Travel Time Data Collection for Measurement of Advanced Traveler Information Systems Accuracy

June 2003

Alan Toppen
Dr. Karl Wunderlich

Contract Sponsor:	Federal Highway Administration
Contract No.:	DTFH61-00-C-00001
Project No.:	0900610-D1
Department:	J190

MTS™ Mitretek Systems
Falls Church, Virginia

Table of Contents

Acknowledgements

The authors would like to acknowledge the following people who reviewed this white paper and provided valuable comments: Don Roberts, Mitretek; Ralph Gillmann, FHWA; Shawn Turner, TTI; Doug Sallman, Cambridge Systematics; and Paul Shuldiner, UMass-Amherst.

1. Introduction

1.1 Background

Evaluating the benefits of Intelligent Transportation Systems (ITS) has been a role of the ITS Joint Program Office (JPO) of the U. S. Department of Transportation for many years. In support of this effort, Mitretek Systems developed HOWLATE (Heuristic Online Web-Linked Arrival Time Estimator), a quantitative modeling methodology based on archived Advanced Traveler Information Systems (ATIS) travel time data, to evaluate the user benefits of regional ATIS deployments (1,2). The HOWLATE methodology quantifies time management benefits of ATIS use on a per trip basis for travelers who need to be on time.

In a recent application of this method, we explored how the accuracy of travel time estimates from ATIS web sites affects user benefit (3). Clearly, users of these services stand to benefit more when the information presented is more accurate. Conversely, if the information is very inaccurate, it may lead a user to do worse than if he had disregarded it and instead followed his habitual route and departure time. In three cities (Los Angeles, CA; Washington, D.C.; and Minneapolis/St. Paul, MN), we developed relationships between ATIS user benefit and ATIS accuracy, showing how the per trip benefit to users of ATIS over non-users declines with declining accuracy in ATIS information provision.

Applying the HOWLATE method, we varied the amount of error in the travel time estimates provided to simulated ATIS users relative to the travel times they actually experienced. We identified the maximum potential benefit (i.e., under a perfectly accurate system) as well as the minimum accuracy required for the average aggregate trip to benefit. We also identified the marginal benefit – the amount of benefit improvement that would result from an improvement in accuracy of 1%. Benefits to the user include reduced in-vehicle travel time, improved on-time reliability, and reduced stress from concerns about being late.

Figure 1 shows the benefit vs. accuracy relationship derived for Los Angeles. Benefit is expressed in monetary terms on a per trip basis, a collective value comprising time savings and stress reduction (4). When accuracy drops below a critical point, one is better off not using ATIS and relying on experience with historical traffic patterns. In Los Angeles, that point is in the range of 13-21% error (error is the standard deviation of the percent difference between the true travel time and that estimated by the ATIS—more detail will be presented on this later).

At the highest levels of accuracy, little is gained by making further improvements. Beyond a certain point (below 5% error), it makes little sense to invest in improved accuracy. In this case, funds for ATIS improvements would be better spent in areas besides improving accuracy, such as expanding surveillance coverage to other roadways.

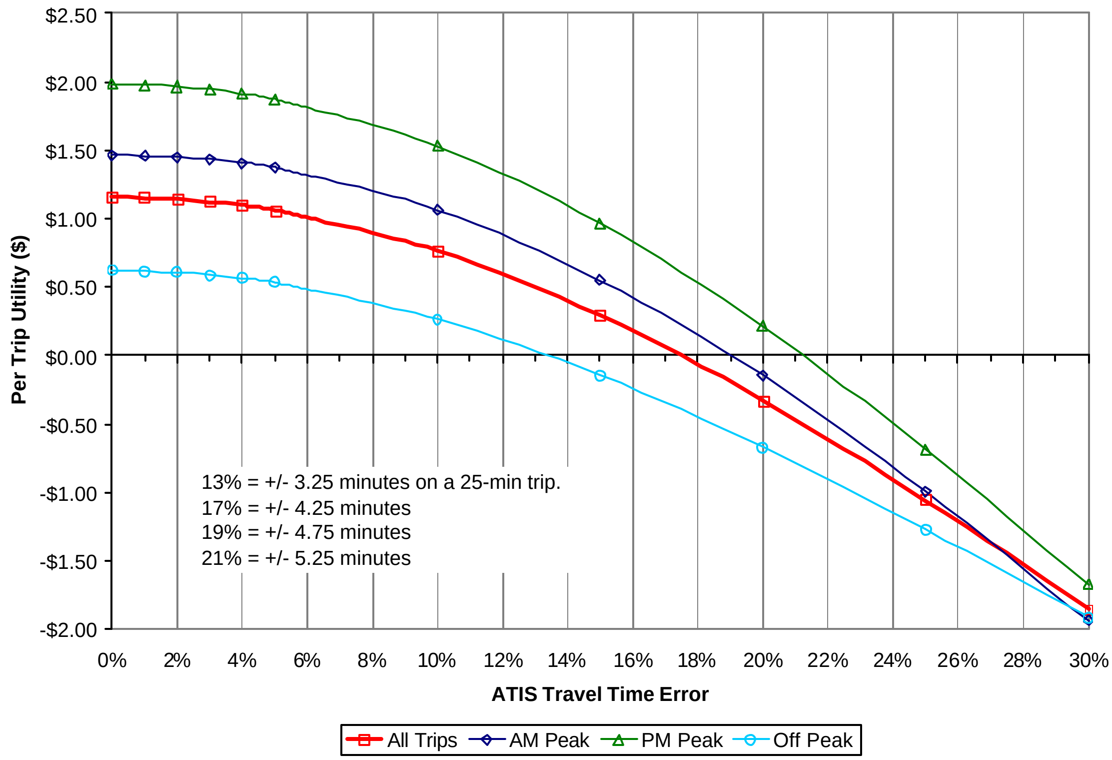

Figure 1. Benefit-Accuracy Relationship for Los Angeles

In a parallel effort, we began studying the relationship between the location and extent of surveillance coverage on user benefit. While that work is still in progress, its objective is to identify a methodology using observable network attributes such as AADT that one might use to prioritize roadway miles for deploying ATIS surveillance based on relative cost-effectiveness. In addition, it is concerned with identifying whether there is some point beyond which it does not make sense to continue instrumenting the region with surveillance because the costs of covering less traveled roadways are not justified by the accompanying benefits.

The joint purpose of these two studies was to guide ATIS deployment decisions. Given limited funds, an ITS planner is faced with the decision of whether to expand surveillance coverage or improve the accuracy of roadways already covered. The most cost effective deployment strategy weighs improving accuracy against expanding coverage. When accuracy is high, the most cost effective deployment decision may be to expand coverage. If surveillance coverage is already extensive, improving accuracy on lane-miles already covered may be the most cost-effective option.

1.2 The Relationship Between Accuracy and Variability

How accurate do travel time estimates need to be? This question depends on a few factors, the most important being regional day-to-day travel time variability. In what may be a counter-intuitive result, users of ATIS in metropolitan areas with less day-to-day travel time variability require more accurate ATIS information. The more predictable traffic conditions are from day to day, the better knowledge of historical conditions is than imprecise real-time information in predicting the travel time for a pending trip. On the other hand, when day-to-day variability is high, even imprecise ATIS travel time estimates can be an improvement over historical averages. Therefore, day-to-day variability is a key indicator in determining how accurate ATIS travel time estimates need to be to provide user benefit.

