

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



This document is distributed with a Creative Commons Attribution-Share Alike 3.0 License (see <http://creativecommons.org/licenses/by-sa/3.0/us/> for terms and conditions of this license). The recommended attribution for this document is as follows:

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

Report No.:	Report Date: June 2011	No. Pages: 76	Type Report: Final Contract	Project No.: TPF-5(200)
			Period Covered: August 2009 -- May 2011	Contract No.:
Title: GUIDELINES FOR EVALUATING THE ACCURACY OF TRAVEL TIME AND SPEED DATA				Key Words: accuracy, benchmark, data quality, evaluation, reidentification, speed, test vehicle, travel time, traveler information
Author(s): Shawn Turner, James Richardson, Mike Fontaine, and Brian Smith				
Performing Organization Name and Address: Virginia Center for Transportation Innovation and Research 530 Edgemont Road Charlottesville, VA 22903		University of Virginia Department of Civil and Environmental Engineering Thornton Hall Charlottesville, VA 22902 Texas Transportation Institute The Texas A&M University System College Station, TX 77843-3135		
Sponsoring Agencies' Name and Address: Virginia Department of Transportation (lead state DOT in pooled fund study) 1401 E. Broad Street Richmond, VA 23219				
Other sponsoring agencies in this pooled fund study include Alabama DOT, Caltrans, FHWA, Maryland SHA, Michigan DOT, and Pennsylvania DOT.				
Supplementary Notes:				
Abstract:				
<p>These <i>Guidelines</i> describe a process for evaluating the accuracy of traveler information in three forms: 1) congestion level (qualitative or color-coded); 2) average link speed; and, 3) average route travel time. Public agencies, private companies, or third-party evaluators can use these <i>Guidelines</i> to evaluate the accuracy of commercial traveler information services (TIS) prior to purchasing data or to ensure that contract requirements have been met. Also, public agencies and private companies can use these <i>Guidelines</i> internally to evaluate and improve the accuracy of their own TIS. These <i>Guidelines</i> define methods for collecting and processing benchmark data, but do not establish the level of accuracy that a traveler information service provider must achieve. These quality levels must be defined by the user or purchaser of the data based on their intended application of the data.</p> <p>These <i>Guidelines</i> outline and describe three main steps in the accuracy evaluation process: 1) develop an evaluation plan; 2) collect and reduce benchmark data; and 3) compute and report accuracy measures. An evaluation plan outlines basic parameters of an accuracy evaluation (e.g., accuracy measures, benchmark data collection method, the roads and time periods to evaluate, etc.). Once an evaluation plan has been developed, benchmark speed and travel time data are independently collected using reidentification (e.g., Bluetooth, license plate, etc.) or test vehicle (GPS, clipboard/stopwatch, etc.) methods. The benchmark is considered to be the best estimate of the true average travel time and speed for the defined roadway link/route and time period. Once benchmark data are summarized, accuracy measures are selected (based on the evaluation scenario), calculated, and reported. Finally, these <i>Guidelines</i> provide a list of resources and several examples to demonstrate the application of principles. A companion <i>Technical Appendix</i> documents the supporting analyses for these <i>Guidelines</i>.</p>				

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

Revision History

Version	Date	Description and/or Summary of Changes
0.1	5/17/2010	Work in progress sent to Study Sponsors
0.2	10/12/2010	First draft of Sections 1, 2, and 4 sent to review panel
0.3	1/21/2011	Complete draft of all sections sent to review panel
0.4	5/6/2011	Final formatted draft of all sections sent to review panel
1.0	6/7/2011	Complete and final version made publicly available at http://pooledfund.org/projectdetails.asp?id=426&status=4

Licensing and Attribution



This document is distributed with a Creative Commons Attribution-Share Alike 3.0 License (see <http://creativecommons.org/licenses/by-sa/3.0/us/> for terms and conditions of this license). The recommended attribution for this document is as follows:

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

Acknowledgments

The authors gratefully acknowledge the sponsorship and comments from the following agencies and persons:

- Alabama Department of Transportation: Emmanuel Oranika
- California Department of Transportation: Nick Compin and Joe Palen
- Federal Highway Administration: Jimmy Chu
- Maryland State Highway Administration: Subrat Mahapatra
- Michigan Department of Transportation: Collin Castle
- Pennsylvania Department of Transportation: Jay Sengoz
- Virginia Department of Transportation: Mike Fontaine

This document has adopted several basic principles identified in “Traffic Information Benchmarking Guidelines” as published by the North American Traffic Working Group (NATWG) at <http://www.itsa.org/natwg.html>. These principles include:

- Categorization of three evaluation scenarios for traveler information: 1) congestion categories (typically color-coded or qualitative labels; 2) link speeds; and, 3) route travel times.
- Evaluation of traveler information at three spatial units: 1) links (NATWG refers to TMC Locations, which is an industry-specific implementation of a link; 2) routes; and, 3) region or citywide (NATWG refers to this as markets).
- Specified features of GPS instrumentation for test vehicle benchmark data collection.
- Inclusion of NATWG-recommended accuracy measures (i.e., root mean square error and seconds per mile) into a broader set of measures.

The NATWG members and other external reviewers also have provided feedback that helped to improve this document. In particular, the following persons provided review comments on several drafts:

- Billy Bachman, Geostats
- Chris Scoffield, INRIX
- Corinne Bradley, NAVTEQ
- Greg Jordan, Skycomp
- J.D. Margulici, Novavia Solutions
- Terri Johnson, NAVTEQ

TABLE OF CONTENTS

1. INTRODUCTION	1
Background	1
Purpose and Scope.....	1
Summary and Overview of Evaluation Guidelines.....	2
Definition of Terms	2
Relation to other Documents.....	5
Before You Buy.....	5
2. DEVELOP AN EVALUATION PLAN	7
Overview of Evaluation Process.....	7
Typical Evaluation Scenarios	9
Roadway Segmentation	11
Time Intervals.....	12
Routes or Links to Evaluate.....	12
Network Stratification Guidelines.....	13
Time Periods to Evaluate	15
Benchmark Data Collection Method.....	16
Selecting Accuracy Measures.....	16
Frequency and Duration of Evaluations.....	17
Evaluation Objectivity and Transparency	18
3. COLLECT AND REDUCE BENCHMARK DATA	19
Selecting a Data Collection Method.....	19
Reidentification Data Collection Methods.....	21
Test Vehicle Data Collection Methods	28
Capping Benchmark Speeds to Correspond with Posted Speed Limits	34
4. COMPUTE AND REPORT ACCURACY MEASURES	35
Gather TIS Data to be Evaluated.....	35
Select Accuracy Measures.....	38
Establish Acceptable Accuracy Targets	39
Calculate and Report Accuracy Measures	40
5. RESOURCES AND EXAMPLE APPLICATIONS OF THE GUIDELINES	55
Resources	55
Overview of Example Applications	56
Example #1.....	57
Example #2.....	68
Example #3.....	71

LIST OF TABLES

Table 1 Characteristics of Benchmark Data Collection Methods	20
Table 2 Example of Reporting Reidentification Link Length Discrepancies: I-95, Maryland	22
Table 3 Example of Reporting Reidentification Link Length Discrepancies: I-35, Austin, Texas	22
Table 4 Table of t-Distribution Values for Typical Two-Sided Confidence Intervals and Sample Sizes	37
Table 5 Recommended Speed Ranges for Computing and Reporting Accuracy Statistics	38
Table 6 Example Calculations for Absolute Speed Error and Resulting Travel Time Error	40
Table 7 Benchmark Evaluation of Normal and Congested Traffic Flow.....	42
Table 8 Illustrative Example of Link Speed Accuracy Calculations.....	45
Table 9 Illustrative Example of Accuracy for Route Travel Time	48
Table 10 Time-Space Diagram to Illustrate Calculation of Path-Based Travel Time.....	50
Table 11 Example Field Data.....	52
Table 12 Example Field Data with Link Length Adjustments	52
Table 13 Sample Calculations for Multiple-Link Path-Based Travel Time and Speed.....	53
Table 14 Other Accuracy Measures	54
Table 15 Example of Segment Definitions from Dayton.....	60
Table 16 Example of Network Stratification	60
Table 17 Example of Floating Car Data Collection	63
Table 18 Example of Travel Time Data Collection	64
Table 19 Example of Data Reduction for Travel Time Samples	65
Table 20 Example of Speed Classifications	66
Table 21 Example of Error Calculations for Travel Times	67
Table 22 Example of Error Bias Calculation	67
Table 23 TMC Segments.....	69
Table 24 Example of Error Metric for Link Speed	71

LIST OF FIGURES

Figure 1 Example of Congestion Category-Based Traveler Information.....	9
Figure 2 Example of Link Speed-Based Traveler Information.....	10
Figure 3 Example of Route Travel Time-Based Traveler Information.....	10
Figure 4 Overview of the Accuracy Evaluation Process	8
Figure 5 Illustration of Node Offset at End of Links or Routes	11
Figure 6 Reidentification Examples: Identifying Outlier Values in Benchmark Data	25
Figure 7 Reidentification Example Showing Difficulty in Identifying Outlier Values When Traffic Speed Variability is High	26
Figure 8 Reidentification Example Showing Overly Aggressive Outlier Algorithm.....	26
Figure 9 Illustration of Overlap Between Benchmark Travel Time and Multiple TIS Intervals.....	34
Figure 10 Inconsistencies in Travel Time Error at Different Speeds for Absolute Speed Error	39
Figure 11 Illustrative Example of Categorical Accuracy Evaluation Results.....	42
Figure 12 Example of Error Distribution Chart Used to Visualize Individual Link Error.....	46
Figure 13 Illustrative Example of Accuracy for Route Travel Time	48
Figure 14 Dayton Study Area (www.buckeyetraffic.com)	58
Figure 15 Example of Segmentation in Dayton	59
Figure 16 Data Collection Time Periods in Dayton	62
Figure 17 Example of a TMC Segment Chain	68
Figure 18 Example of a Speed Confidence Interval	70

1. INTRODUCTION

Background

Traveler information is an established industry, with a wide variety of services and products offered by government agencies as well as private companies. A key part of this traveler information is real-time traffic updates in the form of congestion levels, traffic speeds, and/or travel times on major roads. This real-time traffic data can be gathered using different collection methods and technologies, each having their own advantages and disadvantages. Formal or informal evaluations of data quality are typically conducted before purchasing or releasing a new traveler information service, and later as an ongoing quality assurance process. However, even when these evaluations have been publicly disclosed, their evaluation methods and data quality measures are often inconsistent or may not allow for fair assessments.

It is difficult for government agencies or private companies to make fully informed decisions about purchasing or providing traveler information services, given the inconsistent data quality evaluations of a wide variety of products and services. Guidelines are needed to describe how to evaluate the accuracy of traveler information, so that evaluation procedures are fair and consistent and evaluation results are comparable. Guidelines are also needed to help government agencies provide accurate, reliable traveler information. These *Guidelines* are intended to address this need.

Purpose and Scope

These *Guidelines* describe a process for evaluating the accuracy of traveler information presented in three possible ways:

1. Congestion level (qualitative or color-coded);
2. Average link speed; and,
3. Average route travel time.

The most common uses for these *Guidelines* are as follows:

- Public agencies (or third-party evaluators) can use these *Guidelines* to evaluate the accuracy of commercial traveler information services prior to purchasing data or to ensure that contract requirements have been met.
- Private companies (or third-party evaluators) can use these *Guidelines* to evaluate the accuracy of commercial traveler information services provided by other companies.
- A commercial data provider or industry organization can use these *Guidelines* to certify that certain traveler information services meet a specified quality level.
- Both public agencies and private companies can use these *Guidelines* internally to evaluate and improve the accuracy of their own traveler information services.

These *Guidelines* are intended to be informative and not prescriptive. “Should” statements are used throughout to indicate typical practices, whereas “shall” or “must” statements indicate consensus-based practices that are absolutely necessary for consistency. The intent is to provide guidance and consistency on the evaluation parameters that matter most. These *Guidelines* define methods for

collecting and processing benchmark data, but do not establish the level of accuracy that a traveler information service provider must achieve. These quality levels must be defined by the user or purchaser of the data based on their intended application of the data.

Summary and Overview of Evaluation Guidelines

The process for evaluating the accuracy of traveler information consists of three major parts:

1. Develop an evaluation plan
2. Collect and reduce benchmark data
3. Compute and report accuracy measures

These three parts form the basis for the organization of these *Guidelines* as follows:

- **Section 2, Develop an Evaluation Plan**, describes how to make decisions about the basic parameters of an accuracy evaluation: the evaluation measures, benchmark data collection method, the roads and time periods to evaluate, and the frequency and duration of evaluations. Section 2 is a must-read if you want to get “started on the right foot.”
- **Section 3, Collect and Reduce Benchmark Data**, outlines methods to independently collect travel time and speed data using reidentification (e.g., Bluetooth, license plate matching, etc.) or test vehicle (GPS, clipboard/stopwatch, etc.) techniques to establish benchmark values. The benchmark is considered to be the best estimate of the true average travel time and speed for the defined roadway link/route and time period. This section also addresses important data reduction techniques for these methods as well.
- **Section 4, Compute and Report Accuracy Measures**, describes the measures of accuracy, their strengths and limitations, and their underlying calculation procedures. The accuracy is defined as the degree of agreement between the benchmark values and the traveler information values. These *Guidelines* define the calculation procedures for several measures of accuracy based on the type of traveler information being provided. These *Guidelines* also present several charts and graphics that can be used to consistently summarize and communicate the results of accuracy evaluations.
- **Section 5, Example Applications of the Guidelines**, provides a list of resources and three examples of how these *Guidelines* can be applied in practice.

Definition of Terms

Several important terms related to the evaluation of traveler information services are defined as follows.

Accuracy: A data quality attribute that measures the degree of agreement between a data value or set of values and a source (i.e., benchmark data) believed to be correct.

Benchmark: A data value or set of data values considered to be the standard against which other data values are compared. The term “benchmark” is preferred over the term “ground truth” because the true values can seldom be measured (i.e., measuring the true average travel time or speed of all vehicles over a section of road), only estimated with some level of confidence.

Coefficient of variation: A normalized measure of dispersion for sample data, expressed as a percentage. The coefficient of variation (CV) is defined as the ratio of the standard deviation of the sample data to the mean of the sample data.

Coverage: The degree to which data values in a sample accurately represent the whole of that which is to be measured. As a measure of traffic data quality, coverage is typically reported in absolute (miles or kilometers of coverage) or relative (percentage) terms.

Completeness: The degree to which data values are present in the attributes (e.g., volume and speed are attributes of traffic) that require them. Also referred to as availability, completeness is typically reported in relative terms (i.e., as a percentage).

Data quality: A general term that describes the suitability of data for the purpose which it was intended. Several attributes can be used to quantify data quality, including accuracy, coverage, and completeness.

Fixed-point sensor: A traffic measurement device that is used to record or estimate the average spot speed of all detected traffic at a designated point of interest. In some cases, spot speeds from fixed-point sensors are used to estimate link travel times for traveler information purposes. Inductive loop detectors and side-fire radar are examples of fixed-point sensors.

Link: A spatial unit of roadway with relatively consistent traffic characteristics. Roadway links are typically shorter than one to two miles in urban areas, and can be as short as connecting two adjacent ramps on an access controlled highway, or connecting two adjacent signalized intersections on an arterial street.

Must: A verb that, when used in these *Guidelines*, indicates a consensus-based practice that is absolutely necessary for consistency.

Reidentification: Any data collection method that relies on uniquely identifying a vehicle at a given location, then reidentifying that same unique vehicle at another location. License plate numbers, mobile device identifiers (e.g., Bluetooth media access control (MAC) addresses), toll tag identifier numbers, and magnetic vehicle signatures are unique attributes that are most commonly used for vehicle reidentification purposes.

Route: A longer spatial unit of roadway that combines two or more links. Routes can be defined for a single named roadway or for a combination of roadways that includes intersection turns and/or interchange ramps. Routes are most analogous to a person's trip, which could be five or more miles in urban areas, or the distance between cities for intercity routes.

Segment: A spatial unit of roadway that is designated only for the purposes of benchmark data collection for long routes. For example, to collect benchmark travel time data via reidentification on a 100-mile intercity route, it will probably be necessary to divide the 100-mile route into shorter segments for ease of data collection. Segments can be between three to ten miles in length in urban areas, but could be much longer in rural areas with very infrequent access points. Segments should have similar traffic conditions on their entire extent, such that traffic is relatively homogenous along its length. If reidentification methods are used to collect benchmark values, segments should be defined such that sample sizes are adequate.

Shall: A verb that, when used in these *Guidelines*, indicates a consensus-based practice that is absolutely necessary for consistency.

Should: A verb that, when used in these *Guidelines*, indicates a typical practice that is advisable but not required for consistency.

Space mean speed: The average speed of all vehicles traveling a defined link of roadway during a specified period of time. Also known as the harmonic mean speed or average link speed, the space mean speed is calculated using the average travel time and the link length (Equation 1).

$$\text{Equation 1} \quad \text{Space Mean Speed, } \bar{v}_{SMS} = \frac{\text{link length}}{\text{average travel time}} = \frac{l}{\frac{\sum t_i}{n}} = \frac{n \times l}{\sum t_i}$$

where l = length of roadway link (distance traveled)
 t_i = travel time of the i th vehicle
 n = number of travel time observations

Test vehicle: The method of data collection in which speed and travel time data are collected by an instrumented vehicle driving solely for the purpose of traffic evaluation. There are several different test vehicle driving styles, but the most common is to “float” with traffic in an attempt to match the average speed of the surrounding traffic flow.

Time mean speed: The average speed of all vehicles at a fixed point along a roadway during a specified period of time. For example, most fixed-point traffic sensors are configured to report a time mean speed at frequent time intervals. At this fixed point, the time mean speed is calculated as the arithmetic average of all vehicles for a specified period of time (Equation 2).

$$\text{Equation 2} \quad \text{Time mean Speed, } \bar{v}_{TMS} = \text{average spot speed} = \frac{\sum v_i}{n}$$

where v_i = spot speed of the i th vehicle
 n = number of speed observations

TMC (Traffic Message Channel) path: A spatial unit of roadway defined by the NAVTEQ-TeleAtlas consortium for the purposes of traveler information. A TMC path is a specific industry-defined instance of a roadway link.

Travel time: The time required to traverse a route between any two specified points of interest. For the purposes of traveler information, travel time can be measured directly (e.g., with reidentification) by a sample of vehicles traversing the route. In other cases, travel time is estimated from average spot speeds measured by one or more fixed-point sensors between the two points of interest.

Traveler information: Any trip-related information provided to a traveler or potential traveler, including but not limited to traffic conditions, road construction, crashes or incidents, the availability and conditions of public transportation, and the availability of parking. These *Guidelines* focus solely on traffic condition information, such as congestion levels, speeds, and travel times.

Relation to other Documents

There are several other documents closely related to these *Guidelines*.

- *Traffic Data Quality Measurement*¹: This September 2004 report issued by the Federal Highway Administration outlines five data quality attributes (accuracy is one of the five attributes) and suggests several measures for each data quality attribute. This 2004 report was the origin for the definition of accuracy and several of the accuracy measures presented in these *Guidelines*; however, this 2004 report does not address benchmark data collection to the degree contained in these *Guidelines*.
- *Traffic Information Benchmarking Guidelines*²: Developed by the North American Traffic Working Group (NATWG), an ad-hoc committee of ITS America. The *NATWG Guidelines* primarily address the definition and calculation of accuracy measures, and the data collection is currently oriented to a single GPS-equipped test vehicle. These *Guidelines* are an adaptation and extension of the *NATWG Guidelines*, and contain more information about planning and executing different types of benchmark data collection, particularly reidentification.
- *USDOT Section 1201 Requirements*³: Section 1201 of SAFETEA-LU required USDOT to establish a Real-Time System Management Information Program that provides the capability to monitor real-time traffic conditions on major highways in the United States. The final rule for this Program was issued November 8, 2010 in the Federal Register (i.e., 23 CFR Part 511) and included accuracy requirements for highway travel times. This final rule indicates that “*States shall develop the methods by which data quality can be ensured to the data consumers.*” These *Guidelines* can be used to fulfill this requirement, such that states can use the methods herein to document the accuracy of highway travel times required in 23 CFR Part 511.

Before You Buy

Several public sector reviewers suggested that these *Guidelines* include advice for those public agencies that have not yet procured private sector traveler information. While it is not the intent of this report to discuss this issue in detail, public sector agencies should carefully consider these factors before beginning data procurement:

- **Define applications and requirements.** How will the public agency use the procured data? Which specific applications? For example, will the procured data be used for just for situational awareness? Could the procured data also be used for traveler information, incident detection, traffic control, and/or performance monitoring? Not only will this be important to delineate in the RFP, but it would also help to set appropriate data quality targets for a performance-based contract.
- **Address data integration issues.** How will the procured data be merged and/or blended with other data sources? For example, does the procured data overlap with existing public agency

¹ Available at <http://ntl.bts.gov/lib/jpodocs/reports/te/14058.htm>, accessed June 2, 2011.

² Available at <http://www.itsa.org/natwg.html>, accessed June 2, 2011.

³ More information can be found at <http://ops.fhwa.dot.gov/travelinfo/about/rtsmip.htm>, accessed June 2, 2011.

sensor coverage? If there is overlap, will one data source be used as a primary input with the other source providing redundant confirmation of traffic conditions? Or will overlapping data sources be combined, and if so, what relative weights are given to each data source? Which data source(s) will be assumed to be correct if there are significant discrepancies or disagreements?

- **Address mapping and location referencing issues.** What mapping platform will be used, and how will data with different geospatial references be integrated? Is this scalable to a regional or statewide level? How will map updates from the private sector data provider be handled?
- **Consider funding stability and service continuity.** How stable are the funding sources for future continuity of data feeds? For example, if operating/maintenance budgets are reduced, will this private sector data service be any more susceptible to periodic outages or interruptions?

2. DEVELOP AN EVALUATION PLAN

Developing an evaluation plan is a critical first step that can save time and evaluation resources. An evaluation plan first defines the context and scope of the evaluation. For example, the context and scope answers the question, “What information is being evaluated?” Are color-coded speed maps being evaluated for accuracy? Are travel times on websites or dynamic message signs being evaluated? Are numeric values of speed being evaluated? Or do several different types of traveler information need to be evaluated?

By clearly defining the context and scope, various evaluation and data collection parameters (e.g., roadway network segmentation, accuracy measures, time periods for data collection, etc.) can then be better defined. In many cases, there is not a single best solution for these evaluation parameters—it depends on the context and scope of the evaluation.

This section describes the process for developing a plan to evaluate the accuracy of link speed and travel time data for a traveler information service (TIS). First, several common evaluation scenarios are presented to help illustrate the process of defining the context and scope of the evaluation.

