TECHNICAL APPENDIX

for

GUIDELINES FOR EVALUATING THE ACCURACY OF TRAVEL TIME AND SPEED DATA
Version 1.0, June 2011

Developed by:
Texas Transportation Institute
University of Virginia
Virginia Center for Transportation Innovation and Research

Pooled Fund Study Sponsors:
Virginia Department of Transportation (Lead State)
Alabama Department of Transportation
California Department of Transportation
Federal Highway Administration
Maryland State Highway Administration
Michigan Department of Transportation
Pennsylvania Department of Transportation

Recommended attribution for this document:

Guidelines for Evaluating the Accuracy of Travel Time and Speed Data: Technical Appendix, Version 1.0, June 2011.

GUIDELINES SECTION 2 – TECHNICAL APPENDIX

Covers technical issues related to Section 2. Develop an Evaluation Plan of the guidelines.

Guidance on Reidentification Link Length

On page 11 (Section 2) and page 21 (Section 3), there is guidance that recommends that "… the benchmark link length should not differ from the TIS link length by more than 10%." This is reported as a "general rule-of-thumb" but is also supported by an error analysis done for the I-95 Corridor Coalition. In their January 2009 evaluation report of INRIX data, the University of Maryland also addressed this and gave the following guidance (bold underline emphasis is added here):

> *The Bluetooth technology measures a sample of travel times between two cross sections of the highway. The distance between a pair of sensors will then be needed to convert the travel times into speeds. In practice sensors were deployed based on the coordinates provided in the traffic message channel (TMC) location database supplied by INRIX. In the deployment process, every effort was made to minimize the distances between sensor deployment locations and the target TMC points. The targeted TMC points were pinpointed on a geographical information system (GIS) map overlaid on top of the aerial photos of the highways of interest so that the TMC points could be identified with ease in the field. Also, the TMC coordinates were inputted into a hand-held global positioning system (GPS) device which was used in the field. Apart from normal device and human errors that can be expected in this type of operation, occasionally lack of shoulders in the right place along the highway, and/or lack of a permanent object in the close proximity of the TMC point to which the sensor could be safely tethered introduced some unwanted but additional errors in the placement of the sensors. Sometimes, the location of the TMC point that was of interest fell in a construction zone that also could be a cause for minor errors.*
>
> *Due to the importance of this issue and its effect on validation results a thorough analysis on the size and causes of the deployment errors were performed. To this end, the recorded coordinates of the sensor deployment locations (as recorded by the GPS component of the sensor upon deployment) were overlaid on the same GIS map as TMC points. Since the great circle distance between two points may not be an accurate estimate of the actual distance traveled by vehicles between two points along the highway, in all cases the highway distances between the sensor location and its corresponding TMC point were measured on a GIS system. These discrepancies are reported as the placement errors in the corresponding result tables which will appear in the following sections. The actual distance between any pair of deployed sensors is estimated using the same GIS map on which the TMC points were located. For each TMC segment and "ground truth" speed estimations, this is reported as the corrected length of the segment under consideration.*
>
> *It can be shown that, if we were to use the standard TMC lengths instead of actual distances between sensors for speed estimation, the acceptable tolerance in sensor deployment would be a function of travel time between two sensors. Specifically, to keep the error in speed estimation to less than one mph, the discrepancy between the length of the segment and actual TMC length (in feet) should be less than 1.47 times observed travel time on the segment in seconds. For a five minute travel time this translates into a 441 ft* ***or 8.3% (on a one mile segment) tolerance which***

for all practical purposes was adopted as the maximum acceptable tolerance in sensor deployments.

Estimation of Ground Truth Travel Time Using Statistical Sampling Theory

Background

Many past data quality evaluations have treated the determination of ground truth travel time as a measurement problem. This is to say, that ground truth travel time was treated as a fixed value. These guidelines take a slightly different approach and suggest that ground truth travel time is a value that can only be estimated. The main difference in these approaches is that the guidelines acknowledge that there is a degree of uncertainty in estimating a random variable. This is in some ways a fairly conservative approach because it does not provide a single point value that can be used for benchmark evaluations of a traveler information service. Rather, a confidence interval, or range of likely values is generated from the sample data. This range represents the benchmark or ground truth estimate of travel time.

Statistical Sampling

The statistical sampling approach to estimating ground truth travel time follows the same approach taken in many statistical sampling applications. First, a population of interest is defined. The population is simply an enumeration of what is to be estimated. For travel time studies, the population is the set of travel times for all vehicles that traverse a specified link during a specified time interval. In reality, there are a finite set of vehicles that traverse a specified link during a specified time interval. So, in theory at least, they could be exhaustively enumerated. In practice it would be extremely costly to observe the travel time of every vehicle so a sample of observations is collected.

Statistical sampling is the process whereby a randomly selected subset of the population is observed and statistics are computed from the sample data. For example, since not every vehicle can be observed, a random subset of the vehicles traversing a link could be observed and the travel times for those vehicles then recorded. This would then be a sample of observations from the population. The sample plays an important role in statistical inference.

Statistical Inference

Statistical inference is the process whereby statistics computed from the sample data are used to infer population parameters. For example, ground truth travel time is generally defined as the average travel time of a link during some time interval. This is equivalent to stating that ground truth travel time is the mean travel time of the population. Using statistical inference, the population mean can be estimated from the sample mean. However, the sample mean is a random variable and therefore is subject to some uncertainty. The Central Limit Theorem states that the distribution of the sample mean is approximately normal and the variance of the sample mean decreases as the sample size increases. This is an important insight into the problem of estimating a value to use as ground truth. For small samples, the variance of the sample mean is quite large. Therefore, there is very little certainty or confidence in the sample mean as an estimate of the population mean. However, as the sample size increases, the sample estimate of the population mean becomes more precise due to the decrease in the variance of the sample mean.

Confidence Intervals

All of these facts can be summarized succinctly in terms of a statistical confidence interval on the sample mean. A confidence interval (CI) is a method of expressing the likelihood that a population parameter falls between two values. For example, a confidence interval can be used to express how likely it is that the ground truth travel time falls between two values.

The expression for a confidence interval is given by:

$$\bar{X} - t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}} \qquad (0\text{-}1)$$

where:

$\bar{X}$ *is the sample mean*
$t_{\frac{\alpha}{2},n-1}$ *is the Student's t statistic*
α *is the acceptance level of the test statistic*
s *is the sample standard deviation*
n *is the sample size*
μ *is the population mean*

The expression states that for some given acceptance level (alpha) the population mean will lie between the computed values in approximately (1 – alpha) percentage of the observed cases. To a test engineer, this is simply a method to determine the most likely range of values for the ground truth.

A Note on Test Statistics. One thing to note is that the expression for a confidence interval uses a "Student's t Test Statistic". In many engineering applications the test statistic is given as "Z" which is the standard unit normal distribution. There are some reasons to consider using the "t" statistic over the "Z" statistic in computing a confidence interval.