1.3 Objective

The objective of this white paper is to recommend an approach to measuring ATIS travel time accuracy so that ITS planners might have the data they need to make cost effective decisions regarding deployment of surveillance technologies to support ATIS. There are at least eleven metropolitan areas with online ATIS services that provide travel time estimates on major freeways (5) and more are expected to come on line in the future. It is not common practice, however, for operators of these systems to measure the accuracy of the travel time estimates they provide to the public on their web sites.

Based on the aforementioned studies, in order to make cost-effective deployment decisions three things must be known:

1. The extent of surveillance relative to full coverage,
2. Regional day-to-day variability, and
3. The prevailing accuracy of ATIS travel time estimates.

Determining the extent of coverage is straightforward once full coverage is defined. This could be all the major freeways in the metropolitan area. Though it does not need to be so, current reliance on point detection has made arterial travel time estimation infeasible in most cases; very few real-time travel time estimates are available on arterials. Another option for the definition of full coverage is what is given on the ITS Deployment Tracking website (6).

Determining accuracy and variability require data to be collected in the field. For ATIS *accuracy* measurement, "ground truth" travel times need to be collected. Error is then calculated by comparing the ground truth travel time with that given by the ATIS for the same segment on the same day at the same time. For *variability* calculations, additional ground truth travel times need only be collected if the ATIS travel time estimates are shown to be unreliable based on the accuracy measurement. Certain ATIS systems revert to a default value when no data is available, others cap travel times at the speed limit, and others aggressively smooth estimates from one time interval to the next. Each of these reduces the amount of variability in the ATIS data and would bias an estimate of regional variability toward there being less variability than there is in reality. If the ATIS travel time estimates accurately match the ground truth data, however, calculating variability on the basis of the daily ATIS travel time estimates is preferable because it reduces the amount of field data required.

In Section 2, we will define, for purposes of consistency, "ground truth" travel time, error and variability. In Section 3, we will describe various technologies that may be used to collect the necessary data in the field and the amount of data that needs to be collected. In

Section 4, we will present cost estimates for data collection with various technologies, and in Section 5 we will present our final recommendations.

2. Definitions

Terms that need to be clearly defined are ground truth travel time, travel time error and variability. Each of these can be measured in a variety of different ways. This section will give alternative definitions for each and give the rationale for the one chosen.

2.1 Ground Truth Travel Time

ATIS users, who are travelers needing to make decisions such as whether to leave earlier than previously planned, take an alternate route, change mode, or cancel their trip altogether, interpret travel time estimates from an online ATIS as predictions of how long their various trip options might take. In their minds, a perfectly accurate ATIS travel time estimate would tell them exactly how long any trip option will take if they were to leave at the current time or at some time in the near future. Of course, it is never possible for such information to be perfectly accurate since it requires insight into the future (i.e., how congestion will build or dissipate over the course of the trip, whether an accident will occur, etc.).

In order to measure accuracy, we need a "ground truth" travel time against which we can compare the ATIS estimate. There are at least four definitions one could use for ground truth travel time over some segment $S(a,b)$ with length $l(a,b)$ at some time t_1 or over some time interval $T(t_1,t_2)$. These are described as follows, with reference to Figure 2.

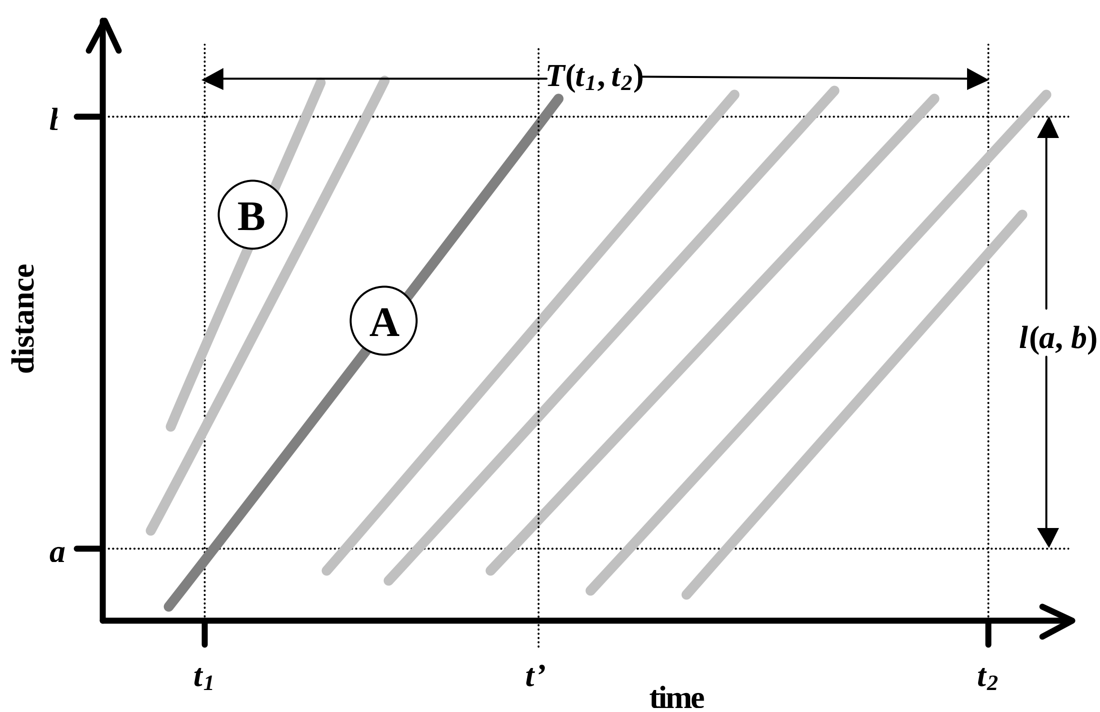

Figure 2. Individual Vehicle Trajectories Over Time and Space

1. The time to traverse segment $S(a,b)$ if entering the segment (i.e., passing point a) at time t_1. This is shown as trajectory A in Figure 2. This is probably the best measure, though impossible for an ATIS to report without error because it requires a prediction of when traffic passing point a at time t_1 will then pass point b at some future time t'. For very long segments and transient conditions this may be difficult to predict. This measure can be obtained through license plate matching or probe vehicle studies (measurement techniques will be discussed in more detail in a later section).
2. The space mean speed V over $S(a,b)$ and $T(t_1,t_2)$:

   $$V = \left[\frac{1}{N} \sum \frac{T_i}{X_i} \right]^{-1}$$

   where
   T_i = the duration of time vehicle i is in $S(a,b)$ over $T(t_1,t_2)$,
   X_i = the distance traveled by vehicle i in $S(a,b)$ over $T(t_1,t_2)$,
   N = the number of vehicles appearing on segment $S(a,b)$ within $T(t_1,t_2)$.

