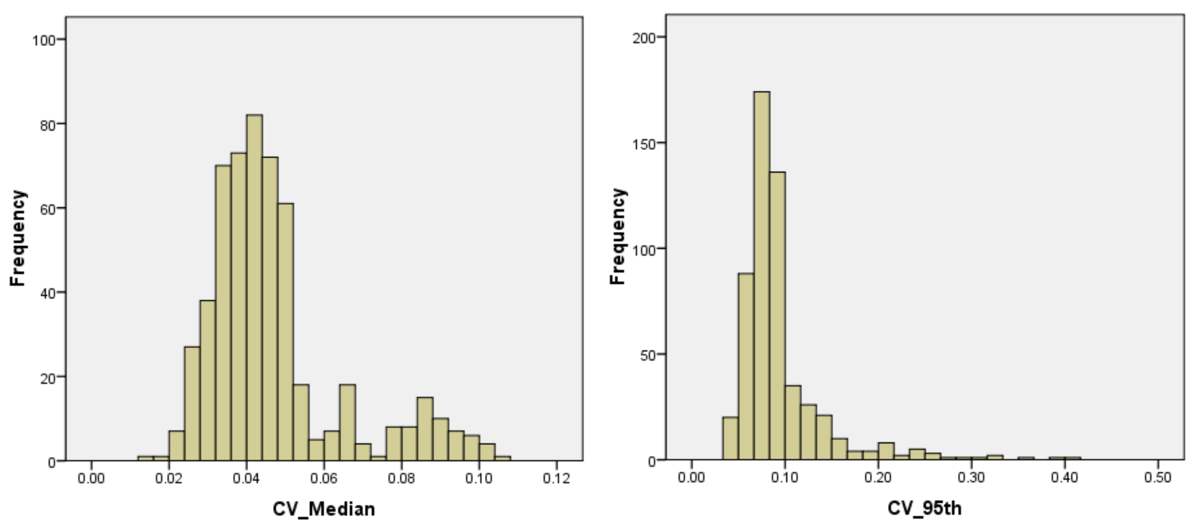
We calculated the coefficient of variation for each route and 5-minute interval in our study in order to characterize individual driving differences. Table 4 shows median CV values for each route. As can be seen, those values are quite small, and 5% appears to be a typical CV. In other words, the standard deviation of trip times within a given sample is about 5% of the trip times themselves. This result is important, because it shows that although individual differences must be taken into account, travel times on urban freeway corridors are not that dispersed.

Table 4. Median CV for Different Routes

Route ID*	1	3	4	6	9	10	11	12	13	15	16	18	19	20	22	23
Median CV	0.03	0.04	0.09	0.04	0.05	0.04	0.05	0.03	0.04	0.04	0.07	0.03	0.04	0.06	0.05	0.04

* Some routes are not reported due to the ground true data being unavailable.

Figure 4 presents the distributions of daily median CV and 95th percentile CV (for all dates and all routes). The daily median CV varies within a range of 0.03 to 0.05 for most cases, with a few higher daily median CV that reach around 0.09. Even daily 95th percentile values are for the most part less than 0.1, and almost exclusively below 0.2.



(a) (b)

Figure 4. Distribution of daily median CV and 95th percentile CV (n=544): (a) Median CV; (b) 95th percentile CV.

4.4. Traffic-Dependent Variations

Table 5 shows results obtained with calculations of ATTI and 95th percentile TTI for different routes. All but a few routes show values of ATTI significantly above 1. This suggests that congestion remains limited to certain times of day only.

