The Eastern Transportation Coalition – Traffic Data Market Place Data Validation Program 2021 Version 2.0

1. INTRODUCTION AND BACKGROUND

An essential aspect of the Eastern Transportation Coalition (ETC), formerly the I-95 Corridor Coalition, Vehicle Probe Project traffic monitoring system is the validation program, which independently evaluates traffic data provided through the marketplace to ensure its quality and accuracy. The traffic monitoring system, previously referred to as the Vehicle Probe Project (VPP), but now labeled the Eastern Coalition's Traffic Data Marketplace (ETC TDM), is procured from private industry (the Vendor/s) based on contract specifications. The intent of the validation is to assure that traffic data conforms to contractual standards and meets the data needs of the Coalition members, a process that involves both quantitative and qualitative analysis of datasets available through the marketplace. Where possible, reference data is independently collected in the field, processed to establish ground truth traffic conditions, and compared to vendor data reported through the ETC TDM.

The remainder of this document focuses primarily on the evaluation methods used to evaluate ETC TDM travel time and speed data, a core data product in procurement, and which has a legacy of established validation protocol dating back to 2008. These data collection, processing, and evaluation techniques based on Wireless Reidentification Traffic Monitoring (WRTM) technology were developed over the course of VPPI and VPPII and are not expected to change significantly moving forward, although the ETC reserves the right to update methodologies to reflect best practices.

Additionally, this document provides a high-level overview of how other data are expected to be validated. Unlike travel time and speed data, the new TDM data items (e.g., traffic volume, origin-destination, waypoint data, freight data, and conflation data sets) do not yet have fully-developed validation methodologies. However, note that *any data procured in the ETC TDM is subject to Quality Control and Quality Assurance (QC/QA) by the ETC validation team, under the direction of the Eastern Coalition's Data Quality Validation Committee*.

2. ROADWAY LOCATION REFERENCING

The CATT Works Georeferencing Protocol (CWGP) is used to describe roadway segment and point locations in a map-free, or map-agnostic, manner. The Coalition validation team will create CWGP location references to describe where reference data is collected (e.g., *segment* locations for quantifying speed and travel time over a distance defined by WRTM sensor placement, and *point* locations to describe traffic volumes at a fixed location along a roadway).

Vendors are responsible for providing data for the corresponding road segment locations specified using a method that conforms to the CWGP protocol.

3. TRAVEL TIME AND SPEED DATA VALIDATION

Travel time and speed data, a core data product in the ETC TDM and the core data item from VPPI and VPPII, are validated by deploying re-identification sensors along a target corridor, collecting travel time data from a sample of vehicles passing by the sensors, and using the sampled data as the basis for comparison with travel time data provided by vendors as part of the ETC TDM. Various techniques are used to compare data sources, which evolved over the course of the first two iterations of the Vehicle Probe Project. These analysis techniques may be improved and modified over time as technology and research evolve, but the general validation approach is expected to stay essentially/structurally the same as the current process described below.

3.1 WRTM Data Collection

Portable Wireless Re-Identification Traffic Monitoring (WRTM) sensors are deployed at strategic locations along selected road segments to record, identify, and later re-identify at a downstream location unique electronic signatures (i.e., MAC Ids) emitted by in-vehicle electronic equipment via Bluetooth, Wi-Fi and other technologies. These portable sensors which were initially developed in the Center for Advanced Transportation Technology (CATT) at the University of Maryland run on rechargeable batteries and are suitable for about 10-14 days of continuous data collection. While deployed in the field, each sensor continuously records the MAC ID and timestamp of Bluetooth and Wi-Fi-enabled electronic devices that pass through its detection radius. Note that the sensors also have embedded GPS location devices for both precise positioning and time references.

After the data collection is complete, raw data is extracted from each device and uploaded to a server where it can be downloaded by the validation team for post processing and further analysis. Matching MAC addresses between consecutive sensors generates travel time samples, which form the basis of comparison to VPP travel time and speed data. This process is illustrated visually in Figure 3-1 and described in further detail later.

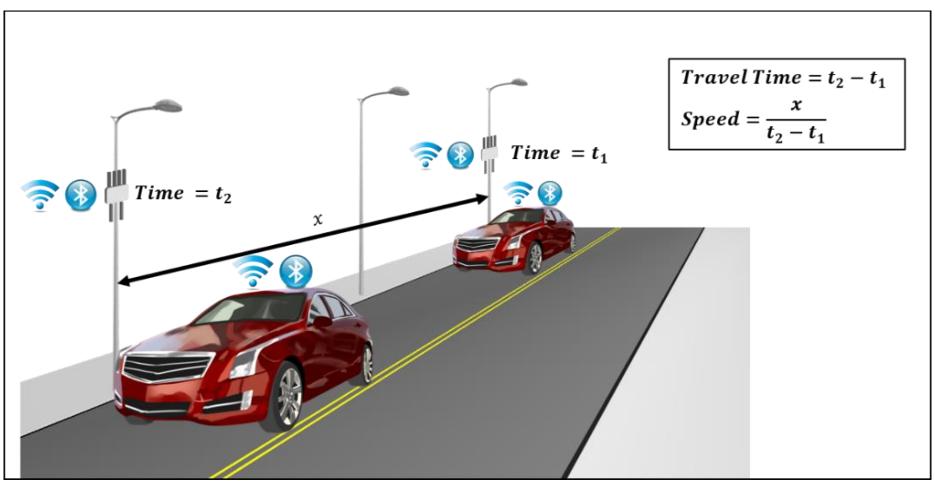


Figure 3-1. WRTM Travel Time and Speed Data Collection

Note, as per the ETC TDM, travel time is the basis for measurement. Speed data, referred to as space-mean speed for a segment, is simply the distance for the segment divided by the travel time. It is NOT spot speed, or average speed measurements along the segment (referred to as time-mean speed in academic literature).

