Initial TDM Validation Activity: Volume & Travel Time Study

TETC Transportation Data Marketplace Data Validation

Prepared by: TETC Validation Team

4/6/2023

Initial TDM Validation Activity: Volume & Travel Time Study

TETC Transportation Data Marketplace Data Validation

Kickoff validation activity for TDM Validation program, focusing on Volume and Travel Time data in North Carolina

TDM-VAL-02

***Data Categories**: Volume, Travel Time*

***Evaluation Period**: Aug/Sep 2022*

***Publication Date:** March 2023*
***Release to Public:** September 2023*

4/6/2023

The Eastern Transportation Coalition is a partnership of 17 states and the District of Columbia focused on connecting public agencies across modes of travel to increase safety and efficiency. Additional information on the Coalition, including other project reports, can be found on the Coalition's website: www.tetcoalition.org

Executive Summary

This report describes the inaugural TDM data evaluation effort conducted by The Eastern Transportation Coalition (TETC) data validation team, which focuses on two of the six data items available in the Transportation Data Marketplace (TDM): Volume and Travel Time/Speed data. The two components of this activity, Volume and Travel Time/Speed validation, were conducted concurrently in North Carolina, but represent distinct activities with different evaluation methods and participation by different (albeit overlapping) sets of data vendors. Although some anonymized results are shared in the report, the primary objective of this inaugural effort is to test methods, metrics, and logistics associated with the revised validation process, which has evolved significantly since Vehicle Probe Project Phase II (VPPII) and now incorporates additional vendors and multiple data types.

Travel Time

Although Travel Time/Speed probe data has been well-studied in previous Coalition validation activities, the new TDM validation program includes several changes that are designed to modernize the program, improve efficiency, and enable more meaningful measurements of latency. These changes include (i) the use of a standardized georeferencing protocol to define custom reporting segments (instead of using Traffic Message Channel (TMC) codes or other segmentation methods), and (ii) a real-time delivery requirement. Given these changes, plus the addition of several new vendors, the emphasis of this activity was to work through the logistics of data delivery and protocols.

The study area consisted of 10 freeway segments along I-40 and I-95 just South of Raleigh, North Carolina. These segments were chosen with the expectation of capturing congestion on a beach route between Raleigh and Wilmington over Labor Day weekend (although no significant congestion was observed) and also because these segments represented a basic freeway scenario for which probe data is known to provide high quality data. This scenario was designed to allow vendors to focus on new data delivery formats and submission logistics instead of challenging traffic conditions. Reference data was collected between August 31 and September 9, 2022, and corresponding travel time vendor data was provided by four of the five vendors, with one unable to prepare data for review in this round.

Although this travel time validation activity did not provide an opportunity for significant validation analysis (due to the lack of congestion on the chosen segments)**, data vendors adjusted well to the new protocols and should be ready for future activities.** Furthermore, **use of the georeferencing protocol greatly improved the efficiency of the validation**, removing the resource burden of conflating data sets from multiple vendors and streamlining the entire process. Key points include the following:

- **New georeferencing / data delivery requirements were successfully implemented**.
    - The CattWorks Geo-referencing Prototcol[1] (CWGP) was used to describe custom reporting segments, and all four vendors who submitted were able to respond using that prototcol.
    - Vendor data was delivered in real-time to the validation server from three of the four vendors who submitted data, allowing the validation team to directly access latency.

[1] CATT Works Georeferencing Protocol (link)

- **Minimal speed fluctuations were observed during the study period.** 99.9% of 5-min time periods recorded in the reference data were at or near free-flow speed. As a result, the typical error metric calculations (AASE/SEB) were not statistically significant to reflect the ability to capture slowdowns. The handful of periods with congestion were assessed visually by comparing vendor provided data overlaid on the reference data. The performance was within the expected bounds for this small sample.

- **New processes did not appear to impact data quality**
  - Vendor-provided travel time and speed data generally tracked the Bluetooth collected travel time and speed data for the handful of speed fluctuations in the data set.
  - Vendors tended to follow the faster mode when multiple travel time modes were present (consistent with observations from VPPII, sometimes referred to as an optimistic bias).

As a result of the increased validation efficiency resulting from implementing the CWGP, future validation activities be timelier and more responsive, and the Coalition will gain capacity to evaluate additional vendor data with minimal resource expenditure.

Volume

The methodology used to evaluate volume data is based on recommendations from a previous report[2] and feedback from the TDM Validation Technical Advisory Committee (TAC). This initial volume validation activity focused on *hourly volume counts* - one of the four 'mandatory' volume data deliverables required per the RFP. The activity encompassed a blind evaluation of vendor data using reference counts collected by North Carolina DOT (NCDOT) and a Coalition contractor using a variety of error metrics and visual plots. As with the travel time and speed validation process, the objective of this first volume validation was to test the procedures and data delivery formats. No vendor specific results will be published.