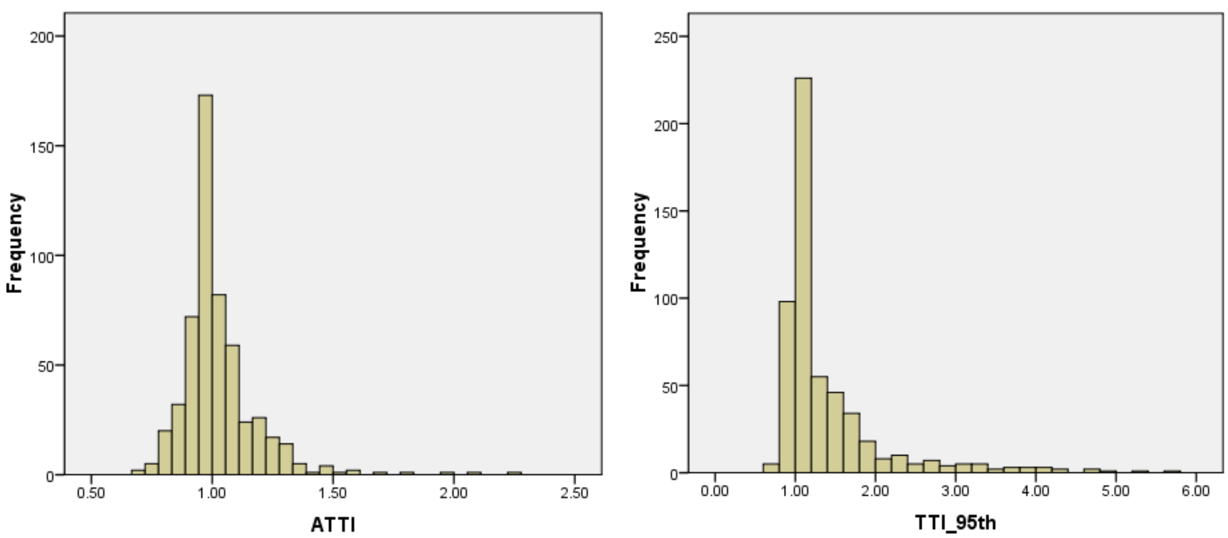
Peaks in traffic are revealed by the 95th percentile TTI values. Some routes (i.e. routes 9, 10, 11, 13, 15, 16, 20 and 23) are barely affected by recurrent traffic congestion. For other routes, a 60% surplus travel time or more is common. Routes 3 and 22 show particularly high peaks, but they also happen to be short routes (around 2 and 3 miles, respectively), which explains a higher volatility.

Table 5. ATTI and 95th Percentile TTI for Different Routes

Route ID*	1	3	4	6	9	10	11	12	13	15	16	18	19	20	22	23
ATTI	1.02	1.00	0.87	1.14	0.95	1.06	0.98	1.15	0.95	0.92	0.95	1.07	1.10	0.99	1.20	0.97
95th TTI	1.28	2.64	1.06	1.94	1.05	1.18	1.05	1.64	1.07	0.97	1.20	1.64	1.61	1.10	2.57	1.11

* Some routes are not reported due to the ground true data being unavailable.

The distributions of daily ATTI and 95th percentile TTI are shown on Figure 5. These are interesting in that they give a glimpse of more extreme cases. Thus ATTI can reach 1.5 and more on some routes and on certain days. The 95th percentile TTI reaches values of 2, 3, 4 and even 5. This means that on those days, trip times are 5 times greater than under free flow conditions over some period of time (you don't want to be driving in this traffic!)



(a) (b)

Figure 5. Distribution of daily ATTI and 95th percentile TTI (n=544): (a) ATTI; (b) 95th percentile TTI.

5. CONCLUSIONS

Our team examined a very large set of ground truth travel time data for the purpose of testing a previously developed benchmarking methodology. In this paper, we focus on descriptive statistics of the data set, which we think is highly instructive for both researchers and practitioners. In effect, few such studies have ever been attempted because large volumes of ground truth data is difficult to come by.

We confirmed the effectiveness of an outlier removal procedure using the MAD method. We used 15-minute time intervals and found the method to be effective and reliable. However, this result depends on the volume of ground truth data available and may not be generalized. After outliers are removed, travel times within a homogeneous sample appear to be normally distributed.