First, the "Z" statistic is used in cases where the population variance is a known value. In some engineering applications, the population variance is fairly stable between sample data collections. In these cases, the population variance can be reasonably assumed to be a known constant value that can be estimated from a large sample collection. However, in travel time studies the population of interest is the set of travel times of vehicles traveling a link during some time interval. The variance of this population is subject to a fair amount of uncertainty due to a number of different factors. For example, the variance of travel times on an arterial link will be different than the variance of travel times on a freeway link. The variance will be affected by time of day, volume of traffic, geometric design, weather conditions, and a number of other factors. Therefore, the variance of the population is typically estimated from the sample variance. From a mathematical statistics perspective, whenever the variance is estimated from sample data, it is suggested that a Student's t statistic be used to account for the additional uncertainty.

Although the Student's t statistic is computationally more complex than the Z statistic, there is a second reason to prefer its use. The "t" statistic provides a more conservative confidence interval than the "Z" statistic because it is sensitive to the sample size. The "Z" statistic is only sensitive to the alpha-level which is typically set at a constant 0.05 or 95% confidence level. The "t" statistic, however, is sensitive to both the alpha-level and the sample size (expressed in degrees of freedom). So, for small samples where the sample variance is also small, a confidence interval expressed using a "Z" statistic will be unreasonably narrow. This is likely a misleading result, because the estimate of the population variance

from the sample variance is dependent on the sample data collected and would likely change if a new sample were taken. The Student's "t" statistic provides a more conservative estimate of the confidence interval width because it "penalizes" very small samples and accounts for the additional uncertainty due to the use of sample variance.

The following chart shows a comparison of the "Z" and "t" test statistics. In particular, one can see that for very small sample sizes (e.g. 3 observations) the width of the confidence interval is far too wide for any meaningful applications when using the "t" statistic. The width of the "Z" statistic at only 10% variation (i.e. CV = 0.10) and a sample size of n = 3 observations is about 11% which is close to the acceptable range for many applications. This indicates that in cases where the number of sample observations is limited, caution should be used when developing a confidence interval using a "Z" statistic.

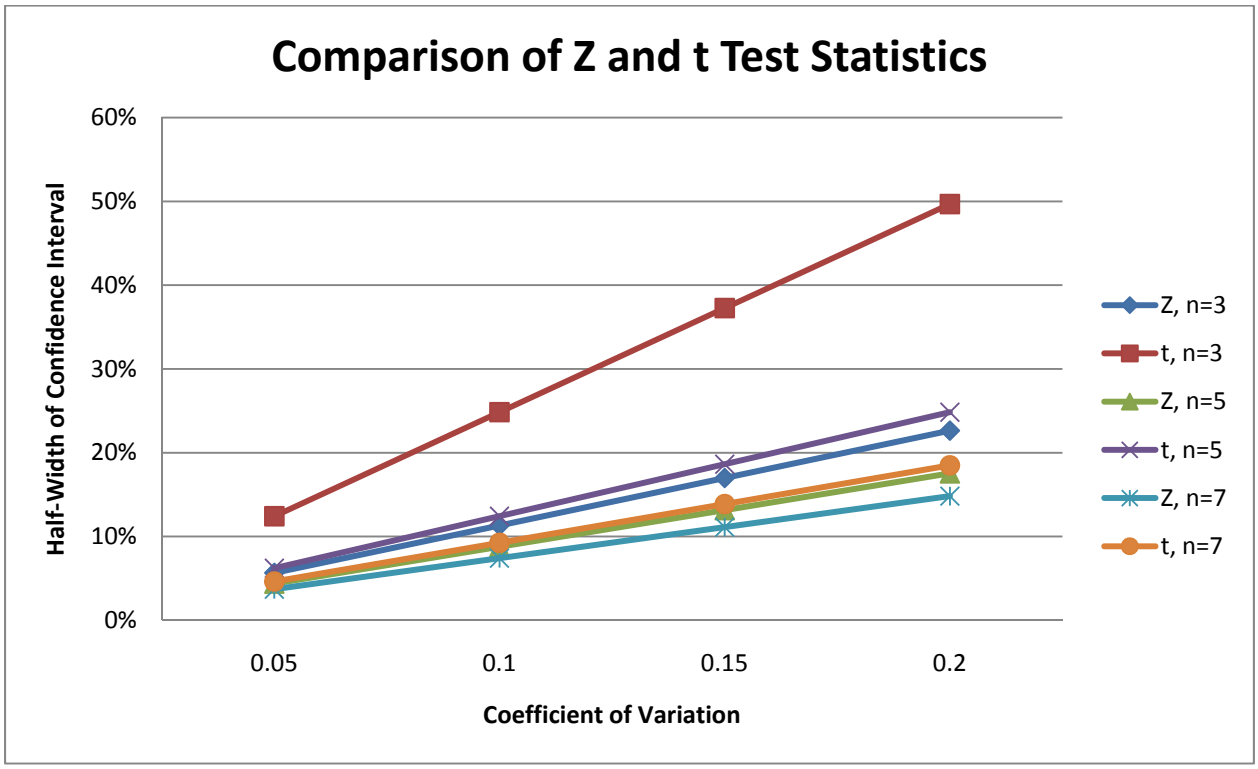

Figure A-1. Comparison of Z and t Test Statistics

Example of a Confidence Interval Application. Assume that sample travel time observations are collected in the field for some link. A total of 3 observations are collected in a 15 minute period using the floating car method. The average travel time from these observations was 120 seconds with a standard deviation of 9 seconds. A confidence interval can be computed using the two methods described in the previous section.

$\bar{X} = 120\ secs$
$\alpha = 0.05$
$n = 3$
$t_{\frac{\alpha}{2},n-1} = 4.30\ (found\ using\ a\ lookup\ table\ or\ statistics\ package)$
$s = 9\ secs$

Using the Student's "t" statistic:

A 95% Confidence Interval is expressed as: $120 - 4.30\ \frac{9}{\sqrt{3}} \leq \mu \leq 120 + 4.30\ \frac{9}{\sqrt{3}}$

Therefore, the range of feasible values for the mean travel time lies between 98 and 143 seconds. In this case, it should be clear that although the standard deviation for the observations was fairly small at only 9 seconds, the confidence interval is approximately 22 seconds on each side of the sample mean.

Using the "Z" statistic:

A 95% Confidence Interval is expressed as: $120 - 1.96\ \frac{9}{\sqrt{3}} \leq \mu \leq 120 + 1.96\ \frac{9}{\sqrt{3}}$

The "Z" statistic at the 95% level is approximately 1.96. Therefore the range of values for the confidence interval when using a "Z" statistic would be between 110 and 131 seconds. This range is a much more optimistic range and does not account well for the small sample size or the uncertainty in estimating the variance from the sample.

The choice of a test statistic is an important decision because it will affect the precision of a statistical estimate. The Z statistic has been used in many studies because it is easier to compute and is generally accepted for many engineering applications. The guidelines do not endorse either test statistic but do encourage users to consider the trade-offs in terms of precision and robustness of estimates when deciding on a test statistic.