Based on these evaluation scenarios, specific details are then provided for making decisions about the following evaluation parameters:

- Roadway segmentation
- Time intervals
- Routes or links to evaluate
- Time periods to evaluate
- Benchmark data collection method
- Accuracy measures
- Frequency and duration of evaluations
- Evaluation objectivity and transparency

Overview of Evaluation Process

While the sequence of steps for developing an evaluation plan is flexible and may be iterative, it is recommended that certain elements (Figure 4) be addressed. The order of the steps does not significantly affect the plan, but the order of steps presented in these *Guidelines* is intended to facilitate a logical development of the evaluation plan.



Figure 4 Overview of the Accuracy Evaluation Process

Typical Evaluation Scenarios

There are several common scenarios that arise when evaluating the accuracy of TIS-based link speed and travel time data. The most common scenarios are described here and will be referenced throughout this document to illustrate the process of logically defining evaluation parameters based on the context. These scenarios are not intended to limit the scope of evaluations to only these scenarios, but simply to illustrate the process of developing an evaluation plan for common examples.

The most basic evaluation scenario is a **congestion category** evaluation. In this scenario, the traveler information to be evaluated is provided in two or more categories, with each category representing a defined range of travel speed values. A common example is a color-coded speed map, with green-colored road links representing the fastest speeds, red-colored road links representing the slowest speeds, and other colors representing in-between ranges (Figure 1). The number of congestion categories and associated speed ranges should be consistent within an evaluation, but among evaluations, these can vary according to agency or company preferences.

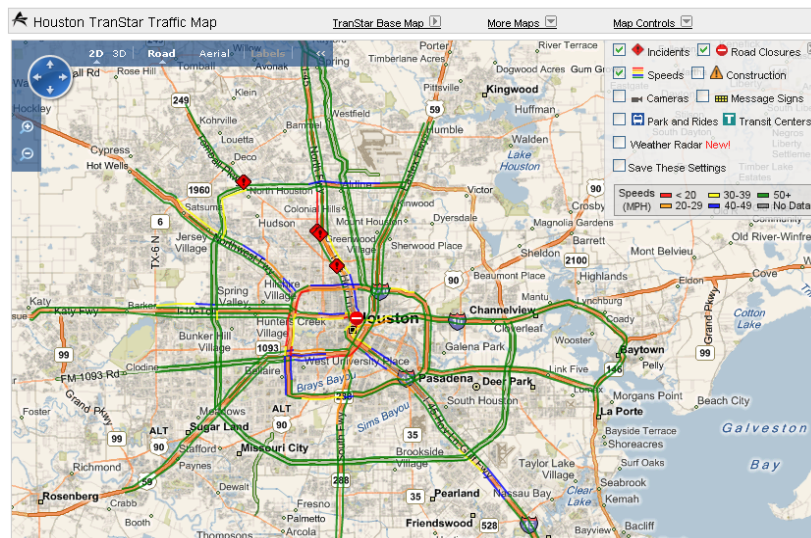


Figure 1 Example of Congestion Category-Based Traveler Information
(Source: <http://traffic.houstontranstar.org>)

The second evaluation scenario is a **link speed** evaluation. In this scenario, the traveler information is provided as numeric speed values for specified roadway links (Figure 2). Another common example is a real-time data feed that is provided by traffic information providers to distribution partners (for this example, each defined traffic message channel (TMC) path is considered a link). Another example is when numeric link speed values are provided in association (via a hyperlink or mouseover/hover box) with a color-coded speed map.

Mainlanes Eastbound						
From	To	Data Age	Distance (miles)	Link Travel Time(m:ss)	Cumulative Travel Time(m:ss)	Speed(mph)
Pin Oak	Grand Parkway	8:04:09 PM	2.10	1:48	1:48	70
Grand Parkway	Westgreen	8:03:28 PM	3.20	2:53	4:41	66
Westgreen	Greenhouse	8:04:09 PM	1.80	1:34	6:15	68
Greenhouse	SH-6	8:04:18 PM	2.90	2:33	8:48	68
SH-6	Eldridge	8:03:57 PM	2.20	1:59	10:47	66
Eldridge	Kidwood	8:04:22 PM	1.20	1:08	11:55	63
Kidwood	Belway S-West	8:04:13 PM	1.90	1:43	13:38	66
Belway S-West	Blalock	8:04:19 PM	2.30	2:05	15:43	66
Blalock	Bingle/Voss	8:04:27 PM	1.70	1:28	17:12	68
Bingle/Voss	IH-610 West Loop	8:04:06 PM	2.40	2:04	19:16	69
IH-610 West Loop	T. C. Jester	8:04:04 PM	2.30	2:08	21:24	64
T. C. Jester	Taylor	8:04:14 PM	2.25	1:57	23:21	69
Taylor	Main	8:02:42 PM	1.35	1:19	24:40	61
Summary			27.6	24:40	24:40	67

Figure 2 Example of Link Speed-Based Traveler Information

(Source: <http://traffic.houstontranstar.org>)

The third evaluation scenario is a **route travel time** evaluation. In this scenario, traveler information is provided as travel times (typically integer values) for specified travel routes (Figure 3). A route may traverse a single named roadway or multiple roadways, but it is always a collection of two or more links. For example, many state DOTs provide route travel times on dynamic message signs. Navigation devices or routing algorithms provide route travel times through a network that typically includes two or more named roadways.

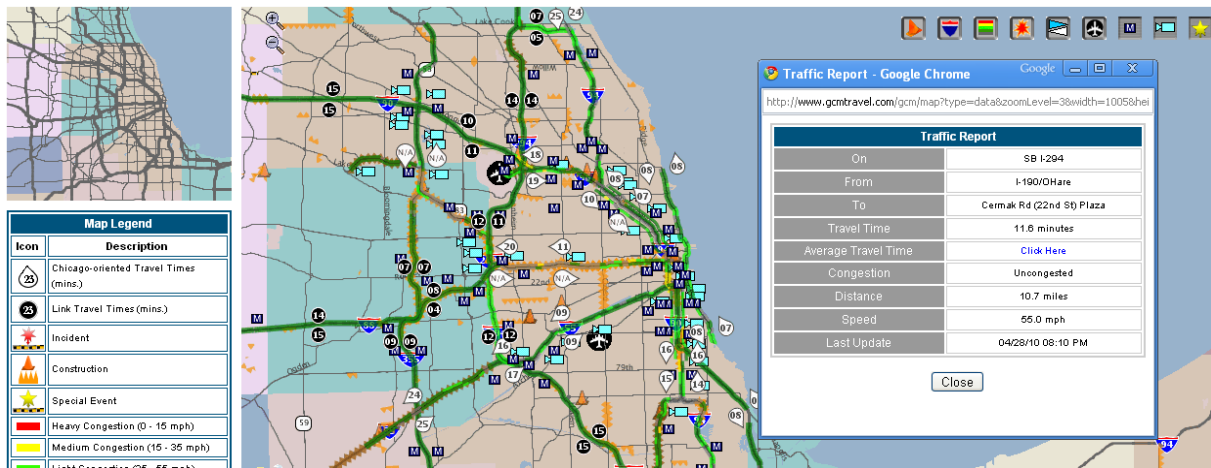


Figure 3 Example of Route Travel Time-Based Traveler Information

(Source: <http://www.gcmtravel.com>)

It is likely that some evaluations will include more than one evaluation scenario. For example, several public agencies provide **congestion categories** on a colored speed map as well as **route travel times** on dynamic message signs. In evaluations with more than one scenario, some of the evaluation parameters may have to be adjusted to account for multiple scenarios.

Roadway Segmentation

The roadway segmentation should be defined by the evaluation scenario. In other words, the same roadway segmentation used to report the speeds or travel times also should be used in collecting benchmark data and calculating accuracy measures. The exception is for TIS links that are shorter than $\frac{1}{2}$ mile, which should be combined with adjacent upstream and downstream links.

There are many variations of segmentation. Public agencies typically report speeds on links (shorter) and travel times on routes (multiple links). Commercial data providers typically report speeds on traffic message channel (TMC) paths (i.e., links) or travel times on defined routes. Regardless of the segmentation, the most important detail is that accurate location information be available for link or route endpoints.

If multiple scenarios are to be evaluated (e.g., link speeds and route travel times), the segmentation should be disaggregate enough to support all scenarios. In certain cases, the segmentation for different scenarios may not have common node points. For example, a link travel time node may be slightly offset from the nearest link speed node. In other cases, a slight adjustment to a node point may be necessary for benchmark data collection. If this node offset cannot be avoided, it may be necessary to estimate travel times or speeds based on extrapolation (see example in Figure 5). A “rule of thumb” is that the benchmark link length should not differ from the TIS link length by more than 10%. Section 3 contains more information on segmentation specific to each benchmark data collection method.

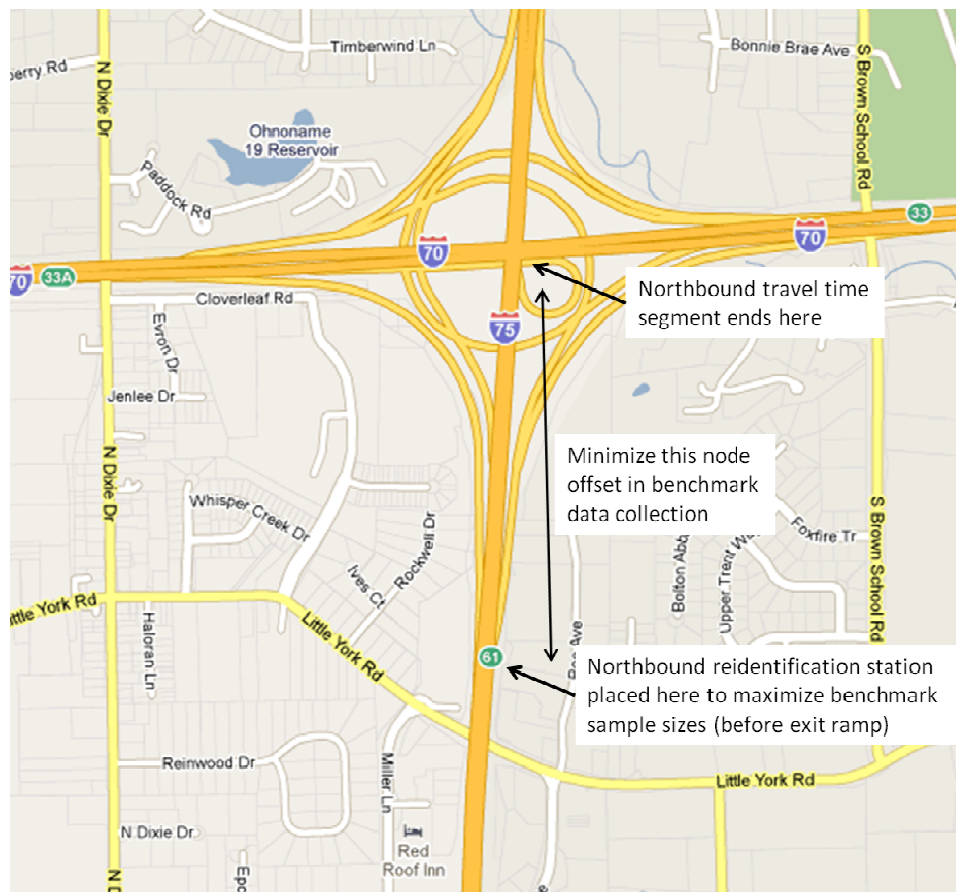


Figure 5 Illustration of Node Offset at End of Links or Routes

Time Intervals

The time intervals used to collect and compare benchmark data should be defined by the evaluation scenario. For example, how often are the travel times or speeds updated or refreshed? What time interval is associated with the reported travel time or speed? Common time intervals for reporting and evaluation are 1, 2, and 5 minutes.

Some travel time or speed data may be reported as “current conditions” without an associated time interval or update rate. In these cases, the data provider should be required to designate a time interval that the “current conditions” represents. If the time interval is not available, then it may be necessary to estimate the time interval from an approximate update rate or other discernable data process.

Travel time or speed data may represent a certain time interval but be updated on a different time interval. For example, a rolling 5-minute average could be updated every minute. In these cases, a 5-minute time interval should be used. However, the evaluators can decide whether to use a rolling 5-minute evaluation interval (e.g., 7:00—7:05, 7:01—7:06, 7:02—7:07, etc.) or a fixed 5-minute interval (e.g., 7:00—7:05, 7:05—7:10, 7:10—7:15, etc.). Regardless of the time interval, the benchmark values should be calculated to correspond to the exact same time interval as the TIS data.

The minimum time interval for analysis should be long enough to contain a sufficient number of samples for calculating benchmark speeds or travel times. A “rule of thumb” used by several different evaluators is a minimum of 5 to 10 individual vehicle samples per time interval. By increasing the time interval it is possible to increase the number of samples collected and thereby improve the statistical confidence in the benchmark value. However, by changing the time interval some problems may be encountered in determining the best method to aggregate the TIS estimates. In such cases, it is recommended that the estimate reported by the TIS provider be averaged over the new time period and compared with the benchmark value.

Routes or Links to Evaluate

Key Points to Consider

- Evaluations should include a number of routes and links with high variance in speeds or travel times as these are the areas that will be most challenging for a TIS provider or monitoring system to measure.
- Links should be classified by level of variance in speeds or travel times (high or low) before data collection begins. This will help determine an appropriate data collection method and sample size. This is covered in the section on Network Stratification.

The routes or links to be evaluated should be determined by the expected use of the data being evaluated. For example, the evaluation may focus on freeways, arterials, or both depending on how the data will be used. The evaluation should also specify geographic boundaries on the scope of the evaluation. Once these parameters are known, the following guidelines can be used to determine which subset of links should be subjected to a detailed evaluation, as well as the methods used to collect data on that link.

Generally speaking, it is desirable to prioritize evaluations to focus on the links where traffic is the most variable, as these will be the most difficult locations to monitor effectively. If traffic remains at free flow for most of the day, it is unlikely that there will be significant problems with travel time or speed estimates. If congestion forms and dissipates in a more unpredictable manner, then it will be much more difficult to estimate conditions accurately.

One way to measure data variability is with the coefficient of variation (CV), which is a normalized measure of dispersion for sample data, expressed as a percentage. The CV is defined as the ratio of the standard deviation of the sample data to the mean of the sample data (Equation 3).

$$\text{Equation 3} \quad CV = \frac{s}{\bar{x}}$$

where s = sample standard deviation
 \bar{x} = sample mean

For freeway links, typical CV values for travel time are in the range of 5% to 10%. For arterial street links, typical CV values can be much higher due to the influence of traffic signals, and are often in the range of 10% to 20%. Link length also has an impact on CV values, as longer links or routes typically have lower CV values.

Empirical experience with data collection in the field shows that certain types of links in the road network are more likely to experience these higher levels of travel time variation over longer periods of time than other links. Identifying these links and evaluating data accuracy on these links is a key part of a TIS evaluation.

Network Stratification Guidelines

Key Points to Consider

- Stratify the network by facility type into freeways, urban arterials, and non-urban arterials.
- Freeway links can be further stratified by classifying links according to the expected level of variance in speeds or travel times. A simple scoring system is suggested based on readily-available roadway inventory data.
- Data collection methods can be determined by the expected level of variance. Urban and non-urban arterials and freeway links with high levels of variance should be sampled using reidentification technologies. Non-urban arterials and freeway links with low levels of variance may be sampled using a test vehicle or reidentification.

Stratification by Facility Type

The road network should first be stratified by type of facility. A typical stratification by type of facility might result in two strata: freeways and arterials. It is also possible to refine the stratification by dividing arterials into urban and non-urban (i.e. suburban and rural) arterials. This would result in three strata to be evaluated: freeways, urban arterials, and non-urban arterials.

An urban arterial is characterized by higher signal density, greater land use intensity, and high access point density. A non-urban arterial is characterized by primarily suburban or rural land-uses with low

signal and access point density. Local expert knowledge may be used to identify arterials as being either urban or non-urban.

Stratification of Freeways

The freeway system should be further stratified to separately categorize higher-variance (with respect to travel time and speed) and lower-variance links. Research into travel time variation has found that freeway networks can be divided into two types of links for the purposes of a data quality evaluation. A higher-variance link is defined as any freeway link that has a relatively high probability of requiring large sample sizes to determine benchmark travel time. Lower variance links are the remaining links of the freeway network not identified as higher variance links.

Freeway Stratification Method

The freeway stratification method is based on a simple scoring method. The scores are determined by the following rules:

- 1 point – ADT per lane > 20,000 vehicles per day per lane
- 1 point – Access point density > 2 on/off ramps per mile
- 1 point – Link length < 2 miles

Each link can be scored based on these rules. **Links receiving 2 or more points are identified as higher-variance links** and should be sampled using a reidentification technology with proper experimental control as identified in Section 3 (i.e., Table 1). Links with only 1 or 0 points are lower-variance links and can be sampled by using any available sampling technology that directly measures travel time. For example, test vehicle methods may be used on lower variance links, although reidentification methods would likely produce larger sample sizes and more accurate estimates.

Freeway Stratification Examples

In this section, two hypothetical links are evaluated to determine the appropriate stratification for each segment. The segment ADT, number of lanes, link length, and access points are given.

Link A – Roadway Data

- ADT = 65,000
- Lanes = 3
- On / Off Ramps = 4
- Length = 1.8 miles

Link A – Scoring

- $ADT / Lane = 65,000 / 3 = 21,666 = 1 \text{ point}$
- $Access \text{ Point Density} = On/Off \text{ Ramps} / Length = 4 / 1.8 = 2.22 = 1 \text{ point}$
- $Link \text{ length} = 1.8 \text{ miles} = 1 \text{ points}$
- $Total \text{ Score} = 3 \text{ points, classify as higher-variance link and use reidentification}$

Link B – Roadway Data

- ADT = 45,000
- Lanes = 3
- On / Off Ramps = 5
- Link length = 2.25 miles

Link B – Scoring

- $ADT / Lane = 45,000 / 3 = 15,000 = 0$ points
- $Access\ Point\ Density = On/Off\ Ramps / Length = 5 / 2.25 = 2.22 = 1$ points
- Link length = 2.25 miles = 0 points
- Total Score = 1 points, classify as lower-variance link and use test vehicle or reidentification

Local Expert Knowledge / Bottlenecks

In addition to the stratification method shown in the previous section, local expert knowledge should also be used in identifying bottlenecks. Bottlenecks are points in a freeway network where capacity decreases when the bottleneck is activated, such as a reduction in traffic lanes, an entrance or exit ramp, or work zones. After the bottleneck activates, congestion starts at the point of the bottleneck and extends upstream of the bottleneck. Empirical data has shown that the links upstream of a bottleneck tend to have higher travel time variance. Due to the higher travel time variance, reidentification methods are recommended for sampling travel times on links upstream of a bottleneck. Local expert knowledge can be used to identify the location of bottlenecks.

Time Periods to Evaluate

There are four general time periods that should be considered for evaluation. These periods are:

1. Morning Peak (6 a.m. to 10 a.m.)
2. Midday (10 a.m. to 4 p.m.)
3. Evening Peak (4 p.m. to 8 p.m.)
4. Overnight (8 p.m. to 6 a.m.)

The times in parentheses are typical values, and it is recommended that links selected for evaluation be sampled during time periods representative of peak travel conditions for each specific link. Sampling during other time periods, such as non-holiday weekends, is also recommended if resources permit. If resources are constrained, evaluations should focus on the peak periods. Off-peak evaluations serve primarily to determine whether the TIS is providing valid estimates during periods of low volume when data points may be sparse.

Non-recurring events (e.g., incidents, special events) are typically the most challenging times for accurate traveler information. Time periods should be selected to include those times during which non-recurring events are most likely to disrupt normal traffic patterns (typically the morning and evening peak periods). In practice, it is difficult to predict when non-recurring events will occur. Automated reidentification methods for benchmark data collection have an advantage in this regard, as they can be positioned to collect data for several days at a time, thereby increasing the probability of capturing non-recurring events in the evaluation.

Benchmark Data Collection Method

The selection of a data collection method depends on many factors: available funds and equipment, staff expertise, required sample sizes, and many of the evaluation parameters described earlier in this section. Because the selection of a benchmark data collection method involves numerous considerations, this process is discussed in depth in Section 3.

The two basic types of benchmark data collection are reidentification and test vehicle. Reidentification methods consist of identifying a unique vehicle attribute, then reidentifying that same unique vehicle attribute at another point on the roadway or roadway network at a later time. The most common vehicle attributes used for reidentification are media access control (MAC) addresses from Bluetooth-enabled devices, license plates, or electronic toll tag identifiers. Time-lapse aerial photography also has been used to evaluate the accuracy of traveler information. Reidentification techniques rely on vehicles that are already in the traffic stream for purposes other than evaluation.

Test vehicle methods are those in which a specially designated vehicle is driven in the traffic stream for the sole purpose of collecting benchmark values. Floating car travel time runs are the most common test vehicle approach. These methods most commonly include a GPS device within the test vehicle to automatically record time and location at periodic intervals. Other equipment that can be used to record time and location is a distance-measuring instrument (DMI) or simply a clipboard and stopwatch.

Another evaluation approach sometimes used is a comparison with readily-available data, such as from fixed-point sensors or pre-existing but limited data collection. These *Guidelines* do not endorse this approach as a legitimate method to establish accuracy. Because the accuracy of these types of data sources are not demonstrably better than the link speed and/or travel time data being evaluated, these data sources are not and should not be considered a benchmark. Most often they are used simply for comparison purposes in less rigorous evaluations.

Selecting Accuracy Measures

Accuracy measures quantify the degree of agreement between a data value (or set of values) and a source (i.e., benchmark data) believed to be correct. The selection of accuracy measures should be based on the evaluation scenario(s).

There is no single best accuracy measure for all possible evaluation scenarios. Section 4 of these *Guidelines* describes the calculation procedures for several accuracy measures. Each of the accuracy measures defined in this section has certain strengths and/or limitations.

From past experience, these accuracy measures are commonly used for each evaluation scenario:

- **Congestion category**
 - Percent of correct category classifications (with a specified boundary tolerance), with the option of reporting the average deviation of incorrect classifications
- **Link speed**
 - Root mean square error, or RMSE (units of mph)
 - Average absolute error (units of mph)
 - Average error, also known as bias or signed error (units of mph)

- Percent of TIS values within X of the benchmark value, where X is a specified error tolerance in mph, percentage, or statistical confidence interval (units of %)
- **Route travel time**
 - Average absolute percent error (units of %)
 - Average absolute error per unit length (units of seconds per mile)
 - Average absolute error (units of minutes)

Accuracy measures other than these preferred measures are described in Section 4, but users are cautioned to understand their strengths and limitations before using them in an accuracy evaluation.

