TDM Validation Activity: Massachusetts Volume Study Part I (AADT)

TETC Transportation Data Marketplace Data Validation

Prepared by: TETC Validation Team

7/31/2026

TDM Validation Activity: Massachusetts Volume Study Part I (AADT)

TETC Transportation Data Marketplace Data Validation

Validation activity for TDM Validation program, focusing on 2024 AADT Volume data in Massachusetts.

TDM-VAL-11

Data Categories*: Volume*

Evaluation Period*: 2024 AADT*

Publication Date: *July 2026*

7/31/2026

The Eastern Transportation Coalition is a partnership of 20 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

The Eastern Transportation Coalition (TETC) has performed validation of industry volume measures as part of its Transportation Data Marketplace (TDM). Previous validation efforts include analyses conducted in collaboration with North Carolina Department of Transportation (DOT) and Georgia DOT in 2023 and 2024, respectively. The current validation effort was executed in collaboration with the Massachusetts DOT, with data collection in 2025.

Five TDM vendors – Bentley (HERE), INRIX, Iteris, StreetLight, and Replica – delivered volume data at five temporal aggregations: *annual average daily traffic (AADT), average daily traffic (ADT), average hourly daily traffic (AHDT), daily volumes, and hourly volumes* at over 1000 evaluation locations across the state of Massachusetts. 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.

The initial intent of this study was to characterize vendor performance across all five temporal aggregations, using metrics, statistical tests, and visualizations consistent with prior studies in GA and NC. However, initial AADT-level analysis uncovered several data issues that warranted further investigation. As such, the analysis was split across two reports: **This report focuses exclusively on analysis at the AADT level**, with other temporal aggregations saved for Part II.

A key part of the AADT analysis includes applying USDOT's updated statistical method for assessing the accuracy of non-traditional volume sources for AADT. Compliance with this method serves as an important benchmark for states seeking to use probe-based volume products for federal reporting purposes. Prior reports included a variation of this method, but FHWA's guidance on this topic has been updated within the last year and the validation team has improved the implementation of this method within its validation procedures.

Initial analysis of AADT volume estimates revealed high errors and major outliers, which were unexpected based on vendor performance in prior validations. As such, the validation team spent considerable effort tracking down the cause of these discrepancies – a process that uncovered several different sources of error (beyond just vendor inaccuracy). These errors are important to understand because they can produce a disproportionate impact on resulting error metrics and distort apparent performance – including the FHWA statistical test. Because some of the erroneous data was included in the calibration dataset, vendors were allowed to resubmit results, and all 'suspect' locations were excluded from the final analysis. Several observations and takeaways emerged from the resulting evaluation:

- **Agency counters and deployment locations are more diverse in Massachusetts than in prior validations**. Like prior validation studies, Massachusetts primarily uses inductive loop counters for its core Continuous Count Station (CCS) network. However, several of these sites are deployed near merge/diverge points, which can add ambiguity around which lanes are actually counted. Additionally, MassDOT has a sizeable inventory of intersection cameras deployed at signal-controlled intersections, intended primarily for traffic detection/control, but also used for counts. Relative to prior validations, *the inclusion of cameras is the most notable difference,* as it represents a new type of counting technology and deployment context.
- **Many sources of error and uncertainty can impact resulting accuracy measures, including several unrelated to vendor data quality.** Although some outliers are attributed to vendor estimation error, other sources of error were equally or more impactful to the results. Examples of such errors encountered in this study include (I) miscoded site locations (a data entry issue in the DOT database), (II) ambiguity in

georeferencing, and (III) reference data error – none of which are reflective of vendor data quality.

Issues associated with ambiguity in georeferencing/conflation are particularly challenging because both agency counts and vendor estimates may be correct within their geographic context but compare unfavorably due to incorrect or ambiguous georeferencing. For this study, CATT Works Georeferencing Protocol (CWGP) location references were used by the validation team to communicate evaluation locations to vendors. In certain complex locations (e.g., at merge/diverge points or arterial intersections with slip lanes), simply knowing the latitude, longitude, and direction of travel was not enough to interpret which lanes, or how many lanes were included in the counts. It is recommended that future validations include number of lanes in georeferencing as an added data quality check.

- **Camera counters at signalized intersections should be treated with caution**. In Massachusetts, cameras are used primarily for traffic control purposes and collect turning movement counts, from which directional volumes on each intersection leg can be derived. Although many intersection sites appeared to produce reasonable segment volumes and the DOT generally trusts the outputs, the data quality is less validated than traditional, inductive loop sensors. Generally, errors at camera locations were greater than those from inductive loop count sites. Whether these errors arise from fundamental technology issues or from the complex environment at intersections where Cameras are typically deployed -- or a combination of the two – is currently unknown.

  In this report, the validation team reports results both with and without camera counters included. Results from locations excluding Camera based technology are deemed most comparable to previous validation results from GA and NC.

- **Most vendors showed high accuracy after accounting for data issues, particularly when focusing on non-camera locations**. After removing the suspect data and processing vendors' final submissions, four out of the five vendors produced directional volume estimates that generally matched reference AADT values, particularly when excluding camera locations. Performance was evaluated under a variety of conditions – including across volume ranges and by functional road class (see Section 6 for full results).

- **Several vendors performed well on the FHWA statistical test, showing industry progression relative to last year's study.** The validation team re-implemented the FHWA statistical tests for AADT volume accuracy based on updated guidance, focusing on bi-directional volumes. For comparative purposes, last year's results from Georga were regenerated with the same methodology to serve as a point of comparison.

  The FHWA test is particularly important for agencies seeking to use AADT volume estimates for Federal reporting purposes. In last year's Georgia study, two vendors were close to meeting the FHWA guidelines for bi-directional AADT values. This year, when evaluating accuracy against continuous counters (but excluding camera locations), which is contextually comparable to prior studies, ***two vendors passed all components of the FHWA statistical test***, with two others approaching (nearly passing) the benchmark.

Ultimately, after accounting for the many sources of error, most vendor products *do* show strong agreement with reference counts – particularly when focusing on non-camera locations. While this is an encouraging result, the data issues themselves are a key part of the story. Many of the statistics used to evaluate vendors' performance – either visualizations, error metrics, or the FHWA test – are highly sensitive to outliers. Understanding the sources of uncertainty and learning to prevent or identify these issues early is critical for structuring future validation activities and fairly evaluating vendor products.

Furthermore, the findings of this validation are additionally applicable not only to validation procedures, but also to any applications of outsourced volume data. The various types of errors and the sources identified in the validation effort would also be relevant for any application that required industry sourced data to be conflated to an agency base-map. For this reason, Coalition members are encouraged to review the findings of this report not only with respect to accuracy of AADT estimates from vendors, but also to understand various error sources relevant to any application of outsourced vendor data.

Table of Contents

1. 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 with respect to the original Vehicle Probe Project; during the first two phases of the Vehicle Probe Project (2008 to 2022), validation activities focused on ensuring that traffic data conform to the contractual standards of travel time and speed, which had highly defined specifications 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 extends to five other core 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 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. 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 fourth Volume validation activity, and the third for which error measures are reported publicly. Prior reports provide useful context on how the volume validation process was developed and the most recent results: TDM-VAL-01 (lit review/recommendations for structuring volume validation), TDM-VAL-02 (inaugural proof of concept activity), TDM-VAL-05 (North Carolina study) and TDM-VAL-09 (Georgia 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.

2. Data Vendors

Five TDM vendors participated in the validation: Bentley (HERE), INRIX, Iteris, StreetLight, and Replica. The validation team set expectations for data submission consistent with RFP specifications, including using a specific georeferencing protocol (CATT Works Georeferencing Protocol, CWGP) to describe validation locations, and requested five volume data deliverables from the participating vendors. These deliverables included: 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.

3. Geographic Scope

Figure 1 shows the geographic scope of the volume study area, which encompasses 1044 directional roadway locations across the entire state of Massachusetts. These locations correspond to MassDOT 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 planned for May 2025 or were collected in 2024 (for AADT). These evaluation locations capture all functional road classes, ranging from Interstate roadways (class 1) through local roads (class 7). *Note that this report focuses on AADT exclusively, which includes 844 of the locations shown below (all except the 200 STC sites from May 2025, which will be included in a follow-up report).*

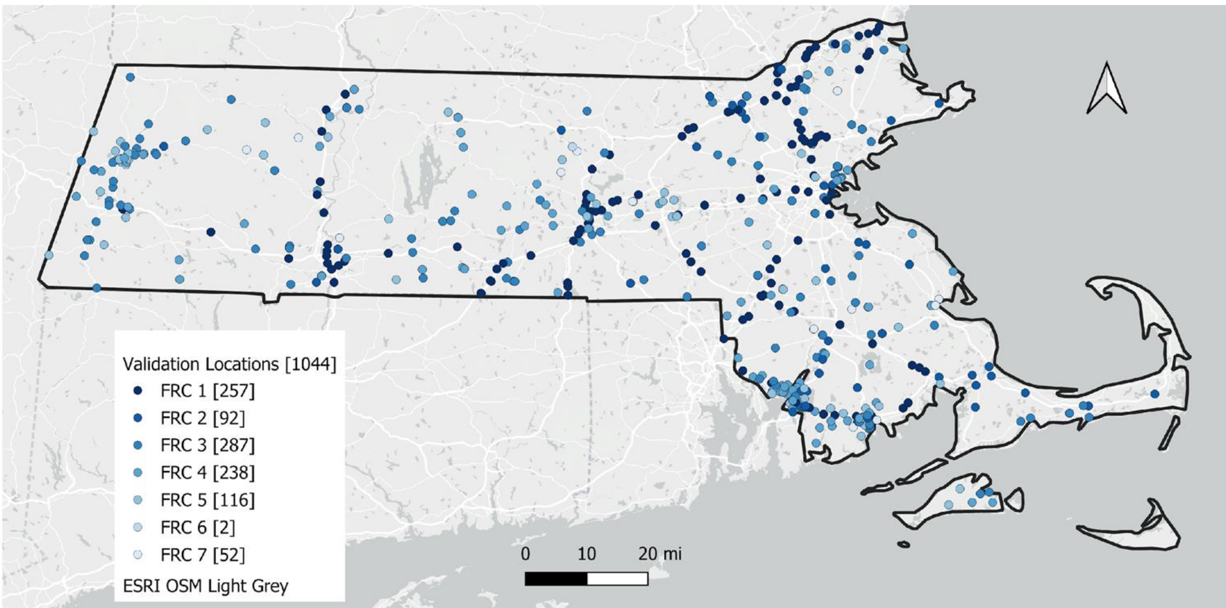


Figure 1 - Evaluation locations

Table 1 shows the key attributes used to communicate volume evaluation locations to vendors via georeferencing protocol (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	-73.202088 42.485620	7	CHESHIRE ROAD	3
P002	-73.202088 42.485615	187	CHESHIRE ROAD	3

Note, as a result of the findings of this validation, this report recommends that data used to communicate locations includes an additional field for the number of lanes and an indicator of lane type (e.g., mainline, ramp). Due to the ambiguity of complex intersections and interchanges with respect to the inclusion or exclusions of ramps and slip lanes, the ‘number of lanes’ parameter and corresponding lane type indicator would assist in precisely identifying the volume estimate requested. This will be implemented in future validation efforts and is recommended in practice for any effort in which volume data from industry is conflated to an agency base map or linear referencing system.

4. Data Description

4.1 Reference Data

Hourly traffic counts from validation locations – consisting of both continuous and short-term count sites – were obtained from Massachusetts DOT for May 2025, the primary study period. Additionally, MassDOT provided 12 months of historical count data (Jan-Dec 2024) at the continuous count sites, which were shared with vendors for model calibration purposes. They also provided ‘factored’ AADT volumes (not the underlying count records) for a subset of short-term counts collected during 2024. Together the historical volumes from 2024 were used for assessing AADT. Note that because hourly counts from 2024 were provided to vendors as calibration/training data (for the May 2025 primary study period), the 2024 AADT evaluation is not a ‘blind’ study. Nonetheless, the TAC indicated that this aggregation level was important to include as a validation exercise.

The MassDOT count data was used to create five different reference datasets (corresponding to the vendors’ deliverables), summarized in Table 2. Note, however, **that this report focuses on AADT**, with other temporal aggregations covered in the follow-up Part II effort.

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

Most aggregate

Data Item	Description	Evaluation period	Continuous Counts	Short-term counts
Annual Avg Daily Traffic (AADT)	Avg volume for a day during period	2024	✓	✓
Avg Daily Traffic (ADT)	Avg volume for a day during period	May 2025	✓	
Avg Hourly Daily Traffic (AHDT)	Avg hourly volume by day-of-week (e.g., Mondays between 8-9am)	May 2025	✓	
Daily Volume	Volume for a specific day during period	May 2025	✓	✓
Hourly Volume	Volume for a specific hourly period (e.g., 05/01/2025 between 8-9am)	May 2025	✓	✓

Most granular

5

Two categories of reference AADT data were obtained: (1) AADT values directly calculated by aggregating the hourly count records from 2024 at continuous count sites and (2) factored short-term counts provided by MassDOT for short-term count sites.

- **AADT Calculated at Continuous Count Sites (CCS)**

  First, hourly CCS counts were subject to MassDOT's internal QC/QA process. Afterwards, CCS locations were filtered based on a 75% data availability requirement; only locations with at least 75% of hourly records present during the study period (2024) were used for analysis. This rule was imposed for consistency with prior validations, and to prevent locations with a small amount of data from being used as representative 'ground truth' averages over the year. After applying this filter, 486 directional CCS locations were kept for analysis, with 131 omitted.

  After filtering, hourly counts were aggregated to AADT volumes using the method detailed on page 3-14 of the 2022 Traffic Monitoring Guide. This approach 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. It produces identical results to a simple arithmetic average of total counts for each day in the year when there is no missing data.

  It should be emphasized that these AADT locations are all *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; 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.

  <u>Note1</u>: During the course of the analysis, several additional 'suspect' data locations (that were originally included in the Reference dataset) were uncovered. Understanding the nature of these suspect data points is central to this report. How suspect data points were identified and excluded will be discussed later in the report.

- **AADT from Factored Short-Term Counts**

  Directional AADT estimates based on factored, short-term counts were provided by MassDOT; no processing was done by the TETC validation team.

Figure 2 shows the distribution of 844 directional locations by (a) volume and (b) road class, with values reported separately for Continuous (blue) and Short-Term (orange) sites. These figures highlight that CCS locations are distributed across a wider range of volumes than STC locations, which tend to be more concentrated at lower volumes. Both site types have representation across most road classes, with about half of CCS locations on interstates (FRC 1). In this case, there are no sites classified as FRC 6 (minor collectors), but all other functional classes are represented.

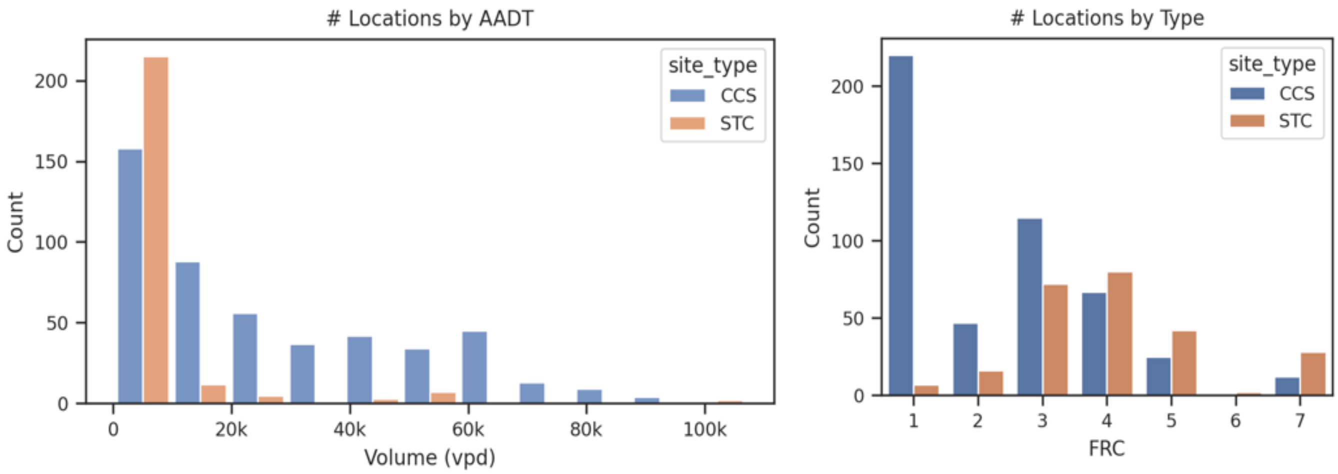


Figure 2 – Distribution of AADT volumes across volume (left) and functional class (right).

Count Location Classification

In contrast to prior validations, MassDOT's continuous count sites encompass multiple types of counters, which include: (1) traditional inductive loop counters on public roadways, (2) inductive loops on automated electronic tolling (AET) facilities, and (3) cameras, typically deployed at signalized intersections. *Note that both Traditional and AET count locations use inductive loop technology and thus are identical in many respects; however, toll counters are categorized separately here because they are sometimes deployed near merge/diverge points, which has the potential to add an additional source of ambiguity with respect to location and number of lanes, as will be demonstrated later.*

Relative to prior validations, **the inclusion of cameras is the most notable difference**, as it represents a new type of counting technology (i.e., counts automatically extracted from a video feed) and monitors traffic volumes at more complicated locations (multiple legs of intersection).

Figure 3 summarizes the number of locations that fall into each of these CCS sub-categories, plus locations where AADT was obtained at short-term count sites. Subsequent analysis in the Results section will investigate vendor performance at these different count site classifications.