Determining the minimum sample size

The expression for a confidence interval can be manipulated to derive an expression for the minimum sample size. Depending on the preferred test statistic there are two possible expressions for minimum sample size:

Table A-1. Minimum Sample Size Using Absolute Width of CI

(1) When the population variance is known:	$n = \left(\frac{Z\sigma}{d}\right)^2$
(2) When the population variance is estimated from sample variance:	$n = \left(\frac{ts}{d}\right)^2$

Where:

$d\ is\ the\ desired\ width\ of\ the\ confidence\ interval\ in\ units\ of\ measure\ (e.g. mph, secs)$

and other terms are given as above.

Note, that in case 2, because the Student's "t" statistic is a function of sample size (i.e. degrees of freedom) the solution for "n" is iterative. The method to solve for "n" is explained well in Engineering Statistics textbooks.

The minimum sample size can also be expressed relative to the population mean. For example, rather than determining the width of the confidence interval in absolute units (e.g. mph or seconds), the

confidence interval can be expressed as a percentage (e.g. 10%). This is sometimes preferred to avoid problems where the magnitude of the units change between populations. The same expressions for minimum sample size can be written using relative widths as:

Table A-2 - Minimum Sample Size Using Relative Width of CI

(1) When the population variance is known:	$n = \left(\frac{Z \frac{\sigma}{\bar{X}}}{e} \right)^2$
(2) When the population variance is estimated from sample variance:	$n = \left(\frac{t \frac{s}{\bar{X}}}{e} \right)^2$

Where:

$$e \text{ is the desired width of the confidence interval, expressed as a decimal value (percentage)}$$

Example of Determining a Minimum Sample Size. The equations for determining a minimum sample size can also be applied to this example.

If the desired confidence interval width is specified as 10 seconds:

Table A-3 - Example of a Minimum Sample Size Calculation

(1) When the population variance is known:	$n = \left(\frac{1.96 * 9}{10} \right)^2 = 4\ obs.$
(2) When the population variance is estimated from sample variance:	$n = \left(\frac{4.30 * 9}{10} \right)^2 = 130\ obs.$

Clearly, the sample size when using the "t" statistic is too small for the desired confidence interval. An iterative solution shows that the minimum sample size using a "t" statistic would be 7 observations.

Similarly, the sample size can be computed as a function of relative width of the confidence interval. The width is specified as a decimal percentage and the standard deviation is replaced by the coefficient of variation.

Coefficient of Variation

Notice, that in this new formulation of the minimum sample size, the standard deviation is divided by the sample mean. This is called the "coefficient of variation" and is heavily used in the research throughout these guidelines. The "coefficient of variation" (CV) is a non-dimensional (i.e. unit-less) measure of relative variation and is the ratio of the standard deviation to the mean.

One of the nice features of the CV is that it is not sensitive to changes in the magnitude of the units of measure. For example if travel time is being studied on a 3 mile link the mean and standard deviation of travel time will tend to be smaller than studies on a 6 mile link. Therefore, comparing statistics between links of different lengths is not helpful. The coefficient of variation is a useful way to express relative variation. The CV can be compared across links of different lengths.

Network Stratification Research

The previous section on estimating ground truth travel time using statistical sampling and inference discussed the basis for viewing ground truth travel time as a statistical estimation problem and how to estimate a confidence interval using statistical inference. This section will present the results of empirical research into the spatial and temporal distribution of travel time variance.

Why Stratify a Network?

The idea of stratifying a network for travel time sampling has been around for a while. In NCHRP Report 398, *Quantifying Congestion*, the authors recommended that congestion studies use a similar stratification method as what is presented in these guidelines. Research has shown that when sampling from different populations, more precise inference on population parameters can be achieved by stratifying the populations. The basic idea is to identify variables that group the populations into more homogeneous categories.

These guidelines presented a simple but effective approach to stratifying a road network. Basically the network is first stratified by facility type such that samples are collected on arterials separately from freeways. Second, the guidelines recommend that freeways be further stratified by using a classification method to separate freeway segments into high and low variance segments. This method is based on analysis of empirical data collected from Houston, Texas. More detailed information about the freeway stratification method is shown in the following sections.

Influence of Time Intervals on Statistics

One of the factors that can affect the statistics of travel time studies is the time interval used to aggregate the sample observations. In the Houston study used for the development of these guidelines, the base time interval used to aggregate observations was 5-minutes. This time interval was selected because it is a commonly used time interval by Traveler Information Services for reporting travel time estimates. It is also the smallest interval of interest for most current applications. The 5-minute interval represents a fairly granular level of analysis. Longer time intervals will have different effects on the magnitude of variance observed in the data. Therefore, it is recommended that the values for ranges of travel time variance reported in this research be considered within the context of a 5-minute sampling interval. Samples collected over longer time intervals may show significantly higher ranges of travel time variance.

Relationship between Coefficient of Variation of Travel Time and Different Traffic States

The CV of travel time was shown in previous sections to be a useful indicator of relative variation in traffic data. When used in traffic engineering studies, CV travel time is also a useful indicator of different traffic states. Empirical data from Houston was analyzed and it was found that the distribution of CV travel time was correlated with different average speeds of traffic flow. The following table summarizes the findings from this analysis.

Table A-4. Relationship Between CV Travel Time and Traffic Conditions

Coefficient of Variation of Travel Time Range	Description of Traffic Flow
"Low" : Values less than 0.10	Generally free flow conditions where average speeds are around 60 mph
"Medium": 0.10 – 0.20	Transition and congested states of traffic flow where speeds range from 20 mph – 50 mph.
"High": Values greater than 0.20	Highly unstable traffic flow with the possible presence of outlier observations influencing the degree of variation. Generally congested or transition speeds of traffic flow.

To further illustrate this relationship, the distribution of CV travel time was computed for 5-minute intervals on a large set of freeway segments in Houston. A histogram for different average speeds was computed and is shown in the following plot.