A significant conclusion that we think has wide applicability concerns minimum sample size. By basing our method on mean estimation, we find that sample sizes of 5-10 during a time interval over which the distribution of travel times can be assumed to be homogeneous are largely adequate. In many cases, a sample size of 3-5 after outlier removal seems sufficient. A related result is that the coefficient of variation for ground truth travel times within a homogeneous sample is quite low, typically less than 5%. In other words, individual differences between drivers on urban freeways tend to be small in relation to average trip times.

We find a variety of traffic conditions and impacts of congestion among the set of routes we studied. This is characterized by two metrics related to the travel time index, namely the average travel time index and the 95th percentile travel time index.

Benchmarking results are presented in a companion paper (*12*). In that paper like in the present one, we point out the fact that our study focuses on urban freeways. Signalized arterials have very different traffic features and could be the object of future work.

ACKNOWLEDGMENTS

We wish to thank the 511 team at the Metropolitan Transportation Commission as well as Caltrans for making ground truth travel time data available for this study. We also thank our colleagues at the California Center for Innovative Transportation who have designed, implemented and maintained the MITTENS system without which the collection and processing of data could not have taken place.

REFERENCES

1. Khattak, A., Kanafani, A., and Colletter, E.L. Stated and reported route diversion behavior: implications of benefits of advanced traveler information system. Journal of Transportation Research Board 1464, 1994, pp28-35.
2. Lindveld, C.D.R., Thijs, R., Bovy, P.H.L, and Van der Zijpp, N.J. Evaluation of online travel time estimators and predictors. Journal of Transportation Research Board 1719, 2000, pp45-53.
3. Peng, Z.R., Guequierre, N., and Blakeman, J.C. Motorist response to arterial variable message signs. Journal of Transportation Research Board, 1899, 2004, pp55-63.
4. FHWA. Travel Time Reliability: Making It There On Time, All the Time. www.ops.fhwa.dot.gov/publications/tt_reliability/TTR_Report.htm. Accessed July, 2009.
5. Oda T. An algorithm for prediction of travel time using vehicle sensor data. In Proceedings of Third International Conference on Road Traffic Control, 1990, pp40-44.
6. Smith, B.L., and M.J. Demetsky. Traffic flow forecasting: comparison of modeling approaches. Journal of Transportation Engineering, 123(4), 1997, pp261-266.
7. Huisken, G., and M.v. Maarseveen. Congestion prediction on motorways: a comparative analysis. In Proceedings of the 7th world congress on ITS (CD-ROM), 2000.
8. Rice, J., and Zwet, E.V. A simple and effective method for predicting travel times on freeways. In Proceedings of IEEE Intelligent Transportation Systems, Oakland, CA, 2001, pp227-232.

