Traffic Volume Validation - Literature Review and Recommendations

TETC Traffic Data Marketplace Data Validation

Prepared by: TETC Validation Team

11/22/2022

Traffic Volume Validation - Literature Review and Recommendations

TETC Traffic Data Marketplace Data Validation

Literature review and recommendations for structuring a volume validation program

TDM-VAL-1

Data Categories: Volume

11/22/2022

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

Executive Summary

The purpose of this document is to highlight the key traffic volume validation studies that have been conducted to date, focusing on approaches that may be suitable for evaluating TETC traffic volume datasets. As such, it is organized as follows:

Section 1, titled *TDM Volume Data Expectations*, reviews the Volume data expectations and quality measures defined in the RFP.

Section 2, titled *Literature Review,* reviews relevant traffic volume validation research, focusing on a handful of recent studies that investigate the quality of probe-based traffic volume estimates. These studies are reviewed with an eye for relevant evaluation strategies, error measures, visual presentation of results, and other learnings that might prove useful for TETC validation purposes.

Section 3, titled *Recommendations*, presents a preliminary path forward for TETC volume validation, seeking to build off the learnings from previous studies and align the evaluation process with the expectations outlined in the RFP.

Table of Contents

Background

This report marks the first validation activity for The Eastern Transportation Coalition (TETC) Traffic Data Marketplace (TDM), which encompasses six transportation data items (Travel Time/Speed, Volume, Waypoint, Origin-Destination, Freight, Conflation Services) and 12 data vendors. Whereas the VPPII (Vehicle Probe Project Phase II, the precursor to TDM) validation program focused on quantifying vendor travel time/speed data accuracy via travel time field measurements and well-established analysis methods, the TDM program will need to develop new validation techniques that are appropriate for the new data items sold within the TDM. As such, the TDM validation reports will contain a broader range of topics and analysis; each report will summarize a discrete “validation activity” – not necessarily quantitative evaluation of vendor data based on comparison with field measurements. Reports may also include literature reviews of best practices for evaluating new types of data, methodological development of validation processes, descriptive analyses of new datasets, and Quality Assurance / Quality Control procedures to sanity check emerging datasets.

Based on guidance from the Validation Technical Advisory Committee (TAC), the validation team’s initial focus area is Traffic Volume data. This inaugural TDM validation report is structured as a technical memo to the TAC containing a detailed literature review of best practices for quantifying volume accuracy and recommendations about how to structure the volume validation process. These recommendations – which were reviewed by the TAC and other key stakeholders – will be used to guide the initial Fall 2022 Volume validation in North Carolina.

TDM Volume Data Expectations

Historical Traffic Volume data is one of the core data items in the Traffic Data Marketplace. However, in contrast to Travel Time and Speed data, which has been studied extensively by the Coalition since 2008 and has well defined accuracy measures, traffic volume datasets are new to the TDM and have not been evaluated. Relevant expectations from the TDM RFP[1] (Section 2, pg. 80-84), are listed below.

Data Elements and Reporting

Vendors must deliver volume data across a wide range of aggregation levels, ranging from Annual Average Daily Traffic (the most aggregate) to much more granular hourly estimates for specific days (e.g., 8-9AM on June 1, 2022). **Each of these data elements – AADT, Avg. Daily Traffic (ADT), Avg. Hourly Daily Traffic (AHDT), and hourly estimates – are considered mandatory (M)**. Reporting at the sub-hourly level (e.g., 5, 15, or 30 min) is considered highly desirable (HD), while breakdown of results by vehicle class or grouping of vehicle class (e.g., light/medium/heavy duty) is considered desirable (D). Real-time volume was requested of the vendors as a non-mandatory data item, but no vendor stepped up to offer a real-time volume data feed.

Data Quality and Latency

This section of the RFP acknowledges that validating data quality is an emerging field, and thus the specific accuracy measures and parameters of the program are not yet established. As such, it lists a series of Highly Desirable (HD) requirements that serve as a starting point for thinking about accuracy based on studies by FHWA and the Eastern Transportation Coalition's Volume and Turning Movement project, which will be covered further in the Literature Review below. Additionally, it notes that each vendor will be required to self-report various accuracy measures using cross-validation techniques associated with the calibration of the inference engine.

[1] **Traffic Data Marketplace RFP (link)**

Literature Review

As shown in Figure 1, traffic volume estimation literature can be categorized according to its level of temporal granularity and the estimation approach. Like the TDM marketplace, which contains volume data items ranging from AADT (aggregate) to hourly/sub-hourly estimates for a particular day (disaggregate), the existing volume estimation literature captures a variety of volume aggregation levels. Using a similar categorization as [1], estimation approaches tend to fall in to one of three categories: (i) Factoring, which involves estimating aggregate volumes from short-duration traffic counts, (ii) Forecasting, which involves predicting future volumes from historical values, and (iii) Historical / Real-time Estimation, which seeks to estimate volumes at unmeasured locations for historical or current periods.

The first approach, Factoring, is only relevant for estimating aggregate volumes, with literature focused on developing approaches to improve AADT estimates from typical short-duration counts (e.g., [2]–[4]). The second approach, Forecasting, is relevant for varying levels of temporal granularity. At the aggregate level it often involves estimating AADTs for future years using historical values [5]–[7], while at the disaggregate level it includes predicting future volumes within a short time horizon (e.g., using time series [8]–[10] or machine learning [11]–[13] methods). However, these first two approaches are generally *not* as relevant to the TDM because vendors are expected to provide network-wide volume data for historical/current periods without the aid of short-term counts. The third category, Historical / Real-time Estimation, represents the type of estimation most relevant to the TDM for both aggregate and disaggregate data. Relevant research encompasses a wide range of estimation approaches, including spatial interpolation of AADTs (e.g., [14]–[16]) and direct estimation of volumes via statistical/machine learning regression (e.g., [17]–[22] for AADTs, [23]–[27] for hourly volumes).

In the context of TDM volume validation, the estimation methodology is the vendors' responsibility, and the evaluation process should not depend on the specific modeling method. Accordingly, **this review focuses on how volume accuracy is evaluated** rather than how estimates are generated. Note that this can include research studies where accuracy is self-reported or independent evaluations of commercial volume products (e.g., [28]–[34]).