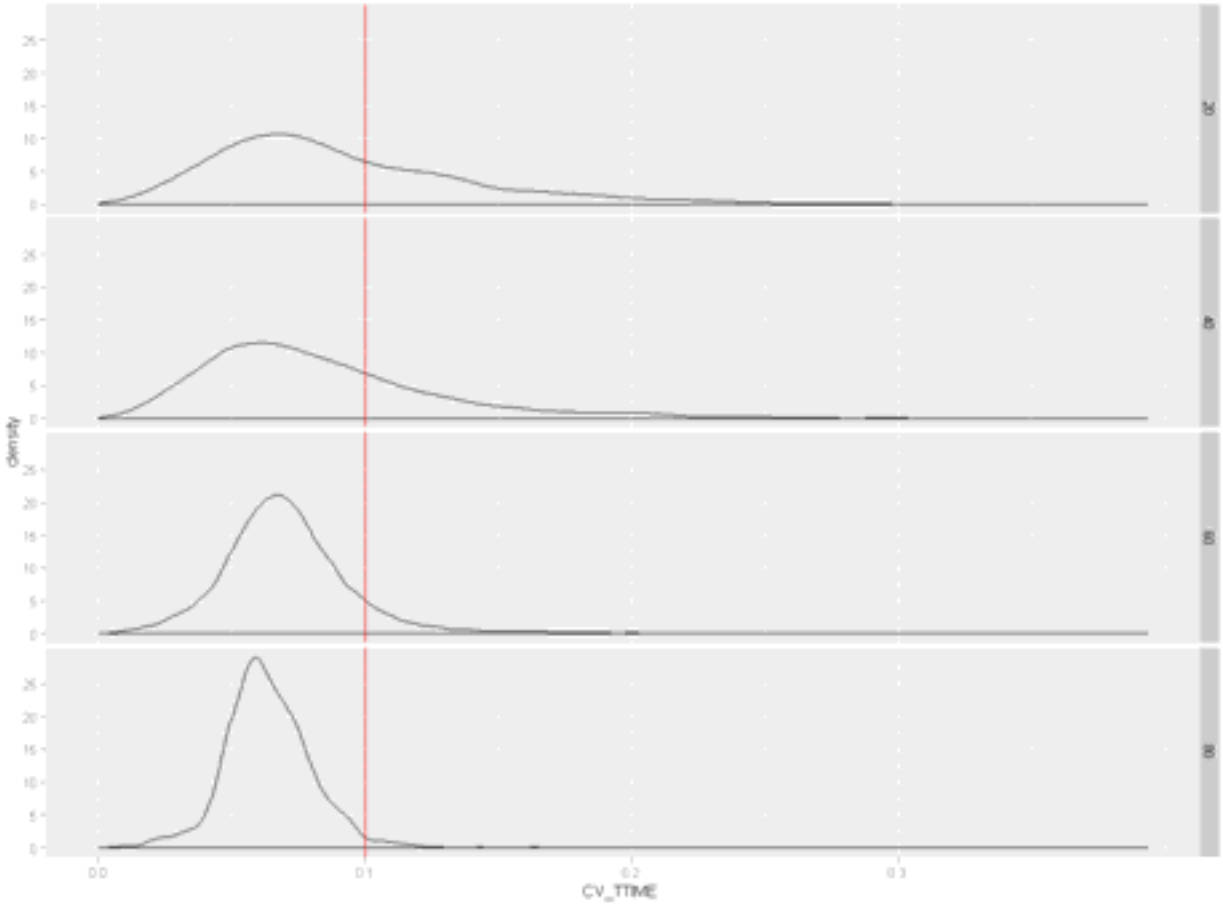

Figure A-2. Relationship between Average Speed (In Bins) and CV Travel Time (Density Plots)

The plot shows that as average speeds get smaller (i.e. congestion forming) the distribution of CV travel time tends to be more skewed towards larger values. This indicates that higher variances in travel time are typically observed when traffic is transitioning between free flow and congested states. The solid red line indicates that point where CV travel time = 0.10.

Classification of High and Low Variance

The previous section showed the relationship between CV travel time and different traffic states. The Houston data shows that higher values of CV travel time are observed when congestion is forming or when traffic is in a congested state. However, for the purposes of developing a classification method for network links a "threshold" value is needed to distinguish between "high" and "low" variance states.

The threshold value used in this research was CV = 0.10. This value was selected because it represented a reasonable point to distinguish between free-flow and congested traffic states. This is also the point where the minimum sample size for estimation of ground truth travel time confidence intervals start to become begin enough to warrant the exclusive use of re-identification technologies.

The following table shows the minimum computed sample size using a "Z" and "t" statistic for different levels of CV travel time at a 95% Confidence Level and 10% precision.

Table A-5. Minimum Computed Sample Size Using a "Z" and "t" Statistic

	Min. Sample Size (95% CI, 10% precision)	
CV	"Z"	"t"
0.04	1	3
0.06	2	4
0.08	3	5
0.1	4	7
0.12	6	9
0.14	8	11
0.16	10	13
0.18	13	15
0.2	16	18

For CV levels less than 0.10, the required sample size for the "Z" and "t" statistic is around 5 or fewer observations. For CV levels greater than 0.10 the required sample size is in the range of 7 – 18 or more observations depending on how large the variance in the sample is. This dichotomy, while somewhat subjective, represents a breaking point that can be used to distinguish between cases where a less statistically robust sampling technique such as the floating car method could be substituted for a technique that generated sufficiently large samples.

Identifying Links with "High" Travel Time Variance

The guidelines make an effort not to endorse any one particular sampling technique. However, the research into travel time variance has shown that certain types of links are more likely to experience high travel time variance during sampling intervals than other types of links. The rationale for providing a freeway stratification method was so that evaluations could easily identify links in the network that are likely to experience high travel time variation during sampling intervals. Once these links are identified, the guidelines recommend that statistically valid samples are collected on these links to estimate the ground truth confidence interval.

The methodology used to develop the network stratification method presented in these guidelines was based on a comprehensive analysis of travel time variance in Houston. Two different classification methods were tested using an open-source data mining tool from the University of Waikato called "Weka". The classification methods tested were the J48 decision tree (a Java-based implementation of Quinlan's C4.5 decision tree) and the Naïve Bayes method. These methods were selected because of their effectiveness and simple interpretation in classification problems.

Data Collection

The freeway network in the Houston metropolitan region is monitored by an extensive network of toll tag readers. These readers record the time a vehicle passes a specific point in the network and from these timestamps the travel time of individual vehicles along specified links can be computed. For this study, one year of travel time data from the year 2008 was loaded into a database for analysis. The original data set covered approximately 200 links. This amounted to approximately 273.9 million individual travel time observations.

Table A-6 shows the data structure for the aggregated travel times. The data was aggregated by link (reader to reader) and 5-minute time interval. For each link and 5-minute interval sample statistics and a confidence interval were computed.

Table A-6. Aggregated Data Structure

Column	Data Type	Explanation
CALENDAR_KEY	Number	Refers to the calendar date
TIME_KEY_FIVEX	Number	Refers to the five minute aggregation period of the day (e.g. from 11:00 – 11:05 AM)
SEGMENT_ID	Number	Refers to the link ORIG_ID and DEST_ID
NUM_OBS	Number	Number of observations in the sample
AVG_TTIME	Number	Arithmetic mean of travel time
AVG_SPEED	Number	Arithmetic mean of speed
STDDEV_TTIME	Number	Standard deviation of travel time
STDDEV_SPEED	Number	Standard deviation of speed
MIN_OBS_TTIME_95_T	Number	Minimum number of observations needed to estimate mean travel time with 95% confidence level and 10% precision using Student's T statistic
CV_TTIME	Number	Coefficient of Variation of travel time
HARMONIC_MEAN_SPEED	Number	Space mean speed computed by distance / (average travel time)

Data Filtering. The data set was filtered to only include samples between 6 am and 8 pm on weekdays. The filtering was done to reduce the impact of weekend and off-peak travel on the cumulative distributions. The data set was also filtered to only include samples where the sample size was sufficient to estimate the mean travel time using a T statistic. This calculation was based on Equation 3. For each sample it was determined if the sample size was sufficient to estimate the mean travel time with a 95% confidence level and 10% precision. The aggregated data resulted in 20.9 million samples. From this set, samples with 2 or fewer observations were rejected, resulting in 16.5 million samples. From the resulting set of samples, approximately 85% of the samples (14 million) were kept.

