

North American Traffic Working Group

TRAFFIC INFORMATION BENCHMARKING GUIDELINES (TIBG)

VERSION 1.0



Version Summary

<u>Version</u>	<u>Date</u>	<u>Highlights</u>
1.00	04-22-2010	Initial guidelines based on a single-trace floating car run for ground truth data collection. Comments and revisions can be submitted found from http://www.itsa.org/natwg.html . Licensed under Creative Commons Version 3.0 http://creativecommons.org/licenses/by-sa/3.0/legalcode

Contents

Background information	iv
Purpose	iv
About NATWG	iv
License.....	iv
Finding the most current version.....	v
Contributions	v
Getting involved.....	vi
Submitting modifications.....	vi
List of abbreviations.....	vii
1 Introduction and general considerations.....	1
1.1 Intended audience	1
1.2 Intended use	1
1.3 Basic premises and scope	1
1.4 Overview of current version and known limitations	3
1.5 Test meaningfulness	4
1.6 Geographical references.....	5
1.7 Reporting results.....	6
1.8 Metrics	6
2 Route selection	7
2.1 Road classes, make-up and length.....	7
2.2 Time of day and traffic conditions	8
3 Test equipment	8
3.1 Description	8
3.2 Recommended specs	9
3.3 Sampling rate	9
4 Driving behavior.....	9
4.1 Route completeness	9
4.2 Speed and acceleration.....	10

4.3	Lane following	10
4.4	Passing guidelines	10
5	Data logs processing	10
5.1	Map matching and outlier removal	11
5.2	Valid segment requirements.....	11
5.3	Generating segment speeds	12
6	Traffic content processing.....	12
6.1	Time stamping and epoch consolidation.....	12
6.2	Posted Speed Treatment and Maximum Speeds.....	13
7	Speed comparison.....	14
7.1	Metric.....	14
7.2	Tolerance and/or smoothing	14
7.3	Aggregation and weighting	14
8	Travel time comparison	14
8.1	Metric.....	15
8.2	Tolerance and/or smoothing	15
8.3	Aggregation and weighting	15
9	Congestion level comparison.....	17
9.1	Congestion levels and definition.....	17
9.2	Speed Tolerance.....	18
9.3	Aggregation and weighting	18
	Appendix A: Key Assumptions	19

Background information

Purpose

This document is published by the North American Traffic Working Group (NATWG) under a Creative Commons license. Its intent is to describe a recommended benchmarking method to evaluate the quality of traffic information estimates on a roadway network. NATWG's purpose with these guidelines is to offer the market a more unified way of assessing and comparing traffic estimates. Our hope is that the application of these guidelines will help buyers of traffic information (i.e. car companies, media companies, navigation devices manufacturers, fleets, roadway operators, etc.) conduct more effective audits and purchasing decisions, which will reduce evaluation costs and improve overall market conditions. This vision hinges on the voluntary adherence of market actors to the present guidelines, and we thus seek to achieve consensus.

About NATWG

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

For more information about NATWG:

- Len Konecny, Clear Channel, lenkonecny@clearchannel.com
- Steve Bayless, ITS America, sbayless@itsa.org

License

The license for Traffic Information Benchmarking Guideline (TIBG) v1.0 is Attribution-Share Alike 3.0 Unported. Anyone may share by copying, distributing or transmitting the work, or adapting the work under the following conditions:

Attribution — You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).

Share Alike — If you alter, transform, or build upon this work, you may distribute the resulting work only under the same, similar or a compatible license.

With the understanding that:

Waiver — Any of the above conditions can be waived if you get permission from the copyright holder, which is ITS America and the North American Traffic Working Group

Public Domain — Where the work or any of its elements is in the public domain under applicable law, that status is in no way affected by the license.

Other Rights — In no way are any of the following rights affected by the license:

Your fair dealing or fair use rights, or other applicable copyright exceptions and limitations;

The author's moral rights;

Rights other persons may have either in the work itself or in how the work is used, such as publicity or privacy rights.

Notice — For any reuse or distribution, you must make clear to others the license terms of this work. The best way to do this is with a link to this web page.

The full license can be found here: <http://creativecommons.org/licenses/by-sa/3.0/legalcode>

Source URL is http://www.itsa.org/natwg_TIBGversion1.html

Finding the most current version

The most current version of this document is linked to from <http://www.itsa.org/natwg>

Contributions

The first released version of this document was developed over the course of about one year, starting in January, 2009. Initially, a 'requirements' committee was assembled under the leadership of David McNamara (MTS, LLC - coachdavemc@gmail.com). Seven organizations¹ were represented and agreed to share their current practices for conducting traffic information quality benchmarks. This effort morphed into a task force that was created at ITS America's annual conference in June. The charter of the task force was to develop best practices or guidelines based on the inputs provided by NATWG members.

The current composition of the task force is as follows:

- J.D. Margulici, California Center for Innovative Transportation, jd@calccit.org (chair)
- Billy Bachman, Geostats, wbachman@geostats.com
- Matt Lindsay, NAVTEQ, matt.lindsay@navteq.com
- Kevin Lu, Telcordia , klu@telcordia.com
- Chris Scofield, INRIX, chriss@inrix.com
- Shawn Turner, Texas Transportation Institute, shawn-turner@tamu.edu

¹ BMW North America, California Center for Innovative Transportation, Clear Channel, Inrix, NAVTEQ, Sirius XM, TrafficCast

Getting involved

NATWG and the Traffic Information Benchmarking Taskforce are dynamic entities and welcome additional members and contributors. The present guidelines need early adopters in order to gain traction and legitimacy. You and your organization can get involved in a number of ways, including but not limited to:

- Joining the NATWG membership and weighing in on further development;
- Taking these guidelines on a ‘trial run’ by employing them the next time you evaluate traffic information;
- Providing existing evaluation data, which may be anonymized and scrambled as necessary, so that the Taskforce can add it to a pool that we are assembling for the purpose of validating the proposed metrics;
- Providing feedback and comments, and suggesting modifications;
- Joining the Traffic Information Benchmarking Taskforce and directly contribute to the next iteration.