The study area encompassed the entire state of North Carolina, with specific evaluation locations chosen based on where reference hourly count data could be acquired. This included a combination of NCDOT Continuous Count Stations (CCS), a sample of NCDOT short term count sites that coincided with the study period, and a handful of short-term counters deployed by TETC's validation team. In total, vendors were asked to report hourly counts at 662 directional roadway locations between August 15 and September 15, 2022. **Two of the four vendors successfully delivered estimated counts for this activity. The remaining two vendors were unable to due to data supply chain or subcontractor issues.**

As was the case with Travel Time/Speed data, a primary emphasis was to work through the data formats and delivery logistics. However, in contrast to Travel Time/Speed evaluations, which have been conducted by the Coalition for 15+ years and have well-defined processes, the Volume evaluation methodology is still under development. As such, a key goal was to test the suitability of various error metrics for quantifying vendor data quality. Several key takeaways emerged from the resulting evaluation and follow-up discussions with vendors:

[2] TDM-VAL-01: Traffic Volume Validation – Literature Review and Recommendations (link)

- **Further work is needed to define appropriate error metrics and provide context for their interpretation.** Some of the error metrics used in this analysis do not match the insight gained from visual inspection of the time series plots. Percent error metrics were found to be highly sensitive to overall traffic volumes and generally not reflect patterns revealed in time-series plots. *Guidance is needed for setting appropriate targets, recognizing that the data may be useful at different levels of accuracy depending on the intended application.*

- **Most vendors are currently concentrating their product development efforts on higher time aggregations such as AADT, ADT and AHDT**. This validation activity, which focused on hourly counts (the most granular of required data items required in the RFP), is most appropriate for operations performance measures. Although a supported product by the vendors, volume products are more mature for the planning level measures. Future validations will incorporate more temporally aggregate volume products such as AADT, ADT, and AHDT.

- **Unlike travel time, volume estimation requires engineering resources for each exercise.** Vendors requested more advanced warning (3+ months) to plan resources to respond to the validation activity. As a result, the validation team will strive to give vendors adequate time to efficiently prepare for future validations, and request data submittal consistent with expectations set within the TDM contractual obligations.

- **Volume estimates, unlike travel time and speed, are highly sensitive to perturbations in the data supply chain.** Delays may occur due to disruptions of base data supply requiring re-engineering of the process.

- **Georeferencing Protocol worked as designed for providing for efficiency of the volume validation process, similar to the travel time and speed validation.** Both vendors who participated were able to locate each volume validation location using the CWGP and submit volume estimates in this protocol. The CWGP will continue to be used in future activities.

- **Need to begin developing cross-validation procedures.** It is expected that self-reported cross-validation metrics will complement blind evaluation studies to characterize the accuracy of probe-based volume estimation models. The validation team plans to work with the TAC, industry partners, and vendors to develop a “cross-validation audit” to integrate into the evaluation framework.

Table of Contents

Introduction

Transportation data sold through the Eastern Transportation Coalition (TETC) Transportation Data Marketplace (TDM) is procured from private industry based on contract specifications. The intent of the Coalition's validation program has evolved from the original Vehicle Probe Project validation which was primarily limited to ensuring that traffic data conforms to contractual standards. The TDM includes that essential function, but also has flexibility to adjust to the needs of the Coalition members as the market evolves and data needs expand. The validation process is overseen by a technical advisory committee that sets general direction and review results. The TDM includes both quantitative and qualitative analysis of datasets available through the marketplace as appropriate for each data type. The marketplace currently contains six core data items: Travel Time/Speed, Volume, Waypoint, Origin-Destination, Freight, and Conflation, with all but one (Travel Time/Speed) being sold through the marketplace for the first time. As such, the validation team, under the guidance of the TETC Validation Technical Advisory Committee (TAC), is beginning to establish standards and methods for effectively evaluating data quality and value across the different data sets.

This validation report evaluates two of the products sold in the TDM: Travel Time/Speed data, a follow-on from VPPII, and Volume data, the new data item identified as highest priority by the TAC. Based on the TAC's direction to focus on Volume data, the validation team previously produced a report, TAC-VAL-1, containing a literature review and recommendations for validating volume data. The strategy outlined in TAC-VAL-1 guides the methodology used to evaluate vendor Volume data in this report.

As part of the validation study, reference Travel Time and Volume datasets were collected and used as the basis for evaluating reported vendor Travel Time/Speed and Volume data. Although a few anonymized accuracy measures are included in the report, the primary objective of this effort was to test methods, metrics, and logistics associated with the validation process, which included new location referencing protocols, data formats and delivery expectations. During this process, the validation team collected feedback from vendors and took notes on what worked well, and areas in need of improvement allowing for improvement and updates to the process moving forward. As such, the focus of this report is on the lessons learned by walking through the data collection, vendor submission, and evaluation process, with future reports tackling data accuracy in more detail on a vendor specific basis.

Data Vendors

All vendors selected through the TDM RFP process in the Travel Time/Speed and Volume categories were invited to participate in the validation study (for Travel Time/Speed Carto, HERE, INRIX, Iteris, Timmons; for Volume HERE, INRIX, Iteris, Streetlight). The validation team set ambitious expectations for data submission, including using newly introduced georeferencing protocol (CWGP) to describe validation locations and requiring Travel Time/Speed data to be submitted in real time. Recognizing the challenges associated with these processes and the quick turnaround time required, vendors were encouraged to meet the more stringent requirements but given the nature of this first exercise late submissions and, in the case of TT/Speed data, non-real time submissions were accepted.

Four of the five Travel Time/Speed vendors successfully submitted data in the travel time evaluation (one was unable to prepare data for review), and two of the four Volume vendors successfully submitted data in the volume evaluation (the other two had data supply chain and subcontractor changes, respectively). The Results section below contains more information about each vendor's submission.

Travel Time

Geographic Scope

Validation roadway locations were communicated to the data vendors using CATT Works Geo-referencing Protocol (CWGP). CWGP provides a simple mechanism for describing roadway locations based on geodetic coordinates, headings, and a few basic attributes, and does not require the sender and receiver to use a common map or agree to a specific pre-coded segmentation scheme (e.g., TMC segments). Exercising this protocol for the first time was a key aspect of the validation activity.