Links Selected for Analysis. A total of 117 links were selected for analysis. These links were identified as the core links in the Houston network. HOV and tolled facilities were excluded from the analysis. The selected links covered approximately 297 center-line miles of roadway with a range of ADT values between 57,000 and 160,000 per link. ADT values were interpreted from published TxDOT ADT maps. The number of access points on each link was determined by using GIS layers to identify on and off ramps. Additionally, each link was visually inspected using satellite imagery (via Google Earth) to identify the average number of through lanes over each link.

Classification Methods Examined

Decision tree methods have been widely used in data mining problems. A decision tree classifier such as J48 attempts to classify by splitting the training data along different dimensions. Each split results in a node with branches. The optimal size of the tree is determined by the algorithm's pruning strategy. Naïve Bayes methods have also been widely used in data mining. The method takes its name from Bayes' rule. Each attribute is assumed to be independent and probability distributions are computed from the empirical distribution of the attribute. These probabilities are combined according to Bayes' rule and a posterior probability for a class given a set of attributes is computed. The posterior probability is used for classification.

Baseline Model. The baseline model used for comparison is a simple classifier that "votes" for the majority class. This method will always have a correct classification rate equal to the relative size of the majority class. In this problem, the majority class was the "low" variance class and comprised 62% of the cases to be classified. So the baseline model achieved a 62% correct classification against the data set. However, since identification of the "high" variance class is of greatest interest in this problem we also want to evaluate models based on how well they classify the target class. A classifier that both exceeds the baseline's correct classification rate (62%) and also improves identification of the minority class would be considered a good candidate.

Kappa Statistic. A popular statistical measure of classifier performance is the Kappa statistic. It is shown in Equation 4.

$$\kappa = \frac{P(a) - P(e)}{1 - P(e)} \tag{4}$$

where $P(a)$ is the relative observed agreement between the model and the data, and $P(e)$ is the hypothetical probability of a random agreement between the model and the data. A baseline model that simply votes for the majority class has a Kappa value of 0. An ideal model that correctly classifies all instances would have a Kappa statistic of 1.

Results of Model Exploration. The J48 decision tree and Naïve Bayes classification methods were explored using the data mining software. Each method was tested using the full set of parameters as well as subsets of the parameters. The models generated by the data mining software were used to guide the development of the proposed model in this paper. Exploration using these methods resulted in the following insights:

1. ADT per lane and link length are the strongest predictors of class membership
2. Thresholds for ADT per lane were found to fall between 20,000 – 25,000 and link length was found to fall close to 2 miles.

3. Access point density thresholds were generally around 2.5 points / mile.
4. Chokepoint was coded as an indicator variable but did not significantly affect the performance of either model. It was dropped from consideration.

Proposed Classification Model

Using the results of the model exploration as a starting point, the following classification model was created. The proposed model uses only ADT per Lane, Access Point Density, and Link Length as inputs. Links are classified by totaling the number of points based on the following criteria:

- ADT per Lane >= 20,000 (1 point)
- Access point density >= 2.5 points per mile (1 point)
- Link length < 2 miles (1 point)

A link which scores 2 or more points can be classified as a "high" variance link. Otherwise the link is classified as "low" variance.

Model Performance

The proposed model was tested against the Houston data set and compared with the baseline model. Table A-7 compares the performance of the proposed model compared with the baseline model.

Table A-7. Model Performance

Model	Description	% Correct Classified	Kappa Statistic
Baseline	Vote for majority class	62%	0
Proposed Model	Decision tree: ADT, Access Point Density, and Link Length	67%	0.175

The results show that the proposed model was able to improve classification of links over the baseline model. In addition, the proposed model achieved a correct classification rate of 60% in the "high" variance class. Table A-8 shows the confusion matrix for the proposed model.

Table A-8. Confusion Matrix, Proposed Model

	Predicted Class	
Actual Class	High	Low
High	27	18
Low	30	42

Validation of Model

A small validation set was collected to test the proposed model's performance. The validation data was assembled from a sample of links tested by the University of Maryland for a validation study of private sector traveler information. Individual travel times of vehicles were measured using Bluetooth re-identification and data was aggregated following the same method (described earlier in this paper) used for the Houston data set. The distribution of CV travel time was computed for each link.

A larger validation data set will be compiled in the future and used to further test the proposed model. Three links were evaluated using the data from the University of Maryland study. Table A-9 shows attribute data for each link along with the 90th percentile CV travel time and the predicted class based on the model. The validation data shows that the model correctly identified these links as "high" variance based on the given attributes.

Table A-9. Validation Data

Link Description	Validation Date	ADT per Lane	Access Point Density	Link Length (mi.)	90th Percentile CV	Predicted Class
MD03-0004	April 2010	26,250	2.30	2.17	.101	High
PA01-0007	March 2010	21,166	1.92	1.56	.103	High
110+04178	February 2010	31,333	1.74	1.23	.102	High

Discussion of Findings

The model exploration resulted in a proposed model that classifies links based on ADT per lane, access point density, and link length. The threshold values for these attributes were found by testing different classification methods using a data mining software package. These values can also be considered in the context of traffic engineering. For example, links with ADT per lane >= 20,000 are more likely to experience volumes which approach capacity. The fact that the model identified short links with higher access point density as "high" variance links also concurs with what one would expect to find in the field. The validation of the model, while limited, indicates that thresholds for the parameters are appropriate. The fact that chokepoints were not incorporated in the model may be due to a problem in coding the parameter. Future research may develop a method for coding this information and incorporating this parameter in the model.

GUIDELINES SECTION 3 – TECHNICAL APPENDIX

Covers technical issues related to Section 3. Collect and Reduce Benchmark Data of the guidelines.

Comparison of Floating Car and AVI Travel Time Estimates

The floating car (FC) method assumes the probe vehicle passes as many cars as it is itself passed. As such, the floating car's speed, etc. should behave like the median of the traffic group in which it is traveling. A quantitative test of this assumption proceeds from a comparison of floating car data to interval estimates derived from AVI data taken contemporaneously and co-located with the FC observations. Interval estimates of the mean and median should roughly "cover" the respective floating car values to the nominal confidence level if the FC observation is actually behaving like a mean or median. If not, then the question becomes whether FC values are behaving like aggregates of observations or are no more privileged than a single random value.

Houston AVI data for 32 segments were matched to an equivalent number of floating car runs. Any AVI matches whose start times fell within +/- 120 seconds of the floating car start times were included. As may be seen in Table 1, each floating car run was matched with between 6 and 94 AVI samples. Three interval estimates were constructed from the AVI data:

1) A 95% Confidence Interval for average (mean) travel time. This is a parametric (normality assumption) confidence interval.
2) A nonparametric (rank-based) 95% confidence interval for the median travel time. This interval is a more robust estimate of travel time since it does not rely on any distribution assumptions (symmetry, etc.).
3) A 95% Prediction Interval for a “new” travel time. Prediction intervals are used to estimate "new" observations based on information about the previous sample.

The first two intervals estimate average or median travel time. We expect the nonparametric interval for median travel time to be slightly “wider” than that for the parametric interval for the mean. While confidence and tolerance intervals estimate present population characteristics, the prediction interval estimates what future values will be, based upon the present sample. They are intervals which estimate likely values for individual observations, not aggregate statistics such as mean, median, etc.