In addition to accuracy, timeliness (i.e., lag) is another important data quality attribute for traveler information services.⁴ However, the calculation and evaluation of timeliness is not addressed in these *Guidelines*, even though timeliness can be calculated with the same benchmark data used to calculate accuracy. It should be recognized that the accuracy measures in these *Guidelines* will indirectly capture timeliness problems. For example, if the timeliness of traveler information lags by 5 to 10 minutes, then the accuracy will be lower during periods of quickly changing traffic conditions. However, to directly quantify timeliness, one must calculate the actual time units by which the reported traveler information lags the benchmark values.

Frequency and Duration of Evaluations

A data quality evaluation is much like a point estimate. The evaluation results indicate the data quality at a point in time, but these results might be different if the evaluation is repeated. The question is: how often should an evaluation be repeated and over how long of a period should an evaluation be conducted?

The question of how long an evaluation should last is perhaps the simpler question. The limiting factor in the duration of an evaluation will generally be cost. The trade-off is between the cost of a longer evaluation period and the number of samples that can be collected. A one day evaluation period may be insufficient to sample from each link of the network. Therefore, enough time should be allocated to ensure that samples from all links of interest on the network can be gathered for the evaluation.

The question of how often to repeat the evaluation is more complex. This is a question of business requirements as well as changes in the network (e.g. new links) or changes in the local travel patterns (e.g. increased population, new job center built, etc.). If there are significant changes to the network such as increased capacity at an interchange or a loss of capacity due to a construction project it may be useful to re-evaluate data quality. If changes in the network occur and data quality worsens it may indicate that the data source is not robust to network changes. This may be an important consideration in an evaluation. Likewise, if significant shifts in travel behavior occur the distribution of travel time variation may change in the network. A re-evaluation of data quality may be necessary in either of these cases. These major changes may also necessitate a change to the links that are to be examined in recurring evaluations.

⁴ See recommended traffic data quality measures at <http://ntl.bts.gov/lib/jpodocs/reports/14058.htm>, accessed June 2, 2011.

Evaluation Objectivity and Transparency

It is important to maintain an objective and transparent evaluation process, especially for performance-based data service contracts. Even the perception of impropriety could call into question the evaluation results. For this reason, the following principles are recommended:

- An independent third-party is typically used to conduct evaluations and report results. Private consultants and university research groups have been the most common third-party evaluators to date.
- In data service contracts, the evaluator is typically hired by the purchaser and not the data provider.
- To the extent permissible through contractual agreements, the entire evaluation process should be documented and made available to all involved parties (including the respective data providers). This includes the data collection and reduction methodology as well as the raw and summary benchmark and TIS data.

Some commercial data providers regularly conduct internal evaluations of their data as part of a quality assurance process. These internal evaluations can be valuable for product development and refinement; however, they are sometimes viewed with skepticism by potential customers (both private and public). The use of a third-party evaluator can help to alleviate concerns; however, the perception of bias may still remain if the evaluator is hired directly and exclusively by the data provider.

The most objective and transparent approach to date has been the use of third-party evaluators hired by the purchaser, with the entire evaluation method and results publicly documented.

3. COLLECT AND REDUCE BENCHMARK DATA

This section contains guidance and information on collecting and reducing benchmark data. As defined in Section 1, the benchmark data are the definitive set of link speed and travel time values considered to be the standard against which TIS values will be compared. Because benchmark data are considered to be the definitive standard, extraordinary care should be used in its collection, reduction, and reporting.

Selecting a Data Collection Method

The two basic methods for benchmark data collection are reidentification and test vehicle. The selection of one of these methods depends on many factors: available funds and equipment, staff expertise, required sample sizes, and other considerations. This section describes the advantages and disadvantages for each method. Using this information, an agency can make an informed decision based on its objectives and available resources. In many cases, there is a fixed budget for evaluation and the goal is to perform the best possible evaluation with the available budget.

Reidentification methods consist of identifying a unique vehicle attribute, then reidentifying that same unique vehicle attribute at another point on the roadway or roadway network at a later time. The most common vehicle attributes used for reidentification are media access control (MAC) addresses from Bluetooth-enabled devices, license plates, or electronic toll tag identifiers. Time-lapse aerial photography also has been used to evaluate the accuracy of traveler information. Reidentification methods rely on vehicles that are already in the traffic stream for purposes other than evaluation.

Test vehicle methods are those in which a specially designated vehicle is driven in the traffic stream for the sole purpose of collecting benchmark values. The “floating car” style of driving is the most common test vehicle approach. These methods most commonly include a GPS device within the test vehicle to record locations at periodic time intervals. Other equipment that can be used to record time and location is a distance-measuring instrument (DMI) or simply a clipboard and stopwatch.

The test vehicle method has sometimes been referred to as the floating car method. However, the term “test vehicle” is preferred for several reasons: 1) there are several different driving styles for test vehicle data collection, and “floating car” is one particular driving style; 2) some traffic professionals outside of North America use the term “floating car” to mean something else (what we in North America call a probe vehicle); 3) test vehicle is a more descriptive term (it communicates that the vehicle is part of the “test” or evaluation).

Table 1 compares the characteristics of these two benchmark data collection methods. The test vehicle method has been used for traffic data collection for more than 50 years, and the supporting tools have evolved from a clipboard and stopwatch to sub-meter accuracy GPS with portable data storage and automated map-matching algorithms. Prior to 2008, the test vehicle method had been used by the private and public sector to evaluate traveler information. Evaluators using test vehicles are able to cover an entire city’s major roadway mileage in a few days, albeit at a “shallow depth” of one vehicle sample per time interval. The low number of test vehicle samples is thought to be offset by the ability of a single trained driver to replicate the prevailing or average speed of traffic. Based on analyses conducted for this project (see *Technical Appendix* accompanying these *Guidelines*), however, this assumption is not always valid.

Since 2008, the use of Bluetooth technology has significantly improved the cost and practicality of using reidentification for accuracy evaluations. Reidentification via Bluetooth typically provides many more samples per time interval than test vehicle, and permits a statistical confidence interval to be used to indicate the certainty of the benchmark values. Analyses conducted for this project indicate a cost per data point ratio for Bluetooth reidentification that is at least several hundred times lower than test vehicle data collection for long-term evaluation activities. In other words, for each dollar spent for data collection, Bluetooth reidentification has the capability to collect at least several hundred times more benchmark data points than test vehicle methods.

Bluetooth reidentification does have some limitations related to breadth of coverage and positioning of portable equipment. Additionally, care must be used to filter outlier values from the valid benchmark data. Despite these minor limitations, Bluetooth reidentification is increasingly being used as a very cost-effective method of traffic monitoring and evaluating other traffic data sources.

Given the overall benefits for Bluetooth reidentification, test vehicle methods are now considered applicable only for very short-term data collection, initial screening, or quick “spot checks” of traveler information. Conversely, Bluetooth reidentification is considered the accepted practice for statistically-valid travel time and speed accuracy evaluations.

Table 1 Characteristics of Benchmark Data Collection Methods

Characteristics	Reidentification	Test Vehicle
Depth of coverage	<ul style="list-style-type: none"> Typically provides many more samples per time interval than test vehicle on high-volume roads and times. 	<ul style="list-style-type: none"> Typically provides very limited samples.
Breadth of coverage	<ul style="list-style-type: none"> Breadth of coverage limited to where equipment can be positioned. 	<ul style="list-style-type: none"> Breadth of coverage much greater than reidentification, but at the cost of depth of coverage. Not limited by positioning of equipment.
Technology maturity/adoption	<ul style="list-style-type: none"> Depends on technology used for reidentification. Bluetooth is a mature communication standard but relatively new for traffic monitoring applications. 	<ul style="list-style-type: none"> GPS and DMI technologies mature and widely used in transportation applications.
Overall experimental control	<ul style="list-style-type: none"> Must establish outlier filtering to remove vehicles that deviate from the designated link/route. Also must filter slower modes (buses, bikes, etc.) and vehicles with multiple passengers (e.g., transit vehicles). Should ensure adequate sensor spacing to minimize error. 	<ul style="list-style-type: none"> Must establish driving protocol for test vehicles and enforce its standard application.
Cost ranges	<ul style="list-style-type: none"> Dependent on reidentification technology. Bluetooth devices can provide at least several hundred more data points per data collection dollar than test vehicle. 	<ul style="list-style-type: none"> Cost is usually the constraining factor, and typically limits test vehicle runs to a statistically inadequate sample size.

Reidentification Data Collection Methods

Reidentification Technology

There are several possible technologies that can be used to identify and reidentify unique individual vehicles in the traffic stream. In early 2011, Bluetooth-enabled devices are one of the most cost-effective technologies that can be used for reidentification. Other technologies include license plate matching (automated using optical character recognition or semi-automated using portable computers with manual plate entry), electronic toll tag identifier matching, vehicle magnetic signature matching, and time-lapse aerial photography.

Because of rapidly changing technology, these *Guidelines* will remain technology-neutral and provide general guidance that is applicable to all reidentification technologies. It is a near certainty that a yet-to-be-invented (or widely implemented) technology will someday join or replace Bluetooth as a cost-effective reidentification technology. By keeping these *Guidelines* as technology-neutral as possible, the longevity and applicability of the guidelines are better ensured.

Reidentification Node Locations

The location of reidentification nodes (i.e., the points at which vehicles are uniquely identified) will be based on the roadway segmentation defined by the evaluation scenario. For example, to evaluate travel times on Interstate 10 between First and Tenth Street, reidentification nodes should be placed as close as possible to the interchange centerline of Interstate 10 at First Street, as well as Interstate 10 at Tenth Street.

A perfect time and location match is an ideal goal for accuracy evaluations; however, in practice, it is sometimes difficult or impractical to place the reidentification technology at the exact same location as the TIS link endpoint (for example, see Figure 5). In these cases, best judgment should be used to position the reidentification technology as close as possible to the link endpoint. A general “rule of thumb” is that the actual reidentification link length should not differ from the TIS link length by more than 10%. Additionally, the evaluation should:

1. Clearly report where and by how much the reidentification link length differs from the TIS link length; and,
2. Make adjustments to the travel time (or use average speed) to account for differences in length between the reidentification link and the TIS link.

For example, evaluations of traveler information by the I-95 Corridor Coalition show differences in the TIS link (TMC path) length and benchmark link (Bluetooth) length (Table 2). Similar length difference reporting has been used by TTI for private sector evaluations on I-35 in Austin, Texas (Table 3).

Table 2 Example of Reporting Reidentification Link Length Discrepancies: I-95, Maryland
(Source: I-95 Corridor Coalition)

Validation Segment	TMC Paths Included					Length (mi)		Length Difference
	1	2	3	4	5	TMC	Bluetooth	
MD04-0001	110+04340	110P04340	110+04341	110P04341	110+04342	2.1	2.05	-3.6%
MD04-0002	110P04342	110+04343	110P04343			1.5	1.68	+8.8%
MD04-0003	110N04343	110-04342	110N04342			1.5	1.45	-6.0%
MD04-0004	110-04341	110N04341	110-04340	110-04339		2.1	2.11	+1.8%
110-04633	110-04633					1.5	1.44	-5.7%

Table 3 Example of Reporting Reidentification Link Length Discrepancies: I-35, Austin, Texas
(Source: Texas Transportation Institute)

Bluetooth Segment	Bluetooth Segment Length	TMC Path Sequence (Length in parentheses)
IH35 NB: Stassney to Woodward	2.11 miles (-9.4% shorter than TMC)	112+05036 (0.33 mi), 112P05036 (0.01 mi), 112+05037 (1.19 mi), 112+05038 (0.21 mi), 112P05038 (0.10 mi), 112+05039 (0.49 mi) = 2.33 miles total length
IH35 NB: Woodward to N. Lady Bird Lake	2.31 miles (-1.7% shorter than TMC)	112+05040 (0.72 mi), 112P05040 (0.12 mi), 112+05041 (0.53 mi), 112+05042 (0.12 mi), 112P05042 (0.86 mi) = 2.35 miles total length

If multiple scenarios are to be evaluated (e.g., link speeds and route travel times), the segmentation should be disaggregate enough to support all scenarios. In certain cases, the segmentation for different scenarios may not have common node points. For example, a travel time link node may be slightly offset from the nearest link speed node.

Minimizing Error and Uncertainty in Uniquely Identifying Vehicles

Measurement error must be minimized in the collection of benchmark data, since it will be the standard against which TIS data are evaluated. To minimize error and uncertainty, evaluators should understand basic working principles of the vehicle identification technology and possible sources of error.

This section reviews possible sources of error for the most common vehicle identification technologies currently available, and provides suggestions for ways to minimize these errors in benchmark data. The errors associated with uniquely identifying vehicles can be broadly classified into two categories: identification error and location error. Errors associated with matching will be discussed in the next section.

Bluetooth matching – An alphanumeric identifier (i.e., MAC address) is transmitted via the Bluetooth wireless communication standard, and the MAC address is either read correctly or there is no MAC address read. So when a MAC address has been read, one can be certain that a Bluetooth-enabled device is within range (note that the converse is not true – there may be cases when a Bluetooth-enabled device is in range but is not detected by the Bluetooth reader). However, the range for Class I Bluetooth devices is at least 100 meters, so this range can introduce location error. For example, the Bluetooth reader may be in a signal cabinet in the northwest quadrant of an intersection, but

reading Bluetooth devices in vehicles at all four intersection approaches. In this case, a matched northbound travel time could include delay experienced by northbound traffic waiting at the southernmost through lane approach, or by Bluetooth devices waiting in turn lanes on the east-west street. To minimize the effects of location error with Bluetooth matching:

- Use selective Bluetooth reader positioning, antenna shielding, and/or lower gain antennas (i.e., shorter range) to focus on the appropriate areas of detection. Exclude certain lanes that are not representative of prevailing traffic at that location (to correspond with TIS reports). Examples may include carpool, express, or managed lanes, as well as intensive weaving areas and queues in intersection turn lanes.
- Space Bluetooth readers far enough apart so that location error does not significantly affect overall travel times, but not so far apart that there is very little through traffic that would be detected by readers on either end of the section. For urban freeways and expressways, 1 to 2 miles is the optimal spacing and 4 to 5 miles is the maximum spacing. For urban arterials, optimal spacing is of ½ to 1 mile with a maximum spacing of 2 to 3 miles. Spacing is shorter on arterials due to the higher density of access points, which can cause sample sizes to quickly decline. Spacing on rural highways can be much longer (greater than 10 to 15 miles), particularly on routes with high levels of through traffic and few intermediate access points.
- The processing software must also deal with duplicate MAC address reads (i.e., the vehicle is within range of the Bluetooth reader during multiple scan cycles), as well as the data scan and delivery delay.

Electronic toll tag matching – The errors with electronic toll tags are similar to Bluetooth-enabled devices, except that toll tags typically have a much shorter and focused read range. Therefore, if a toll tag identifier has been read, it is a near certainty that the identifier has been read correctly. Like Bluetooth, there is a small possibility that a toll tag is within range but not detected. This would not produce erroneous data, but would simply result in fewer tag reads and/or matches. Because of the shorter read range, toll tags are not as susceptible to location error as Bluetooth-enabled devices. Some tag reader antennas are omni-directional and could be configured to read from multiple lanes. To minimize the effects of location error with toll tag matching:

- Exclude certain lanes that are not representative of prevailing traffic at that location (to correspond with TIS reports). Examples may include carpool, express, or managed lanes, as well as intensive weaving areas and queues in intersection turn lanes.

License plate matching – Identification error is the main concern for license plate matching, in the form of incorrectly recognizing the alphanumeric license plate characters. If plates are being read automatically using machine vision, there will be errors in the optical character recognition process (e.g., a “T” being read as an “I”). If plates are read and transcribed manually by human observers, then there are several points at which errors could occur. In many cases, an identification error at one node never gets matched and so never results in an erroneous matched travel time. Occasionally, though, an incorrectly-read plate may get matched with a correctly-read plate, resulting in an erroneous matched travel time. These erroneous matches can usually be filtered out with the outlier algorithms discussed in the next section. The following should be considered to minimize the effects of identification error with license plate matching:

- If human observers are used, have them read and transcribe the first four plate characters on the plate. Four characters are easier to collect than six or more characters, and those first four characters are often unique during the data collection time period.
- Collect the plate characters while the vehicle is approaching. This technique allows the observer to record the plate number at the instant the vehicle crosses the designated checkpoint. Headlight glare during low-light periods may require collecting plate characters while the vehicle is traveling away from the observer.
- Collect as many plates as possible, as the probability of matching plates increases for each additional plate entered.
- Focus on a representative sampling of through traffic lanes. Auxiliary, weaving, and turn lanes should be avoided because of the high percentages of turning traffic.

Minimizing Error in Reidentification Matches

Even when identification and location errors have been minimized at the reader nodes, there are still several situations that may produce erroneous data when the vehicle identifiers are matched between adjacent nodes. This section describes how to minimize the errors that can occur with travel time matches.

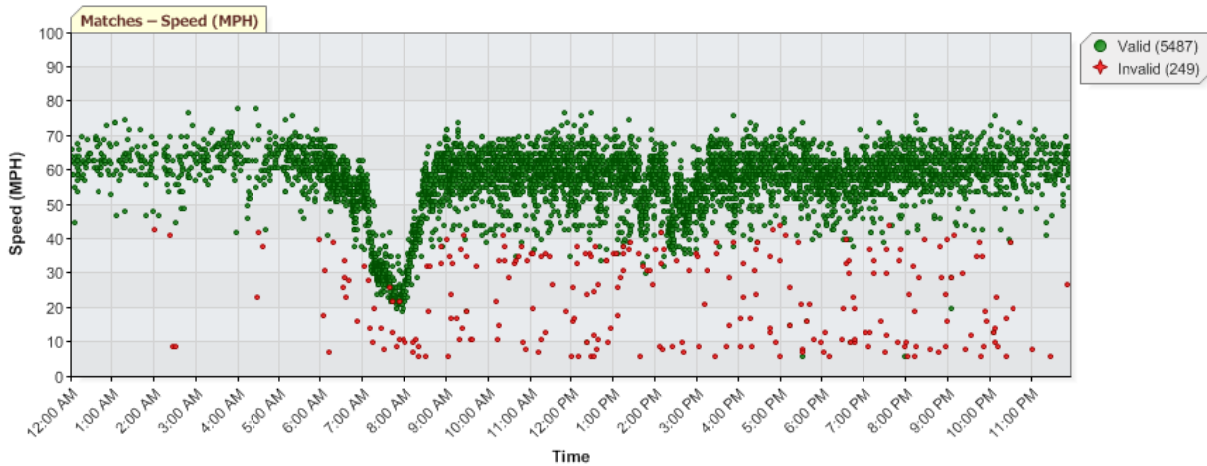
The most common post-match error with reidentification occurs when vehicles experience additional travel time between identification nodes that is not due to heavy traffic conditions. For example, a commuter may make a quick stop for an errand or at a convenience store. Bluetooth-enabled devices on a bus may be read and matched such that a matched travel time includes the bus' stopped time along the segment. For longer segments, drivers may take alternate paths (other than the route being evaluated) between the identification nodes. In all of these examples, the matched travel time is considered an invalid outlier and should not be used in the calculation of benchmark travel time and speed values.

In most cases, the outlier travel time and speed values will be obvious when viewed in the context of other valid values. For example, Figure 6 shows two charts in which an outlier algorithm has already been used to identify invalid values. In certain situations, it may be easier to identify invalid outliers by using a chart that shows travel speeds (rather than travel times) by time-of-day. Or a combination of both travel speed and travel time charts can be used, as shown in Figure 6.

In certain traffic situations, it may be difficult to distinguish invalid outliers from valid data values (see Figure 7). This may occur when there is a lot of traffic variability, or when several travel modes are operating at overlapping speed ranges (e.g., bike and bus traffic combined with congested car traffic on an arterial street). These mixed-mode situations are less common in the United States where the predominant travel mode is the car. A more common scenario in the United States would be when express or carpool lanes are operating at similar speeds as the adjacent main lanes. A weigh station that allows most traffic to bypass may also cause some difficulties in identifying invalid outlier values. Likewise, cash-only lanes at a toll plaza may blur the boundary between valid benchmark values and invalid outlier values. Evaluators will need to proceed cautiously when using reidentification methods for benchmark data collection in situations like these.

Another possible source of error can occur when multiple reidentified devices are within a single vehicle (e.g., multiple passengers with Bluetooth devices on a bus). Multiple identical bus travel times could skew the average travel time as reported by other vehicles in the traffic stream. This error can be detected by looking for multiple devices read at the exact same origin and destination within a few seconds of each other.

Identifying Outliers Using Speed



Identifying Outliers Using Travel Time

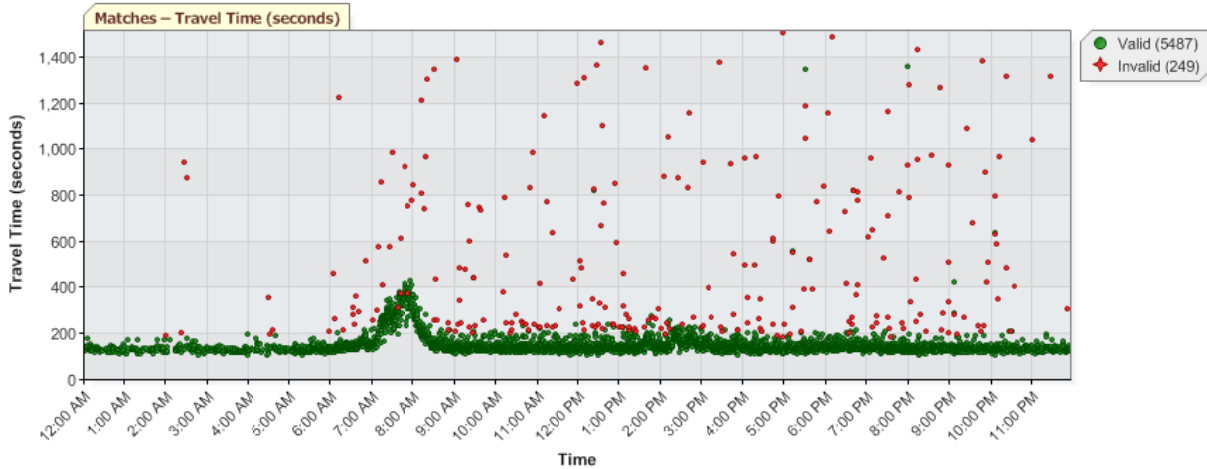


Figure 6 Reidentification Examples: Identifying Outlier Values in Benchmark Data
(Source: Texas Transportation Institute Bluetooth Data Analysis Tools)

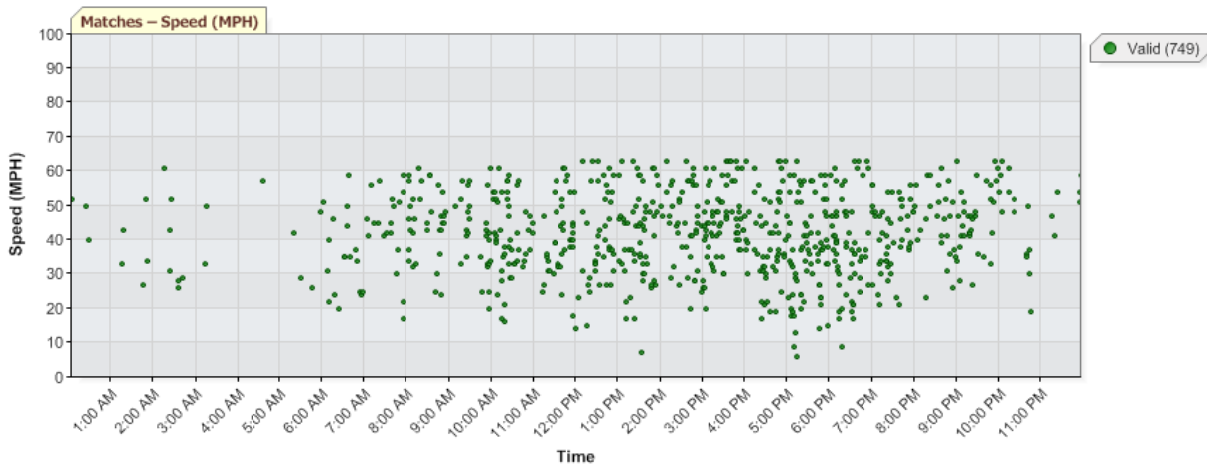


Figure 7 Reidentification Example Showing Difficulty in Identifying Outlier Values When Traffic Speed Variability is High
 (Source: Texas Transportation Institute Bluetooth Data Analysis Tools)

A computer-based, automated algorithm should be used to identify invalid outliers in benchmark data. However, the results of automated algorithms should be visually reviewed to ensure that the algorithm is functioning effectively. For example, Figure 8 shows an example in which the automated algorithm was too aggressive in designating outlier values. In certain situations, manual review and editing may be necessary to ensure the integrity of the benchmark data.