Figure 1 shows the geographic scope of the travel time study area, which consists of 10 freeway segments along I-40 and I-95 just South of Raleigh, NC. This study area was chosen for two reasons: (1) the expectation of capturing traffic congestion on a beach route between Raleigh and Wilmington over Labor Day weekend and (2) because it represented a basic freeway scenario for which probe data is known to be high quality from former VPI & II validation activities, thus allowing vendors to focus on new data delivery formats and submission logistics. The segment definitions were chosen to coincide with available volume counter locations, and endpoint locations were chosen where Bluetooth re-identification sensors could be safely mounted.

Figure 1 – Validation segments shown relative to nearby cities (left) and zoomed in (right)

Table 1 shows the key attributes used to communicate the validation segment locations to vendors via CWGP. This information was shared with vendors as a GeoJSON file with MultiPoint coordinate geometry (representing the start and end points of each segment) so that it could be readily processed with a computer or displayed via web or Geographic Information System (GIS) tools.

Table 1 - CWGP attributes for travel time validation segments

ID	Start Location	Start Heading	End Location	End Heading	Road Name	Road Class	Length (Feet)
NC2022-01	-78.53274 35.43259	167	-78.52824, 35.41874	153	I-40	1	5233
NC2022-02	-78.52824 35.41874	153	-78.46637 35.34118	124	I-40	1	34858
NC2022-03	-78.46637 35.34118	126	-78.32878 35.21419	128	I-40	1	65686
NC2022-04	-78.32878 35.21419	126	-78.30583 35.20481	109	I-40	1	7696
NC2022-05	-78.3058 35.20499	290	-78.32866 35.21437	306	I-40	1	7671
NC2022-06	-78.32866 35.21437	307	-78.46618 35.34129	307	I-40	1	65668
NC2022-07	-78.46618 35.34129	304	-78.52805, 35.41886	333	I-40	1	34856
NC2022-08	-78.52805, 35.41886	334	-78.53251 35.43266	347	I-40	1	5213
NC2022-09	-78.53648, 35.37894	60	(-78.4624 35.41751	70	I-95	1	26246
NC2022-10	-78.46245, 35.41768	250	-78.53661 35.37908	240	I-95	1	26275

Data Collection

Reference Data

Reference travel time samples were collected along the 10 directional freeway validation segments from August 31 to September 9, 2022. Bluetooth re-identification sensors were deployed by a company that specializes in field collection, and the resulting raw data was processed by the Coalition Validation team to obtain travel time measurements.

Vendor Data

Vendor travel time/speed data was delivered in real time to a submission server managed by TETC's Validation Team. Vendors were instructed to report the average travel time/speed values for all validation segments during each minute of the collection period. In contrast to how data submission was conducted in previous iterations of validation, this approach introduced two key changes: vendors provided data along CWGP-defined validation segments (rather than TMCs or other segmentation) and did so in real time. These changes are expected to improve validation efficiency, help enable fairer comparisons with reference data, and provide a common time base for computing latency.

Vendors demonstrated a high degree of compliance with the new data delivery requirements and geo-referencing protocol. As Table 2 shows, four of the five vendors delivered data for this activity. All vendors that delivered data adhered to the georeferencing protocol and delivery specifications, and all but one delivered in real time. The validation team met with each vendor to discuss this process and is working with the one vendor that was unable to deliver to resolve data pipeline issues that were unearthed in the exercise.

Table 2 – Summary of travel time data delivery by anonymized vendor

Vendor	Data Delivery		Used Geo-Referencing Protocol	Followed JSON delivery spec
	Delivered?	Real-time		
TT-A	Yes	Yes	Yes	Yes
TT-B	Yes	Yes	Yes	Yes
TT-C	Yes	Yes	Yes	Yes
TT-D	Yes	No	Yes	Yes
TT-E	No	-	-	-

Evaluation Methodology

The travel time validation followed the basic methodology outlined in Section 3 the Data Validation Program document[3], focusing on the Traditional Analysis, which is most suitable for freeway road types. At a high level, this approach involves aggregating the reference and vendor data to the 5-minute level and computing two error measures, Average Absolute Speed Error (AASE) and Speed Error Bias (SEB).

The primary goal of this study was for vendors to become familiar with the new data formats and submission method, so the data quality evaluation was a much lower priority. Nonetheless, the validation team generated the Traditional Analysis error measures and corresponding time series plots to visually sanity check the extent to which vendor data agreed with the reference data.

Results and Discussion

The primary intent of the travel time validation was to exercise the georeferencing protocol and test the real-time data delivery procedure. As such, the high degree of compliance (see Table 3 in previous section) showed the process to be successful. The data quality evaluation proved to be less relevant for this particular data collection. Of the roughly 20k observations, 99.9% were close to free-flow speed. As a result, the calculated AASE/SEB error metrics for speed ranges other than free flow were not statistically significant. Visually inspecting the travel time and speed data for each path provided more useful insight.

Despite limited opportunities to observe congestion, the vendors that provided captured the general pattern of congestion in the few instances where present, indicating that the new delivery formats and georeferencing protocol did not impact data quality. Furthermore, as was the case in VPPII, when multiple modes were present in the reference data, vendor data tended to track the faster mode ("optimistic bias"). Figure 2 shows an example slowdown with vendor data (red dots) reporting more optimistic, higher speed / lower travel time values during the event.