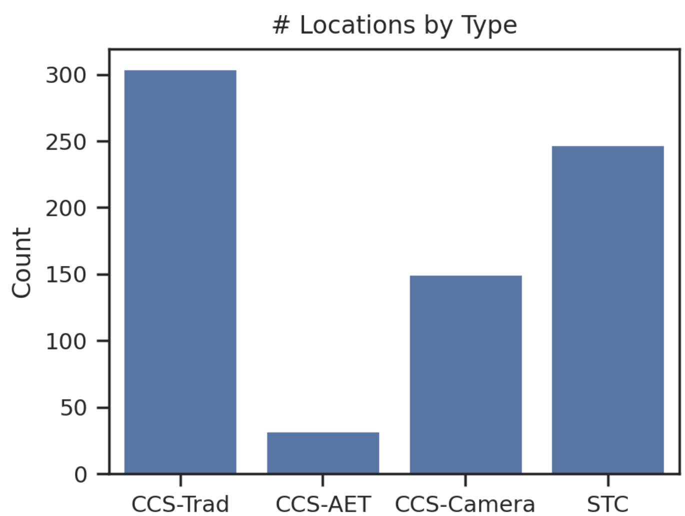


Figure 3 – Distribution of locations by counter classification

4.2 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. Additionally, vendors were also asked to optionally provide "cross-validation" results – an estimate of their model performance based on internal training and testing routines.

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 and highlights that all vendors submitted AADT estimates – the subject of this report.

Table 3 – Vendor Submission Summary

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

Note 1: *During the course of the study, several data anomalies were uncovered in the reference data (see Results section). Because these issues also impacted the calibration data that was made available to vendors ahead of time, the validation team – in consultation with the TAC – allowed vendors to re-submit their volume products. The results section will discuss this further and show performance for both data submittals.*

Note 2: *Historical 2024 hourly counts at CCS locations were provided to vendors for May 2025 model calibration purposes (for ADT, AHDT, Daily, and Hourly products). As such, the 2024 AADT analysis is not a truly 'blind' study, as the vendors had access to 2024 count information both from the provided calibration data as well as published through the FHWA TMAS system. However, Vendor E (which only submitted AADT data) noted that they did not use the 2024 MassDOT calibration data when producing AADT volume estimates. Additionally, Vendors A-D submitted volume data for the May 2025 study period, so as a sanity check, each vendor's 2024 AADT performance was compared to May 2025. Data quality was found to be reasonably consistent across aggregations, indicating that 2024 AADT results are most likely representative of industry capability. It is recommended that future validation activities try to prevent calibration data from aligning with the AADT study period.*

5. Evaluation Methodology

Vendor AADT submissions are compared to reference data (agency counts) through multiple approaches, including error metrics, FHWA's statistical test for evaluating non-traditional volume sources, and visual inspection. Note that although these comparisons generally treat the agency count data as the 'ground truth', the reference source also has inherent uncertainty that should be acknowledged. As such, major discrepancies were manually investigated and checked across multiple vendors.

5.1 Error Metrics (Directional AADT)

Table 4 summarizes the validation metrics used to quantify the discrepancy between the values reported by vendors and the 'reference' source. Given a paired dataset, the error at each location is calculated as the difference between vendor and reference volumes and may be expressed in the same units as the reported volume (vehicles) or as a percentage of the reference volume (percent error, often abbreviated as PE).

The mean and various percentiles of errors 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 communicates how much the values vary.

In most cases, results are typically grouped 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.

Table 4: Error metrics

Given a set of *n* volume estimates ($\hat{\boldsymbol{Y}}$) and corresponding reference volumes ($\boldsymbol{Y}$), the resulting signed errors, (***E***) are used as the basis for calculating various error metrics.

$$\mathbf{E} = \{e_1, e_2, \ldots, e_n\} = \{(\hat{y}_1 - y_1), (\hat{y}_2 - y_2), \ldots, (\hat{y}_n - y_n)\}$$

Metric	Formula	Explanation & Notes
Error Percentiles	$\mathrm{P}_X(\{e_1, e_2, \ldots, e_n\})$	The Xth percentile error represents the *signed* error value below which X% of observations lie (expressed as volume). • P_{50} is the median error (**MdE**) • P_{25} to P_{75} bracket middle 50% of observations
Percent Error Percentiles	$\mathrm{P}_X\left(100 \cdot \left\{\frac{e_1}{y_1}, \frac{e_2}{y_2}, \ldots, \frac{e_n}{y_n}\right\}\right)$	Similar to **Error Percentiles**, but errors are expressed as a percentage of reference volume. • P_{50} is the median error (**MdPE)**
Absolute Percent Error (APE) Percentiles	$\mathrm{P}_X\left(100 \cdot \left\{\left\lvert\frac{e_1}{y_1}\right\rvert, \left\lvert\frac{e_2}{y_2}\right\rvert, \ldots, \left\lvert\frac{e_n}{y_n}\right\rvert\right\}\right)$	Similar to **Error Percentiles**, but only the magnitude (not sign) of errors is considered, and errors are expressed as a percentage of reference volume. All values are positive. • P_{50} is the median error (**MdAPE)** • P_{90} represents 'tail risk'
Mean Absolute Percent Error (MAPE)	$\frac{100}{n}\sum_{i=1}^{n}\left\lvert\frac{e_i}{y_i}\right\rvert$	Avg. magnitude of error (as percent of reference volume)
Normalized Root Mean Square Error (NRMSE)	$100 \cdot \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2}}{\frac{1}{n}\sum_{i=1}^{n} y_i}$	Expresses avg. magnitude of error (as percent of average reference volume) but penalizes large errors more than MAPE.

5.2 FHWA Method (Bi-Directional AADT)

FHWA's report, titled "Guidelines for Obtaining AADT Estimates from Non-Traditional Sources, Revised" (FHWA-HPL-23-018), outlines a framework for evaluating the quality of "Non-Traditional" volume data sources (such as the probe-based volume estimates included in this study). The FHWA's method includes a series of statistical tests that can be used to determine whether emerging probe-based volume estimates are suitable for Federal reporting purposes. Given agency interest in this particular use case, the validation team implemented this evaluation approach method to gain additional insight into vendor data quality. One thing to note about the FHWA analysis method is that it was developed and intended to be implemented on bi-directional traffic. As such, **directional volumes were aggregated to bi-directional volumes at each reference location to create a bi-directional dataset for application of the FHWA method.**

Before describing the FHWA evaluation approach, it is first helpful to understand how agencies currently estimate AADT at locations where there are not permanent counters. Aside from limited CCS sites where AADT can be computed directly, agencies typically estimate AADT by annualizing short duration counts. This process – often referred to as "factoring", involves (1) grouping similar CCS sites (often by road type and traffic characteristics), (2) developing factors to summarize how each group's counts vary by season and day-of-week, (3) assigning each short-term count location to the most similar factor group, and (4) multiplying daily counts by seasonality, day-of-week (and sometimes additional) factors to obtain estimates of annual average daily traffic. The FHWA's guidelines for minimum accuracy of non-traditional sources are based on (referenced to) the accuracy of these short-term counts.

Accuracy and Precision Limits associated with factored Short-Term Counts

USDOT has previously benchmarked the accuracy and precision associated with the method of factoring 48-hour counts to annual averages (see Table 5 below). The concept embedded in the FHWA guidelines is that if an agency wants to procure an AADT dataset (such as the vendor products included in this study), these products should be *at least as accurate and precise* as factored, short-term counts. Thus, the FHWA statistical method involves calculating the accuracy and precision of vendor provided volume estimates to see if they compare favorably with the values associated with the 'status quo' short-term counter approach.

Table 5 (copied directly from Table 14 on page 49 of the FHWA report) summarizes the accuracy and precision acceptance limits based on USDOT's historical analysis of factored 48-hour counts, reporting limits separately for four different volume bins.

- **Accuracy Limit #1: '*Minimally 95% Probability, Traffic Count Error (TCE) Median Error Bias (%)*'**: This metric provides a near-worst-case interval (middle 95%) of sample Median TCE values expected from factored AADTs, given a sample of N sites. At higher sample sizes the confidence interval becomes narrower, and the sample median more closely approximates the population median. This accuracy metric is testing for bias (consistent over or under-estimation).
- **Accuracy Limit #2**: '**Minimally 95% Probability, MAPE Upper 95% Confidence Bound (CB) (%)**': This metric reports the near-worst case MAPE (95^{th} percentile) that would be expected from factored AADTs given a sample of N sites. At higher sample sizes this upper limit gets smaller, as large samples are less sensitive to being inflated by individual outliers.
- **Precision Limit**: ***Minimally 95% Probability, 95% TCE Population Error Range (%)***':

The precision limit describes how spread out (i.e., middle 95%) the site errors are when assessing factored AADTs. Unlike the accuracy limits above, this is a description of the underlying population of count site errors (not central tendency sample statistic like median PE and MAPE), so the limits do not depend on sample size. However, note that the testing procedure (described later) *does* depend on sample size.

The accuracy and precision limits as exhibited in Table 5 from the FHWA guidance were obtained (from previous USDOT studies) by analyzing a sub-sample of 48-hour counts obtained from CCS. These sub-samples were treated as 48-hour short-term counts, factored appropriately, and then compared to the official AADT as derived from the full data set from their respective CCS. These fundamental studies of sub-samples of 48-hour synthesized STC, formed the basis for accuracy expectations for non-traditional volume estimates.

Table 5: Accuracy and Precision Limits from Factored Short-Term Counts (from FHWA report)

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 (%)
100 – 499 (very low)	N ≥ 1,000	± 2.7	18.8	± 64.4
	N = 200	± 4.1	19.9	
	N = 100	± 5.6	20.9	
	N = 50	± 7.6	22.3	
	N = 25	± 10.1	24.0	
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	

†Roadways with AADT less than 100 have no pre-defined standards.

Evaluating Accuracy and Precision of Vendor AADT Estimates

The limits in Table 5 serve as the basis for assessing the accuracy and precision of vendor AADT estimates. Evaluating accuracy is straightforward; Median TCE (i.e., Percentage Error) and MAPE are computed for each volume bin and compared to the limits provided in Table 5, resulting in a pass/fail designation for each bin. Note that this table only reports accuracy limits for specific sample sizes (N=25, N=50). In practice, it is often useful to know the precise limits for sample sizes in between the values listed in the table. In such cases it is acceptable to either interpolate between records in the table or use a set of equations provided by FHWA (see Table 14a in FHWA report), for the specified sample size.

The precision test is also straightforward to execute but relies on more sophisticated underlying statistical quality control theory. To execute this test, the TCE (i.e., Percentage Error) is calculated for each record in the dataset and compared to the precision interval limit (last column in Table 5). Each record is flagged as “in” or “out” of the precision interval limit, and the total number of observations that fall outside the limits are summed across the whole dataset (or subset to be included in the precision test). This total is then compared to the number of allowable failures based on specified sampling plan, which should be determined ahead of time. If the number of observations that fall outside the precision limits is less than or equal to the allowable limit, the precision test passes, otherwise it fails. This technique is often used in manufacturing to determine whether to accept or reject a production lot, based on inspecting a sample.

At a high level, the sampling plan seeks to determine the number of ‘failures’ that can be considered acceptable, while balancing risk to both the agency (a potential consumer of the data) and vendor (the producer). A sampling plan requires two binomial equations to be solved simultaneously:

$$1-\alpha = \sum_{d=0}^{c} \frac{n!}{d!\,(n-d)!} p_1^d (1-p_1)^{(n-d)}$$

$$\beta = \sum_{d=0}^{c} \frac{n!}{d!\,(n-d)!} p_2^d (1-p_2)^{(n-d)}$$

- *n*: sample size
- *c*: number of allowable failures (**what we want to solve for)**
- *p1*: failure rate that is considered “good performance”. Here failure rate is defined as the proportion of records that fall outside of precision intervals from Table 5.
- *p2*: failure rate that is considered “bad performance”. Here failure rate is defined as the proportion of records that fall outside of precision intervals from Table 5.
- α: vendor risk (usually set to 0.05, or 5%). This is the probability that the test fails even though the actual failure rate (proportion of records that fall outside precision intervals) is less than *p1* (the threshold for “good performance”)
- β: agency risk (usually set to 0.05, or 5%). This is the probability that the passes even though the actual failure rate (proportion of records that fall outside precision intervals) is greater than *p2* (the threshold for “bad performance”)

Solving these equations is not trivial because some parameters – like sample size – are automatically determined based on the number of available validation locations and must be discrete values. The value for c (number of failures) is what we want to solve for and must also be a discrete value. Alpha and beta are often set to 0.05 (5%) to balance agency and vendor risk, which just leaves p1 and p2 (the definitions of “good” and “bad” failure rates, respectively). *Evaluators are free to come up with sampling plans that satisfy these constraints; however, they must determine the value of* ***c*** *(number of failures) before running the precision test.*

To make things easier, FHWA has prepared a table of reasonable sampling plans – each of which uses alpha = 0.05 and beta = 0.05 and tries to pick p1 and p2 failure rates that are as fair as possible. This is summarized in Table 6 below. For clarity, the validation team uses the sampling plan in Table 6 for our implementation of the FHWA method. In cases where the sample size falls between observed records, we perform linear interpolation and round the value of c (number of allowed failures) up to the nearest integer – giving the benefit of the doubt to the vendors.

Table 6: Acceptance sampling plan for precision test (from FHWA report)

n	c	p_1	p_2	α	β
25	0	0.002	0.113	0.05	0.05
50	1	0.007	0.091	0.05	0.05
80	2	0.01	0.077	0.05	0.05
100	4	0.02	0.089	0.05	0.05
200	7	0.02	0.065	0.05	0.05
400	16	0.027	0.060	0.05	0.05
1000	50	0.04	0.063	0.05	0.05

With this information in mind, the FHWA test is conducted as follows:

1. Create a paired dataset with reference and vendor volumes (bi-directional) and calculate Percentage Error (PE) for each record.
2. Assign each record a volume bin based on reference volume value. (Very low = 1-499, Low = 500-4,999, Medium = 5,000-54,999, High = 55,000+)
3. Group records by volume bin. For each bin:
   - **Accuracy Test #1:** Calculate Median PE (called TCE) across records and compare it to "*Minimally 95% Probability Traffic Count Error (TCE) Median Error Bias (%)"* limit.
     - If Median PE is within interval → **Pass**
     - If Median PE is outside interval → **Fail**
   - **Accuracy Test #2:** Calculate MAPE across records and compare it to *"Minimally 95% Probability, MAPE Upper 95% CB (%)"* limit.
     - If MAPE is less than or equal to limit → **Pass**
     - If MAPE exceeds the limit → **Fail**
4. **Precision Test:** For each record in the full dataset (or subset of data to be included in precision test), compare the PE value to '*Minimally 95% Probability, 95% TCE Population Error Range (%)'* and count the total observations outside the interval.
   - If number of 'failures' is less than or equal to limit from sampling plan → **Pass**
   - If number of 'failures' exceeds limit from sampling plan → **Fail**

Changes to FHWA Test Implementation

The implementation of the FHWA method in this report is slightly different than how it was executed in the prior reports. The prior implementation was based on the best available information at the time, but the validation team's understanding has evolved as FHWA has put out updated guidance. The major changes are as follows:

- **Focus on bi-directional volumes**. The FHWA accuracy and precision limits were developed from prior USDOT research with total volume (rather than directional volume) in mind. Additionally, a prior report from the FHWA pooled fund study indicated that total volume should be used. As such, although most of our volume validation focuses on directional volumes, we aggregate directional volumes by count site to evaluate total volume on each roadway.

- **(Slight) Update to accuracy limits.** To improve the accuracy limits, we now use equations provided by FHWA to specify the limit based on sample size. This change is slightly more stringent, but better handles values between sample sizes reported in Table 5. *Impact of changes is minor.*
- **Update to precision limits**. In the previous implementation, the validation team applied the precision test to each volume bin. Although this is allowable, it results in many tests with small sample sizes. When sample sizes are small, the precisions limits typically allow for only 1 or 2 outliers (e.g., 1-2 outliers could cause a failure). In the updated implementation, the precision limits are applied across multiple volume bins, that is the entire data set (or at least the part subject to evaluation). This maximizes the sample size, making passing or failing less dependent on a single outlier. Additionally, we made a slight improvement to how we choose the acceptable number of failures in the sampling plan. *The impact of precision test changes was more significant than the accuracy test changes. The most significant changes were aggregating (pooling) the data across volume bins to increase the sample size. This updated implementation is fairer to vendors, as TDM validation studies typically have statistically small sample sizes when disaggregated to specific volume bins.*

6. Results and Discussion

The AADT results are presented below, beginning with analysis of directional volumes (Section 6.1) and followed by an analysis of bi-directional volumes using FHWA's statistical test (Section 6.2). As will be described in more detail in Section 6.1, initial analysis of directional volumes revealed high errors and major outliers, which were unexpected based on observed vendor performance in prior validations. As such, the validation team spent considerable effort tracking down the cause of these discrepancies – a process that uncovered several different sources of error (beyond just vendor inaccuracy), including miscoded count locations, georeferencing/conflation issues, and errors in agency count data. These errors are important to understand because they can have a disproportionate impact on resulting error metrics and distort apparent performance (including the FHWA statistical test in Section 6.2). Accordingly, Section 6.1 provides concrete examples and explanations for the different sources of error.

This investigative process resulted in the validation team identifying a set of 'suspect locations' to be omitted from the analysis. However, because historical counts at these locations were also provided to vendors for model calibration purposes, errors in the training data may have impacted vendors' volume estimates. To account for this possibility, vendors were given a list of suspect locations to omit from calibration procedures and provided the opportunity to re-submit volume estimates.

The analysis also investigated performance at different types of agency counters (traditional inductive loop counters, counters on tolled facilities, and intersection cameras) and for different levels of road complexity. This analysis found that camera counters were associated with higher errors (even beyond obvious outliers that were omitted from the analysis). However, because cameras are typically deployed at more complex locations (typically signalized intersections), it is unclear whether this is due to the camera technology itself or the environment in which they are deployed, where it is difficult to know exactly which lanes are being counted, or a combination of the two.

Many initial and intermediate results were produced to investigate errors and sources of uncertainty; however, for clarity, the majority of results presented in this report will focus on the final, resubmitted dataset. Due to the remaining questions about camera count locations, the final dataset is reported both with and without cameras included. The non-camera CCS locations are expected to be most consistent with the validation locations from the prior validation activity in Georgia and North Carolina.

6.1 AADT (Directional)

Preliminary Results

The following preliminary results reflect performance prior to identifying various sources of error not attributable to vendor inaccuracy.