Submitting modifications

Members of the NATWG may make changes through the ITS America Wikispaces interface. Changes are accepted automatically, but may be subject to final review by the current Task Force. Members making changes are encouraged to use the SAVE WITH COMMENT feature to provide rationale for changes submitted.

Requirements Task Force members may convene periodically to determine whether to accept changes and re-compile *Traffic Information Benchmarking Guideline (TIBG)* into a new version. The current version is uniquely identified on the wiki as version 1.0, followed by the Wikispace generated "revision IDs" of each chapter found on the main [Traffic Information Benchmarking Guideline \(TIBG\) v1.0](#)

In submitting changes, members must adhere to ITS America's [Anti-Trust Guidelines](#) and WikiSpaces [Acceptable Use Policy](#).

In submitting changes, members must adhere to ITS America's Anti-Trust Guidelines and WikiSpaces Acceptable Use Policy.

To join this wiki, please go to the [Join NATWG Page](#).

List of abbreviations

GIS	Geographical Information System
GPS	Global Positioning System
HOV lane	High-Occupancy Vehicle (i.e. carpool) lane
ITS America	Intelligent Transportation Society of America
MPH	Miles Per Hour
NATWG	North American Traffic Working Group
PHS	Position, Heading, Speed
RDS	Radio Data System
RMSE	Root Mean Squared Error
SPM	Seconds Per Mile
TIS	Traffic Information Source
TMC	Traffic Messaging Channel

1 Introduction and general considerations

1.1 Intended audience

The intended audience for these guidelines includes producers of traffic information, both public and private, buyers such as automobile manufacturers, personal navigation solution providers, mobile phone network operators and other media companies, or roadway network operators, as well as all intermediaries, third-party stakeholders and facilitators such as government agencies.

1.2 Intended use

The guidelines outline a set of methods and metrics that can be used to evaluate the quality of traffic information that may include incident or event messages, flow speed information, and travel time estimates. This information is assumed to be broadcast in near real-time to reflect either current or future conditions.

The metrics proposed in the guidelines can be applied toward quality assurance and data validation purposes. They can also serve as benchmark measurements in order to compare data quality between several information sources or commercial providers on a given roadway or within a metropolitan region, or to establish comparisons across multiple regions.

The word ‘guidelines’ aptly describe the intent of this document as a first step toward practice harmonization. While they may end up setting a de facto standard, they are not intended as one. A standard may ultimately constitute the most beneficial outcome for the traffic information industry, and it may be developed on the basis of these guidelines, but it is still too early to tell whether this is a doable or even a desirable goal. This will depend in large part on the adoption of the present guidelines, and the response from traffic information customers.

As guidelines, the methods presented in this document leave room for interpretation and often balances general principles with formal rules. Only with usage and feedback can we determine what constitute the best approach on a case-by-case basis. Thus the most important recommendation of all is that those reporting test results clearly document their methodology and its adherence to or departure from the proposed guidelines. We suggest that test results should be as transparent as possible in order to be recognized as legitimate benchmarks.

1.3 Basic premises and scope

These guidelines were developed based on a consensus that there exist no commonly accepted metric or methodology to measure and compare the quality of traffic information. Institutional purchasers of traffic information (e.g. car manufacturers, Departments of Transportation, etc.) conduct benchmarks with mostly ad-hoc methodologies. This increases the overall costs of those benchmarks and prevents effective comparisons between test results. A broader implication of this situation is that absent a standard lexicon and consistent measurements, traffic data quality issues are generally not well understood, if at all, by its consumers. The postulate that drives the development of the present guidelines is that harmonized benchmarking methods would benefit both suppliers and customers by a)

improving the consistency and fairness of evaluations; b) lowering their overall costs by eliminating duplication of efforts; and c) establishing recognition for true value-added which will pull quality upward.

Note that in order to fully appreciate the experience of a traffic information consumer, one needs to include such factors as timeliness (i.e. how fast is information provided), ease of use (i.e. delivery format), and add subjective variables on top of that. These guidelines do not attempt to capture any of those elements. Their primary concern is with the accuracy of traffic estimations that are produced and transmitted.

The most basic assumption about the proposed methodology is that there exists a way to collect information about current conditions on a roadway that is deemed trustworthy enough to qualify as ‘ground truth’ and thus serve as a referent against which to compare other traffic information sources². Once ground truth data has been collected, the exercise consists in scoring information sources against it using a set of metrics. These metrics should ideally meet the following criteria:

- Formally defined and easy to compute (i.e. not too many exceptions / fringe cases);
- Relevant to the end-user experience;
- Easy to interpret;
- Good balance of synthetic vs. exhaustive (i.e. tells the story concisely);
- Normalized and scalable (i.e. independent from route length, sample size, etc.)

In their present form, the guidelines apply to controlled-access roadways, essentially centrally-divided expressways with no intersections. The focus is on estimating the accuracy of velocity-based information, as this is arguably the most determining factor of how drivers experience traffic (as opposed to say, vehicle density measurements that are the main currency of traffic engineers³). Such information can take one of the forms outlined below:

- 1) Overall speed of the traffic flow at a particular point and a given time;
- 2) Average travel time along a roadway segment or route;
- 3) A descriptive qualification, such as ‘free flowing’ or ‘heavily congested’, alternatively presented with color codes on a map of the roadway network.

Since these information types can all be derived from an estimate of flow speed, the evaluation method presented in this document boils down to performing the comparison illustrated by Figure 1. However, we still consider each information type separately and provide corresponding evaluation metrics in order to offer a direct relationship between common applications of traffic data and means to assess their quality.

² This document intently avoids setting a formal definition or standard for ground truth, but it does provide practical recommendations on how to collect referent data.

³ As such, these guidelines apply primarily to traveler information services. While this type of information is also of great use to transportation authorities and roadway traffic managers, it does not cover all of their needs, which typically include vehicle counts, classification, traffic density measurements, and other variables.