Travel time and speed data are typically collected on a quarterly basis. Data collection locations are selected based on discussion between the Coalition validation team and host member states. The goal is to identify road segments that are of interest for the member states while also advancing broader validation research initiatives. Initial validation efforts primarily targeted freeway facilities, and over time have increasingly placed greater emphasis on arterials (i.e., interrupted flow facilities due to traffic signals) and other more complex traffic flow scenarios.

Currently, the Coalition contracts with a data collection company that is responsible for deployment logistics. Thus, the Coalition validation team focuses primarily on planning the deployment and analyzing the subsequent data, while the responsibility of managing the sensor fleet, obtaining necessary permits, and installing and tearing down sensors is handled externally.

3.2 Data Processing Overview

WRTM reference travel time data requires a sequence of processing steps to transform raw MAC Id detections into filtered travel time and speed records that can be subsequently aggregated and used to characterize ‘ground-truth’, or reference, traffic conditions. Likewise, travel time and speed data provided by the vendor, and reported in 1-minute epochs along road segments defined by WRTM sensor locations, often needs to be adjusted to be aligned with the WRTM reference data. The validation team attempts to place sensors at logical junctions and landmarks that typically align with map segments from Traffic Message Channel base maps, however, often time field conditions, safety measures, and available mounting opportunities does not allow for an exact match. With the Traffic Data Marketplace, the Coalition is requiring support of the CATTWorks Georeferencing Protocol (CWCP). This will provide a common method to reference sensor and segment locations, and thus make validation procedures less resource intensive.

Figure 3-2 provides a high-level visual overview of the data processing steps. The WRTM processing workflow involves matching raw MAC IDs to form travel time records, filtering travel time observations to remove outliers, and grouping or aggregating the records to characterize ground truth traffic conditions for relevant analysis periods. The specific data processing details – particularly how the travel time observations are aggregated or grouped (e.g., 5-minute average travel/time speed estimates, percentile-based statistics summarizing typical hour-of-day travel times) – vary across analysis methods and are explained in further detail in the sections that describe each analysis. The data processing workflow ETC TDM vendor data mainly involves ensuring that the data feed corresponds to the same time periods as WRTM data and is aggregated and grouped in a similar manner.

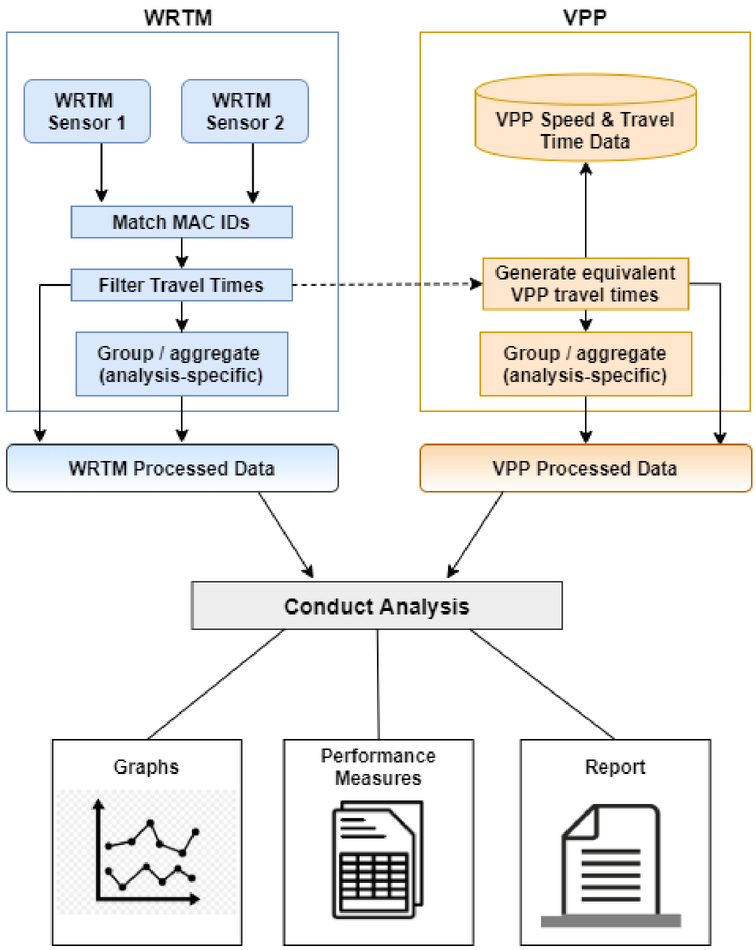


Figure 3-2. WRTM and VPP Data processing framework.

3.2.1 *WRTM Data*

Match MAC Ids and filter WRTM travel time observations

Raw travel time observations obtained by matching MAC IDs from adjacent WRTM sensors are first filtered to remove outliers. The filtering step is necessary because WRTM sensors sometimes re-identify vehicles that stop between sensors or record travel times from pedestrians or non-motorized vehicles that are not representative of actual traffic conditions.

Aggregate WRTM records

After outlier observations are removed, the remaining data represents valid travel time observations, and are grouped or aggregated to characterize traffic conditions for a given analysis period (defined by trip end times). Note that the aggregation process is analysis-dependent; some analyses involve summarizing data into highly granular five-minute bins, while others group observations across many days to quantify time-of-day patterns.

Various descriptive statistics can be used to summarize the central tendency and variability of WRTM samples for a specified analysis period. Examples of descriptive statistics include mean, median, standard deviation, coefficient of variation, and inter-quartile range. Additionally, WRTM sample data can be used to make inferences about the mean travel time and speed of all vehicles on the road via inferential statistics. Although aggregation details can vary slightly depending on the analysis context, the most common approach used are described below.