   This is the harmonic mean of the speeds of individual vehicles comprising the traffic stream (7). In practicality, this is difficult to measure since to be precise, it must include vehicles that do not traverse the entirety of $S(a,b)$ in $T(t_1,t_2)$, such as vehicle trajectory B in Figure 2. This vehicle would have been detected at point a before t_1. Therefore, its position at t_1 would not be known if detectors were only located at points a and b. For purposes of field data collection, this is a theoretical entity. The first definition relates better to the traveling public.
3. The average instantaneous speed at time t_1 for all vehicles in $S(a,b)$, converted to travel time with $l(a,b)$. This can be computed from the distance traveled by each vehicle in $S(a,b)$ during a short time interval δt, which can be obtained from successive aerial photographs or video frames. Radar is another method, though not very practical for capturing multiple vehicle speeds over a segment at a single instant in time.
4. The average speed at discrete points in $S(a,b)$ over $T(t_1,t_2)$, converted to travel time with $l(a,b)$. ATIS travel time estimates are typically measured in this way since most are constrained to estimate segment travel times with point detection.

The way users of ATIS interpret the information needs to be the basis for our selection of baseline or "ground truth" travel time since the deviation from this is error from the user's perspective. Definition 1 relates most closely with the experience and perspective of users of ATIS. However, measurement of this quantity is still not straightforward. Different vehicles travel at different speeds. Further, defining a travel time over $S(a,b)$ at in instant in time t_1, is problematic. For a precise instant in time (1 second or less), there is a good chance that no vehicle will enter segment $S(a,b)$. The problem is to define a travel time for a hypothetical vehicle entering $S(a,b)$ at time t_1, given the traffic stream is comprised of vehicles traveling at different speeds and entering $S(a,b)$ around time t_1.

Ideally, "ground truth" is based on a smoothed average of all vehicles traversing $S(a,b)$. Since different vehicles travel at different speeds, the traffic stream is a scatter of travel times. A value for the instant in time, t_1, could be obtained by smoothing the data points using commonly used statistical techniques (e.g., exponential smoothing, moving average) or by simply averaging all data points over the interval $T'(t\text{-}d,\ t\text{+}d)$. The selection of d would depend on the flow rate, the capture rate, and engineering judgment. The lower the flow rate, the larger d would need to be for a sufficiently large sample size. We estimate a d of 30 seconds is adequate when the flow rate and capture rate are high. A d of 5 minutes may be needed when flow rate and capture rate are low.

2.2 Travel Time Accuracy/Error

There are at least four different ways to represent error. In each of these options, error is the difference between the observation and ground truth travel time and percent error is the error divided by the ground truth value.

Four possible representations of error are:

- Mean Absolute Deviation (MAD) (also known as the mean absolute error) – the average of errors. The quantities $\hat{y}_t$, y_t, and n are the ATIS estimate of travel time, the ground truth travel time, and the number of observations, respectively.

  $$MAD = \frac{1}{n}\sum\left|\hat{y}_t - y_t\right|$$

- Mean Absolute Percent Error (MAPE) – the average absolute percentage difference between the estimate and ground truth.
  - $$MAPE = \frac{1}{n}\sum\left|\frac{\hat{y}_t - y_t}{y_t}\right|$$
- Root Mean Squared Error (RMSE) – the square root of the average of the squared errors.
  - $$RMSE = \sqrt{\frac{1}{n}\sum(\hat{y}_t - y_t)^2}$$
- The Standard Deviation of Percentage Error (SDPE) – the square root of the average of the squared percentage errors.
  - $$w_t = \frac{\hat{y}_t - y_t}{y_t}$$
  - $$s_W = \sqrt{\frac{1}{n-1}\sum w_t^2 - n \cdot \overline{w}^2}$$

The SDPE is an attractive choice because the benefit vs. error curves in (3) defined error as the standard deviation of the percent difference between ground truth and the travel time estimate. Monte Carlo simulation was used to generate hypothetical travel time estimates based on the following equation:

$$\hat{y}_t = y_t + b \cdot y_t + r \cdot e \cdot y_t$$

where:

$\hat{y}_t$ = the travel time estimate, t

y_t = the corresponding ground truth travel time, t
b = the estimation bias, the amount which the ATIS over or under reports, on average.
e = error
r = a randomly generated number from $R \sim N(0,1)$

We can solve to get:

$$b + e \cdot r = \frac{\hat{y}_t - y_t}{y_t}$$

If we define $w_t = \dfrac{\hat{y}_t - y_t}{y_t}$ and W as the random variable from which the sample w_t for all t is drawn, then

$$b + e \cdot R = W$$
$$Var[b + e \cdot R] = Var[W]$$
$$e^2 \cdot Var[R] = Var[W]$$
$$e^2 = Var[W]$$

This is consistent with the SDPE. If the bias is positive the ATIS underestimates travel time, on average. This may be the case during congestion if loop detectors are the primary means of detection. Loop detectors are less accurate at low speeds and may overestimate speed. If the bias is negative, the ATIS overestimates travel time. This may be the case during free flow conditions if the ATIS has a policy of not reporting travel times implying faster than speed limit travel.

2.3 Variability

Besides the “ground truth” travel time measure, which is needed to measure error in the ATIS travel time estimates, day-to-day variability measures must also be standardized. Variability is the standard deviation of segment speed across many days for the same segment at the same time of day. Segment speed is simply the segment distance divided by the travel time. If variability is measured for multiple segments, network variability is the average of the segment standard deviations, weighted by segment length. This is:

$$s_{s,t} = \sqrt{\frac{1}{D-1} \sum_d (u_{s,t}^d - u_{s,t})^2}$$

$$V = \frac{1}{S} \frac{1}{T} \sum_s \sum_t s_{s,t}$$

where:

$u_{s,t}^d$ = the speed on segment s, at time t, on day d—the segment length divided by the segment travel time.
$s_{s,t}^d$ = the standard deviation of travel time on segment s, at time t, on day d.
S = the total number of segments.
T = the total number of time periods.
D = the total number of days.
V = day-to-day variability.

$u_{s,t}$ = the average speed on segment s at time t.

3. Technologies and Techniques for Ground Truth Travel Time Measurement

The Travel Time Data Collection Handbook describes a number of different techniques for collecting travel time data in the field. They can be classified as: probe vehicle methods, license plate matching, and emerging ITS technologies such as cell phone tracking, Advance Vehicle Identification (AVI) and inductive loop signature matching (7). Each of these will be discussed in turn with the primary focus being on probe vehicle methods and license plate matching.

3.1 Probe Vehicle Techniques

Probe vehicle techniques involve the use of a data collection vehicle within which an observer records his travel time at predefined checkpoints (7), or in the case of data collected using Global Positioning Systems (GPS), the precise location of the probe vehicle is captured at specific time intervals. There are several different methods, depending on the technology and driving style used. The three most common driving styles are:

Average Car – the probe vehicle tries to capture the average of the traffic stream by passing as many vehicles as passes it,
Chasing Car – the probe vehicle selects one vehicle to be representative of the traffic stream and follows it, and
Maximum Car – the test vehicle attempts to drive at the posted speed limit unless impeded by traffic.

Any of these methods are suitable as long as a single one is used consistently. In terms of technology or instrumentation, three different approaches are

Manual – travel times are manually recorded,
Distance measuring instrument (*DMI*) – this device is linked to the transmission of the vehicle to automatically record speed and distance, and
Global positioning system (*GPS*) – a GPS receiver records speed, vehicle latitude and longitude, and time, at short intervals.

As with any data collection technique, there are pros and cons to probe vehicles techniques.

Pros:

- *Data can be collected over a wide area.* Unlike other techniques, probe vehicles do not require instrumentation to be set up on the roadway. Therefore, probes can easily collect data on any part of the network. For ATIS data accuracy studies, it is advantageous to get the widest sample of the network as possible—as many segments as possible at different times of day.
- *Initial costs are low.* Relatively inexpensive equipment is required and little specialized knowledge is required to collect the data. The actual data collection

can be easily outsourced to any number of local data collection firms that conduct these studies.

