TDM Validation Activity: North Carolina Volume Study

TETC Transportation Data Marketplace Data Validation

Prepared by: TETC Validation Team

11/14/2023

TDM Validation Activity: North Carolina Volume Study
TETC Transportation Data Marketplace Data Validation

Follow-up validation activity for TDM Validation program, focusing on Volume data in North Carolina.

TDM-VAL-05

Data Categories: *Volume*

Evaluation Period: *Dec 2022- Jan 2023*

Publication Date: *November 2023 (updated Oct 2024[1])*

11/14/2023

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

[1] AADT volume metrics and visuals were regenerated to fix a data input error (impacting only AADT) that was discovered after initial publication. The impact to the results was minor, and did not change overall conclusions.

Executive Summary

This report describes the second TDM data evaluation effort conducted by The Eastern Transportation Coalition (TETC) data validation team and focuses on quantifying Volume data accuracy from four TDM vendors: HERE, INRIX, Iteris, and Streetlight. Building off the previous (inaugural) validation activity[2], which focused on testing methods, metrics, and logistics associated with the revised validation process, this follow-up activity extends the analysis to additional Volume data items and provides the first public release of vendor accuracy measures.

The error metrics and visuals used to evaluate volume data follow the approach taken in the inaugural activity, which was based on recommendations from a comprehensive literature review[3] and feedback from the TDM Validation Technical Advisory Committee (TAC). Whereas the initial volume validation activity focused on *hourly volume counts* - one of the four 'mandatory' volume data deliverables required per the RFP, this report considers all four volume aggregations specified in the TDM marketplace: *hourly volume counts, average hourly daily traffic (AHDT), average daily traffic (ADT), and average annual daily traffic (AADT).* The first three – hourly counts, AHDT and ADT – were the primary focus of the study and subject to blind evaluation (i.e., reference data was withheld from vendors prior to data submission), while AADT was also included in the analysis acknowledging that vendors had access to the majority of count data during the calendar year of interest. Additionally, vendors were required to provide responses to a questionnaire that provided clarification on their product characteristics.

The study area focused on all continuous count stations in North Carolina, which correspond to 263 directional evaluation locations. Vendors were asked to deliver four different types of volume data at these locations for specific date ranges; hourly counts were evaluated between 12/17/22 – 1/31/23 (capturing holiday travel), AHDT and ADT focused specifically on January 2023, and AADT targeted all of 2022. All four Volume vendors actively participated in the validation activity, with three delivering all of the mandatory data items requested. The remaining vendor delivered AHDT, ADT, and AADT deliverables, but declined to report hourly volumes.

Several key takeaways emerged from the resulting evaluation and follow-up discussions with vendors:

- **All 4 vendors were able to deliver data to support validation, reflecting a major milestone for industry.** Whereas this type of traffic data was in research and development pre-pandemic, it is emerging and maturing quickly with four vendors providing data to the TDM for evaluation across several hundred continuous count stations. Although there were some differences from vendor to vendor, all were able to report on a high majority of the CCS, in a standard method with minimal to no manual interaction with respect to geo-location, conflation, or interpretation of results. This is a major milestone for the industry.
- **A standardized set of visuals, metrics, and benchmarks are emerging.** Metrics and methods used to compare and contrast performance from vendor to vendor are beginning to 'gel', meaning that a standardized set of visualizations, metrics, and benchmarks are emerging from research and development for use in evaluating industry grade volume data. Although benchmarks will need to be set based on application needs,

[2] TDM-VAL-02: Initial TDM Validation Activity – Volume and Travel Time Study (link)
[3] TDM-VAL-01: Traffic Volume Validation – Literature Review and Recommendations (link)

the tool set (statistical measures and visualization) is converging to common practice. *The need to observe time series data* (i.e., plotting vendor-supplied data against reference data over time) *remains critical to understand trends and individual vendor characteristics* with respect to the volume estimate methodologies and reporting practices.

- **Industry appears to be in an early product maturity stage,** though it is expected that vendor products will evolve with changes in base data sources, algorithms, and reporting. Accuracy and fidelity of vendor supplied data are measurable, with moderate differences observed between vendors, and varying across reported data elements and volume ranges. The questionnaire revealed that most vendors (three of four) take a “top-down” approach to volume estimation starting with a more aggregate volume measure, typically ADT, and then disaggregating appropriately for other measures (and aggregating for AADT). One of the four vendors used a “bottom up” strategy, that is they directly estimated volumes for specific dates/time periods and then aggregated to other measures). The results were consistent with these findings, as the one “bottom up” vendor was the only one that produced meaningful temporal patterns at the hourly level, while the “top down” ones did not. Even the more aggregate data elements – the focus area for most vendors – had mixed performance, and for certain vendors, inconsistent accuracy between urban and rural locations.

Table of Contents

Introduction

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

This validation report evaluates one of the core data products available in the TDM: Volume data, which was not part of previous Coalition marketplace procurements and was identified as highest priority by the TAC for validation. Based on the TAC's direction to focus on Volume data, the validation team previously produced a report, TDM-VAL-1 (see footnote 3 above) containing a literature review and recommendations for validating volume data. The strategy outlined in TDM-VAL-1 guides the methodology used to evaluate vendor Volume data.

As part of the validation study, reference Volume datasets were collected and used as the basis for evaluating reported vendor Volume data. Whereas the inaugural TDM validation study focused on test methods, metrics, and logistics associated with the validation process, this study seeks to quantify data accuracy in more detail and provides the first opportunity to share vendor-accuracy measures publicly. The intent of this report is to provide initial insight into vendor accuracy for the different Volume products sold in TDM.

Data Vendors

All vendors selected through the TDM RFP process in the Volume category were invited to participate in the validation study (HERE, INRIX, Iteris, Streetlight). The validation team set expectations for data submission consistent with RFP specifications, including using a specific georeferencing protocol (CWGP) to describe validation locations, and requested four Volume data deliverables: Hourly Volume, Average Hourly Daily Traffic (AHDT), Average Daily Traffic (ADT), and Average Annual Daily Traffic (AADT). 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 with the Coalition, but here are labeled 'A/B/C/D'.

Geographic Scope

Figure 1 shows the geographic scope of the volume study area, which encompasses roadway locations across the entire state that correspond to NCDOT continuous count stations. Count locations are present on roads across all functional classes, with highest representation on interstates (FRC 1).

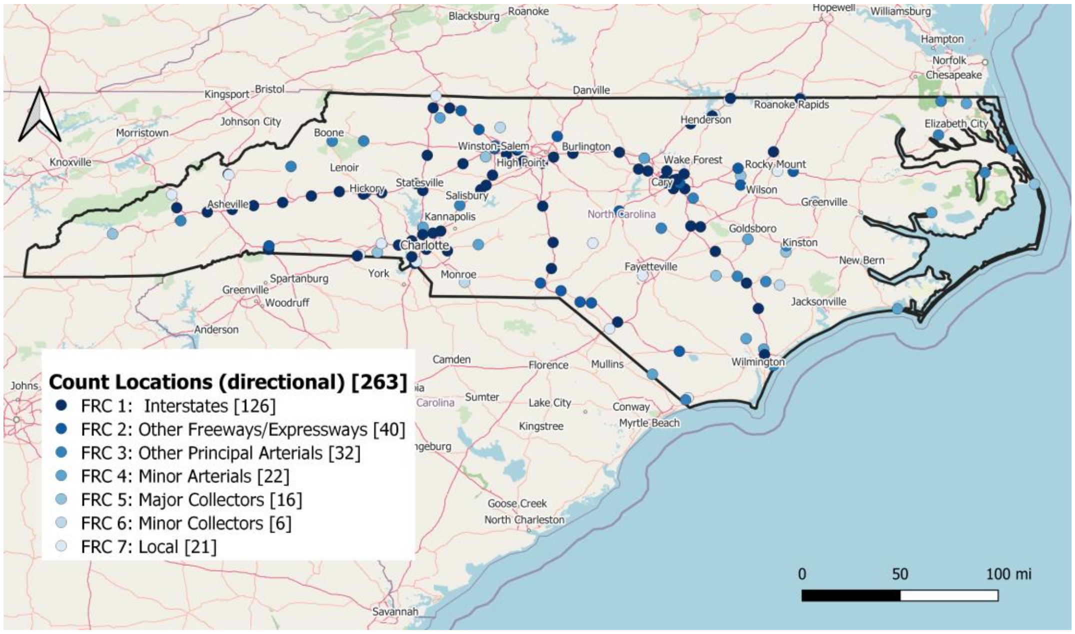

Figure 1 - Count location sites

Table 1 shows the key attributes used to communicate the volume evaluation locations to vendors via CWGP. Each Point location reference is defined by the latitude/longitude of the center-line of the road and its heading **for a specific direction of travel.** This means that a permanent count station that reports traffic counts in two directions would correspond to two CWGP point location references: one for each direction of travel. On divided highways, such as freeways, two distinct location references would represent the centerline of each set of lanes, along with two heading estimates approximately 180 degrees apart, one for each direction.

Table 1 - CWGP attributes for volume validation points

ID	Location	Heading	Road Name	Road Class
NC2022-P0033	-81.4299 35.17107	259	I-85	1
NC2022-P0034	-78.6315 34.34235	111	US 74	2

Data Collection

Reference Data

Reference hourly traffic counts were acquired by NCDOT during the study period of December 17, 2022 – January 31, 2023. During this period NCDOT shut off public access to the continuous count data, which is usually published online, thus enabling blind evaluation of results. Additionally, NCDOT provided 17 months of historical count data at the Continuous Count Stations (CCS), which was shared with vendors for model calibration purposes.

This reference count data was used to create four different reference datasets – consistent with the four mandatory data items from the TDM RFP. The four data elements used in the analysis are detailed below and summarized in Table 2, listed in order of decreasing temporal granularity.

Table 2 - Summary of data items and study periods

Data Item	Description	Study period	Reference Source
Hourly Volumes	Volume for a specific hourly period (e.g., 12/17/22 between 8-9am)	12/17/22 – 1/31/23	Continuous Count Stations (NCDOT)
Avg Hourly Daily Traffic (AHDT)	Avg hourly volume by day-of-week (e.g., Mondays between 8-9am)	1/1/23 – 1/31/23	Derived from CCS counts
Avg Daily Traffic (ADT)	Avg volume for a day during period	1/1/23 – 1/31/23	Derived from CCS counts
Avg Annual Daily Traffic (AADT)	Avg volume for a day during period	1/1/22 – 12/31/22	Derived from CCS counts

Hourly Volumes

Hourly volumes were provided directly by NCDOT for each location and did not require further aggregation. The date range used for evaluation (12/17/22 – 1/31/23) was chosen to include holiday periods where traffic may differ from typical weekly patterns. **Error! Reference source not found.** shows the distribution of hourly volumes across 250 locations and about 271k observations, which highlights the fact that there were many low-volume periods (half of all hourly observations have less than 350 vehicles per hour, with 90% less than 1000 vehicles per hour).