Traditional Analysis Aggregation

The aggregate speed estimate for each five-minute interval serves as the basis of comparison with the corresponding ETC TDM travel time/speed data feed, and therefore plays a crucial role in determining the results of the traditional validation process.

The best single-value estimate of the true space-mean speed is the space-mean speed of the WRTM travel time observations $\boldsymbol{tt} = \{tt_1, tt_2, \dots, tt_n\}$, which is shown in Eq. 3-1 and computed as the ratio of the segment length L to the average observed travel times samples $\overline{tt}_W$.

$$S_W = \frac{L}{\overline{tt}_W} = \frac{L}{\frac{1}{n}\sum_{i=1}^{n} tt_i} \quad (3\text{-}1)$$

To account for uncertainty in the sampling process, an interval estimate is constructed using a band whose width is proportional to the standard error of the mean speed. The bounds of this band (i.e., lower/upper bound speed, $[S^{lb}, S^{ub}]$) are defined in Eq. 3-2, where S_W is the space-mean speed of the WRTM samples from Eq 1 and SE_W is the standard error that is estimated based on Eq. 3-3 using the standard deviation of WRTM speeds SD_W, and sample size *n*. Eq. 3-4 defines the standard deviation SD_W, where speed observations $\{s_1, s_2, \dots, s_n\}$ represent the equivalent speed for each travel time observation $\{tt_1, tt_2, \dots, tt_n\}$, and $\bar{S}_w$ is the arithmetic mean speed of the samples from the interval, which is computed as shown in Eq. 3-5.

$$[S^{lb}, S^{ub}] = [S_W - 1.96 \cdot SE_W, \; S_W + 1.96 \cdot SE_W] \quad (3\text{-}2)$$

$$SE_W = \frac{SD_W}{\sqrt{n}} \quad (3\text{-}3)$$

$$SD_W = \sqrt{\frac{\sum_{i=1}^{n}(s_i - \bar{S}_w)^2}{n-1}} \quad (3\text{-}4)$$

$$\bar{S}_w = \frac{1}{n}\sum_{i=1}^{n} s_i \quad (3\text{-}5)$$

This band (Eq 3-2) serves as a proxy for the 95% confidence interval, and gets wider when the sample size is small or there is more variation in the observed WRTM speeds.

3.1.2 Vendor Data Processing

Vendor data is spatially aligned with the reference data through use of CWGP, but may not be aligned temporally. Vendor data is typically reported at regular intervals (e.g., 1-minute), with each record representing the average travel time / speed *during that specific period*. In contrast, WRTM travel time observations can span multiple time periods, but each trip is associated with the timestamp corresponding to its end time. Thus, WRTM sample data for a given interval represents the average travel time / speed *for trips ending during that specific period*. To facilitate a fair comparison with WRTM data, the reported vendor data is adjusted to ensure temporal alignment as explained in the next section.

Adjust vendor data records to reflect trip end times

The vendor data which reports average speed/travel time *during* each 1-minute interval is adjusted to reflect average speed and travel time for a vehicle that *ends* during the 1-minute interval. The simple example below illustrates this process.

Consider a 1-mile road segment with vendor data reported as shown in Table 3-1. Note the 8:05-8:06 period, which reports an average speed of 29 mph for trips observed *during the 1-min interval*. To compute the equivalent average speed of trips *ending* during this interval, begin by choosing the middle time value during this interval and consider it the trip end time. Using this as the ending point, work backwards in time to figure out the total travel time to traverse the segment using the speeds provided for each interval – a process shown step-by-step in Table 3-2. For example, the vehicle would travel 0.242 miles during the 30 seconds spent in the 8:05-8:06 interval, 0.533 miles in the 60 seconds spent in the 8:04-8:05 interval. At this point only 0.265 miles are remaining in the 1-mile segment, which will be traversed in approximately 28 seconds at an average speed of 34 mph. Given this information, the equivalent speed can be computed as the ratio of the total distance traveled to the total travel time; 30.5 mph (i.e., 118 s / 1.0 mi * 3600).

Table 3-1: Example VPP feed

8:01-8:02	8:02-8:03	8:03-8:04	8:04-8:05	8:05-8:06	8:06-8:07
36 mph	35 mph	34 mph	32 mph	29 mph	27 mph

Table 3-2: Adjusted Speed for 8:05-8:06 period to reflect trips *ending* during epoch

End	Start	Interval	Speed	Time	Distance
8:05:30	8:05	8:05-8:06	29 mph	30 s	0.242 mi
8:05	8:04	8:04-8:05	32 mph	60 s	0.533 mi
8:04	8:03:32	8:03-8:04	34 mph	28 s	0.265 mi
Sum				118 s	1.0 mi
Adjusted Speed			30.5 mph	(total distance / total time)	

Aggregate VPP records

Afterwards, the adjusted vendor records are aggregated and grouped to the same temporal period as the reference data. The specific aggregation process is analysis-specific, but the most common use case is computing the average travel time and speed for each 5-minute period for the traditional validation analysis.

Note that unlike the WRTM aggregation (which deals with individual travel time observations), VPP aggregation involves grouping records that each represent the mean travel time and speed for a more granular time period. Most commonly, this means aggregating 1-minute records into 5-minute records. When computing average speeds in this manner, the harmonic mean speed is used. Harmonic mean speed is equivalent to computing the ratio of total distance to total travel time for all records.

3.3 Validation Methods

The following subsections describe the various validation techniques that have been developed and used for validating travel time and speed probe data in VPPI and VPPII. The so-called "traditional analysis" -- initially developed to quantify accuracy on freeway facilities and subsequently adjusted for lower-speed facilities -- has been used primarily for contract compliance in past iterations of the Vehicle Probe Project. Over time, other analyses were developed for specific purposes (e.g., focusing on congested time periods, investigating latency in data feeds, quantifying performance on signalized arterials).