- *Data can simultaneously be collected for both directions of travel*. Since the probe vehicle must return to the starting point, it is logical that data can typically be collected in the inbound direction, as well the outbound direction, using the same vehicle.
- *Little data reduction effort is required.* Especially with DMI and GPS instrumentation, once data collection is completed, very little effort is required to convert that data to segment travel times.
- *Data may be easily collected for subsections of the study corridor*. Travel time and/or speed data may be obtained for subsections of a longer corridor, especially if GPS data collection is utilized. This adds the additional capability to match the collected data with speed data from point speed collectors placed at the corridor subsections to provide validation of the automated detectors.

Cons:

- *A single vehicle represents the traffic stream.* This, in a sense, violates the premise that ground truth takes into consideration the entire traffic stream. However, this is compensated for by using the average car method, which approximates the average flow of traffic. Furthermore, as this paper will make clear, there is no method that assures the entire traffic stream will be captured.
- *Measurement of day-to-day variability is difficult.* Measuring day-to-day variability requires travel time measurements to be taken across multiple days *at the same time.* It is not easy to control precisely when the probe vehicles enters the segment for which travel time is measured.
- *Continuous data collection may require significant resources*. The data collection effort requires a significant amount of manual labor and it is difficult to further automate the process. The collection of a large number of data points or collecting data over a long period may require the commitment of a significant amount of labor resources.

3.2 License Plate Matching Techniques

License plate matching techniques involve matching the license plates of vehicles at two points and measuring travel time by the time difference. As with probe vehicle techniques, there are various approaches depending on the technology and level of instrumentation employed. Three different approaches are:

Manual – Observers in the field manually record license plates as vehicles pass by, either on paper, into a tape recorder, or into software a laptop computer. Travel times are measured by the difference in timestamp for matched license plates at successive checkpoints. Benefits are that little expensive equipment is required and data reduction is minimal. However, this approach is only feasible for low speed, low volume situations. Even in such cases operator error can be high, particularly due to fatigue.

Video with manual transcription – Video cameras are positioned in the field to capture images of license plates of passing vehicles. Later, individuals manually enter license plates into a computer as the video is viewed. This is the most robust license plate matching technique. Nearly all vehicles can be detected by video observers with full control over stopping, starting, and pausing the tape. This allows for the best possible measurement of ground truth travel time. However, data reduction is time consuming. Between four and ten hours of data reduction are required for every one hour of video, depending on the experience level of the viewer and the quality of the video recording (7).

Video with character recognition – High quality video is used to capture license plate images of passing vehicles. Later, the video is run through an automated license plate reader (LPR) that uses optical character recognition technology to read license plates from the video. This method combines the high vehicle capture rate of video while reducing the data reduction effort. It does not eliminate the need for data reduction, however. While fixed installations use a trigger—typically a loop detector—to tell the camera when a vehicle is present, for short term studies where there is no trigger, the license plate reader has to use less robust video imaging techniques to determine when a vehicle has entered its line of sight. As a result, an operator is required to confirm or correct each license plate image. Software brings up each license plate image with the ASCII interpretation from the reader so this is considerably faster than viewing the entire video footage, but it takes time nonetheless (7,8).

Pros:

- *A high percentage of vehicles can be captured.* Particularly with video-based techniques, a very high percentage of the traffic stream can be measured. Since our objective is ground truth, this aspect is better than using probe vehicles where the traffic stream is summarized by only one vehicle.
- *Video with manual transcription provides the most robust measurement of ground truth available.* This method directly measures travel time of nearly every vehicle in the traffic stream, though occlusion may cause some vehicles to be missed if more than one lane is being viewed with a single camera. In this way, this is the ideal method for ground truth measurement of a segment.
- *Day to day variability can be accurately measured*. Because instrumentation is installed on a single segment, day-to-day variability can be directly measured as long as data collection equipment is operational at the same time each day.

Cons:

- *Methods that do not use video still only detect a sample of the vehicles in the traffic stream.* Field observers can not possibly record license plates for every passing vehicle. Furthermore, a vehicle as to be captured by both upstream and downstream observers to be measured.
- *For video with manual transcription, data reduction is time-consuming and labor-intensive.* Up to ten hours of data reduction may be needed for every hour of video.
- *Video with character recognition is costly.* A separate camera is required for each lane, LPRs depend on good weather and operator experience to ensure good quality video, and the equipment is expensive. Because training and experience are required to get capture good license plate images, subcontracting to private firms that specialize in these studies is common.

- *It is more difficult to cover a wide area.* Because it requires equipment to be set up at a location, license plate matching techniques are not well suited to covering a large number of segments. While this is well suited for measuring variability, for ATIS accuracy measurement it is better to sample as many different segments as possible for the best characterization of system accuracy.

3.3 Extrapolation From Inductive Loops and Other Point Sensors

The most common traffic measurement technique uses inductive loops and to a lesser extent, other types of point sensors. Single loops can directly measure volume and occupancy. Speed can be measured directly with dual loops or by calculation from single loop measurements with an estimate of average vehicle length. Other types of point sensors such as radar, microwave, video, or infrared measure speeds at a point (9). Inductive loops, because of their ubiquity, are often used to estimate point-to-point travel times though they are merely detectors of point speeds. Most jurisdictions apply the speed at a detector station (a location where loops cover all lanes) to a wider area—typically half the distance to the next detector. While there is no guarantee this represents the average speed over this segment, it is a reasonable estimate. Some jurisdictions have developed prediction algorithms that use historical information to make short-term forecasts based on current speeds at point along a segment (7).

As mentioned previously, this is the most common form of ATIS travel time estimate. It is not suitable for use as ground truth travel time because it does not directly measure travel time. In addition, loops have been shown to be unreliable for measurement of low speeds (7).

3.4 Other

For the sake of completeness, other travel time measurement techniques include signpost-based Advanced Vehicle Location (AVL), Automatic Vehicle Identification (AVI), cellular phone tracking, and vehicle signature matching with inductive loop detectors. Each of these is briefly discussed here and is presented in more detail in (7) and (9). These are technologies that have been used for travel time measurement applications throughout the world because they provide more direct measurement of travel time than interpolation from point speeds. However, for ATIS travel time accuracy measurement, it is sufficient to say these are emerging technologies that either have not been proven reliable enough, can not guarantee sufficient market penetration, or are too costly for short term studies to be considered suitable for near-term ground truth travel time measurement. In addition, many require infrastructure (tag readers for AVI), have institutional or privacy barriers that must be overcome (cell phones), or are not technologically proven (inductive loop signature matching), making them unsuitable for short term studies of the type discussed in this white paper.

AVL – AVL is most commonly found on buses and is used to manage headways and alert operators in case of emergencies. AVL is a promising travel time data collection technique because buses cover major portions of the urban street network and it does not require any additional infrastructure beyond what is already used by the bus system. However, buses are not representative of the traffic stream due to their many stops and starts (7). For our purposes we are mainly interested in freeway travel times, which makes AVL less attractive since buses travel mostly on arterial streets.