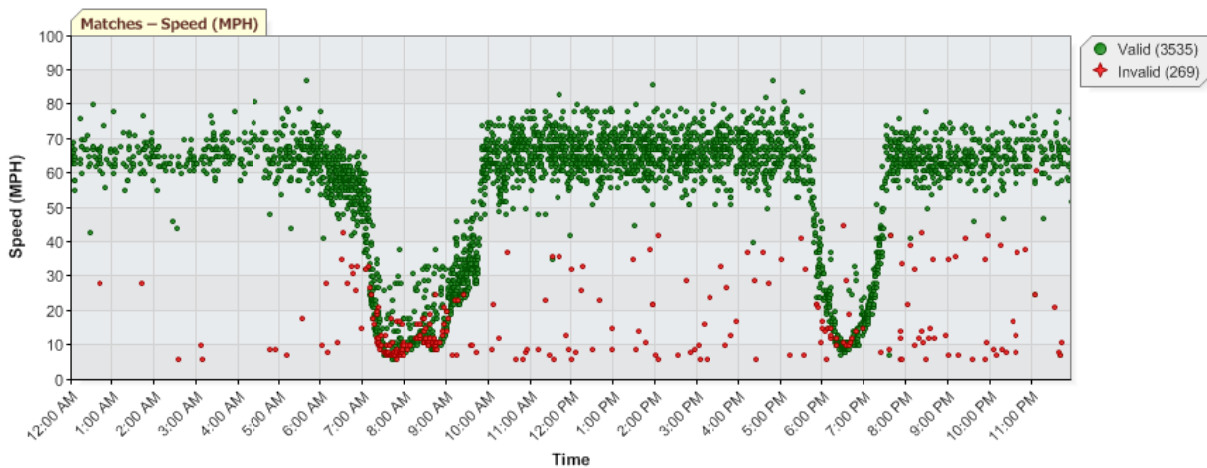


Figure 8 Reidentification Example Showing Overly Aggressive Outlier Algorithm
 (Source: Texas Transportation Institute Bluetooth Data Analysis Tools)

There are several different automated algorithms that can be used to identify invalid outliers. One of the most comprehensive outlier algorithms has been implemented in work by the University of Maryland.⁵ Other algorithms that have been used include the following:

- The **interquartile range (IQR) algorithm** calculates the 25th and 75th percentiles of the last X matches (where X is user-specified, typically about 10 to 15 matches). The IQR is the difference between those two percentiles. The lower acceptable limit is the 25th percentile minus 75% of the IQR. The upper limit is the 75th percentile plus 75% of the IQR.
- An acceptance interval of **twice the standard deviation**. For this algorithm, the average and the standard deviation of the last X matches are computed (where X is user-specified, typically about 10 to 15 matches). Any matches that fall outside of ± 2 standard deviations are designated as outliers.
- Other buffer values in the **acceptable buffer algorithm**. For example, a 20% buffer has worked well on access-controlled highways, but a higher buffer value (e.g., 35%) may be more appropriate on arterial streets.

Data Collection and Processing Issues Unique to Reidentification

There are several data collection and processing issues that are applicable only if reidentification methods are used for benchmark data collection. This section summarizes the guidance in these situations.

Minimum number of reidentification samples – To minimize sampling error and ensure confidence in the benchmark data, a minimum sample size should be established for the calculation of benchmark values. This sample size is applicable for each time interval and link/route for which a benchmark value is calculated for comparison purposes. Typical practice to date has set this minimum sample size between 5 to 10 valid reidentification samples.

In some cases, these minimum samples sizes may not be attainable, especially for roads and time periods with very low traffic flows. In cases like these, traffic congestion may be very unlikely given the low traffic flows, so the need to evaluate real-time traveler information is a lower concern than for more heavily-traveled roads and time periods. Therefore, low-volume roads and time periods are typically not included in most evaluations of real-time speeds and travel times.

Another way to address sampling error is to establish a maximum confidence interval for benchmark values. With this approach, if the benchmark value has a wide confidence interval (e.g., greater than 20 to 30 mph), then there is lots of uncertainty in the benchmark value, and it should not be used for evaluation purposes. For example, assume that 5 reidentification samples were used to calculate a 5-minute average speed estimate of 40 mph, but the 95% confidence interval is between 25 and 55 mph. This wide confidence interval indicates that there is a lot of uncertainty in the 40 mph estimate and it should not be used as a benchmark value.

⁵ Haghani et al, Data Collection of Freeway Travel Time Ground Truth with Bluetooth Sensors. *Transportation Research Record 2160*, Transportation Research Board, 2010, pp. 60-68.

Use the correct formula to calculate average space mean speed – Reidentification methods are used to calculate an individual vehicle’s travel time over a known distance, which is then used to calculate that vehicle’s average speed. To accurately calculate the average speed between two identification nodes, one must use the equation for space mean speed (Equation 4). A space mean speed is calculated by first computing the average travel time, then dividing that into the link length. The relationship between space mean speed and time mean speed is discussed in more detail in the **Definitions** section of Section 1. Space mean speed is typically less than time mean speed; in practice, the difference is typically less than 1 to 3 mph, depending on the variability of travel times.

$$\text{Equation 4 Space Mean Speed, } \bar{v}_{SMS} = \frac{\text{link length}}{\text{average travel time}} = \frac{l}{\frac{\sum t_i}{n}} = \frac{n \times l}{\sum t_i}$$

where l = length of roadway link (distance traveled)
 t_i = travel time of the i th vehicle
 n = number of travel time observations

Additional guidance on individual technologies – There are several publications that have more detailed guidance on each individual reidentification technique. These include:

- *Travel Time Data Collection Handbook*: <http://www.fhwa.dot.gov/ohim/timedata.htm>
- *Data Collection of Freeway Travel Time Ground Truth with Bluetooth Sensors*, Transportation Research Record 2160.

Time interval alignment – For accurate comparisons, it will be necessary to align the benchmark value time interval ($T_{benchmark}$) with the time interval used by the traveler information service (T_{TIS}). Because reidentification methods will produce several benchmark samples for comparison with a single TIS time interval, the reidentification samples should be summarized to match the TIS time interval. The link exit time (as opposed to the link entry time) should be used to summarize the reidentification samples to the corresponding TIS time interval.

Test Vehicle Data Collection Methods

As mentioned previously, reidentification via Bluetooth is increasingly being used as a cost-effective method of traffic monitoring and evaluating other traffic data sources. However, there are a few situations in which test vehicle techniques may be appropriate:

- Roads or time periods in which an adequate number of reidentification samples cannot be collected.
- Roads or locations at which reidentification equipment cannot be positioned (e.g., due to construction, lacking of suitable mounting location, lack of public agency permission, etc.).
- Roads on which reidentification cannot reliably differentiate between different types of lanes (e.g., carpool, express, managed lanes, etc.) or different travel modes (e.g., congested street with bus stops and heavy bike traffic).

- Legacy test vehicle evaluation processes and applications that would be prohibitively expensive to convert to reidentification methods.

In situations like these, test vehicle methods may be the only practical way to collect benchmark data. If this is the case, then extra care should be taken to minimize the sampling error associated with a small number of test vehicle runs (e.g., in many cases, one sample/test vehicle run per time interval). Several recommendations to minimize test vehicle sampling error are provided later in this section.

Several parts of this test vehicle section have been adapted directly from the *NATWG Traffic Information Benchmarking Guidelines*.

Test Vehicle Instrumentation

There are several technologies that can be used to instrument and record the time and location of a test vehicle. Global Positioning System (GPS) data loggers are the most commonly used tool for several reasons: cost-effectiveness, simplicity, and analysis flexibility, particularly integration with geographic information systems (GIS). Electronic distance measuring instruments can also be used, but the resulting data cannot be as easily integrated with GIS. Manual methods (e.g., stopwatch and clipboard) should be avoided because of higher likelihood of measurement and data entry error.

Because GPS is the preferred technology for test vehicle instrumentation, the rest of this section focuses on the use of GPS data loggers for collecting benchmark data. GPS receivers are capable of reporting several attributes, but the most relevant ones for benchmark data collection are the location (i.e., latitude and longitude), heading, and current speed, along with a date and time stamp. These GPS attributes are logged onto a portable storage device at a fixed time sampling rate, typically 1 to 5 seconds. The logged GPS data can then be downloaded at a later time to an office computer, where it is processed and analyzed for benchmark data.

NATWG Guidelines recommend GPS data loggers with at least these features (at a minimum):

- Built-in, rechargeable battery capable of 10 hours of continuous use;
- 12V DC car and 110V AC home charging capability;
- 2-status LED to indicate power and GPS signal;
- USB 2.0 or Bluetooth 2.0 connectivity to host computer; and,
- Built-in antenna.

NATWG Guidelines also recommend the following minimum performance specifications for GPS loggers:

- GPS Chipset: SiRF Star III or Evermore, 20 channel receiver;
- ± 5 meter horizontal accuracy;
- Time To Fix (cold start) < 30s;
- Time To Fix (reacquisition) < 0.1s;
- Sensitivity: -159dBm;
- Track Point Capacity > 1,000,000 data points; and,
- Sample Rate: NEMA rate or programmable 1 to 5 seconds.

The recommended sampling rate for logging GPS data is 1-second intervals. That is, the test vehicle's time-stamped location, speed, and heading should be recorded at least every second.

Test Vehicle Sample Sizes

To have statistical confidence in benchmark values, a minimum sample size is required for each link/route and time period comparison. The minimum sample size is the number of measurements (also referred to as “runs”) that must be made by test vehicles. The minimum required sample size is determined by three factors: the variability of the traffic stream speeds, the desired confidence level, and the permissible error⁶.

In practice, however, minimum statistical sample sizes are rarely met for congested traffic conditions, simply because these sample sizes exceed practical limits and would make test vehicle data collection prohibitively expensive. In these cases, the recommendation is to collect the maximum possible number of test vehicle samples given available resources and logistics. The corresponding error can then be calculated by using the coefficient of variation and desired confidence interval (see Equation 5).

Equation 5 **Sample error, e (%) = $\frac{c.v. \times z_{\alpha}}{\sqrt{n}}$**

where e = sample error (%) based on sample less than required
 $c.v.$ = coefficient of variation (%)
 z_{α} = z-statistic based on confidence level (1.96 for a 95th% confidence)
 n = the actual collected sample size

If only a single run is used to calculate a benchmark value, then a coefficient of variation and the sample error cannot be determined. If the sample size is very small (3 or fewer test vehicle runs per time interval), then the variation could be high, resulting in unacceptably high sample error.

Test Vehicle Driving Style and Behavior

Since each driver of a test vehicle is a member of the data collection team, the driving style and behavior can be controlled to match the desired data collection approach. The test vehicle driving style and behavior is absolutely critical to establishing an accurate benchmark of travel time and speed. Adequate training of test vehicle drivers should be established, and their performance should be monitored or audited periodically to ensure valid benchmark data. Finally, safety for the test driver and other drivers on the road should remain the primary consideration at all times.

Floating car and passing guidelines –The floating car technique should be used in collecting test vehicle benchmark data. In the floating car technique, the test driver “floats” with the traffic stream by attempting to safely pass as many vehicles as pass the test vehicle. In heavy traffic, it will be difficult to count passing vehicles and vehicles being passed. Test drivers should generally avoid passing on single-lane roads. The exception is in the case where the driver is stuck behind a slow-moving vehicle which all or most other vehicles are passing.

Test vehicle speed and acceleration – Unless a route specifically includes a ramp (e.g., freeway-to-freeway connector), drivers should not quickly accelerate or decelerate to insert their vehicle

⁶ *Travel Time Data Collection Handbook*, <http://www.fhwa.dot.gov/ohim/timedata.htm>, accessed April 18, 2011.

into the traffic flow while on a recorded run. Data collection routes should be designed to avoid this by starting or ending at locations that provide an upstream or downstream buffer for the test driver. While collecting benchmark data, test drivers should maintain a speed similar to other vehicles in view, such that the speed of the test vehicle emulates the average speed of all vehicles. There is one exception to this rule: posted speed limits should be obeyed, even if the prevailing vehicle speed is higher than the posted speed limit.

Lane following – Test vehicle drivers should drive in the middle lane (or middle lanes for roadways with four or more lanes). Test vehicle drivers should avoid frequent lane changes and attempt to drive at the average speed of the through traffic stream. When visibility is obscured in the middle lane(s), moving to another lane is desired so that the test vehicle can maintain the same average speed as most vehicles in sight. Travel in the leftmost lanes should be limited because speeds in these lanes often are higher than average and may exceed the speed limit.

Multiple traffic flow patterns - At some multi-lane locations, there may be more than one distinct traffic flow pattern. For example, there may be one or more lanes of traffic waiting to exit onto a ramp while the leftmost through lanes are nearly free-flowing. In situations like this, the test vehicle driver should “float” with the through traffic stream. The test vehicle driver can narrate situations like this (as well as the presence of incidents or other events) into a voice recorder, which will be helpful to the data analyst when interpreting the evaluation results.

Auditing/monitoring test driver style – The floating car technique minimizes the variability between test vehicle travel times, which in turn reduces the minimum required samples. It is essential that test drivers be aware of the importance of driving style in collecting accurate benchmark data. It may be necessary to periodically monitor or audit driving style to ensure that test vehicle drivers are neither too aggressive nor too passive.

Route completeness - Test drivers should try to complete each route on their itinerary in one, uninterrupted run. Test drivers may exit a route or stop along the way if safety or other important circumstances require it, but the data for the corresponding links (i.e., those on which deceleration and acceleration occurred) must be dismissed for the purpose of quality assurance.

Calculating Link/Route Travel Times from GPS Traces

This section describes how to process the time-stamped GPS data points (referred to as a GPS trace) to calculate a benchmark travel time and speed. In general terms, only those GPS points that are at or very near the link/route endpoints will be used to calculate the elapsed travel time. However, GPS data points along the link/route are also typically used for quality assurance purposes and should not be discarded.

Note that the practice of matching all GPS data points to the nearest evaluation link or route and calculating an average link speed from individual instantaneous speeds is not recommended, irrespective of link or route length. Instead, short links (i.e., less than ½ mile) should be combined for evaluation purposes.

Quality assurance for GPS traces – Once the GPS data points are extracted from the field equipment to the analysis computer, several quality assurance processes should be used to review the GPS point data prior to further analysis and calculations:

The GPS traces should be imported into a GIS-based mapping application that also contains the evaluation routes of interest. There are several commercially available and proprietary GIS-based applications that can be used; however, evaluators should select those applications that are well-established and accepted within the industry. Because of the wide variety of GIS-based applications and road map databases that can be used, the evaluation reports should disclose the choice of application, map publisher and version, and map matching methods.

The GIS-based application should be used to visually verify that the test vehicle runs correspond to the correct evaluation routes and itinerary. This step will ensure that obvious problems in data collection or test driver errors are identified so that inadequate runs can be discarded with no further scrutiny.

Individual GPS points should be scrutinized to ensure a valid run, and these are considered best practices by NATWG:

- If a GPS point appears 75 feet or more from the roadway segment on which it is supposed to be snapped, the record should be dismissed as invalid;
- If a GPS record is within the vicinity of the roadway segment to which it belongs, but ends up being snapped to the wrong roadway by the map-matching software, it is also recommended that this record be dismissed;
- If GPS records are not contiguous to a route, (i.e. they do not line up in sequential TMC paths), they should be dismissed;
- Even though speed readings from the GPS data collection device are not used directly in these *Guidelines*, it is recommended that GPS records that feature unreasonable speed values be dismissed;

For each link being evaluated, the NATWG Guidelines state that these conditions must be met:

- The link must have been entered, traveled and exited at prevailing traffic speed (not including acceleration or deceleration phases due to ingress/egress);
- Based on the recommended sampling rate of 1 reading per second, at least 90% of the records supposed to be collected over the length of the segment must have passed the filters in the previous paragraph;
- Gaps in the GPS trace must remain less than 10 seconds. However, there are cases in which such gaps may not be a cause for concern (e.g. if the distance traveled is small). Ultimately, testers should apply judgment in deciding whether or not gaps in the data should invalidate a segment. If gaps greater than 10 seconds remain in a run that is used for benchmarking, a note should be added to the reporting of results.

Unless otherwise noted, data collected along the link or route must be discarded all together when the above conditions are not met.

Calculating benchmark travel time and speed – Benchmark travel times are calculated by using those GPS data points that are at or nearest to the link endpoints to determine the elapsed time (i.e., travel time). It is unlikely (but not impossible) that one of the GPS data points will land exactly on the link endpoint; therefore, it will be necessary to identify the time (to the nearest tenth-second) that best approximates the test vehicle position at each link endpoint. The difference between the link entry and exit time, the elapsed time, is then the benchmark travel time for that link.

Errors of one or two seconds can have a noticeable impact on link speeds, particularly for short distances. To capture the exact time at each link endpoint, the two GPS points straddling each link endpoint shall be used to estimate the crossing time to the nearest tenth-second.

Once the benchmark travel time has been calculated, the benchmark speed can be calculated using Equation 6. The link distance in this equation should be obtained from an established road inventory or mapping database maintained by a public agency or sold by a commercial entity. Alternatively, the link or route distance may be calculated using dynamic segmentation techniques within a GIS environment.

$$\text{Equation 6} \quad \text{Benchmark Speed, } v = \frac{\text{link or route length}}{\text{benchmark travel time}} = \frac{l_i}{t_i}$$

where l_i = length of roadway link (distance traveled)
 t_i = travel time of the test vehicle

Time interval alignment – For accurate comparisons, it is necessary to align the benchmark travel time interval ($T_{benchmark}$) with the time interval used by the traveler information service (T_{TIS}). If the benchmark travel time both begins and ends within (T_{TIS}), then the benchmark travel time and the TIS value may be directly compared.

However, when $T_{benchmark}$ represents more than one TIS time interval, an equivalent TIS speed must be computed for the benchmark period. The following procedure should be used to compute the equivalent TIS speed:

1. Determine the amount of time the test vehicle spent in each TIS time interval.
2. Calculate the weighted average speed based on the TIS reported speeds in the TIS time intervals, where the weights are computed from the ratio of test vehicle time spent in each TIS interval to the total benchmark travel time, $T_{benchmark}$.

As an example, consider a benchmark time interval defined by

$$T_{benchmark} = 9:12:22 - 9:02:03 = 10:19, \text{ or } 619 \text{ seconds}$$

Now suppose that the TIS reports speeds every 5 minutes, aligned with the start of each hour. Then $T_{benchmark}$ overlaps 3 TIS time intervals, as illustrated in Figure 9.

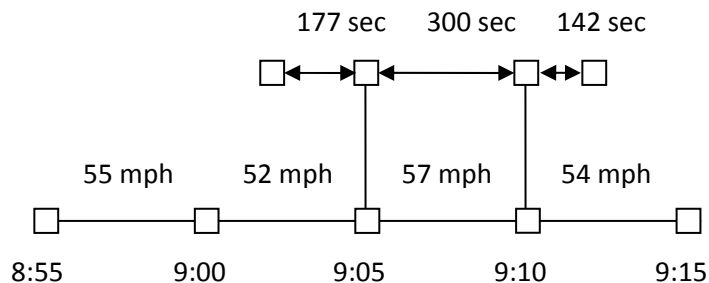


Figure 9 Illustration of Overlap Between Benchmark Travel Time and Multiple TIS Intervals
(Source: NATWG Guidelines)

Thus in this example, the benchmark time spent in the first time TIS interval of the 9:00 hour is 177 seconds, the time spent in the second interval is 300 seconds, and the time spent in the third interval is 142 seconds. The equivalent TIS speed for the ground-truth interval is then

$$V_{TIS} = \frac{(177 \times 52) + (300 \times 57) + (142 \times 54)}{619} \cong 55 \text{ mph}$$

V_{TIS} is then used when comparing the benchmark speed. In this example, the estimated TIS speed of 55 mph (i.e., estimate of average TIS speed from 9:02:03 to 9:12:22) is compared to the benchmark travel time run which during that same time interval.

Synchronize clocks - The clocks on all data collection equipment should be synchronized to a common time standard at the beginning of every day of data collection. The recommended time standard for all equipment is the NIST Internet (<http://tf.nist.gov/timefreq/index.html>). Data collection personnel often find it useful to also have a synchronized wristwatch available during days of data collection.

Capping Benchmark Speeds to Correspond with Posted Speed Limits

Although posted speeds limits do not always reflect prevailing free-flow speed, they represent a standard reference speed against which reported and measured speeds may be compared. Rather than requiring each benchmark testing agency to measure a standard reference speed for each road segment, we recommend that benchmark testing use map-based posted speed limits, and simply report the map type and version number for purposes of comparison.

Although TIS and benchmark speeds may exceed the posted speed limit, it is recommended that all speeds be set to this maximum value of posted speed limit. Thus, if a TIS provider reports a speed of 72 mph on a freeway for which the posted speed limit is 70 mph, the TIS speed will be set to 70 mph for purposes of comparison to benchmark speeds.