Note that the validation team reserves the right to update the methodology based on best practices and introduce any additional techniques that improve the evaluation of vendor data quality. For example, in an extremely low-volume setting, temporal aggregation may need to be expanded from 5 minutes to 15 minutes to have an adequate number of observations for comparison.

3.3.1 Traditional analysis

The traditional validation analysis compares vendor speeds to estimated ground truth speeds over 5-minute intervals (any reference to 'speed' implies space-mean speed unless otherwise noted), and quantifies the discrepancy in terms of two error key error metrics: Average Absolute Speed Error (AASE) and Speed Error Bias (SEB).

Computing Error Measures

In the following equations S_R^{ij} represents the reference space-mean speed for trips ending in 5-minute interval *i* along segment *j*, while S_V^{ij} denotes the probe vendor's space-mean speed from the corresponding period and segment.

Note that S_R may be obtained via single-value or interval estimates (as described in Section 3.2.1) In cases where an interval estimate is used for S_R while computing error measures (as is the case for evaluating vendor compliance), the error between S_V and S_R is considered zero when S_V falls anywhere within the confidence band limits.

Average Absolute Speed Error (AASE)

AASE is calculated by summing the absolute difference between probe vendor speeds and reference speeds for each time interval *i* and segment *j*, and taking the average over all T observations as shown in Equation 3-6. Because the absolute value is used, positive and negative errors cannot cancel, and the result is always positive.

$$\text{AASE} = \frac{1}{T}\sum_{i=1}^{n}\sum_{j=1}^{m}\left|S_V^{ij} - S_R^{ij}\right| \qquad (3\text{-}6)$$

Speed Error Bias (SEB)

Speed Error Bias is calculated similarly except the absolute value of the errors is not taken. Thus, positive and negative errors can cancel each other out, and the resulting value can provide insight into whether there is a consistent positive or negative error.

$$SEB = \frac{1}{T}\sum_{i=1}^{n}\sum_{j=1}^{m}(S_V^{ij} - S_R^{ij}) \quad (3\text{-}7)$$

Reporting considerations

Table 3-3 summarizes a typical summary of results in table format, where results are summed over all validation segments. For each speed range (which are defined separately for freeways and arterials) the AASE and SEB are calculated against the 1.96 SEM band as well as against the mean.

Table 3-3. Typical reporting of contract quality metrics.

Speed Bin	Average Absolute Speed Error (<10mph)		Speed Error Bias (<5mph)		Number of 5 Minute Samples
	Comparison with SEM Band	Comparison with Mean	Comparison with SEM Band	Comparison with Mean	
0-30	1.66	2.57	1.2	1.56	2451
30-45	3.05	4.87	0.48	0.96	2036
45-60	2.5	5.56	1.17	2.91	8678
60+	1.23	3.86	0.04	-0.04	34651
All Speeds	1.56	4.15	0.32	0.62	47816

Results are also reported separately for each validation segment, and are occasionally also computed for different scenarios – examples of which include peak vs off-peak periods, special events (e.g., before/after a signal re-timing test, hard-shoulder running during specific times of day), varying levels of speed variation, or for different subsets of the data feed (e.g., different confidence scores).

Freeway and Arterial Accuracy Specification

Data quality requirements are based on AASE and SEB error metrics, which are computed using the WRTM interval estimate as the reference speed against which the VPP data is compared. These error measures are calculated separately in four distinct speed bins, with each 5-minute period assigned to the speed bin associated with the WRTM space-mean speed. Table 3-4 defines the speed bins for freeway and arterial facilities, and specifies the allowable values for AASE and SEB measures.

Table 3-4. Data quality requirements for freeway and arterial facilities

Road Type	Speed Bins	AASE Specification	SEB Specification
Freeway	0-30, 30-45, 45-60, 60+ mph	5 mph	4 mph
Arterial	0-15, 15-25, 25-35, 35+ mph	8 mph	4 mph

3.3.2 Slowdown Analysis

The slowdown analysis is an offshoot of the traditional analysis, developed to provide a more intuitive measure of probe data's ability to capture congestion events– a perspective particularly relevant for operations applications and arterial facilities.

At a high level, it involves visually analyzing 24-hour time series plots to identify slowdown events in the WRTM reference data, and then classifying VPP data based on its ability to capture each slowdown. The plots used to facilitate this process are generated as part of the traditional analysis, and thus this this method not require any additional data processing or aggregation techniques.

Slowdown Definition

The definition of a slowdown event is when the mean WRTM speed decreases by at least 15 mph for a minimum duration of an hour (although on slower arterials this definition may be adjusted to speed reductions of 10 mph for at least 30 minutes).

Slowdown Classification

This analysis is a manual classification task performed by an analyst, who first identifies slowdown events based on the previously-described criteria, and then visually compares ground truth and vendor data during each relevant period, using a set of rules to determine whether each vendor "Fully Captures", "Partially Captures", or "Fails to Capture" each slowdown, resulting in a summary table as shown in Table 3-5. These classification rules are based on the magnitude and duration of speed reduction (shown visually in Figure 3-3), and are defined as follows:

- A *Fully Captured* slowdown indicates that the probe data accurately characterized both the reduction in speed, and duration of the slowdown (within 20%)

- A *Partially Captured* slowdown indicates that the probe data reported a significant disruption to traffic, but the extent of speed reduction or duration of time were in error by more than 20%.

- *Failed to Capture* indicates that the probe data either completely missed the slowdown, or the extent of speed reduction or duration of the event were significant in error such that the slowdown would not be interpreted as a significant disruption to traffic.

Figure 3-3: Magnitude and duration differences in WRTM and VPP data during slowdown.