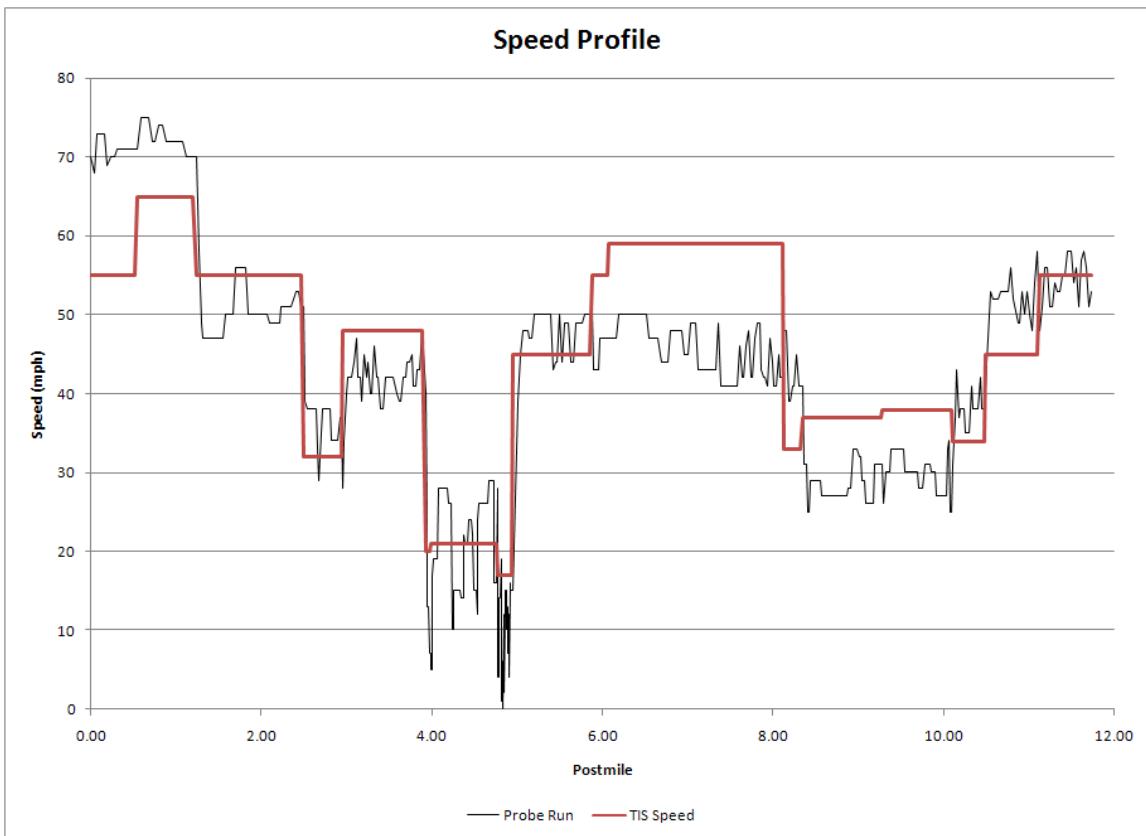


Figure 1 - Example comparison of probe vehicle speed and speed from a traffic information service provider along a route

Accordingly, the document covers the following items:

- What constitutes an acceptable test route?
- How should ground truth data be collected?
- How should ground truth data and traffic estimation data be paired up for comparison?
- What metrics provide the most meaningful comparisons, and how should they be reported?

1.4 Overview of current version and known limitations

As of the current version, the proposed method for collecting referent or ‘ground truth’ data against which to compare estimates is to use floating cars. This means that vehicles outfitted with a GPS receiver and a data recorder take trips on a given itinerary for the purpose of comparing their observed travel speed with the estimated speed broadcast by a traffic information provider, or source (TIS). In other words, the traffic conditions met by those floating cars are considered to constitute the observable ground truth against which estimates are benchmarked at any point in time and space.

The floating car method poses a statistical challenge because sampling is typically scarce unless a whole fleet can be dedicated to benchmark runs. However, it is also very popular because of its relative ease of implementation and the feeling of indisputability provided by data that is collected through direct

immersion into the traffic flow (aka. ‘out of the window’ testing). The current iteration of this document considers a *single vehicle* driving a *single run* on a preset route. Taken in isolation, such a test would not constitute a valid benchmarking method, but it is the **building block** for a more complete test aimed at a roadway network. Once these guidelines meet a reasonable level of acceptance, they will be expanded to consider multiple-run tests and network-level benchmarking. Following is a list of specific issues that we intend to address in future versions:

- Further work by NATWG may consider an alternative ground-truth data collection method based on travel time sampling extracted from technologies that individually identify vehicles (e.g. toll tag readers, automated license plate readers, Bluetooth readers or magnetometers with a unique tagging capability). It is also possible that the massive influx of probe data stemming from mobile phones and GPS receivers into traffic information collection systems will result in the ability to self-validate (specifically, real-time estimates may be checked after the facts with bread-crumb data).
- For each set of metrics proposed in these guidelines, a complete methodological explanation is provided. However, the resulting outcome is not ‘plug-and-play’. Specific formulas are conditional to setting certain parameters which will still require fine-tuning. Further, ultimate test results should aggregate multiple runs over a region, and we have not yet provided guidelines on how to perform such aggregations.
- As we move beyond a single-route / single-run test description, probably the most critical issues will be to determine adequate sampling rules that attach a high degree of confidence to test results. As pointed out above, data collection alternatives to floating cars will be considered, notably including reidentification technologies.
- As it stands, the guidelines are aimed at measuring traffic information quality on controlled-access roadways where traffic flows are relatively homogeneous. The traffic information industry is expanding its coverage to signalized arterials and data quality will need to be assessed on those roadways as well. However, it is still early to move in that direction.