4. COMPUTE AND REPORT ACCURACY MEASURES

This section describes the computation and reporting of accuracy measures. These calculation procedures assume that benchmark data have already been collected and reduced to the appropriate roadway segmentation and time interval (as described in Section 3). This section also includes several examples to illustrate how the accuracy evaluation results can be effectively reported.

Gather TIS Data to be Evaluated

The calculation of accuracy measures requires that benchmark values and traveler information service (TIS) values be collected under the exact same conditions (i.e., location, date, and time). In practice, this may be difficult, but as much effort as possible should be made to minimize the differences in the benchmark and TIS locations and times. For example, TIS providers always include a time-stamp in real-time traffic information, but rarely indicate the exact start or end time for the travel time population being measured. One can infer the start and end time from their update cycle; however, this inference is not a certainty. Section 3 discussed this in detail for the collection of the benchmark values, and this section provides guidance for gathering the TIS values.

Use generalized schedules: The scheduled dates of the evaluation should be kept at a general level so that the TIS provider cannot “game the system” (i.e., put extraordinary effort into data quality for only those dates being evaluated). For example, if the evaluation data will include several different days, only the month should be announced to the TIS provider.

Gather TIS data in real-time: The TIS data should be gathered by the evaluator in real-time, especially when the TIS data provider is a separate entity. This real-time capture of TIS data ensures its legitimacy and prevents the possibility of data fabrication “after the fact.” Gathering the TIS data in real-time is less critical for internal evaluations (evaluator and TIS provider are the same entity) or when there is an established credible archive of real-time data reports.

Evaluate data provided to consumer: Many TIS providers do additional processing of their TIS estimates before providing to the end consumer. For example, TIS speeds are sometimes rounded down so that they do not exceed the posted speed limit. Another example is rounding TIS travel times to the nearest integer value, or providing a range of travel time values. To address this post-processing discrepancy, the benchmark values should be post-processed in the same way as the TIS estimates. If TIS speeds are rounded to the posted speed limit, then the benchmark values should be rounded to the same posted speed limit. If TIS travel times are rounded down or communicated as a range, then the same rounding or range algorithms should be applied to the benchmark values.

Comparing benchmark values to intermediate TIS values (i.e., before they are post-processed for the consumer) is not recommended except when both the evaluator and the TIS provider agree that intermediate values are to be evaluated for accuracy. For example, a TIS may provide integer link speed values to a state DOT, who then post-processes the link speeds into a color-coded congestion map, link travel times, and integer speed values rounded down to the posted speed limit. In this case, the state DOT is a consumer of the intermediate TIS speed values and the state DOT should evaluate those TIS speed values. Unless significant post-processing is done by the state DOT, the accuracy of the end consumer information (color-coded congestion map, link travel times, and integer speed values) can be inferred from the accuracy of the intermediate TIS speed values. Accordingly, the acceptable accuracy ranges of the TIS speed values should be informed by the end consumer uses.

Use confidence intervals to indicate uncertainty in benchmark values: Even with reidentification methods that collect larger sample sizes, the resulting benchmark values will always have some degree of uncertainty. Large samples and/or low variation among samples create less uncertainty and narrow confidence intervals. Small samples and/or high variation among samples create more uncertainty and wider confidence intervals. Using statistical confidence intervals (see Equation 7) addresses this issue of uncertainty in the benchmark value and provides a fair assessment of TIS data.

For example, a contract requirement may state that the benchmark values be computed by using an upper or lower confidence interval, rather than a single average benchmark value. For a given link and time period, assume a benchmark speed value of 38 mph, but due to low sample size and high variability, the lower and upper 95% confidence intervals are 33 and 43 mph. Now assume a TIS value of 45 mph was provided for this same link and time period. Using the confidence interval method, this TIS value has an error of 2 mph (45 – 43). Now assume a TIS value of 42 mph. Since this value is within the confidence interval (33 to 43 mph), the error is computed as 0 mph.

Equation 7 *Confidence Interval* = $\bar{x} \pm t_{(n-1, \alpha/2)} \times \left(\frac{s}{\sqrt{n}} \right)$

where \bar{x}	= evaluation sample mean, also called benchmark value
$t_{(n-1, \alpha/2)}$	= the t-value associated with n-1 degrees of freedom and α confidence interval (see Table 4 for selected t-values). For example, when n=31, the t-value is 2.045 for a 95% confidence interval and 1.70 for a 90% confidence interval.
s	= sample standard deviation
n	= number of estimate-to-benchmark comparisons

Report accuracy separately for different speed ranges: To appropriately characterize TIS accuracy, evaluation results should be calculated and reported separately in different traffic conditions ranging from light flow to heavy congestion. A real-time TIS could provide static historical speeds and still be reasonably accurate during many hours of the day when traffic is not congested. Therefore, combining accuracy statistics from light flow and heavy congestion can provide misleading summary results, as light flow conditions are more common and are more heavily weighted in the calculation of average error.

Table 5 presents recommended speed ranges for use in computing and reporting accuracy statistics for link speeds and route travel times. For congestion category evaluations, the predefined TIS speed categories should be used when calculating the percent of correct category classifications. The ranges in Table 5 are considered “rules of thumb” and are chosen to represent different traffic conditions. These ranges can be adjusted for local conditions; however, selecting different ranges may decrease the comparability of your results with other evaluation results.

Table 4 Table of t-Distribution Values for Typical Two-Sided Confidence Intervals and Sample Sizes

n-1	t-Value Corresponding to Two-Sided Confidence Interval				
	80%	90%	95%	98%	99%
1	3.078	6.314	12.710	31.820	63.660
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
50	1.299	1.676	2.009	2.403	2.678
60	1.296	1.671	2.000	2.390	2.660
80	1.292	1.664	1.990	2.374	2.639
100	1.290	1.660	1.984	2.364	2.626
120	1.289	1.658	1.980	2.358	2.617
∞	1.282	1.645	1.960	2.326	2.576

Note: t-values for other combinations can be calculated using the TDIST or TINV functions within Microsoft Excel®.

Table 5 Recommended Speed Ranges for Computing and Reporting Accuracy Statistics

Range	Benchmark Speed Range	
	Access-Controlled Roads (e.g., freeways)	Arterial Streets
1 (light flow)	Greater than 60 mph	Greater than 40 mph
2 (transition)	45 to 60 mph	30 to 40 mph
3 (slowing)	30 to 45 mph	20 to 30 mph
4 (stop-and-go)	Less than 30 mph	Less than 20 mph

Select Accuracy Measures

As indicated in Section 2, the selection of accuracy measures will depend on the evaluation scenario. There is no single best accuracy measure for all possible evaluation scenarios. Each of the accuracy measures defined in this section has certain strengths and/or limitations.

From past experience, these accuracy measures are commonly preferred for each evaluation scenario:

- **Congestion category**
 - Percent of correct category classifications (with a specified boundary tolerance), with the option of reporting the average deviation of incorrect classifications
- **Link speed**
 - Root mean square error, or RMSE (units of mph)
 - Average absolute error (units of mph)
 - Average error, also known as bias or signed error (units of mph)
 - Percent of TIS values within *X* of the benchmark value, where *X* is a specified error tolerance in mph, percentage, or statistical confidence interval (units of %)
- **Route travel time**
 - Average absolute percent error (units of %)
 - Average absolute error per unit length (units of seconds per mile)
 - Average absolute error (units of minutes)

Accuracy measures other than these preferred measures are described in this section, but users are cautioned to understand their strengths and limitations before using them in an accuracy evaluation.

In addition to accuracy, timeliness (i.e., lag) is another important data quality attribute for traveler information services.⁷ However, the calculation and evaluation of timeliness is not addressed in these *Guidelines*, even though timeliness can be calculated with the same benchmark data used to calculate accuracy. It should be recognized that the accuracy measures in these *Guidelines* will indirectly capture timeliness problems. For example, if the timeliness of traveler information lags by 5 to 10 minutes, then the accuracy will be lower during periods of quickly changing traffic conditions. However, to directly

⁷ See recommended traffic data quality measures at <http://ntl.bts.gov/lib/jpodocs/reports/14058.htm>, accessed April 14, 2011.

quantify timeliness, one must calculate the actual time units by which the reported traveler information lags the benchmark values.

Establish Acceptable Accuracy Targets

As indicated in Section 1, these Guidelines do not establish the level of accuracy that a TIS provider must achieve. These accuracy targets (i.e., acceptable accuracy values) must be defined by the user or purchaser of the data based on their intended application of the data. However, the following paragraphs will highlight several factors that should be considered when establishing accuracy targets.

The most important consideration relates to the use of speed-based accuracy measures. For example, assume that the accuracy target is to have TIS speeds within 10 mph of the benchmark value. In terms of travel time, however, a 10 mph error at 60 mph is less detrimental to the traveler than a 10 mph error at 30 mph.

Figure 10 and Table 6 better illustrate this inconsistency of absolute speed error impact at different speeds. For example, consider a 5-mile roadway segment in which the absolute acceptable error is 10 mph. At 60 mph, this 10 mph acceptable error can result in a travel time estimate that is either 0.7 minute too fast or 1 minute too slow (no error results in a 5-minute travel time). However, at 30 mph, this 10 mph acceptable error can result in a travel time estimate that is either 2.5 minutes too fast or 5 minutes too slow (no error results in a 10-minute travel time).

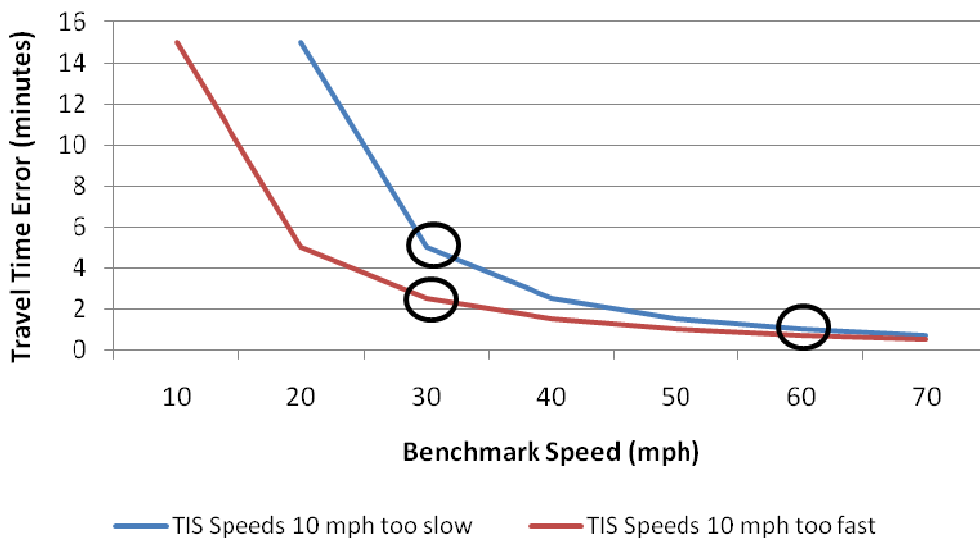


Figure 10 Inconsistencies in Travel Time Error at Different Speeds for Absolute Speed Error (5-mile segment, ± 5 mph acceptable error)

Table 6 Example Calculations for Absolute Speed Error and Resulting Travel Time Error
(5-mile segment, ± 5 mph acceptable error)

Benchmark Speed (mph)	Benchmark Travel Time (minutes)	TIS Travel Time		TIS Travel Time Error		TIS Travel Time % Error	
		10 mph too fast	10 mph too slow	10 mph too fast	10 mph too slow	10 mph too fast	10 mph too slow
20	15.0	10.0	30.0	5.0	15.0	33%	100%
30	10.0	7.5	15.0	2.5	5.0	25%	50%
40	7.5	6.0	10.0	1.5	2.5	20%	33%
50	6.0	5.0	7.5	1.0	1.5	17%	25%
60	5.0	4.3	6.0	0.7	1.0	14%	20%
70	4.3	3.8	5.0	0.5	0.7	13%	17%

Note: Shaded rows represent examples described in text.

For short segments, this inconsistency can be small and sometimes negligible. But for longer segments, an error that is acceptable at higher speeds can result in a very high percentage error at low speeds. One way to address this inconsistency is the use of separate speed ranges (as shown in Table 5) and a sliding scale for the accuracy targets. For example, a 10 mph acceptable error on access-controlled roads could be specified for the fastest 2 speed ranges (i.e., greater than 60 mph, 45 to 60 mph), whereas a 5 mph acceptable error could be specified for the slowest 2 speed ranges (i.e., 30 to 45 mph, less than 30 mph).

Another way to address this inconsistency is the use of relative measures besides absolute speed error. For example, the percent of TIS values within X of the benchmark could be used, where X is stated in percentage terms or a statistical confidence interval. Average absolute percent error can also be used to specify acceptable error levels. For example, the Section 1201 traveler information requirements (i.e., 23 CFR Part 511) establish an acceptable travel time error level of less than 15%.

Calculate and Report Accuracy Measures

This section describes the calculation and reporting of the preferred accuracy measures, as well as a few other basic accuracy measures. For most of these accuracy measures, comparisons are made for each unique roadway link/route and time interval, then aggregated or averaged over all such link/route and time comparisons within a designated speed range (Table 5).

Congestion Category Accuracy Measure

In some cases, TIS is provided in the form of generalized or color-coded congestion categories, with each category representing a defined range of congestion or travel speed values. A common example is a color-coded speed map, with green-colored road links representing the fastest speeds, red-colored road links representing the slowest speeds, and other colors representing in-between ranges (see Figure 1). **The preferred measure for congestion categories is the percent of correct category classifications (with a specified boundary tolerance)**, with the option of reporting the average difference of incorrect classifications. This measure is simple, easy to calculate (Equation 8), and easy to understand.

A correct category classification refers to those instances in which a TIS provider reports a congestion category correctly as compared to the benchmark category value (and appropriate boundary tolerances).

For example, an online TIS map shows link congestion as being “yellow - slowing” for speeds between 30 and 45 mph, and the benchmark speed value of 38 mph confirms that the true average speed is in the “yellow – slowing” category. This indicates a correct congestion category classification by the TIS provider.

When evaluating the categorical accuracy of congestion levels, it is possible for the benchmark value to differ from the TIS value by 1 mph and yet still be considered an incorrect classification. For example, the measured benchmark value is 32 mph and the TIS category is severe congestion, which ranges from 20 to 30 mph for this TIS. In this case, the deviation of the benchmark from the moderate congestion boundary is only 2 mph (i.e., 32 mph – 30 mph) yet it is considered an incorrect classification. This result does not fairly represent the accuracy of the TIS value.

For this reason, boundary tolerances should be used when evaluating congestion categories, so that TIS values outside but close to the category count as a correct classification. These tolerances should be predefined by the evaluator and reported clearly in evaluation results. The tolerance is subtracted from the lower boundary and added to the upper boundary (Equation 9). In the case of a 2 mph boundary tolerance, for example, a TIS value in the 20 to 30 mph category would be counted as a correct classification if the benchmark value was between 18 and 32 mph. Note that the same tolerance concept could be applied to travel time ranges, such as that placed on dynamic message signs.

To evaluate these congestion categories, the underlying or estimated travel speed ranges must be provided for each category (even if the TIS consumer does not see these speed ranges on their TIS interface). Benchmark values can then be compared to the TIS category speed range (with a specified tolerance at the lower and upper boundary) to determine a correct classification. In addition to the measure of percent of correct classifications, the deviation between the respective TIS category boundary and the benchmark value can be calculated. This deviation can provide additional insight into the question “When the TIS value is not classified correctly, how far off is the TIS estimate?”

Equation 8 *Percent of correct classifications (in each congestion category)* = $\frac{\text{Number of correct classifications}}{\text{Total number of comparisons}}$

Equation 9 *Correct classification when* $\mu_{lower} - \theta \leq \mu \leq \mu_{upper} + \theta$

where μ = benchmark value (mph)
 μ_{lower} = lower category boundary value (mph)
 μ_{upper} = upper category boundary value (mph)
 θ = boundary tolerance (mph)

A single benchmark value (instead of a statistical confidence interval) should be used in congestion category evaluations. The use of boundary tolerances accounts for benchmark value uncertainty and makes the confidence interval calculation unnecessary for categorical analysis.

Figure 11 shows an example of how categorical accuracy results can be presented. This example shows results for a single evaluation period; however, a similar layout (clustered columns) can also be used to show changes in accuracy over time for each of the TIS-defined congestion categories.

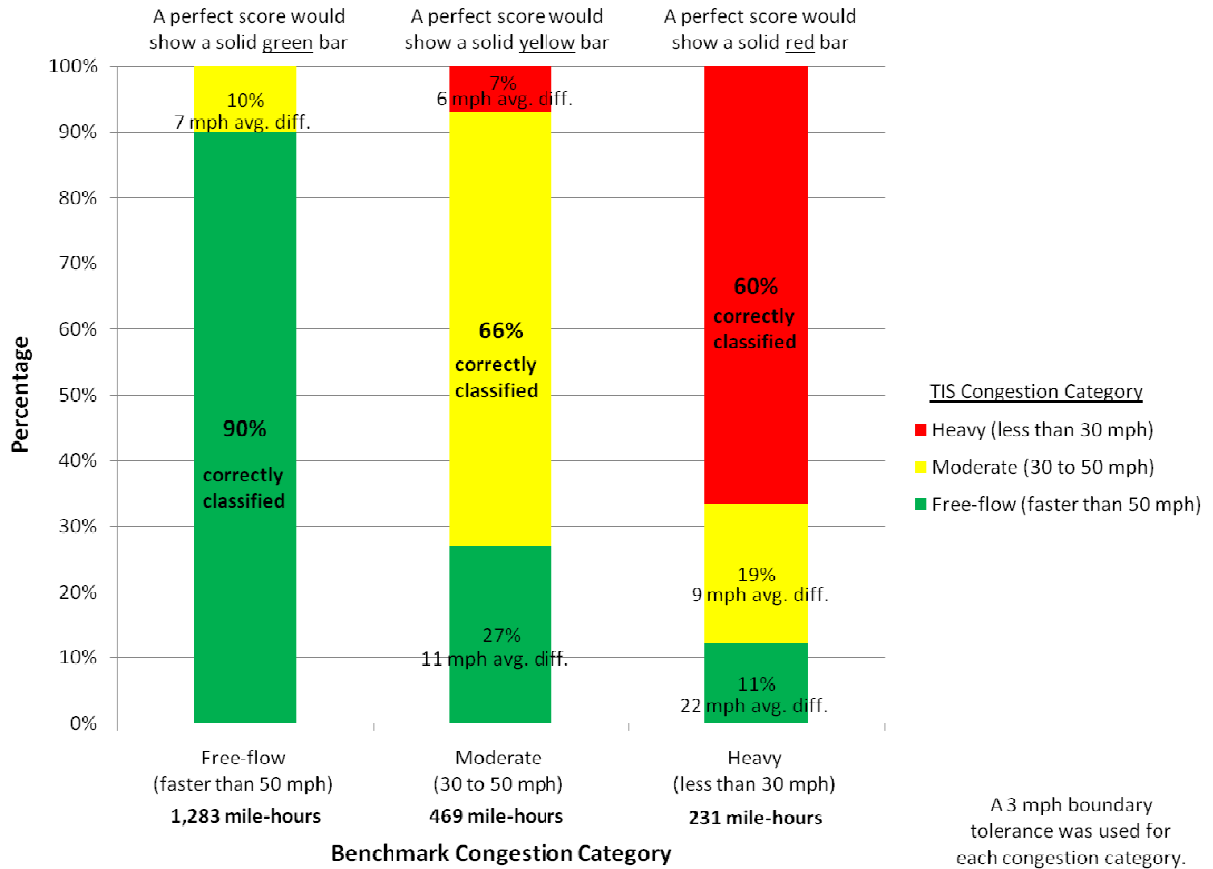


Figure 11 Illustrative Example of Categorical Accuracy Evaluation Results

In some cases, traffic congestion levels may be reported in only two categories: 1) normal traffic flow; and 2) congested traffic flow. In this situation, correct classifications only occur when the benchmark and TIS values are in the “congested” category (Table 7). If the benchmark value is “normal flow” and the TIS value also is “normal flow,” this should not be counted as a correct classification or an incorrect classification. Because normal traffic flow occurs during many hours of the day, including this as a correct classification could provide misleading summary results, as these normal flow conditions are more common and are more heavily weighted in the calculation of percent correct classifications.

Table 7 Benchmark Evaluation of Normal and Congested Traffic Flow

Benchmark Value Classification	TIS Classification	
	Normal flow	Congested
Normal flow	Do not count	Incorrect Classification (secondary metric)
Congested	Incorrect Classification (secondary metric)	Correct Classification (primary metric)

Link Speed Accuracy Measures

In some cases, TIS data are provided as numeric speed values for specified roadway links (see Figure 2). For example, a TIS provider may publish a real-time data feed of link speeds to distribution partners (in this example, each defined TMC path is considered a link). Another example is when numeric link speed values are provided in association (via a hyperlink or mouseover/hover box) with a color-coded speed map.

The preferred measures for computing link speed accuracy are:

- **Root mean square error, RMSE** (units of mph)
- **Average error**, also known as bias or signed error (units of mph)
- **Average absolute error** (units of mph)
- **Percent of TIS values within X of the benchmark value**, where X is a specified error tolerance in mph, percentage, or statistical confidence interval (units of %)

Each of these measures is described in the following paragraphs, and Table 8 shows calculation examples.

The root mean square error (RMSE) is calculated as shown in Equation 10. This measure is relatively easy to calculate and is standard output for data analysis software. However, the underlying concept may not be well understood or related to by non-technical audiences. The RMSE does give greater weight to higher magnitude errors (by squaring the error term), so it is an effective way to identify cases where some data values diverge from the benchmark by large amounts.

Equation 10 *Root mean square error, RMSE (mph)* = $\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu_i)^2}$

where μ_i = benchmark value for the *i*th comparison (mph)
 x_i = the *i*th TIS estimate (mph)
 n = number of estimate-to-benchmark comparisons

The average absolute error (Equation 11) is a simple accuracy measure that is easy to calculate and easy to understand. Average absolute error conveys the average magnitude of the error (in mph), regardless of whether the error is positive or negative. However, the average absolute error does not indicate if there is a consistent negative or positive bias. Average absolute error should not be used for route travel times, as it provides misleading results when comparing different route lengths.