9. Hartley, J.K. Prediction of link travel times in the context of Nottingham's urban road network. In Proceedings of the 17th European Simulation Multi-Conference, 2003, pp423-428.
10. Hinsbergen, C.P.IJ, Lint, J.W.C., and Sanders, F.M. Short-term prediction models. In Proceedings of the 14th world congress on ITS (CD-ROM), 2007.
11. Margulici, J.D., and Ban, X. Benchmarking travel time estimates. IET Journal of Intelligent Transportation Systems, 2008.
12. Margulici, J.D., Pan, X., and Mellet, A. An Extended Benchmark of Travel Time Estimates on Urban Freeway Corridors. Submitted for presentation and publication to 89th Annual Meeting of the Transportation Research Board, Washington, D.C., 2010.
13. Ban, X., Li, Y., and Skabardonis, A. Local MAD Method for Probe Vehicle Data Processing. In Proceedings of the 14th ITS World Congress, 2007.
14. Chen, A., Yang, H., Lo, H.K., and Tang, W. A capacity related reliability for transportation networks. Journal of Advanced Transportation, 33, 1999, pp183-200.
15. Chen, C., Skabardonis, A., and Varaiya, P. A system for displaying travel times on changeable message signs. In Proceedings the 83rd Transportation Research Board Annual Meeting (CD-ROM), Washington, DC, 2004.
16. Fisher, K. et al. Performance Measures of Operational Effectiveness for Highway Segments and Systems. NCHRP Synthesis 311, 2003.
17. Emam E. B. and Al-Deek, H. Using Real-Life Dual-Loop Detector Data to Develop New Methodology for Estimating Freeway Travel Time Reliability. Journal of Transportation Research Board 1959, 2006, pp140-150.
18. Liu, X.H., He, X.Z., Recker, W. Estimation of the time-dependency of values of travel time and its reliability from loop detector data. Transportation Research, Part B, 41(4), 2007, pp448-461.
19. Lyman, K. and Bertini, R. L. Using Travel Time Reliability Measures to Improve Regional Transportation Planning and Operations. Journal of Transportation Research Board 2046, 2008, pp1-10.
20. Qi, Y., and B. L. Smith. Travel Time Data Quality Assessment: Development of Design to Test New Generation Traffic Monitoring Systems. Presented at 84th Annual Meeting of the Transportation Research Board, Washington, D.C., 2005.
21. Guo, J., Smith, B. L., and Zhang, G. Information Technology Used to Evaluate Nontraditional Traffic Monitoring Systems. Journal of the Transportation Research Board, No. 2024, 2007, pp. 1-7.

APPENDIX IV: COLLABORATION WITH THE NORTH AMERICAN TRAFFIC WORKING GROUP

The North American Traffic Working Group (NATWG) works collaboratively to define, accept and advocate for the unique needs of North America traffic information services. NATWG seeks to develop a coordinated, proactive market driven implementation of traffic and travel information services and products by both influencing international standards efforts and coordinating the development of non-competitive commercial agreements. NATWG was established within the Intelligent Transportation Society of America in late 2007. JD Margulici, who is the project manager of this task order, is one of the board members. Since mid-2009, JD has also been leading a task force within NATWG that aims to establish a unified methodology for traffic information benchmarking. Benchmarking traffic information with clear and consistent metrics is an essential requirement for the development of data services that may complement the data collected by states through their own infrastructure.

This appendix contains two documents:

- A presentation of NATWG and the activities under the traffic information benchmarking task force as of this writing. – 22 pages.
- The most current draft guidelines that have been developed by the task force. – 21 pages.

HARMONIZING TRAFFIC INFORMATION BENCHMARKS

North American Traffic Working Group

January 12th, 2010

About NATWG

2

The North American Traffic Working Group (NATWG) works collaboratively to define, accept and advocate for the unique needs of North America traffic information services. NATWG seeks to develop a coordinated, proactive market driven implementation of traffic and travel information services and products by both influencing international standards efforts and coordinating the development of non-competitive commercial agreements.

Members sampling:

Traffic Information Quality

- Traffic information has become abundant but quality remains seldom monitored
  - End users are relatively clueless about information quality
  - Margins of error are not well understood and used in practice
- There are no widespread metrics or evaluation procedures to measure data quality
  - Each customer (e.g. car manufacturer, DOT…) conducts its own benchmark
  - Evaluation results cannot be readily compared

<u>Postulate:</u>

- Harmonized benchmarking methods would benefit both suppliers and customers
  - Improve consistency and fairness of evaluations
  - Lower overall costs by eliminating duplication of efforts
  - Better recognize true value-added and pull quality upward

NATWG's Data Quality Efforts

- Objective: agree on and publish guidelines on how to measure and report traffic information quality

- Process to date:
  - January-June 2009: Committee-level discussions
    - Each provider disclosed its data evaluation procedures
    - Concluded with synthesis at ITS America's annual conference
  - July-December 2009: Task force
    - Starting point: single floating car as ground truth collector
    - Developed draft guidelines that include procedures and metrics