1.5 Test meaningfulness

Much of this document focuses on ensuring that benchmark tests are conducted in ways that produce meaningful results. A few general principles can be highlighted here:

- Floating car runs conducted in the absence of any traffic congestion have very little value. It is well documented that merely broadcasting historical traffic conditions or even posted speed limits can yield accurate results well in excess of 90% of the time. Therefore the value of real-time traffic reporting can only be assessed during changing or unusual conditions.
- As a corollary, traffic information benchmarks must be conducted on roads that host significant flows. This excludes most residential streets and minor rural roads.
- A further necessity of making benchmarking results meaningful is to bind test areas to a set of relatively homogeneous roadway sections. In this respect, the most important distinctions are

rural areas vs. urban or suburban areas, and controlled-access roadways vs. signalized arterial roads⁴.

1.6 Geographical references

These guidelines consider three useful geographical reporting levels: TMC location codes, routes, and networks (or markets). As depicted on Figure 2, these three levels are nested: a market is a collection of routes, itself a collection of TMC codes. At each level, traffic information quality is a function of time – roughly speaking a score can be attributed at regular time steps and then aggregated over longer periods.

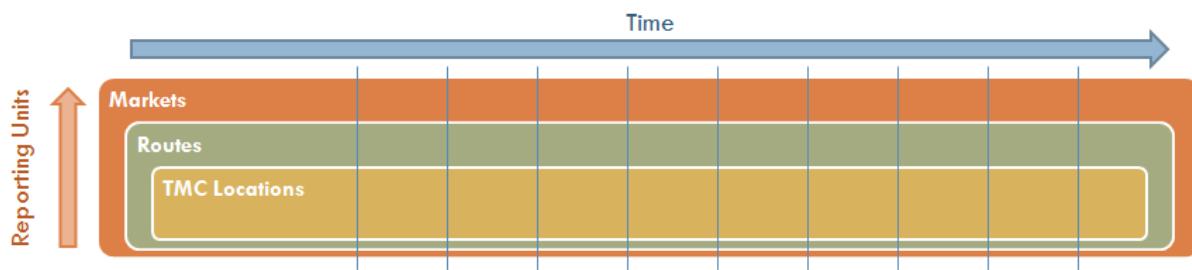


Figure 2 - Benchmark reporting units: TMC codes, routes and markets

Traffic Message Channel location codes (commonly referred to as **TMC codes**) were developed for the Radio Data System (RDS) Service in order to standardize the reporting of traffic events on major roadways under a unique set of geographical references. TMC codes are typically assigned at significant decision points and intersections, in an unambiguous format that is independent of existing digital maps for the area. In North America, TMC location codes are assigned and maintained through a collaborative effort between map publishers NAVTEQ and TeleAtlas. While those TMC location codes may be defined with slightly different coordinates in both map systems, the physical reference (typically an intersection) is common. Note that the incidence of using either map system on test results will be so small that is considered negligible. However, for disclosure sake, we recommend reporting the mapping system and version in test results.

Although the TMC standard was initially designed to deal primarily with events (accidents, closures, weather, etc.), traffic information providers have adopted it so widely that it is also used to report flow information on segments defined by two adjacent location codes. In practice, TMC codes thus designate both a location and the directional segment that originates from that location through the next TMC code. As of this writing, TMC codes constitute the reporting unit of choice for the entire traffic information industry. Therefore, they are the natural set of references with which to assess traffic information quality.

Note that while TMC location codes have been a boon to the traffic information industry by bringing a standard to the market, they are not a panacea for reporting complex congestion patterns. If a traffic

⁴ Although these guidelines are not specifically developed for signalized arterials, many general principles still apply.

information provider reports data with a finer spatial granularity, the methodology described in the present guidelines still applies: the comparison between a TIS and a set of records believed to represent ground truth can be performed along any set of elementary road segments, TMC location codes or others. The remainder of the text primarily assumes that TMC codes are used, but one may substitute TMC codes for a different roadway segment definition.

A **route** is a meaningful itinerary, and can also be defined as a sequence of contiguous TMC location codes.

Finally, a **network** is a collection of roadways in a given geographical area, which may define a ‘market’. A network will typically encompass hundreds of TMC location codes, and it can be broken down into routes for benchmarking purposes. Benchmarking at the network level is not covered in the present version of this document but will be tackled in a subsequent version. A key consideration at the network level is the volume of ground truth data that is necessary to achieve statistical significance and thus ensure reliable and undisputable test results.

1.7 Reporting results

As already mentioned, an overarching recommendation is to provide as much information and transparency as possible on the methods with which test results are obtained. One way to do this is to follow these guidelines and to clearly indicate where they are observed and where the test method deviates from them.

At a minimum, test results for a single-run test should report the following information:

- Route description, including TMC codes covered
- Type of GPS receiver and data recording device
- Map provider and version used during the test

1.8 Metrics

The metrics presented in these guidelines are designed to measure differences between a traffic information source and data collected by a floating car. These metrics will provide a set of scores that can be easily captured and understood, and allows for an immediate appreciation of the quality of an information source by a reasonably attuned but non-expert professional. Our design philosophy is to strike an adequate balance between a broad synthesis that may hide relevant nuances in the data, and a level of detail that may reveal all kind of interesting features but fails to directly tell a story to the naked eye. Of course, raw numbers can be misleading without a proper understanding of context, and we do not suggest that metrics can be a substitute to a rounded view of any given situation. In fact, that argument precisely militates for standardized metrics that can become a lead-in to a meaningful conversation about context rather than be the object of the discussion.

Three concepts were used in the development of metrics intended to assess the quality of a TIS:

1. One concept focuses on traffic speeds and measures the aggregated difference between estimated values and ground truth values. The result is a numerical value or a set of them that is

expressed in miles per hour. Such a metric provides a reading of the overall discrepancies between actual traffic speeds and broadcasted speed information, as well as their distribution.