Equation 11 *Average absolute error (mph)* = $\frac{1}{n} \sum_{i=1}^n \text{abs}(x_i - \mu_i)$

where μ_i = benchmark value for the *i*th comparison (mph)
 x_i = the *i*th TIS estimate (mph)
 n = number of estimate-to-benchmark comparisons

The average error (Equation 12) is also referred to as the bias or the signed error, as the sign (positive or negative) of this measure indicates whether the TIS values are consistently higher or lower than the benchmark values. This measure is simple, easy to calculate, and easy to understand. However, large

positive and negative errors can cancel each other and provide a misleading estimate of the magnitude of the error. Therefore, average error should only be used to indicate a consistent bias in TIS values. Average error should not be used for route travel times, as it provides misleading results when comparing different route lengths.

Equation 12 *Average error or bias (mph)* = $\frac{1}{n} \sum_{i=1}^n (x_i - \mu_i)$

where μ_i = benchmark value for the *i*th comparison (mph)
 x_i = the *i*th TIS estimate (mph)
 n = number of estimate-to-benchmark comparisons

The **percent of TIS values within X of the benchmark value** (where X is a specified error tolerance in mph, percentage, or statistical confidence interval, Equation 13) is similar to categorical accuracy but is used when the TIS reports discrete speed or travel time values. The use of an acceptable error tolerance or error interval indicates that “close counts.” The acceptable error tolerance or error interval should be predefined by the evaluator and reported clearly in evaluation results.

Equation 13 *Percent of TIS values within X* = $\frac{\text{Number of TIS values within X}}{\text{Total number of comparisons}}$

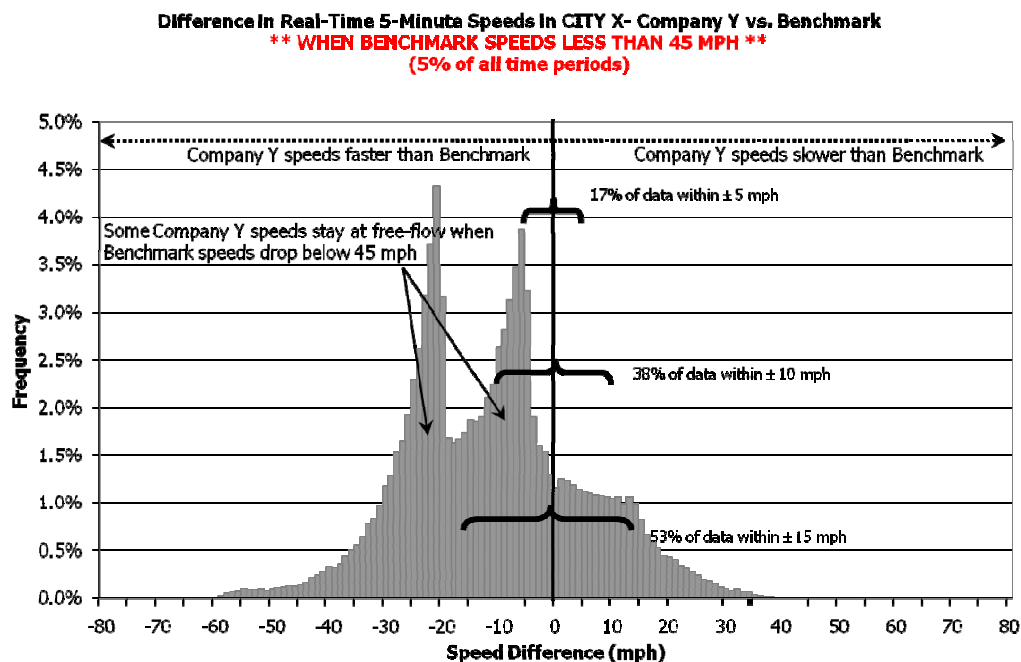
This accuracy measure provides the ability to vary the error tolerance on different roads (freeways vs. arterials) or area types (urban vs. rural) yet still report a consistent measure. Additionally, the percentage value with a maximum score of 100% is easy to understand. However, the calculation details are a bit more difficult to communicate concisely. Another possible limitation is that statistical confidence intervals for this measure cannot be computed with single test vehicle data collection.

Table 8 Illustrative Example of Link Speed Accuracy Calculations

Start Time	Ending Time	TIS Link Speed	Benchmark Speed	Difference	Difference Squared	Benchmark +10 mph	Benchmark -10 mph	Within ± 10 mph
7:20:17	7:21:17	55	44	11	121	34	54	False
7:21:17	7:22:17	48	39	9	81	29	49	True
7:23:17	7:24:17	43	39	4	16	29	49	True
7:26:17	7:27:17	37	32	5	25	22	42	True
7:27:17	7:28:17	35	31	4	16	21	41	True
7:28:17	7:29:17	31	32	-1	1	22	42	True
7:29:17	7:30:17	31	32	-1	1	22	42	True
7:30:17	7:31:17	30	37	-7	49	27	47	True
7:32:17	7:33:17	35	29	6	36	19	39	True
7:34:17	7:35:17	28	29	-1	1	19	39	True
7:35:17	7:36:17	27	21	6	36	11	31	True
7:36:17	7:37:18	26	29	-3	9	19	39	True
7:37:18	7:38:18	26	23	3	9	13	33	True
7:39:18	7:40:18	29	28	1	1	18	38	True
7:40:18	7:41:18	27	22	5	25	12	32	True
7:41:18	7:42:18	29	26	3	9	16	36	True
7:42:18	7:43:18	25	22	3	9	12	32	True
7:43:18	7:44:18	27	24	3	9	14	34	True
7:45:18	7:46:18	25	34	-9	81	24	44	True
7:47:18	7:48:18	29	25	4	16	15	35	True
7:49:18	7:50:18	28	27	1	1	17	37	True
7:50:18	7:51:18	29	32	-3	9	22	42	True
7:53:18	7:54:18	35	28	7	49	18	38	True
7:55:18	7:56:18	32	23	9	81	13	33	True
7:56:18	7:57:18	29	24	5	25	14	34	True
7:57:18	7:58:18	29	27	2	4	17	37	True
7:59:19	8:00:19	31	30	1	1	20	40	True
				Avg = +2.5	Sum = 721			
Summary								
RMSE = Square root (721/27) = 5.2 mph								
Average Signed Error (Bias) = +2.5 mph								
% of TIS Values within 10 mph = (26/27) = 96%								

The first three link speed accuracy measures (i.e., root mean square error, average absolute error, and average error) can be calculated for a single link-date-time combination, as well as aggregated over several links or routes for a corridor or citywide average value. The arithmetic averages should be computed for each predefined speed range (see Table 5) as shown in Equations 10, 11, and 12. The traffic volume or vehicle-miles of travel (VMT) can be used as a weighting factor for each link-date-time combination in calculating a citywide average. This weighting factor is not required for corridor or citywide averages, as detailed traffic volume data at the time of the evaluation may not be available for all roadway links.

In addition to a single aggregated average, the frequency distribution of individual errors can be used to visualize the error patterns (see Figure 12). Although these distribution charts can be difficult to explain to non-technical audiences, they do convey more information about error (such as the bias, range and magnitude) than a single average value.



The remaining measure, the percent of TIS values within X, already accommodates aggregate corridor or citywide values in its formula. In fact, this accuracy measure becomes less suitable when only a few benchmark comparisons are made. As with the other accuracy measures, the “percent of TIS values within X” should be computed for each predefined speed range (see Table 5). Otherwise, higher accuracy in light traffic flows can overwhelm lower accuracy in congested traffic conditions (due to averaging many hours of TIS accuracy in free flow with only a few hours of TIS accuracy in congestion).

Route Travel Time Accuracy Measures

The third evaluation scenario is a **route travel time** evaluation. In this scenario, traveler information is provided as travel times (typically integer values) for specified travel routes. A route may traverse a single named roadway or multiple named roadways, but it is always a collection of two or more links. For example, many state DOTs provide route travel times on dynamic message signs. Navigation devices or routing algorithms provide route travel times through a network that typically includes two or more named roadways.

The preferred measures for computing route travel time accuracy are:

- Average absolute percent error (units of %)
- Average absolute error per unit length (units of seconds per mile)
- Average absolute error (units of minutes)

The average absolute percent error (Equation 14, also called mean absolute percent error or MAPE) is a simple measure that is easy to calculate. This measure conveys only the magnitude of the error, and not whether the error is consistently positive or negative. However, for short travel times, small changes in magnitude can create large changes in the average percent difference.

$$\text{Equation 14} \quad \text{Average absolute percent error (\%)} = \frac{1}{n} \sum_{i=1}^n \frac{\text{abs}(x_i - \mu_i)}{\mu_i}$$

where μ_i = benchmark value for the i th comparison (minutes or seconds)
 x_i = the i th TIS estimate (minutes or seconds)
 n = number of estimate-to-benchmark comparisons

The average absolute error per unit length (Equation 15) normalizes the error by route length and permits comparisons across or aggregation of different routes. This measure conveys only the magnitude of the error, and not whether the error is consistently positive or negative. This measure is ideal for summarizing the evaluation results of numerous routes, as in an aggregated citywide average.

$$\text{Equation 15} \quad \text{Average absolute error per unit length (seconds per mile)} = \frac{1}{n} \sum_{i=1}^n \frac{\text{abs}(x_i - \mu_i)}{l_i}$$

where μ_i = benchmark value for the i th comparison (seconds)
 x_i = the i th TIS estimate (seconds)
 l_i = route length for the i th TIS estimate (miles)
 n = number of estimate-to-benchmark comparisons

The average absolute error (Equation 16) is identical to the above measure except that it is not normalized by route length. This measure is recommended only for scenarios in which the actual time error for a specified route is desired (such as evaluating specific route travel times posted on dynamic message signs). The average absolute error should not be used to aggregate accuracy results from different routes (because of the varying route lengths).

Equation 16 *Average absolute error (minutes)* = $\frac{1}{n} \sum_{i=1}^n \text{abs}(x_i - \mu_i)$

where μ_i = benchmark value for the *i*th comparison (minutes)
 x_i = the *i*th TIS estimate (minutes)
 n = number of estimate-to-benchmark comparisons for **the same route**

Table 9 and Figure 13 show examples of the results of these calculations.

Table 9 Illustrative Example of Accuracy for Route Travel Time

Link 18: I-75 SB: Timber Lane to SR 4
 Length: 2.9 miles (4.7 km)

Benchmark Speed Range	Average Absolute Error		Mean Difference (Bias)		Comparison Samples
	Travel Time (minutes)	Speed (mph)	Travel Time (minutes)	Speed (mph)	
0 to 30 mph	1.1	6.8	-1.1	+6.8	37
30 to 45 mph	0.6	4.6	+0.2	+1.3	48
45 to 60 mph	0.5	7.1	+0.4	-2.3	87
60+ mph	1.0	10.5	+1.0	-10.5	3

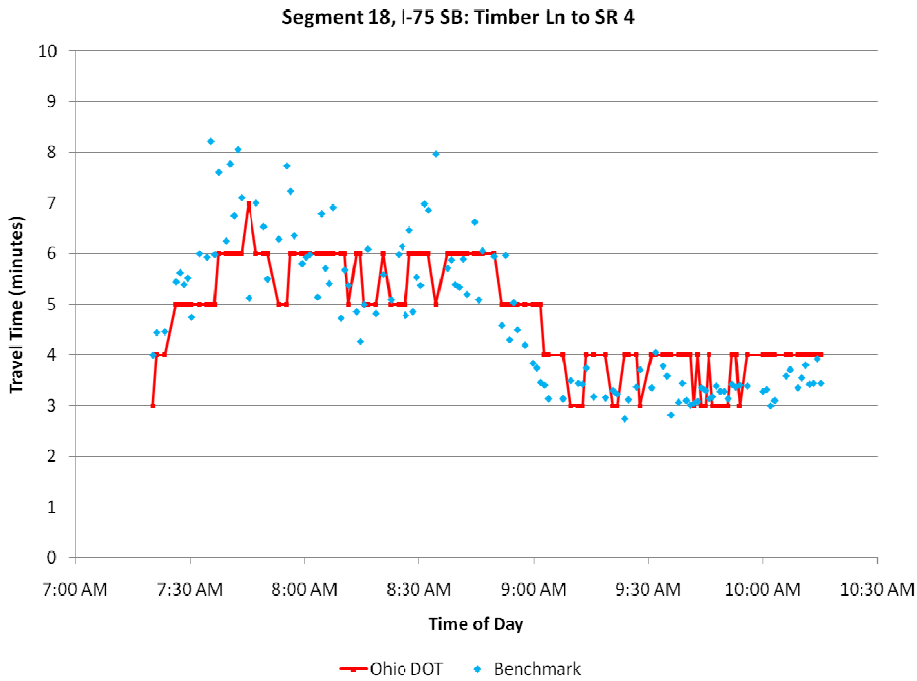


Figure 13 Illustrative Example of Accuracy for Route Travel Time
 (Source: *Statistical Validation of Speeds and Travel Times Provided by a Data Service Vendor*⁸)

⁸ Report is available at http://www.dot.state.oh.us/Divisions/TransSysDev/Research/reportsandplans/Reports/2010/Traffic/134416_FR.pdf, accessed June 2, 2011.

Calculating TIS Route Travel Times with TIS-Provided Link Information

In some evaluations, the available TIS data may consist only of link speeds and travel times and not an aggregated route travel time. Or, there may be multiple short links that should be combined for the purposes of accuracy evaluation (like on arterial streets). In this scenario, travel times for multiple links must be combined into a single route travel time. This is accomplished using a path-based calculation that accounts for the changing traffic conditions as a hypothetical vehicle travels along the route being evaluated. The following calculation procedures have been adapted with permission from the University of Maryland in their validation work for the I-95 Corridor Coalition.⁹


Estimation of the equivalent path-based speeds

In order to compare individual TIS link speeds with a benchmark route travel time, one equivalent TIS route travel time should be estimated. The TIS route travel time estimation may be based on either the observed entrance or exit time of the vehicle to/from the path. For the sake of simplicity in illustration, though, the initial example given herein and the proposed algorithm are based on the exit times. However, in the final example provided with real-data, both forward and backward procedures are applied and results are reported.

First, the time period to which the end time of a Bluetooth match belongs should be identified. Given the distances and speeds of each segment in that time period an average travel time for each segment can be calculated. Backtracking based on these travel times we can determine where the vehicle has been at the beginning of the current time period. Obviously, in the case that vehicle has been midway on the path the procedure can be repeated with data from previous period until it is determined that vehicle has entered the path during the current time period. The concept is illustrated through an example in Table 10.

⁹ See “Estimation of Travel Times for Multiple TMC Segments, February 2010” at <http://www.i95coalition.org/i95/Projects/ProjectDatabase/tabid/120/agentType/View/PropertyID/107/Default.aspx>, accessed June 2, 2011.

Table 10 Time-Space Diagram to Illustrate Calculation of Path-Based Travel Time



TMC Path Sequence	0.1	45(0.13)	30(0.2)	40(0.15)	25(0.24)
	0.2	30(0.4)	40(0.3)	25(0.48)	45(0.27)
	0.3	15(1.2)	30(0.6)	45(0.4)	25(0.72)
	0.4	25(0.96)	15(1.6)	30(0.8)	45(0.53)
		8:00	8:05	8:10	8:15
		Time Interval			

Table 10 shows an aggregated space-time diagram of a path comprised of 4 smaller segments. Segments are 0.4, 0.3, 0.2 and 0.1 mile long, which together form a one-mile path. Time periods are each 5 minutes long. Sample TIS link speeds (mph) for each segment and each time period is shown in the corresponding cells in this figure followed by corresponding average travel times inside the parentheses in minutes. Two reidentification sensors are placed at either end of this longer path. If we have a reidentification match with exit time at 8:11, then we know it belongs to the fourth time period above. In the one minute time that this vehicle has spent in the fourth time period it has covered the last two segments and parts of the second one:

- $1 - 0.24 \geq 0$, thus our vehicle has totally covered the fourth segment during this period,
- $1 - 0.24 - 0.27 \geq 0$, thus our vehicle has totally covered the third segment during this period,
- $1 - 0.24 - 0.27 - 0.72 < 0$, thus our vehicle has spent $1 - 0.24 - 0.27 = 0.49$ min on the second segment during this period. This translates to $0.3 * 0.49 / 0.72 = 0.2$ mile of the second segment. In other words at the beginning of this period, our vehicle has already covered $0.3 - 0.2 = 0.1$ mile of the second segment,

Given the average speed on the second segment in the third time period, our vehicle would have required $0.1 / 0.3 * 0.4 = 0.13$ min to cover 0.1 mile on the second segment,

- $5 - 0.13 \geq 0$, thus our vehicle has covered the 0.1 mile on the second segment during the previous period,
- $5 - 0.13 - 0.8 \geq 0$, thus our vehicle has totally covered the first segment during the previous period.

In total, our vehicle has spent $1+0.13+0.8=1.93$ min to go through the segments. This is the equivalent travel time estimated using average reported speeds on each segment which results in an average travel speed equal to $(0.4+0.3+0.2+0.1)/1.93*60=31.0$ mph for our sample vehicle.

Accordingly, for the purpose of estimating equivalent travel times the following algorithm is proposed:

Given a sequence of segments $i = 1, \dots, N$ forming a path, each with length $l(i)$, and speed $S(i, t)$ at time period t :

For each valid reidentification observation pair $(StartTime, EndTime)$ ending at the same time period $\{T: (T - 1) * 5 < EndTime \leq T * 5\}$, let: $TE = mod(EndTime, 5)$; $n = N$; $t = T$; $PTT = 0$;

While (True)

 If $TE \geq \sum_{i=1}^n TT(i, t)$ then

$PTT = PTT + \sum_{i=1}^n TT(i, t)$; Break.

 Find segment j for which, $TE \geq \sum_{i=j+1}^n TT(i, t)$ and $TE < \sum_{i=j}^n TT(i, t)$

$PTT = PTT + TE$;

$TT(j, t - 1) = TT(j, t - 1) * \frac{\sum_{i=j}^n TT(i, t) - TE}{TT(j, t)}$

$TE = 5min$; $n = j$; $t = t - 1$;

Loop while,

Repeat.

Compute the equivalent average path travel time, \overline{PTT}

Compute the equivalent average speed for comparison with benchmark speed:

$$S_{TIS}(T) = \frac{\sum_{i=1}^N l(i)}{\overline{PTT}}$$

Example Using Real Data

To further illustrate this method, an example with real field data along I-95 is given below. Two consecutive short TMC paths (103-04311 and 103N04311) are considered. The TIS provider reported speeds and average travel times (minutes) on these two segments are shown in Table 11. Please note that times are in UTC.

Table 11 Example Field Data

TMC Path Sequence	0.19	49(0.24)	49(0.24)	46(0.25)	7(1.65)	6(1.93)
	0.57	49(0.70)	49(0.70)	46(0.74)	7(4.89)	6(5.70)
		21:25	21:30	21:35	21:40	21:45
		Time Interval				

To account for deviation from standard TMC path lengths during reidentification sensor deployments, adjusted travel times reported in Table 12 will be used in all ensuing calculations.

Table 12 Example Field Data with Link Length Adjustments

TMC Path Sequence	(0.23)	49(0.28)	49(0.28)	46(0.30)	7(1.97)	6(2.30)
	(0.52)	49(0.64)	49(0.64)	46(0.68)	7(4.46)	6(5.20)
		21:25	21:30	21:35	21:40	21:45
		Time Interval				

Table 13 shows the outcomes of applying the method on a reidentification sample and summarizes the procedure to estimate equivalent TIS speeds when both forward and backward algorithms are utilized.

Table 13 Sample Calculations for Multiple-Link Path-Based Travel Time and Speed

Time Period	MAC ID	Start Time	End Time	BT TT		BT Speed		TIS Equivalent TT (sec)		TIS Equivalent Speed (mph)			
				(sec)	(mph)	Agg.	Forward	Backward	Forward	Agg.	Backward	Agg.	
21:30	00:22:65:F2:90:87	212456	212537	41	66		55	55	49		49		
	00:23:7A:C1:7D:84	212624	212709	45	60		55	55	49		49		
	00:13:6C:4B:95:D8	212639	212725	46	59	61	55	55	49	49	49	49	
	00:0E:9F:22:4A:42	212807	212857	50	54		55	55	49		49		
	00:25:67:DF:56:0A	212848	212928	40	68		55	55	49		49		
21:35	00:1D:F6:9E:70:83	213200	213426	146	18	18	59	59	46	46	46	46	
21:40	00:1C:43:09:1B:C0	213252	213502	130	21		59	61	46		44		
	00:1F:CD:66:AF:7C	213359	213629	150	18	16	59	134	46	15	20	17	
	00:13:6C:95:CB:11	213430	213746	196	14		220	200	12		14		
	00:0F:E4:53:82:10	213542	213842	180	15		386	247	7		11		
21:45	00:05:4F:42:37:AD	213817	214020	123	22	22	433	389	6	6	7	7	

Other Accuracy Measures

Although they are less commonly used and reported, there are several other accuracy measures that can be calculated for the link speed and route travel time scenarios (Table 14).

Table 14 Other Accuracy Measures

Accuracy Measure	Strengths	Limitations
<p>Average percent error (bias)</p> $\frac{1}{n} \sum_{i=1}^n \frac{x_i - \mu_i}{\mu_i}$	<ul style="list-style-type: none"> • Simple measure, easy to calculate • Indicates whether there is consistent bias • Percentage value permits comparison of accuracy on different link lengths 	<ul style="list-style-type: none"> • Large positive and negative errors cancel each other, providing a misleading estimate of the error magnitude • At low speeds, small changes in magnitude can create large changes in percent difference.
<p>Root mean square percent error</p> $\sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \mu_i}{\mu_i} \right)^2}$	<ul style="list-style-type: none"> • Relatively easy to calculate, standard output for data analysis software • Provides greater weight to higher magnitude error, even in a limited quantity • Percentage value permits comparison of accuracy on different link lengths 	<ul style="list-style-type: none"> • Underlying concept not well understood or related to by non-technical audiences
<p>Where: μ_i is the ith benchmark value x_i is the ith TIS estimate n is the number of observations for each unique link/route, date, and time period</p>		

5. RESOURCES AND EXAMPLE APPLICATIONS OF THE GUIDELINES

Resources

In addition to the technical guidance provided in the first four sections of these *Guidelines*, it often is helpful to have specific examples to review. The following list of actual field travel time evaluations can be useful in developing an evaluation plan specific to your needs.

