# TDM Validation Activity: Georgia Volume Study

## TETC Transportation Data Marketplace Data Validation

Prepared by: TETC Validation Team

5/6/2025

# TDM Validation Activity: Georgia Volume Study

# TETC Transportation Data Marketplace Data Validation

*Validation activity for TDM Validation program, focusing on Volume data in Georgia.*

**TDM-VAL-05**

***Data Categories****: Volume*

***Evaluation Period****: April 2024 (and 2023 AADT)*

***Publication Date:*** *May 2025*

5/6/2025

The Eastern Transportation Coalition is a partnership of 19 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 documents a TDM data evaluation effort conducted by The Eastern Transportation Coalition (TETC) data validation team which quantifies the accuracy of Volume data provided by five TDM vendors: HERE, INRIX, Iteris, Streetlight, and Replica. Building off the prior validation activity [1] in North Carolina, the first comprehensive multi-vendor volume study for which accuracy measures were publicly reported, this follow-on activity reassesses accuracy in a new location. The time period for this evaluation was April 2024 (and all of 2023 for AADT estimates), noting that at this time the industry was recovering from broad-based data supply chain disruptions.

The error metrics and visuals used to evaluate volume data generally follow the approach taken in the prior North Carolina activity. This report considers five volume products: *annual average daily traffic (AADT), average daily traffic (ADT), average hourly daily traffic (AHDT), daily volumes, and hourly volumes.* The first three – AADT, ADT, and AHDT, are aggregate planning-level volume products, while the final two -- daily and hourly volumes, focus on specific dates and times, and are oriented towards operational applications. A key focus of the analysis was incorporating USDOT's updated statistical method for assessing the accuracy of non-traditional volume sources for AADT, which serves as an important benchmark for states seeking to use probe-based volume products for federal reporting purposes.

The study area includes 963 directional validation locations across the state of Georgia, primarily corresponding to the locations of (i) continuous count stations and (ii) short-term count sites for which counts were collected during April 2024. Vendors were asked to deliver five different types of volume data at these locations for specific date ranges; four (ADT, AHDT, daily volumes, hourly volumes) focused on April 2024, while the fifth (AADT) targeted all of 2023. Each of the five volume vendors actively participated in the validation activity, with three delivering all five data items requested, and two submitting only AADT. Several key takeaways emerged from the resulting evaluation and follow-up discussions with vendors:

- **Industry has progressed over the past year.** Despite multiple vendors weathering data supply chain issues, this study showed that probe-based volume estimation products have matured in the past year. Several vendors are on the verge of meeting FHWA's accuracy/precision benchmarks for AADT, a key use case for the data. Also, there has been growth in the quality and availability of more granular data elements. Specific improvements in vendor product offerings include the following:
    - All vendors are now capable of reporting volumes on all road classes (FRC 1-7). In the prior validation, Vendor C reported only FRC1-5, but has now added coverage for FRC 6-7 (typically low-volume and often rural facilities).
    - All three vendors who submitted AHDT data are now able to regularly capture average day-of-week and time-of-day volume patterns. In the prior validation, Vendor A struggled with differentiating between weekend versus weekday patterns for AHDT, but their product has noticeably improved in this regard.
    - Two vendors (Vendors C and D) are now capable of producing meaningful daily and hourly volume estimates that are sensitive to atypical volumes. In the prior validation, only Vendor C had the ability to do so, but over the past year Vendor D introduced a meaningful hourly/daily product.

[1] TDM-VAL-05: North Carolina Volume Study (link)

- **There is measurable separation among the vendors**. Although details differ across the different temporal aggregations, several vendors stand out as being particularly strong in specific areas. In the previous studies it was concluded that the differences between vendors were not substantial. At the planning level, three vendors are getting close to FHWA benchmarks for precision/accuracy (A,C,D), with vendor D most consistently close to the benchmark regardless of whether volumes are treated directionally or bi-directionally. These vendors also demonstrate the ability to capture average day-of-week and time-of-day patterns (AHDT). From an operational perspective, vendors C and D stand out has having meaningful Daily and Hourly volume products that can detect abnormal volume patterns.
- **Several vendors are on the verge of meeting FHWA guidelines for AADT accuracy/precision, although the results differ slightly based on whether tests are run on directional or bi-directional volumes**. The validation team intends to make the FHWA test a core part of future analyses and plans to seek further clarification on how to handle volume directionality.
- **There are multiple sources of error and uncertainty that impact the accuracy of results.** In addition to estimation error as a byproduct of each vendors' modeling processes, conflation/geo-referencing issues can lead to erroneous volumes when validation locations are translated incorrectly between map representations and/or geo-referencing protocols. These georeferencing processes also require properly interpreting roadway directionality, and if handled improperly, can lead to significantly over or underestimating volumes. Additionally, the reference data from both continuous and short-term counters require careful inspection to deal with abnormalities and missing data that can impact annual average calculations. As vendor product offerings continue to mature, it will be particularly important to monitor and quantify theses different sources of error.

Overall, it is encouraging to see industry improvement across several vendors, despite the study period aligning with disruptions to several vendors' data supply chains. At least one vendor was still in the midst of resolving major disruption to its source data (Vendor B), while two others were just emerging from integrating new data sources (Vendors C and D). As of early 2025, it appears that all vendors have recovered from this disruption, and full participation is expected in the next validation scheduled for 2025. Results from this study will serve as important point of comparison to assess product accuracy trends over time.

# 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. During the first two phases of the Vehicle Probe Project which ran from 2008 to 2022, validation activities were limited to ensuring that traffic data conform to the contractual standards of travel time and speed with highly defined specification using prescribed methods and metrics. The TDM, which became active in 2022, continues to validate travel time and speed using that protocol, but it also validates various other data sets.

The methodology and format of the validation is flexible based on current industry practice, and adjusts to the needs of the Coalition members as the market evolves and data needs expand, under the direction of the TDM Technical Advisory Committee (TAC). The validation process is overseen by the TAC that sets general direction and reviews results. The TDM validation program 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. The TDM also contains various ancillary products to complement and extend the core data products. As such, the validation team, under the guidance of the TETC Validation TAC, is beginning to establish standards and methods for effectively evaluating data quality and value for these various data sets.

This validation report evaluates one of the core data products available in the TDM: Volume data, which was identified by the TAC as the highest priority for validation. This report is the third Volume validation activity, and the second for which error measures are reported publicly. The prior reports provide useful context on how the volume validation process was developed and the most recent results: TDM-VAL-01 (literature review and recommendations for structuring volume validation), TDM-VAL-02 (inaugural proof of concept activity), and TDM-VAL-05 (North Carolina study). The intent of this report is to provide insight into vendor accuracy for the different Volume products sold in TDM and show how performance has changed over the course of the past year.

# Data Vendors

All vendors selected through the TDM RFP process in the Volume category were invited to participate in the validation study (HERE, INRIX, Iteris, Streetlight), as was another vendor that serves as a subcontractor with Coalition customers in the TDM (Replica). The validation team set expectations for data submission consistent with RFP specifications, including using a specific georeferencing protocol (CWGP) to describe validation locations, and requested five Volume data deliverables: Annual Average Daily Traffic (AADT), Average Daily Traffic (ADT), Average Hourly Daily Traffic (AHDT), Daily Volume, and Hourly Volume. The Data Collection section below contains more information about each vendor's submission. Vendors are not explicitly identified in this report; vendor-specific results were shared verbally with the Coalition, but vendors are anonymously labeled A, B, C, D, and E in this report.

## Geographic Scope

Figure 1 shows the geographic scope of the volume study area, which encompasses 926 directional roadway locations across the entire state of Georgia. Most of these locations correspond to GDOT count sites, including all continuous count sites, which collect traffic counts 24/7 for each day of the year, and a subset of short-term counters that were active during April 2024, typically collecting a 48-hour count. A handful of additional locations were used to sanity-check the internal consistency of vendor estimates (e.g., checking for spatial consistency as will be defined later). These evaluation locations capture all functional road classes, ranging from Interstate roadways (class 1) through local roads (class 7).

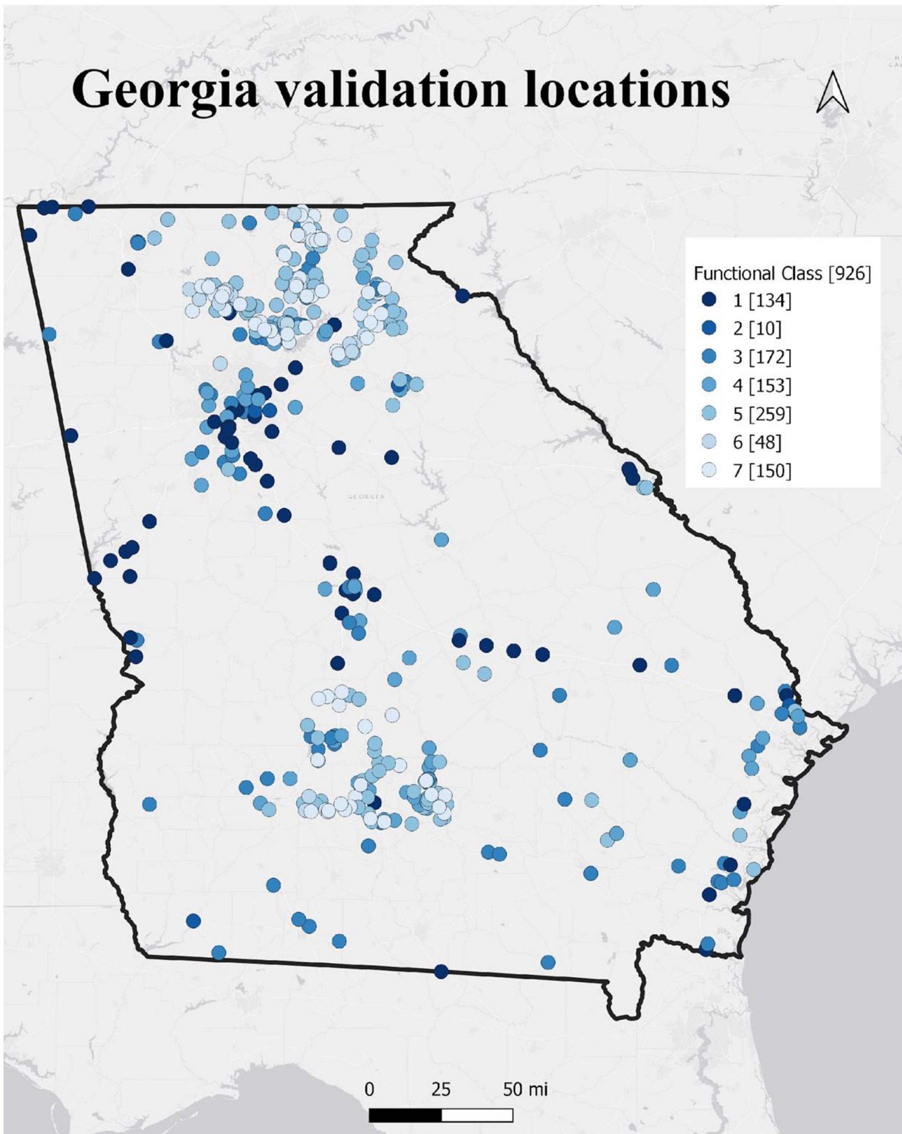


**Figure 1 - Evaluation locations**

Table 1 shows the key attributes used to communicate volume evaluation locations to vendors via CWGP. Each point location reference is defined by the latitude/longitude of the centerline of the road and its heading **for a specific direction of travel.** This means that a 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. Note that heading is measured as degrees clockwise from North.

**Table 1 - CWGP attributes for volume validation points**

| ID | Location | Heading | Road Name | Road Class |
|---|---|---|---|---|
| **P001** | -83.758607<br>30.799667 | 101 | US-84/SR38 W of Hallman Rd Thomas Co. | 3 |
| **P002** | -83.758573<br>30.799847 | 281 | US-84/SR38 W of Hallman Rd Thomas Co. | 3 |

# Data Collection

## Reference Data

Hourly traffic counts from validation locations – consisting of both continuous and short-term count sites – were obtained from Georgia DOT for April 2024, the primary study period. These volumes were *not* available to vendors, as GDOT shut down public access to their public count site until vendors submitted their volume estimates. Additionally, GDOT provided 15 months of historical count data (Jan 2023 – Mar 2024) at the continuous count sites, which were shared with vendors for model calibration purposes. This historical information was also used for assessing AADT from 2023.

The GDOT count data was used to create five different reference datasets, consistent with the mandatory data items from the TDM RFP plus the addition of daily volumes. The five data elements used in the analysis are detailed below and summarized in Table 2, ordered by increasing temporal granularity. With the exception of AADT, each reference volume dataset focuses on April 2024, with both continuous count stations (CCS) and STC locations used for Daily and Hourly products, and only CCS locations used for the measures that require aggregating over a full month (AHDT and ADT). The AADT dataset is based on 2023 data at continuous count sites.

**Table 2 - Reference datasets used as the basis for evaluation.**

Most aggregate ↕ Most granular

| Data Item | Description | Evaluation period | Continuous Counts | Short-term counts |
|---|---|---|---|---|
| Annual Avg Daily Traffic (AADT) | Avg volume for a day during period | 2023 | ✓ | |
| Avg Daily Traffic (ADT) | Avg volume for a day during period | April 2024 | ✓ | |
| Avg Hourly Daily Traffic (AHDT) | Avg hourly volume by day-of-week (e.g., Mondays between 8-9am) | April 2024 | ✓ | |
| Daily Volume | Volume for a specific day during period | April 2024 | ✓ | ✓ |
| Hourly Volume | Volume for a specific hourly period (e.g., 04/01/24 between 8-9am) | April 2024 | ✓ | ✓ |

The reference counts were subject to GDOT's QC/QA process and further investigated by the validation team. Apart from a single GDOT count site (2 directional validation locations) that appeared to be only counting a subset of traffic lanes, all locations were considered valid and used for Hourly and Daily volume analysis. However, when creating reference datasets for more aggregate measures (AHDT, ADT, AADT), locations were omitted when there were less than 75% of hourly records available for the study period. As an example, Figure 2 below shows the distribution of missing hourly records at directional validation locations in 2023 (used for AADT), with locations left of the line kept for analysis, and those to the right omitted.

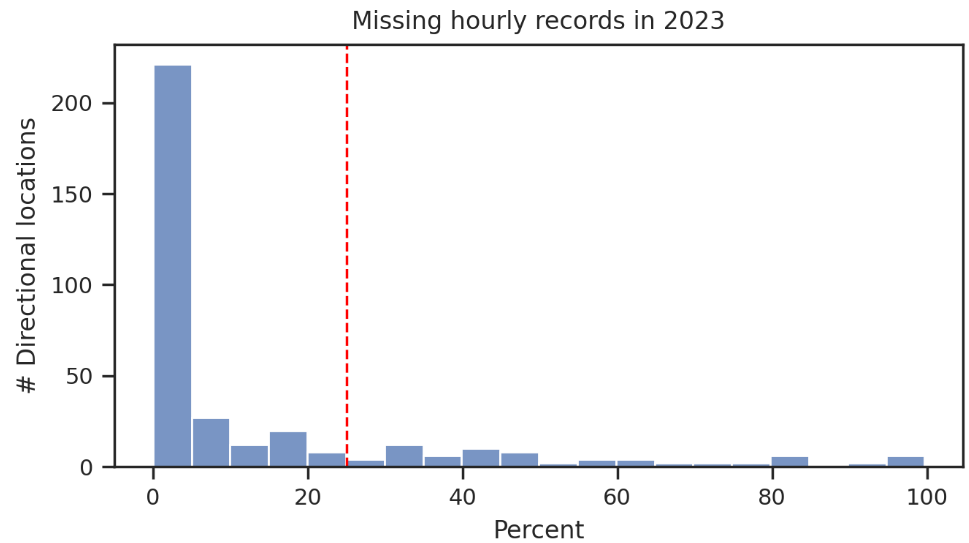


**Figure 2 – Applying missing data thresholds**

This rule was imposed to prevent locations with a small amount of data from being used as representative 'ground truth' averages over a month or year. Future work will explore suitable thresholds, but a conservative threshold of 75% was used here to ensure reference values were suitably reliable.

Hourly volumes from a subset of validation locations (corresponding to GDOT's continuous count site) were provided to vendors for calibration purposes. Historical count records were provided from Jan 2022 through Mar 2023, representing the 15 months immediately preceding the study period of April 2024.

## Annual Average Daily Traffic (AADT)

AADT values were obtained by aggregating the hourly counts from 2023 using the FHWA recommended method at validation locations corresponding to continuous count sites. Only locations with sufficient data coverage over the course of the year (defined here as 75% of hourly records present) were used for analysis. The aggregation approach, detailed on page 3-14 of the 2022 Traffic Monitoring Guide (see footnote 2), first computes the Monthly Average Daily Traffic for each month and then takes a weighted average of all 12 months, taking into consideration the number of days in each month. This method produces identical results to a simple arithmetic average of total counts for each day in the year when there is no missing data. Note that because hourly counts from 2023 were provided to vendors as calibration/training data, AADT evaluation is not a blind study. Nonetheless, the TAC indicated that this aggregation level was important to include as a validation exercise.

Figure 3 shows the distribution of AADT volumes across 286 directional locations. Half of AADT volume observations have values under 8k with about a quarter under 3k.

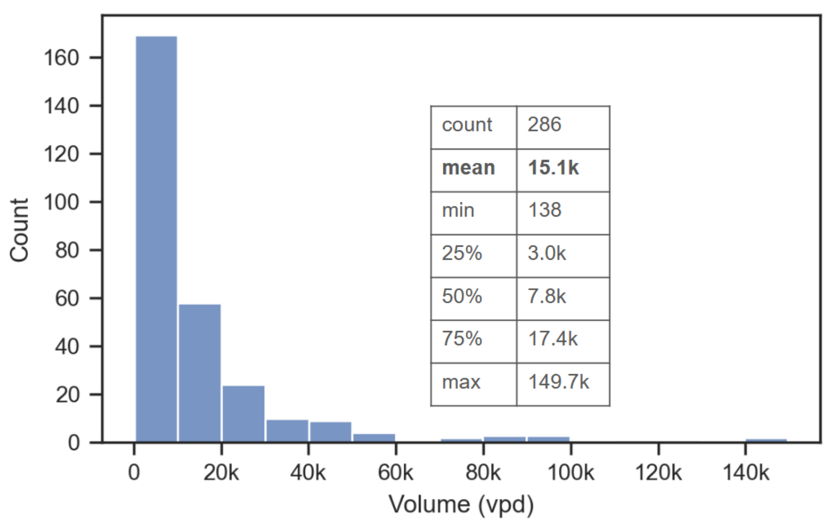


**Figure 3 – Distribution of AADT volumes.**

It should be emphasized that these AADT locations are *directional;* there is one AADT volume provided per direction – regardless of whether a roadway is divided or undivided. This is consistent with how all other data aggregations are analyzed for this validation activity; however, it differs from how FHWA represents AADT (described later), which is generally bi-directional, meaning both directions of travel are included in an AADT value. Thus, when analyzing AADT, the validation team will provide results for both directional and bi-directional representations. All other data aggregations will be based on directional analysis only.

## Average Daily Traffic (ADT)

ADT values were obtained by grouping the hourly CCS counts from April 2024 by location and computing the volume corresponding to an average day, following the FHWA-recommended method outlined on page 3-14 of the 2022 Traffic Monitoring Guide[2] (see MADT definition). This method produces identical results to a simple arithmetic average of daily counts when CCS stations records do not have missing data but is well-suited to handle small gaps that may be present. Figure 4 shows the distribution of values across 274 directional locations with at least 75% of hourly records present. Half of observed observations have AADT values under 8.1k with about a quarter under 3.2k.