Figure 4 plots vendor AADT volumes (y axis) against reference AADT values (x axis) for each vendor's AADT estimates, focusing in particular on the 0-50k vehicle per day (vpd) volume range where approximately 85% of the data lies. The 45-degree dashed line represents points where there is zero error (i.e., estimates are equal to reference values). Data points above the line represent overestimates, and data points under the line reflect underestimates. These plots are color-coded to show Continuous Count Sites (CCS) in blue and Short-Term Counters (STC) in red. These visual plots are useful to observe if any bias exists in the data (more data above or below the 45-degree line), the variance in the data, and prevalence of outliers.

The presence of major outliers is the main takeaway from the plots of the Mass DOT Validation. Two vendors show some systematic errors (Vendor A often underestimates across a range of volumes, while vendor C significantly overestimates on certain low-volume roads, especially STC); however, all vendors show sporadic individual points that disagree significantly – in some cases by thousands of vehicles per day.

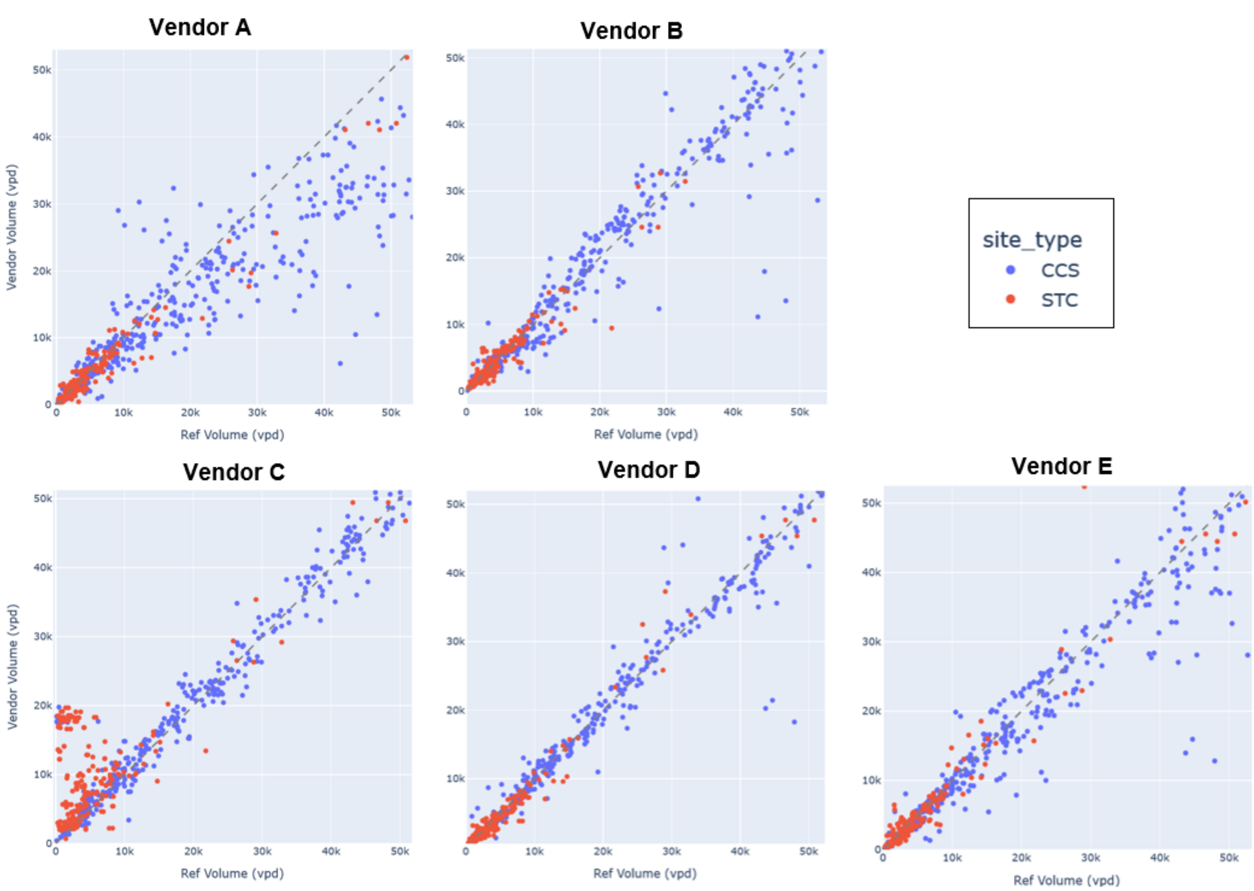


Figure 4: AADT scatterplot for vendors' initial submissions

To quantify performance further, Figures 5-9 reflects the individual scatterplots for each vendor, with separate plots and summary metrics (median and mean APE) reported for CCS and STC. Accompanying boxplots summarize the distribution of errors in units of volume (vehicles).

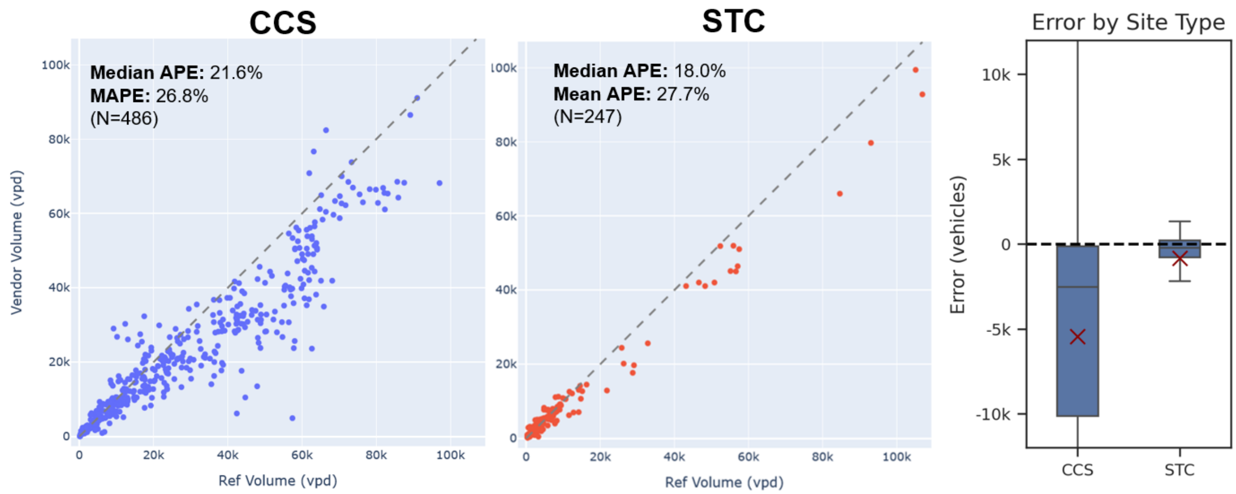


Figure 5: Vendor A scatter plots and error distribution for all CCS and STC locations

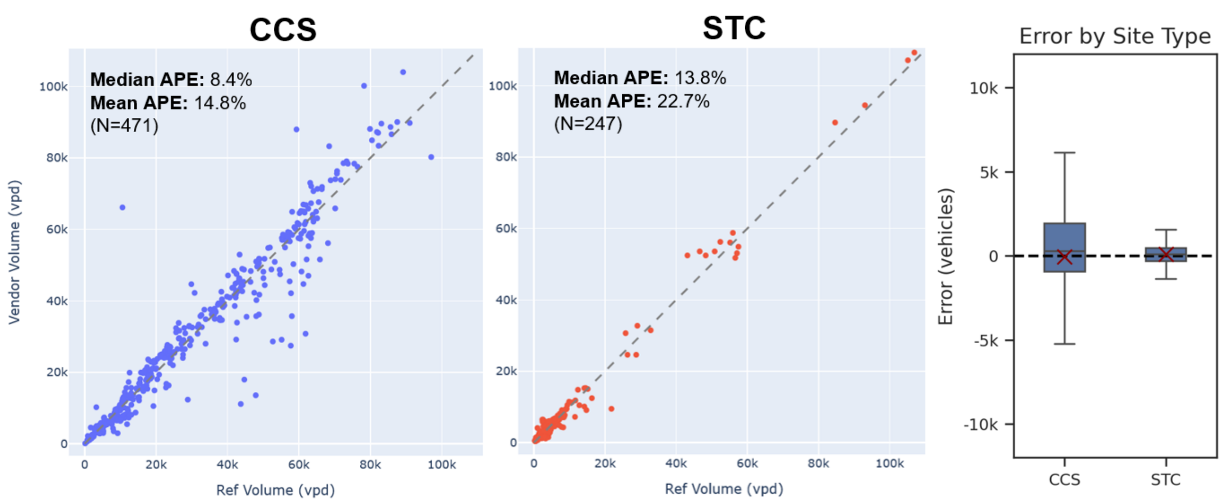


Figure 6: Vendor B scatter plots and error distribution for all CCS and STC locations

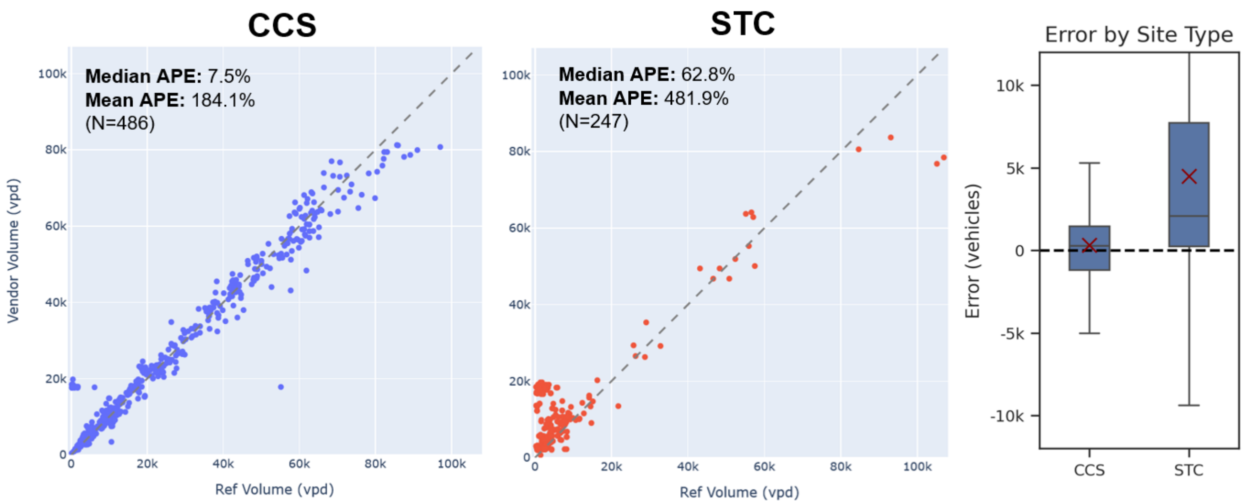


Figure 7: Vendor C scatter plots and error distribution for all CCS and STC location

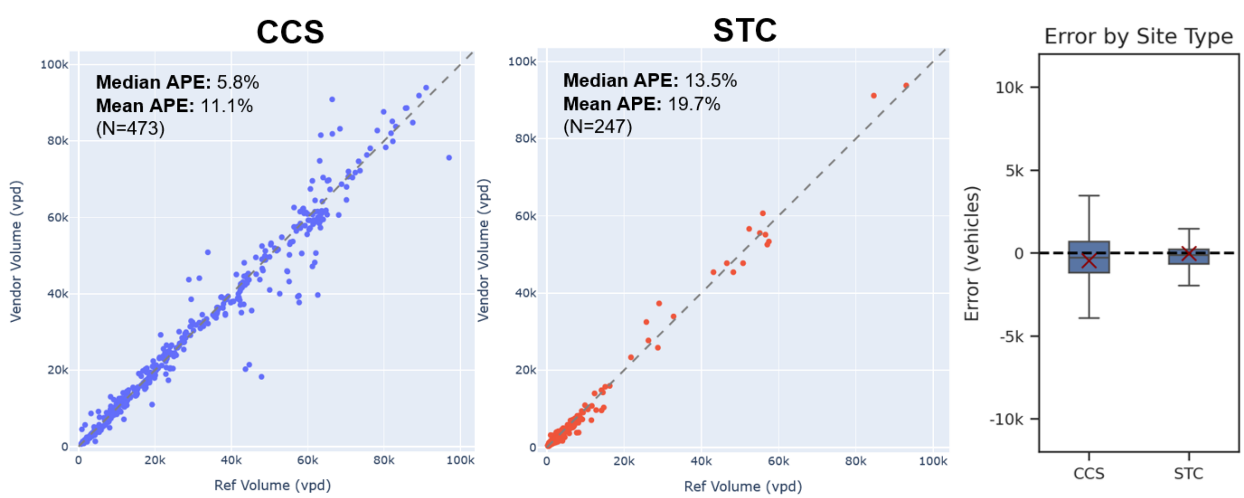


Figure 8: Vendor D scatter plots and error distribution for all CCS and STC locations

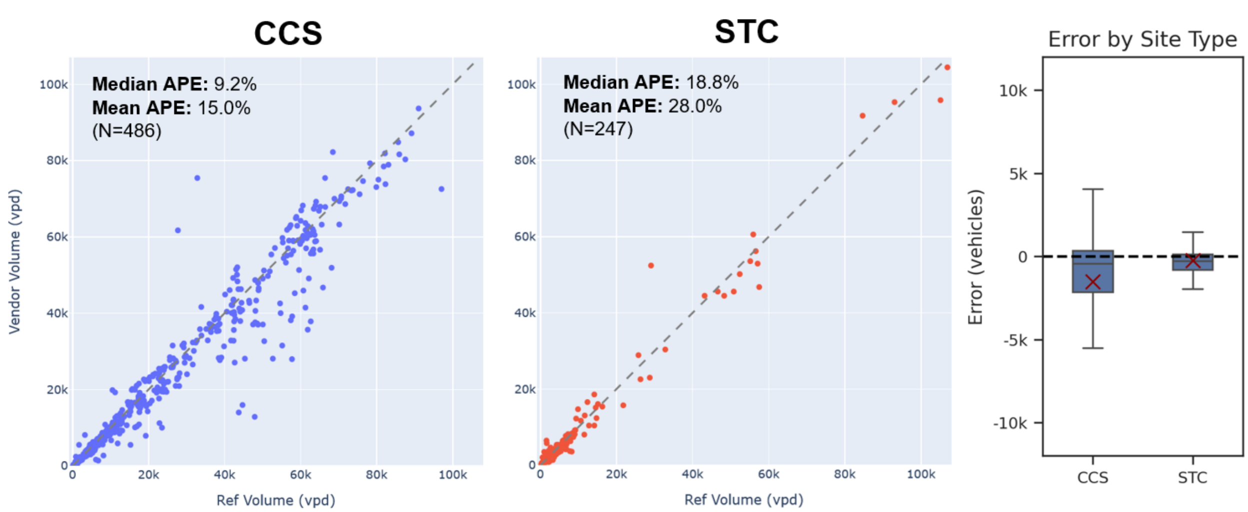


Figure 9: Vendor E scatter plots and error distribution for all CCS and STC locations

Common to all five vendors in the preliminary results were the prevalence of significant outliers, that is multiple data points that fall significantly far from the 45-degree dashed lined. These plots reinforce the discrepancies between vendor estimates and agency counts and show the impact that large outliers can have on error metrics. For example, consider vendor C in Figure 7 above. In the CCS plot, the majority of points are reasonably close to the ideal 45-degree line, resulting in a low median (typical') APE value of 7.5%. However, because there are a number of points with over 1000% error (when agency counts were very low and the vendor reported estimates around 20,000 vpd), the mean APE (i.e., MAPE) is over 100%.

Because outliers can be quite impactful – both for the common statistics metrics reported above, and also in FHWA's statistical tests -- the validation team determined that it was necessary to understand the source of these discrepancies to determine whether the issues were related to vendor estimates or other factors. Note the high prevalence of outliers depicted in the previous charts differed substantially from validation results seen in Georgia 12 months prior. Reasoning that overall industry performance would not decay over time, this added to the concern that additional issues were at play within the MassDOT data set.

Outlier Investigation

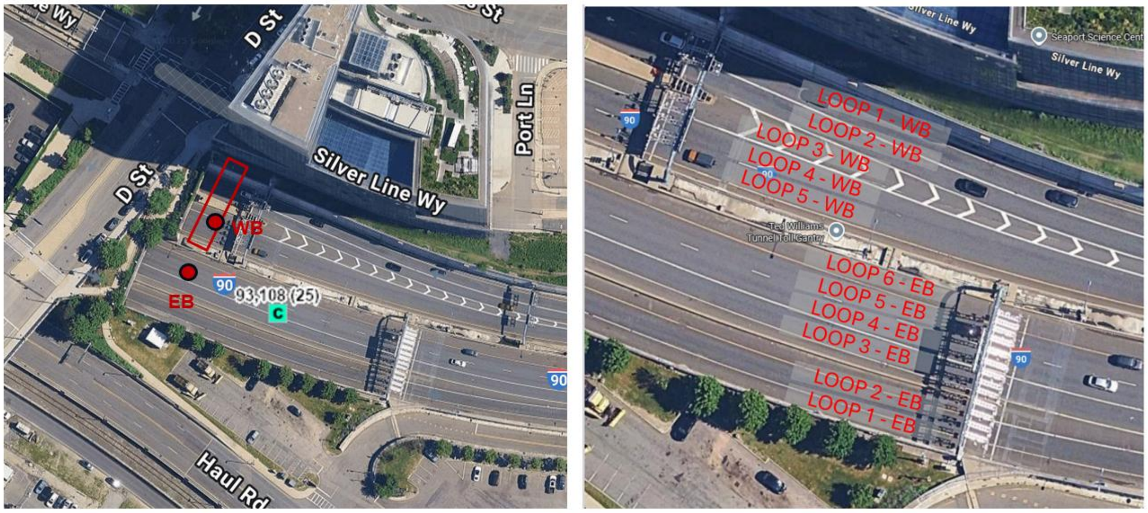
The validation team identified the most notable discrepancies between agency volume counts and vendor estimates for each vendor's AADT product. For each vendor, the most significant outliers (such as the top ten) were analyzed. The validation team double checked the agency counts (and subsequent aggregation process to obtain AADT), sought to understand the type of counter and its exact deployment location, and sanity checked vendor estimates. Having data from 5 vendors proved useful for understanding whether outliers were associated with a single vendor or shared by many. If excessive error at a specific count site was common from multiple vendors, it implied a potential problem with the reference data at that location. It also proved useful to compare volumes from both directions of travel on the same roadway segment. This analysis focuses on AADT. Volume counts on opposing travel directions are usually balanced over a prolonged period. Discrepancies between directional flows at the same location, for either agency counts or vendor estimates, flagged anomalies in the data. Broadly speaking, the source of error for major outliers were categorized as follows:

- **Miscoded Site Location**: Manual investigation of several high-error sites common to all vendors led to the discovery that a subset of locations had been miscoded in the DOT's database system. MassDOT discovered these location errors independent of the validation team's investigation and updated the database after the validation activity had started. Because of this issue, vendor volume estimates were provided at different locations than the actual location of the CCS., often resulting in substantial differences in volume. *In total, 39 sites (52 directional locations) were impacted – all of which were 'traditional' inductive loop counters*
- **Location Referencing / Conflation Errors**. Additionally, another set of errors were related to location referencing or conflation. Recall that the validation team provided location references to each vendor to describe where they should report volume estimates. These location references (encoded in CATT Works Georeferencing Protocol) – which contain latitude, longitude, heading, road name, and road classification – did not always contain sufficient information to precisely identify where the agencies collected reference counts.