- **I-95 Corridor Coalition – Vehicle Probe Project**
<http://www.i95coalition.org/i95/Projects/ProjectDatabase/tabid/120/agentType/View/PropertyID/107/Default.aspx>, select “Highlights” tab.
 - The University of Maryland has prepared extensive documentation and evaluation results for a real-time data feed along the I-95 corridor in several states (2008-2011).
 - The I-95 evaluation activities meet all of the “should” recommendations in these *Guidelines*. In fact, several of the “should” or “shall” recommendations were adapted directly from the I-95 evaluation activities as a best practice in these *Guidelines*. The difference is that these *Guidelines* provide a comprehensive menu of practices and approaches. For example, the I-95 evaluation effort uses two specific accuracy measures, whereas these *Guidelines* describe several other accuracy measures and their most suitable application (for example, if you want accuracy for travel time instead of speed).
- **Maryland DOT – Baltimore Multimodal Traveler Information System (MMTIS)**. Ali Haghani, Saini Yang, Masoud Hamed, "Cellular Probe Data Evaluation, Case Study: The Baltimore Multimodal Traveler Information System (MMTIS)", Technical Report, Submitted to Maryland Department of Transportation, January 2007. (e-copy available)
 - The University of Maryland evaluated a cell phone-based traveler information system.
- **Ohio DOT – Statistical Validation of Speeds and Travel Times Provided by a Data Service Vendor**
http://www.dot.state.oh.us/Divisions/TransSysDev/Research/reportsandplans/Reports/2010/Traffic/134416_FR.pdf
 - The University of Akron and Texas Transportation Institute evaluated travel time accuracy on selected freeways and expressways in Dayton, Ohio. (January 2010)
- **Maricopa County DOT and Arizona DOT – Travel Time Estimates, Displays, and Forecasts – Final Report**, Technical Report #2, Prepared for Maricopa County DOT and Arizona DOT by OZ Engineering and Motion Maps, December 2004. (e-copy available)
- Eisele, W.L. and L.R. Rilett. Travel-Time Estimates Obtained from Intelligent Transportation Systems and Instrumented Test Vehicles: Statistical Comparison. In *Transportation Research Record 1804*, TRB, National Research Council, Washington, D.C., 2002, pp. 8-16.
- Chan, K.S., M.L. Tam, M.L. Tam, and W.H.K. Lam. *Using Automatic Vehicle Identification Data for Estimating Current Travel Times in Hong Kong*. Paper submitted for presentation at the Transportation Research Board Annual Meeting, January 2007.

- Bertini, R.L. and S. Tantiyanugulchai. Transit Buses as Traffic Probes: Use of Geolocation Data for Empirical Evaluation. In *Transportation Research Record 1870*, TRB, National Research Council, Washington, D.C., 2004, pp. 35-45.
- Li, R., G. Rose, and M. Sarvi. Evaluation of Speed-Based Travel Time Estimation. *Journal of Transportation Engineering*, Vol. 132, Issue 7, July 2006, pp. 540-547.

Overview of Example Applications

This section contains hypothetical examples of how the guidelines may be applied to evaluate the quality of traveler information. The data presented in these examples are hypothetical but are representative of real-world applications. In some cases, examples are taken from existing applications for the purposes of demonstrating the use of these *Guidelines*.

1. A state DOT collects real-time traffic data from their own sensors for the state highway network within the largest city in their state. Using this data, the state DOT provides an online color-coded speed (red-yellow-green) map and also posts travel times on dynamic message signs along major routes. The state DOT would like to ensure the accuracy of the color-coded speed map and the travel time messages.
2. An MPO purchases real-time traffic data from a private data provider for 15 major routes (300 centerline miles) in their region. Using this data, the MPO provides an online speed map that shows exact link speeds (to nearest mph). The MPO also provides point-to-point travel times for major interchanges through 511 and their website. The MPO would like to ensure the accuracy of their link speed values and their point-to-point travel times.
3. Evaluation of travel time and speed data on arterial links has been challenging. Estimating a benchmark travel time on an arterial facility typically requires larger sample sizes than freeways due to greater variance in the data, more detection and filtering of outliers, and longer sampling of time intervals due to the lower volumes. This example will present a discussion of these challenges.

Example #1

A state DOT collects real-time traffic data from their own sensors for the state highway network within the largest city in their state. Using this data, the state DOT provides an online color-coded speed (red-yellow-green) map and also posts travel times on dynamic message signs along major routes. Note that the source of this data does not drive this example – the traffic data supporting the applications could be purchased from a travel data service provider. The state DOT would like to ensure the accuracy of the color-coded speed map and the travel time messages.

Introduction

This example will show in detail how to develop an evaluation plan, how to collect and reduce benchmark data, and finally how to apply error metrics to evaluate the data. To illustrate some of the details of the process we will refer to a study conducted by the Ohio Department of Transportation to evaluate travel time and speed data in Dayton, Ohio (ODOT, 2009). In this study, benchmark travel times were collected on 103 centerline miles in Dayton using a combination of floating car and Bluetooth reidentification. In this example, the DOT is considering how to make the best use of limited resources and therefore must use a combination of floating car runs and Bluetooth reidentification. The acceptable error for travel times has been set by the DOT to be 10% of the relevant benchmark value. Similarly, for the evaluation of the color coded speed map the evaluation will consider up to a 10% misclassification rate acceptable.

Task 1: Develop an Evaluation Plan

Section 2 of these *Guidelines* presents a systematic approach to developing an evaluation plan. In this use case, we will follow this approach to develop an evaluation plan for the Dayton case study. Since the state DOT is evaluating both a color coded speed map as well as travel times on dynamic message signs we should plan to collect benchmark data for both evaluations. The following steps should be followed to develop the evaluation plan.

Step 1: Roadway Segmentation In this step we segment the network to align with the intended applications. There are 103 centerline miles in Dayton where travel times are provided. Figure 14 shows an overview map of the Dayton study area. Sections 2 and 3 of these *Guidelines* lay out the recommended approach to segment a network. Here, we will show how this approach was applied in the Dayton case study.

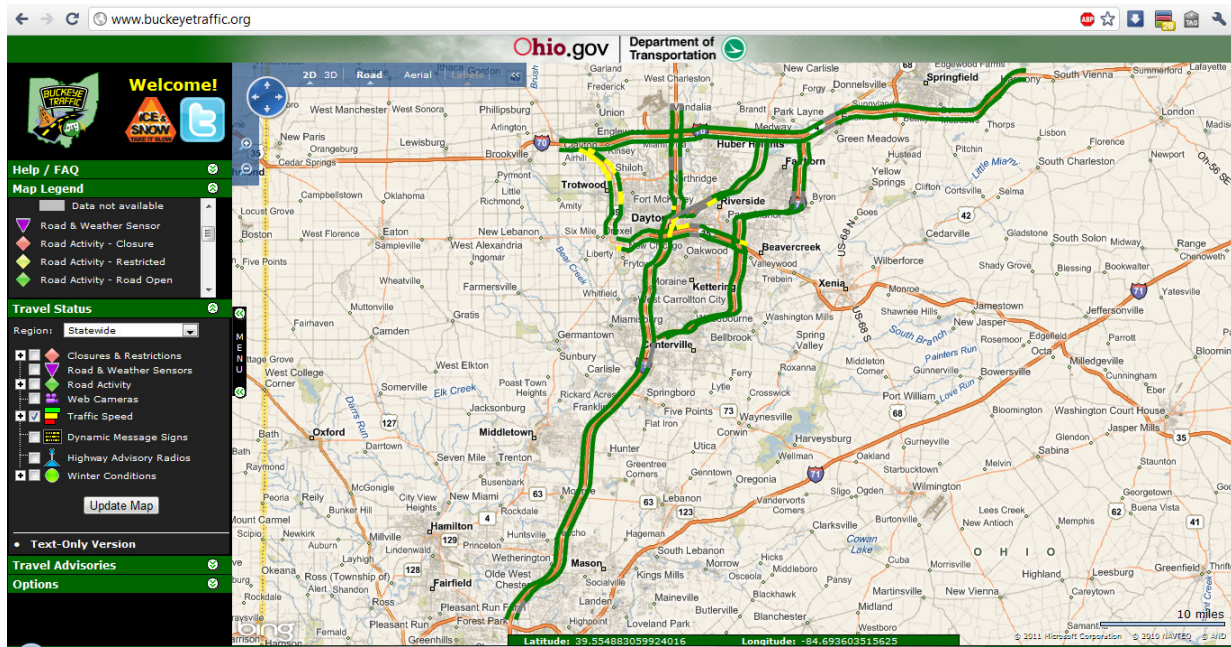


Figure 14 Dayton Study Area (www.buckeyetraffic.com)

A total of 36 segments were defined in this study. These 36 segments cover 100 of the 103 centerline miles in the evaluation area. A mix of freeways and arterials were covered. In the Dayton study the end points for each link were determined based on definitions provided by ODOT. In general, the link end points can be determined by aligning the benchmark link definitions with the links being evaluated. In cases where long travel time routes are being evaluated it is recommended that the route be divided into shorter benchmark links. The guidelines suggest that optimal freeway benchmark link length is between 1 – 2 miles in length and arterial link length is between $\frac{1}{2}$ and 1 mile. Typically, benchmark link end points are positioned close to freeway access points or at signalized intersections on arterials. When using reidentification techniques it is sometimes necessary to offset the link end point due to safety concerns or infrastructure constraints. Information on optimal link node positioning is provided in Section 3. Figure 15 shows an example of the positioning of segment end points for I-70 in Dayton.

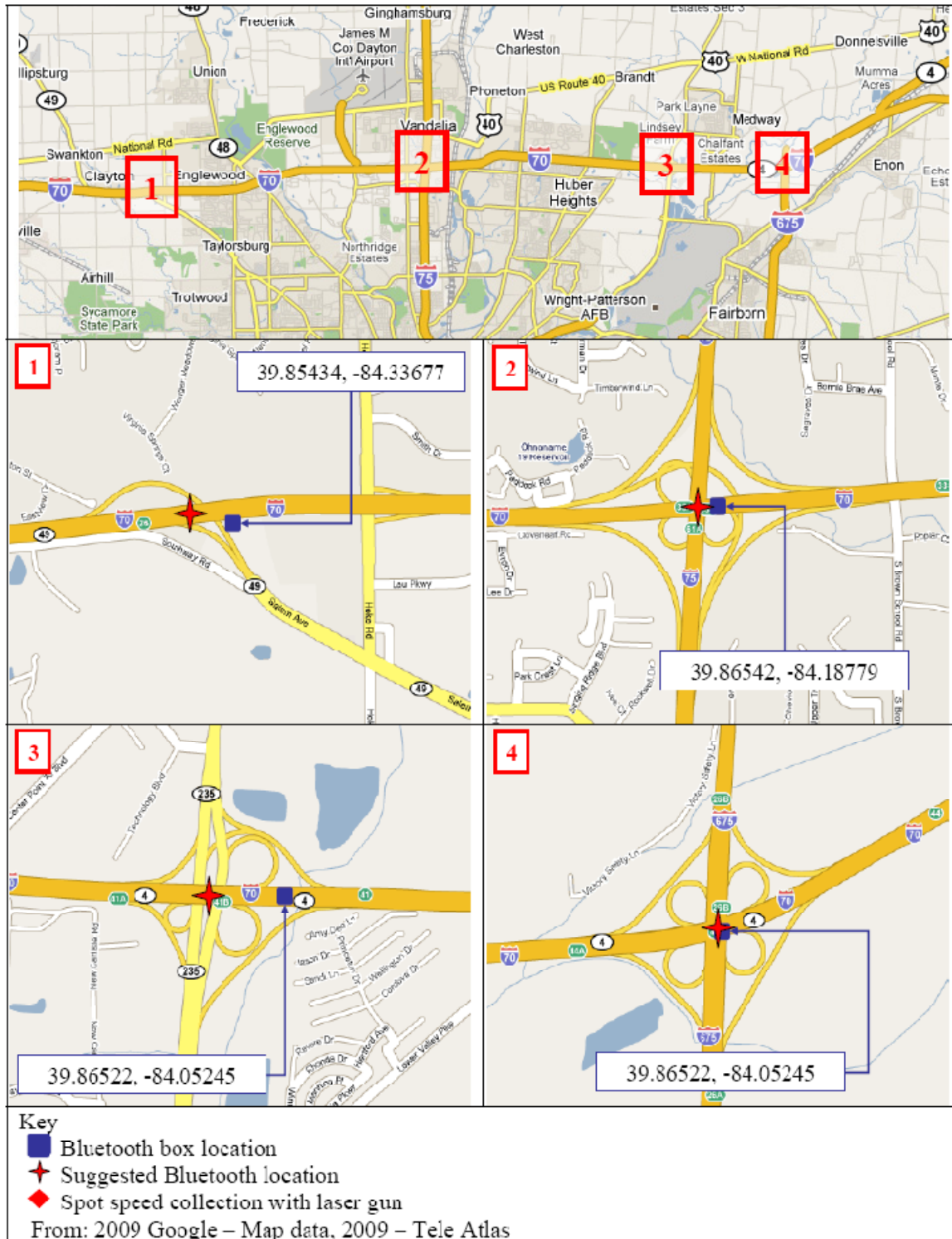


Figure 15 Example of Segmentation in Dayton

Table 15 shows the results of the segmentation process for I-70. Each segment is clearly defined using the relevant milepost numbers for the facility and the length of the segment. One thing to consider, however, when defining segments is that the segment length will have an effect on benchmark data

collection. For example, in the Dayton study, a number of segments were close to 10 miles in length. While this is a feasible length for floating car studies, it is longer than the recommended length for reidentification. There are a few problems with long segments that can be avoided by benchmarking shorter segments. The first problem is that the percentage of vehicles that can be reidentified will decrease as the segment length increases. This is due to vehicles exiting the traffic stream. Second, the number of outlier data points will tend to increase due to vehicles either stopping along the way or possibly exiting and re-entering the traffic stream. And third, as the length of the segment increases the travel time variance tends to decrease. This can cause an evaluation to possibly overlook problem areas in the network. These *Guidelines* recommend that benchmark segments not be greater than 5 miles in length when using reidentification on freeways and 3 miles on arterials.

Table 15 Example of Segment Definitions from Dayton

Segment ID	Road Name	Description	Segment Length (mi)	Start Milepost	End Milepost
1	I-70	SR 49 to I-57 E	8.0	25.9	33.9
3	I-70	I-75 to SR 4 (South) E	7.2	33.9	41.1
5	I-70	SR 4 (South) to I-675 E	3.2	41.1	44.3
7	I-70	I-675 to SR 4 (Enon) E	2.9	44.3	47.2
2	I-70	SR 49 to I-57 W	8.0	25.9	33.9
4	I-70	I-75 to SR 4 (South) W	7.2	33.9	41.1
6	I-70	SR 4 (South) to I-675 W	3.2	41.1	44.3
8	I-70	I-675 to SR 4 (Enon) W	2.9	44.3	47.2

Step 2: Time Intervals The time intervals for benchmark data collection should correspond with the frequency update intervals of the applications being evaluated. In this example, we will assume that the color coded speed map and the travel time messages are being updated at 5 minute intervals. Therefore, benchmark data will be reduced to a 5-minute aggregation level.

Step 3: Stratify Network Network stratification is used to group together segments with similar traffic flow characteristics so as to reduce the overall variance in the evaluation data. The Dayton study evaluated both arterial and freeway segments. These *Guidelines* recommend first stratifying by facility type and further stratifying freeway segments by expected variance. The guidelines recommend that reidentification always be used for benchmarking arterials. However, for benchmarking freeways there is some flexibility. Travel time segments that are classified as low variance can be benchmarked using a floating car method whereas high variance segments should be benchmarked using reidentification. To illustrate this method, two hypothetical segments will be evaluated. Let’s assume that the segments have the following characteristics:

Table 16 Example of Network Stratification

Segment ID	ADT	Lanes	ADT / Lane	Access Points	Length (mi)	Access Density	Classification
Ex1	65,000	4	16,250	4	3.3	1.21	Low Variance
Ex2	120,000	4	30,000	6	2.9	2.06	High Variance

The ADT / lane and access point density of the first segment are $65,000 / 4 = 16,250$ and $4 / 3.3 = 1.21$ and therefore this segment would be classified as a “low variance” segment. The ADT / lane and access point density of the second segment are $120,000 / 4 = 30,000$ and $6 / 2.9 = 2.06$ and therefore this segment would be classified as “high variance”.

Step 4: Select Benchmark Data Collection Methods The results from Step 3 can be used to help determine an appropriate data collection method for each benchmark segment. In the Dayton study, both reidentification and floating car techniques were used to collect benchmark data on 34 out of the 36 monitored segments in the evaluation. The Dayton study used floating car as the primary benchmark value and used Bluetooth reidentification data to validate the observed speeds obtained from the floating car runs. These *Guidelines* do not endorse one specific technique for benchmark data collection, but encourage the use of methods that can collect large samples when travel time variance is greater. The guidelines also recognize reidentification as a valid source for benchmark data. The guidelines endorse the use of reidentification as a means of generating confidence intervals for mean speed and travel time. This is especially important where travel time variance can reasonably be expected to be relatively high. Thus, the stratification method recommends that both arterials and high variance freeway segments be monitored using a reidentification technique. Refer to Section 2 and 3 for more detail on this issue.

Step 5: Time Periods to Evaluate Benchmark data should be collected during peak travel periods. If it is known that a segment is more heavily traveled during a particular peak period (e.g. AM Peak) then data collection should be conducted during the heaviest peak period. Bluetooth reidentification will collect data over multiple periods (usually several days at a time). However, the floating car runs should be scheduled during peak travel periods. In the Dayton study, benchmark data collection was scheduled during the peak travel periods for that region. Figure 16 shows the time periods evaluated in the Dayton study.

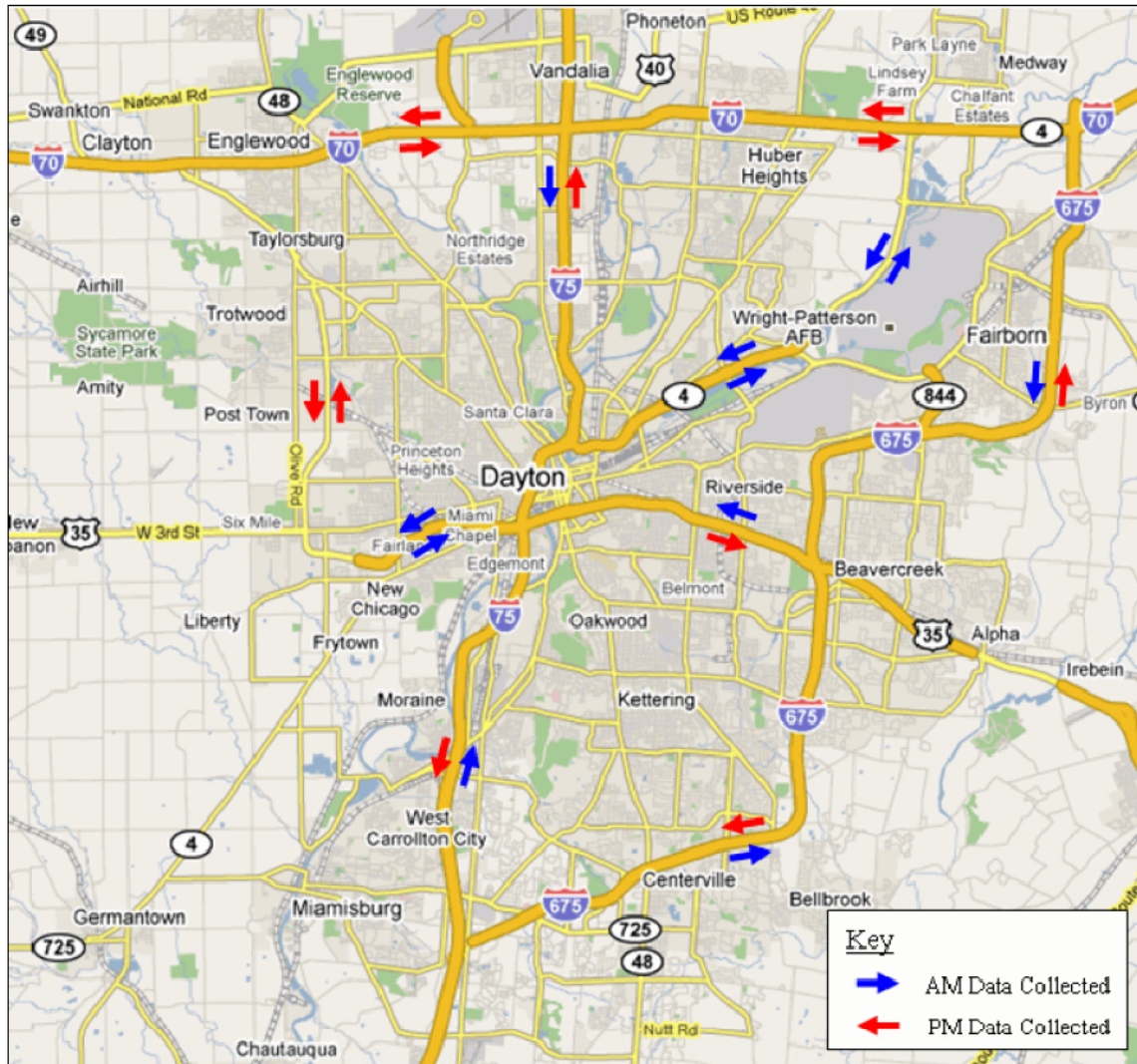


Figure 16 Data Collection Time Periods in Dayton

Final Evaluation Plan The Dayton study analyzed speed and travel time accuracy over 36 segments and 103 centerline miles in the Dayton region. In this example we examined how to segment the network, choose an appropriate data collection time interval, when to collect benchmark data, and how to choose an appropriate benchmarking technology. Next, we will look at how to collect and reduce benchmark data for this study.

Task 2: Collect and Reduce Benchmark Data

In Task 1, we focused on the methodology for segmenting the network, determining appropriate time intervals for data collection, and determining appropriate data collection methods. In this task we will focus on the collection and reduction of benchmark data for the two example segments described in Task 1.

Segment Ex1: Test Vehicle Technique

Data Collection Here, we refer to *Section 3: Collect and Reduce Benchmark Data* and focus on the *Section: Test Vehicle Data Collection Methods*. Following the guidelines presented in *Section 3*, we select an appropriate GPS technology to instrument the test vehicle. The driver of the vehicle has been instructed as to how to drive the vehicle at the average speed of the traffic stream. In the Dayton study, the GeoLogger GPS unit from GeoStats was used to record benchmark data. The GPS unit logs position, instantaneous speed and time at one second intervals. After data are collected they are downloaded to a computer for processing. More detailed information on collecting travel time data using a GPS can also be found in the *Travel Time Data Collection Handbook* (FHWA-PL-98-035, 1998).

Data Reduction After the data from the GPS has been extracted and loaded into a computer for analysis, the benchmark values may be computed. However, first a visual inspection of the GPS data should be conducted to ensure that the data conforms to the intended segment. This can be accomplished by importing the data into a Geographic Information System (GIS) such as ArcMap. Once the data has been visually verified it should also be inspected for any outliers in time. If there are unreasonable or unexplainable deviations in the times recorded by the device the data should be discarded and another run should be conducted. Average speeds can be calculated by computing the distance traveled and the time taken to travel the distance. To illustrate, average speed calculations for Segment #Ex1 are shown in the next table.