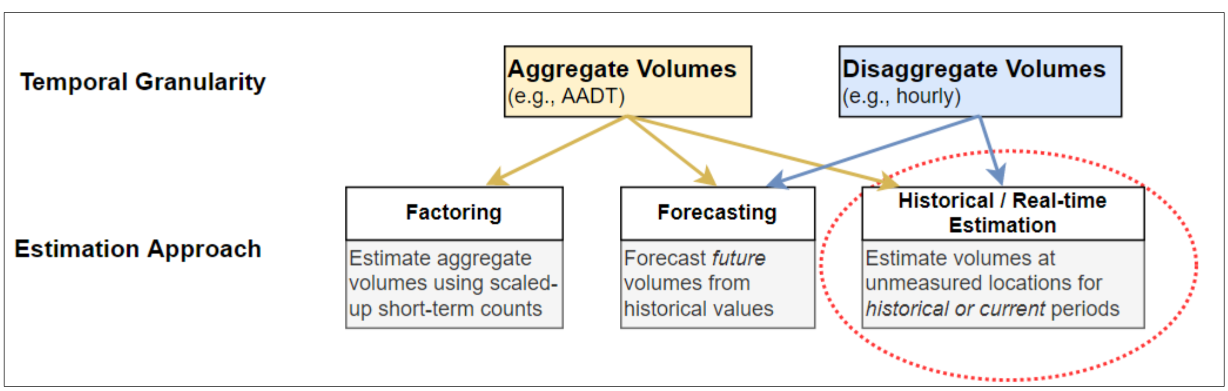

Figure 1 – Framework for categorizing volume estimation literature.

The following subsection, Validation Characteristics, reviews existing volume validation literature with an eye for concepts that may be relevant for the TMD program, categorizing validation approaches according to several key factors. Afterwards, the "Summary of Most Relevant Studies" section provides a more in-depth review of several key studies that are most closely related to evaluating the type of volume data sold in the TDM.

Validation Characteristics

Many different methods and approaches are used to quantify accuracy, but at a high level, the strategies can be categorized according to the following factors:

Reference data source and location. Volume estimates – whether produced independently by a commercial vendor or internally as part of a research effort – need to be compared to a reference source to compute accuracy measures. Ideally this comparison would take place across a variety of roadway locations, encompassing different road classes and traffic conditions. Reference volume data can be obtained from a variety of sources, including:

- ***Continuous Count Stations (CCS):*** Well-calibrated CCSs managed by state DOTs tend to be the most reliable source of continuously-collected count information in the United States [35], and thus serve as an excellent candidate for reference volume data. A complicating factor, however, is that CCS volumes are often publicly available and used by vendors/researchers for calibrating volume estimation models, making it challenging to know whether models would generalize well elsewhere. Ideally, a subset of high-quality CCS locations could be withheld from the public to support a truly blind evaluation. While this is certainly possible and has been done on occasion (see [30] and [34], for example), it requires a "big lift" from agencies and is generally not practical, particularly at a national scale. A more common approach for evaluating accuracy with CCS data is to employ cross-validation (e.g., k-fold or leave-one-out), where count data are repeatedly partitioned into training and testing groups, with the volume estimator calibrated at the the training data and evaluated on the testing data (see, for example, [25]–[29]). This allows high-quality volume data from CCS locations to be used as a reference source.

- ***Other permanent traffic sensors*:** Other permanently traffic sensors – either explicit traffic counters or other sensors which count vehicles as part of another function (e.g., properly equipped traffic signal controllers) -- may be useful as a reference volume source. The advantage of such sources is that the resulting count data may be unseen to the data vendor, and thus may be useful for conducting a blind evaluation. However, unlike CCSs, which are carefully managed by state DOTs and whose counts are submitted regularly to FHWA, the quality of such sources is often unknown, and should be checked carefully prior to use. For example, Tsapakis, et al. [30] looked into using a variety of non-CCS counters including traffic signal control sensors, Doppler radar, and ITS-based sites (inductance loops, side-fire radar, video machine vision) for AADT evaluation. They ruled out a few technologies immediately and validated others against manually reduced video feeds and found somewhat erratic results (mean absolute error of 21% for Doppler radar and 5%/5%/14% for the ITS based solutions), ultimately deciding that none were suitable for validation.

- ***Short Duration Counts*:** Short-duration counts typically refer to the 24 hour – 1 week traffic counts conducted along various roadways by state DOTs for compliance with Federal requirements. These short-term counts are collected on each road segment on a rotating basis (e.g., 48-hour counts every 3 years) and are expanded to obtain official AADT estimates for the public roadways. This scaling process is often referred to as factoring, and involves computing average daily traffic (ADT) from short term counts, and then multiplying by various adjustment factors to account for seasonality and temporal patterns to get to the AADT level [36], [37]. While agency counts may serve as a useful source for reference data in a variety of different locations, it is important to note that they are intentionally collected during "typical" conditions, and thus might not capture the type of variation that would be most useful for evaluating performance – at least for temporally granular volumes.

  Short-duration counts can also be collected independent of state DOTs via one-off field deployments, as is expected for TDM validation efforts. The advantage of this approach is that sites can be selected to intentionally capture atypical or seasonal traffic patterns that might not show up in official short-duration counts. Care should be taken to make sure sensors are properly calibrated and capable of producing trustworthy reference data before using as the basis for evaluation.

Temporal granularity and collection duration. The temporal granularity at which vendor and reference data are compared plays a large role in the evaluation and interpretation of results.

- ***Hourly (and sub-hourly) volume estimates:*** Evaluation of hourly (and more granular) volume estimates is generally straightforward because corresponding hourly reference data can easily be collected for the duration of the estimation period; even 24 or 48-hour short-duration counts can provide opportunities for direct comparison.

- ***Aggregate volume estimates:*** Evaluation of aggregate volume estimates (e.g., AADT, ADT, AHDT) is more challenging because corresponding reference volumes, when computed from counts, requires much more data – up to 1 year in the case of AADT. Because it is hard to withhold the high-quality continuous count data (i.e., the one source that can provide a full year of data for evaluation), it is common to either employ cross-validation with CCSs (e.g., [20], [23]) or simply use scaled-up short-term counts as "ground truth" (e.g., [29]). On one hand using AADT estimates from short-term counts is understandable; it is hard to obtain a year's worth of count data for blind evaluation and the resulting comparison can still serve as a useful exercise. However, expanding short-term counts can introduce additional estimation error, and can make it difficult to determine whether discrepancies between reference and vendor data sources are due to true differences or errors introduced by factoring.

