Guidelines for Obtaining AADT Estimates from Non-Traditional Sources

(Supersedes FHWA Publication: FHWA-PL-21-031)

FHWA-HPL-23-018

U.S. Department of Transportation

Federal Highway Administration

August 2024

Notice

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The U.S. Government assumes no liability for the use of the information contained in this document.

Non-Binding Contents

Except for the statutes and regulations cited, the contents of this document do not have the force and effect of law and are not meant to bind the States or the public in any way. This document is intended only to provide information to the public regarding existing requirements under the law or agency policies.

Quality Assurance Statement

The Federal Highway Administration (FHWA) provides high-quality information to serve Government, industry, and the public in a manner that promotes public understanding. Standards and policies are used to ensure and maximize the quality, objectivity, utility, and integrity of its information. FHWA periodically reviews quality issues and adjusts its programs and processes to ensure continuous quality improvement.

Disclaimer for Product Names and Manufacturers

The U.S. Government does not endorse products or manufacturers. Trademarks or manufacturers' names appear in this document only because they are considered essential to the objective of the document. They are included for informational purposes only and are not intended to reflect a preference, approval, or endorsement of any one product or entity.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA-HPL-23-018		

4. Title and Subtitle	5. Report Date
Guidelines for Obtaining AADT Estimates from Non-Traditional Sources	August 2024
	6. Performing Organization Code
7. Author(s) Robert Krile, Filmon Habtemichael	8. Performing Organization Report No.
9. Performing Organization Name and Address Battelle 505 King Avenue Columbus, Ohio 43201	10. Work Unit No. (TRAIS)
	11. Contract or Grant No. Contract 693JJ322A000010 Order 693JJ322F00296N
12. Sponsoring Agency Name and Address Federal Highway Administration Office of Highway Policy Information 1200 New Jersey Avenue SE Washington, DC 20590	13. Type of Report and Period Covered
	14. Sponsoring Agency Code

15. Supplementary Notes

This is the Task 2 report for the *Technical Assistance to Federal, State, and Local Highway Agencies for Effective Adoption of Big Data and Big Data Analytics Based Highway Traffic Parameters* project, which was performed in cooperation with the U.S. Department of Transportation, Federal Highway Administration (FHWA), Steven Jessberger (Task Manager) and a Technical Advisory Committee (TAC) consisting of Alaska DOT, Caltrans, Colorado DOT, Georgia DOT, Idaho DOT, Illinois DOT, Maryland DOT, Minnesota DOT, Nebraska DOT, New Jersey DOT, North Carolina DOT, North Dakota DOT, Oregon DOT, Ohio DOT, Pennsylvania DOT, South Carolina DOT, Texas DOT, and Virginia DOT.

16. Abstract

This document provides decision making guidance that State DOTs and local transportation agencies can use to plan and execute purchases of traffic volume estimates, such as annual average daily traffic (AADT), from the private sector when those estimates are based on non-traditional data sources, such as vehicle probe or smartphone data. It also provides guidance for validating the quality of such estimates. This report replicates some of the content but adds new guidance and expands on the methods of FHWA Publication FHWA-PL-21-031. As such, this document supersedes FHWA-PL-21-031.

17. Key Words	18. Distribution Statement
AADT, accuracy, precision, probe data, volume, Big Data, LBS, GPS, Machine Learning,	No restrictions. This document is available to the public.

19. Security Classif.(of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	96	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

GUIDELINES FOR OBTAINING AADT ESTIMATES FROM NON-TRADITIONAL SOURCES

INTRODUCTION

This document provides decision-making guidance that State DOTs and other transportation agencies can use to set specifications, assess quality, and procure traffic volume estimates such as annual average daily traffic (AADT) from the private sector based on non-traditional data sources, such as vehicle probe or smartphone data. It provides specific guidance for evaluating data quality from such providers with respect to initial and ongoing data validation efforts.

This guidance is the outcome of a Federal Highway Administration (FHWA) led pooled fund study (TPF-5(384)) that included eighteen state partners, representing the states of Alaska, California, Colorado, Georgia, Idaho, Illinois, Maryland, Minnesota, Nebraska, New Jersey, North Carolina, North Dakota, Ohio, Oregon, Pennsylvania, South Carolina, Texas, and Virginia.

The guide provides material designed to help agencies:

- prepare for purchasing data,
- interact with potential providers of those data,
- determine the quality and accuracy of the data being offered and evaluate the value of those offerings.

The remainder of this section discusses what AADT is and how it is measured, identifies advantages and limitations of AADT estimates from non-traditional sources, and then provides introduction to each of the four sections of guidance as:

A. Preparing to Purchase AADT Estimates
B. Data Required to Compare Bids
C. Testing the Quality of the Data
D. Understanding the Data to be Licensed

Accompanying checklists are provided for each of the four sections in Appendix A.

- Checklist A is designed to help an agency clarify their requirements prior to requesting bids or evaluating an offering.
- Checklist B is designed to make sure the bid(s) include the data required to compare different offers.
- Checklist C is designed to help judge the quality of the data being supplied and then confirm acceptable quality or to select between offers for best quality.
- Checklist D is designed to help the agency understand the technical procedures of the vendor providing AADT estimates.

What is AADT and how it is measured?

AADT is the mean daily traffic volume across all days for a year for a given location along a roadway. AADT is a basic measurement that indicates vehicle traffic volume on a road segment. It measures how busy a road is. It is used in many traffic engineering and travel planning applications, e.g., examining the performances of transportation facilities, monitoring traffic, supporting policy decision, allocating funds to transportation agencies, optimizing traffic operations, and performing safety and environmental impact analyses.