2. The second concept considers travel time estimates. There is an inverse relationship between travel times and speed, and thus the information conveyed by measuring travel time errors is essentially the same as the one obtained by measuring speed differences. However, broadcasting travel times is a very common and popular application of traffic data, and the inverse relationship is not intuitive. Therefore, we offer a methodology to compare estimated travel times with ground truth travel times and report the difference in a normalized way by using seconds per mile as a unit.
3. The third concept looks more broadly at the estimated traffic conditions in terms of congestion levels. Most information sources report traffic conditions either in descriptive terms (e.g. ‘free flow’, or ‘heavy congestion) or by using color codes on a digital map (e.g. green, yellow, and red). Thus a useful indicator of information quality is to count the number of instances in which the reported congestion level matches or doesn’t match reality. Another way to think about such reporting is in terms of error type with respect to the presence of congestion, i.e. type I errors (congestion is reported when it is in fact not present) and type II errors (congestion is missed).

All three concepts and the corresponding metrics formulation are described in section 7 (numerical speed comparisons), section 8 (travel time comparisons) and section 9 (congestion levels) of this document.

2 Route selection

This section describes how to design test routes that are suitable for benchmarking traffic information.

2.1 Road classes, make-up and length

For the purpose of designing test routes, roadway sections may be divided loosely into three types:

- Controlled-access highways: roads connected only by ramps, with no at-grade crossings
- Major arterials: roads with at-grade crossings and posted speed over 40 mph
- Other arterials: roads with posted speed less than 40 mph

A cohesive route should be primarily made up of one type of roadway. Further, routes should span primarily urban/suburban areas or rural areas, but not both. The make-up of a route per the above categories should be indicated in the test results.

Ideally, routes should be designed such that they can be defined as a sequence of contiguous TMC location codes, i.e. along the same road as defined in the TMC table. The object of the test is to measure through-traffic. Thus including ramps into a test route is not recommended, though it may be needed in some cases (e.g. freeway-to-freeway ramps). For TMC location codes included in the route, the entire corresponding segment should be driven and included in test results. In some cases, this may mean that

the driver needs to enter and exit a freeway at ramps that are located outside the route boundaries in order to not skew results with acceleration/deceleration and merges.

In order to provide meaningful reports that are distinct from TMC-level results, routes should be at least 2 miles and 3-TMCs in length. Routes can be much longer provided that their make-up follows the aforementioned guidelines and driving times do not typically exceed an hour, at which point fluctuations within a single run may become too large for the route to constitute a meaningful unit.

2.2 Time of day and traffic conditions

As pointed out in section 1.5, test routes should be known to feature a reasonable likelihood of traffic congestion on a substantial portion of their length and during extended periods of time. There is never an absolute guarantee ex-ante that a particular roadway section will be congested (in fact we would always hope that it is not!) However, time of day and known historical patterns should guide both route selection and run scheduling.

Congestion may result from very regular patterns or more exceptional ones, and both instances may constitute good test cases for the purpose of benchmarking. Examples of the former would include:

- Weekday rush hours (typically defined as 6-10 AM and 4-8 PM) on urban arterials and inbound (morning)/ outbound (afternoon) regional connectors;
- Weekday rush hours between major urban points of interest, e.g. airport, central business districts or shopping malls;
- Weekend vacation routes.

Examples of routes for which congestion is possible but less certain include:

- Midday, evening or weekend routes;
- Routes to or past special events, e.g. trade shows or ball games.

3 Test equipment

This section specifies the type of test equipment that should be employed to log test data.

3.1 Description

Floating car probes will use Global Positioning System (GPS) data loggers to record point, heading, speed (PHS) triplets. GPS data loggers record PHS data at a fixed sampling rate, and store the time-stamped data points. Typically, these devices may be plugged into a USB port for downloading the data.

The great advantage of GPS is the relatively low-cost of the receivers, making large-scale traffic monitoring possible. Most low-cost devices today exhibit horizontal error better than +/- 15m. This horizontal uncertainty prevents precise detection of roadway and direction of travel using lat/long position alone. Section 5 provides guidelines on how to identify the best roadbed on which to ‘snap’ GPS data points.

3.2 Recommended specs

Multiple inexpensive data loggers are available for consumer use. We recommend selection of a device with characteristics at least as good as the following:

- Built-in, rechargeable battery with capable of a minimum of 10 hours of continuous use from a full charge;
- 12VDC, 110VAC car and home chargers, respectively;
- 2-status LED to indicate power, GPS signal;
- USB 2.0 or Bluetooth 2.0 connectivity to host computer;
- Built-in antenna.

Performance Specifications:

- 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;
- Sample Rate: NEMA rate or programmable 1s to 5s.

3.3 Sampling rate

The recommended sampling rate for logging floating car GPS data is the default NEMA rate of 2s (equivalent to 0.5Hz).

4 Driving behavior

This section describes recommended driving behavior that will result in more reliable tests. These recommendations notwithstanding, safety for both the test driver and other drivers on the road should remain the primary consideration and guide behavior at all times.

In general, test drivers should attempt to mimic the experience of a ‘regular’ driver. This is a somewhat elusive notion, but in practice, it means that test drivers should try to be following the bulk of the traffic, not driving in the slowest lane, nor attempting to always be in the fastest lane. Test drivers are free to change lanes and negotiate traffic in a manner consistent with the majority of other drivers present on the road at that time. If one thinks of driver behavior as distributed across a typical bell curve, then test drivers are trying to hit the center of that curve –figuratively of course: safety first!

4.1 Route completeness

Test drivers must ensure that they complete each route on their itinerary in one, uninterrupted run. A route that is started must be finished with no stop while adhering to the driving guidelines presented in the remainder of this section. Of course, test drivers should not hesitate to exit a route or stop along the

way if safety or another important circumstance requires it, but then the entire data for the corresponding run must be dismissed for the purpose of data quality benchmarking described in this document.

4.2 Speed and acceleration

Unless a route specifically includes a ramp (e.g. freeway to freeway connector), drivers should not accelerate (or respectively, decelerate) to insert their vehicle into the traffic flow (respectively, exit the traffic flow) while on a recorded run. As indicated in section 2, routes will be designed to avoid such incidence by starting / ending at locations that provide an upstream / downstream buffer for the driver.