AVI – AVI has been shown to be very accurate for the travel time measurement of individual vehicles equipped with transponders or toll tags (11,12). However, for an ATIS based on AVI technology, accuracy depends on sufficient market penetration. When equipped vehicles traverse a segment in quick succession, the system can provide up-to-date travel time estimates that are representative of the traffic stream. When market penetration is low (i.e., there is a long time between equipped vehicles), travel time estimates are less up-to-date and are based on the measurement of fewer vehicles, increasing the chance that travel time estimates are skewed by outlier data points.

Cell Phones as Probes – Studies by researchers at the University of California at Berkeley (13) and the University of Virginia (14) have tested the suitability of cell phones for use as traffic probes. This approach to travel time measurement is attractive because it takes advantage of existing infrastructure and market penetration of cell phones is high and increasing. This approach depends on the adoption of "Enhanced 911" or E-911, which is a mandate by the FCC that carriers provide caller locations within 125 meters. While these studies have shown this to be a promising approach for the future, the technology currently does not support it. In addition, there are many institutional barriers, such as privacy concerns and cost-sharing, that limit this concept at the current time as a feasible travel time data collection technique.

Vehicle Matching – In addition to the more common technique of license plate matching, there are other travel time measurement techniques also based on reidentification of individual vehicles or platoons. These are described in detail in (7). A promising method for use in the near term is based on research from Ohio State University and the University of California at Berkeley (15). These researchers have developed methods to match vehicles at successive loop detectors based on their lengths and the order in which they appear in platoons. This is an attractive approach to travel time measurement because makes use of existing infrastructure. However, this technique is still in the research and development stage.

4. Data Collection Approach and Minimum Sample Size

As our purpose is to obtain a measure of ATIS system-wide accuracy, it is important to obtain a representative sample of ATIS accuracy measurements. Accuracy may vary by time of day (e.g., accuracy may be lower at congested times) and by location due to uneven detector reliability. Therefore, a representative sample involves measurements over different times of day, different segments, and different days.

We have discussed two different measurement objectives. The first is to sample the network over different segments and times of day to get a system accuracy measurement. The second is to select one or a few segments to measure travel time, each day, at the same time. This will provide a measure of day-to-day variability that will indicate how accurate the ATIS system needs to be in order for the average trip to benefit. Both measurements are standard deviations of a quantity. The first is the standard deviation of percent error according to the SDPE defined above. The second is the standard deviation of link speed. Therefore, the methodology for determining the minimum required sample size for each is the same.

4.1 Minimum Sample Size for Error Measurement

We define s_E as the true error – the error of all ATIS travel time estimates at all times of day for all days over all segments in the network. That is, s_E is the SDPE of E_t if we could continuously measure ground truth travel time on all segments in the network. In reality however, we can only make a small number of measurements to generate a sample SDPE, s_E. The larger the sample, the more likely it is the sample SDPE, s_E, will be close to the true error, s_E. Selection of a sample size is determined by using confidence intervals. If we define E as previously,

$$E_t = \frac{\hat{y}_t - y_t}{y_t}.$$

Then, assuming E is normally distributed[1], for a sample size of n,

$$\frac{(n-1)\cdot s_E^2}{s_E^2} \sim c^2(n-1).$$

Therefore, the confidence interval for s_E is

$$\sqrt{\frac{(n-1)}{c^2(n-1)_{a/2}}}\cdot s_E < s_E < \sqrt{\frac{(n-1)}{c^2(n-1)_{1-a/2}}}\cdot s_E.$$

We can determine the necessary sample size using this equation. Figure 3 shows how the 90% confidence interval narrows with increasing sample size. That is, for larger samples, we can be more precise about whether s_E is close to s_E.

Defining "close" is a matter for engineering judgment; it depends on how precisely we want s_E to estimate s_E. This is defined by the "bound." For instance, we may want to know the required sample size in order to be 90% confident the sample error, s_E, is within 0.02 of the true error, s_E. The bound in this case is 0.02, and it is half of the width of the confidence interval, *W*, shown in Figure 3. Figure 4 inverts the axes of Figure 3 to show sample size as a function of the bound for a 90% confidence interval where s_E is 0.20, which is a reasonable peak period error for a moderately accurate system (16).

Based on Figure 4, if we want to say with 90% confidence that the measured error is within 0.03 of the true error, we need to collect a sample of 65 error measurements; 140 observations are needed to be within 0.02. That is, if we collect 140 observations and calculate a sample error s_E of 0.20, we can be 90% confident the true error s_E is in the

[1] It is well known that travel time distributions are skewed and are therefore more lognormally distributed than normally distributed. However, error as defined above can be assumed to be normally distributed.

interval 0.20±0.02. The sample size we will recommend also depends on the cost of data acquisition, a topic which will be discussed in the next section.