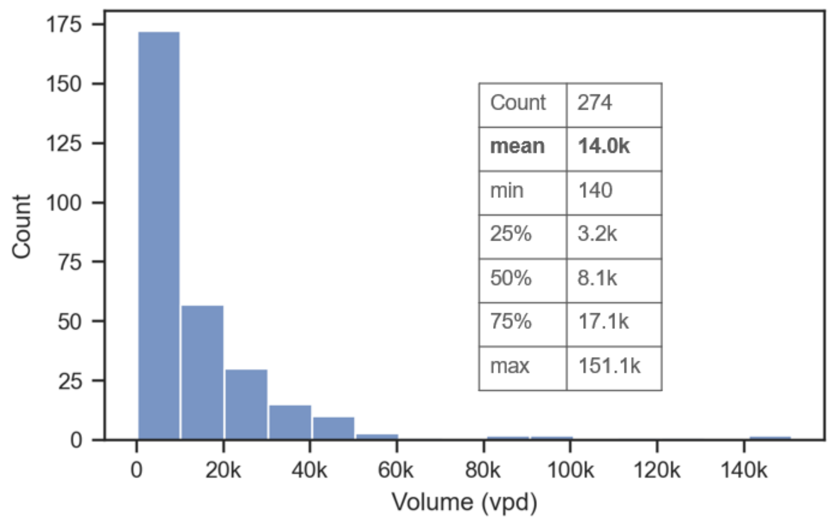


**Figure 4 – Distribution of ADT volumes.**

## Average Hourly Daily Traffic (AHDT)

AHDT values were obtained by grouping the hourly CCS counts from April 2024 by location, day of week, and hour of day, and then taking the arithmetic average of all counts in each group (e.g., average volume on Mondays between 8-9am). No holidays or anomalous days were excluded from the averages. Figure 5 shows the distribution of values across the same 274 locations and about 42k hourly observations. About half the hourly observations are under 250 vph, with a quarter under 65.

[2] 2022 FHWA Traffic Monitoring Guide (link)

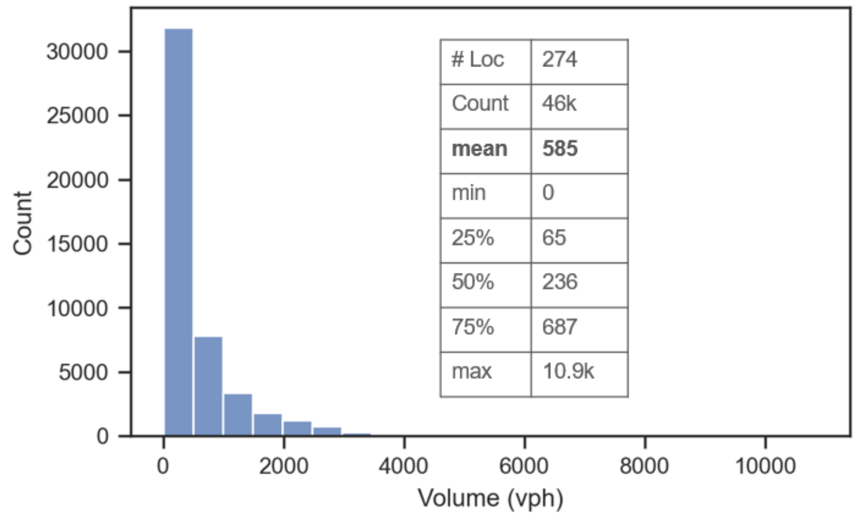


**Figure 5 – Distribution of ADHT volumes.**

## Daily Volumes

Daily volumes were obtained by aggregating hourly counts for each date in April 2024 at all validation locations (both CCS and STC). Only dates with 24 hourly records were included in the aggregation to prevent partially counted days from being interpreted as full daily volumes. Figure 6 shows the distribution of counts across 557 locations and about 8.6k total records, with Figure 7 breaking down the results by functional class. Figure 7 highlights that short term count sites are primarily concentrated on functional classes 3-7, and despite representing almost 50% of locations, make up only a small percentage of daily observations (since CCS locations have 30 daily observations over the course of April ‘24).

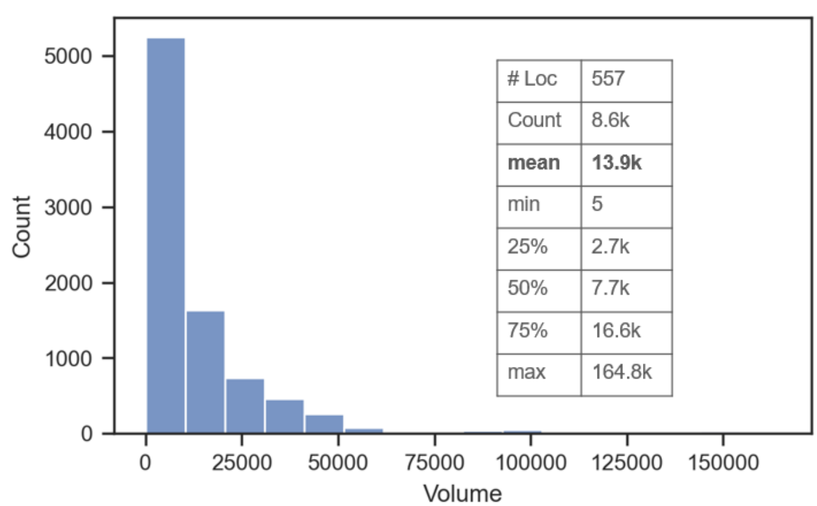


**Figure 6 – Distribution of Daily volumes.**

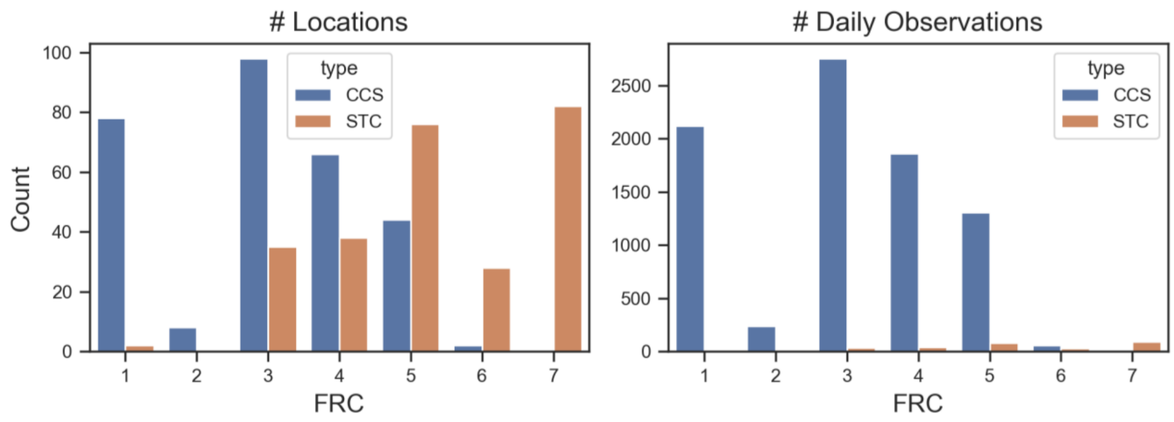


**Figure 7 – Distribution of Daily Volume locations and observations by functional class.**

## Hourly Volumes

Hourly volumes were provided directly by GDOT for each location during April 2024 and did not require further aggregation. Figure 8 shows the distribution of hourly volumes across 561 locations and about 213k observations, which highlights the fact that there were many low-volume periods (half of all hourly observations have just over 200 vehicles per hour, with about a quarter less than 50 per hour). Figure 9 indicates that short term count sites are primarily concentrated on functional classes 3-7, and despite representing almost 50% of locations, make up only a small percentage of hourly observations (since CCS locations have 30*24=720 hourly observations over the course of April '24). The data in Figures 8 and 9 represents directional hourly volumes.

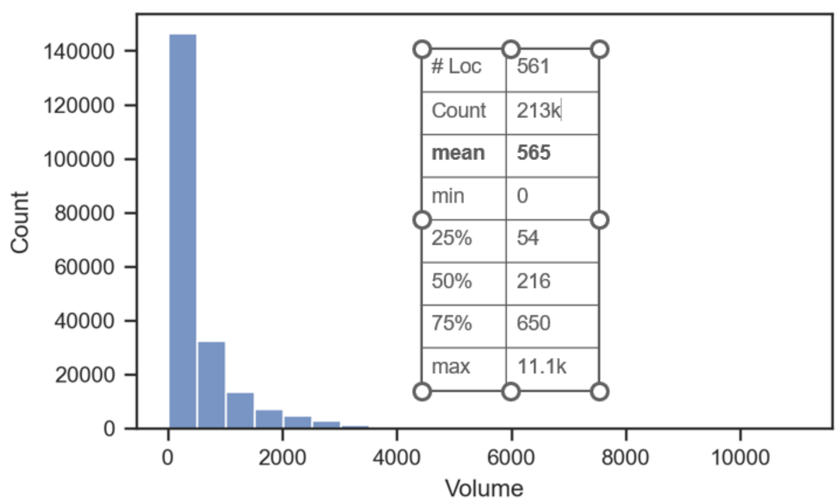


**Figure 8 – Distribution of Hourly volumes.**

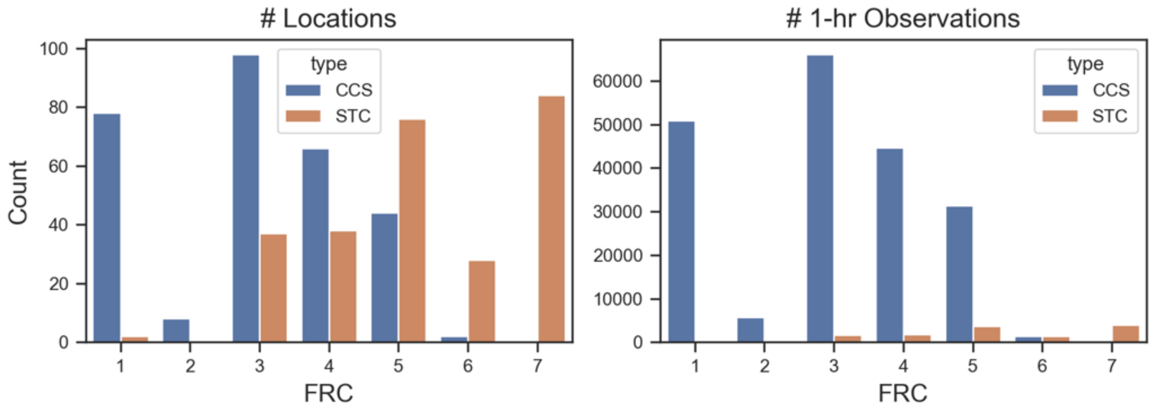


**Figure 9 –Distribution of Hourly Volume locations and observations by functional class.**

## Vendor Data

Each vendor was instructed to prepare five volume data deliverables corresponding to the data items listed in Table 2: AADT, ADT, AHDT, Daily, and Hourly volumes. All five vendors actively engaged with the validation activity and delivered data within the specified timeframe and according to the specified data exchange format. Table 3 summarizes the vendor submissions for each volume product; three of the vendors (A,C,D) delivered all requested data items, while the remaining two (B,E) delivered just the most aggregate product: AADT volumes.

Several vendors indicated that their products were impacted by disruptions to underlying source data supply chains. Impacted vendors were at different points in the process of re-establishing data supply chains at the time of this activity, and this impacted at least one vendor's ability to submit data across all aggregations.

**Table 3 – Vendor Submission Summary**

| Vendor | AADT | ADT | AHDT | Daily Volumes | Hourly Volumes |
|---|---|---|---|---|---|
| A | ✔ | ✔ | ✔ | ✔ | ✔ |
| B | ✔ | | | | |
| C | ✔ | ✔ | ✔ | ✔ | ✔ |
| D | ✔ | ✔ | ✔ | ✔ | ✔ |
| E | ✔ | | | | |

# Evaluation Methodology

Vendor data quality was evaluated separately for all five deliverables (AADT, ADT, AHDT, Daily, Hourly) using error metrics. For Daily and Hourly aggregations a visual inspection method was also used based on recommendations from the Volume Validation technical memo TDM-VAL-1 and prior volume validation report TDM-VAL-5.

The primary error metric used to quantify differences between vendor and reference volume values is percent error. Percent error (PE) is defined as

$$\mathrm{PE} = \frac{V_V - V_R}{V_R} \cdot 100,$$

where $V_R$ is the reference volume and $V_V$ is vendor volume. The mean and various percentiles of PE are calculated to communicate the central tendency and distribution (spread) of values across the dataset (or subset of the dataset), respectively. For example, the mean absolute percent error, referred to as MAPE, communicates the average absolute error expressed as a percentage of the reference volume, whereas a summary of absolute percent error (APE) percentiles (e.g., 25%, 50%, 75%) communicates how much the values vary (often expressed visually in a box-and-whisker plot). In most cases, it makes sense to group aggregate results by volume bin, road class, or another data characteristic of interest. Percentage-based errors are sensitive to low volumes, so separating results by reference volume range can provide useful context for interpreting results. Note, the methods above were incorporated initially for directional volumes and are now also for bi-directional volumes when needed.

## FHWA Method

Additionally, the validation team consulted the updated USDOT methodology[3] for evaluating the accuracy and precision of AADT estimates obtained from emerging technologies. This method, which was updated in 2024, establishes accuracy and precision benchmarks in three volume ranges (Low: 500-4999 vehicles per day (vpd), Medium: 5000-54999 vpd, High: 55000+ vpd). These benchmarks for acceptable precision and accuracy reflect the expected precision/accuracy of factoring 48-hour counts to AADT values (based on prior FHWA research) and were designed to ensure that probe-based volume estimates are *at least as accurate and precise as factoring*. The same error measures for each vendor's AADT product are compared with FHWA's benchmarks in each volume range to determine whether acceptance thresholds are met.

There are three statistical tests conducted for each volume range – two focusing on accuracy, and one on precision. These tests are briefly described below and the benchmark values used for assessing vendor performance are summarized in Table 4 (based on evaluation of vendor data limited to CCS sites). Note that the benchmarks depend on volume range and the number of available evaluation sites.

- **Accuracy Test #1**: ***'Minimally 95% Probability, Traffic Count Error (TCE) Median Error Bias (%)'***: Calculate the percentage error (called TCE using FHWA terminology), take the median value, and determine whether it falls within the established range.

[3] "Validation of Estimated Annual Average Daily Traffic (AADT) for Non-Traditional Data Sources" (link)

- **Accuracy Test #2**: '**Minimally 95% Probability, MAPE Upper 95% CB (%)**': Calculate the Mean Absolute Percentage Error (MAPE) and determine whether it is less than the allowable upper bound.
- **Precision Test**: '***Minimally 95% Probability, 95% TCE Population Error Range (%)***': Calculate the percentage error (PE) for each location, determine whether each location's PE falls within the established range (pass/fail), and count the number of 'fails' and compare it to the allowable limit determined by the sampling plan. The sampling plan is based on statistical quality control theory, and specifies the acceptable performance of error (e.g., balance between type I and type II errors) for different sample sizes. For this validation activity, the default sampling plan from FHWA's report was used.

**Table 4 -- FHWA accuracy/precision thresholds when evaluating using CCS locations.**

| AADT Volume Range† | Sample Size for Evaluation | Minimally 95% Probability, TCE Median Error (Bias) (%) | Minimally 95% Probability, MAPE Upper 95% CB (%) | Minimally 95% Probability, 95% TCE Population Error Range (%) |
|---|---|---|---|---|
| 500 – 4,999 (low) | N ≥ 1,000 | ± 1.6 | 13.0 | ± 43.4 |
| | N = 200 | ± 3.1 | 14.5 | |
| | N = 100 | ± 4.5 | 15.5 | |
| | N = 50 | ± 6.4 | 16.7 | |
| | N = 25 | ± 8.7 | 18.1 | |
| 5,000 – 54,999 (medium) | N ≥ 1,000 | ± 2.0 | 10.5 | ± 33.6 |
| | N = 200 | ± 2.9 | 11.3 | |
| | N = 100 | ± 3.8 | 11.9 | |
| | N = 50 | ± 5.2 | 12.7 | |
| | N = 25 | ± 6.9 | 13.6 | |
| 55,000 + (high) | N ≥ 1,000 | ± 2.7 | 8.2 | ± 22.0 |
| | N = 200 | ± 4.0 | 8.9 | |
| | N = 100 | ± 5.3 | 9.8 | |
| | N = 50 | ± 7.1 | 11.2 | |
| | N = 25 | ± 9.4 | 12.9 | |

## Visual Inspection

While ADT and AADT data items involve only one statistical error metric (and thus one opportunity for comparison) per location, more granular data items provide opportunities to investigate temporal patterns – both recurrent (AHDT, focusing on time-of-day patterns by day of week) and non-recurrent (hourly and daily volumes, which can capture perturbations in traffic demand).

Vendor performance can be assessed visually by overlaying vendor and reference traces on a time series plot. The nuances of how vendor traces match the reference data counts across time can be intuitively understood through visual inspection, even when it is more difficult to capture quantitatively via error metrics. A key goal of this approach is to identify which of the reported error metrics best capture how well vendor-provided data agrees with reference data as revealed by the "eye test".

# Results and Discussion

The results are reported below, starting with AADT and moving sequentially through the data elements with increasing temporal granularity. The first section, AADT, involves all 5 vendors, while the remaining sections focus on the 3 vendors (A,C,D) who submitted the full set of aggregations for April 2024. Given the emphasis on quantifying accuracy for FHWA reporting purposes, the AADT section is covered in more detail, with other sections focusing primarily on top-level visuals and metrics.

When analyzing the results, note that there are several different sources of error that can contribute to discrepancies between vendor and reference volumes (beyond just model inaccuracy). **A source of error that emerged in this study was related to location referencing**; evaluation locations were defined by the validation team via a georeferencing protocol (CWGP), but behind the scenes each vendor needed to translate these to their own road network representation – presenting an opportunity for conflation errors. For example, one vendor's initial results showed significant errors on certain high-volume facilities – an unintuitive result as volumes are typically more accurate for higher volume flows. Upon further inspection, the source of the error was rooted in location referencing. In this case, the referenced location corresponded to multiple roadway facilities (i.e. lanes) on their road network, and they were only reporting volumes from one of those facilities. In cases such as this example -- where data issues were identified and communicated clearly -- the validation team encouraged vendors to resubmit results. Instances such as these reveal that in some cases, errors in volume measures may be attributed to errors in roadway location referencing, and that greater emphasis is needed in systematic methods for ensuring accurate conflation procedures.

The validation team also inspected the reference data for sources of error. Validation locations for which all vendors showed high levels of error were analyzed closely to understand potential issues with reference counts or aggregation processes. In one instance, this investigation uncovered a count site that was reporting on only a subset of lanes, resulting in volumes much smaller than what vendors were reporting. Both directional locations at this site were omitted from the analysis. Additionally, even when reference counts are accurate, aggregating them to measures such as AHDT, ADT, and AADT require making decisions about how much missing data can be tolerated. In a preliminary analysis the validation team used all locations for where measures could be calculated, but later imposed data availability restrictions to ensure aggregate measures were representative (see 'Reference Data' subsection of Data Collection above).

## AADT

Figure 10 plots vendor AADT volumes (y axis) against reference AADT values (x axis) for each vendor, focusing in particular on the 0-35k veh/day volume range where approximately 90% of the data lies. The dashed red line represents points where there is zero error (i.e., estimates are equal to reference values), while the dotted black lines represent +/- 15% error. In general, all vendors consistently produced estimates that frequently fell within the 15% bounds, with some notable outliers (e.g.. overestimates by Vendor C at a handful of locations) observed. It should be noted that the 15% bounds are visually indistinguishable from the 45-degree line at very low volumes, meaning that AADT estimates can easily have sizeable percentage errors even when vendor estimates match the reference data well.