  Figure 10 below is one example. The image on the left shows EB and WB directions at a particular count site along I-90. The latitude and longitude clearly distinguish EB from WB lanes, but there is some ambiguity about how many lanes should be included in the

WB traffic – the reason being that it is near an exit ramp, and is unclear what vendors should report (e.g., through lanes vs all lanes vs ramp lanes). MassDOT provided loop assignment diagrams (right image) that clarified that the WB direction was including all lanes (through plus ramp), which differed from most vendors' interpretation.

Figure 10: Example of errors due to georeferencing/conflation.



The validation team observed that ambiguous locations tended to be associated with more complex locations – typically signalized arterials (intersection cameras) and to a lesser extent, tolled facilities with entrance/exit ramps, like the example above. Traditional counters on highways were typically located in less complex locations, minimizing the probability of error. See the subsection below called "Complexity Analysis" for further discussion on site complexity. As a result of this discovery, it is recommended that geo-referencing for future validation also includes number of lanes and an indication of lane type (e.g., mainline, ramp, other).

- **Vendor Data Error**. In some cases, the source of the error was traced to vendors' processes, which truly reflect inaccuracy in vendor data and should be included in the analysis. To identify vendor errors, we focused on errors that were unique to a single vendor at a specific location. Differences in directional flow also contributed to identifying outliers that are attributed to vendor error. Although directional differences are to be expected during times of the day or days of the week, directional imbalances at an annual average scale often indicate an anomaly.

- **Reference Data Error.** Finally, some outliers were associated with errors in the agency count data. We identified these errors by searching for locations where all (or most) vendors showed strong internal agreement but differed from the reference counts. At these locations, we then ran sanity checks on the reference data (e.g., checked for imbalanced directional AADT) and consulted with MassDOT staff who had a better understanding of local traffic patterns.

  Figure 11 shows one example in which all vendors disagreed with the reference counts on the North leg of the intersection. Further investigation into the reference data showed that the Southbound AADT value was three times larger than the Northbound AADT. This location was monitored by an intersection camera, whose primary purpose was to provide turning movement detection for traffic control. The validation team pulled TMC reports and consulted with MassDOT to determine the source of the issue. It turned out that the slip lane (outlined in the rectangle below, between the East and North legs of the intersection) was occluded from the camera view (not needed for traffic control) – resulting in a substantial portion of total traffic being undercounted.

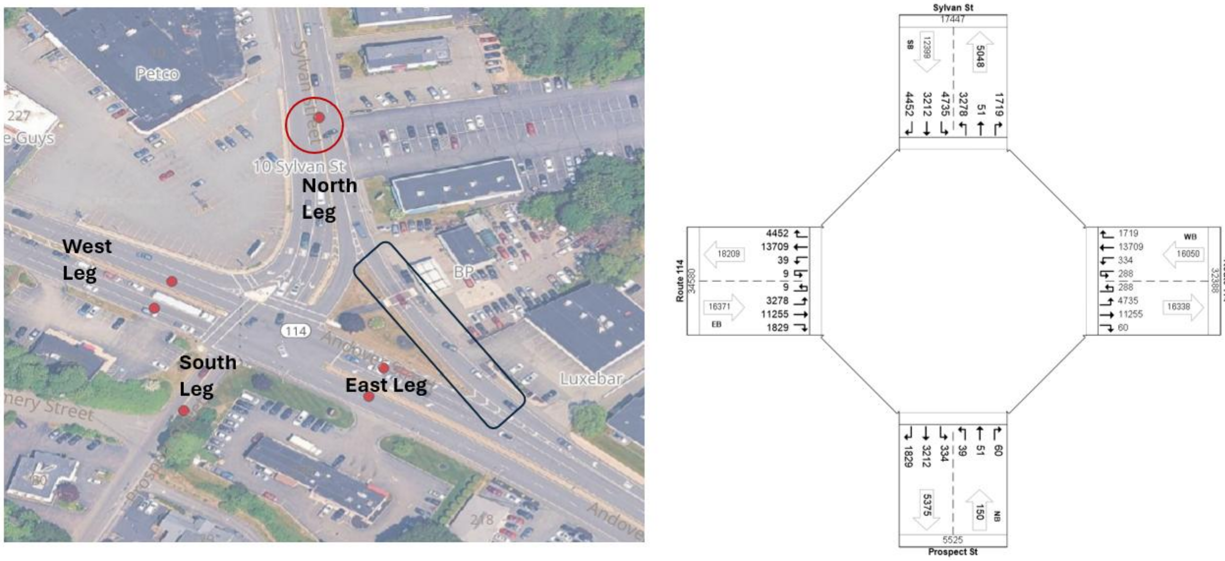


Figure 11: Example of error in reference data due to occlusion in camera counter

Performance by Counter Type

Figures 12-14 visually depict performance across Traditional, Toll (AET), and Camera locations. *Again, note that Traditional and Toll (AET) locations use the same detection method (inductive loops), with Cameras differing most significantly from the other two.* In the interest of keeping the plots readable, this paper presents the results for just three of the five vendors to highlight performance across industry. The validation team coded each scatter plot by color to flag 'suspect locations' – a subset of the outliers that were deemed unrelated to vendor performance. These suspect locations will be omitted from the dataset in the final results; however, in this section the validation team includes the suspect data points in the plots and metrics. These figures and corresponding metrics show consistently higher errors for Camera locations with respect to percentage errors. Note that camera counters tend to be stationed at locations with lower volume, so the magnitude of the error will be lower although the percentage error is higher. In all three sets of plots across traditional, AET and camera location, many, if not most, of the large outliers were deemed not to be vendor error and are marked as 'suspect' as indicated by the red data point.

Traditional CCS

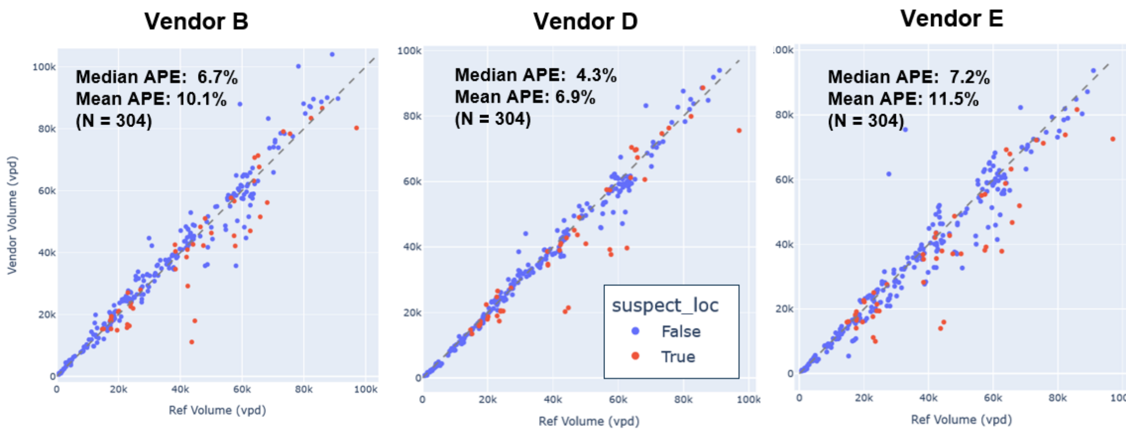


Figure 12: Scatter plots showing Traditional CCS locations for three vendors.

Automated Electronic Tolling (AET) CCS

Figure 13: Scatter plots showing AET CCS locations for three vendors.

Camera CCS

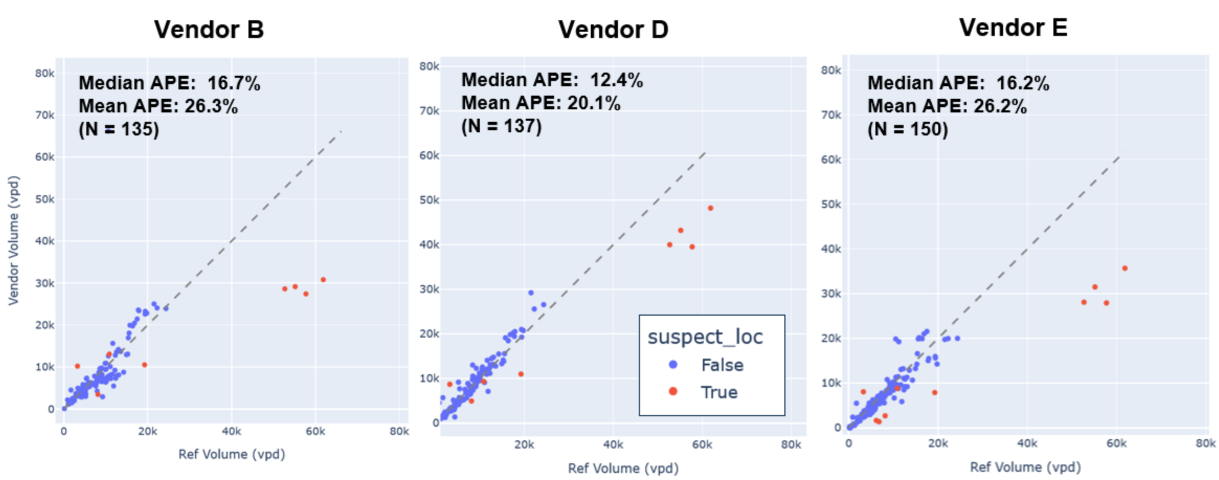


Figure 14: Scatter plots showing Camera CCS locations for three vendors.

Impact of Site Complexity

The outlier review process suggested that many georeferencing/conflation issues were found in complex environments where it was more difficult to determine how the validation location should be assigned to the road network. To test this theory, the validation team classified each of the CCS validation locations as low, medium, or high complexity based on the number of OpenStreetMap segments that could be reached within a 30m (~100ft) radius. Figure 15 summarizes the classification scheme and associated thresholds.

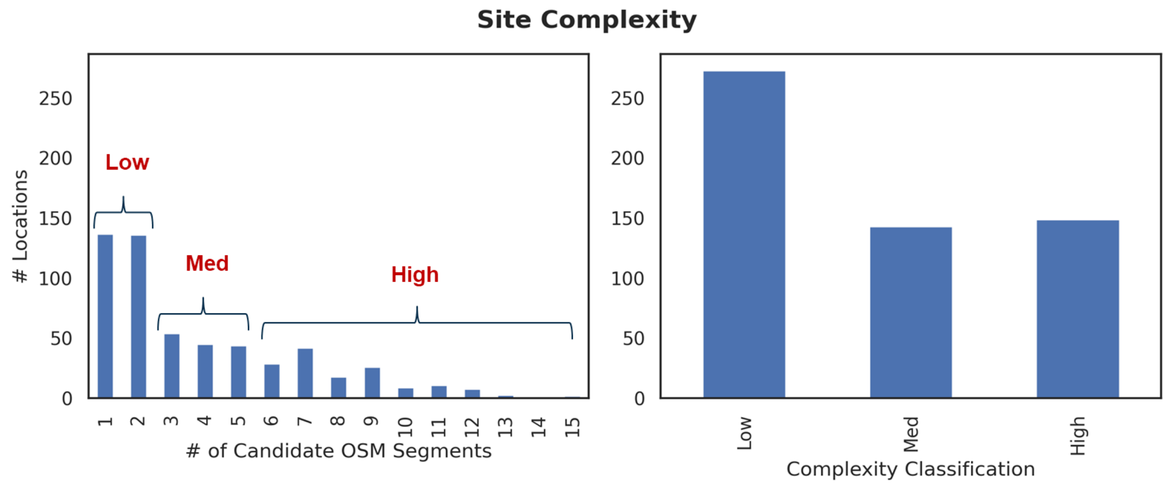


Figure 15: Classification of CCS locations as low/med/high complexity.

Figure 16 shows examples of low, medium, and high complexity locations. Low complexity locations are typically simple to interpret (1-2 candidate segments, typically on interstates and similar facilities) with little to no ambiguity with respect to mapping of the location onto the appropriate roadway segment, while high complexity locations present many more challenges and are often found in urban environments. The High complexity example in Figure 13 shows an intersection in an urban environment with 15 candidate segments within the specified radius.

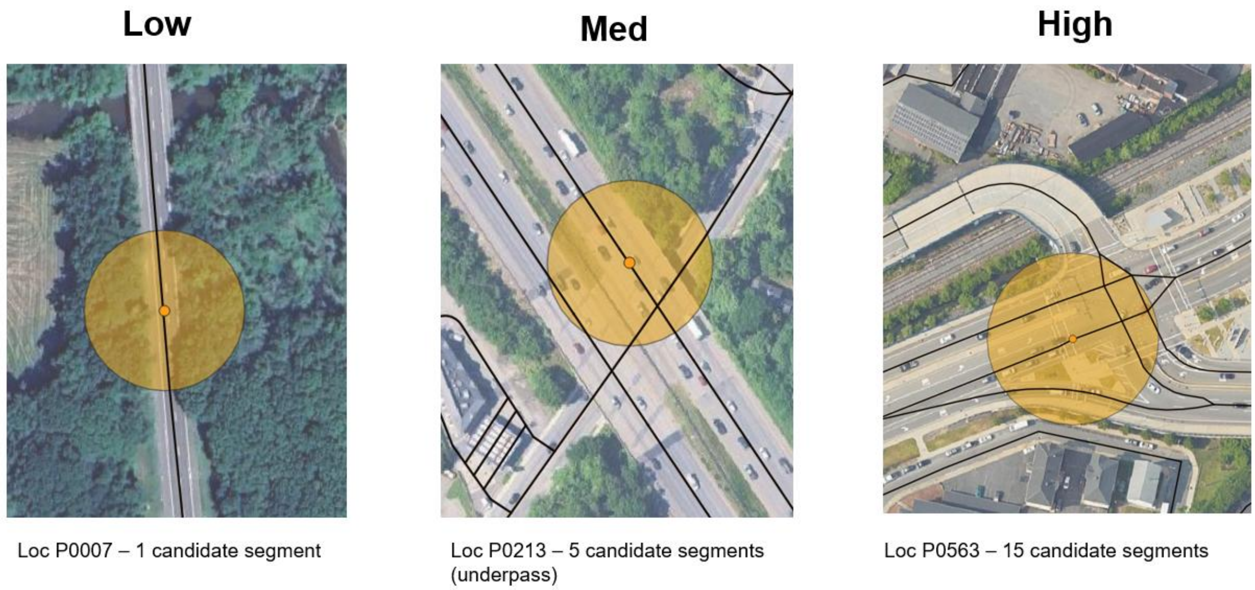


Figure 16: Examples of low/med/high complexity.

Figure 17 summarizes the number of locations that fall in each complexity bin for the three CCS counter types. Traditional counters – which represent the greatest share -- are typically low complexity, Toll counters are split between low and medium complexity, and camera counters are generally associated with higher complexity. Because cameras tend to be deployed in more complex environments, it is unclear whether lower accuracy at camera counters is due to the technology itself or related to their deployment context, or perhaps a combination of both.

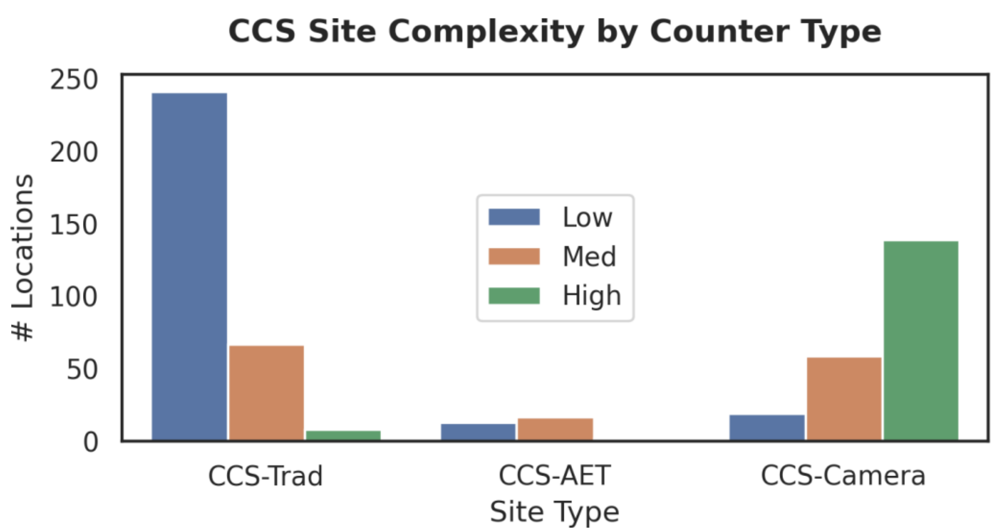


Figure 17: Distribution of low/med/high complexity locations by site type

One of the challenges with camera locations is that **it is not always clear where the camera is measuring vehicle counts** relative to the intersection. Figure 18 shows an example of an intersection with camera deployment. The orange dots show the location of count sites as reported in the DOT database system. Note that these are not "traditional" count stations, but rather individual legs of an intersection treated as individual sites (with volumes derived from aggregated turning movement counts). Consider the validation location represented by red dot on the South leg of the intersection, representing traffic flow in the SE direction (i.e., in the direction of the arrow). The intent of the validation location is to capture this traffic flow, but it's unclear whether the camera is truly counting traffic precisely where the red vector is placed or also including the slip lane from the West leg of the intersection. If the intent was to capture all traffic from the intersection proceeding on this leg of the intersection, the locations would be set back farther from the intersection to more clearly communicate intent. However, it is not likely that the cameras at the intersection are placed and configured to capture the referenced slip lane if they were installed and configured primarily for traffic control purposes.