[3] The Eastern Transportation Coalition Transportation Data Marketplace Data Validation Program Overview (link)

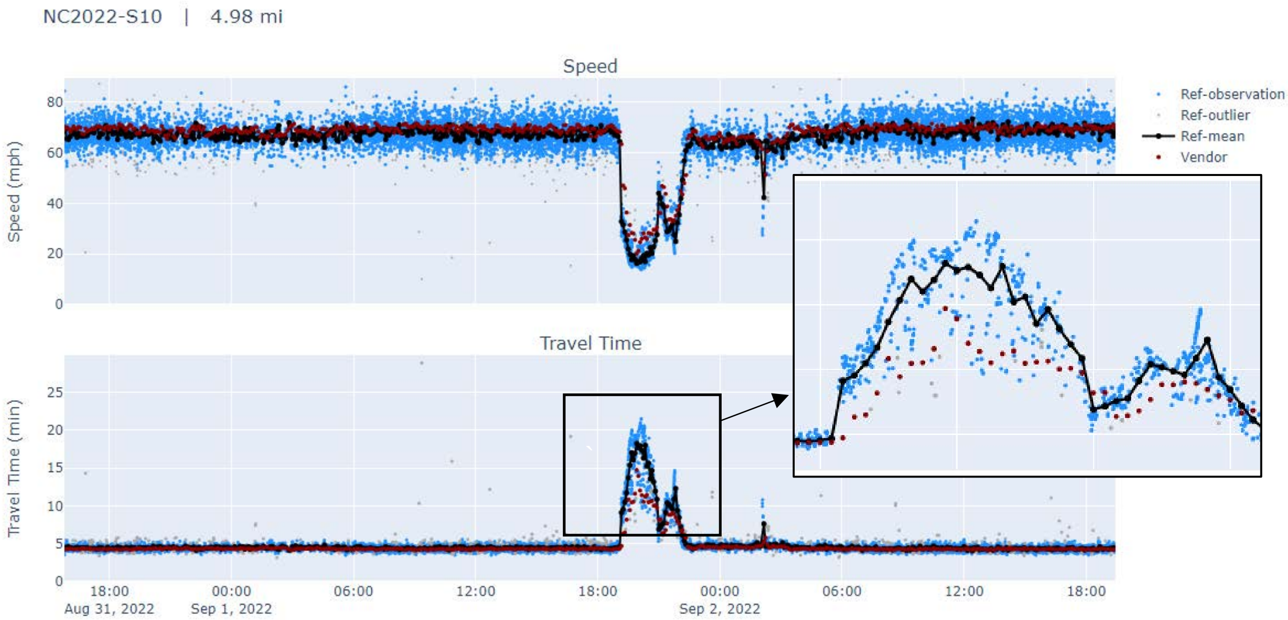

Figure 2 – Example slowdown event correctly identified in vendor data feed (with optimistic bias)

One travel time/speed vendor uncovered a data issue through the evaluation process, but it appears to be related to their delivery/submission pipeline, not underlying probe data quality. This will be monitored in future activities where vendor data quality is assessed and reported.

Key Findings

Although this travel time validation activity did not provide an opportunity for significant validation analysis (due to the lack of congestion on the chosen segments)**, data vendors adjusted well to the new protocols and should be ready for future activities.** Furthermore, **use of the georeferencing protocol greatly improved the efficiency of the validation**, removing the resource burden of conflating data sets from multiple vendors and streamlining the entire process. Key points include the following:

- **New georeferencing / data delivery requirements were successfully implemented**.
  - The CattWorks Geo-referencing Prototcol[4] (CWGP) was used to describe custom reporting segments, and all four vendors who submitted were able to respond using that prototcol.
  - Vendor data was delivered in real-time to the validation server from three of the four vendors who submitted data, allowing the validation team to directly access latency.

- **Minimal speed fluctuations were observed during the study period.** 99.9% of 5-min time periods recorded in the reference data were at or near free-flow speed. As a result, the typical error metric calculations (AASE/SEB) were not statistically significant to reflect the ability to capture slowdowns. The handful of periods with congestion were assessed visually

[4] CATT Works Georeferencing Protocol (link)

by comparing vendor provided data overlaid on the reference data. The performance was within the expected bounds for this small sample.

- **New processes did not appear to impact data quality**
    - Vendor-provided travel time and speed data generally tracked the Bluetooth collected travel time and speed data for the handful of speed fluctuations in the data set.
    - Vendors tended to follow the faster mode when multiple travel time modes were present (consistent with observations from VPPII, sometimes referred to as an optimistic bias).

Volume

Geographic Scope

Figure 3 shows the geographic scope of the volume study area, which encompasses roadway locations across the entire state. These evaluation locations were chosen based on where reference count data could be acquired, which included a combination of NCDOT permanent count stations, a sample of NCDOT short term count sites that coincided with the study period, and 6 short term counters deployed by a TETC contractor.