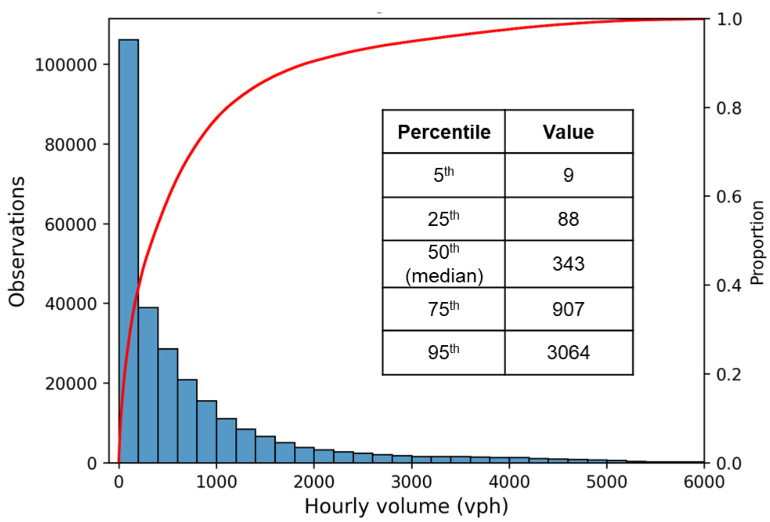

Figure 3 – Distribution of Hourly volumes.

Average Hourly Daily Traffic (AHDT)

AHDT values were obtained by grouping the hourly CCS counts in January 2023 by location, day of week, and hour of day, and then taking the arithmetic average of all counts in each group (e.g., average volume on Mondays between 8-9am). No holidays or anomalous days were excluded from the averages. Figure 3 shows the distribution of values across 250 locations and about 42k observations, with reported percentiles very similar to hourly values reported above.

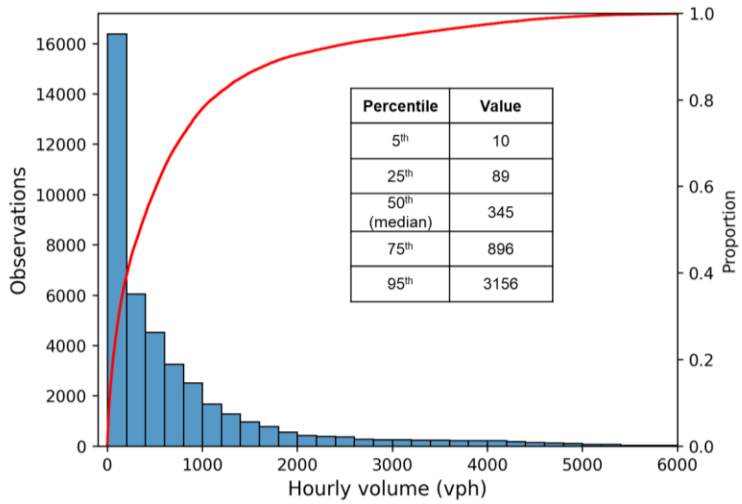

Figure 3 – Distribution of ADHT volumes.

Average Daily Traffic (ADT)

ADT values were obtained by grouping the hourly counts in January 2023 by location and computing the volume corresponding to an average day, following the FHWA-recommended method outlined on page 3-14 of the 2022 Traffic Monitoring Guide[4] (see MADT definition). This method produces identical results to a simple arithmetic average of daily counts when CCS stations records do not have missing data (as is generally the case with NCDOT CCS stations) but is well-suited to handle small gaps that may be present. Figure 4 shows the distribution of values across 250 locations, with one observation per location. Half of observed observations have AADT values under 11k with about a quarter under 5k.

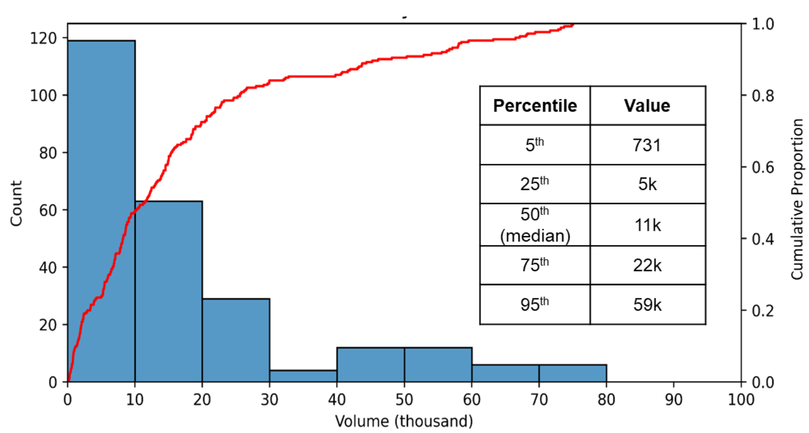

Figure 4 – Distribution of ADT volumes.

[4] 2022 FHWA Traffic Monitoring Guide (link)

Average Annual Daily Traffic (AADT)

AADT values were obtained by aggregating the hourly counts from all of 2022 using the FHWA recommended method. This approach, detailed on page 3-14 of the 2022 Traffic Monitoring Guide (see footnote 4), 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. As with ADT above, this method produces identical results to a simple arithmetic average of total counts for each day in the year when there is no missing data. **Note that hourly counts from 11 of the 12 months in 2022 were provided to vendors as calibration/training data, so AADT evaluation is not a blind study.** Nonetheless, the TAC recommended that this level of aggregation be included as a calculation exercise. Figure 5 shows the distribution of values across 261 locations, with one observation per location. Half of observed observations have AADT values under 14k with about a quarter under 7k.

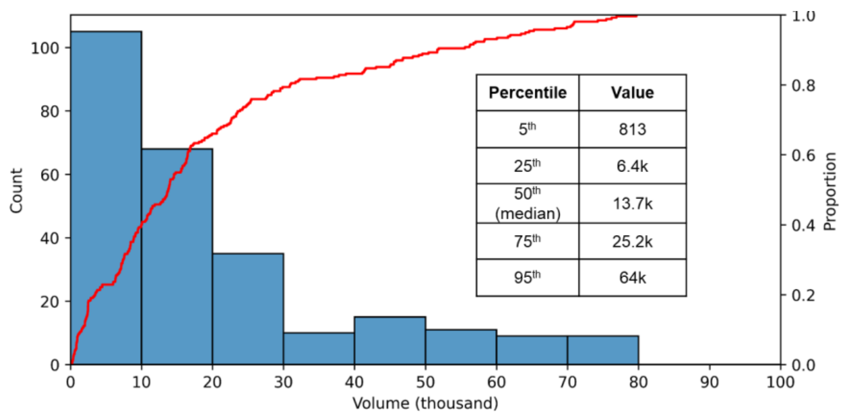

Figure 5 – Distribution of AADT volumes.

Vendor Data

Each vendor was instructed to prepare four Volume data deliverables corresponding to the data items listed in Table 2: hourly counts, AHDT, ADT, AADT. Additionally, they were asked to complete a brief questionnaire clarifying key aspects of their data products.

All four vendors actively engaged with the validation activity and delivered data within the specified timeframe and according to the specified data exchange format. Three of the vendors (A,B,C) delivered all requested data items, while the remaining vendor, D, delivered three of the items (AHDT, ADT, AADT). Vendor C was unable to provide data on roads with functional class 6 and 7 (minor collectors and local roads).

Questionnaire Responses

The focus of the Vendor Questionnaire was to understand how each vendor obtained volume estimates for each data element (AADT, ADT, AHDT, Hourly). In particular, it sought to understand which elements were modeled directly via statistical estimation, and which were calculated (or derived) by aggregating or disaggregating modeled values.

Table 3 summarizes the questionnaire responses for each vendor, which highlights that three of the vendors use a “top down” approach to volume estimation; they estimate volumes at a more aggregate level (typically a variation of ADT at the monthly level), and then generate more granular data items (AHDT and/or Hourly Volumes) afterwards using some form of a disaggregation algorithm. In contrast, one vendor uses a “bottom up” approach, starting by estimating sub-hourly volumes and then rolling these estimates up to more aggregate measures.

Table 4 - Summary of questionnaire responses

	Vendor A	Vendor B	Vendor C	Vendor D
AADT				
1a. Is the statistic directly modeled or calculated?	Calculated	Calculated	Calculated	Calculated
1b. If calculated, what AADT method is used?	Avg of monthly ADT	Avg of quarterly ADT	Simple avg of daily volumes	FHWA (Modified)
1c. Are any days excluded from average?	No	No	No	No
ADT				
2a. Is the statistic directly modeled or calculated?	Modeled	Modeled	Calculated	Modeled
2b. Are any days excluded from average?	No *Modeled for 4 day types, then combined to single value*	No	No	No
AHDT				
3a. Is the statistic directly modeled or calculated?	Modeled	Calculated	Calculated	Calculated
3b. Are any days excluded from average?	No *Hourly avg modeled for 4 day types, then applied to each DOW*	No	No	No
Hourly				
4a. Is the statistic directly modeled or calculated?	Calculated	Calculated	Modeled *Modeled for 15-min periods then summed to hourly*	N/A
Summary				
Core modeling/estimation focuses on:	Variations of **ADT** and **AHDT**	**ADT**	**Hourly** (or sub-hourly)	**ADT**

NOTE: The terminology "m*odeled*" vs "*calculated*" refers to whether a data element is estimated directly from a model ("modeled") or is calculated from a model output at a different level of temporal granularity ("calculated").

Evaluation Methodology

Vendor data quality was evaluated separately for all four deliverables (hourly volumes, AHDT, ADT, AADT) using error metrics and in some cases, visual inspection – an approach based on recommendations from the Volume Validation technical memo TDM-VAL-1 (footnote 3) and inaugural volume validation report TDM-VAL-2 (footnote 2). These approaches are detailed below.

Error Metrics

The following common error metrics are used to quantify differences between vendor and reference volume values, where V_R is the reference volume and V_V is vendor volume:

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

The first three metrics are relevant for all data items (hourly volumes, AHDT, ADT, AADT), whereas EMFR only makes sense for hourly volumes and AHDT because it involves normalizing the error relative to the maximum observed volume at a location. Unlike hourly volume and AHDT, which have about 1100 and 168 records per location, respectively, ADT and AADT have only one value per location.

For each metric, the mean and various percentiles 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 APE, referred to as MAPE, communicates the average absolute error expressed as a percentage of the reference volume, whereas a summary of APE percentiles (eg., 25%, 50%, 75%) communicates how much the values vary (often expressed visually in a box-and-whisker plot). In most cases, it makes sense to group aggregated results by volume bin, road class, or another data characteristic of interest. In particular, percentage-based errors are sensitive to low volumes, so separating results by reference volume range can provide useful context for interpreting results.