  Accordingly, researchers have sought to characterize the amount of error that can be expected when short-term counts are collected and scaled up to AADTs. Gadda et al. [38] investigated this type of estimation error using CCS locations in Florida and Minnesota and reported errors in the 10-15% range, with some variation depending on road type, urban/rural designation, day of week, and count duration. More recently, the 2015 pooledfund study TPF-5(292) *Assessing Roadway Traffic Count Duration and Frequency Impacts on AADT Estimations* [39] used 14 years' worth of high quality CCS data to quantify accuracy and precision of AADT estimates obtained from short-term counts. They looked at the distribution of errors and found that while the median bias was around 1% (strong agreement), there was a wide distribution of values, indicating that there can be large errors in some cases. The 95% confidence interval of error for AADT estimates estimated from 48-hour counts was (-24%, 32%), meaning that there was a 95% probability that a 48-hour count scaled up to the AADT level would have an error within -24% and +32%. In addition to highlighting why short-term counts may be unsuitable for computing reference volumes, this line of research helps benchmark the accuracy for the state of the practice; any viable AADT alternative should meet or exceed such accuracy levels.

Evaluation Methods. The methods used to evaluate AADT estimates tended to fall into one of two categories: (i) quantifying performance by computing error measures or (ii) statistical tests.

- ***Error Metrics.*** Table 1 summarizes key error measures used to quantify differences between estimated and reference volumes, focusing on metrics that were frequently used in in the most relevant studies. No one metric is optimal; some quantify systemic bias in the estimates (e.g., Mean Error), while others look at the absolute value of the error (e.g., MAE) or measure error as a percentage relative to true volume (e.g., MAPE) or roadway capacity (e.g., ETCR/EMFR). Furthermore, care must be taken to interpret the results with proper context; for example, an absolute error of 100 vehicles represents 0.5% error when the true volume is 20k vehicles, but 50% error if the true volume is 200. Accordingly, multiple metrics are typically reported to provide diverse insights into model accuracy, and often computed separately for different subsets of the data (e.g., by volume level, for urban and rural areas, and road class). As suggested in [34], it can be useful to visualize the results, with common visualizations including scatter plots showing predicted vs reference values (with a 45 degree line showing perfect correlation) as well as histograms and boxplots of error distributions.

- ***Statistical Tests*.** Several statistical tests were used for evaluation purposes, including hypothesis tests to determine whether the mean or median estimated volumes were statistically different from the reference volumes. [34] and [42] used paired t-tests to compare means – an approach that assumes that differences between estimated and reference volumes are roughly normally distributed. An alternative non-parametric approach is the Wilcoxon signed-ranks test, which compares the median instead of the mean without requiring any parametric assumptions, examples of which include [30], [34].

Table 1 – Common 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. $\hat{\boldsymbol{Y}} = \{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n\}$, $\boldsymbol{Y} = \{y_1, y_2, \dots, y_n\}$, $\boldsymbol{E} = \{e_1, e_2, \dots, e_n\} = \{(\hat{y}_1 - y_1), (\hat{y}_2 - y_2), \dots, (\hat{y}_n - y_n)\}$			
Metric	**Formula**	**Comments**	**Selected References**
Mean **Error**	$\frac{1}{n}\sum_{i=1}^{n} e_i$	Captures signed error (vehicles)	[28], [30], [31], [34]
Xth Percentile **Error**	$\mathrm{PCTL}_X(\{e_1, e_2, \dots, e_n\})$		[30],[34]
Mean **Absolute Error** (MAE)	$\frac{1}{n}\sum_{i=1}^{n} \lvert e_i \rvert$	Captures magnitude of error (vehicles)	[28], [31], [33]
Xth Percentile **Absolute Error**	$\mathrm{PCTL}_X(\{\lvert e_1\rvert, \lvert e_2\rvert, \dots, \lvert e_n\rvert\})$		[20]
Mean **Percent Error**	$\frac{100}{n}\sum_{i=1}^{n} \frac{e_i}{y_i}$	Captures signed error (%) normalized to reference volume. Percent errors are sensitive to volume level.	[30], [40]
Xth Percentile **Percent Error**	$\mathrm{PCTL}_X\left(100 \cdot \left\{\frac{e_1}{y_1}, \frac{e_2}{y_2}, \dots, \frac{e_n}{y_n}\right\}\right)$		[20], [30], [32], [34]
Mean **Absolute Percent Error** (MAPE)	$\frac{100}{n}\sum_{i=1}^{n} \left\lvert\frac{e_i}{y_i}\right\rvert$	Captures magnitude of error (%) either normalized to reference volume or combination of reference and estimate values. Percent errors are sensitive to traffic volume level.	[20], [23], [30], [34]
Symmetric Mean **Absolute Percent Error** (SMAPE)	$\frac{100}{n}\sum_{i=1}^{n} \frac{\lvert e_i\rvert}{(\lvert y_i\rvert + \lvert\hat{y}_i\rvert)/2}$		[41]
Xth Percentile **Absolute Percent Error**	$\mathrm{PCTL}_X\left(100 \cdot \left\{\left\lvert\frac{e_1}{y_1}\right\rvert, \left\lvert\frac{e_2}{y_2}\right\rvert, \dots, \left\lvert\frac{e_n}{y_n}\right\rvert\right\}\right)$		[30], [32]–[34]
Error to Capacity Ratio (ETCR)	$\frac{100}{n}\sum_{i=1}^{n} \frac{e_i}{\mathrm{capacity}_i}$	Captures signed error (%) normalized to road capacity (from Highway Capacity Manual) or the maximum observed volume (a proxy for capacity).	[23], [25]
Error to Max Flow Ratio (EMFR)	$\frac{100}{n}\sum_{i=1}^{n} \frac{e_i}{\max(Y)}$		[23], [25], [26]
Normalized Root Mean Square Error (NRMSE)	$100 \cdot \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2}}{\frac{1}{n}\sum_{i=1}^{n} y_i}$	Captures magnitude of error (%) but penalizes larger errors more than MAPE.	[20], [30], [34]
Coefficient of Determination	$1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n}\left(y_i - \left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)\right)^2}$	Measures explanatory power / goodness of fit (generally 0-1, with higher=better)	[20], [23], [25], [34]
Pearson Correlation	$\frac{n\sum_{i=1}^{n} y_i\hat{y}_i - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} \hat{y}_i}{\sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}\sqrt{n\sum_{i=1}^{n} \hat{y}_i^2 - \left(\sum_{i=1}^{n} \hat{y}_i\right)^2}}$	Measures strength of linear relationship between estimated and reference values (-1 to 1, with higher=better)	[20], [40]

Summary of Most Relevant Studies

The following studies were identified as being most closely related to the TDM validation task – all of which focused on probe-based volume products, and many of which incorporated independent validation analyses. Key aspects from these studies are summarized below.