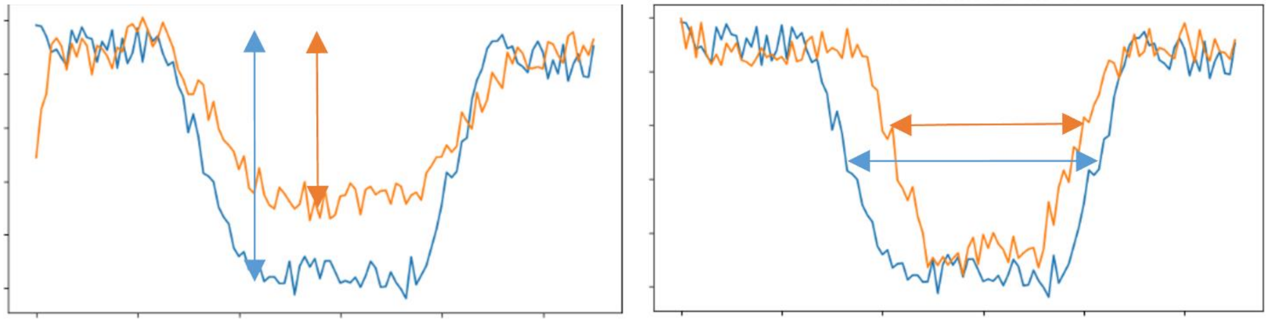


Table 3-5. Example **s**lowdown analysis summary results.

Vendor	Significant Slowdowns	Fully Captured	Partially Captured	Failed to Capture
Vendor X	66	43	21	2

Although this process is currently conducted manually, it has been found to be one of the most useful ways for quantifying vendor probe performance on arterial facilities. Initial efforts have been undertaken to partially automate this process, which may be incorporated into future validation efforts.

Additionally, the validation team envisions correlating the propensity to capture slowdowns with road characteristics (e.g., signal density, AADT) and arterial traffic signal performance measures when signal data is available.

Slowdown Contract Specifications

Data quality requirements are based on the following slowdown classification percentages. Because slowdown observations are sometimes sparse, the prevailing practice is to accumulate over multiple validations, and report when there are at least 10 or more slowdown observations to increase statistical significance.

- Failed to Capture: < 15% (mandatory)
- Fully Captured: > 70% (desired)

3.3.3 Sampled Distribution Method

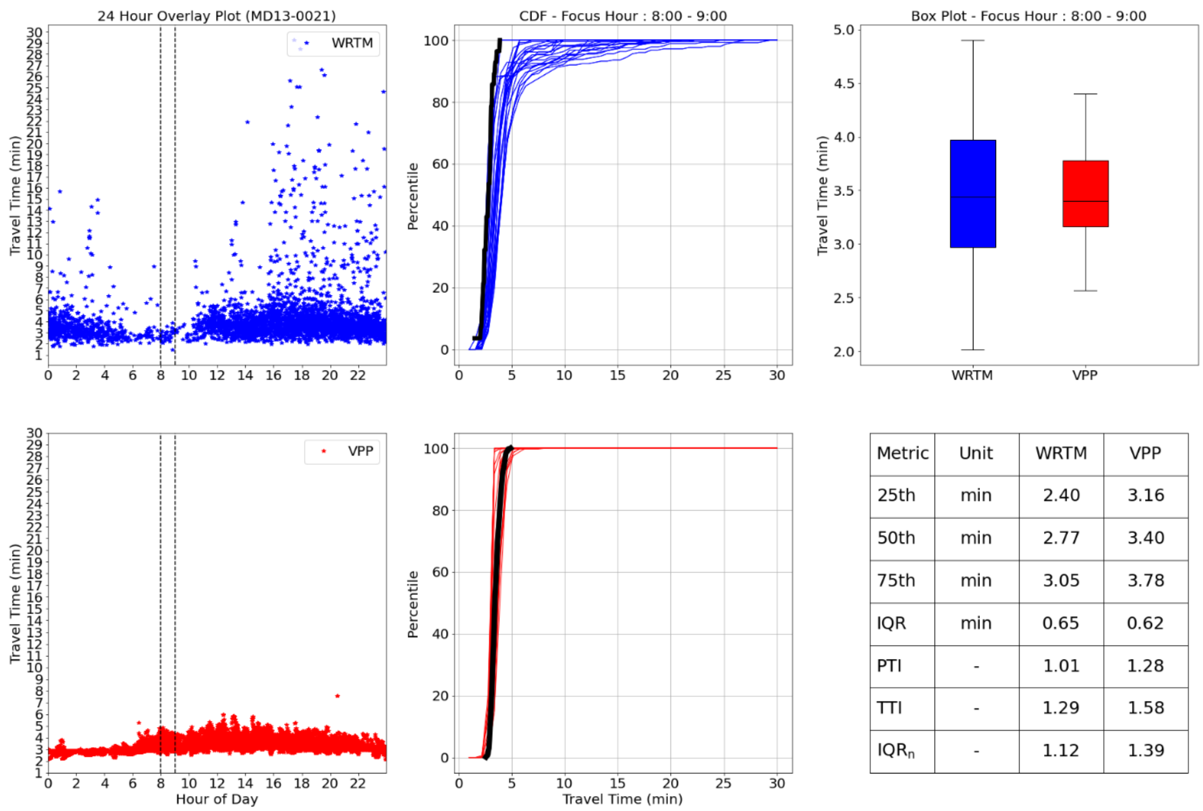
In contrast to the traditional and slowdown analyses that aggregate data in 5-minute bins and quantify data quality for each point in time, the sampled distribution method seeks to quantify the extent to which probe data captures *repeatable* travel time patterns. Accordingly, it aggregates the data over multiple days, grouping WRTM travel time observations and VPP records into time-of-day periods (e.g., 7-8 AM on weekdays), using overlay plots and empirical cumulative distribution function (CDF) curves to summarize the travel time distributions of each analysis period. By comparing empirical cumulative distribution function (CDF) curves for WRTM and

VPP data – both visually and via percentile-based statistics, this method seeks to quantify the extent to which probe data captures repeatable traffic phenomena.