FHWA Method

Additionally, the validation team consulted the recommended FHWA methodology [5] for evaluating the accuracy and precision of AADT estimates obtained from emerging technologies. Their approach uses rank-choice statistics to establish confidence intervals around the median percentage error (measure of accuracy) as well as 2.5% and 97.5% absolute percentage errors (measure of precision). The idea, then, is to compare the resulting accuracy and precision values to see if there are better than those associated with scaling up 48-hour counts to AADT values (obtained from previous FHWA research). Unfortunately, the sample size required to implement this method (93 locations per volume bin) was much larger than our AADT dataset; only one volume bin met this criterion, even when using very large volume bins.

[5] FHWA-PL-21-031 (2021) "Guidelines for Obtaining AADT Estimates from Non-Traditional Sources" (link)

However, the FHWA method provides an alternative, simpler approach that can be used when sample sizes are too small. The alternative approach involves computing specific percent error (PE) and absolute percent error (APE) percentiles – plus two aggregate error measures, MAPE and Normalized Root Mean Square Error (NRMSE). As before, FHWA provides target values associated with the errors that result from estimating AADT values from 48-hour counts. This simplified approach is consistent with our previously listed metrics, albeit with the addition of NRMSE. NRMSE is similar to MAPE (the average APE), but aggregates the error in a way that involves squaring the error term, thus placing a more severe penalty on large errors.

- $\textbf{Normalized Root Mean Square Error (NRMSE)} = \frac{\sqrt{\text{avg}((V_V - V_R)^2)}}{\text{avg}(V_R)} \cdot 100$

Visual Inspection

While ADT and AADT data items involve only one data point (and thus one opportunity for comparison) per location, hourly volumes and AHDT include about 1100 and 168 data points, respectively. As such, these more granular data items provide opportunities to investigate temporal patterns – both recurrent (AHDT, focusing on time-of-day patterns by day of week) and non-recurrent (hourly volumes, which can capture perturbations in traffic).

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

Results and Discussion

The results are reported first for the three data elements for which blind evaluation was possible – ADT, AHDT and Hourly (in order of increasing temporal granularity), followed by AADT. Aggregate temporal scales such as ADT and AADT represent simpler data products, and thus are a good starting point for benchmarking accuracy. Subsequent increases in temporal granularity – AHDT and then hourly volumes, involve increasing complexity and in some cases mark a switch in emphasis from planning to operational applications.

Although a full set of results for all error metrics were initially produced, the following sections will focus on key visuals with and percent error plots. Percent error was chosen because it is easily interpretable across different levels of traffic, and unlike APE or EMFR, it captures the sign of the error to highlight positive and negative biases.

ADT

Figure 2 plots vendor ADT volumes (y axis) against reference ADT values (x axis) obtained from NCDOT CCS data for each vendor. The dashed red line represents points where there is zero error (i.e., estimates are equal to reference values), while the dashed black lines represent +/- 15% error. In general, all vendors consistently produced estimates that fell within the 15% bounds, with some slight differences between vendors. It should be noted that the 15% bounds are visually indistinguishable from the 45-degree line at very low volumes, meaning that ADT estimates can easily have sizeable percentage errors even when vendor estimates match the reference data well. As such, it is important to consider the reference volume level when using percentage-based errors to measure quality.

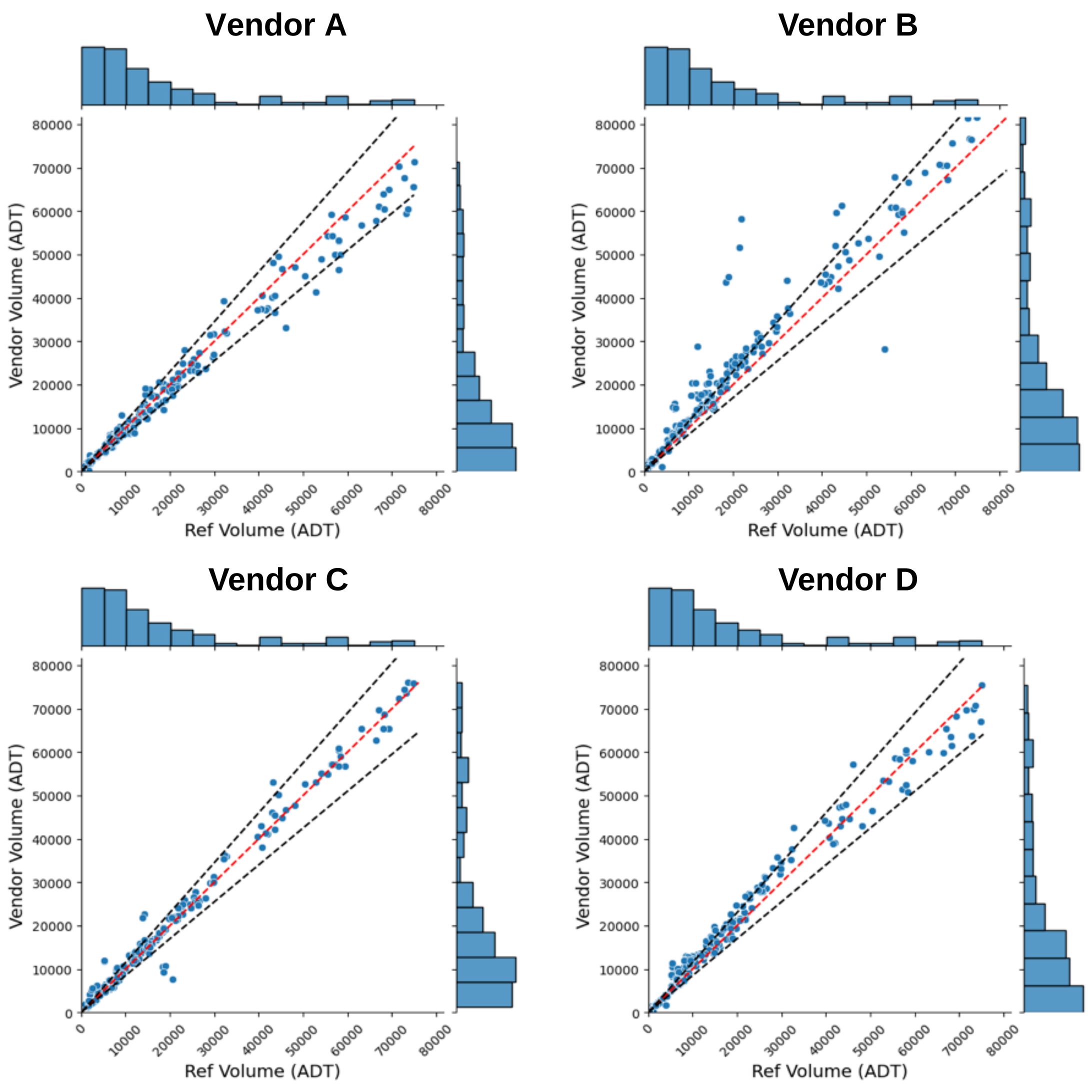

Figure 2 – ADT scatter plot plots by vendor

Figure 3 shows the distribution of ADT error percentages for each vendor via boxplots, separated by volume bin. For Vendors A-C, the magnitude of errors were consistently high at low volumes (especially under 5k), with precision and accuracy tending to improve at higher volume levels. This is consistent (as expected) with statistical theory when the volume estimate is based on a fixed percentage of observations (typically 2% to 10% for probe data). In contrast to Vendors A-C, Vendor D showed reasonable accuracy at lower volumes too – generally within 15% error under 5k AADT. Figure 3 also highlights that the direction of error sometimes "switches" across error bins, as shown by Vendor D, and to a lesser extent, B. In contrast, Vendors A and C overestimate lower volume bins and improve steadily in the same direction at higher volumes (with A overcorrecting at higher volumes). Finally, it should be noted that Vendor C does not have data in the lowest volume bin because all locations in this AADT range are functional class 6-7, which were not included in their delivery.

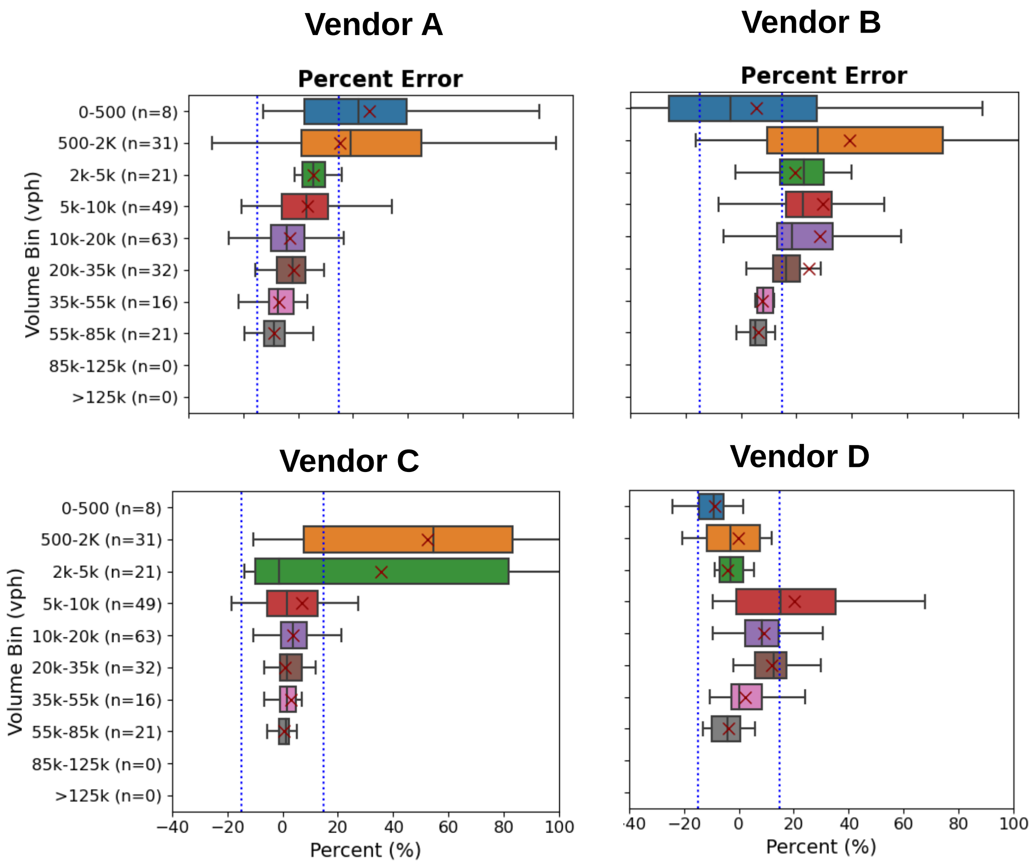

Figure 3 - Distribution of ADT percent error metrics by volume bins

Urban vs Rural

Recognizing that the results from Figure 9 encompass both urban and rural locations, it can be instructive to separate the results to determine whether the results are consistent across location types. Using NCDOT's classification for each count station, Figure 10 shows the results separated by urban/rural location for each vendor, with urban results on the left and rural results on the right. When interpreting the results visually, note that the urban and rural bar charts for a given volume bin may be based on different sample sizes.