While on a route, drivers attempt to maintain speed with the vehicles in the visible field of view and pace with this vehicle group so that the speed of the test vehicle emulates the average flow of all vehicles. There is however, one notable exception to this rule. Speed limits should be obeyed, even if the flow of vehicles is higher than the speed limit. Section 5 and 6 include data processing provisions that ensure that this discrepancy will not adversely impact test results.

4.3 Lane following

On a multi-lane roadway, the driver is to keep to the middle lane at the average flow of traffic whenever possible. When visibility is obscured in the center lane, moving to another lane is desired so that the test vehicle can maintain the same average speed as all the vehicles in sight. Travel in the left lanes should be limited because the left lanes often exceed the speed limit and these lanes maintain speeds which are higher than average.

4.4 Passing guidelines

The average flow of traffic is best judged by keeping a balance between the number of vehicles the test driver overtakes and the number of vehicles who overtake the driver. This number should be kept as even as possible. Test drivers should generally avoid passing on single-lane roads. The exception would be in the case where the driver is stuck behind a slow moving vehicle which all or most other vehicles are passing.

5 Data logs processing

This section describes how to process data collected during a floating car test. Once a run is completed, a log of data records is extracted from the test equipment. Each record essentially includes a set of GPS coordinates, a time stamp, and a speed reading. In order to perform comparisons between data collected from the floating car and speed estimates that are pulled from a TIS, individual records need to be aggregated to match the spatial resolution of the traffic service. In today's market, this naturally points to TMC location codes. In other words, GPS records collected along the road segment defined by a TMC location code during a test run are grouped together to compute a single speed value, which can then be directly compared to the data feed provided by a TIS for that TMC location code.

As a preliminary step to processing data logs, it is recommended that visualization software be used to verify that the runs correspond to specified routes and itinerary. This step will ensure that blatant

problems in data collection or driver errors are captured right away so that inadequate runs can be discarded with no further treatment.

5.1 Map matching and outlier removal

The first step in processing data logs from a floating car test is to match individual GPS records onto a road map. There is abundance of software available to perform this task including commercial and proprietary solutions. Given the impracticality of enforcing a single version of a single map database on the target audience of these guidelines, testers have latitude in their choice of map publisher and vintage as well as map-matching technology. These choices must be disclosed (i.e. publisher, quarter, version or other naming convention), and must preferably remain bounded to technologies that are well-established and accepted within the industry. In particular, testers should have access to a digital representation of the roadway that contains link-by-link reference to adequate segmentation units such as TMC location codes. It would not be appropriate to manually interpret the TMC location table on a base map which does not display TMC codes (e.g. using an online mapping application). In such practice it is highly likely that nuances in the exact placement of the segments used by both the map databases and the TIS could be misrepresented.

Individual GPS records must be filtered upfront to ensure a valid run. Regardless of the technology involved at this stage, the following practices constitute minimum best practices:

- If a GPS record appears 25 m 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 segments, 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;
- Under some circumstances ‘divergent lane conditions’ may occur where the driver cannot emulate the average traffic flow (such as congestion on one side of the road due to slow ramp traffic, toll gates with automated and manual collection, HOV conditions, etc.). GPS records for these conditions should be discarded.

5.2 Valid segment requirements

For each TMC location code or segment used to aggregate individual records, the following conditions must be set:

- The segment must have been entered, traveled and exited at full traffic speed (ruling out acceleration or deceleration phases due to ingress/egress);

- Based on the recommended sampling rate of 0.5Hz, at least 90% of the records supposed to be collected over the length of the segment must have passed the filters set in the map matching step described in section 5.1;
- 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 TMC location code or other segment must be discarded all together when the above conditions are not met.

5.3 Generating segment speeds

The calculation of the space-mean speed for a TMC segment from a GPS trace requires an accurate TMC segment distance value d and an accurate calculation of the segment travel time T^{GT} from the GPS data. Both NAVTEQ and TeleAtlas centerline road databases have TMC segment distances calculated based on their positioning of segment endpoints. When using other road databases, or the raw TMC location codes, the distance must be calculated using GIS techniques.

The calculation of travel time must estimate the exact time that the probe vehicle entered the segment and the exact time that the probe vehicle exited the segment. Errors of one or two seconds can have noticeable impact estimated segment speeds, particularly over short segments. To capture the exact time, the two GPS points straddling each segment end point must be used to interpolate the crossing time. Let t_1 and t_N be the interpolated time stamps at which the probe vehicle entered and exited the TMC segment, then the ground truth segment travel time is defined as:

$$T^{GT} = t_N - t_1$$

And the ground truth speed V^{GT} for the corresponding segment is calculated as:

$$V^{GT}(t_1, t_N) = \frac{d}{t_N - t_1}$$

Simple match-matching routines that assign each GPS point to a segment are not recommended unless there are limits placed on segment length. When such technique is used, we advise to exclude segments less than .25 miles in length from the analysis.

6 Traffic content processing

This section describes how to process traffic information content sources to fit into the benchmark method.

6.1 Time stamping and epoch consolidation

Section 5.3 defines a time period during which an average ground-truth speed is computed:

$$T^{GT} = t_N - t_1$$

Define T^{TIS} as the time interval for which the TIS provides a single speed. One or more TIS time intervals may overlap the ground-truth time period. When T^{GT} both begins and ends within T^{TIS} , then the ground-truth speed and TIS reported speed may be compared directly.

However, when T^{GT} represents more than one TIS interval, an equivalent TIS speed must be computed for the ground-truth period. The following procedure should be used to compute the equivalent speed:

- 1) Determine the amount of time the ground-truth 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 ground-truth time spent in each TIS interval to the total time T^{GT} .