Webinar Goals

- Present NATWG's product to date
  - Guiding principles and process
  - Premises of the guidelines
  - Content and organization of the document at a glance
  - Gaps, voluntary omissions and next steps
- Recruit stakeholders to participate in the task force
  - Obtain further process buy-in and legitimacy
  - Collect feedback on content to move forward

Task Force

- Current Members:
  - J.D. Margulici, California Center for Innovative Transportation
  - Matt Lindsay, NAVTEQ
  - Kevin Lu, Telcordia
  - Chris Scofield, Inrix
  - Shawn Turner, Texas Transportation Institute
- Ex-Officio Member:
  - David McNamara, AutoTech Insider

Data Quality Measurements: Basic Premises

- Customers
  - Auto OEMs, PND manufacturers, data distributors, DOTs…
- Benchmarking purposes
  - Quality assurance, data validation
  - Comparison between providers, markets, traffic conditions…
- What gets assessed?
  - Incident / traffic event messages
  - Instantaneous flow data, i.e. speed-colored maps
  - Travel times
- What gets measured?
  - Timeliness [how fast conditions are transmitted]
  - Accuracy [degree of fit with a trusted source ('ground truth')]
  - User satisfaction [ultimate perception by the end user]

Different Flavors of 'Data Quality'

- FHWA metrics:

<u>accuracy</u>, validity, <u>coverage</u>, <u>timeliness</u>, completeness, <u>accessibility</u>

  - Accuracy
    - Most straightforward
  - Coverage
    - More difficult to articulate – only relevant with regards to a given level of accuracy
  - Timeliness
    - Seems more of an internal / SLA issue
  - Accessibility
    - Notion of usefulness / perception by end-user
    - Essential business feature, but ancillary to benchmarking

- NATWG Guidelines will initially focus on <u>information accuracy</u>

Measuring Information Accuracy

- Collect 'ground truth' traffic data
  - Define a set of technology and procedures
  - Key issue is level of confidence / statistical significance

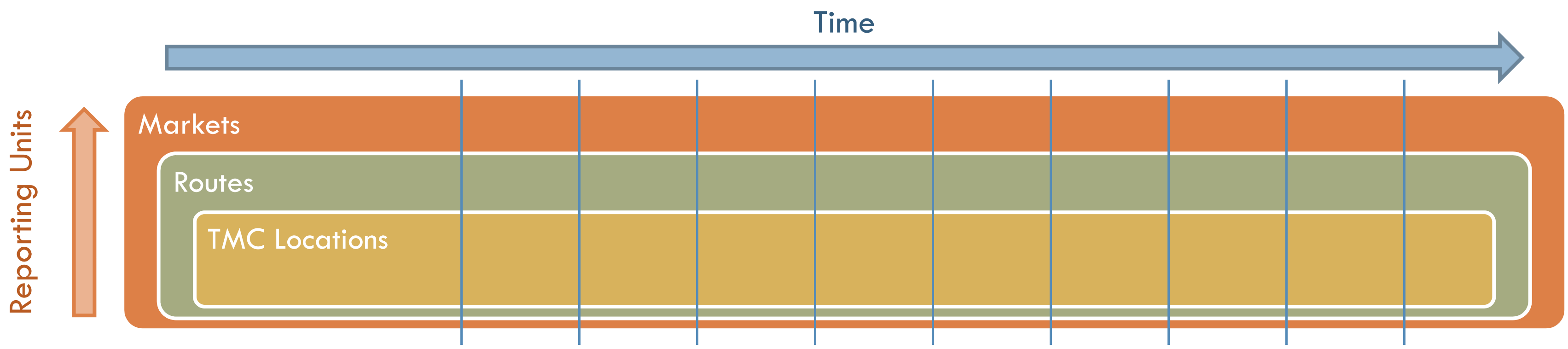
- Compare a traffic information source against ground truth
  - Requires metrics that are ideally:
    - Formally defined and easy to compute (no exceptions / fringe cases)
    - Relevant to the end-user experience
    - Easy to interpret
    - Good balance of synthetic vs. exhaustive (i.e. tells the story concisely)
    - Normalized and scalable (i.e. independent from route length, sample size, etc.)