Vendor D appears to be most consistent across urban/rural locations, with Vendor A also showing similar performance across location types. In contrast, Vendors B and C show a significant degradation in performance in performance for rural locations, relative to their urban counterparts. The stark difference in precision and accuracy warrants further investigation in future studies.

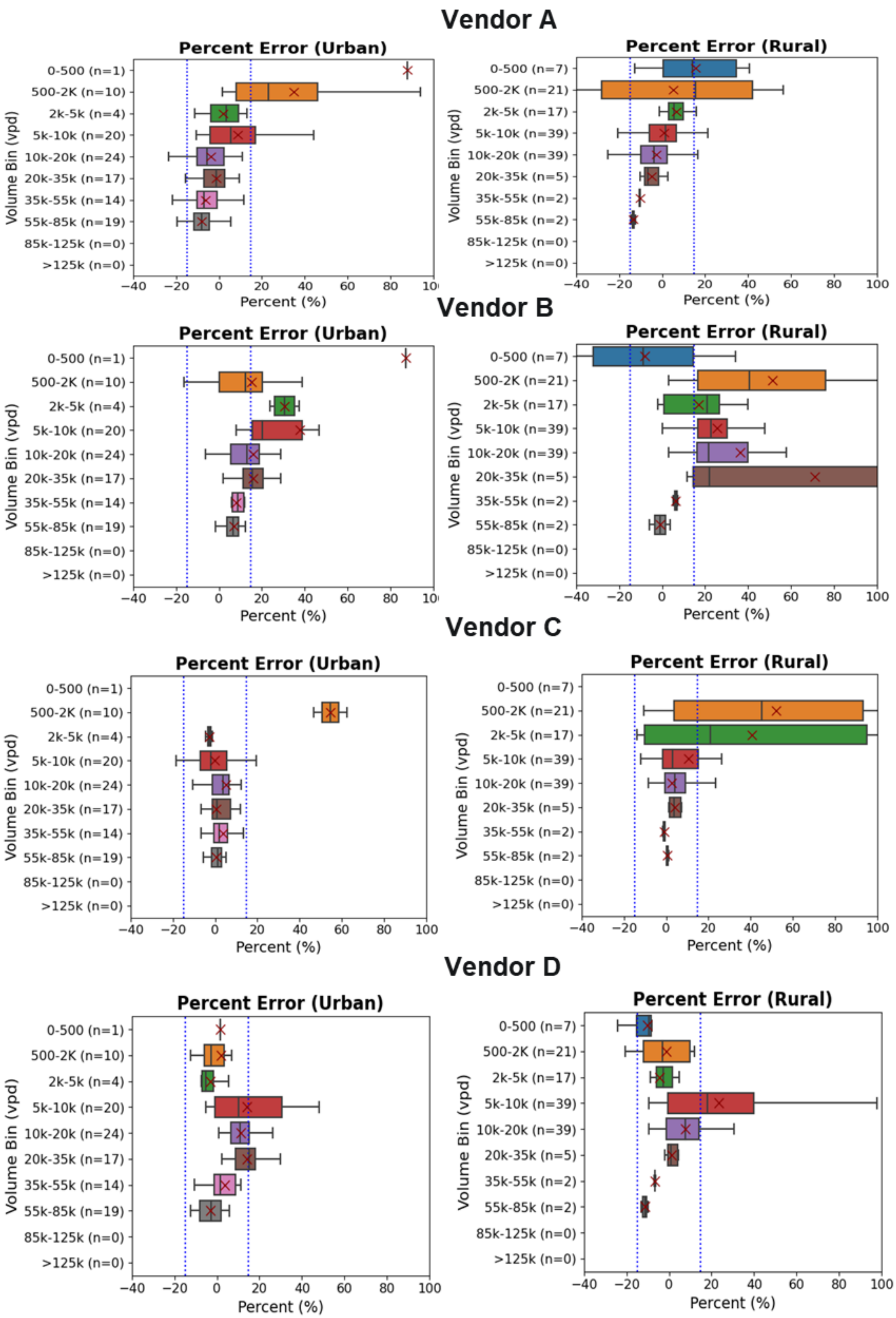

Figure 10 - Distribution of ADT percent error metrics by volume bin for Urban/Rural locations

AHDT

AHDT values represent a step up in temporal granularity from ADT, with 168 hourly records at each location rather than one (24 hours * 7 days). This enables an analyst to visually inspect time series plots, which can help build an intuition for how well vendor data captures recurrent patterns. Inspecting plots at dozens of locations showed that while it is possible to find several positive and negative examples for each vendor, there are noticeable differences in how each vendor handles temporal patterns by day of week.

Error! Reference source not found.-14 provide a representative example for each v endor at a location that shows an AM peak signature during weekdays. Figure 11 (Vendor A) shows that Vendor A is not sensitive to the temporal patterns by day of week; it repeated a very similar temporal pattern each day of the week, but did not capture the changing traffic dynamics over the weekend. In contrast, Figures 12-14 (Vendors B-D, respectively) are more responsive to time-of-day and day of week patterns, albeit with some observed error during specific days (Vendor B) or peak periods (Vendor C).