As an example, consider a ground truth period defined by

$$T^{GT} = t_N - t_1 = 9:12:22 - 9:02:03 = 619s$$

Suppose that the TIS reports speeds every 5 minutes, aligned with the start of each hour. Then T^{GT} overlaps 3 TIS time intervals, as illustrated in the following diagram:

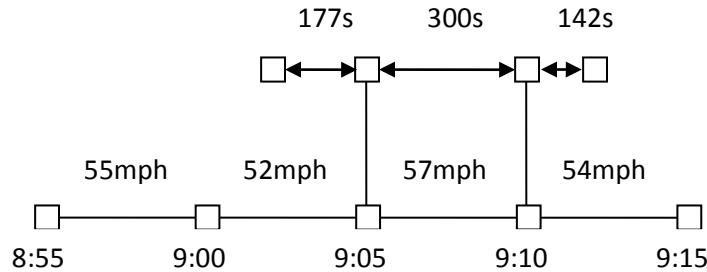


Figure 3 - An illustration of the overlap of a ground-truth time period with the TIS time intervals

Thus in this example, the ground-truth 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

$$V^{TIS} = \frac{54 * 142 + 57 * 300 + 52 * 177}{619} \approx 55 \text{ mph}$$

V^{TIS} is then used when comparing ground-truth speed the TIS reported speed.

6.2 Posted Speed Treatment and Maximum Speeds

Although posted speeds limits are not always reliable, they represent a standard reference speed against which reported and measured speeds may be compared. Rather than requiring each ground truth testing agency to measure a standard reference speed for each road segment, we recommend that

ground-truth testing use map-based posted speed limits, and simply report the map type and version number for purposes of comparison.

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

7 Speed comparison

This section focuses on analyzing the accuracy of speed estimates provided by a traffic information feed by direct comparison with the speed profile recorded by a floating car run. In other words, the metrics described in this section gets closest to measuring the degree of fit between the two curves shown on Figure 1.

7.1 Metric

Speed comparisons shall be performed and reported based on miles per hour (mph) differences between estimated speeds and ground truth speeds.

7.2 Tolerance and/or smoothing

Sections 5 and 6 describe the pre-processing of ground truth data logs and traffic information streams that ensure fair and meaningful comparisons with regards to such issues as speed limits and very short segments.

7.3 Aggregation and weighting

The recommended method to aggregate speed differentials measured at the TMC location code or other segment level is to compute a root mean squared error (RMSE). The RMSE is the square root of the average squared speed differences between the floating car run and the TIS for each TMC location code:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (V_i^{GT} - V_i^{TIS})^2}$$

The RMSE has the advantage to encompass both the systematic error (i.e. an across-the-board overestimation or underestimation), and the error dispersion (i.e. it captures the noise in the data by adding all errors even if the estimate is unbiased) into one single index.

8 Travel time comparison

This section focuses on analyzing the accuracy of travel time estimates provided by a traffic information feed by direct comparison with the actual travel time recorded by a floating car run.

8.1 Metric

Travel time comparisons shall be performed using total travel time and harmonized to seconds per mile of error, which provide a normalized benchmark unit.

8.2 Tolerance and/or smoothing

Sections 5 and 6 describe the pre-processing of ground truth data logs and traffic information streams that ensure fair and meaningful comparisons with regards to such issues as speed limits and very short segments.

8.3 Aggregation and weighting

For each road segment considered (TMC location code or otherwise), four basic elements are calculated:

- L - The length in miles of the cumulative distance between each GPS point on the segment
- T^{ACT} - The actual travel time of the test vehicle.
- T^{REF} - The reference speed travel time estimate of the vehicle.
- T^{TIS} - The estimate of travel time using data from the Traffic Information Service are calculated.

For example, three sections of a 60 MPH limit freeway might yield the following:

Table 1 - Example calculations for seconds-per-mile determination

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

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

$$D^{REF} = T^{REF} - T^{ACT} \text{ and } D^{TIS} = T^{TIS} - T^{ACT}$$

$$E^{REF} = D^{REF}/L \text{ and } E^{TIS} = D^{TIS}/L$$

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

In this fashion, the performance of traffic information can be compared in a normalized fashion to create a metric of the value of the traffic information in the context of the amount of travel time lost due to the vehicles reduced speed where 'Improvement' (I) is defined as:

$$I = ABS E^{REF} - ABS E^{TIS}$$

Results can be further aggregated by considering the performance of the TIS and reference speed estimates over the course of a route of multiple segments. If we consider the three segments above to be a route, we need to sum the length of the segments on the route and each of T^{ACT} , T^{REF} and T^{TIS} and recalculate the relative and absolute deltas, harmonized error per mile values and measures of improvement.

We can further calculate a second value of improvement (I^{PC}) as a percentage of the total absolute error of the reference speed estimate removed using the total TIS estimate for the entire route – e.g.:

$$I^{PC} = I / \text{ABS } E^{REF}$$

Route results can likewise be summed and relative and absolute deltas recalculated to provide comparative statistics at any level of aggregation such as all roads of similar characteristics, whole cities etc (recommendations for which are outlined elsewhere in this document). Per requirements set out in section 2, a route must be at least 2 miles in length and no more than 1 hour in duration.

Our example returns the following:

Table 2 - Example calculations for seconds-per-mile determination (2)

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

On Segment-ID 1, we see that a slower than actual TIS speed creates a longer estimated travel time than was seen in actuality. Despite the fact that a slowdown was identified, the 12 seconds per mile of additional travel time effectively cancels out the opportunity to improve the estimated travel time of the reference speed which was 12 seconds per mile too fast.

On Segment-ID 2, we see a positive improvement of 30 seconds per mile using the TIS estimate over the reference speed, despite the continued existence of 30 seconds per mile of error in the TIS estimate.

On Segment-ID 3, we see a degradation in performance of some 10 seconds per mile by using the TIS estimate over the reference speed.

Another useful metric that can be derived from this numbers is the performance of both the reference speed estimates and the TIS estimates of travel time to the actual travel time observed, where:

$$P^{REF} = 1 - (\text{ABS } D^{REF} / T^{ACT})$$

These metrics provide the context for the amount of congestion observed in the test and the impact of the improvement in performance in the context of the total actual drive time. These metrics as all travel time metrics tend to provide more clarity when aggregated at the route level.