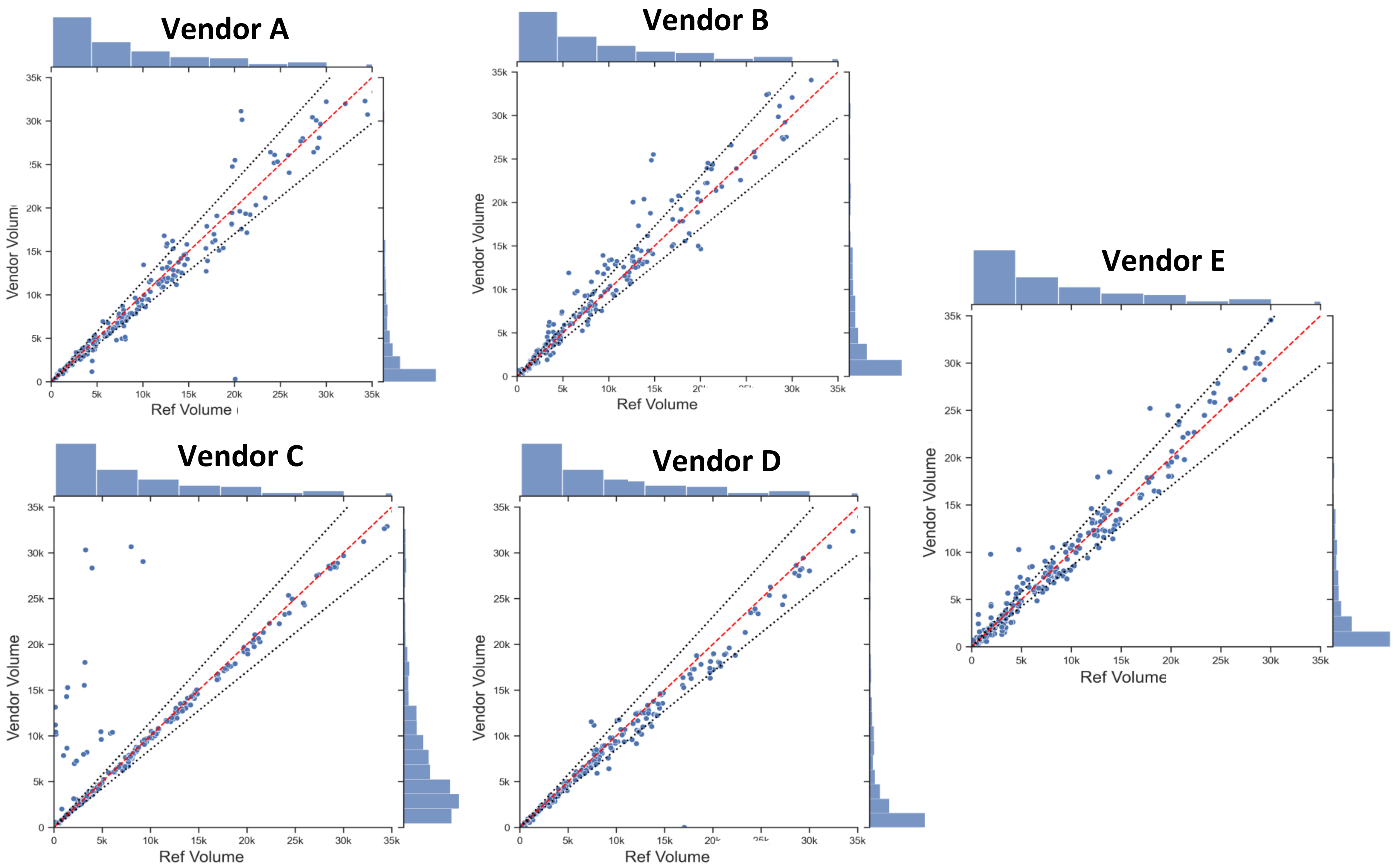


**Figure 10 – AADT scatter plots by vendor.**

Figure 11 shows the distribution of AADT error percentages for each vendor via boxplots, separated by volume bin. These volume bins were chosen to coincide with FHWA's volume bin definitions, so observations are not distributed equally across ranges, and several – particularly the highest two bins, have smaller sample sizes than would typically be used for boxplots. Each boxplot summarizes the distribution of error, with the box showing the middle 50% of data (ranging from the 1st to the 3rd quartiles – often referred to as the interquartile range), with the vertical line in the middle of the box representing the median value, 50th percentile, the whiskers extending below the 25th percentile (Q1) and above 75th percentile (Q3) by 1.5 times the inter-quartile range (Q3-Q1) to represent the estimated non-outlier range, and red 'X's represent the mean value of the error distribution. Including the mean value is useful for identifying the presence of large outliers.

Visually, vendors A, C and D stand out as consistent performers across most volume ranges, with Vendor C showing very high accuracy across mid-to-high volume ranges, and Vendors D and A showing consistent performance across all volume ranges. In all but the lowest volume bins, Vendor C stands out as having very tight box plots centered close to zero, indicating highly accurate estimates whose errors do not vary across locations. However, vendor C has very high errors in the lowest volume range (the box exceeds plot limits), and very large outliers in the next two lowest (mean value is far to the right of the median). In these low-volume cases, Vendor C significantly overestimates volumes relative to the reference data, resulting in high percent error. Vendors D and A exhibit errors that are consistent across all volume ranges, even the lowest volume bins where percentage error tends to be higher (by the nature of calculating error relative to a small number), with both vendors having mean and median values within +/- 15%. The other vendors, B and E, produced estimates that were accurate in some ranges, but less so in others, without clear over/underestimation trends.

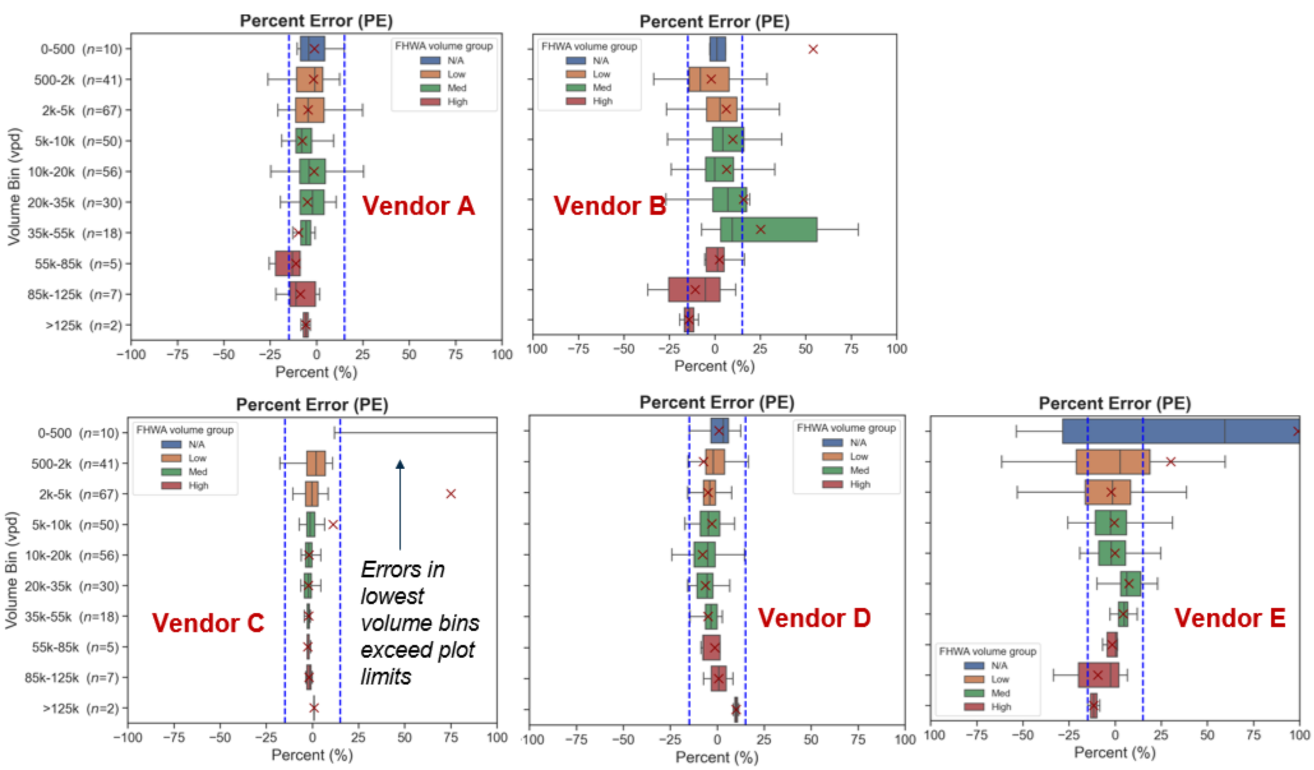


**Figure 11 – AADT box plots.**

It should be noted that Vendor A's results in Figure 11 reflect re-submitted data after a conflation issue was identified in the initial submission. The left plot in Figure 12 shows the initial submission, with significant underestimates in the highest volume bins. After fixing the mapping between validation locations and their internal road network representation, the high-volume locations were more accurate, shown in the right plot of Figure 12 (and corresponding to what was shown in Figure 10).

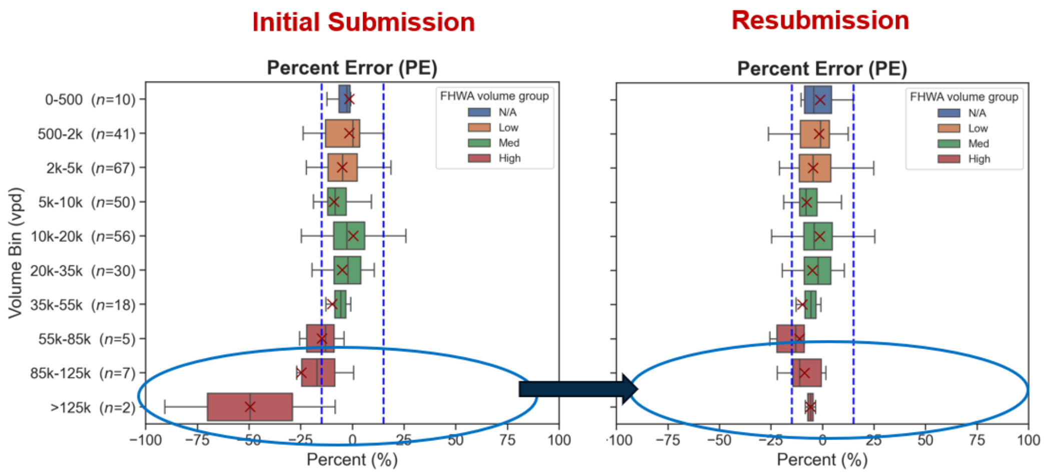


**Figure 12 – Impact of Vendor A fixing conflation issue**

## Urban vs Rural

Recognizing that the results from Figure 11 encompass both urban and rural locations, it can be instructive to separate the results to determine whether performance is consistent across location types. Using GDOT's classification for each count station, Figure 13 shows the results separated by urban/rural location for each vendor, with urban results on the left and rural results on the right. When interpreting the results visually, note that the urban and rural bar charts for a given volume bin may be based on different sample sizes. In some cases, the small sample sizes are sub-optimal for summarizing a distribution via boxplot; however, these results are still instructive for investigating possible differences in performance between urban and rural locations. Based on the plots below, no systematic evidence exists to suggest degradation in performance between urban and rural location types.

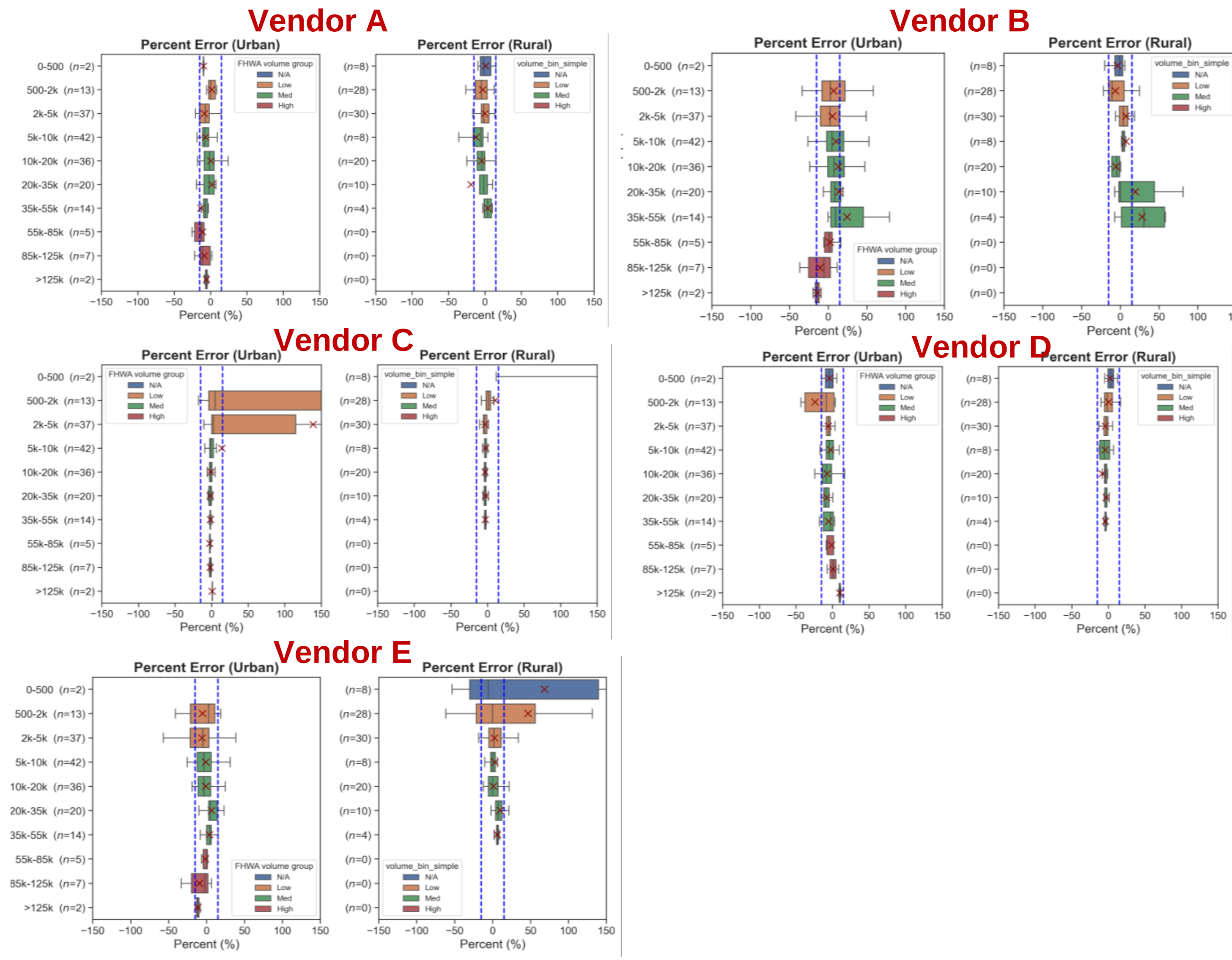


**Figure 13 – AADT Boxplots broken down by Urban vs Rural locations**

## Spatial Consistency Test

Based on feedback from the prior validation activity, the validation team included an additional 'sanity check' test to assess spatial consistency. The idea is to see whether vendor estimates at multiple locations are internally consistent – a test which can be run even if reference data is not available to quantify accuracy measures at specific locations. Figure 14 shows one such interchange where the spatial consistency test was executed; eight validation locations were defined at this interchange, and vendor estimates were used to test conservation of flow (i.e., volume from four legs entering the interchange should roughly equal volumes from the four legs leaving the intersection). This test was executed for all temporal aggregations (hourly counts to AADT), and each temporal aggregation exhibited similar results as the AADT results.

Preliminary calculations indicated that four of the five vendors showed basic conservation of flow, with volumes into and out of the interchange agreeing within about 3%, with the remaining vendor (Vendor C) having about a 25% difference. After bringing this to the vendor's attention, they discovered an error in data processing for one of the locations. After addressing this issue, their performance was consistent with other vendors. It is expected that variations of this test will be integrated more fully in future validation activities.

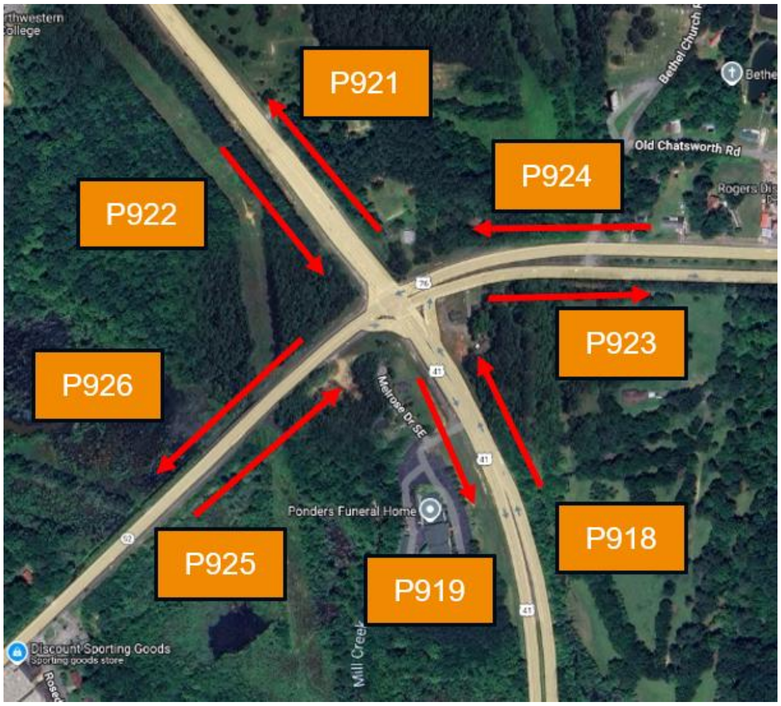


**Figure 14 – Spatial Consistency Test**

## FHWA Method

As described above, the FHWA method establishes accuracy and precision benchmarks in three volume ranges (Low: 500-4999 vpd, Medium: 5000-54999 vpd, High: 55000+ vpd). These benchmarks reflect the expected precision/accuracy of factoring 48-hour counts to AADT values and are based on prior FHWA research. The idea, then, is to calculate the same measures for each vendor's AADT product and compare with FHWA's benchmarks to see whether the estimates are at least as accurate/precise as factoring 48-hr counts (the prevailing method). Overall, there are three benchmarks – two accuracy and one precision test – that are applied in each of the three volume bins bin, resulting in 9 opportunities for evaluation.

Before presenting the results, one major caveat is noted related to directionality of volumes. The Coalition volume validation activity was explicitly designed to test *directional* traffic volumes; regardless of facility type, each validation location refers to a single direction of travel on a specified roadway segment or location. All results presented thus far – and all volumes associated with all other temporal aggregations (ADT, AHDT, Daily, Hourly) are directional volumes. However, FHWA's statistical method is based on bi-directional volumes, using the total volume associated with both directions of traffic flow at a count site. At a state agency level, there is some ambiguity about how agencies define count sites (e.g., whether volume counts on median-separated interstate facilities are captured and recorded as direction or bi-directional.

Accordingly, the validation team generated results both for directional volumes (in line with how vendors submitted data) and bi-directional volumes (aligned with FHWA's perspective). To obtain bi-directional volumes, we summed direction volumes at each site – both for reference and vendor datasets. There are two key implications of aggregating directional volumes to bidirectional volumes for validation: (1) the sample size is cut in half and (2) the distribution of observations in each volume bin (low/medium/high) shifts because volumes roughly double at each location. These shifts in sample size and distribution within volume bins result in different benchmarks, as noted in Table 4.

Noting this caveat, Table 5 summarizes the results using *directional volumes*, while Table 6 summarizes the results for *bi-directional volumes*. In both cases, results are presented for each of the three tests and three volume bins for each vendor, resulting in nine tests per vendor. Benchmark limits are also shown to provide context for how close each vendor's result was to the pass/fail threshold, and a summary tally reports the total number of tests passed (just right of the table).