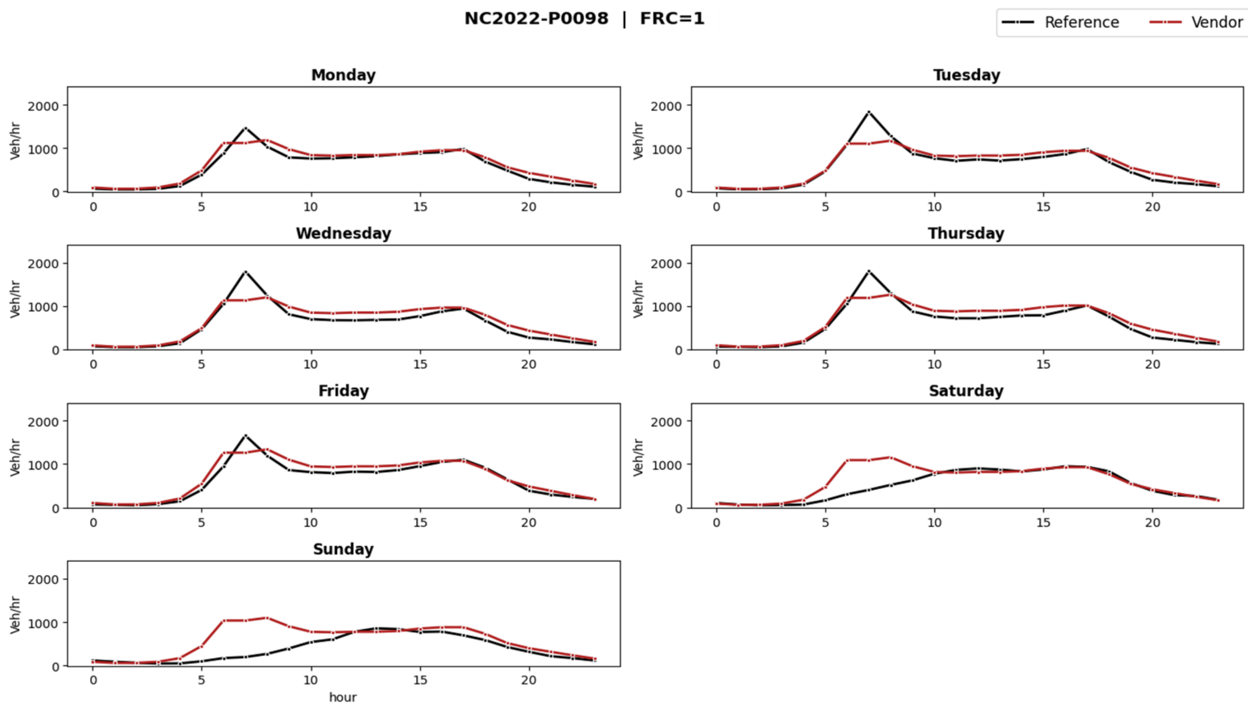

Figure 11 – AHDT time series by day of week for Vendor A

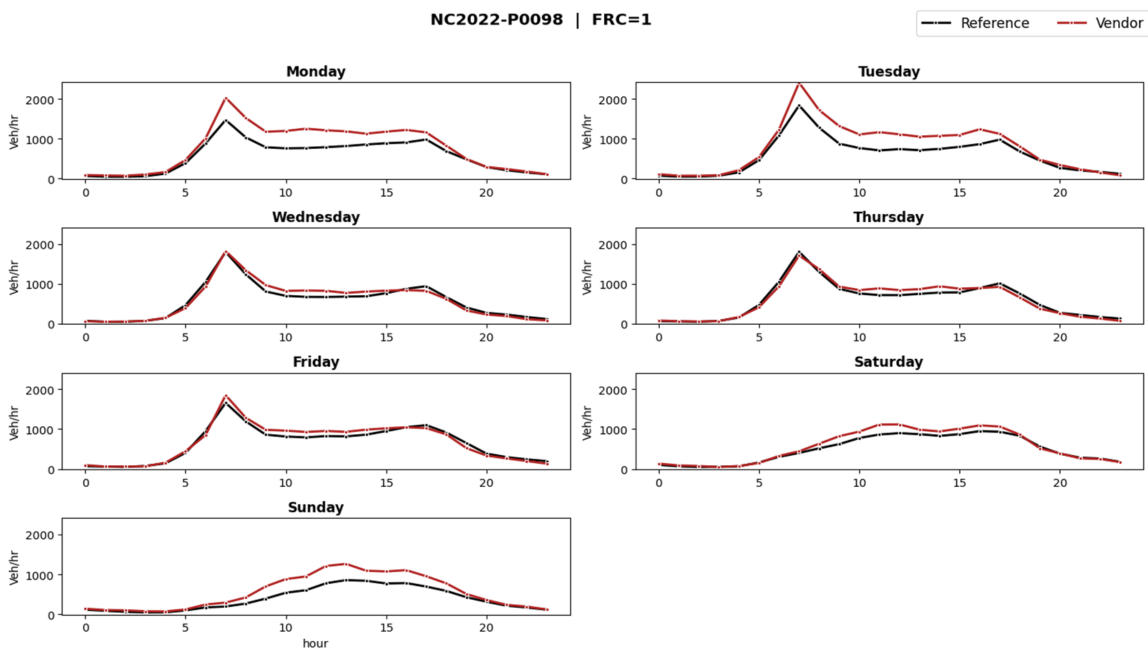

Figure 12 – AHDT time series by day of week for Vendor B

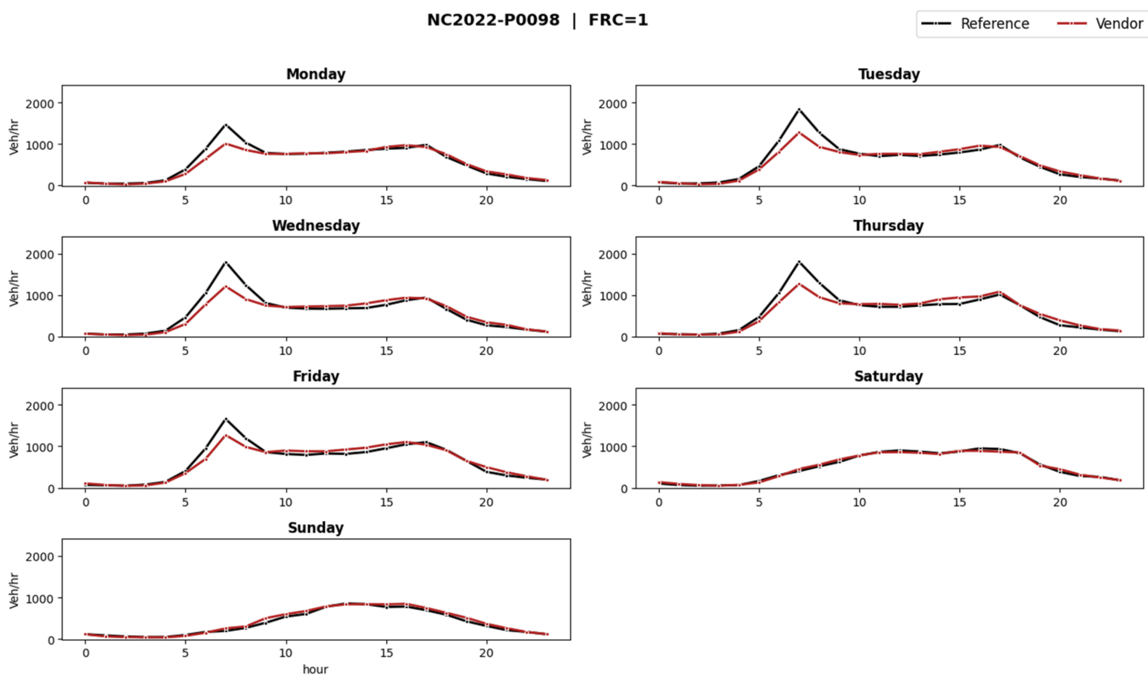

Figure 13 – AHDT time series by day of week for Vendor C

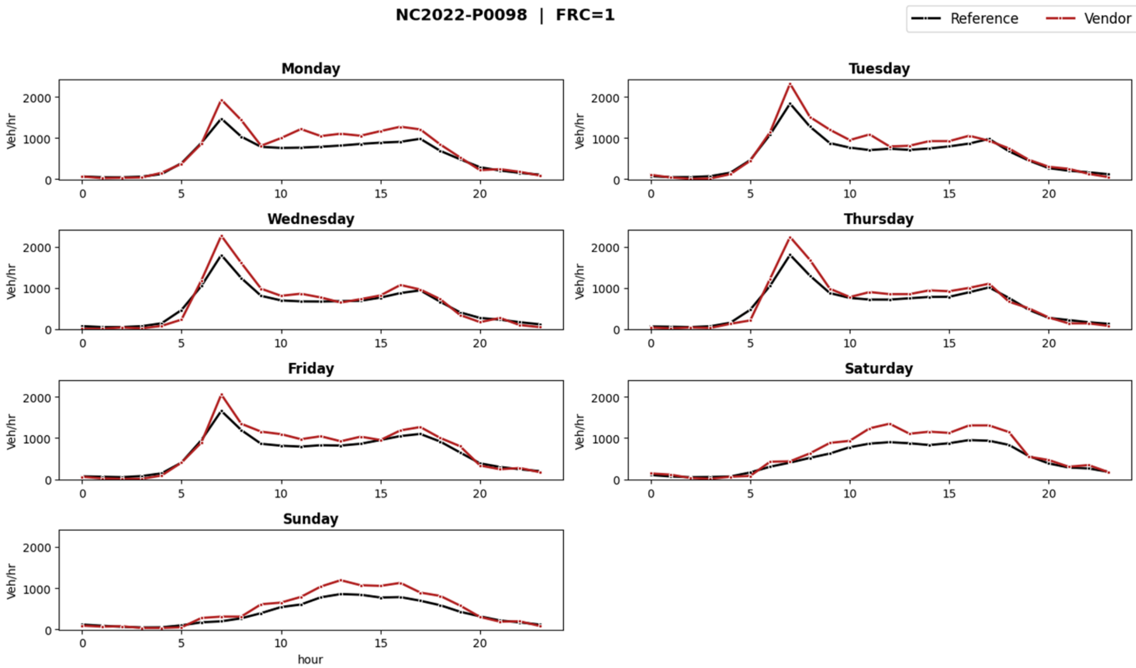

Figure 14 – AHDT time series by day of week for Vendor D

Figure 15 shows vendor AHDT volumes (y axis) plotted against reference AHDT values (x axis) that were obtained from NCDOT CCS data - similar to above for ADT. As before, the 45-degree line represents perfect estimation, while the dashed black lines show +/- 15% error. However, since there are about 42k observations in the AHDT dataset, the data is visualized using a hex-bin plot with darker colors indicating more observations in a region, which reinforces the fact that the majority of observations are quite low volume (see **Error! Reference source not found.** a bove). All four vendors generally track the 45-degree line, although there are differences in how consistently they are within the 15% band. At least two vendors have instances where values were about 2x or ½ times the reference volume, suggesting that there may be directionality issues (e.g., summing both directions of traffic rather than reporting each separately).

Figure 316 shows the distribution of AHDT error percentages for each vendor via boxplots, separated by volume bin. Vendors C and D appear most consistently precise and accurate across volume levels, with B showing consistent overestimation in most of the lower bins. While the overall error metrics may be a useful way to summarize accuracy, it is important to interpret this information with the time series plots in mind, as the summary metrics alone do not tell the whole story – particularly for a data element that has temporal patterns.

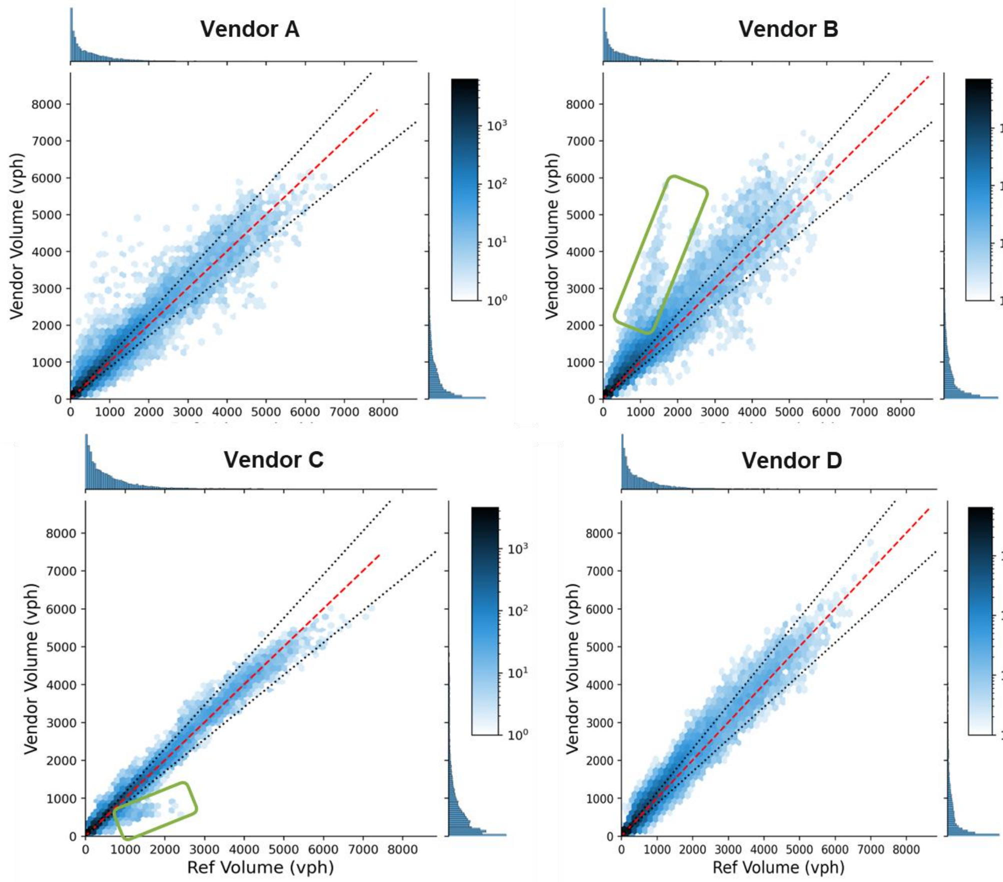

Figure 4 - AHDT scatterplots by vendor

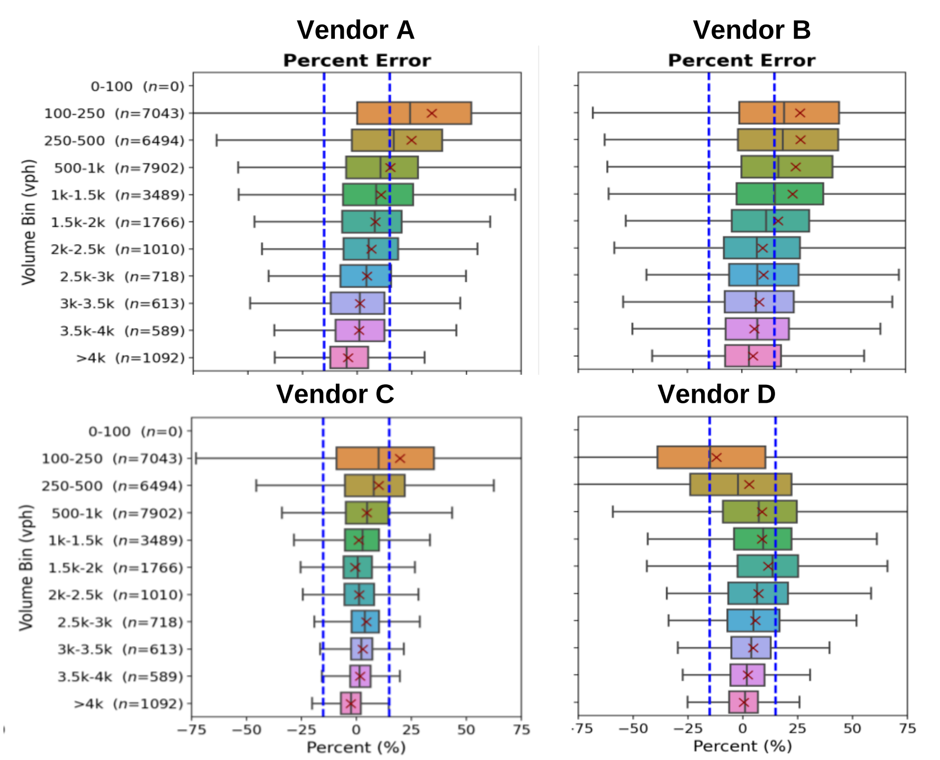

Figure 5 - Distribution of AHDT percent error metrics by volume bin

Hourly Volumes

Hourly volumes represent the most temporally granular data item evaluated. Unlike AHDT, which focuses on temporal patterns repeated across multiple days, hourly volumes allow an analyst to investigate performance during anomalous dates and periods. Manual inspection of time series plots proved most useful for gaining an intuition into how well vendors tracked reference data across time.