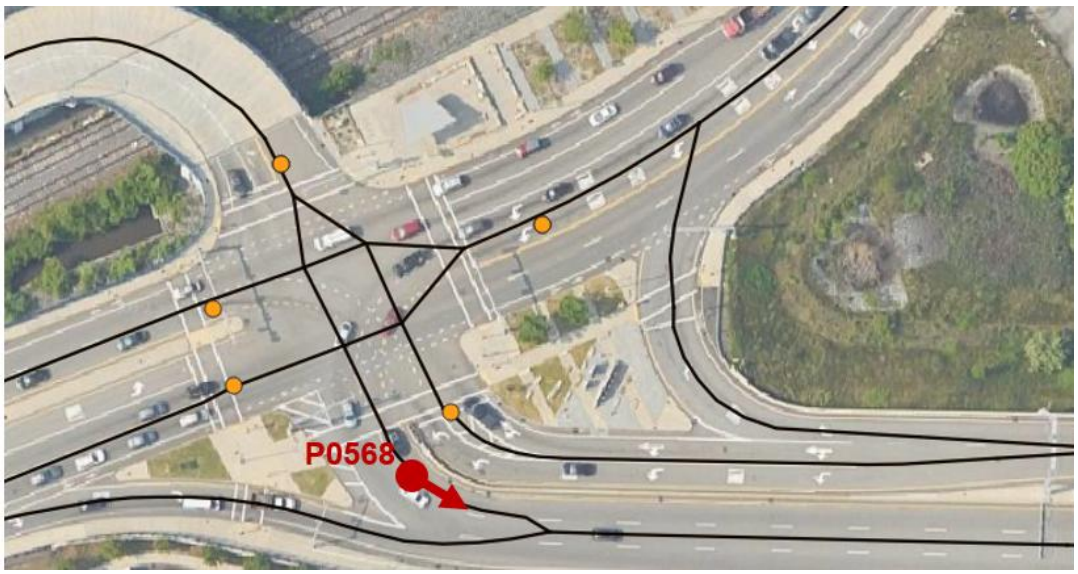


Figure 18: Examples of challenges associated with intersection camera.

Outlier Removal and Resubmission

Based on the preliminary analysis and outlier investigation described above, the validation team omitted 72 directional locations from the analysis. These locations encompassed the sites whose locations had been miscoded, locations with known errors in the reference data, and ambiguous locations where it was not obvious which lanes or facilities were being described.

Note, however, that these suspect locations were included in the calibration data provided by the validation team to the vendors (and thus used to train vendor model estimates). For this reason, vendors offered the opportunity to regenerate estimates and resubmit results. The remainder of the results (see "Final Results" below, as well as Section 6.2) will focus on resubmitted volume estimates with the suspect locations removed.

Final Results

The validation team generated final results from data sets resubmitted by vendors after 72 locations were flagged and subsequently removed. All but one vendor (Vendor B) re-submitted volume estimates. For the vendor that did not re-submit, the suspect locations were simply removed from the original submission.

The report details overall summary error metrics in Appendix A. The following subsections present highlights from these results. Due to unresolved questions about the accuracy of camera count locations, this report summarizes final results with and without camera locations included. The non-camera CCS locations are expected to be most consistent with evaluation locations from prior validation activities in GA and NC.

Overall Performance (All Sites)

Figure 19 plots vendor AADT volumes (y axis) against reference AADT values (x axis) for each vendor at all sites (including CCS Camera locations), focusing on the 0-50k vpd volume range where approximately 85% of the data lies. The plots in Figure 19 show better performance than the preliminary results from Figure 4; more data points fall close to the 'ideal' 45-degree line, and there are fewer outliers, and vendors A and C have less systematic bias (although A still has a negative bias than in Figure 4, just to a lesser degree).

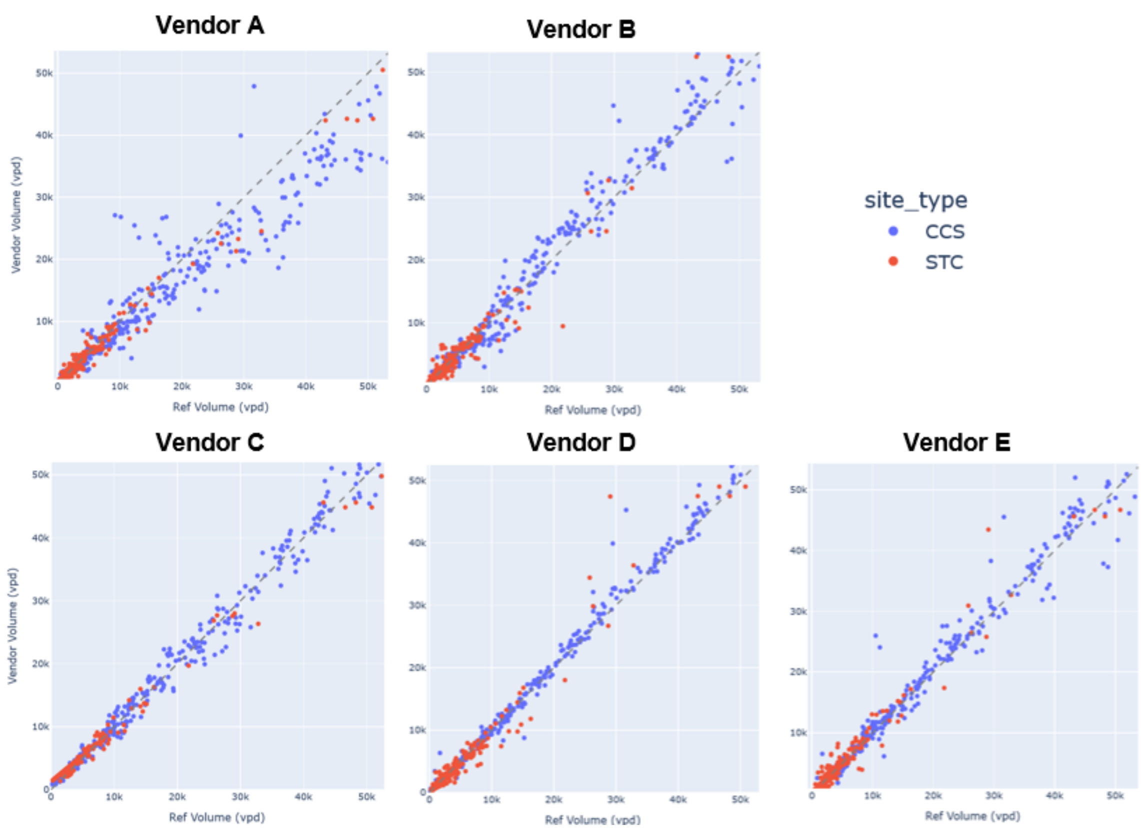


Figure 19: Final scatter plots across all vendors

Figures 20-24 show the individual scatterplots for each vendor, with separate plots and summary metrics (median and mean APE) reported for CCS and STC. Accompanying boxplots summarize the distribution of errors in units of volume (vehicles).

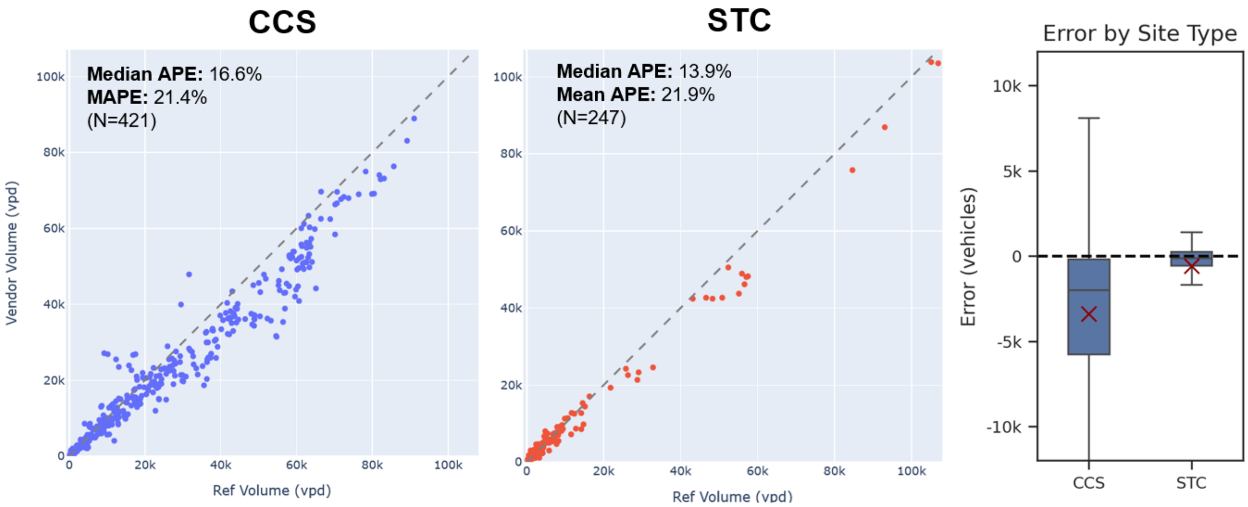


Figure 20: Final scatter plots and error distribution for Vendor A

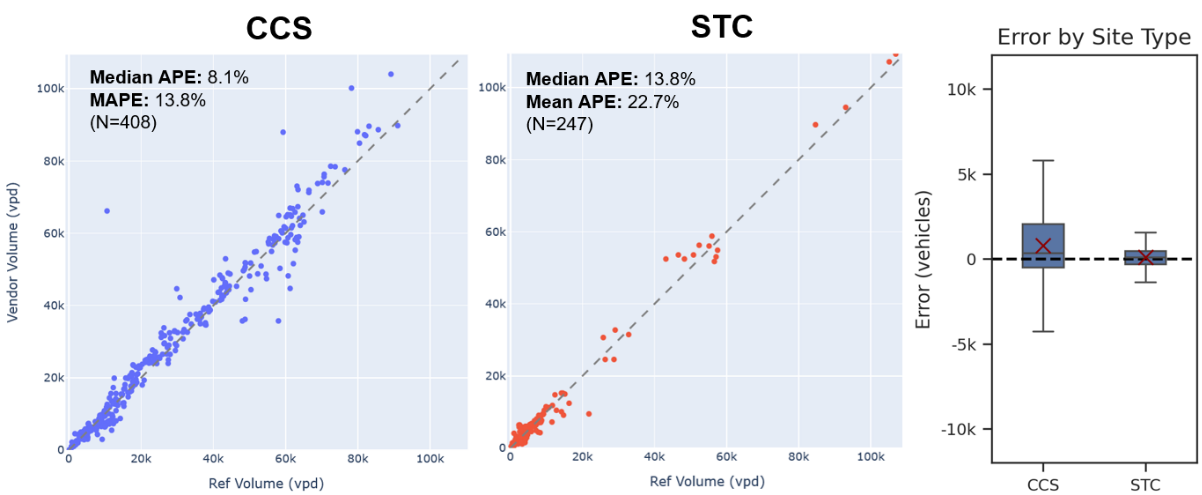


Figure 21: Final scatter plots and error distribution for Vendor B

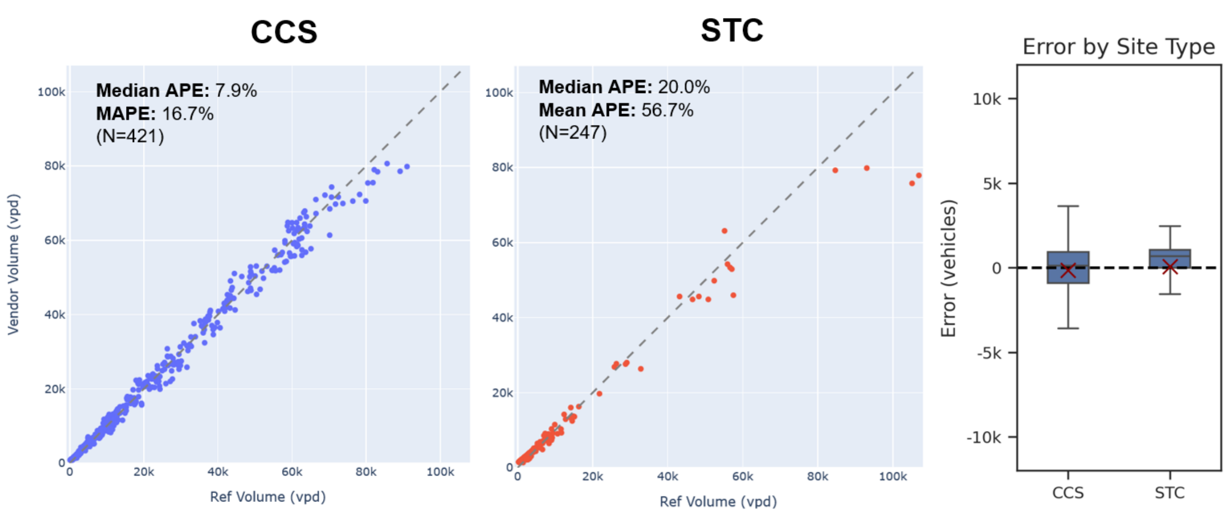


Figure 22: Final scatter plots and error distribution for Vendor C

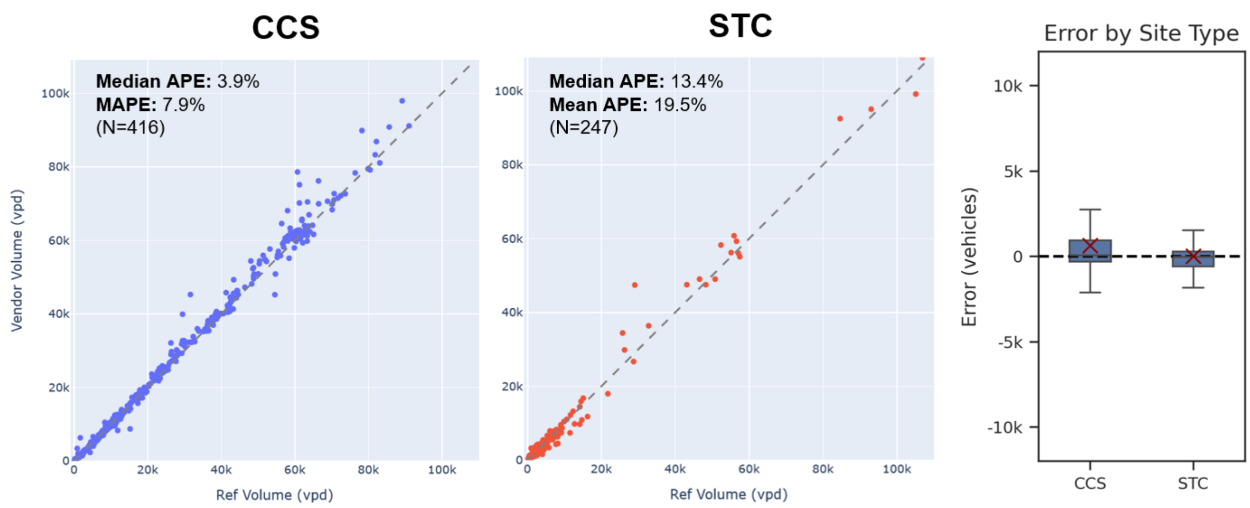


Figure 23: Final scatter plots and error distribution for Vendor D

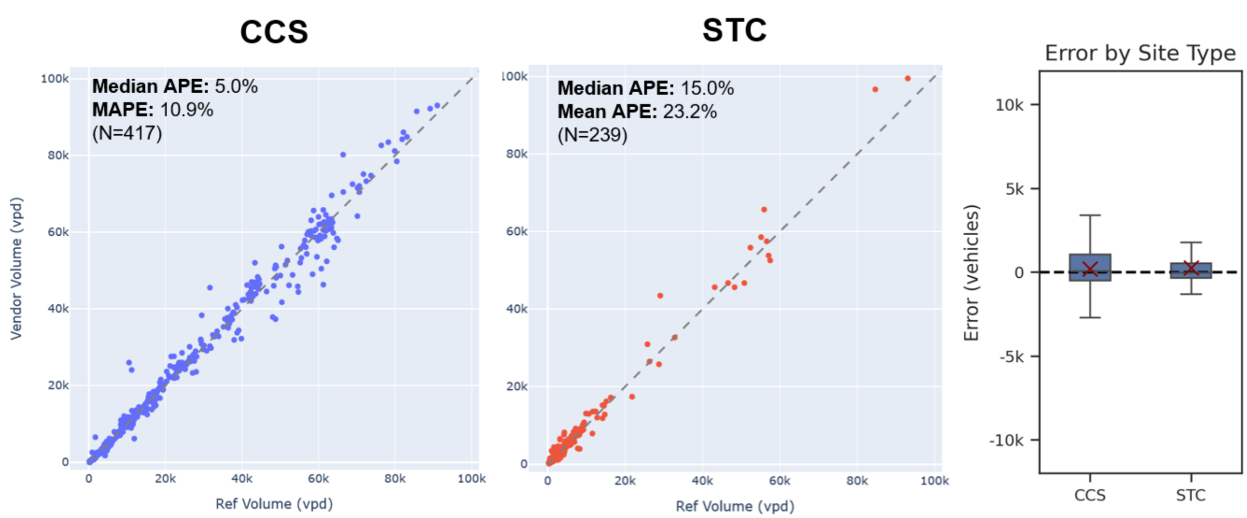


Figure 24: Final scatter plots and error distribution for Vendor E

The summary statistics (Median and Mean APE) from the final analysis indicate improved accuracy – particularly at CCS locations. Additionally, the two vendors with systematic biases in the preliminary analysis showed improvement; Vendor A still systematically underestimates volumes at CCS locations, but to a lesser degree, while Vendor C's low-volume overestimation issue is no longer present.