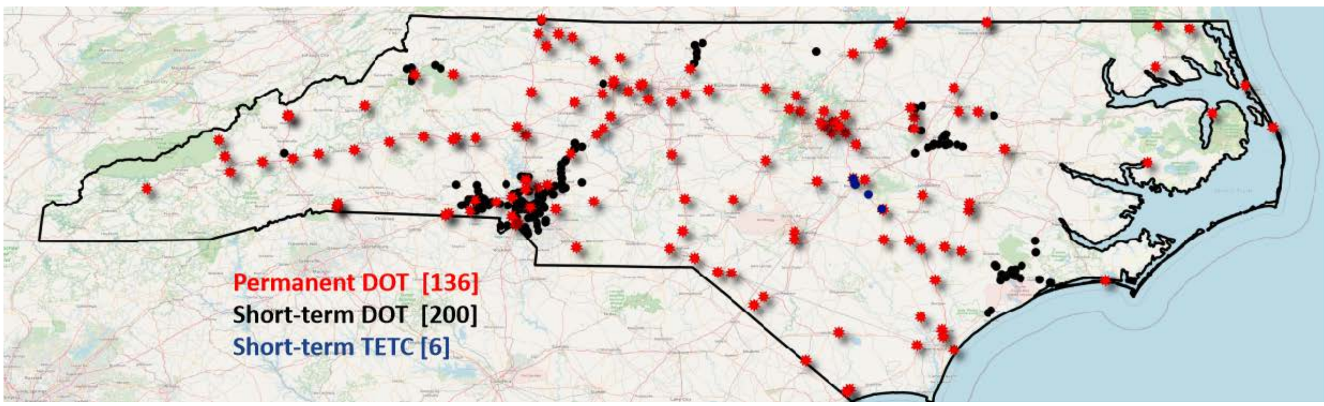

Figure 3 – Count location sites

Table 3 shows the key attributes used to communicate the volume evaluation locations to vendors via CWGP. Each Point location reference is defined by the latitude/longitude of the center-line of the road and its heading **for a specific direction of travel.** This means that a permanent count station that reports traffic counts in two directions would correspond to two CWGP point location references: one for each direction of travel. On divided highways, such as freeways, two distinct location references would represent the centerline of each set of lanes, along with two heading estimates approximately 180 degrees apart, one for each direction of travel.

Table 3 - CWGP attributes for volume validation points

ID	Location	Heading	Road Name	Road Class
NC2022-P0033	-81.4299 35.17107	259	I-85	1
NC2022-P0034	-78.6315 34.34235	111	US 74	2
NC2022-P0035	-78.63142 34.34266	290	US 74	2
NC2022-P0036	-78.86585 34.14242	46	US 701 BYP	4

Data Collection

Reference Data

Reference traffic volumes were acquired by NCDOT and a Coalition contractor during the study period of August 15 - September 15, 2022. NCDOT continuous counters captured hourly counts during the entire study period, while short term counters from NCDOT (NCDOT-STC) and the Coalition Contractor captured shorter periods (often 48 hours counts). Additionally, NCDOT provided 13 months of historical count data at the Continuous Count Stations (CCS), which was shared with vendors for model calibration purposes. It should be noted that during the study period NCDOT shut off public access to the continuous count data, which is usually published online, thus enabling blind evaluation of results.

Figure 4 shows the distribution of locations and hourly counts by reference data source during the study period. While there are a sizeable number of both CCS and Short-term count locations, most hourly records come from continuous counters, since they collected data during the full 1-month period rather than just 48 hours. Likewise, Figure 5 shows the distribution of locations and hourly counts by functional class. Count locations are present on roads across all functional classes, with highest representation on FRC 1-3 (interstates, other freeways and expressways, and other principal arterials). However, the majority of hourly observations are on FRC 1 roads (interstates) because that is where most continuous counters are located.

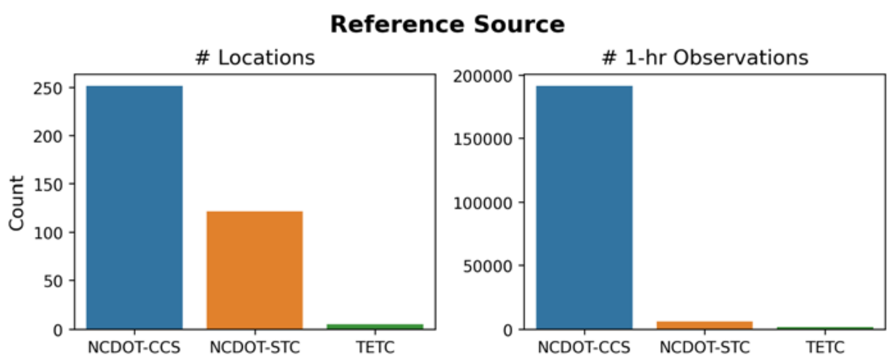

Figure 4 – Summary of locations and hourly counts by reference source

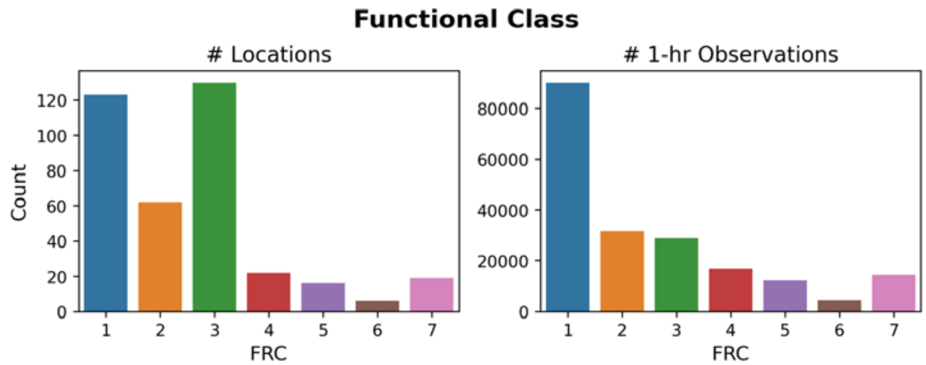

Figure 5 – Summary of locations and hourly counts by functional class

Vendor Data

Each vendor was instructed to prepare hourly count estimates (volume estimates for each date and hour) for the August 15 - September 15, 2022 study period and submit to the validation server by October 1, 2022. This date was chosen such that vendors would submit their volume estimates before NCDOT submitted the counts to HPMS.