Of the three vendors that delivered Hourly data, only one (Vendor C) was attuned to non-recurrent fluctuations in volume. The other two vendors repeated the same time series pattern for each week in the month, and thus did not capture fluctuations that deviated from typical patterns. Figures 17-19 illustrate this finding during the week of Christmas 2022; Vendors A and B (Figures 17 and 18, respectively) show a repeating time series pattern that does not change during Christmas day (i.e., volume is reported identically as the previous Sunday), whereas vendor C (Figure 19) identifies the drop in Volume. These results are consistent with the survey findings reported earlier, which showed Vendor C to be the only one to directly estimate volumes at the granular hourly/sub-hourly level. However, it should be noted that Vendor C – despite being able to track hourly reference volumes most dynamically – had several instances where there was significant error.

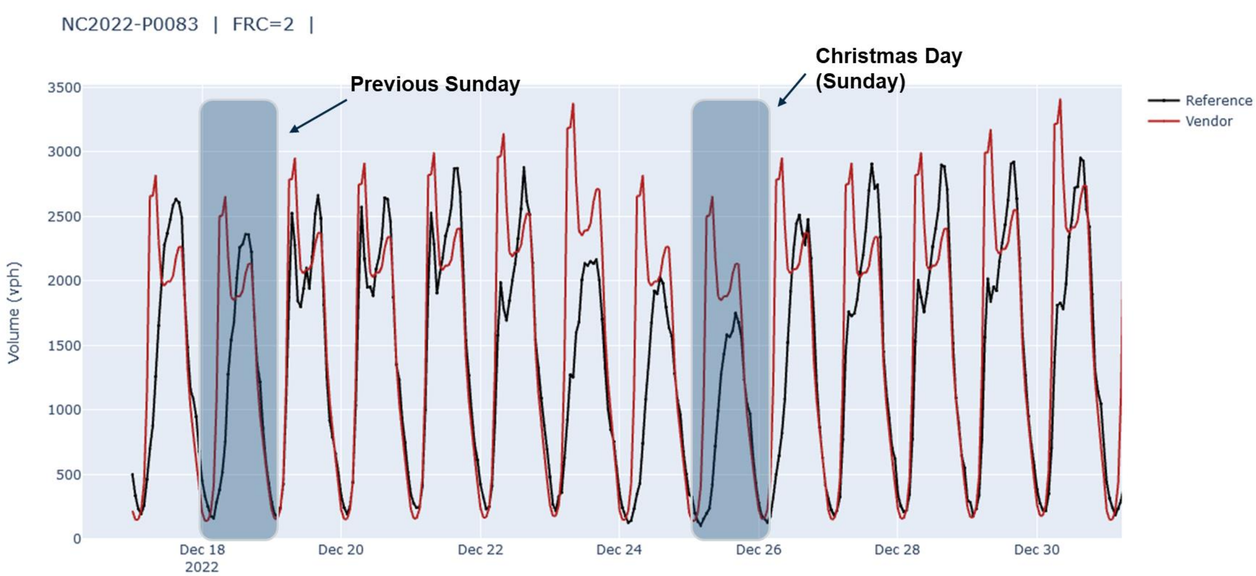

Figure 67 – Hourly time series plot for Vendor A

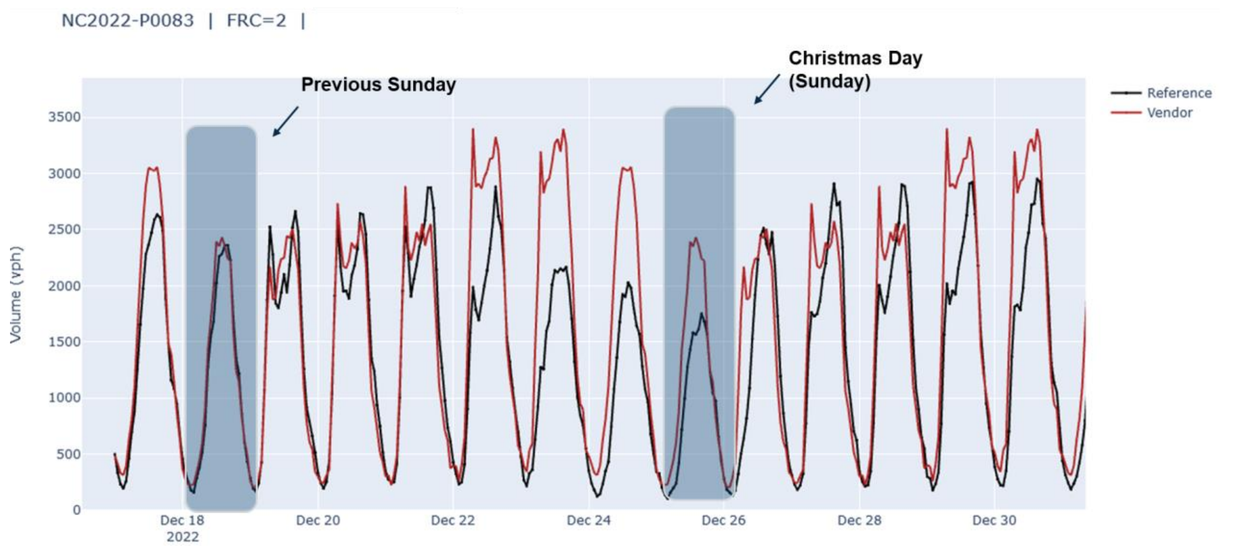

Figure 78 – Hourly time series plot for Vendor B

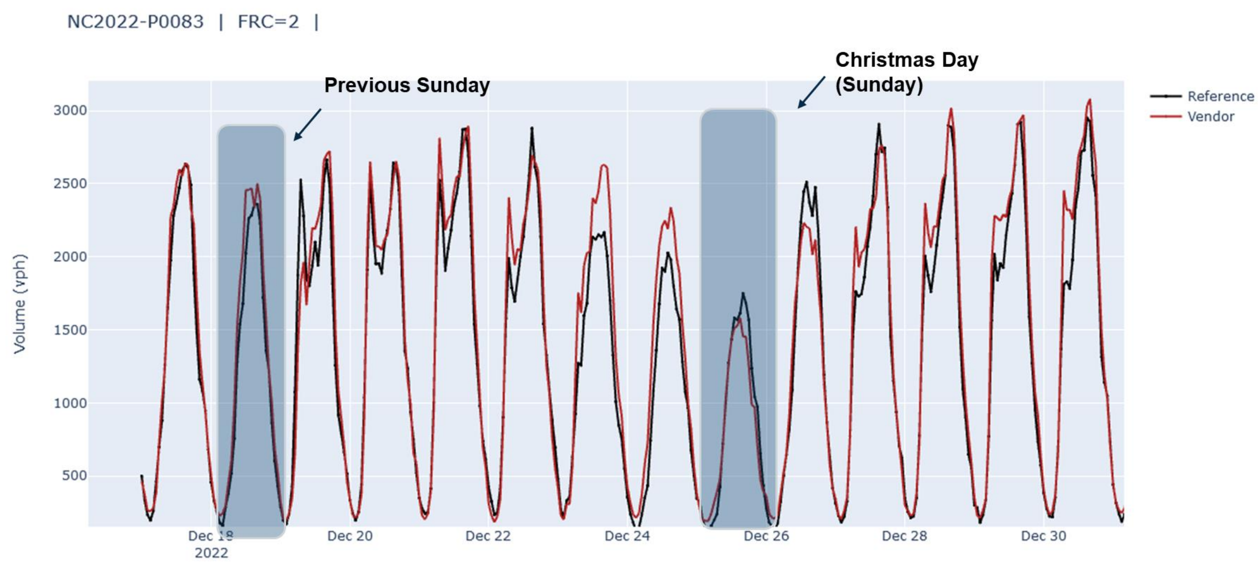

Figure 89 – Hourly time series plot for Vendor C

Figure 20 shows hourly vendor estimates plotted against reference values for each of the three vendors, with darker values representing a higher density of observations. This plot highlights that most of the hourly volumes are very low (dark color), and that the majority of Vendor C's visible points are tightly clustered within the +/- 15% boundaries. The other vendors track the general shape of the 45-degree line but have much more variation – consistent with the findings from inspecting the time series plots.