FHWA Transportation Pooled Fund Study TPF-5(384): "Exploring Non-Traditional Methods to Obtain Vehicle Volume and Class Data"

FHWA's pooled fund study focused on developing non-traditional (probe-based) approaches to estimate AADT volumes and related attributes (e.g., hourly profiles, temporal factors) by vehicle type [43]. This large-scale research effort involved several key entities, including a data vendor who provided the so-called "non-traditional" volume estimates (Streetlight) and multiple research teams who provided independent validation to benchmark data quality (Texas A&M Transportation Institute (TTI), the National Renewable Energy Lab (NREL)). Key publications that came out of this project include [20], [35], [30], [34], [42] . The following paragraphs summarize key points from several reports.

Streetlight published two key reports: [20] provided technical details about their volume estimation method (e.g., input data, machine learning models) and internal validation results, while [35] summarized guidance for agencies interested in purchasing AADT estimates from entities using non-traditional (probe-based) data sources. In the main report [20], Streetlight employed k-fold cross-validation (with groups stratified by AADT) to evaluate their estimates relative to AADTs from permanent count stations, reporting Pearson correlation, MAPE and NRMSE error metrics, as well as various error percentiles (median bias, 68th and 95th percentile absolute error), and the 95% Traffic Count Error (TCE, which is percent error) range. Following the method outlined in Appendix B of [35], the median bias and 95% TCE range measures were used to compare performance to the expected errors associated with scaling up short-term counts reported by Krile, et al. in [39]. Streetlight's internal analysis suggested that AADT estimates were at least as good as short-term counts for traffic volumes over 2000 vehicles/day, with mixed results in the 500-2000 vehicles/day range. They also developed models to estimate monthly average daily traffic (MADT), average daily traffic (ADT) for specific days, and hourly volumes for specific days. Streetlight noted that error was correlated with total probe trips, so shorter time periods and lower class (less traveled) roads had higher error. Along these lines, they concluded that hourly estimates could capture hourly trends on high and medium size roads, but not for low-volume roads when probe samples were sparse.

TTI's independent validation of Streetlight data is summarized in [30], and involves several different analyses comparing Streetlight AADT estimates with reference volumes computed directly from high-quality continuous count stations. Notably, the continuous count data from six states was carefully withheld from the vendor – a task that involved significant coordination with state DOTs and various entities, thus allowing for a truly blind evaluation. Additional sources of count data from traffic signal control sensors, Doppler radar, and ITS-based sites were evaluated as supplemental reference data sources, but none were deemed suitable for quantifying ground truth volumes. TTI's analysis approach involved three parts: (1) exploratory analysis of AADT values and error distributions (2) computation of common error metrics and (3) various statistical tests. Error metrics were computed between estimated and actual AADT values and included mean/median algebraic difference (i.e., error), mean/median traffic count error (TCE, which is percent error), mean/median absolute percent error, normalized root mean square error (NRMSE), coefficient of determination (R^2), and 68% and 95% TCE ranges (measures of precision). Their statistical analysis included (i) the previously referenced method from [35], referred to here as the TPF-5(384) method, to determine whether errors are within accuracy and precision ranges associated with scaling up short-term counts, (ii) a TTI modified version of the TPF-5(384) method, and (iii) the Wilcoxon Signed Rank test. The error metric analysis indicated that Streetlight's AADTs were overestimated (positive bias) at low volumes and underestimated (negative bias) at high volumes, with accuracy improving at higher volumes (using mean/median percentage error and NRMSE). For example, the MAPE across all directional sites was around 15% overall but ranged from 40% in the lowest volume bin (0-499 AADT) to 10% in the highest (55k+ AADT). The statistical testing results required nuanced interpretation as the TFP-5(384) method sometimes produced counterintuitive results that did not agree with the other tests and error metrics. However, using their professional judgement, and giving the most weight to the Wilcoxon Signed Rank test and TTI-modified TPF-5(384) methods, TTI concluded that Streetlight AADTs were suitable for traffic monitoring on roads where bi-directional AADT values are 5000 and above.

NREL also independently validated Streetlight's AADT estimates, using high quality volumes from continuous count stations and toll plazas to compute reference AADT values, and quantifying accuracy using a variety of error metrics and statistical tests [34]. Although there was some overlap with the TTI validation team, NREL acquired reference data in different locations to provide an independent look at data quality. To this end, they started by identifying TMAS continuous count stations from 24 states, keeping only ones that were not used by Streetlight for calibration purposes (including counters withheld by FHWA/ State DOTs and others that Streetlight agreed to exclude), and augmenting with trusted toll system counters in several states. Other data sources were considered as well (e.g., ITS systems, road weather information systems, doppler radar variable speed signs), but counts were ultimately unavailable or deemed unreliable, and thus not used as a reference source. To evaluate the estimates, they used several error metrics (error percentiles and error percentile ranges, median percent error, mean/median absolute percent error, normalized root mean square error, spearman correlation coefficient) and hypothesis tests (paired t-test and Wilcoxon matched pairs signed-rank test). NREL found that Streetlight's AADT estimates correlated well with reference data – both overall and for different reporting scenarios. However, statistically significant differences in the mean/median values at a significance level of 0.05 were observed via hypothesis testing (paired t-test and Wilcoxon signed-rank test), and inspection of the error distributions showed the presence of significant outliers. In general, Streetlight's AADT estimates tended to be positively biased relative to reference values (median error of 5.7% across all sites), with accuracy higher at higher volumes and in rural areas. MAPE values ranged from 30% in the lowest volume bin (< 500 AADT) to 10% in the highest one (55k+ AADT), with Rural and Urban MAPES at 16% and 21%, respectively. Additionally, NREL found that toll locations had noticeably higher error than TMAS sites (34% vs 12% MAPE) and theorized that this could be due to complex geometries at toll facilities. Ultimately, no firm verdict was reached on whether the accuracy was "good enough"; NREL argued that this question is context-specific and should be determined on a case-by-case basis.

The Eastern Transportation Coalition's Volume and Turning Movement (VTM) project

The VTM project [23], led by UMD and NREL research teams under the guidance of the Coalition and a steering committee, initially involved the following tasks: surveying members of state agencies to understand target volume accuracy levels, collaborating with industry partners to source GPS probe data, developing processes to estimate probe-based hourly traffic volumes, and characterizing the quality of estimates through cross-validation. Later, a second phase of the project focused on refining volume estimation methods, evaluating data quality on lower-class roads, and taking steps towards commercialization.