First, consider the directional results from Table 5. Note that results are presented for all three volume bins, although the 'High' volume bin has only 14 observations, while a minimum of 25 are recommended. All vendors passed at least 4 of the 9 tests, while two vendors, Vendors D and C, separated themselves by passing 8 and 7 tests, respectively. Vendor D's only failed test was close to the limit (4.8% median percent error, about 1% higher than the benchmark), so it came close to passing in every category. Vendor C easily passed every test in the medium and high bins but significantly failed two tests in the Low volume bin. This result for vendor C is consistent with the box plots in Figure 9, which highlight significant errors in the lower volume bins, and low error in the medium and high volume ranges.

**Table 5 – FHWA method results (directional).**

| | | | Accuracy Test #1 | | | Accuracy Test # 2 | | | Precision Test | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Vendor | Volume Category | Number of Locations | Absolute Median TCE | Limit | Result | MAPE | Limit | Result | # Failures | Limit | Result | # Tests Passed |
| A | Low | 108 | 4.0 | 4.5 | Passed | 10.5 | 15.5 | Passed | 2 | 4 | Passed | 5 |
| | Medium | 154 | 5.1 | 3.8 | Failed | 11.5 | 11.9 | Passed | 10 | 4 | Failed | |
| | High | 14* | 10.0 | 9.4 | Failed | 11.6 | 12.9 | Passed | 3 | 0 | Failed | |
| B | Low | 108 | 1.7 | 4.5 | Passed | 15.9 | 15.5 | Failed | 10 | 4 | Failed | 4 |
| | Medium | 154 | 4.5 | 3.8 | Failed | 16.6 | 11.9 | Failed | 22 | 4 | Failed | |
| | High | 14* | 5.1 | 9.4 | Passed | 12.2 | 12.9 | Passed | 2 | 0 | Failed | |
| C | Low | 108 | 0.2 | 4.5 | Passed | 102.9 | 15.5 | Failed | 19 | 4 | Failed | 7 |
| | Medium | 154 | 1.9 | 3.8 | Passed | 6.9 | 11.9 | Passed | 4 | 4 | Passed | |
| | High | 14* | 1.8 | 9.4 | Passed | 2.0 | 12.9 | Passed | 0 | 0 | Passed | |
| D | Low | 108 | 4.1 | 4.5 | Passed | 8.6 | 15.5 | Passed | 2 | 4 | Passed | 8 |
| | Medium | 154 | 4.8 | 3.8 | Failed | 8.3 | 11.9 | Passed | 3 | 4 | Passed | |
| | High | 14* | 1.0 | 9.4 | Passed | 6.7 | 12.9 | Passed | 0 | 0 | Passed | |
| E | Low | 108 | 1.5 | 4.5 | Passed | 32.5 | 15.5 | Failed | 17 | 4 | Failed | 5 |
| | Medium | 154 | 1.6 | 3.8 | Passed | 9.9 | 11.9 | Passed | 5 | 4 | Failed | |
| | High | 14* | 3.5 | 9.4 | Passed | 8.8 | 12.9 | Passed | 2 | 0 | Failed | |

*Note: Results are reported for 'High' volume category although sample size is less than 25.*

Table 6 reports results from the same tests but focuses on *bi-directional volumes.* As with Table 5, Table 6 presents accuracy and precision results in all three volume bins, although the 'High' volume category now has 20 observations rather than 14, which is still less than the 25 minimum recommended observations. As was the case with directional volumes, all vendors passed at least 4 of the 9 tests, while two vendors, Vendors D and A, separated themselves by passing 7 tests. Vendor D's performance was similar to the directional results and came close to passing all 9 tests (only two fails, one by a small margin for Accuracy Test #1 and by 1 failure on the precision test with 0 allowable failures). Combining directions helped Vendor A pass 7 tests as well, as compared to 5 tests in the directional version, with the two 'fails' coming from the precision test in medium and high bins. Vendor C, which passed 7 of 9 tests in the directional version, passed 5 of 9 for bi-directional volumes. It appears that the significant lower-volume outliers, which previously confined to the Low volume category, spilled into the Medium category (bidirectional volumes shifted some locations from low to medium volume category).

**Table 6 – FHWA method results (bi-directional).**

| | | | Accuracy Test #1 | | | Accuracy Test # 2 | | | Precision Test | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Vendor | Volume Category | Number of Locations | Absolute Median TCE | Limit | Result | MAPE | Limit | Result | # Failures | Limit | Result | # Tests Passed |
| A | Low | 27 | 0.5 | 8.7 | Passed | 8.9 | 18.1 | Passed | 0 | 0 | Passed | 7 |
| | Medium | 93 | 4.2 | 5.2 | Passed | 10.7 | 12.7 | Passed | 5 | 2 | Failed | |
| | High | 20* | 6.1 | 9.4 | Passed | 10.4 | 12.9 | Passed | 2 | 0 | Failed | |
| B | Low | 27 | 8.1 | 8.7 | Passed | 16.8 | 18.1 | Passed | 1 | 0 | Failed | 4 |
| | Medium | 93 | 3.1 | 5.2 | Passed | 15.0 | 12.7 | Failed | 11 | 2 | Failed | |
| | High | 20* | 3.1 | 9.4 | Passed | 19.8 | 12.9 | Failed | 5 | 0 | Failed | |
| C | Low | 27 | 1.4 | 8.7 | Passed | 149.5 | 18.1 | Failed | 5 | 0 | Failed | 5 |
| | Medium | 93 | 1.7 | 5.2 | Passed | 21.0 | 12.7 | Failed | 6 | 2 | Failed | |
| | High | 20* | 2.0 | 9.4 | Passed | 2.2 | 12.9 | Passed | 0 | 0 | Passed | |
| D | Low | 27 | 3.0 | 8.7 | Passed | 10.1 | 18.1 | Passed | 1 | 0 | Failed | 7 |
| | Medium | 93 | 5.5 | 5.2 | Failed | 7.7 | 12.7 | Passed | 2 | 2 | Passed | |
| | High | 20* | 2.3 | 9.4 | Passed | 4.7 | 12.9 | Passed | 0 | 0 | Passed | |
| E | Low | 27 | 1.3 | 8.7 | Passed | 44.7 | 18.1 | Failed | 6 | 0 | Failed | 5 |
| | Medium | 93 | 0.3 | 5.2 | Passed | 12.3 | 12.7 | Passed | 8 | 2 | Failed | |
| | High | 20* | 2.5 | 9.4 | Passed | 7.0 | 12.9 | Passed | 1 | 0 | Failed | |

*Note: Results are reported for 'High' volume category although sample size is less than 25.*

## Comparison with Prior Results

The prior validation in North Carolina did not assess AADT accuracy/precision with the same FHWA statistical method because the current guidance was not available at the time. An alternative method recommended by FHWA for cases with smaller sample sizes was used, which involved simply reporting a variety of error metrics across detailed volume ranges.

With the new guidance now available, the validation team revisited the Noth Carolina datasets and calculated FHWA statistical tests for both directional and bi-directional volumes – similar to Tables 5-6 above. Additionally, to make results more comparable, additional data filtering (using the same data availability threshold as in GA) was implemented when re-running the FHWA analysis.

Tables 6 and 7 summarize the results for directional and bi-directional results, respectively. The results are fairly consistent across all four vendors, ranging from 4-6 of the 9 tests passed for both directional and bi-directional tables. Relative to these results in North Carolina, we see that several vendors have taken a step forward in the Georgia validation, with three vendors passing at least 7 of the 9 tests (in the directional Table 5 or bi-directional Table 6).

**Table 6 – FHWA method results for prior study in NC (directional).**

| Vendor | Volume Category | Number of Locations | Accuracy Test #1 Absolute Median TCE | Accuracy Test #1 Limit | Accuracy Test #1 Result | Accuracy Test # 2 MAPE | Accuracy Test # 2 Limit | Accuracy Test # 2 Result | Precision Test # Failures | Precision Test Limit | Precision Test Result | # Tests Passed |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | Low | 45 | 14.2 | 8.7 | Failed | 27.7 | 18.1 | Failed | 8 | 0 | Failed | 6 |
| | Medium | 164 | 2.1 | 3.8 | Passed | 9.2 | 11.9 | Passed | 0 | 4 | Passed | |
| | High | 21* | 1.8 | 9.4 | Passed | 6.6 | 12.9 | Passed | 0 | 0 | Passed | |
| B | Low | 48 | 10.8 | 8.7 | Failed | 16.2 | 18.1 | Passed | 4 | 0 | Failed | 4 |
| | Medium | 165 | 6.5 | 3.8 | Failed | 10.4 | 11.9 | Passed | 6 | 4 | Failed | |
| | High | 21* | 2.6 | 9.4 | Passed | 7.9 | 12.9 | Passed | 2 | 0 | Failed | |
| C | Low | 13* | 19.3 | 8.7 | Failed | 27.8 | 18.1 | Failed | 1 | 0 | Failed | 6 |
| | Medium | 153 | 0.7 | 3.8 | Passed | 7.0 | 11.9 | Passed | 0 | 4 | Passed | |
| | High | 20* | 3.6 | 9.4 | Passed | 5.7 | 12.9 | Passed | 0 | 0 | Passed | |
| D | Low | 50 | 4.2 | 8.7 | Passed | 7.3 | 18.1 | Passed | 1 | 1 | Passed | 6 |
| | Medium | 165 | 10.9 | 3.8 | Failed | 13.1 | 11.9 | Failed | 8 | 4 | Failed | |
| | High | 21* | 3.7 | 9.4 | Passed | 4.6 | 12.9 | Passed | 0 | 0 | Passed | |

**Results are reported for volume category although sample size is less than 25.*

**Table 7 – FHWA method results for prior study in NC (bi-directional).**

| Vendor | Volume Category | Number of Locations | Accuracy Test #1 Absolute Median TCE | Accuracy Test #1 Limit | Accuracy Test #1 Result | Accuracy Test # 2 MAPE | Accuracy Test # 2 Limit | Accuracy Test # 2 Result | Precision Test # Failures | Precision Test Limit | Precision Test Result | # Tests Passed |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | Low | 20* | 16.1 | 8.7 | Failed | 33.0 | 18.1 | Failed | 5 | 0 | Failed | 6 |
| | Medium | 75 | 0.7 | 5.2 | Passed | 10.1 | 12.7 | Passed | 1 | 1 | Passed | |
| | High | 31 | 4.0 | 9.4 | Passed | 8.6 | 12.9 | Passed | 1 | 0 | Failed | |
| B | Low | 23* | 9.5 | 8.7 | Failed | 16.7 | 18.1 | Passed | 2 | 0 | Failed | 4 |
| | Medium | 76 | 7.0 | 5.2 | Failed | 11.5 | 12.7 | Passed | 3 | 1 | Failed | |
| | High | 31 | 2.5 | 9.4 | Passed | 7.0 | 12.9 | Passed | 2 | 0 | Failed | |
| C | Low | 2* | | | | | | | | | | 5 |
| | Medium | 68 | 0.5 | 5.2 | Passed | 7.9 | 12.7 | Passed | 2 | 1 | Failed | |
| | High | 30 | 2.4 | 9.4 | Passed | 6.4 | 12.9 | Passed | 0 | 0 | Passed | |
| D | Low | 26 | 5.5 | 8.7 | Passed | 11.7 | 18.1 | Passed | 2 | 0 | Failed | 6 |
| | Medium | 76 | 9.2 | 5.2 | Failed | 12.3 | 12.7 | Passed | 2 | 1 | Failed | |
| | High | 31 | 0.0 | 9.4 | Passed | 8.0 | 12.9 | Passed | 0 | 0 | Passed | |

**Results are reported for volume category although sample size is less than 25.*

## AADT Summary

Overall, the AADT analysis shows that industry has taken a step forward in the past year, with several vendors' products showing improvement since they were last analyzed in North Carolina. There is more differentiation between vendors than in the prior study, with Vendor D and Vendor A standing out as consistently reporting AADT within 10-15% of reference values across a wide range of volumes and Vendor C showing even higher accuracy across mid-to-high volume ranges (though less accurate at low volumes). Several vendors are on the verge of meeting FHWA guidelines, though the results differ slightly based on whether tests are run on directional or bi-directional volumes. Further clarification is needed on this issue, and the validation team intends to investigate this further during the next study.

One caveat of these results is that AADT is not a truly blind study; historical CCS data was provided to each vendor for calibrating their models for the primary study period (April 2024). However, all indications point to vendors' volume estimates reflecting actual model outputs. For example, the results from ADT (the same aggregation as AADT, but for a month) in April 2024, which was not included in the historical calibration data, were consistent with AADT accuracy.

Finally, this AADT analysis reinforces that there are several possible sources of error, of which vendor model performance is just one. Another key source of error includes location referencing/conflation – particularly when vendors translate validation locations to their native map representations, reference data processing. Additionally, the reference data requires careful inspection to deal with abnormalities and missing data that can impact annual average calculations.

## ADT

Like AADT, ADT focuses on average daily traffic, but computes averages over a month (April 2024) rather than a full calendar year (2023). Figure 15 plots vendor ADT volumes (y axis) against reference AADT values (x axis) for the three vendors who submitted data in this category, zooming in to the 0-35k veh/day volume range where approximately 90% of the data lies. The dashed red line represents points where there is zero error (i.e., estimates are equal to reference values), while the dotted black lines represent +/- 15% error. Visually, the plots look similar to their AADT counterparts (as expected) – with a notable difference being the lack of significant overestimation outliers for vendor C that were present in the AADT dataset.

Figure 16 shows the distribution of AADT error percentages for each vendor via boxplots, arranged by volume bin. There are some slight differences between vendors, but in general, all ADT estimates are consistently within +/- 15% across volume ranges. As was observed visually in the scatter plots, the boxplots communicate that Vendor C's ADT estimates for April of 2024 are much more accurate and less impacted by large outliers in low-volume bins than for AADT volume estimates for 2023.

### ADT Summary

The results from the three vendors who submitted ADT data are consistent with -- and in some cases better than – AADT accuracy. In general, all three vendors' ADT is typically within 15% of reference values, and Vendor C's low-volume estimates appear to be more accurate than that of the AADT. Although historical counts were made available at these locations, the ADT estimates were generated for a time period not contained in historical calibration data (April '24). The fact these results are consistent with AADT, which focused on a time period that aligned with historical calibration data made available to vendors, supports the argument that AADT estimates are representative of typical model performance.

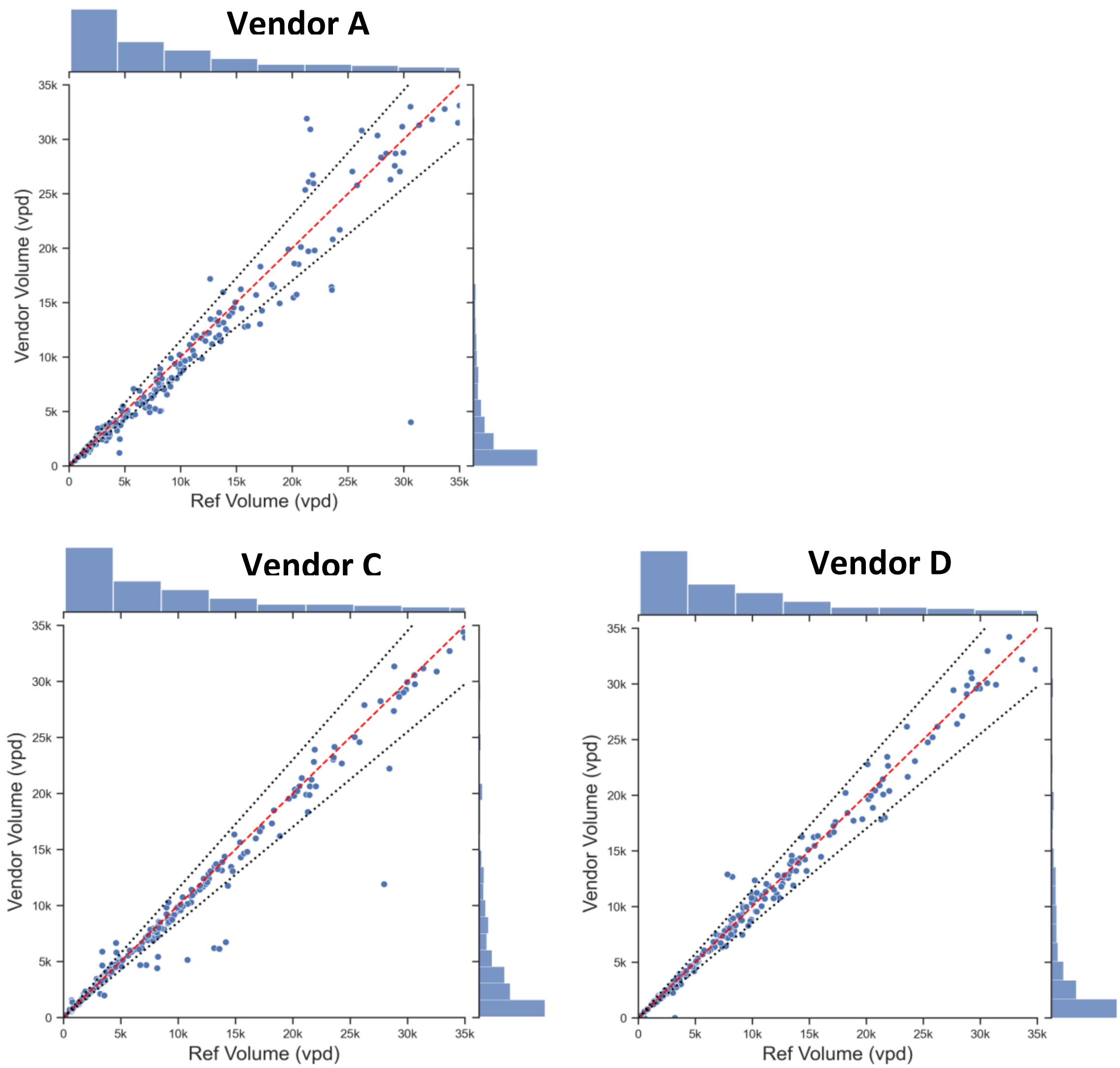


**Figure 15 – ADT scatter plots by vendor.**

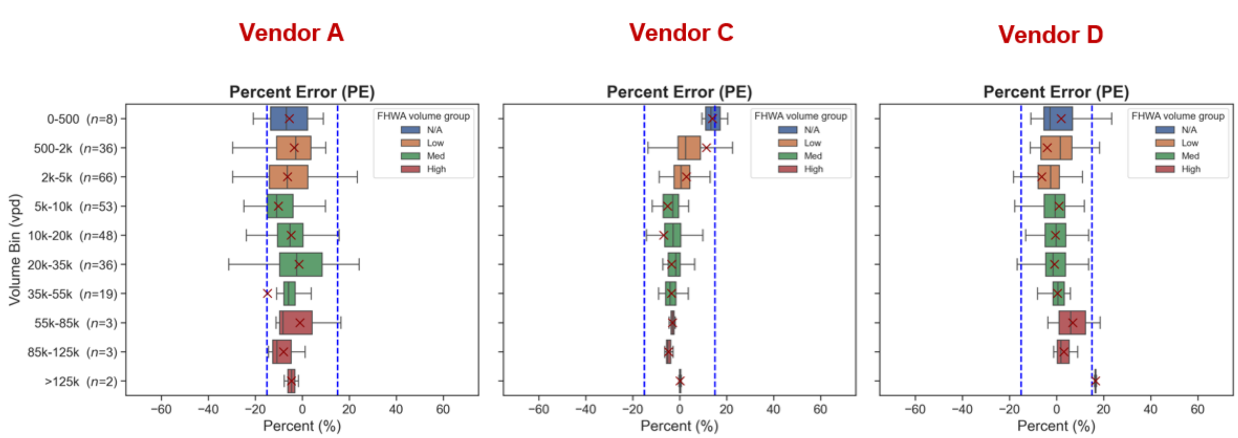


**Figure 16 – ADT box plots.**

## AHDT

AHDT values represent a step up in temporal granularity from ADT, with 168 hourly records at each location rather than one (24 hours * 7 days). This enables an analyst to visually inspect time series plots, which can help build insight for how well vendor data captures recurrent patterns. Inspecting plots at dozens of locations highlighted that in general, **all three vendors were capable of capturing temporal patterns by day-of-week**. The following example shows representative performance from vendors at locations with distinct AM/PM peak patterns.