Performance by Volume Bin

Figures 25 and 26 show the distribution of percent error for each vendor aggregated by volume bins. Figure 25 plots results for all CCS locations, and Figure 26 focuses on non-camera sites (Traditional and AET locations). The volume bins were chosen based on FHWA's volume bin definitions; FHWA bi-directional volume bins were divided in half to create a directional version of bin thresholds.

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), the vertical line in the middle of the box representing the median value (50th percentile) and red 'X's representing the mean value of the error distribution. The whiskers extending to the left of the 25th percentile (Q1) and to the right of the 75th percentile (Q3) are an estimate of the 95th percentile limits (obtained by extending the 25th and 75th percentile by 1.5 times the inter-quartile range (Q3-Q1)). Data points within the whiskers represent the estimated non-outlier range. Note that boxplots may not be reliable for scenarios with limited samples, so this report omits them when there are fewer than 20 observations. In such cases, this report details only the mean and median values (black diamond and red 'X' markers, respectively). For all volume bins, the validation team plots the individual data points (representing percent error at a specific location) as light blue points to give a visual indication of the distribution.

Visual inspection of Figures 25 and 26 result in the following observations:

- **Vendor A** consistently underestimates volume across volume ranges – whether or not camera locations are included (i.e., Fig 25 vs 26). While the level of underestimation is better than the initial submission, there is still notable bias relative to the reference values.
- **Vendor B**'s errors are typically within +/- 15%, but the sign of the error switches across volume bins when all locations are included (Fig 25). However, when camera locations are excluded (Fig 26), the magnitude of Vendor B's error decreases slightly and there is a slight tendency to overestimate.
- **Vendor C** shows strong agreement across mid-to-high volume ranges (especially above 5k vpd) and has very high errors in the lowest volume ranges when cameras are included (Fig 25). If camera locations are excluded, the performance is improved and more consistent.
- **Vendor D** shows high accuracy with minimal bias across all volume bins in Fig 25, and the magnitude of the errors drop even further when excluding cameras (Fig 26). This dataset consistently matches the reference AADT values across all volume ranges.
- **Vendor E** shows strong agreement across mid-to-high volume ranges, with a slight tendency to underestimate at lower volumes when all locations are considered (Fig 25). When cameras are excluded, performance improves.

Performance by Road Class

Figures 27 and 28 show the distribution of percent error for each vendor, aggregated by road class. This is similar to Figures 25-26, but these plots focus on functional class rather than volume range. Figure 27 shows results for all CCS locations, while Figure 28 focuses on non-camera sites (Traditional and AET). Although there is some overlap with the prior figures (i.e., FRC 1 locations generally correspond to high volume bins), it provides a slightly different perspective.

In Figure 27, FRC 3 (Other Principal Arterials) stands out as having notable error/bias across all vendors when camara locations are included. This bias largely disappears if camera locations are omitted.

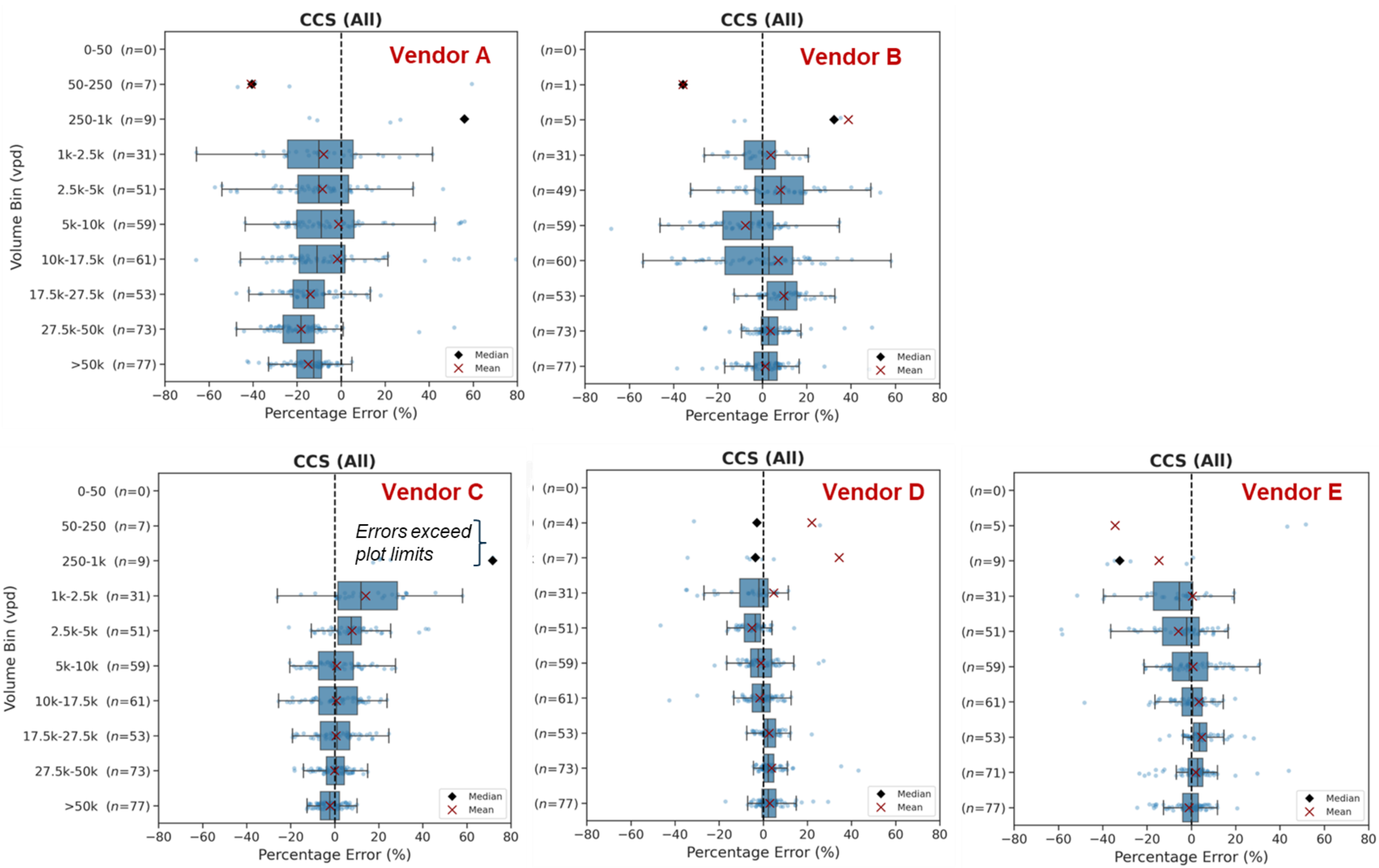


Figure 25: Percent Error boxplots by volume bin (all CCS locations)

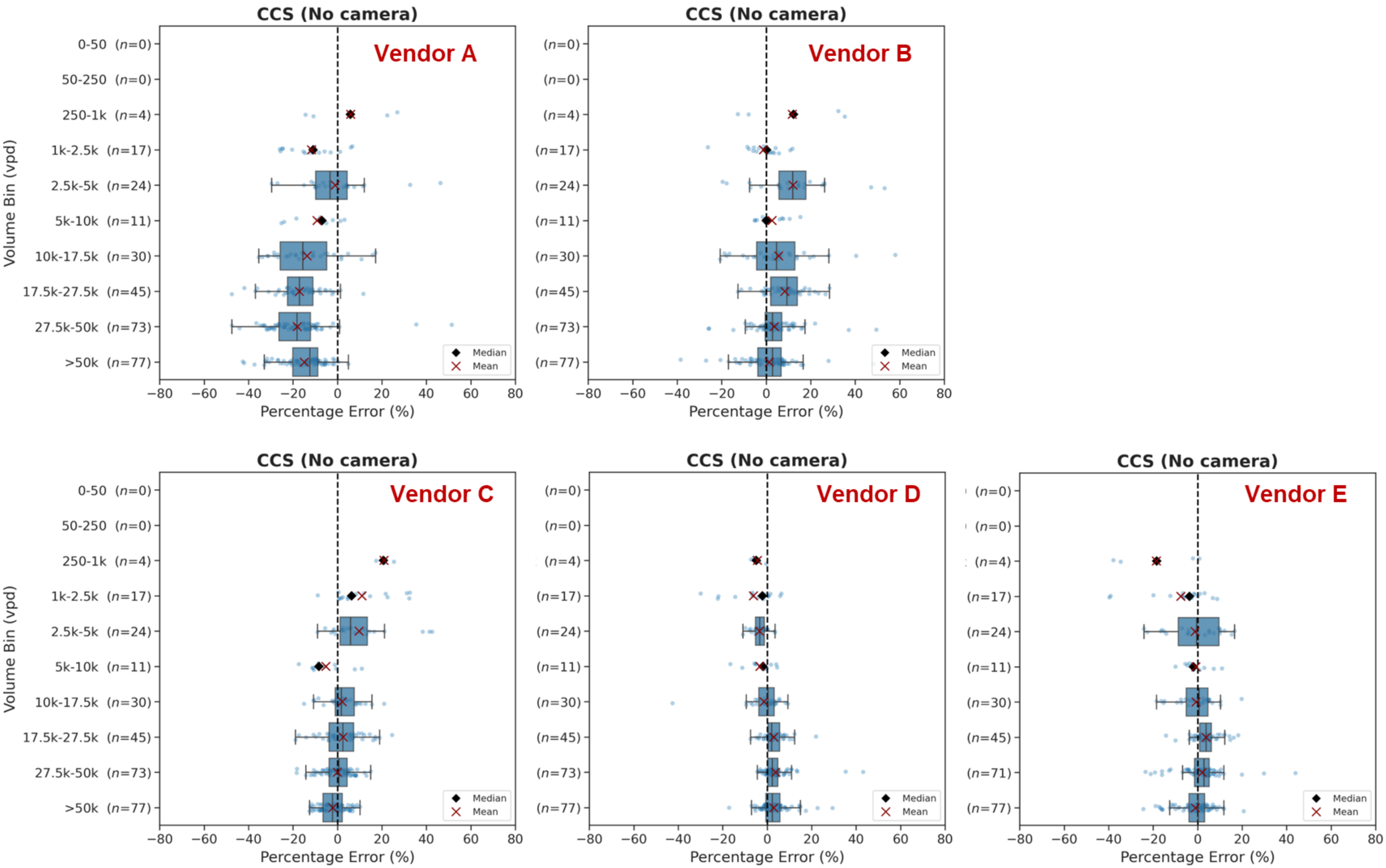


Figure 26: Percent Error boxplots by volume bin (CCS locations with cameras omitted)

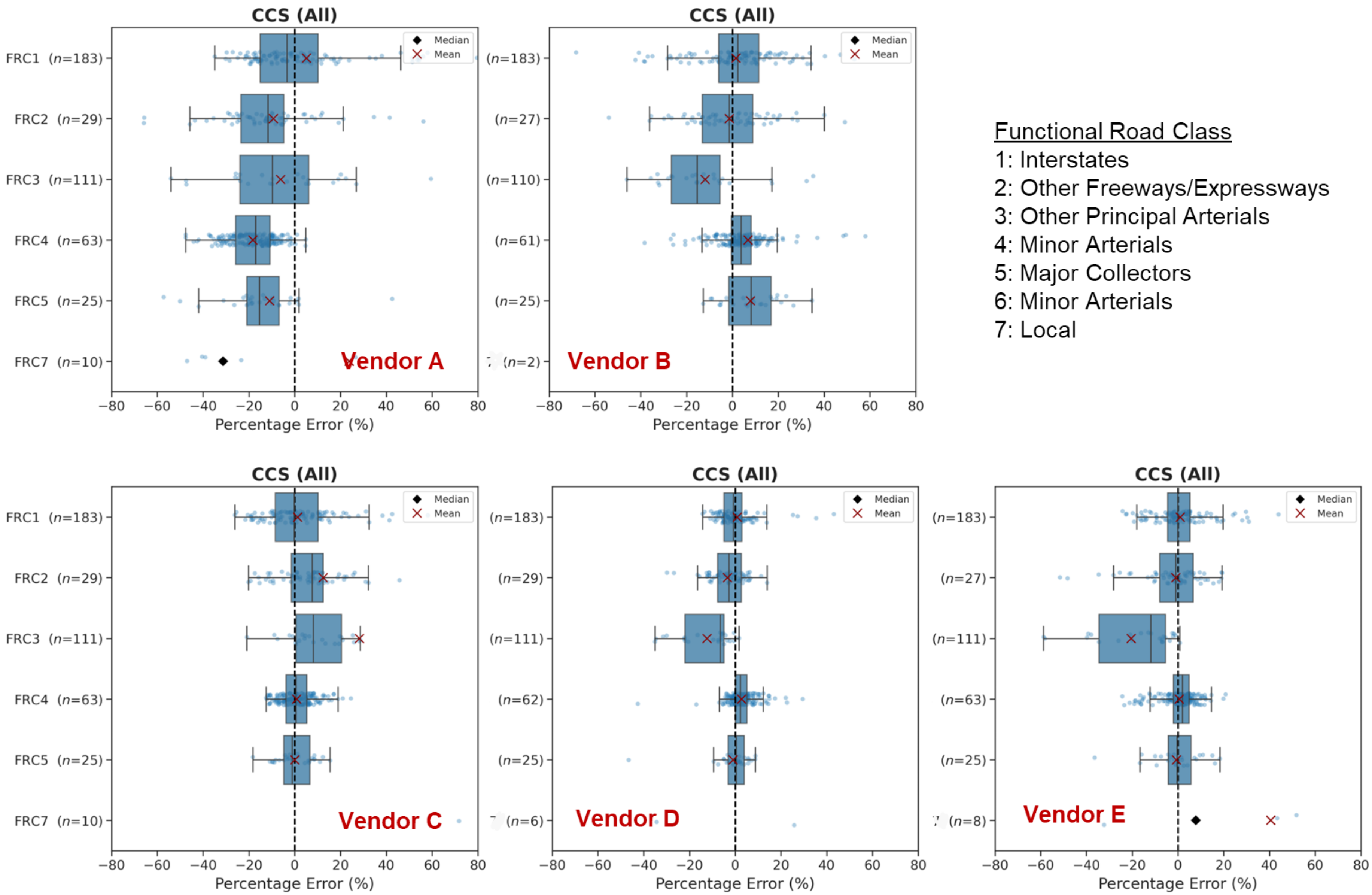


Figure 27: Percent Error boxplots by functional roadway classification (CCS locations)

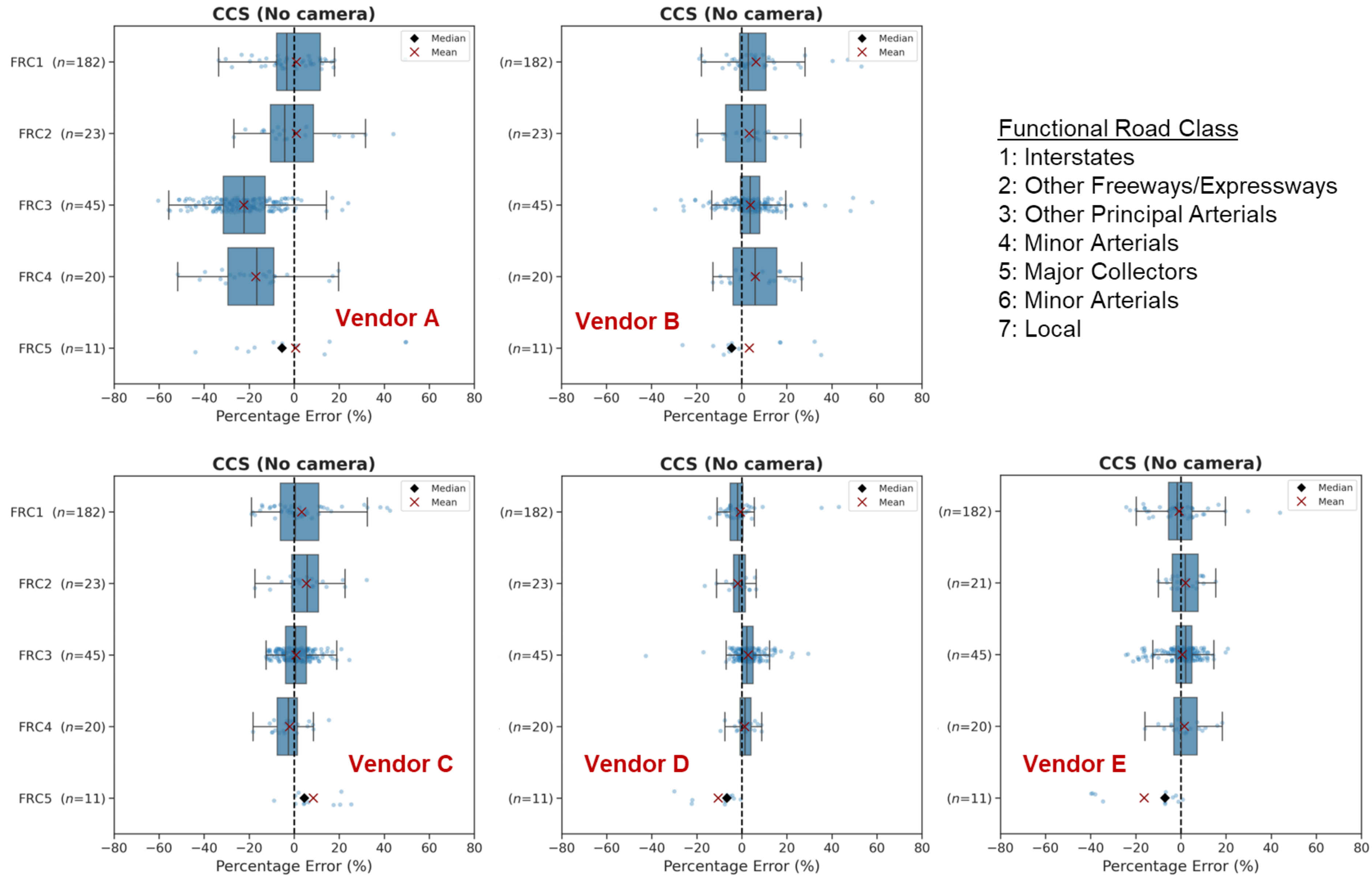


Figure 28: Percent Error boxplots by functional roadway classification (CCS locations with cameras omitted)

6.2 FHWA Test (Bi-Directional)

As described in Section 5, the FHWA evaluation method establishes accuracy benchmarks for four volume ranges (“Very low”, “Low”, “Med”, and “High) plus an overall precision test for the full dataset. These benchmarks – developed by FHWA in a prior pooled fund study -- reflect the expected precision/accuracy of factoring 48-hour counts to AADT values. Conceptually, the FHWA minimum accuracy guidelines for non-traditional volume measures reflect the anticipated accuracy of short term counts relative to continuous counters.