NATWG Guidelines: General Considerations

- Focus on speed information
  - MPH value on a given segment at a given time
  - Median travel time along a route
  - Qualitative description such as ‘free flow’ or ‘heavy congestion’
- Guidelines, not standard (yet)
  - Leaves room to interpretation, balances principles with formal rules
  - Most important is transparency in assumptions, methods and results
- Insistence on meaningful tests
  - Reporting units (routes / time of day) must be homogeneous
  - Information quality matters most when roads are congested

NATWG Guidelines: Additional Considerations

- Guidelines developed for freeway environment primarily
  - Extension to signalized arterials possible
- Ground truth collection
  - Either floating cars or reidentification technology
  - To date, guidelines developed for a single floating car run
- Reporting units
  - By default, the most granular reporting unit is TMC location code
  - However the guidelines will work with any segment definition



NATWG Guidelines: Overview

- Preamble

1. General Considerations
2. Route Selection
3. Test Equipment
4. Driving Behavior
5. Data Logs Processing
6. Traffic Content Processing
7. Speed Comparison
8. Travel Time Comparison
9. Congestion Level Comparison

Speed Comparison

- For each TMC, we can compute the speed differential between the Ground Truth speed (V^{GT}) and the Traffic Information Service speed (V^{TIS})
- Differentials are aggregated across route TMCs, producing a single score
- We recommend using the Root Mean Squared Error (RMSE) aggregation:

$$\boldsymbol{RMSE} = \sqrt{\frac{\mathbf{1}}{\boldsymbol{N}}\sum\nolimits_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{N}}\left(\boldsymbol{V}_{\boldsymbol{i}}^{\boldsymbol{GT}} - \boldsymbol{V}_{\boldsymbol{i}}^{\boldsymbol{TIS}}\right)^{\mathbf{2}}}$$

- Consider the following example:

TMC Code	Description	GT Speed	TIS Speed	Squared Difference
105N04414	I-80 / I-580 Merge to I-80 x Gilman-University	70.9	55.0	254.2
105N04413	I-80 x Gilman-University to I-80 x University	72.5	65.0	56.4
105N04412	I-80 x University to I-80 x Ashby	50.3	55.0	22.3
105N04411	I-80 x Ashby to I-80 x Powell	36.1	32.0	16.5
105P18840	I-80 x Powell to McArthur Maze	41.7	48.0	39.1
105N04409	McArthur Maze to I-80 / I-580 Merge	12.4	20.0	57.9
105N04408	I-80 / I-580 Merge to I-80 / I-880 Merge	19.6	21.0	1.8
105N04407	I-80 / I-880 Merge to Bay Bridge Toll Booths	8.9	17.0	65.5
105N04406	Bay Bridge Toll Booths to Bay Bridge East End	46.2	45.0	1.5
105N04405	Bay Bridge East End to Yerba Buena Island	45.6	58.7	171.9
105N04404	Yerba Buean Island to Bay Bridge West Span	41.7	33.0	75.6
105N04403	Bay Bridge West Span to I-80 x Embarcadero	29.2	37.5	68.8
105N04402	I-80 x Embarcadero to I-80 x Harrison	37.9	34.0	15.3
105N04401	I-80 x Harrison to I-80 x 5th Street	51.9	45.0	47.2
105N04400	I-80 x 5th Street to I-80 x 7th Street	54.4	55.0	0.4
			RMSE	7.72

Travel Times Comparison (Components)