The first example focuses on a relatively high-volume freeway facility (I-20 near Augusta, about 40k AADT in each direction), with locations P345 and P346 representing Eastbound and Westbound traffic, respectively. Figure 17 shows these locations relative to downtown Augusta.

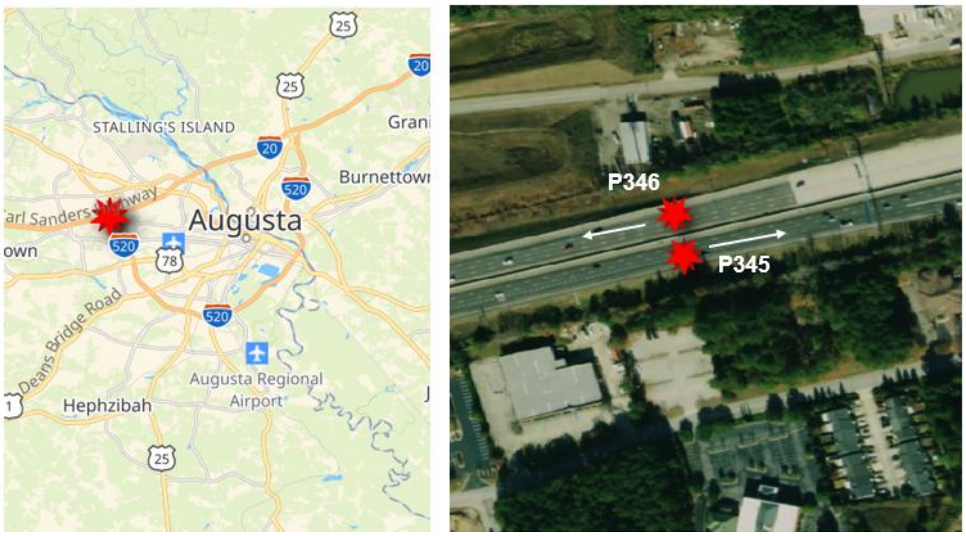


**Figure 17 – Locations P345 and P346.**

Figures 18-23 show time series plots by day-of-week for each of the vendors at these locations (Figs 18-19 = Vendor A, Figs 20-21 = Vendor B, Figs 22-23 = Vendor C). For each vendor, the first figure shows day-of-week volume profiles for the Eastbound direction (P345), while the second figure shows Westbound traffic (P346). As can be observed by the reference data (black curves) in any of the figures, weekday traffic volumes have a larger AM peak in the Eastbound direction (i.e., into Augusta) and larger PM peak in the Westbound direction. Although there is slight error in the magnitude of reported volumes in some cases, all three vendors clearly capture these temporal patterns and differentiate weekday and weekend behavior.

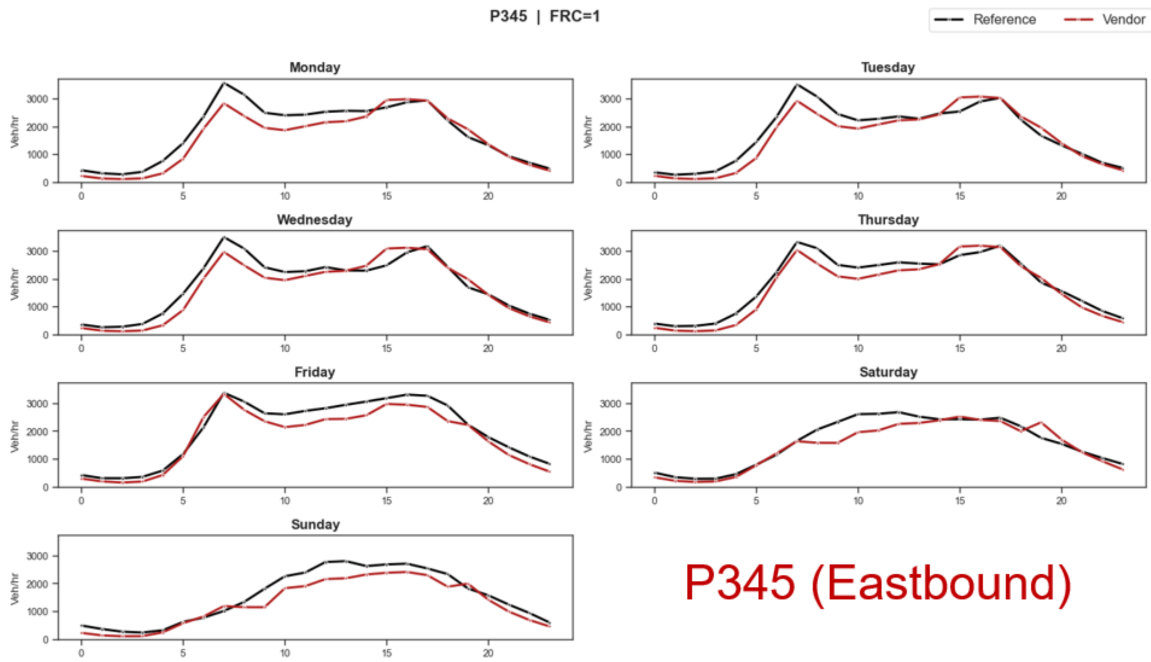


**Figure 18 – AHDT time series by day of week for Vendor A at P345 (Eastbound) .**

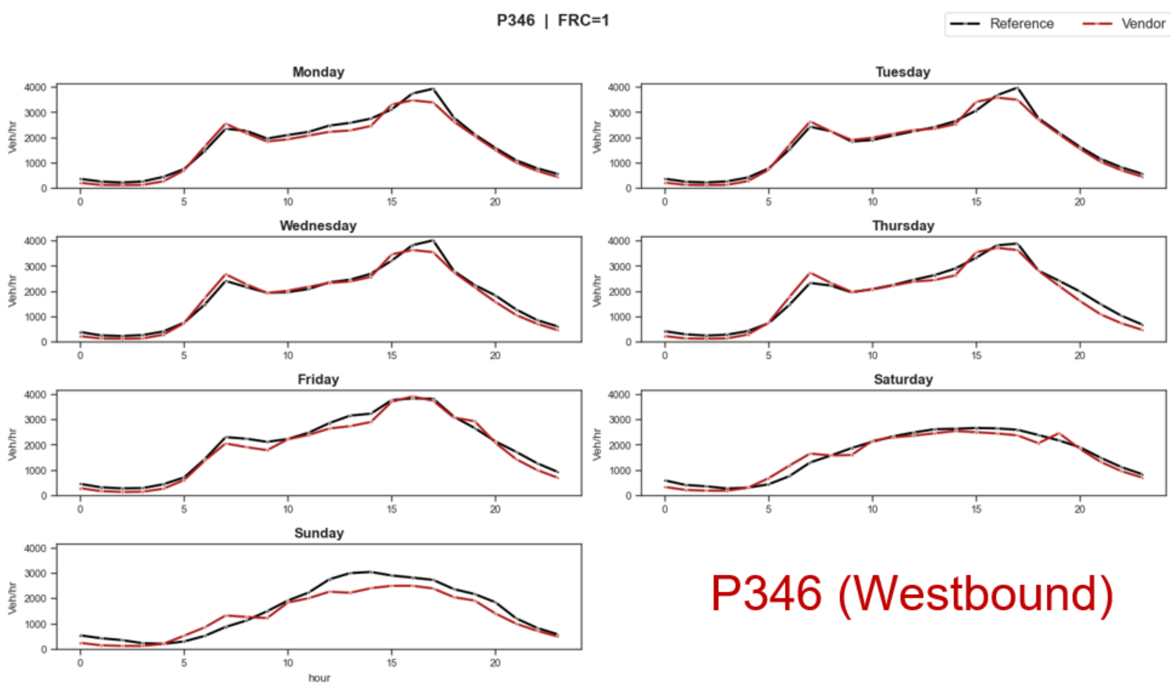


**Figure 19 – AHDT time series by day of week for Vendor A at P346 (Westbound) .**

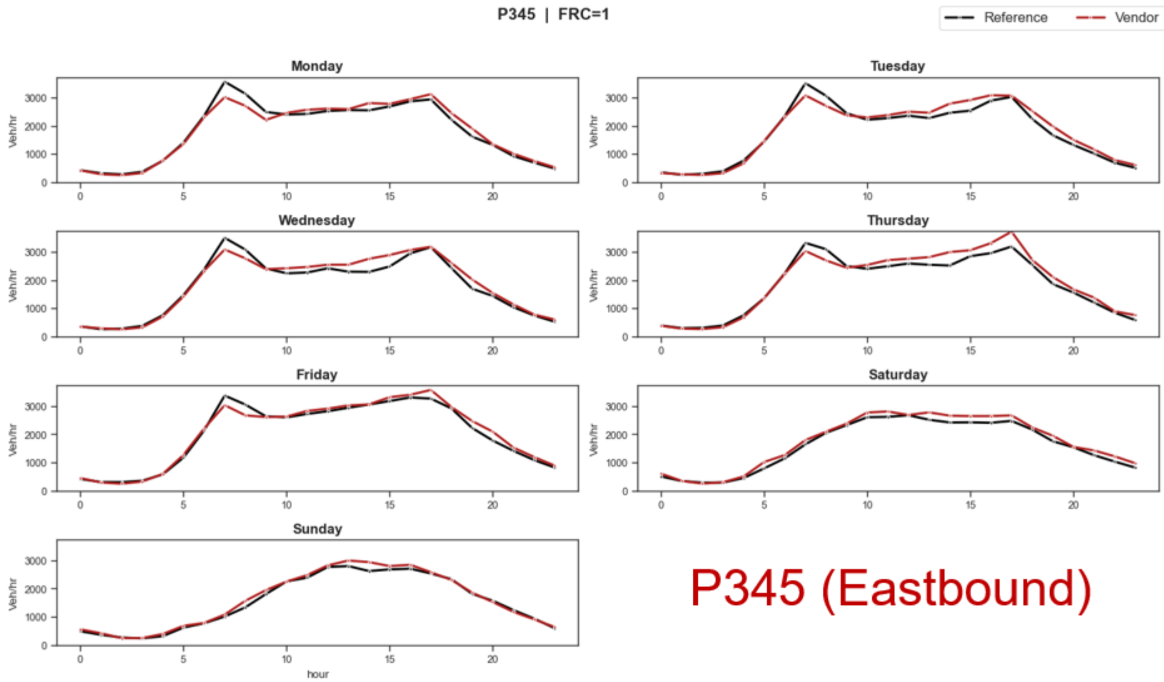


**Figure 20 – AHDT time series by day of week for Vendor C at P345 (Eastbound) .**

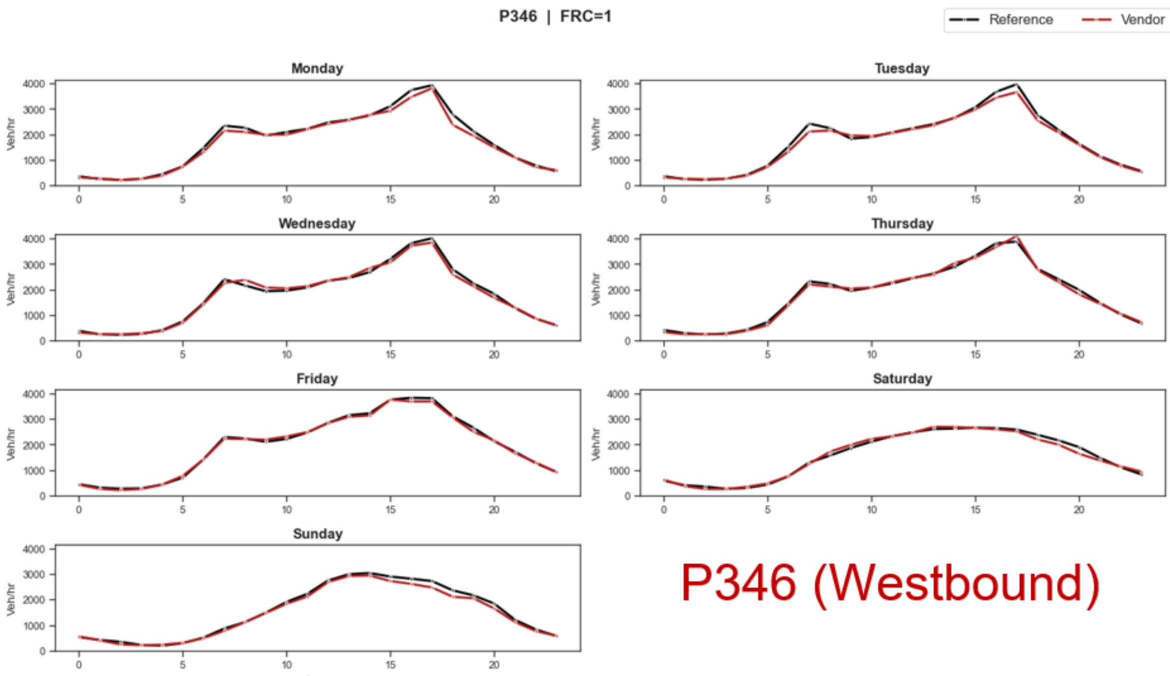


**Figure 21 – AHDT time series by day of week for Vendor C at P346 (Westbound) .**

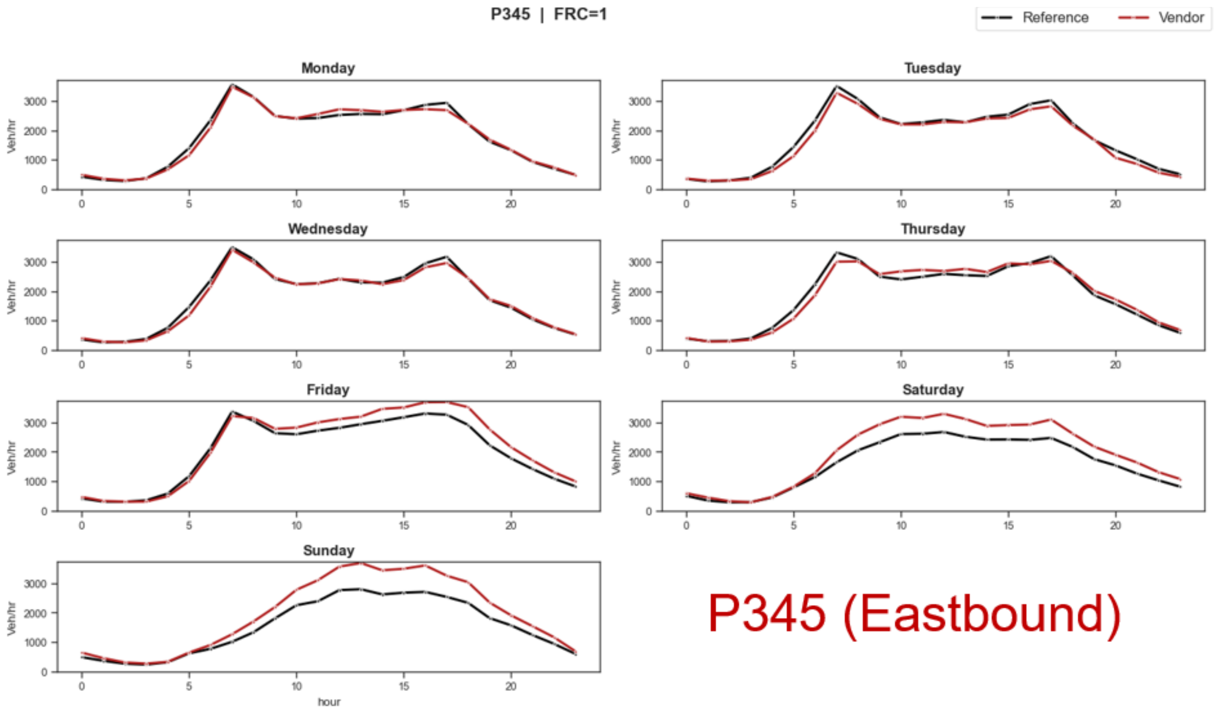


**Figure 22 – AHDT time series by day of week for Vendor D at P345 (Eastbound) .**

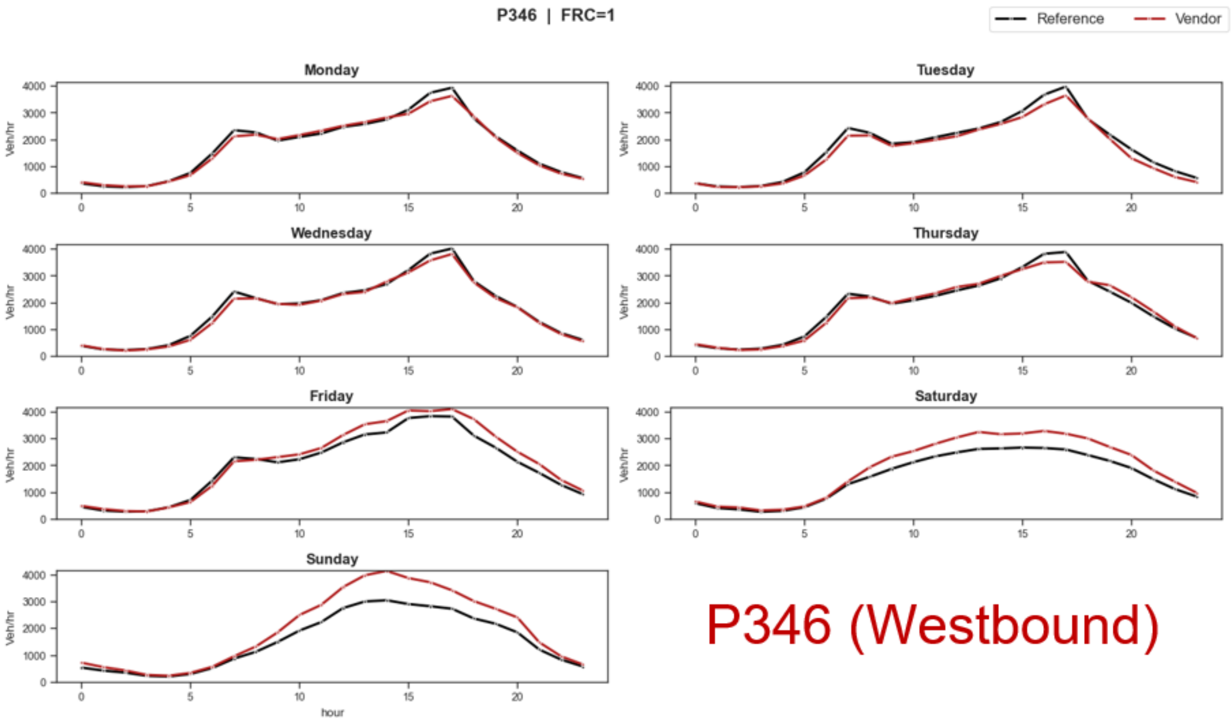


**Figure 23 – AHDT time series by day of week for Vendor D at P346 (Westbound) .**

A key takeaway from this visual inspection was that **Vendor A showed marked improvement in capturing temporal patterns relative to the prior validation in North Carolina**. In the prior validation activity, Vendor A's AHDT product had trouble distinguishing differences in weekday/weekend patterns and sometimes showed directional errors. The ability to capture temporal patterns was noticeably better in this study – in particular differentiating temporal patterns on weekdays and weekends, with fewer directional issues.