Before presenting the results, one caveat to be noted is related to the directionality of volumes. The Coalition volume validation methodology is based on *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. In contrast, FHWA’s statistical method is based on bi-directional volumes, using the total volume associated from both directions of traffic flow at a count site. As such, to ensure alignment with FHWA benchmarks, the validation team summed directional volumes at each site, both for reference and vendor datasets, to generate bidirectional results.

Noting this caveat, Table 5 summarizes the results using *bi-directional volumes for all CCS locations*, while Table 6 summarizes the results for *bi-directional volumes for non-camera CCS locations.* In both cases, 7 statistical tests are conducted: two accuracy metrics applied to three volume bins (low, medium, high), and one overall precision test. Note that the “very low” volume bin was omitted due to insufficient observations in this volume range. Benchmark limits are also shown to provide context for how close each vendor’s results were to the pass/fail threshold. A tally of the total number of tests passed is given to the right of the table.

All CCS Locations

Tables 5 depicts the result based on all CCS sites (including Camera sites). Results are presented for all three volume bins. Note that the ‘Low’ volume bin has only 20 observations, while a minimum of 25 observations are recommended. Three vendors passed at least 5 of the 7 tests; Vendors D and E passed 6 of 7 tests (5/6 accuracy, 1/1 precision), with Vendor C passing 5 of 7 (4/6 accuracy, 1/1 precision). All failures from these three vendors occurred in the lowest volume bin where the majority of camera count sites exist. Vendor B passed 4 of 7 tests (4/6 accuracy, 0/1 precision), with several outliers causing a failure on the precision test (limit = 8, observed outliers = 16). Vendor A didn’t pass any tests due to the issue with systematic underestimation.

Traditional CCS and AET Locations Only

Table 6 depicts the results if the Camera locations are removed from the data set. Results are presented for all three volume bins, although the ‘Low’ volume bin has only 11 observations, while a minimum of 25 observations are recommended. With camera CCS locations removed, **Vendors D and E passed all 7 tests** (6/6 accuracy, 1/1 precision), while **Vendor C passed 6 of 7 tests** (5/6 accuracy, 1/1 precision) and Vendor B passed 5/7 (5/6 accuracy, 0/1 precision). Vendor A passed 1/7 (1/6 accuracy, 0/1 precision) again due to the systematic bias issues. These results, with Camera locations removed, are deemed to be most comparable to the methodology used in previous validations from GA and NC.

Table 5 – FHWA method results for bi-directional AADT (All CCS)

			Accuracy Test #1			Accuracy Test # 2			Precision Test			
Vendor	Volume Category	# Obs	Absolute Median TCE	Limit	Result	MAPE	Limit	Result	# Failures	Limit	Result	Tests Passed
A	Low	20*	9.1	8.8	Failed	33.0	18.1	Failed	4	8	Failed	**0/7**
	Med	121	12.3	3.5	Failed	19.6	11.7	Failed	15			
	High	78	14.9	5.9	Failed	17.3	10.3	Failed	22			
B	Low	18*	0.8	8.8	Passed	19.7	18.1	Failed	1	8	Failed	**4/7**
	Med	120	3.2	3.5	Passed	18.8	11.7	Failed	11			
	High	78	2.4	5.9	Passed	7.4	10.3	Passed	4			
C	Low	20*	16.1	8.8	Failed	32.4	18.1	Failed	3	8	Passed	**5/7**
	Med	121	2.8	3.5	Passed	8.3	11.7	Passed	1			
	High	78	1.3	5.9	Passed	4.8	10.3	Passed	0			
D	Low	18*	3.0	8.8	Passed	25.2	18.1	Failed	1	8	Passed	**6/7**
	Med	121	0.7	3.5	Passed	4.6	11.7	Passed	0			
	High	78	2.6	5.9	Passed	4.4	10.3	Passed	2			
E	Low	20*	5.8	8.8	Passed	29.7	18.1	Failed	2	8	Passed	**6/7**
	Med	121	0.9	3.5	Passed	8.5	11.7	Passed	2			
	High	77	1.6	6.0	Passed	5.5	10.3	Passed	2			

Results are reported for 'Low' category although sample size is less than recommended 25

Table 6 – FHWA method results for bi-directional AADT (CCS, cameras omitted)

			Accuracy Test #1			Accuracy Test # 2			Precision Test			
Vendor	Volume Category	# Obs	Absolute Median TCE	Limit	Result	MAPE	Limit	Result	# Failures	Limit	Result	Tests Passed
A	Low	11*	12.4	8.8	Failed	14.6	18.1	Passed	0	6	Failed	**1/7**
	Med	61	13.1	4.8	Failed	15.8	12.4	Failed	4			
	High	78	14.9	5.9	Failed	17.3	10.3	Failed	22			
B	Low	11*	1.5	8.8	Passed	10.0	18.1	Passed	0	6	Failed	**5/7**
	Med	61	7.4	4.8	Failed	11.5	12.4	Passed	3			
	High	78	2.4	5.9	Passed	7.4	10.3	Passed	4			
C	Low	11*	13.3	8.8	Failed	13.6	18.1	Passed	0	6	Passed	**6/7**
	Med	61	2.4	4.8	Passed	7.1	12.4	Passed	1			
	High	78	1.3	5.9	Passed	4.8	10.3	Passed	0			
D	Low	11*	4.3	8.8	Passed	8.1	18.1	Passed	0	6	Passed	**7/7**
	Med	61	0.1	4.8	Passed	3.8	12.4	Passed	0			
	High	78	2.6	5.9	Passed	4.4	10.3	Passed	2			
E	Low	11*	5.1	8.8	Passed	11.0	18.1	Passed	0	6	Passed	**7/7**
	Med	61	1.6	4.8	Passed	6.2	12.4	Passed	0			
	High	77	1.6	6.0	Passed	5.5	10.3	Passed	2			

Results are reported for 'Low' category although sample size is less than recommended 25

Comparison with Prior Results

The prior validation in Georgia assessed AADT accuracy/precision using a similar, but slightly different implementation of the FHWA statistical method as was used in this report (see Section 5 for a description of the changes). To make the results for Georgia comparable, the validation team revisited the Georgia datasets and calculated FHWA statistical tests for bi-directional volumes using the current implementation (i.e., consistent with Tables 5-6 above).

Table 7 summarizes the bi-directional results based on CCS locations in Georgia for 2023 AADTs. Note that there were no camera counters in Georgia, so these results are most comparable to Table 6 (non-camera CCS locations in Massachusetts). The results are fairly consistent across all four vendors, ranging from 4-6 of the 7 tests passed, with Vendor D leading with 6 of 7 tests passed. Relative to these results in Georgia, we see that 4 of the 5 vendors improved – in some cases significantly. Vendor D has consistently passed the most tests in both validations (6/7 in GA, 7/7 in MA at non-camera locations), and the most notable increase came from Vendor E, which improved from 4/7 in GA to 7/7 in MA at non-camera locations.

Table 7 – FHWA method results for bi-directional AADT in prior GA study

			Accuracy Test #1			Accuracy Test # 2			Precision Test			
Scenario	Volume Category	# Obs	Absolute Median TCE	Limit	Result	MAPE	Limit	Result	# Failures	Limit	Result	Tests Passed
Vendor A	Low	27	0.5	8.5	Passed	8.9	18.0	Passed	0	6	Failed	**5/7**
	Medium	93	4.2	4.0	Failed	10.7	12.0	Passed	5			
	High	20*	6.1	9.4	Passed	10.4	12.9	Passed	2			
Vendor B	Low	27	8.1	8.5	Passed	16.8	18.0	Passed	1	6	Failed	**4/7**
	Medium	92	3.1	4.0	Passed	14.4	12.0	Failed	10			
	High	20*	3.1	9.4	Passed	19.8	12.9	Failed	5			
Vendor C	Low	27	1.4	8.5	Passed	149.5	18.0	Failed	5	6	Failed	**4/7**
	Medium	93	1.7	4.0	Passed	21.0	12.0	Failed	6			
	High	20*	2.0	9.4	Passed	2.2	12.9	Passed	0			
Vendor D	Low	27	3.0	8.5	Passed	10.1	18.0	Passed	1	6	Passed	**6/7**
	Medium	93	5.5	4.0	Failed	7.7	12.0	Passed	2			
	High	20*	2.3	9.4	Passed	4.7	12.9	Passed	0			
Vendor E	Low	27	1.3	8.5	Passed	44.7	18.0	Failed	6	6	Failed	**4/7**
	Medium	93	0.3	4.0	Passed	12.3	12.0	Failed	8			
	High	20*	2.5	9.4	Passed	7.0	12.9	Passed	1			

Results are reported for 'High' category although sample size is less than recommended 25

These results show a consistent improvement in the industry with the ability of most vendors to meet or nearly meet the accuracy thresholds recommended by FHWA.

Conclusions

This report explores the accuracy of AADT volume estimates derived from probe data, focusing specifically on products from five vendors who sell volume data in TETC's transportation data marketplace. The initial intent of this study was to characterize vendor performance across five different temporal aggregations, using metrics, statistical tests, and visualizations consistent with prior studies. However, initial AADT accuracy analysis uncovered high errors and major outliers, which were unexpected based on prior vendor performance. These errors warranted further investigation. This exploration resulted in rich insights into the various sources of uncertainty when using outsource probe-based volume estimates. For this reason, the MassDOT validation was separated into two reports. This first report focuses on the various sources of error and reports exclusively on analysis at the AADT accuracy, once the suspect locations are removed. The accuracy of other temporal aggregations will be reported in Part II. The following takeaways emerged from analysis of AADT data across vendors:

- **Agency counters and deployment locations are more diverse in Massachusetts than in prior validations**. Like prior validation studies, Massachusetts primarily uses inductive loop counters for its core Continuous Count Station (CCS) network. However, several of these sites are deployed near merge/diverge points, which can add ambiguity around which lanes are actually counted. Additionally, MassDOT has a sizeable inventory of intersection cameras deployed at signal-controlled intersections, intended primarily for traffic detection/control, but also used for counts. Relative to prior validations, *the inclusion of cameras is the most notable difference,* as it represents a new type of counting technology and deployment context.
- **Many sources of error and uncertainty can impact resulting accuracy measures, including several unrelated to vendor data quality.** Although some outliers are attributed to vendor estimation error, other sources of error were equally or more impactful to the results. Examples of such errors encountered in this study include (I) miscoded site locations (a data entry issue in the DOT database), (II) ambiguity in georeferencing, and (III) reference data error – none of which are reflective of vendor data quality.

  Issues associated with ambiguity in georeferencing/conflation are particularly challenging because both agency counts and vendor estimates may be correct within their geographic context but compare unfavorably due to incorrect or ambiguous georeferencing. For this study, CATT Works Georeferencing Protocol (CWGP) location references were used by the validation team to communicate evaluation locations to vendors. In certain complex locations (e.g., at merge/diverge points or arterial intersections with slip lanes), simply knowing the latitude, longitude, and direction of travel was not enough to interpret which lanes, or how many lanes were included in the counts. It is recommended that future validations include number of lanes in georeferencing as an added data quality check.
- **Camera counters at signalized intersections should be treated with caution**. In Massachusetts, cameras are used primarily for traffic control purposes and collect turning movement counts, from which directional volumes on each intersection leg can be derived. Although many intersection sites appeared to produce reasonable segment volumes and the DOT generally trusts the outputs, the data quality is less validated than traditional, inductive loop sensors. Generally, errors at camera locations were greater than those from inductive loop count sites. Whether these errors arise from fundamental

technology issues or from the complex environment at intersections where Cameras are typically deployed -- or a combination of the two – is currently unknown.

In this report, the validation team reports results both with and without camera counters included. Results from locations excluding Camera based technology are deemed most comparable to previous validation results from GA and NC.

- **Most vendors showed high accuracy after accounting for data issues, particularly when focusing on non-camera locations**. After removing the suspect data and processing vendors' final submissions, four out of the five vendors produced directional volume estimates that generally matched reference AADT values, particularly when excluding camera locations. Performance was evaluated under a variety of conditions – including across volume ranges and by functional road class (see Section 6 for full results).
- **Several vendors performed well on the FHWA statistical test, showing industry progression relative to last year's study.** The validation team re-implemented the FHWA statistical tests for AADT volume accuracy based on updated guidance, focusing on bi-directional volumes. For comparative purposes, last year's results from Georga were regenerated with the same methodology to serve as a point of comparison.

  The FHWA test is particularly important for agencies seeking to use AADT volume estimates for Federal reporting purposes. In last year's Georgia study, two vendors were close to meeting the FHWA guidelines for bi-directional AADT values. This year, when evaluating accuracy against continuous counters (but excluding camera locations), which is contextually comparable to prior studies, ***two vendors passed all components of the FHWA statistical test***, with two others approaching (nearly passing) the benchmark.

After accounting for the 'suspect' location sites, most of the vendor products showed strong agreement with reference counts – especially at non-camera locations. This is encouraging and indicates that industry is continuing to mature, proving higher quality results.

The validation experience in Massachusetts revealed that there can be several sources of error and beyond those related to vendor modeling/estimation. It is critical to understand the different sources of error, such as ambiguity in georeferencing/conflation that correlate with network complexity, location error within an agency database, and varying accuracy of different types count technologies. All of these are unrelated to quality of a vendors' volume estimation processes. This is a critical finding, as *these same types of errors will be factors in the use and application of industry sourced, probe-based data of any kind*.

Recommendations

The validation team recommends the following with respect to the methodology of evaluating third-party volume estimates:

- Continue base data collection with directional volume counts. Even though the FHWA method is based on bi-directional counts, the relative agreement of bi-directional counts at the same location for higher temporal aggregations (AADT, ADT) can help flag anomalies in both agency and vendor data.

- Cross-check results across multiple vendors. If the vendor data shows industry agreement, but disagrees with agency reference data, it tends to indicate an error source apart from vendor error.
- Systematically evaluate the outliers in a data set (such as the top ten) to assess if additional error sources are present in the analysis.
- Be mindful of the reference data count technology and context. Massachusetts introduced toll road counts and camera counts, both of which had performance peculiarities not previously encountered.
- When communicating validation locations to vendors, include additional fields to communicate the number of lanes and an indicator of lane type (e.g., mainline, ramp) to clarify the request.
- Apply FHWA Guidelines using bi-directional volume counts. Apply FHWA precision metric to the entire data set (not volume bins), unless sample sizes are ample.

Moving forward, the validation team will proceed with Part II of the analysis, focusing on the other temporal aggregations (ADT, AHDT, Daily, and Hourly volumes for May 2025) using the final submissions from the participating vendors. These aggregations will automatically exclude the suspect locations identified in this analysis and continue to investigate the differential performance at camera count locations.

Appendix A: Directional AADT Metrics by Vendor

The tables below summarize directional AADT error metrics for each vendor, broken out by scenario. Each vendor's results are reported in two tables; the first uses all locations, while the second reports results with Camera counters omitted.