In our sample, the P^{REF} values vary from 50% to 84% accuracy of travel time prediction and average only 59% accuracy for the whole route. P^{TIS} also varies from 56% to 84%, but in the context of the route, the travel time estimate using the traffic data is 92% accurate.

Table 3 - Example calculations for seconds-per-mile determination (3)

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

9 Congestion level comparison

This section focuses on ‘congestion level analysis’, i.e. assessing the degree of fit between the congestion level (e.g. uncongested, mild congestion, etc.) called by a traffic information feed and what the floating car driver experiences.

9.1 Congestion levels and definition

Some applications of traffic information only require a coarse estimate of congestion. For these applications, flow rate can be characterized by a congestion level. Table 4 illustrates such a taxonomy in which average traffic speeds are organized into four levels. Note that the congestion levels are defined as percentages of the reference speed on the road segment. However, the congestion levels in this section are provided as an example – users of this document may choose to use fewer or more levels, with different speed boundaries.

Table 4 - Illustration of the congestion level concept based on reference speed

Estimated speed as percentage of reference speed	Index	Congestion Level
92+	4	Green
62-92	3	Yellow
31-62	2	Red
0-31	1	Black

Also shown in Table 4 is the ‘color category’ of each congestion level. Colors are often used when labeling a map with traffic data because they are easier to interpret than actual reported speeds.

In order to perform congestion level comparison, floating car data must be converted to a congestion level for each TMC location code or segment, by using a key similar to the one presented in Table 4. If

the information provided by the traffic source being evaluated consists of numerical speed values only, then the same process is applied to that data.

9.2 Speed Tolerance

This section describes how to treat speed differences that occur at the boundaries of two distinct congestion levels. The floating car speed and the speed reported by a traffic information source may stand across one of these boundaries while being very close. Penalizing the traffic information provider for a wrong estimate in such a situation is neither fair nor desirable. To minimize the impact of such incidences, a speed tolerance is applied when comparing the congestion levels derived from ground truth speed and the traffic information speed.

Let the speed transmitted by the traffic service correspond to congestion level L_i^{TIS} , with lower and upper boundary speeds v_i^L and v_{i+1}^L . Then, if v^{GT} is the ground truth speed, compare the congestion level and ground truth speeds according to:

$$\begin{aligned} &\text{if } v^{GT} > v_i^L - \theta \text{ and } v^{GT} < v_{i+1}^L + \theta, \text{ then } L^{GT} = L^{TIS}, \\ &\text{otherwise } L^{GT} \neq L^{TIS}, \end{aligned}$$

where θ is the speed tolerance. That is, if the ground truth speed is within the same speed band as the traffic service speed \pm a tolerance θ , then the two congestion levels match. Only if the ground truth speed falls outside that band is the congestion level estimated by the TIS considered erroneous. The value selected for θ is at the discretion of the tester but must be clearly called out in the test results.

9.3 Aggregation and weighting

Simple error analysis might consist in counting the number of mismatches between the congestion level estimated by a traffic information source and a floating car run among each traversed TMC location code or segment. However in some cases one is only interested in the ability to distinguish congestion from normal traffic flow. In this case, an error is only counted if the ground truth and TIS categories are on either side of the Yellow/Red boundary. For example an error would be counted if the ground truth category were Yellow and the TIS category Red, but not if the TIS category was Green.

In a similar fashion, correct detection of congestion can be computed on a road segment basis. Note that correct detection of free or mild congestion is not normally counted. These are much easier states to detect, and not typically of interest when evaluating the quality of a traffic information service. For example, if the ground truth congestion level is Red or Black and the TIS congestion level is either Red or Black, this is counted as a correct detection, however, if the ground truth congestion level is Green or Yellow and the TIS congestion level is Green or Yellow, this is not counted as a correct detection (nor as an error).

Appendix A: Key Assumptions

This appendix lists key assumptions and decisions made by the contributing Taskforce. It is provided here as a shorthand reference that can serve as a checklist or a point of entry for those who wish to offer critiques.

- These are guidelines and not a standard
- The overall philosophy is that tests should report their assumptions and method in detail. Not sticking to the guidelines is fine if it is clearly called out.
- As of version 1, the document assumes a single car and a single run. This does not reflect a realistic test but serves as the elementary building block upon which we can expand.
- These guidelines will undoubtedly work well for freeways. For surface street arterials, the evidence may be more mixed.
- Individual GPS records must be within 25m of the roadway segments to which they correspond in order to be considered valid. If a record is snapped to the wrong roadway, it is preferable to discard it.
- In order to be considered valid, each road segment for which test data is reported must be entered, traversed and exited within mainline traffic at the dominant speed. Further, at least 90% of the GPS records expected for that traversal must be collected and appear correct, and gaps of 10 seconds or more in the GPS trace should either invalidate the run or be called out.
- TMC location codes are considered the default reporting unit, but the methodology does not preclude selecting a different, i.e. more granular set of segments.
- Default ground truth sampling rate is 2 seconds.
- On each segment / TMC location code, ground truth data is averaged into a single speed value based on a map-provided distance and time traveled. Accurate segment entry and exit times should be calculated by interpolating the closest neighbors in the GPS trace, especially on shorter segments.
- Along a given segment / TMC location code traversed by a floating car, multiple epochs are consolidated using a time-weighted average.
- Posted speed limits as provided by the digital map used for the test serves as reference speed for uncongested travel conditions.
- Metrics are of three kinds: aggregation of numerical speed comparisons, travel time comparisons, and number of correct/incorrect matches in estimating congestion levels.
- Reporting of numerical speed comparisons is done using root-mean-squared errors.
- Travel time comparisons are normalized to seconds per mile. Differences are reported by averaging the absolute values of errors (average absolute error).
- For congestion levels, it may be difficult to agree on a set of uniformly accepted levels and this discussion has been postponed.
- The final reporting of congestion levels may need additional discussion (i.e. # of misses, vs. a more contrasted picture that counts hits and misses by congestion levels).