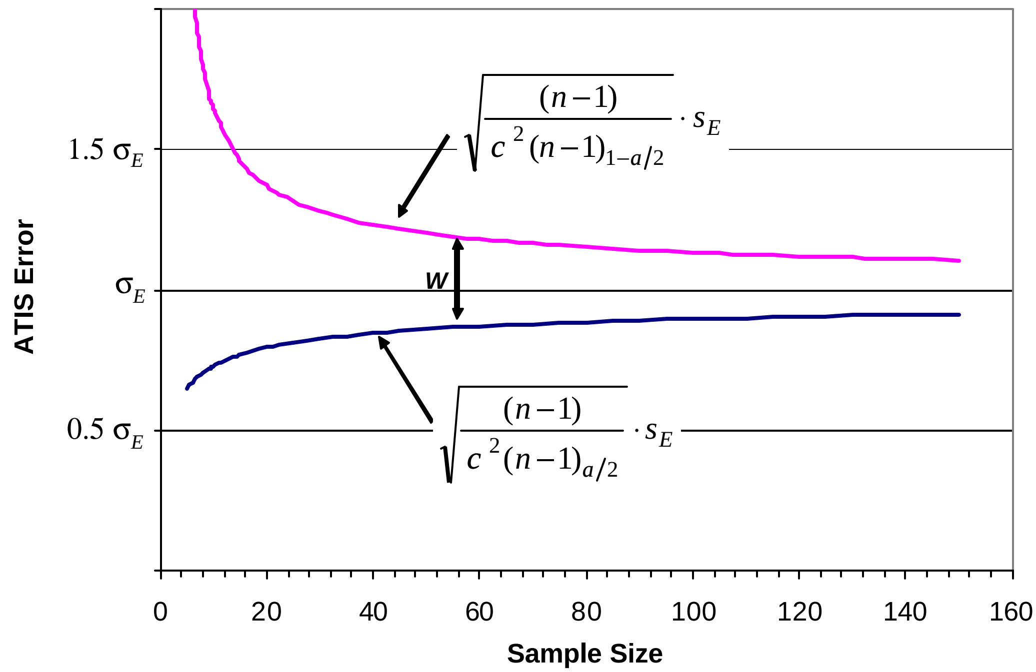

Figure 3. 90% Confidence Interval for s_E as a Function of Sample Size

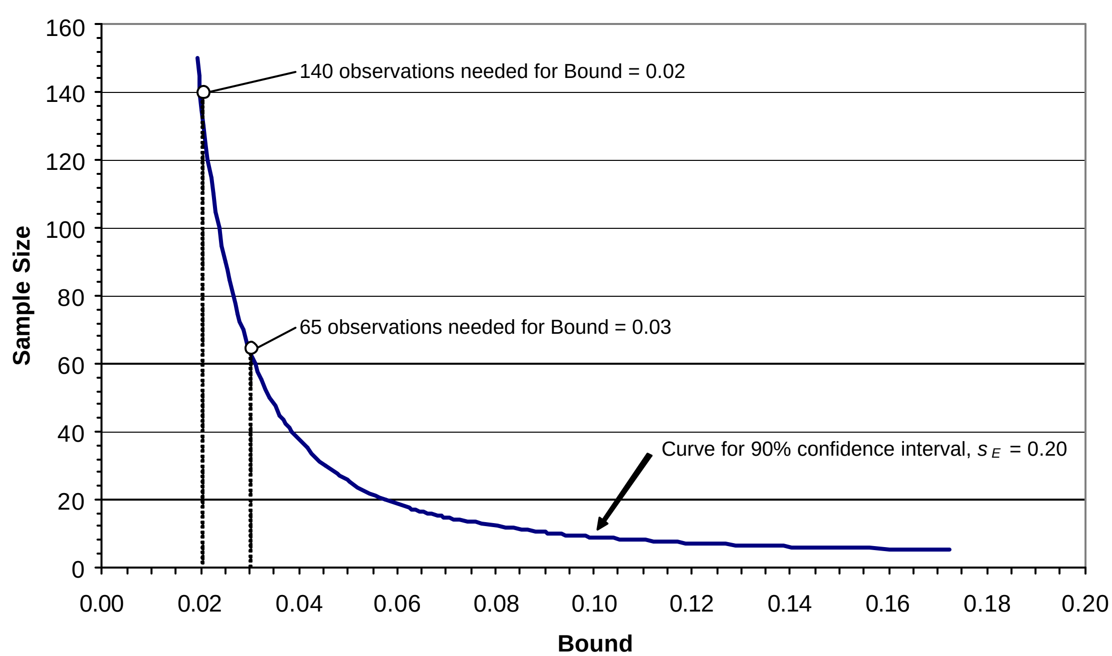

Figure 4. Necessary Sample Size For Accuracy Measurement as a Function of the Bound of a 90% Confidence Interval Assuming the Sample Error is 0.20.

4.2 Minimum Sample Size for Day-to-Day Variability Measurement

The minimum sample size for day-to-day variability is determined the same way as for error. Variability, which is the standard deviation of speed from day to day, can be expected to be approximately 8 miles per hour in the peak periods and 5 miles per hour in the off peak as shown by a sampling of data in Figure 5. (4). Assuming a sample standard deviation of 8 miles per hour and a 90% confidence level, Figure 6 shows the minimum required sample size as a function of the bound. If we want to be 90% sure the measured variability is within 1 mile per hour of the true value, we need to collect 90 days of data. Taking data for 90 days is very labor and cost intensive. More reasonably, with 20 days of data we can be 90% sure the true peak period standard deviation lies within 2.3 miles per hour of the true standard deviation. This is adequate for our application.

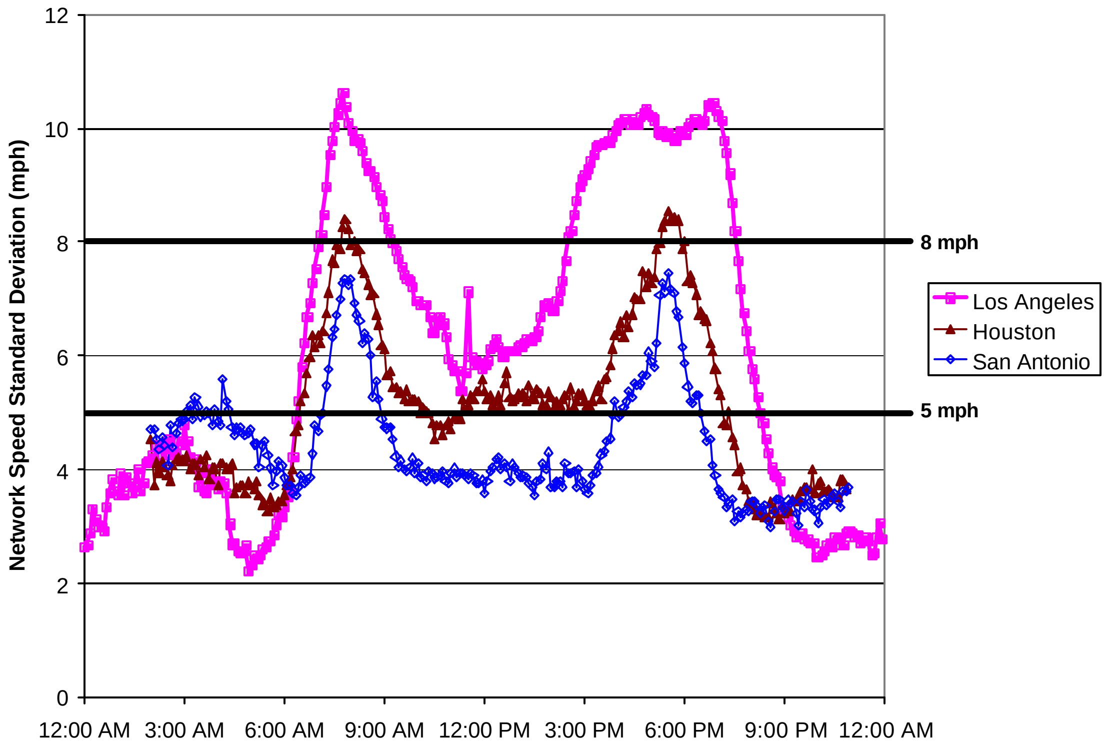

Figure 5. Average Network Speed and Variability in Various Cities From a Sampling of Data from Traveler Information Web Sites (5)

In the off-peak, we can expect variability to be approximately 5 miles per hour. With 20 days of data, we can be 90% sure the measured variability is within 1.44 miles per hour of the true value. Note that the bound is directly proportional to the sample standard deviation, i.e., when variability is twice as high, double the sample size is required.