Several key publications came out of this project, including a joint NREL/UMD paper describing the survey findings and preliminary proof-of-concept results [44], the final NREL report [23], research papers by NREL and UMD groups [24]–[26], [45], and several webinars and conference presentations (e.g., [41], [46]). One of the key findings from the initial survey was that volume estimates needed to be within 10% of roadway capacity to be useful for highway operations, and ideally closer to 5% (see [23], [44]). With this information in mind, and in coordination with the Coalition and other stakeholders, the research team identified several metrics that would be used to quantify estimation accuracy: Mean Absolute Percentage Error (MAPE), Coefficient of Determination (R2), Error to Theoretical Capacity Ratio (ETCR), Error to Maximum Flow Ratio (EMFR), with Symmetric MAPE (SMAPE) added later.

Although NREL and UMD teams used different probe data sources (TomTom and INRIX, respectively), focused on distinct geographical areas and estimated traffic volumes using different machine learning models, both teams worked closely to evaluate accuracy in the same way. High quality counts from trusted permanent count stations were used for both calibration and validation purposes, with a "leave-one-out" cross-validation procedure employed to quantify accuracy. In this framework, the volume estimator was calibrated at all count stations except one location that was held out and used to generate estimates at the held-out location. This process was repeated until each station in the dataset was held out and volume estimates were produced for all locations, thus allowing error metrics to be computed for each station and hourly period. As noted in [23], this project initially envisioned volume estimates being produced by probe vendors, with NREL and UMD teams responsible for conducting independent evaluation. As it turned out, both research teams ended up developing the estimation algorithms themselves, so the accuracy measures are self-reported rather than independently measured; however, the validation methodology and results were cross-checked internally across UMD and NREL research teams and independently reviewed by a technical steering committee.

Errors were reported for various scenarios (overall, by road class, geographical area, probe volume, etc.) but in general, NREL and UMD teams regularly reported errors within 5-10% of capacity, which are consistent with expectations identified by the survey. In earlier efforts in Maryland, Florida, and New Hampshire ([25], [26], [45], [46]), the UMD research team reported typical MAPE values around 25% for hourly estimates, with higher road class, higher average traffic volumes, and higher average probe counts associated with improved accuracy. In a more recent study in Harrisburg, PA [41], UMD produced 15-minute estimates that achieved typical MAPEs around 15%, with most estimation locations falling on freeways. The more limited estimation opportunities on principal and minor arterials showed reduced accuracy – closer to 30% error, yet still within 6-7% error relative to maximum flow. NREL's earlier work was focused primarily on Colorado (e.g., [24], [44]), but they also participated in the Harrisburg, PA study, providing an opportunity to cross-check results with the UMD team. They obtained similar accuracy at freeway locations (15% SMAPE, EMFR values close to 5%), thus demonstrating consistency in the level of accuracy. NREL conducted additional off-freeway analyses in several states and found mixed results; accuracy in PA was extremely degraded (60% SMAPE, likely related to data issues with short-term counts), but slightly more promising in subsequent locations (closer to 32% in North Carolina and 40% in Chattanooga, TN). Thus, while UMD and NREL teams demonstrated viable proof-of-concept solutions in freeway and other higher traffic conditions, it is likely that further follow-up study is needed on low-volume roads where probe vehicles are sparse.

TTI's evaluation of Streetlight volumes for Minnesota DOT

TTI conducted multiple evaluations of Streetlight volume data in coordination with Minnesota DOT, producing an initial report based on Streetlight's beta product in 2017 [29] and following-up in 2020 with a subsequent study [28]. In the initial 2017 evaluation, Streetlight's AADT estimates were calibrated using data from permanent counters and evaluated by the TTI research team at thousands of locations where MnDOT had short-term count data, with MnDOT's official AADT values – obtained by scaling up the short-term counts, used as ground truth. Additionally, Streetlight produced Average Annual Hourly Volumes (AAHV) at both permanent count sites where they had calibrated the AADT model and several new count sites that were not used for calibration. In both cases, TTI quantified accuracy using Mean Absolute Percent Error (MAPE), Mean Absolute Difference (MAD) and Mean Signed Difference (MSD). TTI reported the overall MAPE for AADT estimates to be 61% with accuracy varying significantly based on the traffic volume (34% to 68% depending on volume level), observing that lower volume roadways had the highest errors, Streetlight estimates were consistently positively biased (i.e., over-estimated the volume), and MAPE calculations were impacted by a small number of severe outliers. They also found that Streetlight's estimates of AAHV exhibited higher error when evaluated at locations that hadn't been used for calibration (49% versus 39% MAPE). Overall, they concluded that probe-based volume estimation has potential, but that analytic improvements were needed.

The follow-up report from 2020 [28] incorporated both permanent and short-term counters as reference sources, thus allowing a more accurate evaluation of ground truth AADT values. Although some short-term counts were still used to estimate reference AADTs, TTI sought to capture the uncertainty in the process by constructing a confidence interval of values based on MnDOT error thresholds. They used the same error measures as before and found significant improvements in accuracy relative to the 2017 study, concluding that Streetlight AADT accuracy is much better than before -- especially in mid-to-high volume ranges. For example, MAPE ranged from 8-10% (above 10k AADT) to 42% (<1k AADT) in the current study, as compared to 34% to 68% in the same ranges in 2017. Additionally, they found that there was noticeable overestimation bias on low volume roadways (< 5k AADT) – a trend that was noticeable in both permanent and short-duration count locations, but particularly prevalent for short-duration counts.

TTI's evaluation of Streetlight volumes at Border Crossings

TTI's Center for International Intelligent Transportation Research (CIITR) evaluated the accuracy of Streetlight's 2017 AADT estimates in US-Mexico border regions [31]. Their analysis encompassed two study areas: (i) ports of entry (POE) along the US-Mexico border and (ii) count locations in three border Texas DOT districts, with the former serving as the main study area. In the case of POEs, traffic volumes were obtained from U.S. Customs and Border Protection and local agencies, and subsequently aggregated to AADT values if reported at a more granular level. For border districts, traffic volumes were available from a mixture of permanent and short-term count sites, meaning that reference AADT values were obtained by aggregating counts over the year (permanent count sites) or applying seasonal adjustment factors (short-term count sites). Reference AADT values were compared with Streetlight AADT estimates using a variety of error metrics: Mean Signed Difference (MSD), Mean Absolute Difference (MAD), Mean Absolute Percent Error (MAPE) and Average Coefficient of Variation (ACV) and results were reported by volume level and urban/rural designation. The overall MAPE was calculated to be 33% and 50% for the POEs and border districts, respectively, with higher accuracy observed at higher traffic volumes and in urban, rather than rural areas. They also found that Streetlight AADT estimates were higher than reference AADT values (i.e., positive MSD) at lower AADT levels (under 10k vehicles/day) and lower (negative MSD) at higher AADT levels (above 10k vehicles/day).