Overlay Plots and Empirical CDFs

Overlay plots are constructed by taking multiple days of observation and graphing them on a single 24-hour timeline, a process that typically only includes weekday data. This method of overlaying data on a single 24-hour plot reinforces repeatable traffic phenomenon, enhancing the density of travel-time samples and thus increasing the visible detail of any recurring congestion. Figure 3-6 (left panel) shows an example of an overlay plot for an arterial segment, where each travel time data point collected using WRTM equipment on a weekday from an approximate two-week period is graphed on a 24-hour timeline.

The relative density of the data provides a visual indication of the probability of traversing the corridor at the travel time indicated on the y-axis. A corresponding distribution analysis is constructed from the data in the overlay plot. Each curve in a distribution analysis (called a cumulative frequency diagrams or CFDs) is constructed from the percentiles of the travel time data in the overlay plot. Figure 3-6 (middle panel) shows example CDF curves based on the overlay plots.



Metric	Unit	WRTM	VPP
25th	min	2.40	3.16
50th	min	2.77	3.40
75th	min	3.05	3.78
IQR	min	0.65	0.62
PTI	-	1.01	1.28
TTI	-	1.29	1.58
IQR_n	-	1.12	1.39

Figure 3-6: Sample distribution analysis -- overlay plots (left), empirical CDF curves (middle), and boxplots/stats (right)

Comparing Empirical Distributions

Because the empirical CDFs summarize the *entire* travel time distribution (rather than just the mean value), this method can be used to identify segments with complex traffic flow patterns and travel time variation that would otherwise be missed when only looking at mean/median values. Empirical CDF curves can be compared visually, and additionally relevant percentiles can be extracted and used to form the basis for performance measures. The right panel in Figure 3-6 shows a table summarizing key percentiles and performance measures as well as boxplots comparing the travel time distributions -- formed using important percentiles from the empirical CDF (e.g., min, 25/50/75 percentile, max). Some key summary statistics include:

- Travel Time Index (TTI)
- Planning Time Index (PTI)
- Percentile summaries: 5^{th}, 10^{th}, …95^{th} percentile
- Median: 50^{th} percentile, a measure of central tendency
- Inter Quartile Range (IQR): 75^{th} percentile minus 25^{th} percentile – a measure of dispersion

Currently, this process relies on manual inspecting overlay plots and empirical CDF curves to gain insight into repeatable patterns. In the future, some of the WRTM and VPP percentile-based statistics may be used as the basis for calculating overall error/performance measures. As an example, averaging the difference between WRTM and VPP (normalized) median travel time across all paths and hours of the day would give an overall sense of the extent to which VPP data captures typical travel time values. Repeating this process with IQR or PTI (travel time reliability-based measures) would focus on the extent to which variation in the WRTM data is reflected in the VPP data.

3.3.4 Latency Measurement

Latency within the Vehicle Probe Project is defined as the difference between the time the traffic flow is perturbed as a result on an incident and the time that the change in speed is reported in the traffic data. Conceptually, latency is measured by observing the time difference between the onset of a slowdown as reported by WRTM traffic monitoring, and that reported by the VPP. A graphical representation is shown in Figure 3-7, where the Bluetooth label represents the WRTM data feed is represented by the blue line

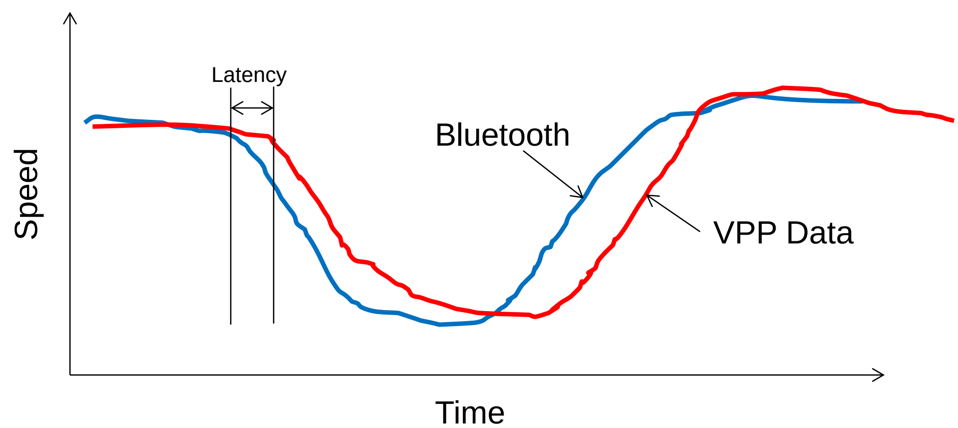


Figure 3-7. Measurement of Latency

Measurement of latency is on an event by event basis, where an event represents a major slowdown. The proposed basis for measuring latency within the VPP is as follows:

- WRTM data is aggregated to a one-minute basis.
- Slowdown events are identified within the Bluetooth data from which to measure latency. Candidate criteria for slowdowns is a decrease in traffic speed of 20 mph within a 10-minute period, and speed remains below the 20 mph for at least 15 minutes as illustrated in Figure 3-8.