Under the assumption that the AVI and floating car travel times constitute samples from the same respective geographic and temporal cohorts, then it should be possible to compare how well the AVI interval estimates “capture” or “cover” the observed floating car travel time. Since the floating car travel time is assumed to estimate median travel time, we should expect it to lay within the bounds of the 95% confidence interval for the median approximately 95% of the time.

Referring to Table A-10, the floating car travel time estimates are compared to the three AVI-based interval estimates. Interval estimates that fail to include the floating car estimates are highlighted in red. From the table, the 95% confidence intervals for the mean only manage to cover the floating car estimate 43.8% of the time, while the coverage of the 95% intervals for the median are slightly better than chance (62.5%). In contrast, the 95% prediction intervals provide the closest coverage to nominal (90.6%). This would seem to indicate that the floating car estimate is not behaving like a median travel time; rather, it is behaving more like a random vehicle from the traffic group. Were this to be the case,

then floating car estimates would necessarily be more variable and less consistent than AVI estimates, whose variability declines with larger sample size.

While the results here are necessarily limited in scope and may therefore constitute more qualitative than quantitative evidence, they nevertheless highlight a possible deficiency in floating car estimates as practiced. If a larger study repeats the initial findings here, then the practice of floating car speed estimation should be reassessed. Either many more runs must be performed, or the runs must be conducted to rigidly observe median speed behavior (e.g., keeping exact running counts of passes). Under current circumstances, AVI estimates can be seen to perform more favorably, in that they provide statistically justifiable estimates.

Table A-10. Coverage Rates of Confidence Intervals Based on AVI Data

AVI Sample Segments				Floating Car	95% Prediction Interval for New TT		95% Confidence Interval for Median TT		95% Confidence Interval for Mean TT	
Orig_ID	Dest_ID	Length	Samples	FC-TT	Lower	Upper	Lower	Upper	Lower	Upper
298	386	3.00	18	290	253	308	272	289	275	286
146	147	1.85	13	110	95	119	101	111	104	110
387	130	2.70	19	295	260	322	288	301	285	297
138	139	1.45	9	86	77	100	85	93	86	92
131	132	2.95	31	375	336	422	368	393	373	386
157	158	1.80	29	146	111	140	123	130	124	128
81	83	4.45	9	317	213	300	242	273	245	267
161	162	5.10	24	338	299	411	339	370	345	364
133	134	1.50	6	89	74	120	85	102	91	104
83	374	3.10	30	232	154	212	175	189	179	188
134	135	1.25	25	74	58	82	68	72	68	72
160	161	1.60	34	105	90	126	104	113	105	111
150	151	1.61	8	94	70	113	87	100	86	97
148	149	3.65	10	225	184	287	217	256	223	248
124	125	1.85	11	115	81	129	99	112	99	110
122	123	3.30	19	194	151	227	180	199	182	196
294	387	2.10	18	387	290	444	357	383	352	382
125	292	3.80	10	233	162	264	197	231	200	225
123	124	1.35	9	77	58	96	70	82	72	81
295	293	1.40	21	84	66	101	78	88	80	86
386	295	2.10	16	195	155	245	188	212	191	209
162	163	1.61	44	135	98	147	120	126	119	125
293	145	1.00	38	62	49	75	58	65	60	64
139	140	2.75	15	157	114	192	139	162	145	161
80	157	1.80	8	203	166	311	212	273	220	258
130	131	2.20	22	202	142	238	180	205	181	198
292	294	1.40	22	89	67	116	86	98	87	96
147	148	1.05	16	57	45	84	61	67	61	68
159	160	1.35	94	183	139	238	179	195	184	193
140	298	2.10	9	193	112	255	152	203	166	202
158	159	2.50	32	463	275	592	444	468	409	457
145	146	2.65	20	207	124	298	196	210	195	227
					COVERAGE:	90.6%	COVERAGE:	62.5%	COVERAGE:	43.8%

Bluetooth Spacing Guidelines

The following is an excerpt from a technical memo prepared by TTI for the Mobility Measurement Pooled Fund Study.

Bluetooth Reader and Antenna Placement

Bluetooth reader placement is somewhat dependent on whether the application is short-term data collection or permanent continuous data collection. For short-term data collection, Bluetooth readers can be deployed in a portable, weather-resistant case (e.g., Pelican case or similar) with a portable battery power source. The Bluetooth reads can be stored on a local processor/computer that is also contained within the weather-resistant case. This is the approach that was first taken by the University of Maryland in validating I-95 travel time data from INRIX (Figure A-3).

Figure A-3. Portable Bluetooth Reader in Weather-Resistant Case
(Source: Stan Young, University of Maryland)

For permanent continuous data collection, Bluetooth readers are most commonly installed in existing traffic signal systems cabinets (Figure A-4). This solution provides the most cost-effective solution, as the signal cabinets offer weather resistance, a power source, and in some cases, a real-time communications link. If a communications link is not available within the cabinet, a cellular modem can be used to communicate the Bluetooth reads to a central database in near-real-time.

Figure A-4. Permanent Bluetooth Reader in Traffic Signal Cabinet
(Source: Darryl Puckett, Texas Transportation Institute)

For permanent installations, most traffic signal cabinets are located in the immediate vicinity of a signalized intersection. With some signal systems, a midblock cabinet also may be provided for data collection for signal system operation. Placement of permanent Bluetooth readers in the immediate vicinity of major signalized intersections is optimal for motorist understanding of link travel times; however, this may be sub-optimal in terms of accurate measurement of individual intersection delay, particularly if several signalized intersections are located between adjacent Bluetooth readers.

If Bluetooth readers cannot be conveniently located within an existing traffic signal cabinet, one low-cost alternative being tested by TTI on other projects is the use of a smaller-sized, solar-powered cabinet that can be located on an existing sign or roadside structure. Another alternative for locations without a traffic signal cabinet is to install/construct a new cabinet or enclosure. However, new cabinet installation significantly decreases the cost advantages of Bluetooth readers.

<u>For both short-term and permanent Bluetooth readers, the antenna should be placed at vehicle windshield height or higher (at least three feet)</u>. This height is ideal so as to minimize obstructions between the antenna and the mobile Bluetooth device. The Bluetooth communication was designed to operate around obstructions, but extensive field experience has shown that Bluetooth read rates are consistently higher with a relatively clear line-of-sight.

For short-term installations, achieving this optimal height may not be possible with a portable case. In these situations, the portable Bluetooth case must be placed on the ground and locked to a sign support, guiderail, or other roadside structure. Some portable Bluetooth installations have included an antenna (contained within a PVC pipe) that can be placed at a more ideal height off the ground (Figure A-5). In these situations, the antenna should be positioned such that the structure it is attached to (e.g., signpost, utility pole, etc.) does not obstruct Bluetooth device reading in both directions of travel.

Figure A-5. Bluetooth Reader Antenna on a Portable Installation
(Source: Darryl Puckett, Texas Transportation Institute)

For permanent installations, antenna height can be achieved in at least two ways: 1) place a connecting antenna on top of the traffic signal cabinet (Figure A-6); 2) place a connecting antenna at a higher elevation on a utility pole or other roadside structure.