TTI / Virginia Tech evaluation of Streetlight AADTs

As part of a project through the Safety through Disruption (Safe-D) National University Transportation Center, TTI and Virginia Tech Transportation Institute sought to characterize the quality of AADT estimates produced using alternative data source (i.e., probe data) and quantify the impact on safety analysis [33]. They evaluated 2017 AADT estimates from Streetlight and compared them to AADTs derived from traffic counts – both permanent and short-duration -- provided by two DOTs. The report acknowledges that reference AADTs obtained via short term counts have inherent estimation error, and thus may not be appropriate for validation purposes, clarifying that this study uses them merely as a point of comparison. AADT accuracy was quantified in terms of several error metrics: Mean Signed Difference (MSD), Mean Absolute Difference (MAD), Mean Absolute Percent Error (MAPE), Median Absolute Percent Error (Median APE), and Average Coefficient of Variation (ACV), reporting them separately by AADT range. They reported an overall Median APE of 77% but noted that it dropped to 25% when only focusing on roads with at least 2000 vehicles/day, as ADT accuracy improved when moving from lower to higher volume roads. Also, Streetlight AADT estimates tended to be higher than reference AADT values (i.e., positive MSD) when reference values were under 10,000 vehicles/day and lower than reference values (negative MSD) above 10,000 vehicles/day.

Oregon DOT's evaluation of Streetlight AADTs

Oregon DOT evaluated Streetlight's 2017 AADT estimates using reference data from both permanent and short-term count stations, quantifying accuracy in terms of Percentage Error and Absolute Percentage Error [32]. Results were reported separately for permanent ATR locations and short-term sites, with permanent counters representing the locations with the best quality reference AADT data. They found the mean and median absolute percentage errors at ATR locations to be 26% and 18%, respectively, with corresponding errors higher at short-term count locations (68% and 32%, respectively). Accuracy tended to improve at higher volumes, particularly at short-term count locations.

Louisiana Transportation Research Center's evaluation of Streetlytics AADTs

The Louisiana Transportation Research Center independently evaluated 2015 Streetlytics' AADT data product to determine whether their AADT estimates were accurate enough to use alongside current methods for systematically reporting AADT [40]. They used a combination of permanent and short-term counters in the Baton Rouge Metro Area to obtain reference AADTs for comparison with Streetlytics estimates, with short-term counts needing to be scaled to the AADT level via seasonal factors. All permanent count stations were "unobserved" by Streetlytics (meaning that they did not have access to prior AADT counts at that location) and short-term counts were split between observed/unobserved status. Two types of analysis were conducted: (i) a bi-variate correlation analysis to investigate the association between Streetlytics and reference AADT estimates and (ii) percentage difference calculations to quantify accuracy, reported separately for different AADT ranges. They found strong correlations between AADT estimates and reference values across both observed and unobserved locations (0.85 to 0.96), with varying accuracy measured reported for different conditions. The percentage difference at unobserved locations was 54% at short-duration count locations and 43% at permanent counters, with high values possibly attributed to Streetlytics reporting a minimum AADT value (300 vpd) that was too low on some rural roads.

Recommendations

After reviewing the existing literature, the TDM validation team makes the following recommendations:

Focus initially on conducting a blind evaluation of hourly, rather than aggregate volume estimates. This course of action is recommended for the following reasons:

- Hourly volume estimates represent the most novel volume data item in the TDM and are relevant for both planning and operations applications.
- Hourly accuracy is not well-characterized in the literature, with existing studies mainly limited to self-reported cross-validation results.
- Evaluation of hourly (and more granular) volume estimates is simpler than aggregate volumes because meaningful "unseen" reference data can be collected in a short period of time via various sources (permanent and short-duration sites)

Later, after characterizing the accuracy of hourly volume counts, it will be important to expand the evaluation to more aggregate volume products such as AADT. These aggregate measures are commonly used for planning applications, and there appears to be strong interest from State DOTs in understanding their viability.

Consider a cross-validation "audit" to complement blind evaluation of vendor data. According to the RFP, vendors are required to self-report cross-validation results, which show how their models perform at held-out locations. There are several advantages to auditing these results, particularly when done in parallel with blind evaluation:

- "Auditing the books" would allow the validation team to see the underlying data used to generate cross-validation results, which would promote transparency and help provide insight into model accuracy.
- If used in parallel with a blind evaluation, there is an opportunity to compare cross-validation accuracy to accuracy obtained via field measurements. Agreement between such values would help build trust that the accuracy reported through cross-validation is representative and generalizes well to other locations.
- If obtaining blind reference volumes is infeasible (e.g., a full calendar year of counts to compute AADT), it can minimally serve as a reasonable check on data quality.

Seek to understand the sources of error in reference volumes before using them to evaluate vendor data. Reference volumes can have several sources of error, including:

- **Measurement error**. Often this is assumed to be negligible at well-maintained and calibrated CCSs. Other permanent counters (e.g., intersection hardware, ITS sensors) and short-duration counters may have more noticeable measurement error, which can depend on several factors (e.g., technology, installation, weather). The Coalition should seek to understand errors associated with any sensor it uses – especially short-term count equipment it will deploy.

- **Error associated with scaling-up short-duration counts**. It is not recommended to use short duration counts for computing reference AADT volumes, but, if necessary, the analysis should take the uncertainty into consideration (e.g., via confidence bands).

Use a combination of permanent and short duration counts to obtain reference volumes. Permanent and short-duration counters each have pros and cons, and it can be beneficial to incorporate both in validation studies.

- **Permanent**: Incorporating CCS data as part of a validation strategy is ideal due to high accuracy and 24-7 counts that can be used to validate any type of volume product. However, it can require coordination with DOTs, as counts are generally publicly available, and thus difficult to withhold from vendors. Other non-CCS permanent counters may be viable too if accuracy is deemed appropriate.

- **Short duration**: Short-term counts can be a useful way to obtain volume data that vendors are unlikely to have used for calibration. They are well-suited for more temporally granular studies (e.g., hourly, sub-hourly) because the data can be used directly without needing to be scaled up. Note that DOT short duration counts will likely be conducted on a typical, “average” day, so if more interesting scenarios are desired, the Coalition should deploy its own portable sensors in strategic locations.