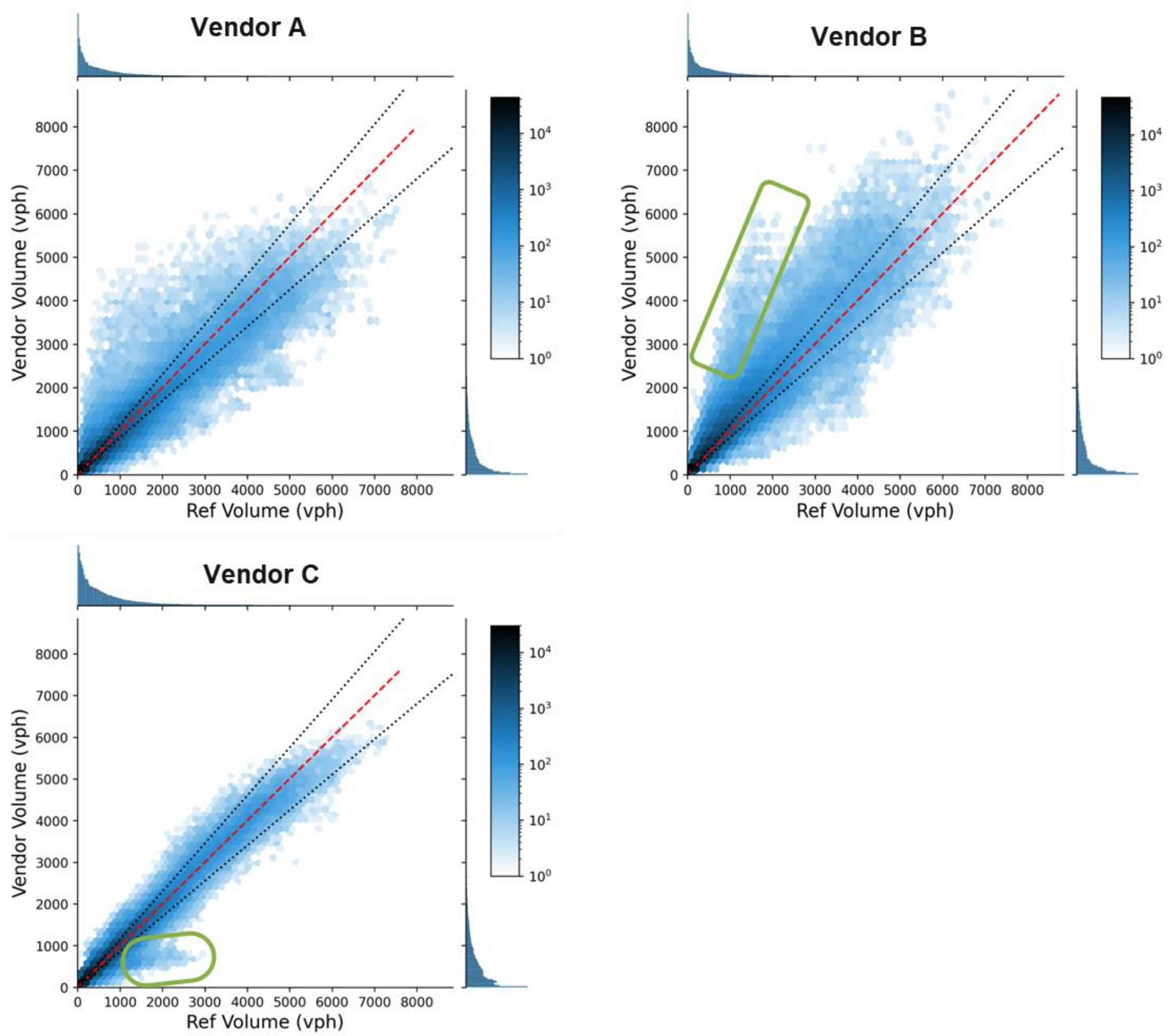

Figure 20 – Hourly scatter plots for all vendors

Error! Reference source not found. summarizes the percent error for the three vendors by v olume bin, with the blue lines indicating +/- 15% error. As was the case with AHDT, volumes less than 100vph are omitted to prevent this range from dominating the plots. Vendors A and C both tend to be within 15% of the reference data in most volume bins above 250 vph, with C showing the tightest error bars. Vendor B tends to overestimate hourly volumes during low-volume periods, but improves as volume grows and the bias disappears. All vendors have a wider error distribution at lower volumes and tighten up as volume levels increase. However, it is important to remember that these summary metrics should not be viewed in isolation; at the hourly level it is important that volume measures track temporal patterns.

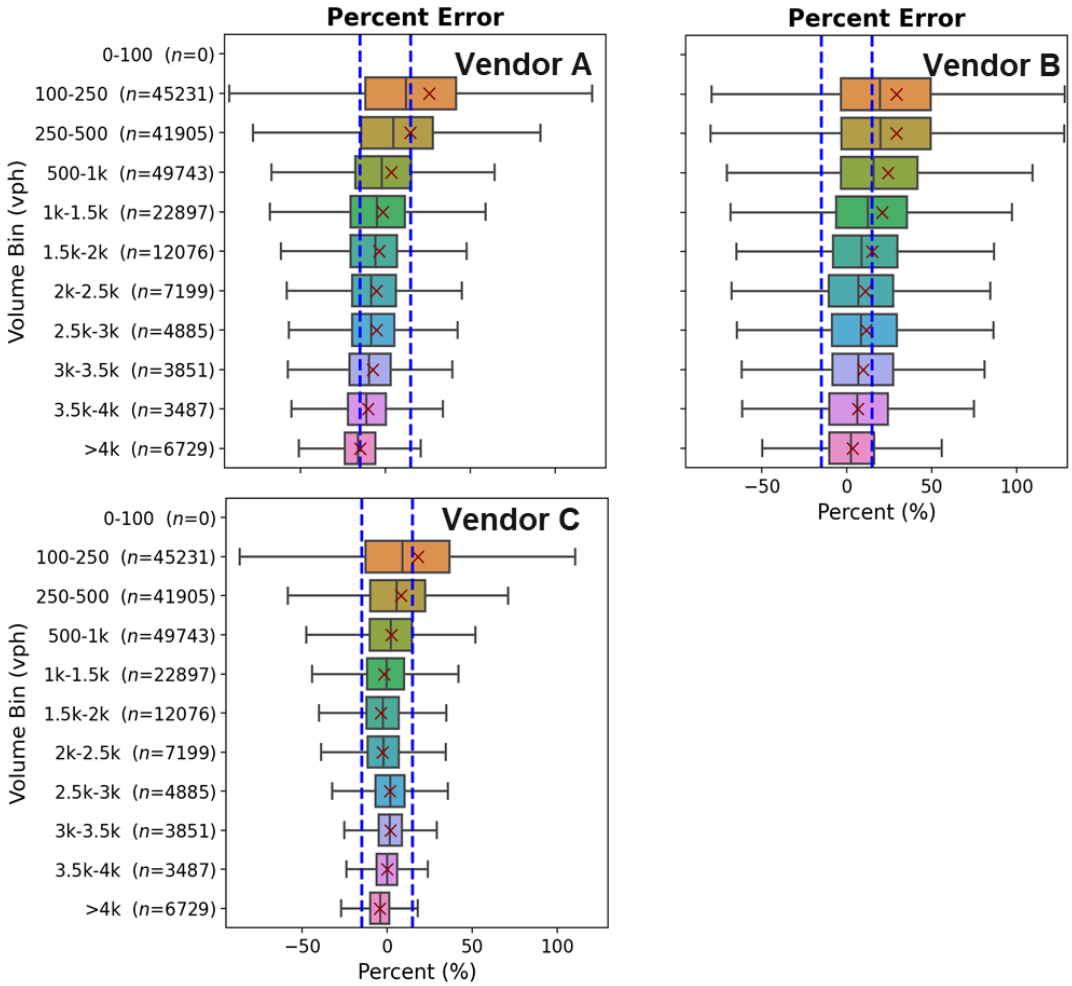

Figure 21 – Distribution of Hourly Volume percent error metrics by volume bin

AADT

AADT represents the final and most temporally aggregate data element. As mentioned previously, unlike the prior three data elements – ADT, AHDT, and Hourly Volumes – the AADT evaluation is *not a blind study*; hourly counts from 11 of the 12 months from 2022 (the calendar year chosen for AADT evaluation) were provided to vendors as calibration/training data. Nonetheless, the TAC recommended that this data element be included as a calculation exercise, and the results presented below suggest that the vendors provided representative data submissions – not relying on the count data provided as a means to estimate AADT directly.

Figure 22 plots vendor AADT volumes against reference AADT values obtained from NCDOT CCS data. Most vendors' estimates fall within the +/- 15% envelope around the ideal 45-degree line, with some differences between vendors. Visually, the higher volume region is easier to see, as there are fewer points in this region and the error envelope is more spread out. Relative to ADT (a similar product), something that stands out is that Vendor B and C have fewer outliers – particularly ones that look like they may be directional issues.

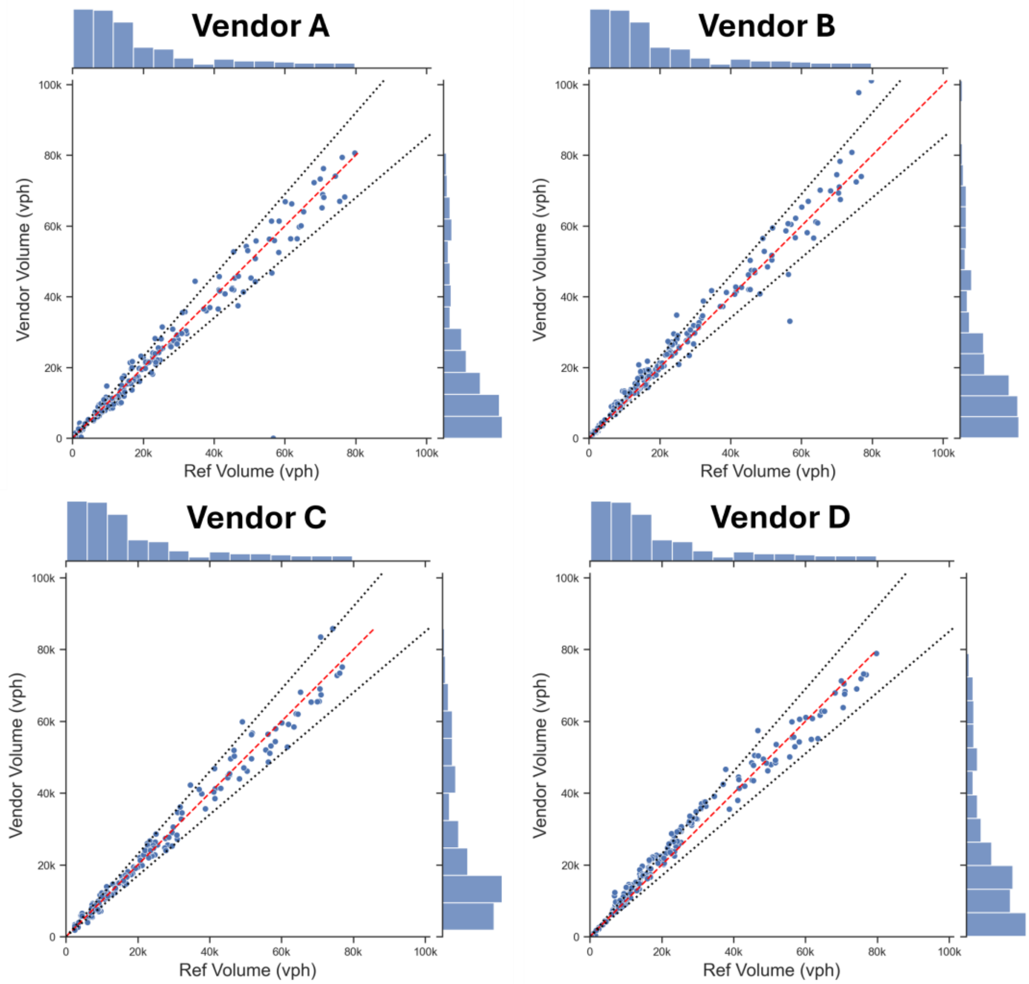

Figure 23 – AADT scatter plots for all vendors

Error! Reference source not found. summarizes the AADT percent error for the three vendors b y volume bin, with the blue lines indicating +/- 15% error. In general, AADT results (calendar year 2022) are similar to ADT (January 2023), which is intuitive because both products are similar, just defined for different time periods. However, Vendor B's estimates are a bit more accurate for AADT; in Figure 9 they consistently overestimated across many volume bins, whereas here the boxplots are often within 15% error. As was the case for other data elements, Vendor C's results are based on only FRC 1-5, and in this case, results are not available for the lowest two bins.