Figure 24 shows vendor AHDT volumes (y axis) plotted against reference AHDT values (x axis). As before, the 45-degree line represents perfect estimation, while the dashed black lines show +/- 15% error. However, since there are many more observations in the AHDT dataset (168 hourly observations per location, rather than 1), the data is visualized using a hex-bin plot with darker colors indicating more observations in a region. The concentration of dark bins near the origin reinforces the fact that the majority of observations are quite low volume (see Figure 4). All four vendors generally track the 45-degree line, although there are differences in how consistently they are within the 15% band. Vendor A is generally centered on the 45-degree line but has the widest band of observations, as well as several outliers (significant underestimation). Vendor C has perhaps the tightest band around the 45-degree line but has a noticeable 'sprite' of observations that represent significant underestimation. This sprite of error values may represent a directional versus bi-directional logic error because roadway volumes reported by the vendor are approximately half of the expected volume. Vendor D has a slightly wider band around the 45-degree line but appears to be most consistent across most volume ranges. However, at very high volumes (likely peak periods on high volume roads), Vendor D tends to overestimate volumes.

Figure 25 shows the distribution of AHDT error percentages for each vendor via boxplots, displayed by volume bin. These results provide more information than could be observed visually through the scatter/hex-bin plot in Figure 23. By looking at the sample size associated with each box, we can see that the lowest volume bin (0-100 vph) has the most samples by far (about 14.5k), followed by several other of the lowest bins. In these low volume ranges, Vendor D is most accurate and unbiased. However, it is worth reiterating that it is important to interpret this information with the time series plots in mind, as the summary metrics alone do not tell the whole story; average/overall accuracy does not fully communicate whether volume estimates capture important temporal patterns.

### AHDT Summary

All three vendors (A,C,D) demonstrated the ability to capture time-of-day and day-of-week temporal patterns across cross a variety of locations and road types. This marks a significant improvement for Vendor A, which struggled with this in the prior validation. Although there are still some occasional missed patterns or directionality issues, the estimation accuracy was consistent – including when distinguishing directional volume trends.

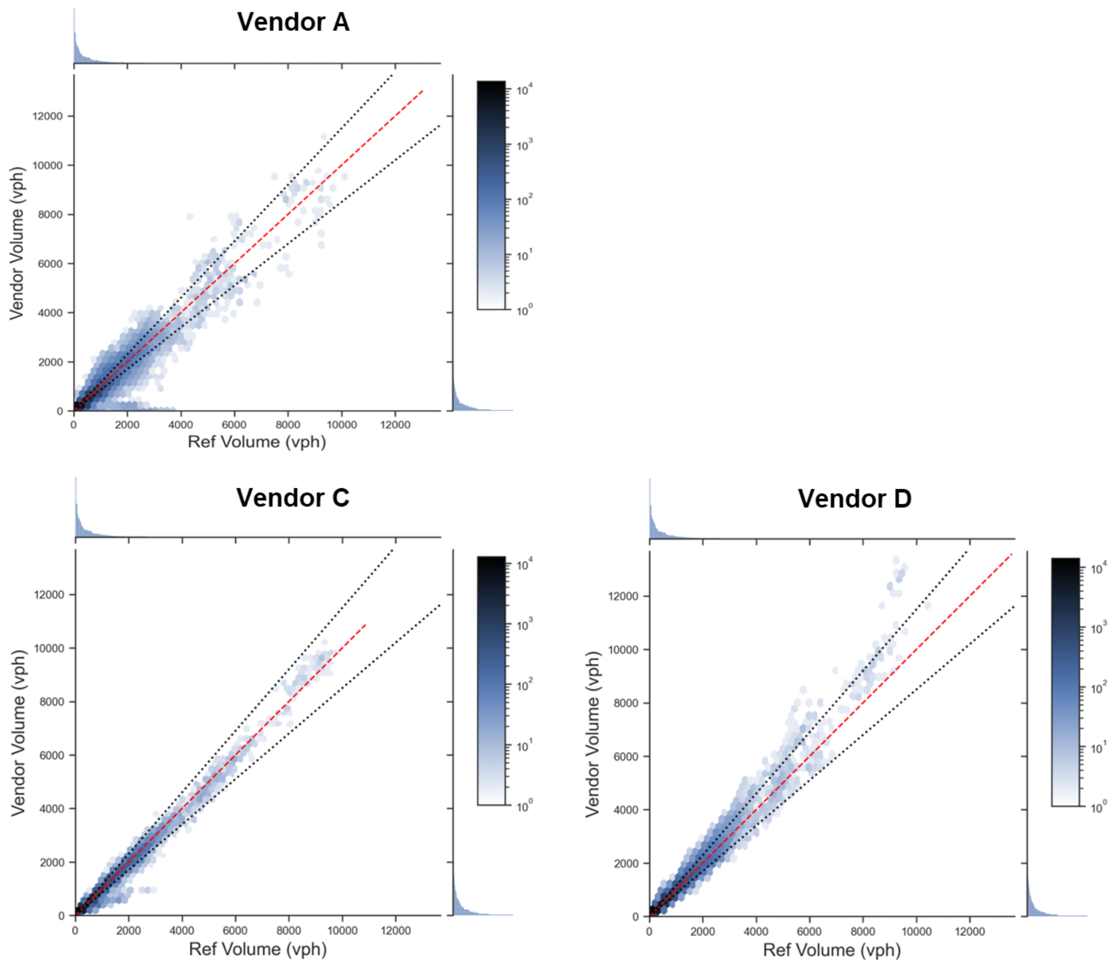


**Figure 24 – AHDT scatter plots by vendor.**

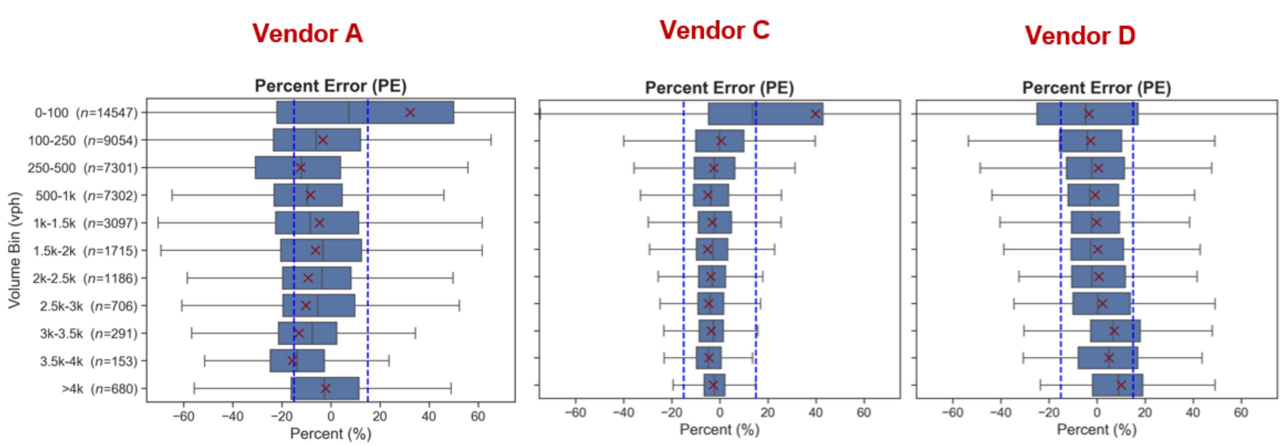


**Figure 25 - Distribution of AHDT percent error metrics by volume bin**

# Daily Volumes

Daily volumes were reported for each date in April 2024 study period, marking the first granular data element that is date-specific rather than based on typical or average behavior (such as ADT, which measures the average daily volume over the month). Data from both CCS and and 48 hour temporary counts (counts taken with portable sensing equipment) were compared to data submitted by the three vendors (A,C,D).

Given that this data element includes volume estimates for specific dates, it is particularly instructive to focus on locations and dates that have anomalous volumes. Also, the value of daily and hourly estimated volumes is to be able to capture atypical (as opposed to average) volume trends. Two locations – one on an FRC 4 and another an a FRC 1 road – were identified as locations with several days of higher-than-usual volumes (April 5-7) that serve as a test of each vendor's ability to capture daily patterns. These locations will be explored in further detail in the next section on Hourly volumes.

Figures 26-28 show daily volumes reported for P060 and P158 (the two locations with volume spikes on April 5-7) for each of the three vendors. In each plot, the black curve shows the reference volume for each date in April, while the red curve shows vendor estimates. From these plots, it is clear that **Vendors C and D are the only two vendors whose product is sensitive to date-specific volumes**. Vendor A's daily estimates appear to be its average volume estimate specific to a day of week, noting the cyclic nature of the daily volume estimates and its insensitivity to the a-typical volumes on April 5-7. Vendors C and D are sensitive to the atypical volumes experience on April 5-7, though not extremely accurate. Vendor C misses one of the volume spikes on P158 and Vendor D significantly overestimates the magnitude of the volume. However, it is still encouraging that Vendors C and D are capable of capturing perturbations in traffic demand.

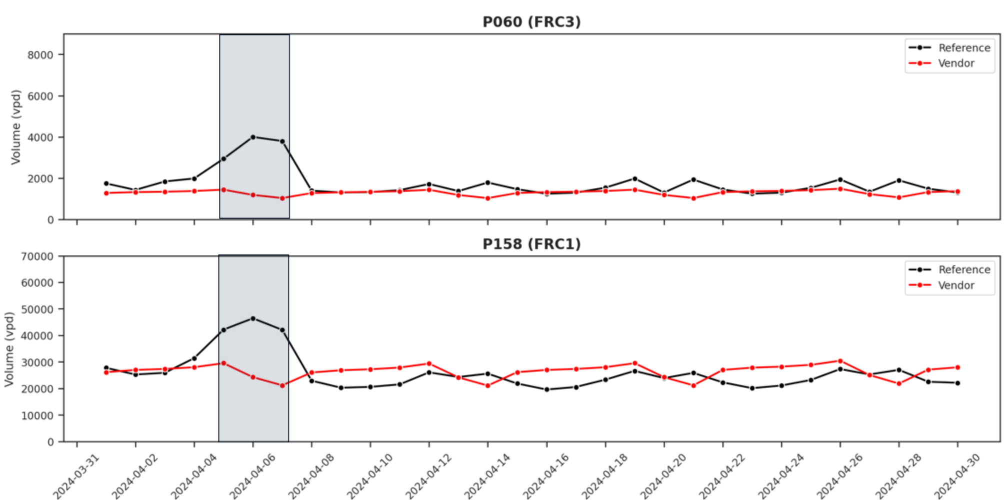


**Figure 26 – Daily Volume timeseries plot for Vendor A**

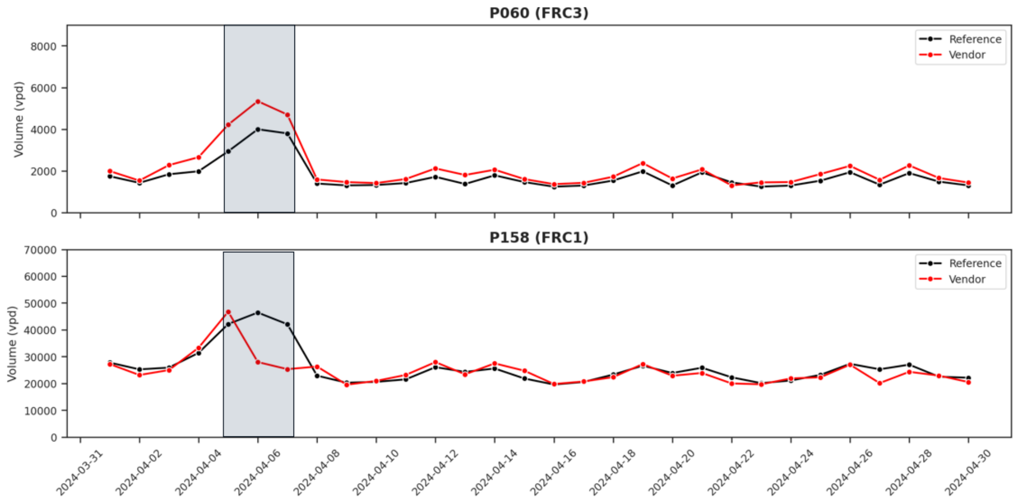


**Figure 27 – Daily Volume timeseries plot for Vendor C**

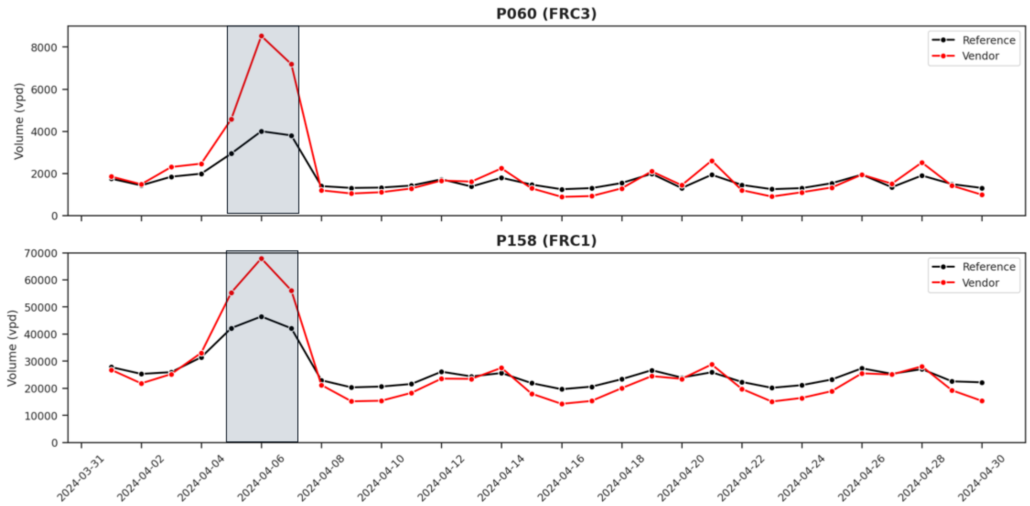


**Figure 28 – Daily Volume timeseries plot for Vendor D**

The overall performance of the three vendors for Daily Volumes is depicted in Figure 29 below, with error distributions plotted across different volume ranges, and +-15% error boundaries shown in dashed blue lines. Note that the estimate of a Daily volume generally worsens with decreasing volume as would be anticipated with a constant sampling or penetration rate. Vendor A's results are fairly consistent across volume ranges, while Vendors C and D have more noticeable biases in the lowest volume bins and Vendor D tends to overestimate in the highest bin. However, note again that these average values encompass mostly typical behavior, and do not communicate how well each vendor can identify days when volumes are atypical.

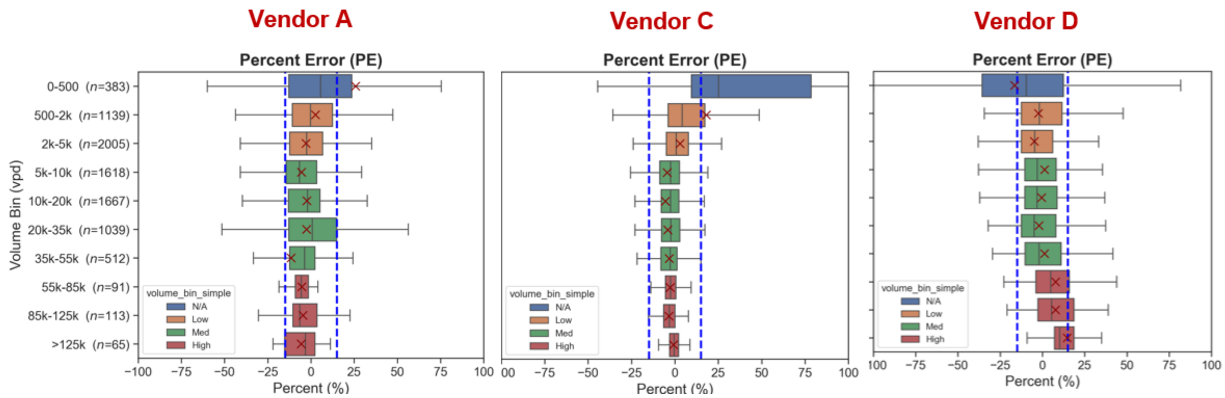


**Figure 29 - Distribution of daily volume percent error metrics by volume bin**

The above results contain both data from continuous counters and temporary counters. The number and distribution of continuous counters versus temporary count sites were shown previously in Figure 8. Note that continuous count stations tend to be located on higher class roadways (FRC 1 through 5) and less representative of lower volume roads (FRC 6 and 7). In contrast, temporary count sites are clustered in lower road classes (FRC 3 through 7). High volume roadway classes (FRC 1 and 2) are nearly exclusively covered by continuous counters, whereas extremely low volume roadways (FRC 6 and 7) are nearly exclusively reported on by temporary count sites. The common process for training machine learning (ML) to estimate volumes is dominated by data from continuous count stations. As such, *better performance would be expected in roadway classes 1 through 5, where significant continuous count data exists for training AI and ML models.*

As a test of vendors' ability to accurately report volumes on roadways beyond that which was used for training, validation performance is separated into continuous count stations (where vendors had access to historic data to train the models) and temporary count stations (where vendors likely did not have access to the data for training the models). The graph to the right in Figure 8 shows the relative size of each data set in actual data points, noting that since temporary counts only collect 48 hours of data as compared to 24x7 for continuous counts, the number of records in each data set differs substantially in size.

Figure 30 shows percent error distributions for Daily Volumes separated into continuous counters versus temporary counters. In all cases the performance on the temporary count sites is worse than on the continuous count sites, as would be expected based on the ML principles because historic data from the continuous count stations were likely included in the training data sets (though not for the time periods tested). Road classes 3 through 5 were common to both continuous count stations and temporary count stations. For Vendors A and D, the difference in performance within these classes was similar between continuous count stations and temporary count stations, though performance at temporary count sites was slightly worse (as would be anticipated). Performance for roadway classes 6 and 7 were substantially worse, suggesting that there may not have been sufficient training data in these classes for good performance. Note that the results for Vendor C were substantially worse across all roadway classes for the temporary count data, suggesting either overlearning of the training data set from continuous counters, or perhaps a processing irregularity for temporary count stations.

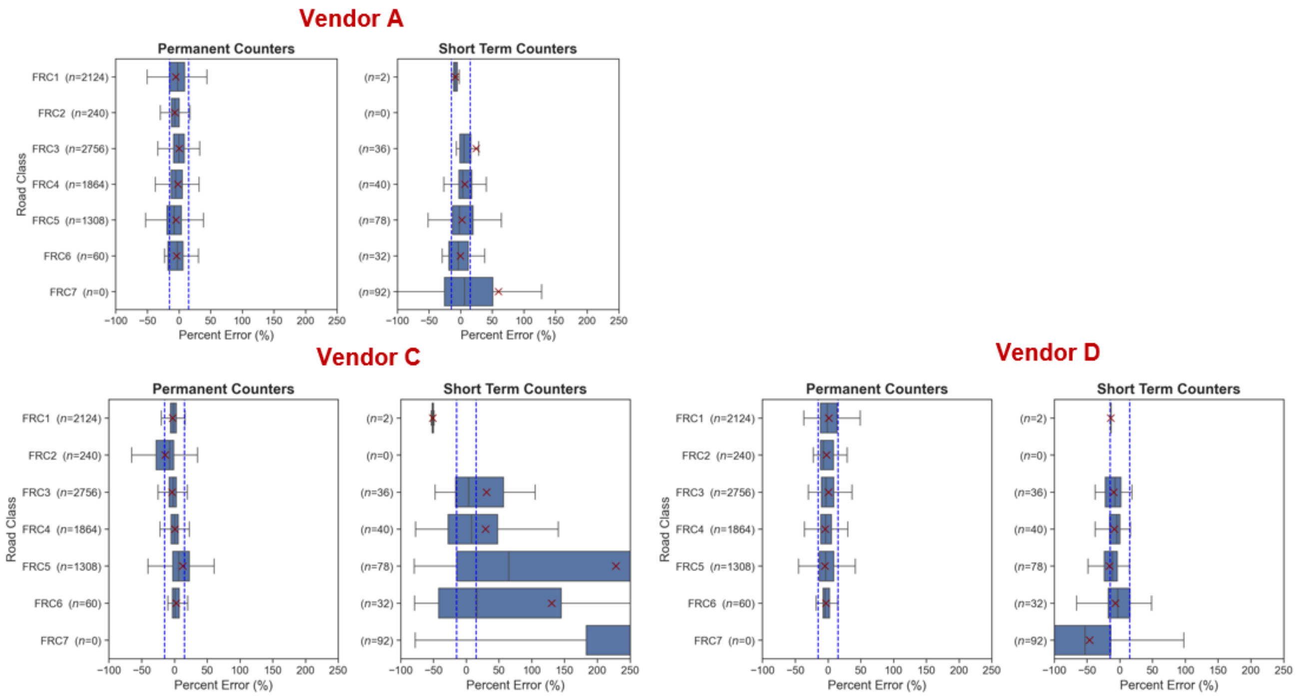


**Figure 30 - Daily Volume Percent Error for permanent vs short-term count sites**

### Daily Volumes Summary

While all three vendors produced reasonable daily estimates on typical days, vendors C and D were the only two capable of identifying unusual volume trends (e.g., accidents, unusual demand patterns). This ability to identify abnormal volumes is critical for operations purposes and motivates the need for this data element – otherwise a simple daily average could be used instead.