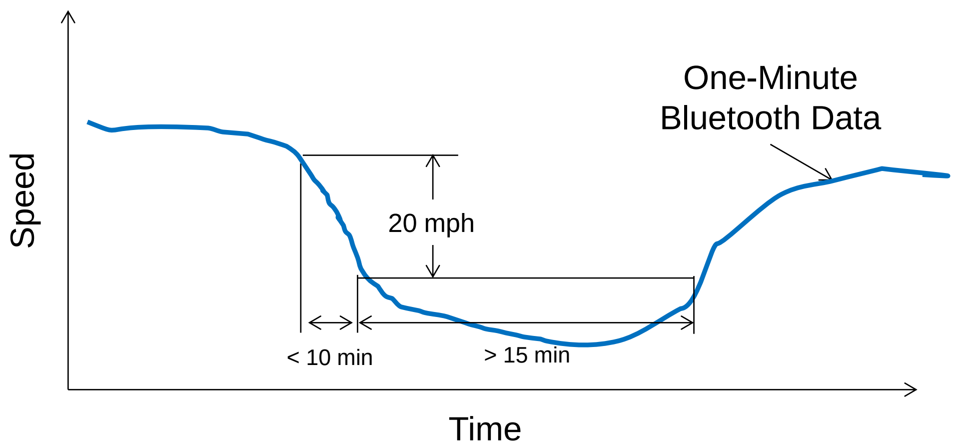


Figure 3-8. Determination of a Slowdown Event

- VPP data feed is correlated to the WRTM data with a variable time offset. The time offset that maximizes the correlation during the slowdown between the Bluetooth and VPP data determines the latency for the slowdown event.
- The average latency is determined by the average of individual latency calculations across the entire validation data set.

4. TRAFFIC VOLUME VALIDATION

Traffic volume data is new to the ETC TDM, and as such has not yet been validated independently by the ETC. Although specific methodologies have not yet been specified, vendor volume data will be subject to highly-quantitative validation processes, and assessed against reference sources for accuracy.

Methodology Development Plan
The volume data validation team will leverage best practices identified from related volume estimation research, including the following:

- Methodology developed for USDOT project: "Non-Traditional Methods to Obtain Average Annual Daily Traffic (AADT)
- Lessons learned from the ETC Volume and Turning Movement (VTM) research initiatives in collaboration with the National Renewable Energy Laboratory and the University of Maryland Center for Advanced Transportation Technology

Additionally, a validation committee consisting of technical representatives from state DOTs will be formed (The ETC Data Validation Committee). This committee will help guide the process for evaluating volume data quality.

Possible Reference Data Sources

The best way to evaluate volume data products is via an independent data source. This may not always be possible or economically feasible, but some options include:

- **Short-term field data collection** (similar to travel-time based WRTM collection). This would most likely involve contracting with a company to provide the data collection service. The physical vehicle counts could be conducted via a variety of technologies (e.g., pneumatic tubes, video-based systems, manual).

- **Obtain short-term count data from local jurisdictions.** Local jurisdictions may be willing to share short-term 24 or 48-hr count data during specified evaluation dates.

- **Obtain video feed from CCTV cameras.** If real-time video can be streamed during a desired collection period (e.g., from DOT CCTV cameras), it can be archived for subsequent analysis. The vehicle counts can then be extracted from the archived footage either via manual counting or video analytics software.

Additionally, it may be worth using publicly-available volume datasets (e.g., count data from CCS / ATR stations) as well. However, it is important to note that accuracy that these locations may not be completely representative of data quality across the road network – particularly if VPP vendors have used this data to train their volume estimation models.

Error Measures

When reference traffic volume data are available, the data quality can be quantified in terms of various error measures. Three error measures used as part of the ETC Volume and Turning Movement project include Symmetric Mean Absolute Percentage Error (SMAPE), Error-to-Max Flow (EMFR) ratio (or Error to Theoretical Capacity Ratio – ETCR for freeways), and Coefficient of Determination (R^2), shown in Eq. 4-1, 4-2, and 4-3, respectively. In each equation $y_i =$ observed volume, $\bar{y}_i =$ average observed volume, $\hat{y}_i =$ model volume estimate, and $y_{\max} =$ max volume (i.e., theoretical capacity in ETCR and observed capacity in EMFR).

$$SMAPE = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{|\hat{y}_i - y_i|}{(|\hat{y}_i| + |y_i|)/2}\right) x\ 100 \tag{4-1}$$

Note: SMAPE may be implemented with weighting factors, in which observed referenced volumes are used as weighting factors, putting greater emphasis on higher volume roadways.

$$ETCR/EMFR = \left(\frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_{max}}\right|\right) x\ 100 \tag{4-2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{4-3}$$

- **Symmetric Mean Absolute Percentage Error (SMAPE)**
  - Reflects absolute volume accuracy
  - Minimum accuracy expected to be:
    - 10-15% for high volume roadways (Freeways)
    - 20-25% for mid volume roadways (Arterials)
    - 30-50% for low volume roadways (Local Roads)

- **Error to Capacity Ratio / Error to Max Flow Ratio (ETCR / EMFR)**
  - Captures accuracy relative to capacity (either theoretical capacity on freeways or max observed flow on non-freeway facilities)
  - Minimum accuracy expected to be:
    - 10% (5% desired) for high volume roadways (Freeways)
    - 20% (10% desired) for mid volume roadways (Arterials)
    - 20% for low volume roadways (Local Roads)

- **Coefficient of Determination (R2)**
  - Shows explanatory power of model
  - Minimum accuracy expected to be:
    - 0.70 for high volume roadways (Freeways)
    - 0.80 for mid volume roadways (Arterials)
    - 0.90 for low volume roadways (Local Roads)
  -

While SMAPE is perhaps most intuitive, it is important to note that its interpretation is volume-dependent; the denominator contains the true traffic volume. Thus, for the same absolute error, it will naturally be higher on lower volume roads and lower on higher-volume ones. In contrast, EMFR does not suffer from the same challenges, as its denominator uses the roadway capacity (ETCR for freeways) or observed maximum flow (EMFR for non-freeways).

Reporting Considerations

It can be useful to compute error measures separately for each validation location, and then consider the distribution of error measures across locations – an approach that minimizes the oversize impact of extreme values (either good or bad) and provides a more typical insight into overall performance. Additionally, evaluating the error measures separately for different road classes or volume levels can provide additional insight into performance.