Compute a variety of error metrics and visualize the results. The literature indicates that there is no one “optimal” metric, and that it is valuable to report various measures to provide various insights into overall accuracy. As such, we make the following recommendations:

- Generate a scatter plot of reference vs estimated volumes to visually identify any systemic bias and gain an intuitive sense of the agreement (perfect agreement falls along the 45-degree line)
- Use the metrics in Table 1 as a starting point and pare them down to a useful subset for reporting purposes.
- When possible, consider the distribution of the error, not just the mean/median value. This can be accomplished via error percentiles or visualizing the distribution (e.g., via histograms, bar charts, box-and-whisker plots).
- Report metrics separately by scenario, including by volume level, urban/rural designation, and road class.

Recognize the need to continuously revise and improve validation methods. It is expected that the first few validation efforts will provide many learning opportunities. Based on these findings and feedback from State DOTs and data vendors, it will likely be necessary to refine the methodology to improve the validation process.

References

[1] M. Lowry, M. Dixon, and National Institute for Advanced Transportation Technology (U.S.), "GIS Tools to Estimate Average Annual Daily Traffic," KLK725, Jun. 2012. Accessed: Jun. 22, 2022. [Online]. Available: https://rosap.ntl.bts.gov/view/dot/24942

[2] S. C. Sharma, P. Lingras, F. Xu, and G. X. Liu, "Neural Networks as Alternative to Traditional Factor Approach of Annual Average Daily Traffic Estimation from Traffic Counts," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 1660, no. 1, pp. 24–31, Jan. 1999, doi: 10.3141/1660-04.

[3] E. Bagheri, M. Zhong, and J. Christie, "Improving AADT Estimation Accuracy of Short-Term Traffic Counts Using Pattern Matching and Bayesian Statistics," *J. Transp. Eng.*, vol. 141, no. 6, p. A4014001, Jun. 2015, doi: 10.1061/(ASCE)TE.1943-5436.0000528.

[4] I. Tsapakis and W. H. Schneider, "Use of Support Vector Machines to Assign Short-Term Counts to Seasonal Adjustment Factor Groups," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2527, no. 1, pp. 8–17, Jan. 2015, doi: 10.3141/2527-02.

[5] M. Dixon, "THE EFFECTS OF ERRORS IN ANNUAL AVERAGE DAILY TRAFFIC FORECASTING:," National Institute for Advanced Transportation Technology, University of Idaho, N04-12, Sep. 2004.

[6] T. Šliupas, "ANNUAL AVERAGE DAILY TRAFFIC FORECASTING USING DIFFERENT TECHNIQUES," *TRANSPORT*, vol. 21, no. 1, pp. 38–43, Mar. 2006, doi: 10.3846/16484142.2006.9638039.

[7] M. Castro-Neto, Y. Jeong, M. K. Jeong, and L. D. Han, "AADT prediction using support vector regression with data-dependent parameters," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 2979–2986, Mar. 2009, doi: 10.1016/j.eswa.2008.01.073.

[8] B. M. Williams, P. K. Durvasula, and D. E. Brown, "Urban Freeway Traffic Flow Prediction: Application of Seasonal Autoregressive Integrated Moving Average and Exponential Smoothing Models," *Transp. Res. Rec.*, vol. 1644, no. 1, pp. 132–141, Jan. 1998, doi: 10.3141/1644-14.

[9] G. Yu and C. Zhang, "Switching ARIMA model based forecasting for traffic flow," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, May 2004, vol. 2, pp. ii–429. doi: 10.1109/ICASSP.2004.1326286.

[10] B. Ghosh, B. Basu, and M. O'Mahony, "Bayesian Time-Series Model for Short-Term Traffic Flow Forecasting," *J. Transp. Eng.*, vol. 133, no. 3, pp. 180–189, Mar. 2007, doi: 10.1061/(ASCE)0733-947X(2007)133:3(180).

[11] S. Sun, C. Zhang, and G. Yu, "A Bayesian Network Approach to Traffic Flow Forecasting," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 1, pp. 124–132, Mar. 2006, doi: 10.1109/TITS.2006.869623.

[12] L. Zhang, Q. Liu, W. Yang, N. Wei, and D. Dong, "An Improved K-nearest Neighbor Model for Short-term Traffic Flow Prediction," *Procedia - Soc. Behav. Sci.*, vol. 96, pp. 653–662, Nov. 2013, doi: 10.1016/j.sbspro.2013.08.076.

[13] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic Flow Prediction With Big Data: A Deep Learning Approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 865–873, Apr. 2015, doi: 10.1109/TITS.2014.2345663.

[14] X. Wang and K. M. Kockelman, "Forecasting Network Data: Spatial Interpolation of Traffic Counts from Texas Data," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2105, no. 1, pp. 100–108, Jan. 2009, doi: 10.3141/2105-13.

[15] M. Lowry, "Spatial interpolation of traffic counts based on origin–destination centrality," *J. Transp. Geogr.*, vol. 36, pp. 98–105, Apr. 2014, doi: 10.1016/j.jtrangeo.2014.03.007.

[16] B. Selby and K. M. Kockelman, “Spatial prediction of traffic levels in unmeasured locations: applications of universal kriging and geographically weighted regression,” *J. Transp. Geogr.*, vol. 29, pp. 24–32, May 2013, doi: 10.1016/j.jtrangeo.2012.12.009.

[17] J. Wu and H. Xu, “Annual Average Daily Traffic Prediction Model for Minor Roads at Intersections,” *J. Transp. Eng. Part Syst.*, vol. 145, no. 10, p. 04019041, Oct. 2019, doi: 10.1061/JTEPBS.0000262.

[18] M. Chowdhury *et al.*, “Cost Effective Strategies for Estimating Statewide AADT,” FHWA-SC-18-10, Feb. 2019. doi: 10.13140/RG.2.2.19033.57440.

[19] S. Das and I. Tsapakis, “Interpretable machine learning approach in estimating traffic volume on low-volume roadways,” *Int. J. Transp. Sci. Technol.*, vol. 9, no. 1, pp. 76–88, Mar. 2020, doi: 10.1016/j.ijtst.2019.09.004.

[20] L. Schewel, S. Co, C. Willoughby, L. Yan, N. Clarke, and J. Wergin, “Non-Traditional Methods to Obtain Annual Average Daily Traffic (AADT),” FHWA-PL-21-031, Sep. 2021.