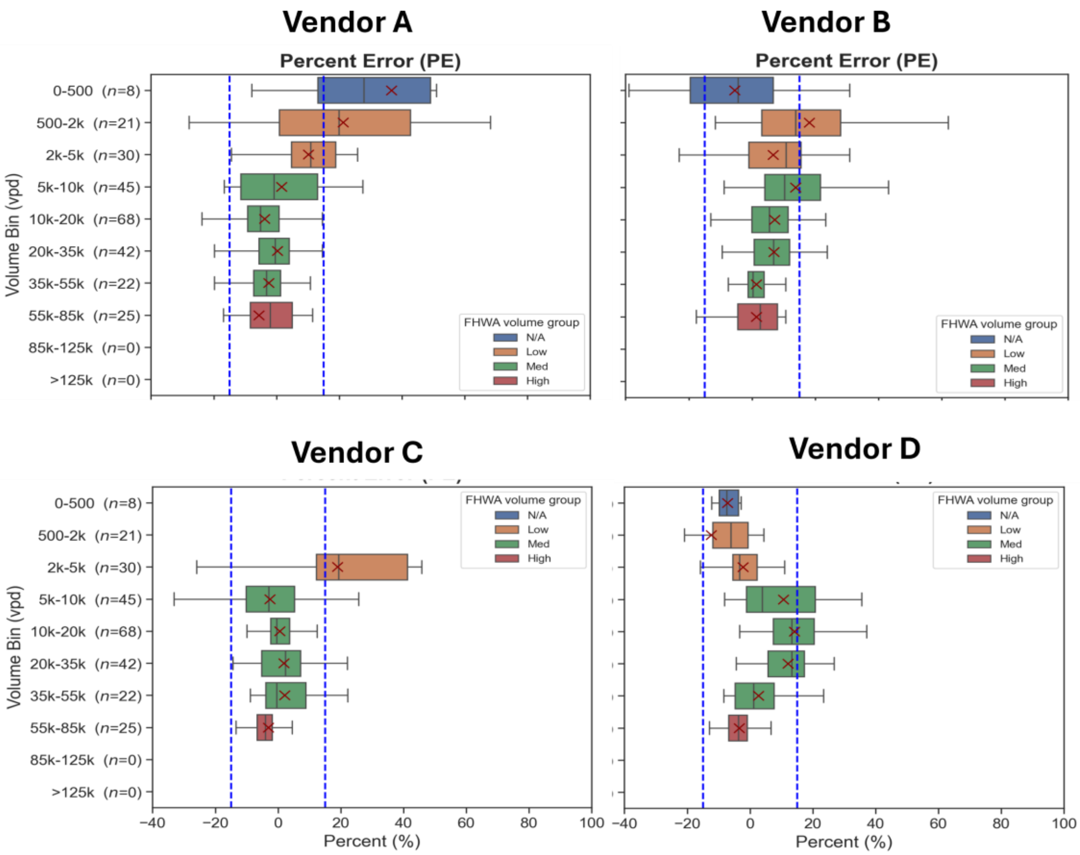

Figure 24 – Distribution of AADT percent error metrics by volume bin

The prevailing practice for estimating AADT values involves factoring short-term counts. As such, it is instructive to consider the accuracy of TDM vendors' AADT estimates relative to this technique. Table 4 lists the accuracy metrics associated with factoring 48 hour counts to AADT volumes[6], which can be used as a point of comparison with vendor data. Note that FHWA recommends a more involved statistical comparison for cases where more than 93 continuous count samples are available per volume bin – a requirement not met for this study.

Table 4 – Accuracy of factoring 48 hr counts when <93 CCS available per bin (source: FHWA)

Road Size Bin	Median Error (Bias) %	68th Percentile Absolute Error (%)	95th Percentile Absolute Error (%)	NRMSE	MAPE
A: 0 – 499	*Unknown*	*Unknown*	*Unknown*	*Unknown*	*Unknown*
B 500 - 1,999	-0.1	11.8	26.9	13.0	10.1
C: 2000 - 4,999	2.3	10.8	33.6	17.3	10.5
D: 5,000 - 9,999	3.2	9.7	28.1	14.0	9.3
E: 10,000 - 19,999	1.3	9.2	27.9	12.9	9.0
F: 20,000 - 34,999	0.9	8.4	24.4	13.3	8.2
G: 35,000 - 54,999	0.5	8.4	19.3	9.8	7.3
H: 55,000 - 84,999	-0.3	6.1	14.5	7.3	5.3
I: 85,000 - 124,999	0	4.9	14.7	6.8	4.7
J: > 125,000+	3.1	7.1	17.7	10.0	6.2

[6] FHWA-PL-21-031 (2021) "Guidelines for Obtaining AADT Estimates from Non-Traditional Sources" (link)

The error metrics in Table 4 are reproduced in Table 5 (column labeled 48hr), with Vendor A through D's results reported alongside each volume bin. For all metrics but the Median Error, results are color coded such that values lower than the benchmark (i.e., better/more accurate) are green and those greater are red. These results suggest that for all vendors, errors in many volume bins are greater than for the reported 48-hour count values, although results vary by metric.

Table 5 – Vendor AADT metrics alongside FHWA-defined benchmarks for factored 48 hr counts.

Volume	Median Error (Bias) %					68th Pct Absolute Error (%)					95th Pct Absolute Error (%)				
Bin	48hr	A	B	C	D	48hr	A	B	C	D	48hr	A	B	C	D
A: 0 – 499		27.9	-4.42		-7.4		47.5	31		9.1		93.7	37.2		11.9
B: 500-2k	-0.1	19.9	13.8		-6.0	11.8	34.1	18		10.6	26.9	68.2	57.3		70.5
C: 2k-5k	2.3	10.9	10.7	19.3	-3.3	10.8	18.9	17	37.8	5.4	33.6	88.8	26	43.5	11.3
D: 5k-10k	3.2	-0.8	10.3	-3.0	3.9	9.7	15	20	12.3	12.8	28.1	23.8	34.5	25.8	35.3
E: 10k-20k	1.3	-5.2	5.44	-0.4	13.3	9.2	13.1	8.69	5.3	18.8	27.9	25.7	29.2	11.9	32.7
F: 20k-35k	0.9	-0.5	6.78	2.3	13.3	8.4	8.05	11.6	9.4	16.1	24.4	21.3	22.6	14.6	20.8
G: 35k-55k	0.5	-3.3	0.32	-0.4	1.1	8.4	10.2	6.58	8.6	7.7	19.3	15.8	15.2	13.4	22.5
H: 55k-85k	-0.3	-2.0	2.63	-4.1	-3.7	6.1	8.64	8.47	6.75	5.5	14.5	15.9	28	15.4	10.5
I: 85k-125k	0					4.9					14.7				
J: >125k	3.1					7.1					17.7				

Volume	NRMSE					MAPE				
Bin	48hr	A	B	C	D	48hr	A	B	C	D
A: 0 – 499		38.9	18.5		7.2		67.4	18.1		7.8
B: 500-2k	13	33.2	20.8		13.4	10.1	39.7	24.7		28.8
C: 2k-5k	17.3	24.6	13.2	27.8	5.3	10.5	31.3	16.5	38.0	6.3
D: 5k-10k	14.0	11.6	14.6	11.2	12.8	9.3	16.3	18.5	14.1	19.3
E: 10k-20k	12.9	9.9	9.5	4.54	14.4	9.9	13.2	13	5.8	16.9
F: 20k-35k	13.3	7.28	9.32	7.31	12.5	8.2	10.7	12.7	9.8	14.6
G: 35k-55k	9.8	7.41	4.95	6.7	7.3	7.3	9.4	7.41	8.7	8.9
H: 55k-85k	7.3	10.3	9.52	6.3	4.7	5.3	18.8	13.5	7.8	5.6
I: 85k-125k	6.8					4.7				
J: >125k	10					6.2				

This comparison marks the first attempt to evaluate volume data quality using the FHWA framework. Further work is needed to align evaluation methods with FHWA's evolving guidelines and make it easier to compare TDM vendor AADT accuracy with FHWA-prescribed benchmarks.

Conclusions

This report explores the fidelity of volume estimates derived from various types of probe data. The study, part of the Eastern Transportation Coalition Transportation Data Marketplace (TDM) validation program, begins to establish evaluation methodologies, metrics, and standard forms of accuracy and fidelity processes to reveal both the accuracy and value of the industry supplied data. The following conclusions emerge from evaluation of volume data at four levels of temporal aggregation (Hourly, AHDT, ADT, AADT) for four TDM vendors:

- **All 4 vendors were able to deliver data to support validation, reflecting a major milestone for industry.** Whereas this type of traffic data was in research and development pre-pandemic, it is emerging and maturing quickly with four vendors providing data to the TDM for evaluation across several hundred continuous count stations. Although there were some differences from vendor to vendor, all were able to report on a high majority of the CCS, in a standard method with minimal to no manual interaction with respect to geo-location, conflation, or interpretation of results. This is a major milestone for the industry.
- **A standardized set of visuals, metrics, and benchmarks are emerging.** Metrics and methods used to compare and contrast performance from vendor to vendor are beginning to 'gel', meaning that a standardized set of visualizations, metrics, and benchmarks are emerging from research and development for use in evaluating industry grade volume data. Although benchmarks will need to be set based on application needs, the tool set (statistical measures and visualization) is converging to common practice. *The need to observe time series data* (i.e., plotting vendor-supplied data against reference data over time) *remains critical to understand trends and individual vendor characteristics* with respect to the volume estimate methodologies and reporting practices.
- **Industry appears to be in an early product maturity stage,** though it is expected that vendor products will evolve with changes in base data sources, algorithms, and reporting. Accuracy and fidelity of vendor supplied data are measurable, with moderate differences observed between vendors, and varying across reported data elements and volume ranges. The questionnaire revealed that most vendors (three of four) take a "top-down" approach to volume estimation starting with a more aggregate volume measure, typically ADT, and then disaggregating appropriately for other measures (and aggregating for AADT). One of the four vendors used a "bottom up" strategy, that is they directly estimated volumes for specific dates/time periods and then aggregated to other measures). The results were consistent with these findings, as the one "bottom up" vendor was the only one that produced meaningful temporal patterns at the hourly level, while the "top down" ones did not. Even the more aggregate data elements – the focus area for most vendors – had mixed performance, and for certain vendors, inconsistent accuracy between urban and rural locations.

Future studies will seek to replicate and advance the processes developed in this report, with further exploration of vendor performance broken out by various conditions, including Urban vs Rural, Peak vs Off Peak scenarios (for temporal data elements). Given state DOT interest in AADT, the next study will likely emphasize more aggregate data elements (likely AADT and/or ADT) and include further integration of FHWA's evaluation framework.