5. VALIDATING OTHER DATASETS

Other datasets procured through VPP may be validated if deemed appropriate by the validation team and/or committee. While the scope of such a validation will likely be limited to QC/QA (rather than an evaluating based on field data collection), the validation team reserves the right to investigate the datasets and publish its findings.

Examples of QC/QA methods may include:

- **Comparing datasets to other known sources**. For example, a dataset detailing vehicle trips can be sanity-checked by comparing the number of trips (or resulting Origin-Destination travel patterns) with those obtained via the National Household Travel Survey (NHTS) or Travel Demand Model. Figure 5-1 shows an example conducted by NREL, which shows agreement between a trips dataset and the NHTS results. While this type of comparison is generally not appropriate for computing quantitative error measures, it can give feedback about whether the data is consistent with other sources.

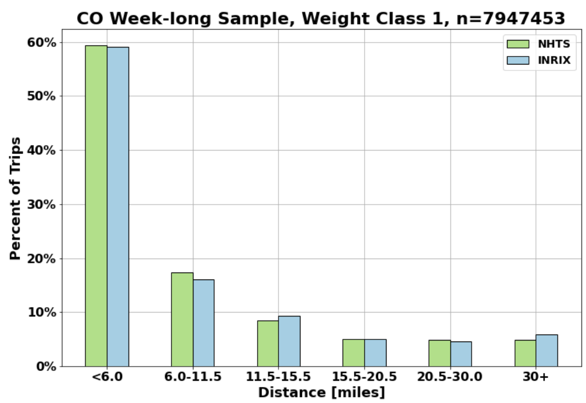


Figure 5-1. Comparing Trips datasets to NHTS
(example from National Renewable Energy Laboratory)

- **Visualizing datasets.** Visually inspecting datasets can often provide useful insight into datasets and help sanity check its characteristics. For example, Figure 5-2 shows an example trajectory dataset in Maryland, with the left plot showing way GPS waypoint locations, and the right one showing a heatmap where data points are most concentrated. The waypoints are densest in the Washington, D.C. and Baltimore, MD areas as well as the corridor that connects them, an intuitive result.

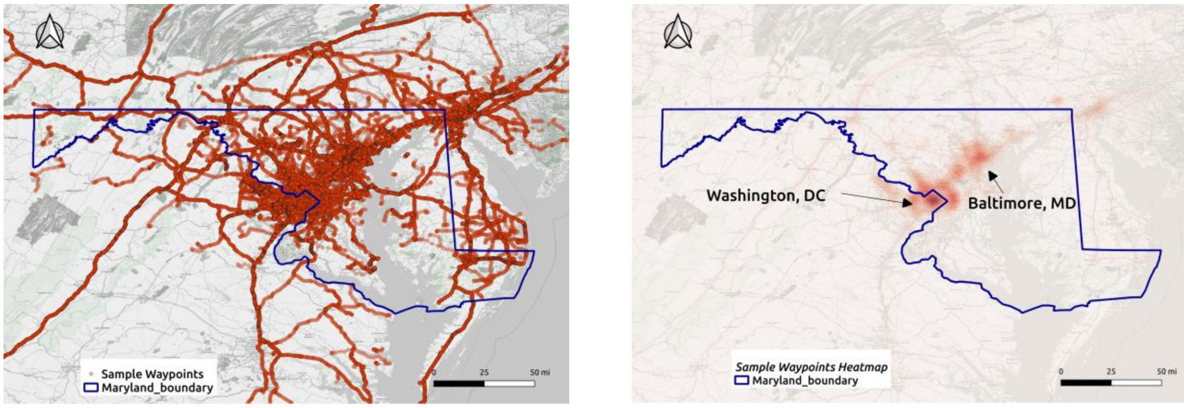


Figure 5-2: Sample visualizations of a trajectory dataset with raw waypoint locations (left) and a heatmap of waypoint density (right).

- **Reporting descriptive statistics.** Even if a reference dataset is not available for comparison, it can be useful to summarize basic descriptive statistics for datasets provided in the ETC TDM. For example, a trajectory dataset could be summarized in terms of the number of waypoints, number of trips, trip distance, distance between waypoints, time gap between waypoints, etc. Figure 5-3 provides an example for a trips dataset, showing the distribution of trip distance and duration via empirical CDF plots.

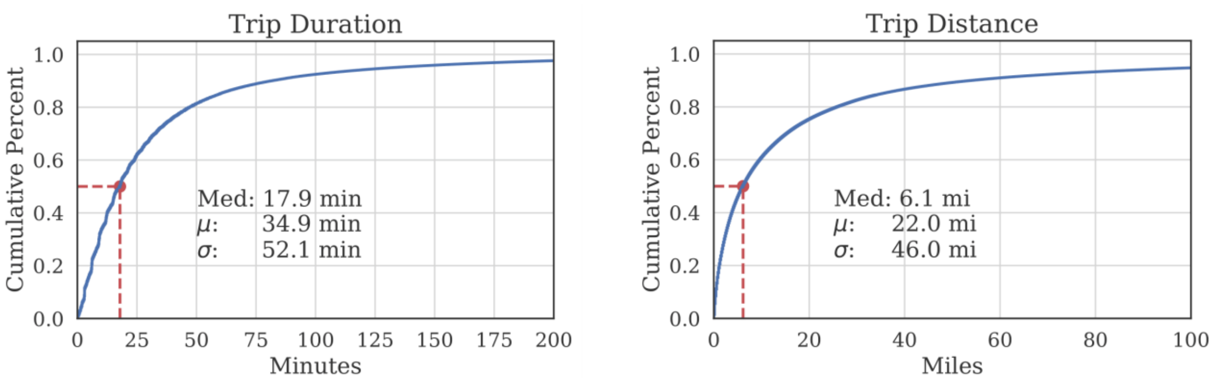


Figure 5-3: Example of descriptive statistics for a trips dataset.