Additionally, all vendors were most accurate when evaluated at continuous count sites (whose historical count data was used for model calibration) and less accurate at short-term count sites that had not been previously used for model training, as expected. However, the degradation in accuracy was particularly pronounced for Vendor C and noticeable across several different road classes.

## Hourly Volumes

Hourly volumes represent the most temporally granular data item evaluated. Unlike AHDT – the other data element that focuses on hourly periods – hourly volumes allow an analyst to investigate performance during anomalous dates and periods. Manual inspection of time series plots, such as shown in Figure 31, proved most useful for gaining insight into how well vendors tracked reference data across time.

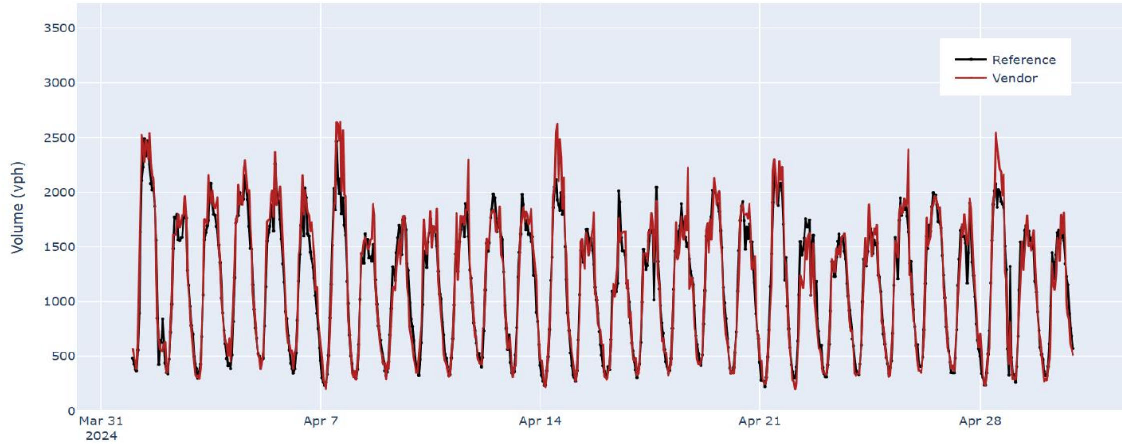


**Figure 31 – Sample hourly time series plot**

Operations applications are concerned primarily with atypical traffic conditions. If traffic were typical, then AHDT for the specific hour and day of week would result in reasonable estimation of real-time hourly volumes, a common method for estimating traffic for specific day and time in lieu of real-time volume data. On roadways without AHDT, the AADT can be factored to hourly estimates based on time-of-day and day of week template using nearby CCS stations in the region for a similar road class. A better test of the quality of hourly real-time volume data for TSMO applications is to observe performance during an atypical period or event. The validation methodology consisted of searching the reference data for atypical roadway volumes and observing the volume estimates from the vendors during those periods.

One such example is location P060, an FRC 3 rural highway in south-east GA located at a GDOT CCS location that experienced a spike in volumes over the weekend of April 5-7. Figure 32 shows the location of P060 relative to Columbus GA and provides a satellite view of the roadway and surrounding area.

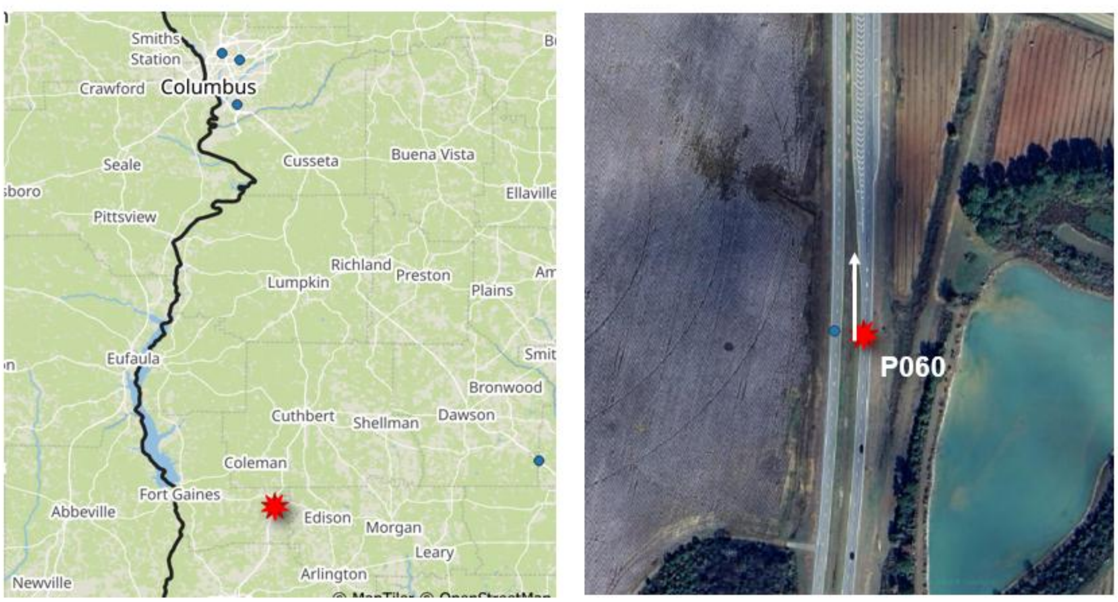


**Figure 32 – Location of atypical volume demand from P060 on April 5-7**

Figures 33-35 show time series plots for each of the vendors. Figure 33 highlights that Vendor A did not register any unusual traffic volumes during the April 5-7 time period. A close inspection of the vendor-reported data indicates that similar, if not identical, volume estimates were reported for April 5-7 as for April 12-14. This was consistent with other locations with anomalous volumes, indicating that their volume feed was not sensitive to atypical volume demand but rather employed either AHDT or a template method for estimating daily traffic. The performance for Vendor C is depicted in Figure 34. Note that Vendor C was able to capture the anomalous event, with relatively high accuracy, reflecting the ability to capture fluctuations in traffic for TSMO applications. Similarly, Vendor D, shown in Figure 35, also was able to detect the anomalous volume fluctuation over this period, though the magnitude of the reported volumes were approximately twice that of the reference data from the CCS.

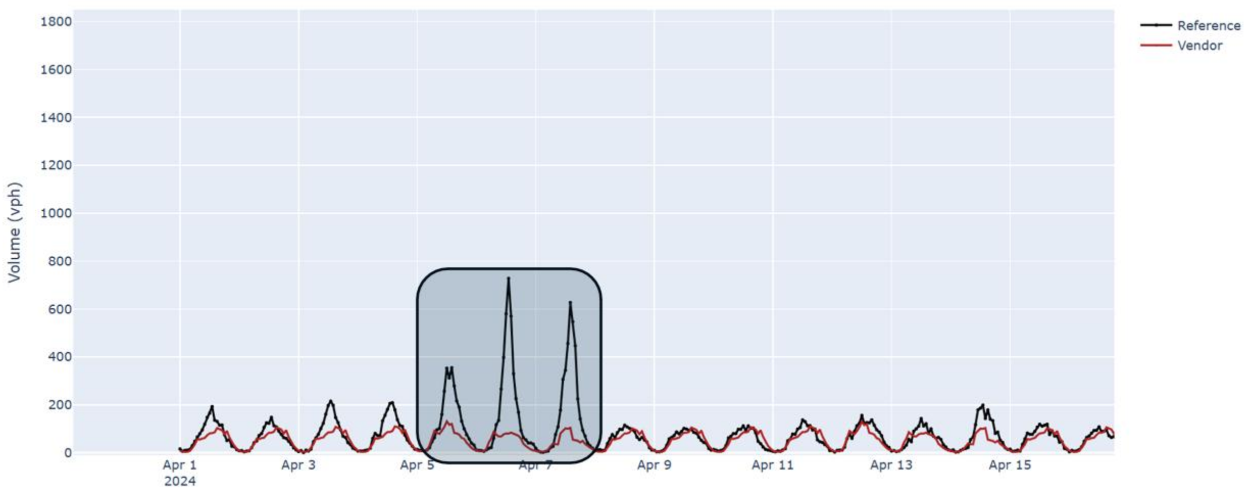


**Figure 33 – Time series plot at P060 during April ’24 for Vendor A**

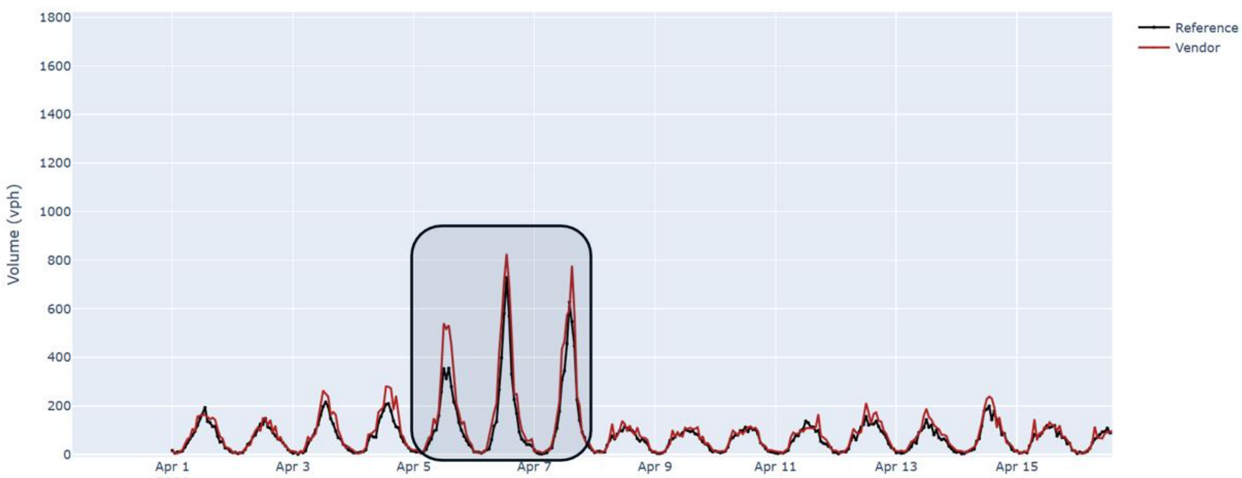


**Figure 34 – Time series plot at P060 during April ’24 for Vendor C**

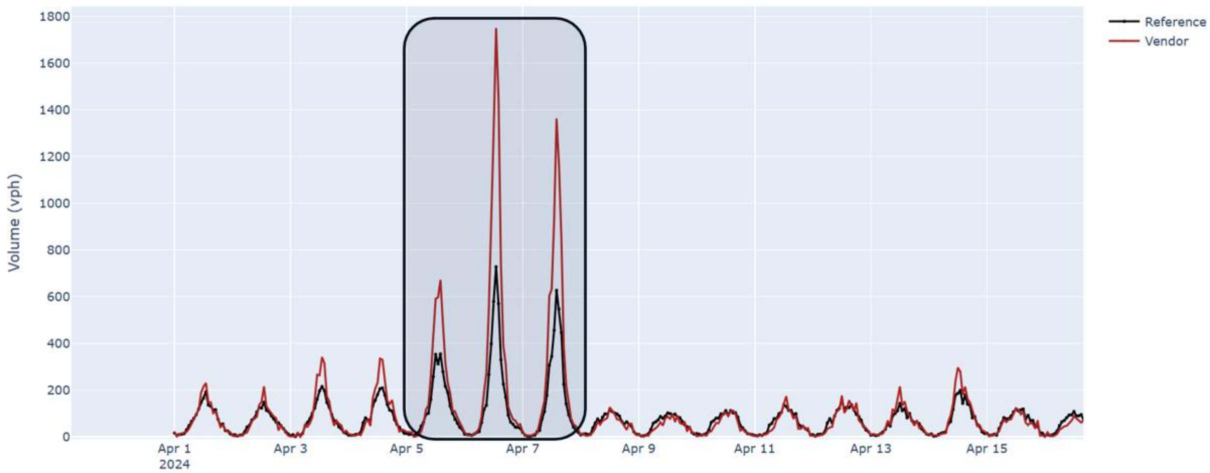


**Figure 35 – Time series plot at P060 during April '24 for Vendor D**

A second example of an unusual traffic volume event is depicted in the figures below. This example is from the same time period as the previous example but focuses on a higher volume freeway location. Figure 36 shows this location, P158, which captures NB traffic along I-75 just North of the Georgia/Florida state line.

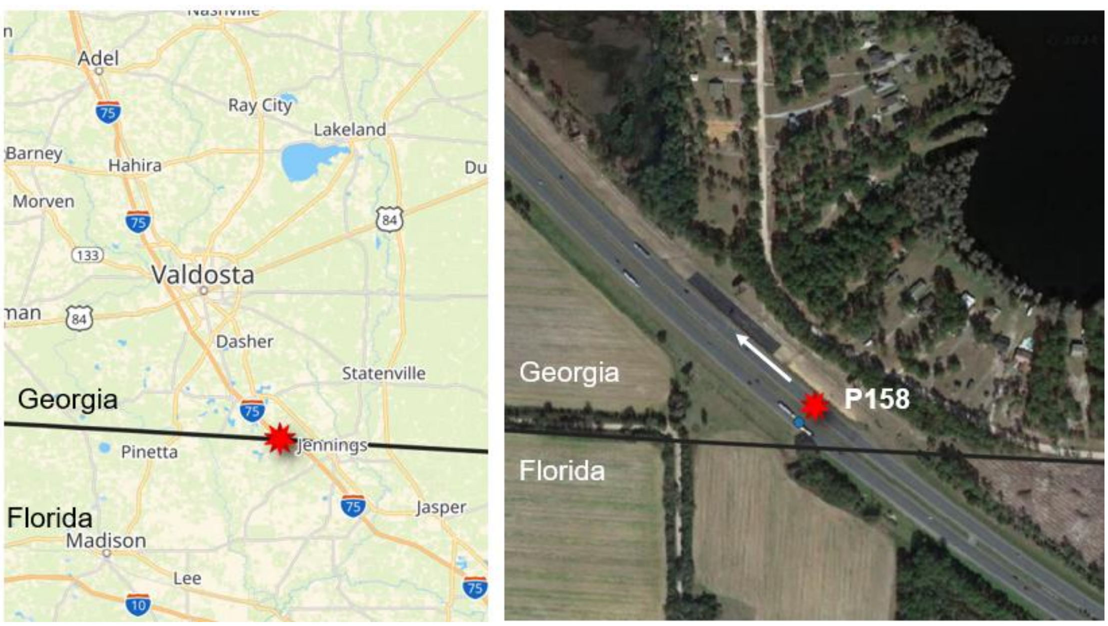


**Figure 36 – Location of atypical volume demand from P060 on April 5-7**

The volumes reported from Vendors A, C and D are provided in Figures 37-39. Similar trends are reflected as were observed in the first example; Vendor A (Figure 37) was not sensitive to any fluctuations in traffic volumes, Vendor C (Figure 38) accurately captured the first day of the heightened volume (April 5) – though failed to capture any unusual volume on April 6 and 7, and Vendor D (Figure 38) reflected the unusual volume event, but over-estimated its magnitude.

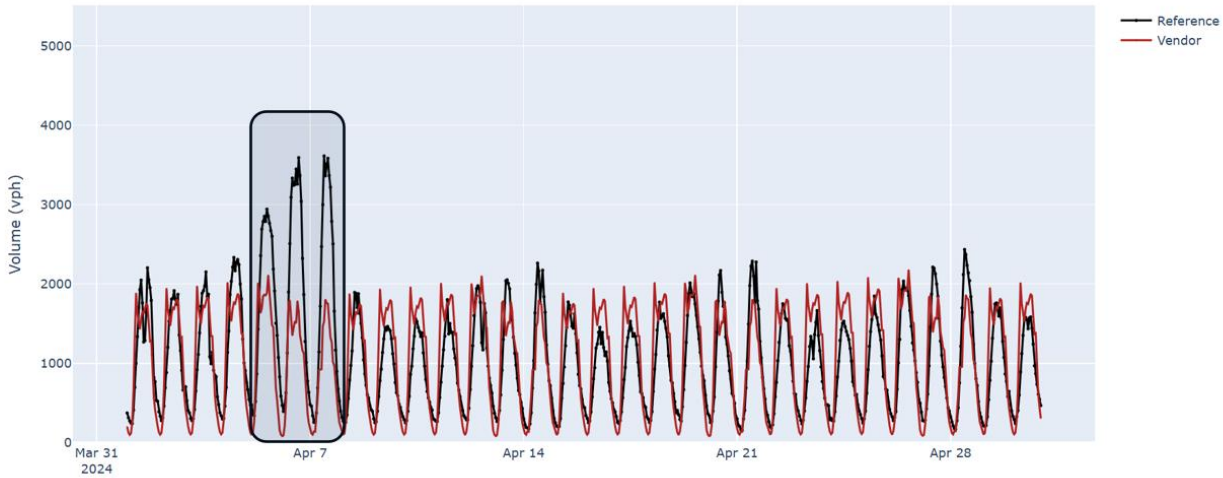


**Figure 37 – Time series plot at P158 during April ’24 for Vendor A**

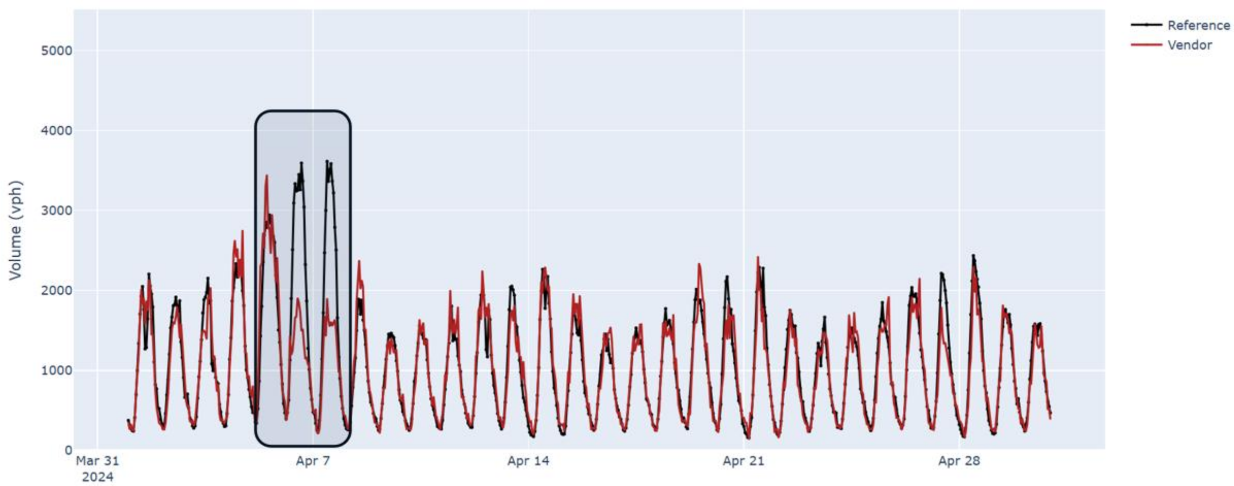


**Figure 38 – Time series plot at P158 during April ’24 for Vendor C**

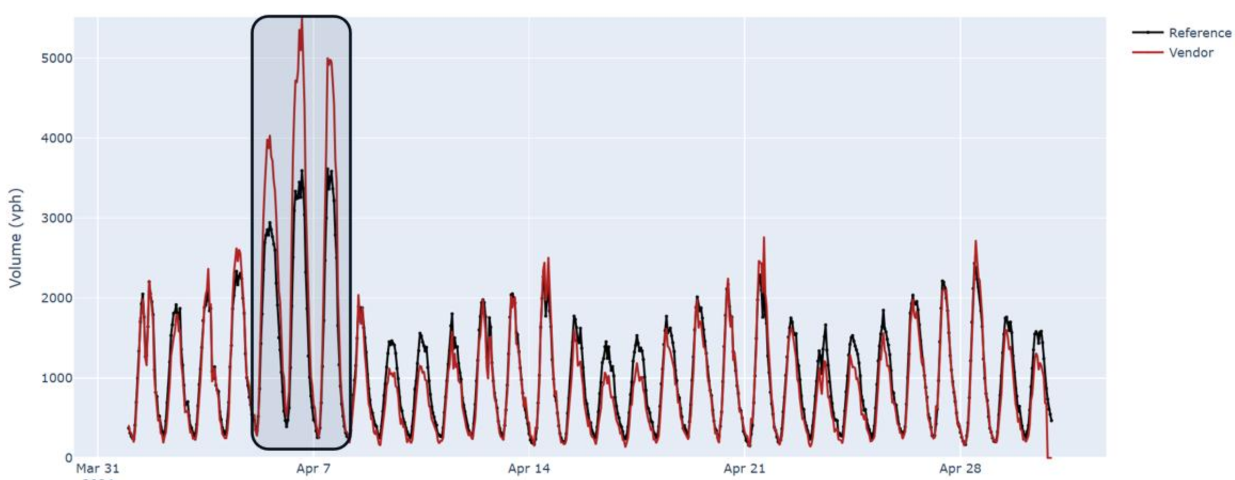


**Figure 39 – Time series plot at P158 during April ’24 for Vendor D**

A third, and last example, is depicted below, coming from a 48-hour STC site on a low volume roadway – shown below in Figure 40. This location, P697, does not include an anomalous event (as did the previous two examples), but captures a 48-hr period at a location not included in the calibration dataset.