- For each road element considered, four basic elements are calculated:
  - *L* - The length in miles of the cumulative distance between each GPS point on the segment
  - T^{ACT} - The actual travel time of the test vehicle.
  - T^{REF} - The reference speed travel time estimate of the vehicle.
  - T^{TIS}- The estimate of travel time using data from the Traffic Information Service are calculated.
- For example, three sections of a 60 MPH limit freeway might yield the following:

ID	Length (Miles) *L*	Entry Time	Exit Time	Actual Travel Time (Seconds) T^{ACT}	Reference Speed Travel Time (Seconds) T^{REF}	TIS Estimated Travel Time (Seconds) T^{TIS}	
1	1	07:00:00	07:01:12	72	60	**84**	← *TIS Too Slow*
2	2	07:01:13	07:05:13	240	120	**180**	← *TIS Too Fast*
3	0.5	07:05:14	07:05:59	45	30	**65**	← *TIS Too Slow*

- From these elements, deltas between T^{ACT} and each of T^{REF} (D^{REF})and T^{TIS} (D^{TIS}) are calculated as relative and absolute values and harmonized by length (E^{REF}, E^{TIS}), where the delta is the est. travel time minus the actual travel time in each case e.g.:
  - $D^{REF} = T^{REF} - T^{ACT}$ and $D^{TIS} = T^{TIS} - T^{ACT}$
  - $E^{REF} = D^{REF}/L$ and $E^{TIS} = D^{TIS}/L$

This creates a relative and absolute metric for each segment equivalent to seconds per mile (SPM) of error for each of T^{REF} and T^{TIS} compared to T^{ACT}.

Travel Times Comparison (Metrics)

- In this fashion, the performance of traffic information can be compared in a normalized fashion to create a metric of the value of the traffic information in the context of the amount of travel time lost due to the vehicles reduced speed where 'Improvement' (I) is defined as:
  - $I = ABS\ E^{REF} - ABS\ E^{TIS}$
- We can further calculate a second value of improvement (I^{PC})as a percentage of the total absolute error of the reference speed estimate removed using the total TIS estimate for the entire route – e.g.:
  - $I^{PC} = I/ABS\ E^{REF}$

ID	L	T^{ACT}	T^{REF}	T^{TIS}	D^{REF}	D^{TIS}	ABS D^{REF}	ABS D^{TIS}	E^{REF}	E^{TIS}	ABS E^{REF}	ABS E^{TIS}	I	I^{PC}	
1	1	72	60	84	-12	12	12	12	-12	12	12	12	0	0.00	*Net tie*
2	2	240	120	180	-120	-60	120	60	-60	-30	60	30	30	0.50	*Net Improvement*
3	0.5	45	30	65	-15	20	15	20	-15	20	30	40	-10	-0.33	*Net Degradation*
R1	3.5	357	210	229	-147	-28	147	28	-42	8	42	8	34	0.81	*Route Net Improvement*

Sum of Individual (L and T) components are then Processed using the same logic to create route based units.
Sum of route based units can be further aggregated within road class to give city wide score.

Travel Times Comparison (Aggregation)

- Another useful metric that can be derived from this numbers is the performance of both the reference speed estimates and the TIS estimates of travel time to the actual travel time observed, where:
  - $P^{REF} = 1 - (ABS\ D^{REF} / T^{ACT})$
- These metrics provide the context for the amount of congestion observed in the test and the impact of the improvement in performance in the context of the total actual drive time. These metrics as all travel time metrics tend to provide more clarity when aggregated at the route level.
- In our sample, the P^{REF} values vary from 50% to 84% accuracy of travel time prediction and average only 59% accuracy for the whole route. P^{TIS} also varies from 56% to 84%, but in the context of the route, the travel time estimate using the traffic data is 92% accurate.