The first table for each vendor initially reports results by Site Type (CCS vs STC), and then further breaks out CCS results by site classification (Traditional, AET, and Camera counters), Urban/Rural designation, AADT range, and road classification. The second table for each vendor (Tables A.2, A.4, A.6, A.8) is similar but focuses on non-Camera locations. Each table reports the following information:

- **# Obs**: Number of observations
- **P25, P50, P75 Error percentiles (vehicles):** Summarizes the signed error in units of volume (vehicles). P25/P50/P75 values correspond to summary percentiles that would be visually represented in a boxplot. P50 is the median error, which communicates the typical bias. A value of 0 indicates no bias, while negative values indicate underestimates and positive values indicate overestimates.
- **P50 Percent Error (%)**: Reports the median (i.e., 50th percentile) percentage error, expressed as a percentage. Similar to the error percentiles above, a value of 0 indicates no bias, while negative values indicate underestimates and positive values indicate overestimates.
- **P50, P90, and Mean APE (%)**: Absolute Percentage Error (APE) expresses the magnitude of error in percentage terms. Median (P50) APE shows 'typical' error, while P90 shows a reasonable "high risk" error (i.e,. 90% of errors are less than this value). Mean APE (i.e., MAPE) is also provided, as it is a commonly reported summary metric. Comparing mean and median (P50) APE values can be useful for identifying the presence of outliers.
- **NRMSE:** Normalized Root Mean Square Error

Vendor A

Table A.1: Vendor A Summary Metrics

Scenario		# Obs	Error (vehicles)			Pct Error (%)	Absolute Pct Error (%)			NRMSE
			P25	P50	P75	P50	P50	P90	Mean	
Site Type	CCS	421	-5756	-1998	-176	-13.0	16.6	38.5	21.4	25.7
	STC (factored)	247	-549	-108	252	-3.5	13.9	49.3	21.9	27.4
Site Class **(CCS)**	Traditional	255	-8299	-4138	-613	-13.4	14.2	28.9	16.0	22.0
	AET (Toll)	26	-13284	-5122	-3909	-26.9	26.9	36.6	22.2	23.0
	Camera	140	-1474	-338	584	-9.1	19.2	58.1	31.1	44.6
Urban/Rural **(CCS)**	Urban	378	-6245	-2271	-243	-13.1	16.8	40.8	21.9	24.7
	Rural	43	-2013	-339	-121	-11.0	15.1	31.8	17.7	44.4
AADT Range **(CCS)**	Very low (50-250)	7	-124	-70	-58	-40.5	47.0	97.4	57.8	60.6
	Low (250-2.5K)	40	-278	-75	217	-6.1	19.7	71.8	35.2	30.6
	Med (2.5k-27.5k)	224	-2404	-867	168	-11.1	16.1	42.4	20.3	29.9
	High (>27.5k)	150	-11089	-7784	-4930	-15.0	15.2	31.9	17.8	18.9
Functional Class **(CCS)**	1: Interstates	183	-10469	-6264	-4027	-17.0	17.0	31.9	18.4	19.9
	2: Other Freeways/Expwy	29	-5141	-2936	-688	-15.4	18.7	44.1	21.9	21.8
	3: Other Principal Arterials	111	-1553	-185	896	-3.4	13.9	51.4	20.4	35.9
	4: Minor Arterials	63	-927	-445	-132	-11.7	14.5	40.9	20.8	33.6
	5: Major Collectors	25	-514	-103	109	-9.8	20.1	47.4	21.1	33.0
	6: Minor Collectors	0	-	-	-	-	-	-	-	-
	7: Local	10	-94	-58	860	-31.2	71.8	206.9	92.8	128.7

Table A.2: Vendor A Summary Metrics with Camera Counters Excluded

Scenario		# Obs	Error (vehicles)			Pct Error (%)	Absolute Pct Error (%)			NRMSE
			P25	P50	P75	P50	P50	P90	Mean	
Site Class **(CCS)**	Traditional	255	-8299	-4138	-613	-13.4	14.2	28.9	16.0	22.0
	AET (Toll)	26	-13284	-5122	-3909	-26.9	26.9	36.6	22.2	23.0
Urban/Rural **(CCS)**	Urban	244	-9262	-4921	-1998	-14.1	14.4	31.1	16.6	21.1
	Rural	37	-2271	-358	-164	-12.9	14.7	30.3	16.9	43.2
AADT Range **(CCS)**	Very low (50-250)	21	-257	-164	-73	-10.8	12.1	25.7	14.3	16.5
	Low (250-2.5K)	110	-3905	-1946	-299	-12.1	13.4	29.8	15.5	22.6
	Med (2.5k-27.5k)	150	-11089	-7784	-4930	-15.0	15.2	31.9	17.8	18.9
	High (>27.5k)	182	-10478	-6323	-4034	-17.1	17.1	31.9	18.4	19.9
Functional Class **(CCS)**	1: Interstates	23	-5180	-3551	-1128	-13.8	13.8	25.6	14.5	20.0
	2: Other Freeways/Expwy	45	-1345	-411	192	-7.6	10.7	28.6	13.6	27.8
	3: Other Principal Arterials	20	-402	-209	-115	-7.5	10.0	16.2	10.3	18.5
	4: Minor Arterials	11	-349	-103	62	-9.8	14.4	26.0	15.5	14.7
	5: Major Collectors	0	-	-	-	-	-	-	-	-
	6: Minor Collectors	0	-	-	-	-	-	-	-	-
	7: Local	0	-	-	-	-	-	-	-	-

Vendor B

Table A.3: Vendor B Summary Metrics

Scenario		# Obs	Error (vehicles)			Pct Error (%)	Absolute Pct Error (%)			NRMSE
			P25	P50	P75	P50	P50	P90	Mean	
Site Type	CCS	408	-499	346	2059	2.8	8.1	28.0	13.8	18.7
	STC (factored)	247	-310	90	465	3.2	13.8	47.2	22.7	22.7
Site Class **(CCS)**	Traditional	255	-204	612	2590	3.3	6.6	20.7	9.5	14.0
	AET (Toll)	26	718	1759	3130	6.2	6.5	11.8	6.6	8.6
	Camera	127	-1183	-140	812	-1.9	16.0	36.5	23.8	66.2
Urban/Rural **(CCS)**	Urban	365	-570	398	2301	3.0	7.9	27.1	13.5	17.8
	Rural	43	-159	4	620	0.2	12.1	34.7	15.7	43.9
AADT Range **(CCS)**	Very low (50-250)	1	-51	-51	-51	-35.7	35.7	35.7	35.7	35.7
	Low (250-2.5K)	36	-105	18	195	0.8	9.6	29.3	19.1	33.1
	Med (2.5k-27.5k)	221	-570	309	1433	4.3	11.3	32.7	17.1	36.7
	High (>27.5k)	150	-901	1313	3520	2.8	5.5	15.6	7.4	11.3
Functional Class **(CCS)**	1: Interstates	183	-158	1428	3310	3.7	6.2	17.3	11.1	15.3
	2: Other Freeways/Expwy	27	-307	1301	3398	8.2	9.2	22.1	11.6	11.2
	3: Other Principal Arterials	110	-587	162	1002	2.4	10.1	33.8	14.1	18.9
	4: Minor Arterials	61	-704	-51	314	-1.4	11.8	28.4	14.7	28.8
	5: Major Collectors	25	-507	-303	-56	-15.4	17.2	34.3	20.2	37.9
	6: Minor Collectors	0	-	-	-	-	-	-	-	-
	7: Local	2	1671	2049	2426	155.7	155.7	162.5	155.7	168.9

Table A.4: Vendor B Summary Metrics with Camera Counters Excluded

Scenario		# Obs	Error (vehicles)			Pct Error (%)	Absolute Pct Error (%)			NRMSE
			P25	P50	P75	P50	P50	P90	Mean	
Site Class **(CCS)**	Traditional	255	-204	612	2590	3.3	6.6	20.7	9.5	14.0
	AET (Toll)	26	718	1759	3130	6.2	6.5	11.8	6.6	8.6
Urban/Rural **(CCS)**	Urban	244	-224	828	3013	3.8	6.2	17.4	8.4	12.4
	Rural	37	-95	238	971	3.1	8.5	36.0	14.9	43.0
AADT Range **(CCS)**	Very low (50-250)	21	-91	4	91	0.2	5.8	26.3	9.2	9.3
	Low (250-2.5K)	110	-62	665	2039	6.9	10.1	22.0	11.7	15.2
	Med (2.5k-27.5k)	150	-901	1313	3520	2.8	5.5	15.6	7.4	11.3
	High (>27.5k)	182	-173	1414	3271	3.7	6.2	17.0	8.3	12.2
Functional Class **(CCS)**	1: Interstates	23	-423	2058	3481	5.9	9.2	20.9	10.5	10.7
	2: Other Freeways/Expwy	45	-114	243	730	2.9	5.9	25.7	10.4	11.1
	3: Other Principal Arterials	20	-121	136	622	5.7	7.7	19.6	10.4	13.4
	4: Minor Arterials	11	-93	-56	253	-4.4	12.7	32.4	15.1	19.2
	5: Major Collectors	0	-	-	-	-	-	-	-	-
	6: Minor Collectors	0	-	-	-	-	-	-	-	-
	7: Local	0	-	-	-	-	-	-	-	-

Vendor C

Table A.5: Vendor C Summary Metrics

Scenario		# Obs	Error (vehicles)			Pct Error (%)	Absolute Pct Error (%)			NRMSE
			P25	P50	P75	P50	P50	P90	Mean	
Site Type	CCS	421	-906	129	939	1.3	6.8	20.5	16.7	9.2
	STC (factored)	247	18	698	1058	17.3	20.0	177.1	56.7	40.4
Site Class **(CCS)**	Traditional	255	-1521	39	1129	0.4	5.2	15.3	7.2	8.3
	AET (Toll)	26	298	1429	2667	5.6	5.6	9.5	5.1	5.4
	Camera	140	-312	272	676	6.0	10.8	34.0	36.2	15.3
Urban/Rural **(CCS)**	Urban	378	-1055	130	1005	1.1	6.7	19.4	17.5	8.9
	Rural	43	-32	125	340	4.6	8.1	21.5	10.0	13.6
AADT Range **(CCS)**	Very low (50-250)	7	690	691	693	397.1	397.1	477.4	399.0	390.7
	Low (250-2.5K)	40	59	233	558	20.1	20.7	78.2	35.6	27.1
	Med (2.5k-27.5k)	224	-439	186	882	2.0	8.0	18.9	9.1	11.8
	High (>27.5k)	150	-2741	-333	1438	-0.7	4.8	10.7	5.1	6.6
Functional Class **(CCS)**	1: Interstates	183	-1700	205	1724	0.6	4.8	11.2	5.5	7.0
	2: Other Freeways/Expwy	29	-1391	-189	353	-1.1	6.7	11.3	6.7	8.1
	3: Other Principal Arterials	111	-1063	-47	673	-0.3	8.9	21.7	10.9	14.2
	4: Minor Arterials	63	-86	292	561	7.6	10.5	24.9	18.2	15.0
	5: Major Collectors	25	10	154	487	8.1	12.0	27.3	31.5	15.9
	6: Minor Collectors	0	-	-	-	-	-	-	-	-
	7: Local	10	665	688	691	326.3	326.3	406.7	269.5	144.4

Table A.6: Vendor C Summary Metrics with Camera Counters Excluded

Scenario		# Obs	Error (vehicles)			Pct Error (%)	Absolute Pct Error (%)			NRMSE
			P25	P50	P75	P50	P50	P90	Mean	
Site Class **(CCS)**	Traditional	255	-1521	39	1129	0.4	5.2	15.3	7.2	8.3
	AET (Toll)	26	298	1429	2667	5.6	5.6	9.5	5.1	5.4
Urban/Rural **(CCS)**	Urban	244	-1588	62	1355	0.3	5.1	13.8	6.7	7.6
	Rural	37	8	115	334	4.3	7.3	21.0	8.8	13.1
AADT Range **(CCS)**	Very low (50-250)	21	39	129	256	11.9	11.9	31.2	13.8	19.6
	Low (250-2.5K)	110	-381	234	1164	2.3	6.2	17.0	8.2	10.4
	Med (2.5k-27.5k)	150	-2741	-333	1438	-0.7	4.8	10.7	5.1	6.6
	High (>27.5k)	182	-1705	220	1725	0.6	4.8	11.2	5.5	7.0
Functional Class **(CCS)**	1: Interstates	23	-1779	-696	258	-2.6	5.9	10.7	6.2	7.5
	2: Other Freeways/Expwy	45	-699	39	595	0.5	8.5	27.2	11.2	15.1
	3: Other Principal Arterials	20	-46	182	361	5.8	8.1	21.7	10.1	11.6
	4: Minor Arterials	11	38	88	149	4.6	6.4	21.0	10.1	7.3
	5: Major Collectors	0	-	-	-	-	-	-	-	-
	6: Minor Collectors	0	-	-	-	-	-	-	-	-
	7: Local	0	-	-	-	-	-	-	-	-

Vendor D

Table A.7: Vendor D Summary Metrics

Scenario		# Obs	Error (vehicles)			Pct Error (%)	Absolute Pct Error (%)			NRMSE
			P25	P50	P75	P50	P50	P90	Mean	
Site Type	CCS	416	-300	74	934	0.4	3.9	12.5	7.9	9.5
	STC (factored)	247	-579	-40	280	-1.8	13.4	44.6	19.5	24.3
Site Class **(CCS)**	Traditional	255	-237	143	1356	0.6	3.1	9.4	4.6	8.4
	AET (Toll)	26	865	2003	3735	6.8	6.8	13.4	6.7	9.2
	Camera	135	-439	-118	179	-2.8	5.6	25.5	14.4	10.7
Urban/Rural **(CCS)**	Urban	373	-300	119	1052	0.8	3.8	12.4	8.0	9.2
	Rural	43	-291	-93	2	-4.2	5.3	19.8	7.6	10.1
AADT Range **(CCS)**	Very low (50-250)	4	-75	1	137	-2.9	63.1	160.7	85.2	125.4
	Low (250-2.5K)	38	-153	-30	29	-2.1	6.0	34.8	26.0	55.6
	Med (2.5k-27.5k)	224	-408	-64	422	-0.7	4.1	11.3	5.6	8.2
	High (>27.5k)	150	-108	1043	2382	2.2	3.0	10.0	4.7	7.3
Functional Class **(CCS)**	1: Interstates	183	-14	740	1961	2.3	3.1	9.9	4.5	7.3
	2: Other Freeways/Expwy	29	-350	197	901	0.5	3.2	8.8	5.2	6.2
	3: Other Principal Arterials	111	-393	-57	393	-0.9	4.4	9.9	6.2	15.4
	4: Minor Arterials	62	-419	-116	80	-2.7	5.2	12.4	6.6	12.3
	5: Major Collectors	25	-413	-259	-53	-6.6	6.6	30.9	12.5	14.3
	6: Minor Collectors	0	-	-	-	-	-	-	-	-
	7: Local	6	-106	228	2002	107.3	141.9	279.4	150.4	358.9

Table A.8: Vendor D Summary Metrics with Camera Counters Excluded

Scenario		# Obs	Error (vehicles)			Pct Error (%)	Absolute Pct Error (%)			NRMSE
			P25	P50	P75	P50	P50	P90	Mean	
Site Class **(CCS)**	Traditional	255	-237	143	1356	0.6	3.1	9.4	4.6	8.4
	AET (Toll)	26	865	2003	3735	6.8	6.8	13.4	6.7	9.2
Urban/Rural **(CCS)**	Urban	244	-154	387	1625	1.5	3.1	9.7	4.6	8.2
	Rural	37	-267	-72	9	-3.5	5.2	10.3	6.3	9.6
AADT Range **(CCS)**	Very low (50-250)	21	-153	-45	-4	-3.5	5.8	22.0	7.2	10.6
	Low (250-2.5K)	110	-222	49	530	0.3	3.4	9.3	4.5	7.9
	Med (2.5k-27.5k)	150	-108	1043	2382	2.2	3.0	10.0	4.7	7.3
	High (>27.5k)	182	-19	720	1967	2.2	3.1	9.9	4.5	7.3
Functional Class **(CCS)**	1: Interstates	23	-129	274	1036	1.3	2.3	8.5	3.6	5.6
	2: Other Freeways/Expwy	45	-335	-114	83	-1.9	4.1	9.7	5.7	21.8
	3: Other Principal Arterials	20	-162	-21	39	-1.2	3.3	7.5	4.0	6.9
	4: Minor Arterials	11	-287	-114	-49	-6.6	6.6	22.2	10.5	13.1
	5: Major Collectors	0	-	-	-	-	-	-	-	-
	6: Minor Collectors	0	-	-	-	-	-	-	-	-
	7: Local	0	-	-	-	-	-	-	-	-

Vendor E

Table A.9: Vendor E Summary Metrics

Scenario		# Obs	Error (vehicles)			Pct Error (%)	Absolute Pct Error (%)			NRMSE
			P25	P50	P75	P50	P50	P90	Mean	
Site Type	CCS	417	-481	75	1079	0.7	5.0	20.7	10.9	11.7
	STC (factored)	239	-326	-26	523	-1.1	15.0	50.5	23.2	24.8
Site Class **(CCS)**	Traditional	253	-506	248	1326	1.1	4.3	16.1	6.8	9.6
	AET (Toll)	26	-1283	650	2189	2.7	4.6	10.8	5.7	9.6
	Camera	138	-436	-74	457	-1.2	9.2	44.8	19.4	28.3
Urban/Rural **(CCS)**	Urban	374	-513	149	1268	1.3	5.0	19.6	10.9	11.4
	Rural	43	-354	-96	6	-3.6	4.2	37.3	11.1	10.2
AADT Range **(CCS)**	Very low (50-250)	5	-140	-121	94	-84.6	84.6	92.0	72.4	70.4
	Low (250-2.5K)	40	-275	-101	3	-6.4	12.8	55.6	29.8	54.7
	Med (2.5k-27.5k)	224	-411	66	681	0.9	5.1	18.7	9.4	15.8
	High (>27.5k)	148	-1421	802	2218	1.7	4.1	15.8	6.1	8.3
Functional Class **(CCS)**	1: Interstates	183	-951	766	1886	2.0	4.0	12.8	5.7	8.3
	2: Other Freeways/Expwy	27	-707	-85	755	-0.3	4.2	16.3	7.7	8.0
	3: Other Principal Arterials	111	-435	-3	626	-0.1	5.1	19.7	8.6	18.0
	4: Minor Arterials	63	-283	-26	240	-0.9	6.8	33.4	16.3	56.8
	5: Major Collectors	25	-564	-272	-96	-11.8	11.8	50.9	20.5	30.1
	6: Minor Collectors	0	-	-	-	-	-	-	-	-
	7: Local	8	-144	-16	496	7.9	69.4	217.5	101.2	338.2

Table A.10: Vendor E Summary Metrics with Camera Counters Excluded

Scenario		# Obs	Error (vehicles)			Pct Error (%)	Absolute Pct Error (%)			NRMSE
			P25	P50	P75	P50	P50	P90	Mean	
Site Class **(CCS)**	Traditional	253	-506	248	1326	1.1	4.3	16.1	6.8	9.6
	AET (Toll)	26	-1283	650	2189	2.7	4.6	10.8	5.7	9.6
Urban/Rural **(CCS)**	Urban	242	-611	438	1690	2.0	4.4	15.7	6.4	9.3
	Rural	37	-250	-62	16	-2.9	4.2	28.3	8.6	6.3
AADT Range **(CCS)**	Very low (50-250)	21	-244	-62	2	-3.6	6.6	37.9	11.5	14.8
	Low (250-2.5K)	110	-321	282	790	2.1	4.9	15.4	6.7	8.4
	Med (2.5k-27.5k)	148	-1421	802	2218	1.7	4.1	15.8	6.1	8.3
	High (>27.5k)	182	-1021	774	1902	2.0	4.1	12.8	5.7	8.3
Functional Class **(CCS)**	1: Interstates	21	-686	113	1075	0.2	4.2	16.0	6.1	7.4
	2: Other Freeways/Expwy	45	-481	-46	563	-1.6	5.1	19.8	9.1	22.5
	3: Other Principal Arterials	20	-100	62	271	2.0	5.6	10.0	6.0	7.1
	4: Minor Arterials	11	-261	-115	-51	-7.0	7.0	39.2	16.3	17.1
	5: Major Collectors	0	-	-	-	-	-	-	-	-
	6: Minor Collectors	0	-	-	-	-	-	-	-	-
	7: Local	0	-	-	-	-	-	-	-	-