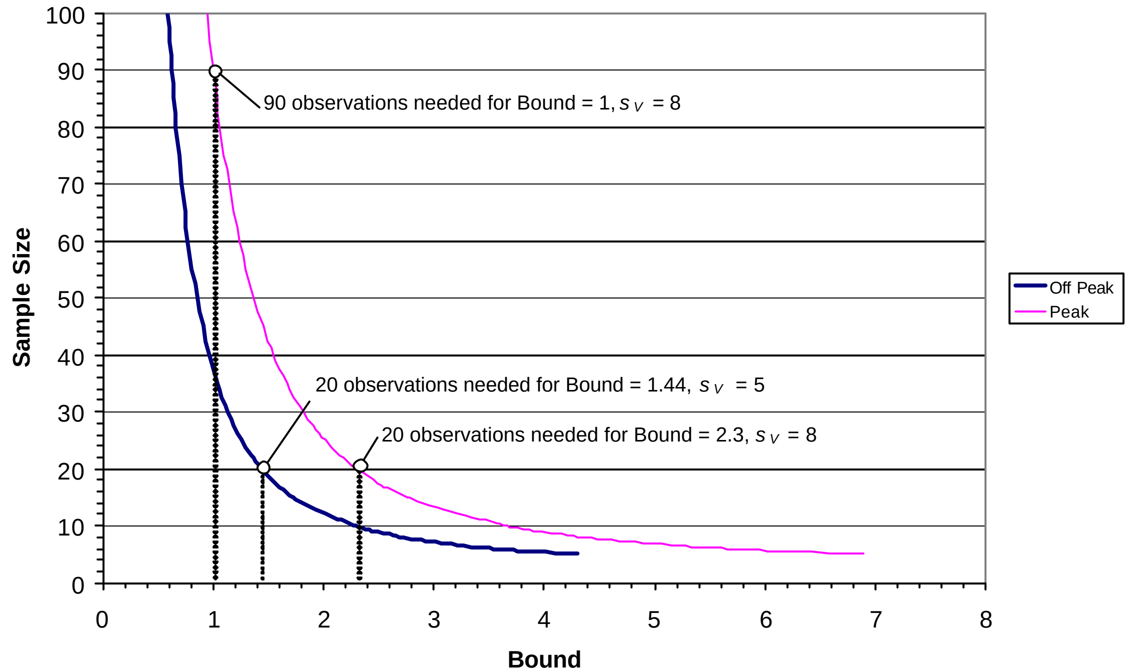

Figure 6. Necessary Sample Size For Day-to-Day Variability as a Function of the Bound of a 90% Confidence Interval Assuming Peak Period Variability is 8 mph and Off-Peak Variability is 5 mph.

5. Costs

The two methods most suitable for our purposes are probe vehicle methods and license plate matching. Approximate costs for variations on these two techniques are given below.

5.1 Probe Vehicle Techniques

Probe vehicle studies are quite common and a number of data collection firms do this type of work. The state of the practice currently is to use GPS receivers. The approximate cost of probe vehicle data collection, based on a few recent projects is \$400-\$500 per vehicle per day, which includes the cost of the vehicle and labor (17). The cost of the GPS equipment is additional and would run approximately \$500 per unit (18).

Based on past probe vehicle data collection efforts, it is reasonable to expect 5 unidirectional runs per vehicle in 3 hours for an approximate 5 mile corridor. Assuming a data collection rate of 10 segment travel time measurements per vehicle per day, the cost of a single measurement is \$500÷10 = \$50. The cost then, of 100 accuracy measurements is \$50×100 = \$5,000. Two probe vehicles could complete this number of runs in a five day work week. A staff-month of effort should be budgeted for planning, data reduction, management and other overhead. The cost elements break down as follows:

Item	Qty	Cost
Planning, management, etc.	1 staff-month	$15,000
GPS equipment	2 vehicles	$500 ea.
Data collection	5 days × 2 vehicles	$500 per vehicle per day

Based on these costs, 100 observations could be collected at a cost of $21,000.

5.2 License Plate Matching Techniques

Because license plate matching is better suited to collecting a lot of data at one or many fixed locations over a period of time, lends itself well to measuring day-to-day variability. The following costs are presented with this type of data collection in mind. They come from (7,8).

Video with manual transcription
We assume we need 20 days of peak period data at one hour per day. By this technique, two lanes can be captures with a single camera as long as some missed reads due to occlusion can be tolerated. Because variability is lower in the off-peak, we need fewer off-peak observations. As variability in the off-peak is approximately half of that in the peak periods, we assume we need 10 days of off-peak data at one hour per day. The cost elements are:

Item	Qty	Cost
Planning, management, etc.	1 staff-month	$15,000
Video cameras:	4 locations	$2,500 ea.
VCR/TV	1	$1,500 ea.
Miscellaneous field supplies	4 locations	$150 ea.
Field personnel	4 loc. × 30 hrs	$25 per person per hour
Data transcription	4 loc. × 30 hrs × 10 hr/hr	$15 per person per hour

Based on these assumptions, the approximate cost for this study is $48,100. Note that of this cost, the labor component (field personnel and data reduction) is $36,000.

Video with optical character recognition
Automated LPRs reduce the amount of costly and labor intensive data reduction that is required. However, LPR equipment is far more costly than video cameras, televisions and VCRs. In the past, travel time surveys with LPRs have been subcontracted to private companies (7,8). The Travel Time Data Collection Handbook estimated the cost of outsourcing data collection with LPRs to be $300-$400 per lane-hour (in 1998).

Item	Qty	Cost
Subcontracting	2 loc. × 2 lanes × 30 hrs	$400 per lane per hour

For the same amount of data as for manual transcription, the cost is $48,000 (120 × $400) per city.

The other option is to purchase the equipment. For multiple studies, this may be more cost effective than subcontracting each time. Approximate costs are:

Item	**Qty**	**Cost**
Planning, management, etc.	1 staff-month	$15,000
Video specialists	4 loc. × 30 hrs	$25 per person per hour
Video camera & accessories	2 lanes × 4 loc.	$2,500 ea.
LPR hardware and software	1 unit	$15,000 ea.
Miscellaneous field supplies	4 locations	$150 ea.
Data reduction	2 lanes × 4 loc. × 30 hrs × 4 hr/hr	$15 per person per hour
VCR/TV	1	$1,500 ea.
Training		$5,000

Based on these cost estimates, the total cost for the first study would be $74,500. For each additional study, the cost of labor is $32,400. This approach would pay for itself over subcontracting with the 2nd study. The breakeven analysis is shown in Figure 7. Note, however, that video with manual transcription is always cheaper than subcontracting LPR data collection. Therefore, the more relevant comparison is between purchasing LPRs and using video with manual transcription for each city. Based on this, purchasing LPRs is most cost-effective if studies are planned for multiple cities. It is worth noting again that some training and experience are required to get the best results.