ID	T^{ACT}	ABS D^{REF}	ABS D^{TIS}	P^{REF}	P^{TIS}
1	72	12	12	0.84	0.84
2	240	120	60	0.5	0.75
3	45	15	20	0.66	0.56
R1	357	147	28	0.59	0.92

Congestion Levels and Speed Tolerance

- Speeds can be put into 'levels' corresponding to degree of congestion. For example:

Speed Levels (% of Reference Speed)	Speed Level Boundaries (Ref Speed = 50)	Speed Level Boundaries (Ref Speed = 65)	Congestion Level	Level Index
92+%	46+	60+	Green	4
62-92%	31-46	40-60	Yellow	3
31-62%	16-31	20-40	Red	2
0-31%	0-16	0-20	Black	1

- The floating car speed and the speed reported by a traffic information source may stand across a speed boundary while being very close

- Penalizing the traffic information provider for a wrong level estimate in such a situation is neither fair nor desirable. This effect is minimized with a speed tolerance threshold θ :

$$if\ |V^{GT} - V^{TIS}| < \theta\ then\ L^{GT} = L^{TIS}\ otherwise\ L^{GT} \neq L^{TIS}$$

Congestion Levels: An Example

- An error count can be computed reflecting the frequency that the GT and TIS speeds correspond to different levels

- Using our example:

TMC Code	Description	GT Speed	TIS Speed	GT Level	TIS Level	GT Level with Threshold	Difference
105N04414	I-80 / I-580 Merge to I-80 x Gilman-University	70.9	55.0	4	3	4	1
105N04413	I-80 x Gilman-University to I-80 x University	72.5	65.0	4	4	4	0
105N04412	I-80 x University to I-80 x Ashby	50.3	55.0	3	3	3	0
105N04411	I-80 x Ashby to I-80 x Powell	36.1	32.0	2	2	2	0
105P18840	I-80 x Powell to McArthur Maze	41.7	48.0	3	3	3	0
105N04409	McArthur Maze to I-80 / I-580 Merge	12.4	20.0	1	1	1	0
105N04408	I-80 / I-580 Merge to I-80 / I-880 Merge	19.6	21.0	1	2	2	0
105N04407	I-80 / I-880 Merge to Bay Bridge Toll Booths	8.9	17.0	1	2	1	1
105N04406	Bay Bridge Toll Booths to Bay Bridge East End	46.2	45.0	4	3	3	0
105N04405	Bay Bridge East End to Yerba Buena Island	45.6	58.7	3	4	3	1
105N04404	Yerba Buean Island to Bay Bridge West Span	41.7	33.0	3	3	3	0
105N04403	Bay Bridge West Span to I-80 x Embarcadero	29.2	37.5	2	3	2	1
105N04402	I-80 x Embarcadero to I-80 x Harrison	37.9	34.0	2	2	2	0
105N04401	I-80 x Harrison to I-80 x 5th Street	51.9	45.0	3	3	3	0
105N04400	I-80 x 5th Street to I-80 x 7th Street	54.4	55.0	3	3	3	0
						Error Count	4

Further Considerations

- Metrics are not fully formalized yet
  - Need to rub against real world data
  - Need buy-in from more stakeholders
- Ground-truth data collection needs revisiting
  - Make determination on adequate sampling
  - Examine alternatives to floating cars
- The testing methodology needs to scale up
  - From a given route to an entire metropolitan market

Implications

- Expand task force
  - Recruit new members who can weigh in on final decisions
  - Generate additional legitimacy
- Need partners to try out the guidelines
  - Use metrics with existing / ongoing validation data
- Need additional technical investigations
  - Fine-tuning of metrics and their parameters
  - Study sample size issues
  - Good news: pooled fund study can provide match

Pooled Fund Study TPF-5(200)

- Objective:
  - Standard test procedure to evaluate the quality of travel time data services
  - Consistent evaluation results
  - Fair comparisons between data services
- Public agency clients, public and private stakeholders
- Sponsors:
  - VA lead state
  - Also AL, CA, FHWA, MD, MI, PA
- Contractors:
  - VTRC, UVA, TTI
- August 2009 to April 2011
- See http://www.pooledfund.org