[21] H. Chang and S. Cheon, “The potential use of big vehicle GPS data for estimations of annual average daily traffic for unmeasured road segments,” *Transportation*, vol. 46, no. 3, pp. 1011–1032, Jun. 2019, doi: 10.1007/s11116-018-9903-6.

[22] X. Zhang and M. Chen, “Enhancing Statewide Annual Average Daily Traffic Estimation with Ubiquitous Probe Vehicle Data,” *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2674, no. 9, pp. 649–660, Sep. 2020, doi: 10.1177/0361198120931100.

[23] S. Young, “Real-time Volume and Turning Movements from Probe Data: Cooperative Research and Development Final Report, CRADA Number CRD-16-614,” NREL/TP-540-77458, 2020.

[24] Y. Hou, S. E. Young, A. Dimri, and N. Cohn, “Network Scale Ubiquitous Volume Estimation Using Tree-Based Ensemble Learning Methods,” Jan. 2018. Accessed: Jun. 29, 2022. [Online]. Available: https://www.osti.gov/biblio/1558882-network-scale-ubiquitous-volume-estimation-using-tree-based-ensemble-learning-methods

[25] P. Sekuła, N. Marković, Z. Vander Laan, and K. F. Sadabadi, “Estimating historical hourly traffic volumes via machine learning and vehicle probe data: A Maryland case study,” *Transp. Res. Part C Emerg. Technol.*, vol. 97, pp. 147–158, Dec. 2018, doi: 10.1016/j.trc.2018.10.012.

[26] S. Zahedian, P. Sekuła, A. Nohekhan, and Z. Vander Laan, “Estimating Hourly Traffic Volumes using Artificial Neural Network with Additional Inputs from Automatic Traffic Recorders,” *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2674, no. 3, pp. 272–282, Mar. 2020, doi: 10.1177/0361198120910737.

[27] N. Caceres, L. M. Romero, F. J. Morales, A. Reyes, and F. G. Benitez, “Estimating traffic volumes on intercity road locations using roadway attributes, socioeconomic features and other work-related activity characteristics,” *Transportation*, vol. 45, no. 5, pp. 1449–1473, Sep. 2018, doi: 10.1007/s11116-017-9771-5.

[28] S. Turner, “Evaluation of StreetLight Data’s Traffic Count Estimates from Mobile Device Data,” MN 2020-30, Nov. 2020.

[29] S. Turner, “Using Mobile Device Samples to Estimate Traffic Volumes,” MN/RC 2017-49, Dec. 2017.

[30] I. “Yianni” Tsapakis, S. Turner, P. Koneman, and P. Anderson, “Independent Evaluation of a Probe-Based Method to Estimate Annual Average Daily Traffic Volume,” FHWA-PL-21-032, Sep. 2021.

[31] I. Tsapakis, L. Cornejo, and A. Sánchez, “Accuracy of Probe-Based Annual Average Daily Traffic (AADT) Estimates in Border Regions,” p. 31.

[32] J. Roll, “Evaluating Streetlight Estimates of Annual Average Daily Traffic in Oregon,” OR-RD-19-11, Jun. 2019.

[33] I. Tsapakis, S. Das, A. Khodadadi, D. Lord, J. Morris, and E. Li, “Use of Disruptive Technologies to Support Safety Analysis and Meet New Federal Requirements,” SAFE-D: Safety Through Disruption National University Transportation Center, Report, Mar. 2021. Accessed: Jun. 29, 2022. [Online]. Available: https://vtechworks.lib.vt.edu/handle/10919/103081

[34] J. Fish, S. Young, A. Wilson, and B. Borlaug, “Validation of Non-Traditional Approaches to Annual Average Daily Traffic (AADT) Volume Estimation,” FHWA-PL-21-033, Sep. 2021.

[35] M. E. Hallenbeck, L. Schewel, S. Co, and J. Wergin, “Guidelines for Obtaining AADT Estimates from Non-Traditional Sources,” FHWA-PL-21-031.

[36] Federal Highway Administration, “Traffic Monitoring Guide,” Washington, DC, 2016.

[37] I. Tsapakis, W. Holik, S. Das, E. Kraus, and P. Anderson, “Informational Guide on Data Collection and Annual Average Daily Traffic (AADT) Estimation for Non-Federal Aid-System (NFAS) Roads,” FHWA-SA-20-06.

[38] S. Gadda, A. Magoon, and K. M. Kockelman, “ESTIMATES OF AADT: QUANTIFYING THE UNCERTAINTY,” p. 25.

[39] R. Krile, F. Todt, and J. Schroeder, “Assessing Roadway Traffic Count Duration and Frequency Impacts on Annual Average Daily Traffic Estimation: Assessing Accuracy Issues Related to Short-Term Count Durations,” Art. no. FHWA-PL-16-008, Oct. 2015, Accessed: Jul. 05, 2022. [Online]. Available: https://trid.trb.org/view/1442739

[40] J. Codjoe, G. Ashley, and W. Saunders, “Evaluating Cell Phone Data for AADT Estimation,” FHWA/LA.18/591, Jun. 2018. Accessed: Jun. 16, 2022. [Online]. Available: https://trid.trb.org/view/1524303

[41] “Volume Estimation using Machine Learning Approaches & Probe Vehicle Data: Case Study in Harrisburg, PA,” Apr. 21, 2020. [Online]. Available: https://tetcoalition.org/wp-content/uploads/2015/02/Coalition-VTM-Harrisburg_Results-21Apr2020-final.pdf

[42] R. Krile and E. Slone, “Evaluating Two Different Traffic Data Methods Based on Data Observed – Analysis of Provided Data – Final Report A,” FHWA-PL-021-040, Nov. 2021.

[43] “Exploring Non-Traditional Methods to Obtain Vehicle Volume and Class Data,” Transportation Pooled Fund Study TPF-5(384). Accessed: Jul. 07, 2022. [Online]. Available: https://www.pooledfund.org/Details/Study/636

[44] S. E. Young, Y. Hou, K. Sadabadi, P. Sekula, and D. Markow, “Estimating Highway Volumes Using Vehicle Probe Data - Proof of Concept,” 2017, p. 12.

[45] P. Sekuła, Z. V. Laan, K. F. Sadabadi, K. Kania, and S. Zahedian, “Transferability of a Machine Learning-Based Model of Hourly Traffic Volume Estimation—Florida and New Hampshire Case Study,” *J. Adv. Transp.*, vol. 2021, pp. 1–15, Nov. 2021, doi: 10.1155/2021/9944918.

[46] Z. V. Laan, “Statewide Traffic Volume Estimation using GPS Traces,” Detroit, MI, Jun. 06, 2018.