<u>In addition to antenna height, the antenna configuration (e.g., type, power level, etc.) is an important parameter that should be optimized at each installation</u>. To indicate the importance of this, TTI conducted a simple field test by placing two Bluetooth readers (call them Reader A and Reader B) with different antenna configurations on a 2-mile road segment with no intermediate entry or exit points. This field test had the following results:

- Antenna 1: 51% of the Bluetooth MAC addresses were read at both Reader A and Reader B.
- Antenna 2: 88% of the Bluetooth MAC addresses were read at both Reader A and Reader B.

It is beyond the scope of this document to indicate optimal antenna and power configurations for all possible situations. However, the importance of antenna configuration is emphasized to provide optimal Bluetooth reading and matching rates.

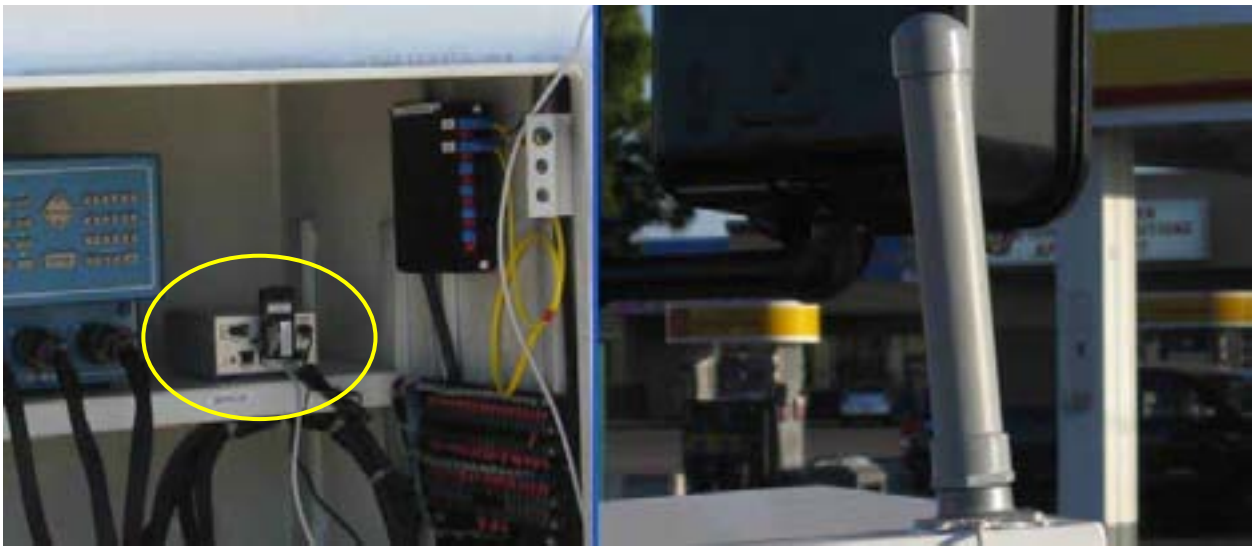

Figure A-6. Bluetooth Reader (left photo) and Antenna (right photo) on a Semi-Permanent Installation
(Source: Darryl Puckett, Texas Transportation Institute)

Bluetooth Reader Spacing

The spacing of Bluetooth readers varies based on the application, the roadway type, and the level of through traffic. The spacing is also influenced by the Bluetooth read radius, which for Class I Bluetooth devices is 100 meters or 328 feet. In other words, because the Bluetooth device could be read anywhere within the 100-meter radius, the spacing should be long enough to keep this location error tolerable as compared to the overall link length. For traveler information, the generally accepted spacing guidance is as follows (based on University of Maryland and TTI research):

- Freeways/expressways (controlled access): optimal spacing of 1-2 miles, maximum of 4-5 miles
- Major arterial streets: optimal spacing of ½-1 mile, maximum of 2-3 miles

The maximum expected error due to the location uncertainty of Bluetooth devices in the read radius can be calculated using standard time-distance calculations, by assuming two separate cases: 1) the Bluetooth device is read 100 meters before the first reader, but 100 meters after the second reader; 2) the Bluetooth device is read 100 meters after the first reader, but 100 meters before the second reader. Figures A-7, A-8, and A-9 illustrate the maximum expected error at different prevailing speeds. These figures are in concert with the generally accepted spacing ranges shown above. That is, as speeds get slower, the Bluetooth readers can be placed more closely together without appreciable increase in speed error.

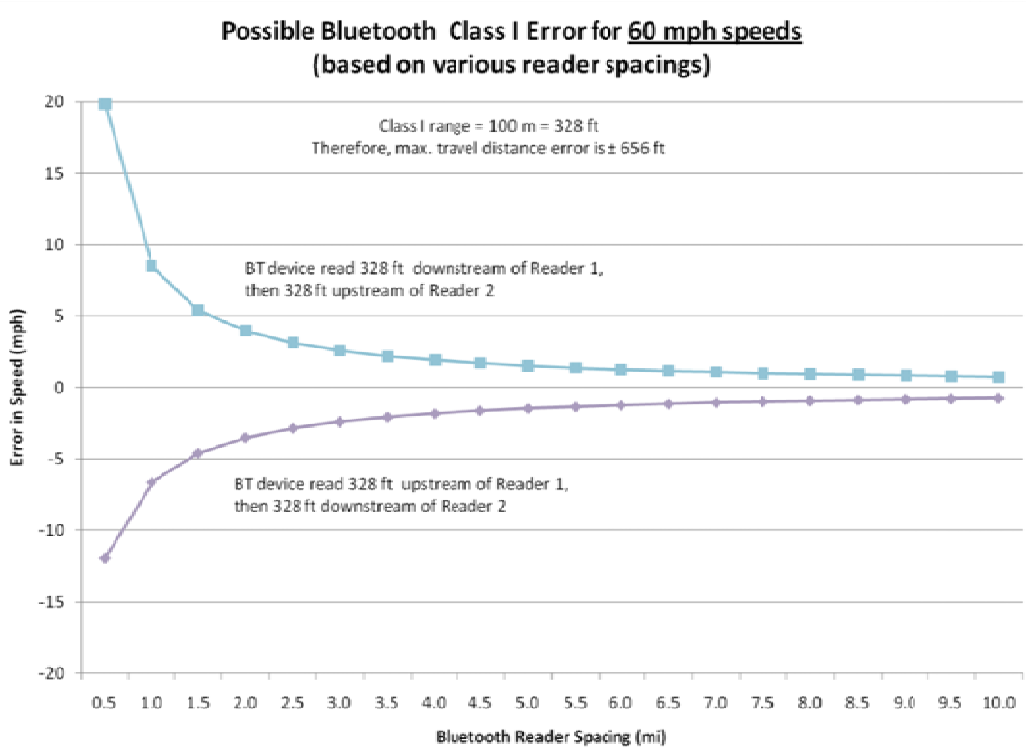

Figure A-7. Maximum Bluetooth-Based Speed Error at 60 mph Due to Location Uncertainty