Table 17 Example of Floating Car Data Collection

Start Time	End Time	Travel Time (s)	Distance Traveled (mi)	Average Speed (mph)
09:10:05	09:13:05	180	3.3	66
10:15:25	10:19:00	215	3.3	55
16:20:11	16:25:40	329	3.3	36
17:15:40	17:22:05	385	3.3	31

Segment Ex2: Reidentification Technique

Data Collection Here, we refer to *Section 3: Collect and Reduce Benchmark Data* and focus on the *Section: Reidentification Data Collection Methods*. The technology selected for this application is Bluetooth reidentification. We will be using devices capable of recording MAC addresses from Bluetooth communication devices and deploy these devices at the end points of Segment Ex 2. The devices are running on battery power and are capable of logging data for a total of 3 days. An example of data collected from the devices is shown in the next table.

Table 18 Example of Travel Time Data Collection

MAC Address	Time at Point 1	Time at Point 2	Travel Time (sec)	Distance Traveled (mi)	Average Speed (mph)
11:DD:22:EE:FF	10:30:25	10:33:20	175	2.9	60
99:XY:88:ZZ:YY	10:30:40	10:33:28	168	2.9	62
...
22:RR:55:TT:SS	16:24:55	16:30:05	310	2.9	34
77:QQ:00:PP:CC	16:25:05	16:30:02	297	2.9	35

To produce this table, it is usually necessary to have specialized computer software capable of merging the output from each Bluetooth recorder and matching the MAC addresses to get the time at each end of the traveled segment. The travel time is then calculated by differencing these times and the average speed (i.e. space mean speed) is computed from the distance traveled and the travel time.

Data Reduction Since we are using 5-minute intervals to evaluate the travel time messages, the data from the Bluetooth recorders should be reduced using a 5-minute aggregation. This can be accomplished by using a simple computer programming language capable of data manipulation such as a database management system (e.g. Microsoft Access or Oracle) or a spreadsheet program (e.g. Microsoft Excel). Typically the data are aggregated according to the time that the vehicle started traveling the segment (i.e. Time at Point 1). The aggregation of the data should generate two important statistics for each time period.

- 1) Average Travel Time
- 2) Standard Deviation of Travel Time

The average travel time and the standard deviation of the travel time can then be used to compute a statistical confidence interval which will give bounds on the true average travel time for the time period being evaluated. Please refer to Equation 7 for details on computing the confidence interval.

To illustrate this calculation, let us assume that the average travel time for this segment during the evaluated time period was 190 seconds (i.e. 55 mph), with a standard deviation of 19 seconds and that the sample size was 10 observations. Inserting these values into the equation for a confidence interval produces the following calculation:

Equation 17 $Confidence\ Interval = \bar{x} \pm t_{(n-1, \alpha/2)} \times \left(\frac{s}{\sqrt{n}}\right)$

$$190 - 2.26 * \frac{19}{\sqrt{10}} < \mu < 190 + 2.26 * \frac{19}{\sqrt{10}}$$

Where:

2.26 = $t_{(n-1, \alpha/2)}$ = Student's T statistic ($\alpha=0.05$ or 95% confidence)

190 = average travel time = \bar{x}

19 = standard deviation travel time = s

10 = number of observations = n

Therefore, the “true” average travel time for this segment during this time period lies between 176 and 203 seconds. We can convert the confidence interval for travel time to a confidence interval for space mean speed by dividing the segment length by the upper and lower bounds of the travel time. Thus, the “true” average speed for the segment lies between 51 and 59 mph. This confidence interval then tells us that approximately 95% of the time we would have observed an average speed between 51 and 59 mph.

The process of calculating a confidence interval should be repeated for each sampled segment and time period. Therefore we should produce a table like the following.

Table 19 Example of Data Reduction for Travel Time Samples

Segment ID	5 minute sampling interval	Average Travel Time (secs)	Standard Deviation of Travel Time (secs)	Sample Size	Upper Bound Travel Time (secs)	Lower Bound Travel Time (secs)
Ex2	10:00:00 – 10:05:00	190	19	10	203	176
Ex2	10:05:00 – 10:10:00	180	15	12	190	170

Task 3: Select and Apply Error Metrics

In Task 2, we focused on data collection and reduction for the two example segments. We collected data using floating car and reidentification techniques. In this task we will focus on selecting and applying error metrics. The selection of an error metric depends on the evaluation and also on the types of errors that are to be detected. For example, to evaluate speeds on a color coded map, these *Guidelines* recommend using percent correct classifications. This error metric will report how well the service did at classifying speeds. An evaluation of travel times might use the mean absolute percent error (MAPE). This error metric will report how far, on average, the service deviated from the benchmark values as a percentage. For example, a MAPE of 10% would indicate that on average it could be expected that reported travel times would deviate from actual travel times by no more than 10% of the true average trip. Thus, a trip that, on average, took 10 minutes could be off by about 1 minute. While the guidelines do not recommend a specific error metric it should be noted that error metrics which measure the absolute magnitude of an error can be somewhat misleading. If the average absolute error for a segment is computed to be 2 minutes but the average travel time is 5 minutes this would be equivalent to a 40% error. Because we want to measure relative error in speeds and travel times, we will be using the Percent Correct Classifications and Mean Average Percent Error as the error measures for this evaluation.

Segment Ex1: Color Coded Speed Map Accuracy

Benchmark data for this segment was collected by a floating car run. The benchmark data will be used to evaluate the accuracy of the color coded speed map application. We will refer to *Section 4: Congestion Category Accuracy Measure*.

The error metric that we will use to evaluate the color coded speed map is given in Equation 18.

Equation 18 *Percent of correct classifications (in each congestion category)* = $\frac{\text{Number of correct classifications}}{\text{Total number of comparisons}}$

However, to avoid inflating the score by including evaluations during free flow traffic conditions, we will only evaluate classifications during congested conditions. The equation for evaluating a classification is given in Equation 19.

Equation 19 *Correct classification when* $x_{lower} - \theta \leq \bar{x}_i \leq x_{upper} + \theta$

Now, let’s assume that we define a “yellow” classification as speeds between 30 – 50 mph and a “red” classification as speed between 0 – 30 mph. Further, let us allow for a small amount of error in the observation and set our “theta” value equal to 5 mph. Therefore, if the benchmark observation lies between 25 and 55 mph and the map displays yellow during this observation, we will score that as a correct classification. Similarly if the benchmark observation lies between 0 – 35 mph and the map is displaying “red” we will score this as a correct classification. The total number of correct classifications is divided by the total number of classifications to compute the percent of correct classifications. Table 19 shows this calculation for a number of segments.

Table 20 Example of Speed Classifications

Observation Time	Benchmark Speed (mph)	Displayed Speed Category	Theta (mph)	Correct Classification?
10:00 AM	35	Yellow	5	Yes
10:30 AM	45	Yellow	5	Yes
11:00 AM	24	Yellow	5	No

The percent of correct classifications would then be $2 / 3 = 67\%$.

Segment Ex2: Travel Time Message Accuracy

In this example, we will assume that benchmark data was collected using Bluetooth reidentification. The benchmark data will be used to evaluate the accuracy of travel time messages posted on a message sign. We will refer to *Section 4: Route Travel Time Accuracy Measures*. For this application, we want to ensure that the average absolute error of the posted travel times does not exceed 10%. We also would like to determine whether the posted times tend to under-estimate or over-estimate the actual travel times. Therefore, we will apply two different error metrics:

- 1) Mean Absolute Percent Error (MAPE)
- 2) Error Bias

The MAPE is calculated by averaging the absolute percent error of each observation period. It is described in *Section 4, Route Travel Time Accuracy Measures*. The equation for MAPE is given in Equation 20 and shown on the next page.

Equation 20 $Average\ absolute\ percent\ error\ (\%) = \frac{1}{n} \sum_{i=1}^n \frac{abs(x_i - \mu_i)}{\mu_i}$

where μ_i = benchmark value for the i th comparison (minutes or seconds)
 x_i = the i th TIS estimate (minutes or seconds)
 n = number of estimate-to-benchmark comparisons

We will use a benchmark confidence interval, described in the Data Reduction step to compute the range of acceptable travel times. Any values that fall outside of the computed confidence interval are considered errors and the percent error is calculated using the closest bound as the reference value. To illustrate this calculation, we will use the following data as an example.

Table 21 Example of Error Calculations for Travel Times

Observation # (i)	TIS Travel Time (secs) (x_i)	Benchmark Confidence Interval (secs)	Benchmark Reference Value (μ_i)	Absolute Error (secs)	Absolute Percent Error
1	220	176 – 203	203	17	17 / 203 = .08
2	155	170 - 190	170	15	15 / 170 = .08
...

To calculate the MAPE we take the average of the absolute percent error column. A reasonable average percent error might be 10% or less. This means that on average the TIS displayed travel times within 10% of the true average travel time for all evaluated segment and time periods.

The error bias will tell us whether the TIS tends to over-estimate or under-estimate the true travel times. This error metric can be calculated by simply averaging the errors without using an absolute value function. If the error bias is positive, we can infer that the TIS tends to overestimate travel times. Table 22 shows an example of error bias calculations.

Table 22 Example of Error Bias Calculation

Observation # (i)	TIS Travel Time (secs) (x_i)	Benchmark Confidence Interval (secs)	Benchmark Reference Value (μ_i)	Error Bias (secs)
1	220	176 – 203	203	17
2	155	170 - 190	170	-15
...

The average error bias can be calculated by averaging the observed error biases. In this example the average error bias would be $(17 + (-15)) / 2 = 1$ sec.

Example #2

An MPO purchases real-time traffic data from a private data provider for 15 major routes (300 centerline miles) in their region. Using this data, the MPO provides an online speed map that shows exact link speeds (to nearest mph). The MPO also provides point-to-point travel times for major interchanges through 511 and their website. The MPO would like to ensure the accuracy of their link speed values and their point-to-point travel times.

Introduction

In Example #1, the details of developing an evaluation plan, collecting benchmark data, and applying error metrics was covered. In this example, we want to highlight some different approaches to network segmentation and benchmark data collection. This example will also demonstrate how to apply error metrics for evaluation of link speed data.

Task 1: Develop an Evaluation Plan

Here we will focus only on network segmentation in order to highlight some other possible approaches to network segmentation.

Step 1: Network Segmentation Another common approach to network segmentation is the use of a widely used standard such as the traffic message channel standard, commonly referred to as TMC segments. This approach to network segmentation has been adopted by many of the private sector data providers and is being promoted as a de-facto standard. We will assume in this example that the private data provider has adopted the TMC segmentation standard. In this example, we will also assume that the MPO has access to toll tag data which can be used to calculate point to point travel times.

In this scenario, it is possible that the toll tag readers are not located at end points of TMC segments and that the distance covered between toll tag readers may cover more than one TMC segment. The network can be segmented by constructing travel time segments between toll tag readers. However, in order to fairly compare benchmark data to estimates from the TIS it is then necessary to apply a spatial aggregation method to combine estimates from the TIS into a single estimate that can be compared with the benchmark.

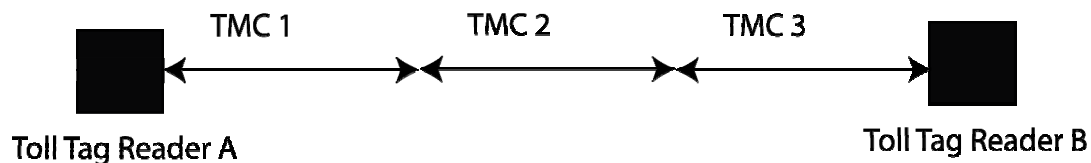


Figure 17 Example of a TMC Segment Chain

Steps 2 – 5 Since we are using toll tag data to determine benchmark values we will assume that data will be available for all time periods (e.g. AM and PM peak periods) and that there is no need to select a data collection method.

Task 2: Data Collection and Reduction

Since we are using travel times based on toll tag readings we can compute a statistical confidence interval following the same methodology outlined in Example #1. This is the benchmark speed or travel time.

However, the calculation of the TIS estimate of mean travel time or speed requires that we combine together the estimates for the three TMC segments (TMC1, TMC2, TMC3). The best way to combine together estimates of smaller segments to create an estimate of a longer segment is by using a distance weighted average. This is a weighted average where the weights are the length of each of the TMC segments. Since we are calculating the space-mean-speed, the average is a harmonic average. The weighted harmonic average can be calculated using Equation 21:

$$\text{Equation 21} \quad \bar{x} = \frac{\sum_{i=1}^n w_i}{\sum_{i=1}^n (w_i/x_i)}$$

where:

\bar{x} is the computed weighted average (space-mean-speed)

w_i is the weight of the i th measurement

x_i is the i th measurement

If we let the weights be equal to the lengths of each TMC segment, we can compute a weighted average for the length of the complete segment between the two toll tag readers. For example, let's assume that the TIS vendor reported the following average speeds for the three TMC segments in this example.

Table 23 TMC Segments

TMC ID	TMC Length (miles)	Report Speed
TMC1	0.9	62
TMC2	1.5	55
TMC3	2.3	50

We can then compute the weighted average as:

$$\text{Equation 22} \quad \bar{x} = (0.9 + 1.5 + 2.3)/(0.9/62 + 1.5/55 + 2.3/50) = 53.5 = 54 \text{ mph}$$

An un-weighted average would be approximately 56 mph. So the weighted average gave slightly more importance to the speed reported for the longer TMC segments and resulted in a lower overall average speed.

To compute benchmark speeds confidence interval, the following procedure should be followed.

- 1) Calculate a confidence interval for the travel time of the segment.
- 2) Divide the distance traveled (miles) by the upper and lower bound for travel time (hours)
- 3) This will give a lower and upper bound for average speed (space mean speed)

Example: If the travel time sample data for a link resulted in the following values:

- Number of observations = 10
- Average travel time = 100 seconds
- Standard Deviation travel time = 10 seconds
- Link Length = 1.5 miles

An alpha = 0.05 confidence interval (see Equation 4-1) would be $100 \pm (2.26 * 10/\sqrt{10}) = 93 - 106$ seconds.

Given a link length of 1.5 miles, the computed confidence interval for space mean speed is then: $(1.5 \times 3600)/106$ and $(1.5 \times 3600)/93$ which is 51 – 58 mph. An example of a confidence interval plot is shown in Figure 18. The lines in blue indicate the confidence interval speeds and the red points indicate the TIS estimates.

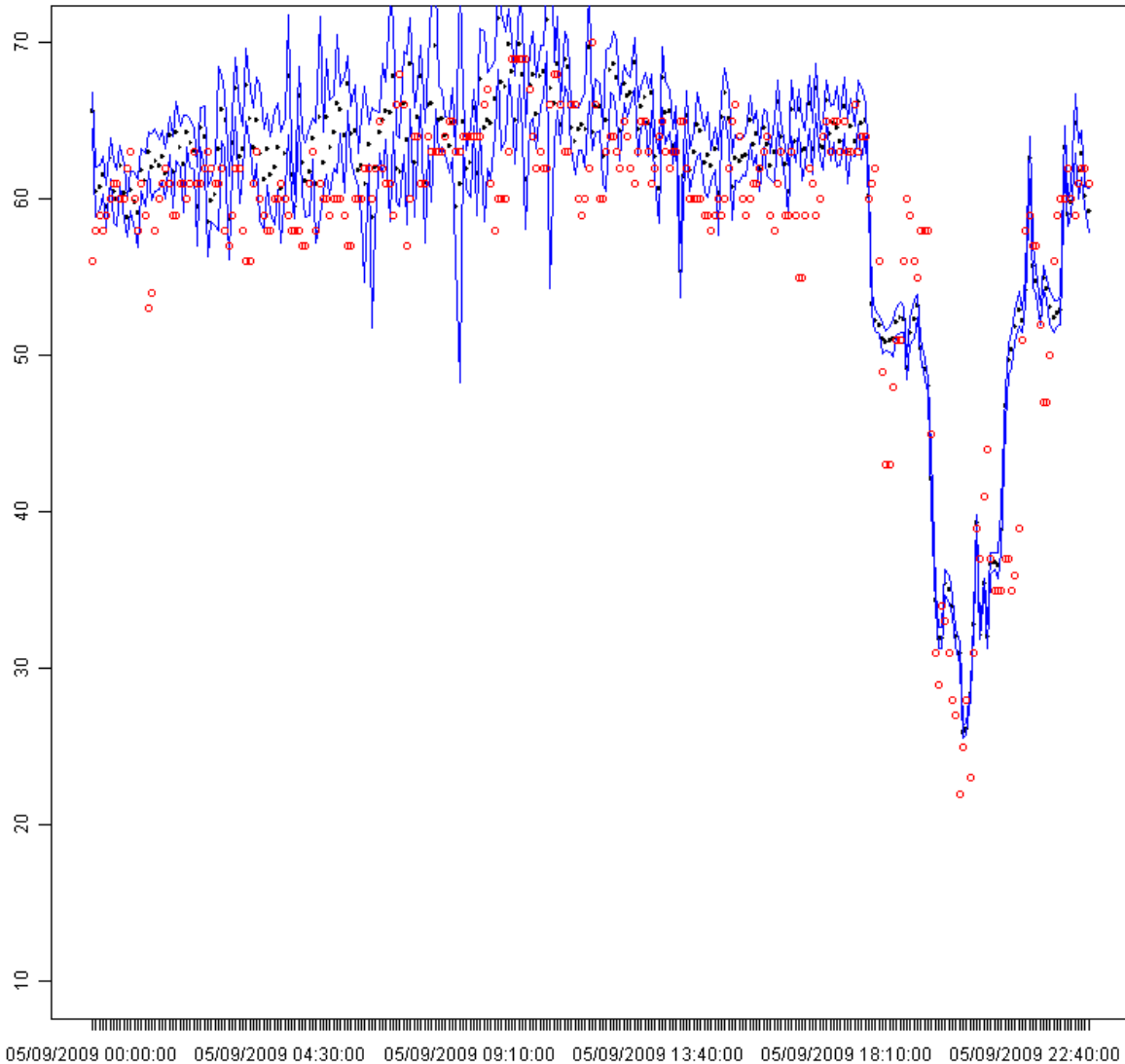


Figure 18 Example of a Speed Confidence Interval

Task 3: Compute Error Metrics

Here, we will focus on computing the error metrics for the link speed estimates. Error metrics for link travel time estimates were shown in Example #1.

In this example, we will use the **Percent of TIS values within X of the Benchmark value**, error metric as described in *Section 4, Link Speed Accuracy Measures*. Let us assume that an error tolerance of 5 mph is

acceptable for our application. Therefore, we will accept any TIS estimates that fall within 5 mph of the confidence interval bounds. To calculate this error metric, first we calculate the confidence interval for each link and time period. Then we compute the weighted average for the TIS using the method described previously. If the weighted average speed falls within 5 mph of the confidence interval we accept this as a correct estimate. Otherwise it is marked as an incorrect estimate. The final percentage is the percentage of estimates that were correct.

Table 24 Example of Error Metric for Link Speed

Time Period	TIS Estimate (mph)	Benchmark Lower Bound (mph)	Benchmark Upper Bound (mph)	Tolerance	Correct?
T1	55	51	58	5	Yes
T2	57	48	55	5	Yes
T3	62	49	56	5	No

The final percentage can be computed by simply summing up the number of correct estimates and dividing by the total number of estimates that were compared. If we have a total of 100 estimates and 80 were marked as correct, we would have an 80% correct estimation rate for the TIS.

Example #3

Evaluation of travel time and speed data on arterial links has been challenging. Estimating a benchmark travel time on an arterial facility typically requires larger sample sizes than freeways due to greater variance in the data, more detection and filtering of outliers, and longer sampling time-intervals due to the lower volumes. This example will present a discussion of these challenges.

Introduction

There are a number of challenging technical issues with statistical estimation of benchmark values on arterial facilities. Many of these issues can be overcome through proper sampling design and data analysis. This example will present some background on the problem as well as some suggestions for approaches to benchmarking arterial facilities.

Arterial facilities present a unique challenge in terms of estimating a benchmark value. Due to the presence of signalized intersections with a multitude of different timing plans, samples of arterial travel time tend to have very high variance. In these guidelines, it is recommended that facilities with high sample variance be monitored using a reidentification technique. However, the design of the sampling plan must be adjusted to deal with some of the specific challenges related to arterial travel.

In addition to the presence of signalized intersections which tend to increase variance in arterial travel times, there are also the issues related to queuing at intersections, turning movements, high access point density, interaction with pedestrians and non-motorized vehicles, and curbside parking. The specific effect of each of these factors on arterial travel times is not yet fully understood. However, there have been some recent efforts to collect arterial travel times and this section will document the findings from these efforts and make some general recommendations on best practices in monitoring arterial facilities.

Findings

Collection of benchmark travel time on arterials may require that the sampling time-interval be increased when compared with the typical 5-minute sampling time-interval on freeways. The increased sampling time-interval allows for collection of more observations which increases the sample size and therefore increases the precision of the benchmark estimate. Additionally, due to lower traffic volumes on arterials the number of vehicles reidentified per time-interval tends to be lower than freeways. Therefore increasing the sampling time-interval is an effective way of improving the precision of a benchmark estimate when sampling on an arterial facility.

The length of the link, the signal density, and the access density are also important considerations when designing an arterial monitoring plan. The interaction of these factors may have an effect on the variance in the observed travel times as well as the number of observations collected. Minimizing variance while ensuring adequate sample observations will improve the precision of any benchmark estimate.

When monitoring arterial links with high signal density or high access point density it may be difficult to collect sufficiently large samples because a percentage of the vehicles identified at the start of the link will leave the traffic stream before they can be reidentified. Selecting links that minimize the number of signals and access points per unit length will ensure that more vehicles remain in the traffic stream and reduce the effect of signals and turning movements on traffic flow. Therefore, when segmenting arterials for benchmark data collection, a balance between link length and access/signal density should be considered.

Another challenge with benchmark data collection on arterials is the presence of outliers in the data. For example, vehicles may exit and re-enter the traffic stream or stop for various reasons. The detection and removal of outliers in arterial data can be more challenging than with freeways. Before attempting to estimate a benchmark value for an arterial link it is important that the sample data be screened for outlier observations. Any outliers should be removed or replaced.

Finally, the distribution of travel times on an arterial is frequently bi-modal. The distribution of arterial travel times is strongly influenced by the cycle of traffic signals. The signals may divide the traffic stream into distinct platoons of vehicles. Therefore, over a given sampling time-interval two distinct distributions of travel times may be observed.

Conclusions

Estimating benchmark values on arterial facilities presents unique challenges. In order to ensure adequate sample sizes for statistical estimation it may be necessary to increase the sampling time-interval. Links should be designed to minimize signal density and access density while ensuring sufficient length for sample data collection. Outliers in the data should be identified and removed before using the data for statistical inference. Finally, the effect of any bi-modal distributions should be considered when estimating a benchmark value.