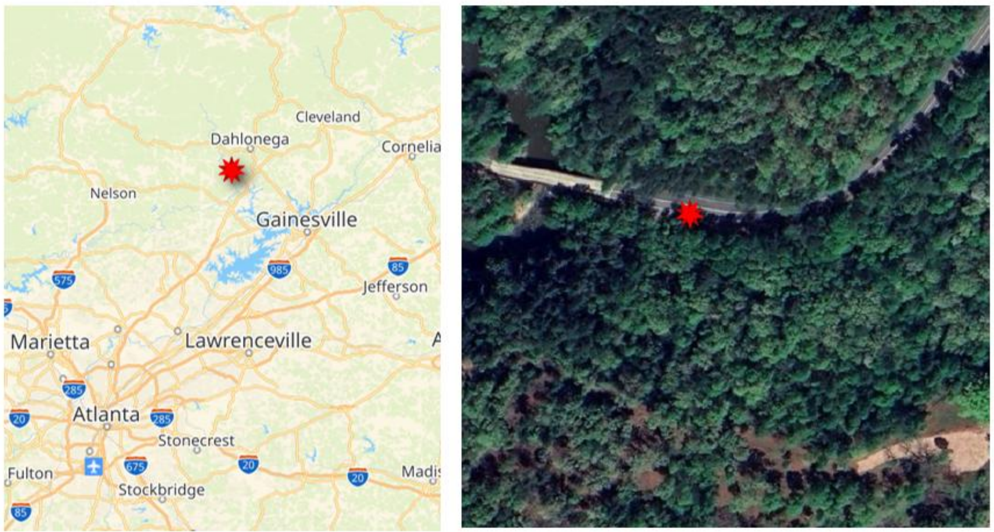


**Figure 40 – Location of short-term count site P697**

The results from vendors A, C and D are shown in succession in Figures 41-43. Vendors A and D reflect the hourly volume to a reasonable degree of accuracy, whereas Vendor C reported volumes that were 2 to 3 times the reference volumes. These results are consistent with the Daily Volume accuracy results for low-volume, short-term counters.

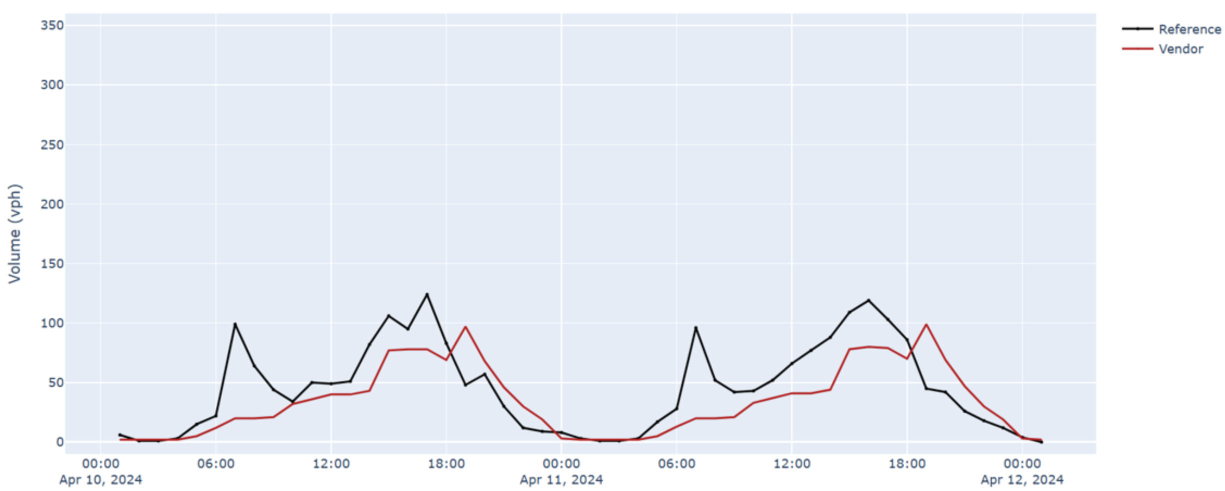


**Figure 41 – Time series plot at short-term count location P697 for Vendor A**

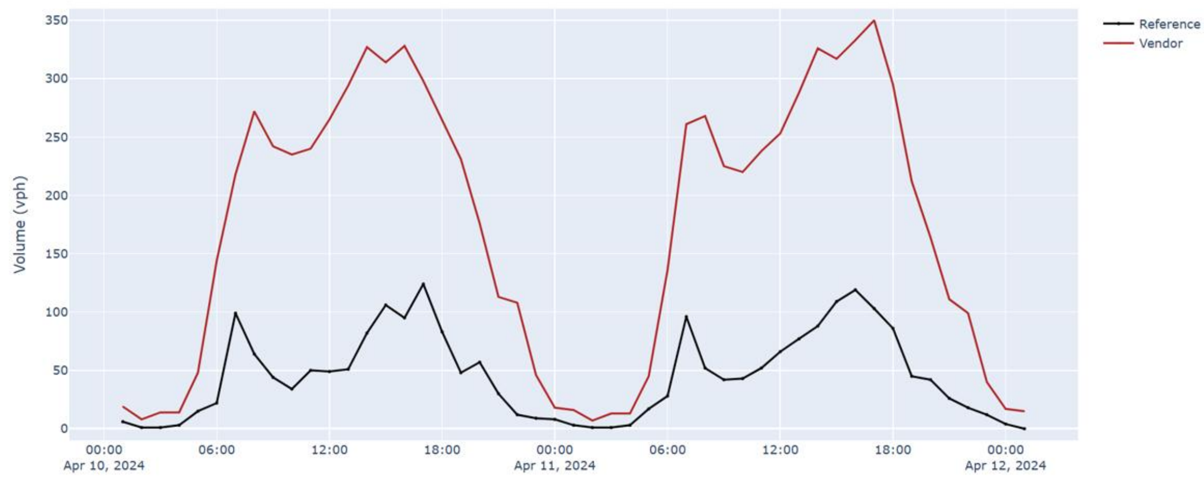


**Figure 42 – Time series plot at short-term count location P697 for Vendor C**

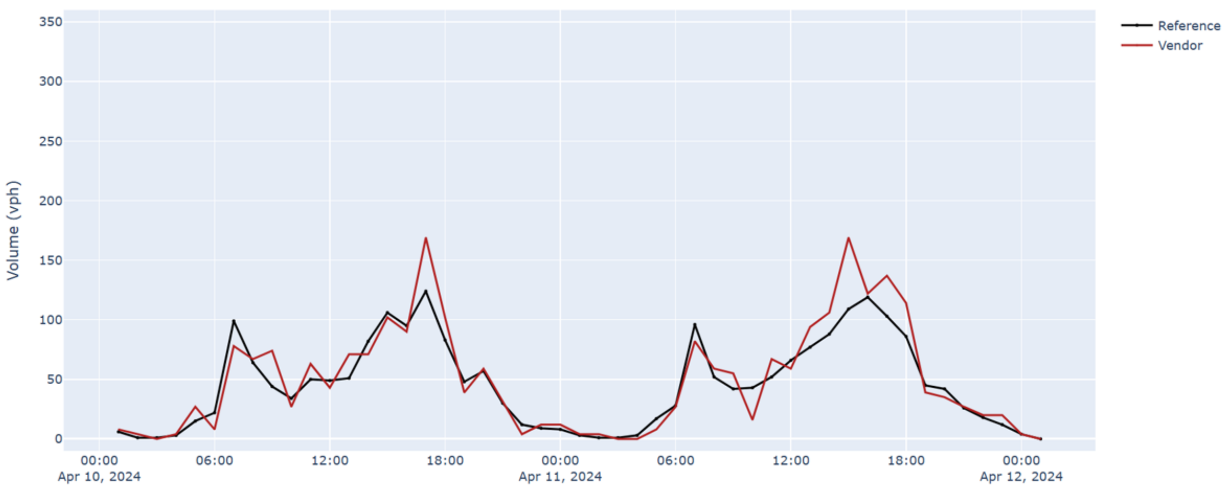


**Figure 43 – Time series plot at short-term count location P697 for Vendor D**

The statistical error distributions for the Hourly Volume validation for the three vendors are depicted in the graphs below, starting first with overall accuracy (Figure 44) and then segregated into CCS and STC locations (Figure 45). The trends in the CCS vs STC plots match findings at the Daily Volume level; vendors A and D are marginally less accurate on STC locations than the CCS locations (as expected), with Vendor C's substantially worse on STC locations across all roadway classes.

Again, it is worth noting that while accuracy distributions are helpful for assessing data quality – especially when breaking down by CCS vs STC locations – the majority of observations in the dataset represent normal traffic conditions. As such, it is important to consider these overall results within the context of each vendor's ability to capture date-specific anomalies. At the present time, Vendors C and D are the only two that are sensitive to such patterns.

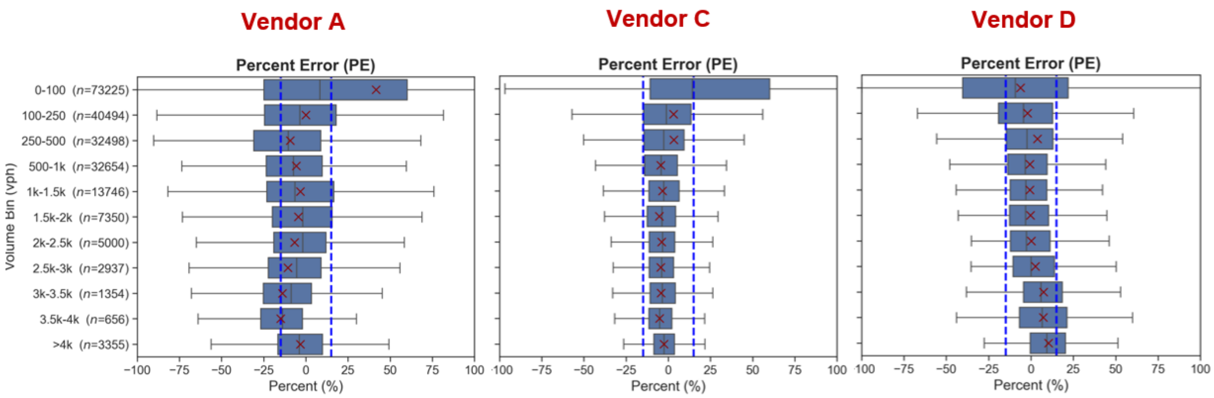


**Figure 44 – Distribution of hourly volume percent error metrics by volume bin**

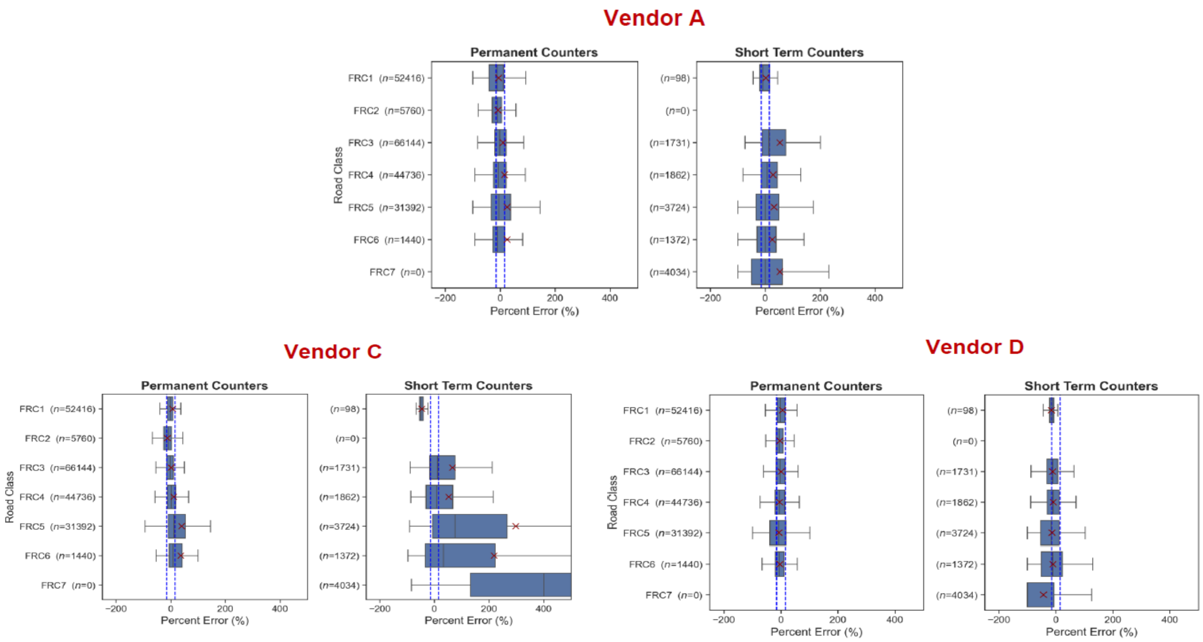


**Figure 45 – Hourly volume percent error for permanent vs short-term count sites**

## Hourly Volumes Summary

The results for Hourly volumes are similar to Daily volumes above, as both emphasize operations applications and vendors' ability to capture anomalous volumes on specific dates. While all three vendors produced reasonable hourly estimates on typical days, vendors C and D were the only two capable of identifying unusual volume trends (e.g., crashes, special events, and other unusual demand patterns). In the prior validation in North Carolina, Vendor C was the only vendor capable of producing such estimates, so Vendor D's introduction of a meaningful hourly volume product is notable industry progress.

Like with Daily volumes, all vendors were most accurate when evaluated at continuous count sites (whose historical count data was used for model calibration) and less accurate at short-term count sites that had not been previously used for model training, as expected. However, the degradation in accuracy was particularly pronounced for Vendor C and noticeable across several different road classes.

# Conclusions

This report explores the fidelity of volume estimates derived from probe data across five levels of temporal aggregation (AADT, ADT, AHDT, Daily, and Hourly), focusing on commercial products from five vendors who sell data in TETC's Transportation Data Marketplace (TDM). This study focuses on data from Georgia in April 2024 (for all data elements except AADT, which used data from 2023) for locations across Georgia. The following conclusions emerge from analysis of data across temporal aggregations and vendors:

- **Industry has progressed over the past year.** Despite multiple vendors weathering data supply chain issues, this study showed that probe-based volume estimation products have matured in the past year. Several vendors are on the verge of meeting FHWA's accuracy/precision benchmarks for AADT – a key use case for the data, and there has been growth in the quality and availability of more granular data elements. Specific improvements in vendor product offerings include the following:
    - All vendors are now capable of reporting volumes on all road classes (FRC 1-7). In the prior validation, Vendor C reported only FRC1-5, but has now added coverage for FRC 6-7 (typically low-volume and often rural facilities).
    - All three vendors who submitted AHDT data are now able to regularly capture average day-of-week and time-of-day volume patterns. In the prior validation, Vendor A struggled with differentiating between weekend versus weekday patterns for AHDT, but their product has noticeably improved in this regard.
    - Two vendors (Vendors C and D) are now capable of producing meaningful daily and hourly volume estimates that are sensitive to atypical volumes. In the prior validation, only Vendor C had the ability to do so, but over the past year Vendor D introduced a meaningful hourly/daily product
- **There is measurable separation among the vendors**. Details differ across the different temporal aggregations, but several vendors stand out as being particularly strong in specific areas. At the planning level, three vendors are getting close to FHWA benchmarks for precision/accuracy (A,C,D), with vendor D most consistently close to the benchmark regardless of whether volumes are treated directionally or bi-directionally. These vendors also demonstrate the ability to capture average day-of-week and time-of-day patterns (AHDT). From an operational perspective, vendors C and D stand out as having meaningful Daily and Hourly volume products that can detect abnormal volume patterns.
- **Several vendors are on the verge of meeting FHWA guidelines for AADT accuracy/precision**, **although the results differ slightly based on whether tests are run on directional or bi-directional volumes**. The validation team intends to make the FHWA test a core part of future analyses and plans to seek further clarification on how to handle volume directionality.
- **There are multiple sources of error and uncertainty that impact the accuracy of results.** In addition to estimation error as a byproduct of each vendors' modeling processes, conflation/geo-referencing issues can lead to erroneous volumes when validation locations are translated incorrectly between map representations and/or geo-referencing protocols. These georeferencing processes also require properly interpreting roadway directionality, and if handled improperly, can lead to significantly over or underestimating volumes. Additionally, the reference data from both continuous and short-term counters require careful inspection to deal with abnormalities and missing

data that can impact annual average calculations. As vendor product offerings continue to mature, it will be particularly important to monitor and quantify the different sources of error.

As of early 2025, vendor data supply chain issues appear to be resolved, and the validation team anticipates evaluating all vendors' updated product offerings in 2025. The next study, scheduled for May 2025 in Massachusetts, will be patterned after this validation activity, prioritizing analysis of multiple temporal aggregations from all vendors. Given state DOT interest in AADT, the next study will continue to emphasize aggregate data elements and include further integration of FHWA's evaluation framework, while continuing to assess the viability of more granular products for operational purposes. Additionally, the 2025 validation effort will incorporate and assess self-reported cross-validation audits, that is a vendor's own estimate of their anticipated accuracy of their volume model.