Two of the four Volume data vendors delivered data for this validation activity, both of whom delivered the required hourly count estimates (plus other optional data items) and adhered to the required protocols and data formats -- see Table 4. The remaining two vendors experienced issues that prevented them from successfully delivering volume estimates. These issues include unexpected data supply chain issues in one case, and a change in subcontractor in the other.

Table 4 – Summary of volume data delivery by anonymized vendor

Vendor	Delivered?	Used Geo-Referencing Protocol	Followed JSON delivery spec	Volume products delivered (hourly counts mandatory)
VOL-A	Yes	Yes	Yes	**Hourly counts**, AHDT, ADT
VOL-B	Yes	Yes	Yes	**Hourly counts**, AHDT, ADT
VOL-C	No	Data supply chain changes		
VOL-D	No	Unexpected change in subcontractor		

Evaluation Methodology

The evaluation methodology used in this study is based on the conclusions from the Volume Validation technical memo (TAC-VAL-1), which recommended computing a variety of metrics used frequently in existing literature and generating various plots. The validation team started by looking at overlay plots of vendors data with the reference data to gain insight of vendor performance, and then computed several different error metrics to assess the degree of correlation of how well vendor provided data agreed with referenced data as revealed by the "eye test".

The following error metrics were used in this study, where V_R is the reference volume and V_V is vendor volume:

- $\textbf{Error (E)} = V_V - V_R$
- $\textbf{Percent Error (PE)} = \frac{V_V - V_R}{V_R} \cdot 100$
- $\textbf{Absolute Percent Error (APE)} = \frac{|V_V - V_R|}{V_R} \cdot 100$
- $\textbf{Error to Max Flow Ratio (EMFR)} = \frac{|V_V - V_R|}{\max(V_R)} \cdot 100$

Previous research has suggested that overall roadway volume plays a large role in the interpretation of results. For example, percentage-based errors (e.g., PE or APE) may be large in low-volume ranges despite showing relatively close agreement from visual inspection and should be interpreted differently than the same percent error at much higher volume levels. As such, multiple error metrics should be reported together, and broken out by volume level, road class, peak versus off-peak, and other scenarios.

Results and Discussion

The two available volume data submissions were evaluated using the previously described approach, first performing a visual comparison of each vendor's volume counts alongside corresponding reference values and then computing a variety of error metrics. The specific results were shared one-on-one with each vendor, but only a handful of anonymized findings are reported here because this inaugural activity was structured as a learning exercise for the vendors and as an opportunity for the validation team to evaluate the usefulness of different error metrics.

Sample anonymized results are shown below for illustrative purposes -- mainly to show the types of plots and metrics used and to communicate at a very high level how well vendor estimates matched reference data. It should be noted that while the specific patterns were different between vendors, both participating vendors produced results that were comparable in accuracy.

CAUTION: The visuals and error metrics presented below are based on an initial delivery from a single vendor and are not intended to be representative of the viability of probe-based volume estimation methods.

Exploratory Visualizations

Figure 6 represents the starting point for visualizing results: a scatter (hex-bin) plot showing reference values and vendor hourly count estimates plotted for each hourly period and location. The 45-degree line (red dashed line) represents perfect estimates; points that are close to it represent instances where the vendor data matches the reference values well; points falling above or below represent underestimates and overestimates by the vendor, respectively. The color or shade of the hexagon dots communicates the density of observations in each region, with darker colors representing more data points at a given location. The main conclusions from this plot are that the vendor data shows general agreement with reference data and that there are many low volume observations (lots of dark-colored points at very low volume).

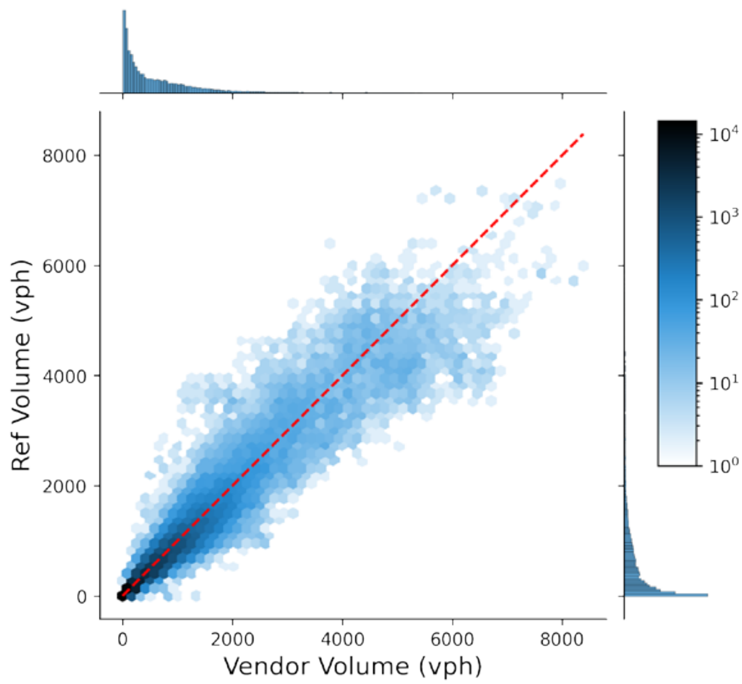

Figure 6 – Scatter plot showing reference versus vendor volume counts

While this provides a useful overall picture of estimation accuracy, it is comprised of many different locations and time periods. As such, it is also useful to focus on individual segments and observe how the vendor data tracks the reference values over time. Figure 7 is one such example, which provides several key insights.