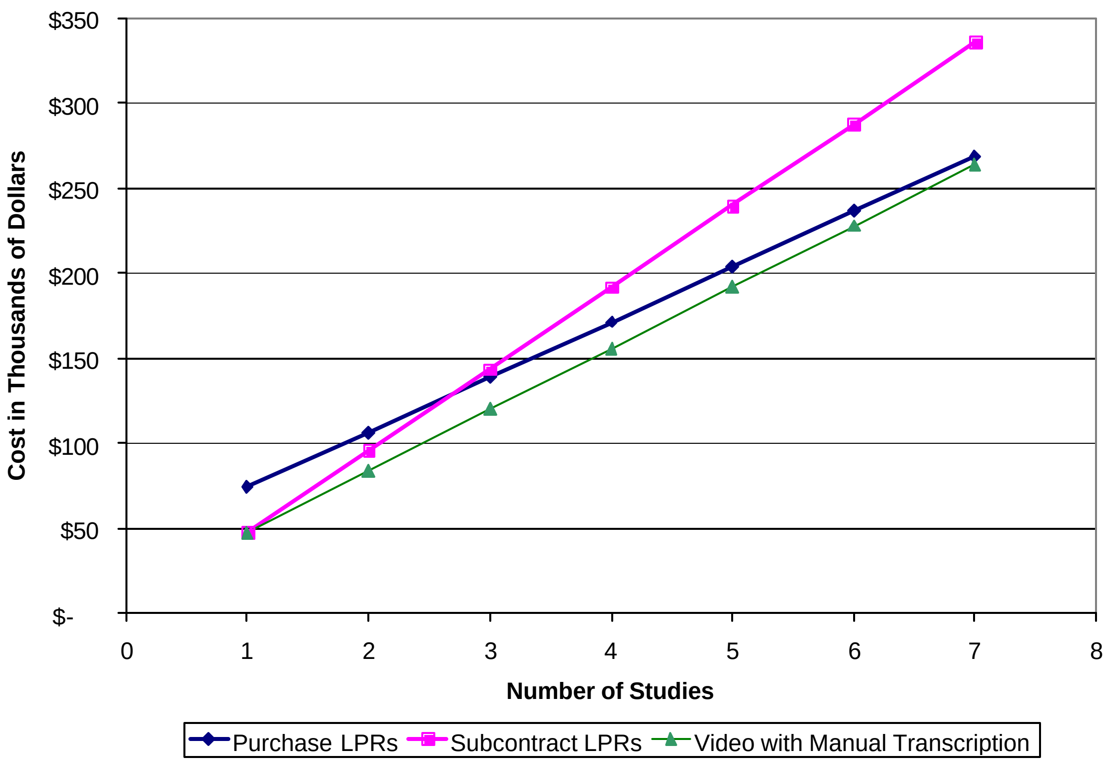

Figure 7. Breakeven analysis for license plat matching data collection options

6. Recommendations

6.1 Two Step Approach

The purpose of this white paper is to recommend an approach to measuring ATIS accuracy in order to guide future deployment decisions. It is not enough just to know accuracy, however. A measure of day-to-day variability is needed in order to know how accurate the ATIS information needs to be. Variability is difficult to measure because data needs to be collected on the same segment at the same time over many consecutive days (at least 20). However, if the ATIS is quite accurate, its travel time estimates may be used to calculate variability, obviating the need to collect additional ground truth data in the field. Therefore, we recommend a two-step approach. The first step is to measure accuracy. Then, only if accuracy is low should field data be collected to calculate variability. A reasonable threshold is an SDPE of 0.15.

Approximately $20,000 to $70,000 should be budgeted per city for an ATIS travel time accuracy study. The amount depends on whether variability needs to be calculated in the field or whether the ATIS is accurate enough that it's travel time estimates can be used to calculate variability.

6.2 Step One: Measuring Accuracy – Probe Vehicle Approach

An important consideration in measuring ATIS travel time accuracy is to ensure a representative sample of data points is collected. The accuracy of different segments can be very different if detector reliability is not even across the network. In addition, accuracy is likely to be lower during peak periods since loops tend to be less accurate at low speeds. Therefore, ground truth travel time should be collected for multiple segments in the network at different times of day. Probe vehicles are best suited to this type of data collection. While they can not collect as much data as license plate matching techniques at any one location and time, it is more important to be able to widely sample the network. Based on the minimum sample size equation and the marginal cost of each data point collected, we recommend collecting approximately 100 data points for the accuracy measurement. In Section 5.1, we estimated this cost to be approximately $21,000 per city.

6.3 Step Two: Measuring Variability – License Plate Matching

Only if the ATIS accuracy is below 0.15 should a second study be undertaken to measure day-to-day variability. First, while license plate matching is the most robust in terms of ensuring reliable and accurate ground truth measurements, it is costly for the amount of data that can be collected. Second, in order to measure variability, data needs to be collected at the same time over multiple days which would involve a lot of setting up and breaking down of equipment. For a single study, it makes the most sense to use video cameras with manual transcription. In Section 5.2, we estimated this cost to be approximately $33,100. If multiple studies are planned, however, it makes sense to consider the added expense of license plate readers with optical character recognition. Based on our cost estimates in Section 5.2, purchasing LPRs makes sense if more than one study is planned.

References

1. Wunderlich, K., M. Hardy, J. Larkin, and V. Shah. On-Time Reliability Impacts of Advanced Traveler Information Services (ATIS): Washington, D.C. Case Study. U.S. Department of Transportation, ITS Joint Program Office, January 2001. EDL#13335.
2. Jung, S., J. Larkin, V. Shah, A. Toppen, M. Vasudevan, and K. Wunderlich. On-Time Reliability Impacts of Advanced Traveler Information Services (ATIS) Volume II: Extensions and Applications of the Simulated Yoked Study Concept. U.S. Department of Transportation, ITS Joint Program Office, March 2002. EDL#13630
3. Shah, V., Jung, S., J. Larkin, A. Toppen, M. Vasudevan, and K. Wunderlich. Accuracy and Coverage Implications of Advanced Traveler Information Services (ATIS) Benefits, Volume III: Extensions and Applications of the Simulated Yoked Study Concept. U.S. Department of Transportation, ITS Joint Program Office, February, 2003.
4. Small, K., Noland, R., and Lewis, D., "Valuation and Travel-Time Savings and Predictability in Congested Conditions for Highway User-Cost Estimation", NCHRP Report #431, National Academy Press, Washington, D.C., 1999.
5. Wunderlich, et. al., Urban Congestion Reporting. Ongoing task for the Federal Highway Administration, U.S. Department of Transportation, Washington, D.C.
6. http://itsdeployment2.ed.ornl.gov/its2000/default.asp
7. Turner, S., Eisele, W., Benz, R., and D. Holdener. Travel Time Data Collection Handbook. Report FHWA-PL-98-035. U.S. Department of Transportation, Federal Highway Administration. March 1998.
8. Shuldiner, Paul. Telephone conversation regarding Calvin Coolidge Bridge reconstruction project, Northampton, Mass. February, 2003.
9. Middleton, D.R., D.L. Jasek, and R.T. Parker. Evaluation of Some Existing Technologies for Vehicle Detection, Research Report 1715-S. Texas Transportation Institute, College Station, TX. September 1999.
10. ITS Engineering and Constructors, Inc. Maintaining and Estimating Travel Times in Work Zones. The ENTERPRISE Program. July, 2002.
11. Carter, M., et al. Automated Vehicle Identification Tags In San Antonio: Lessons Learned From The Metropolitan Model Deployment Initiative – Unique Method For Collecting Arterial Travel Speed Information. U.S. Department of Transportation, Washington, D.C. Report FHWA-OP-01-037. EDL #13230.
12. Eisele, W.L. and L.R. Rilett. A Statistical Comparison of Travel Time Estimates obtained from Intelligent Transportation Systems and Instrumented Test Vehicles. Presented at the 81st Annual Meeting of the Transportation Research Board, Washington, D.C., January 2002.
13. Ygnace, J., Drane, C., Yim, Y.B., and R. de Lacvivier. Travel Time Estimation on the San Francisco Bay Area Network Using Cellular Phones as Probes. UCB-ITS-PWP-2000-18. 2000.
14. Lovell, David J. Accuracy of speed measurements from cellular phone vehicle location systems. ITS Journal. Vol. 6, No. 4, pp. 303-325. 2001.
15. Coifman, B., and Cassidy, M. "Vehicle Reidentification and Travel Time Measurement on Congested Freeways", Transportation Research: Part A, vol 36, no 10, 2002, pp. 899-917.

16. Hardy, M., Larkin, J., Shah, V., and K. Wunderlich. Accuracy of Travel Time Estimates Obtained From Advanced Traveler Information Services, to appear in the proceedings of the 9th Annual Meeting of the Intelligent Transportation Society of America, November 2000.
17. Sallman, Doug. Cambridge Systematics, Inc. Email correspondence regarding past probe vehicle data collection subcontracting efforts. March, 2003.
18. Battelle, Global Positioning Systems for Personal Travel Surveys: Lexington Area Travel Data Collection Test. U.S. Department of Transportation, Washington, D.C. September 1997.