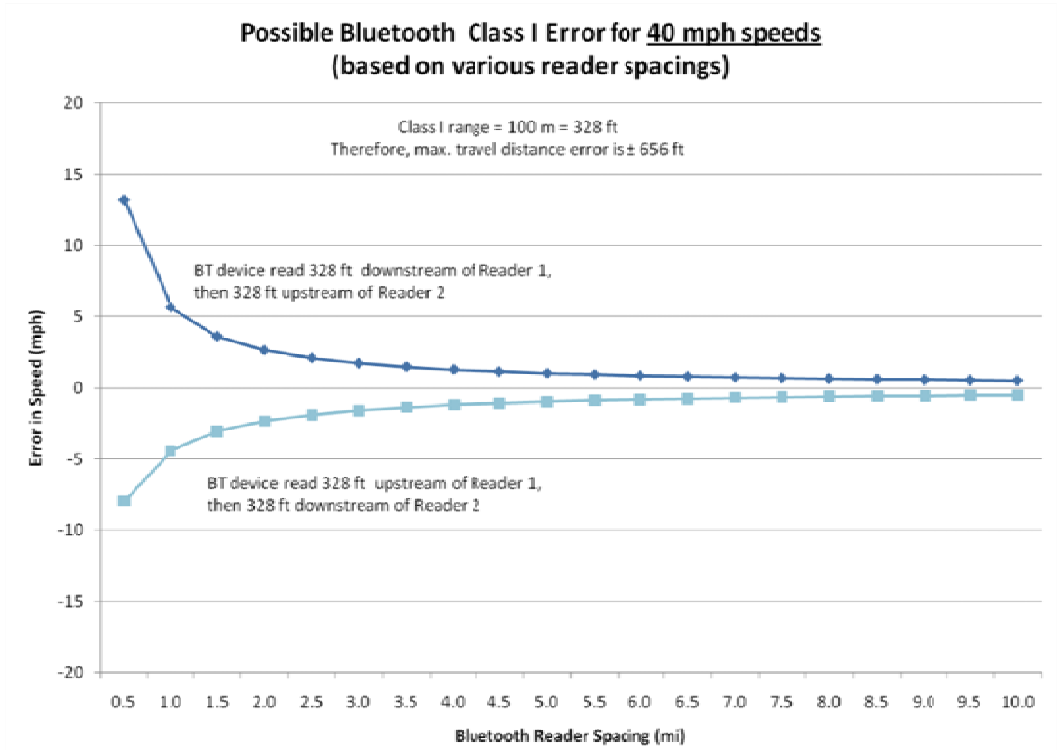

Figure A-8. Maximum Bluetooth-Based Speed Error at 40 mph Due to Location Uncertainty

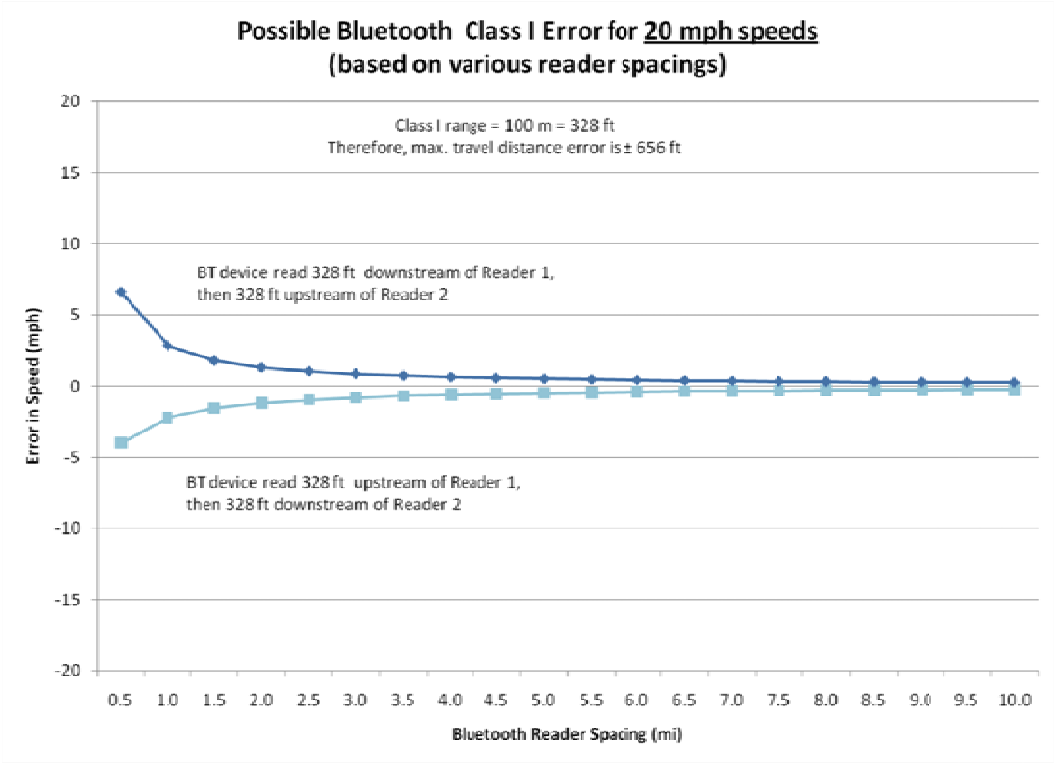

Figure A-9. Maximum Bluetooth-Based Speed Error at 20 mph Due to Location Uncertainty

To confirm that the generally accepted spacing for Bluetooth readers produced acceptable reading and matching results, TTI calculated matched sample sizes and match rates as part of this pooled fund study. TTI has been collecting Bluetooth travel time data for other research projects in Houston and College Station since 2009, and we used this data to analyze the results of the current spacing.

Table A-11 shows the summary results for this spacing and match rate analysis. The Bluetooth reader spacing used by TTI in Houston and College Station is producing match rates that are in the same range as other reported match rate results. For arterial streets in Houston where the link lengths range from 0.7 to 1.3 miles, the average match rate is about 5% of all through traffic, or 36 samples per hour. The results on College Station arterials (0.5 to 1.1 miles in length) are similar, with a 4% match rate of through traffic; however, on these lower volume arterials, that corresponds to about 17 samples per hour. For a single 2.2 mile freeway link in Houston, the results are similar at 3.4% match rate of all through traffic, but on a higher-volume freeway that means 185 samples per hour.

Table A-11. Summary of Bluetooth-Based Match Samples and Rates in Houston and College Station

City / Road Type	# links	Range in AADT	Range in Length	Hourly Valid Matched Samples		Hourly Valid Match Rates	
				Average	Daytime Avg.	Average	Daytime Avg.
Houston arterial streets (4-8 lanes)	20	10,000 to 30,000	0.7 to 1.3 mi	36	49	5.1%	4.8%
Houston freeways	2	115,000 to 120,000	2.2 mi	185	195	3.4%	2.8%
College Station arterial streets (4-6 lanes)	10	11,500 to 16,750	0.5 to 1.1 mi	17	21	3.8%	2.8%

Note: Match rates shown are for a single Bluetooth reader covering both directions of travel on each street.