- Vendor data tracks the reference data closely during low-volume periods and has similar maximum flow values, but shows deviations during the daytime. Note that this visual observation may not always be apparent from the error metrics – especially percent error-based measures, which can be high during low-volume periods (when there is strong agreement) due to numerical issues.
- The key discrepancies occur during peaks and midday periods, with vendor data not clearly differentiating midday volumes from the PM peak.
- Error is correlated in time.   For example, if an estimate is low, it remains low throughout the peak period. Additionally, if the peak period is low one day, it is likely low the next day too.

NC2022-P0020

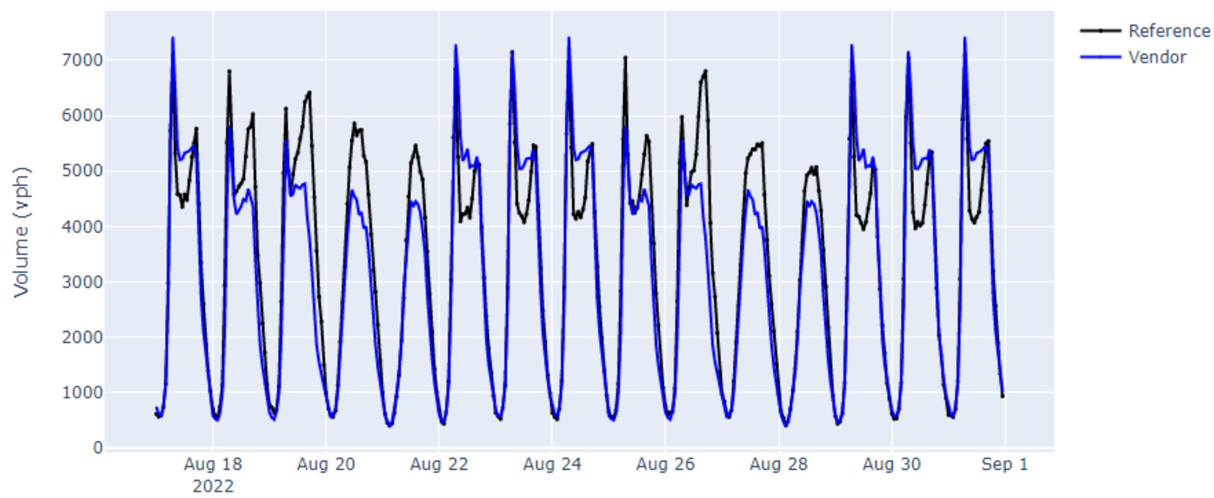

Figure 7 – Time series plot showing reference vendor volume counts for a single segment.

Error Metrics

The four error metrics -- Error (E), Percent Error (PE), Absolute Percent Error (APE), and Error to Max Flow Ratio (EMFR) were computed for all hourly periods across all locations in the dataset and then summarized in different ways. Figure 8 summarizes the distribution of the four error metrics across all hourly periods for one of the vendors. For each boxplot the blue box represent the middle 50% of distribution, the line in the middle of the box is the median value, the red X is the mean, and whiskers extend +/- 1.5*IQR (Inter Quartile Range) from the edges of the box. Note that APE and EMFR by definition do not go negative, thus the lower whisker is capped at zero. The data underlying the plots (i.e., all hourly periods in the dataset) have several high-value outliers which fall outside the whisker extents – and beyond the scale of the graphs in the figure. A few observations can be made from these boxplots:

- *Error is slightly positively biased.* In the E and PE plots (i.e., the metrics that consider the sign of the error) the majority of the distribution is above zero. Note, however, that this does not tell us anything about *where* the vendor data overestimates volumes.
- *Accuracy is roughly 25% APE across all volume levels.* The median APE is right around 25% and the mean is a bit higher, although likely impacted by several high-value outliers.
- *EMFR is quite stable -- around 5-9% error relative to the max observed flow.*
- *Distributions can have long tails*. In several plots -- especially PE and APE, the mean value (red X) is quite a bit larger than the median value because large values can pull up average. More context is needed for interpreting large percentage values, as high percent errors when volumes are extremely low are more of a numerical stability issue than a matter of inaccuracy.

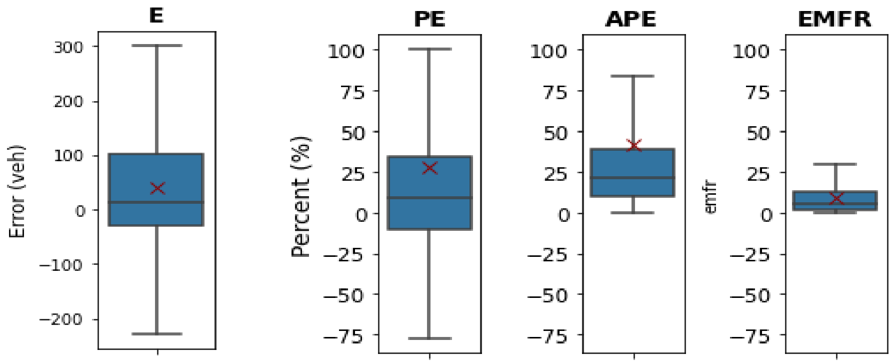

Figure 8 – Box plots summarizing distribution of error metrics across all records

Figure 9 summarizes the mean values of all metrics (I.e., mean error (ME), mean percent error (MPE), mean absolute percent error (MAPE), and mean EMFR) across different volume bins. One point of clarification is that the error metrics for EMFR are not recalculated by volume bin, rather the original EMFR as displayed in Figure 8 were grouped by volume bins. Note that the lowest volume bin has high MPE and MAPE of 87% and 100%, respectively, despite a low ME of less than 14 vehicles per hour -- reflecting the numerical stability issues when division is performed with small values in the denominator. The high MPE and MAPE in the 0 – 100 volume bin are not reflective of the quality of the data, but rather an artifact of the metric calculation.

Volume Bin	# Records	# Locations	ME (veh)	MPE (%)	MAPE (%)	Mean EMFR (%)
0-100	21366	251	13.6	86.5	100.1	8.4
100-250	14629	280	21.9	13.6	27.9	6.5
250-500	13795	278	45.0	12.0	26.9	8.5
500-1000	17715	255	77.1	10.5	23.9	11.1
1000-1500	9001	192	87.1	7.4	20.9	11.4
1500-2000	4276	140	4.4	0.4	18.4	10.8
2000-2500	2472	95	-29.3	-1.4	18.5	11.2
2500-3000	1767	72	-10.7	-0.4	19.3	11.9
3000-3500	1407	54	-1.2	-0.1	18.9	12.3
3500-4000	1410	38	80.9	2.1	23.6	16.8
>4000	2905	33	138.4	3.4	17.2	13.4

Figure 9 – Mean error metrics separated by volume bin

Figure 10 contains the same data with the lowest volume bin omitted, which better reveals the trends across volume bins. ME and MPE (which consider the sign of the error) show slight positive bias at lower volumes, almost no bias at volumes from 1500-3500 vph, and very minor positive bias at higher volumes above 3500 vph. MAPE and EMFR (which only consider magnitude of error, not sign) generally move in opposite directions volume is increased; MAPE tends to decrease (with the exception of 3500-4000 vph), whereas EMFR increases slightly. Note that MAPE calculates errors relative to reference volume, so if the ME remains relatively constant, higher volume bins will naturally decrease the APE and MPE statistics. In contrast, EMFR, which calculates error relative to the max observed flow on the segment, exhibits less variability across volume bins. EMFR shows more agreement with the "eye test" (such as Figure 6). When observing time series plots, most accuracy issues seemed to be concentrated during daytime and peak periods when volume was closer to the maximum volume, and exhibited strong agreement during off-peak low-volume periods. **Further work is needed to contextualize these error metrics and specify target ranges that indicate suitability for different applications.**

Volume Bin	# Records	# Locations	ME (veh)	MPE (%)	MAPE (%)	Mean EMFR (%)
100-250	14629	280	21.9	13.6	27.9	6.5
250-500	13795	278	45.0	12.0	26.9	8.5
500-1000	17715	255	77.1	10.5	23.9	11.1
1000-1500	9001	192	87.1	7.4	20.9	11.4
1500-2000	4276	140	4.4	0.4	18.4	10.8
2000-2500	2472	95	-29.3	-1.4	18.5	11.2
2500-3000	1767	72	-10.7	-0.4	19.3	11.9
3000-3500	1407	54	-1.2	-0.1	18.9	12.3
3500-4000	1410	38	80.9	2.1	23.6	16.8
>4000	2905	33	138.4	3.4	17.2	13.4

Figure 10 – Mean error metrics aggregated by volume bin (lowest bin omitted).

Key Findings

The validation team reviewed the results with each volume vendor, including in-depth review of the validation metrics with the two vendors that successfully submitted data. Key takeaways from analyzing the data and discussing the result with the vendors and the TAC include the following points:

- **Further work is needed to define appropriate error metrics and provide context for their interpretation.** Some of the error metrics used in this analysis do not match the visual observations gained from inspecting time series plots. Percentage-based measures are highly sensitive to traffic volume level, causing instability at low volume levels. Additional guidance is needed for setting appropriate targets (what is good enough), recognizing that the data may be useful at different levels of accuracy depending on the intended application.

- **Most vendors are currently concentrating their product development efforts on higher time aggregations such as AADT, ADT and AHDT**. This validation activity, which focused on hourly counts (the most granular of required data items required in the RFP), is most appropriate for operations performance measures. Although a supported product by the vendors, volume products are more mature for the planning level measures. Future validations will incorporate more temporally aggregate volume products such as AADT, ADT, and AHDT.

- **Unlike travel time, volume estimation requires engineering resources for each exercise.** Vendors requested more advanced warning (3+ months) to plan resources to respond to the validation activity. As a result, the validation team will strive to give vendors adequate time to efficiently prepare for future validations, and request data submittal consistent with expectations set within the TDM contractual obligations.

- **Volume estimates, unlike travel time and speed, are highly sensitive to perturbations in the data supply chain.** Delays may occur due to disruptions of base data supply requiring re-engineering of the process.

- **Georeferencing Protocol worked as designed for providing for efficiency of the volume validation process similar to travel time and speed validation.** Both vendors who participated were able to locate each volume validation location using the CWGP and submit volume estimates in this protocol. The CWGP will continue to be used in future activities.

- **Need to begin developing cross-validation procedures.** It is expected that self-reported cross-validation metrics will complement blind evaluation studies to characterize the accuracy of probe-based volume estimation models. The validation team plans to work with the TAC, industry partners, and vendors to develop a “cross-validation audit” to integrate into the evaluation framework.