There are different methods of AADT computation. They include: 1) Simple average method, 2) AASHTO method (average of averages method), and 3) FHWA AADT method. Under the simple average method, AADT is estimated as the total traffic volume passing a point (or segment) of a road in both directions for a year divided by the number of days in the year. The AASHTO method incorporates 84 averages, i.e., 7 averages for the days of week for each of the 12 months. The FHWA method is a modification and improvement to the AASHTO method, and it incorporates weights to account for the occurrences of available traffic volumes for the hours of the day, days of the week, and months of the year. The FHWA method produces statistically significantly better results and improved AADT estimates. In computing AADT using the FHWA method, one first computes the monthly average daily traffic (MADT) for each month of the year and then the weighted average of the MADTs will be the AADT. The FHWA-recommended procedures for AADT and MADT computation are documented in the technical report *Improved Annual Average Daily Traffic Estimation Processes*.[1] The FHWA method is the recommended method of AADT computation and can be summarized as follows:

$$MADT_m = \frac{\sum_{j=1}^{7} w_{jm} \sum_{h=1}^{24} \left[\frac{1}{n_{hjm}} \sum_{i=1}^{n_{hjm}} VOL_{ihjm} \right]}{\sum_{j=1}^{7} w_{jm}}$$

$$AADT = \frac{\sum_{m=1}^{12} d_m * MADT_m}{\sum_{m=1}^{12} d_m}$$

Where:

$AADT$ = annual average daily traffic

$MADT_m$ = monthly average daily traffic for month m

VOL_{ihjm} = traffic volume for i^{th} occurrence of h^{th} hour of day within j^{th} day of week in m^{th} month

i = occurrence of a particular hour of day within a particular day of the week in a particular month (i = 1,…n_{hjm}) for which traffic volume is available

h = hour of the day (h = 1,2,…24) – or other temporal interval

j = day of the week (j = 1,2,…7)

m = month (m = 1,2,…12)

n_{hjm} = the number of times the h^{th} hour of day within the j^{th} day of week during the m^{th} month has available traffic volume (n_{hjm} ranges from 1 to 5 depending on hour of day, day of week, month, and data availability)

w_{jm} = the weighting for the number of times the j^{th} day of week occurs during the m^{th} month (either 4 of 5); the sum of the weights in the denominator is the number of calendar days in the month (i.e., 28, 29, 30, or 31)

d_m = the weighting for the number of days (i.e., 28, 29, 30, or 31) for the m^{th} month in the year

Traditionally, most AADT estimates are produced by on-site traffic counters. However, advancements in transportation and communication technologies have created an opportunity for collecting and processing large amounts of data from non-traditional sources (e.g., probe vehicle-based data) from private vendors. Therefore, transportation agencies have been exploring the feasibility and accuracy of AADT estimated from non-traditional sources.

[1] Jessberger, S., Krile, R., Schroeder, J., Todt, F., & Feng, J. (2016). Improved annual average daily traffic estimation processes. *Transportation Research Record*, *2593*(1), 103-109.

Advantages and Limitations of AADT Estimates from Non-Traditional Sources

The advantages of AADT estimated from non-traditional data sources are:

- It eliminates the need to carry out some traffic count surveys in the field.
- It creates an opportunity to reduce costs, save time, and improve efficiency.
- It reduces risks to employees and contractors who place sensor devices in and on the roadways to collect this data in the traditional way.
- Some traditional methods may be intrusive to the infrastructure and can disrupt traffic flow when installing, setting, and maintaining the devices. However, collection of probe-based data is not infrastructure intrusive.
- Non-traditional data sources allow transportation agencies to obtain the data they need for their traffic monitoring programs without the agencies having to own, repair, and maintain so many traffic counting devices.
- Non-traditional data sources can provide networkwide AADT estimates (better spatial coverage).

Although AADT estimated from non-traditional sources can offer many benefits, it also has some limitations:

- Accuracy of estimated AADT depends on the penetration rate of probe vehicles, the algorithms and models employed to convert the probe vehicle count to AADT estimates, and the benchmark data “big data analytics” require for learning. In situations where infrastructure has complex geometric layout (e.g., interchanges), mapping and conflation of the GPS pings of probe vehicles may be issues.
- Validation of the AADT estimates is also another challenge. However, with the increase in prevalence of probe vehicles, the non-traditional data sources and traffic parameters estimated from these sources are expected to be of better quality and more accurate.
- Another limitation is that state DOTs have no control over the non-traditional data sources used for AADT estimates. The predominant sources of the non-traditional data are vehicles equipped with location-tracking, mainly vehicle fleets in ride-sharing businesses, new vehicles from leading original equipment manufacturers (OEMs), fleets from the trucking industry, and crowdsourced through internet connectivity and mobility-related applications (e.g., from insurance companies and public transportation providers). If an agency entirely depends on non-traditional data sources, then any changes in the way these data are collected will directly affect the agency’s traffic data monitoring program.
- Other issues related to data being outside of agency control include regulatory and legal changes that affect navigation and location tracking technology, which is provided by Department of Defense’s satellite technology. There may be some liability and risks agencies will face due to changes in terms of use of the Department of Defense’s satellite technology, e.g., selective availability (intentional degradation of accuracy by the government) leading to changes in GPS ping rate and GPS localization precision that will directly affect an agency’s traffic monitoring program. Phone or vehicle location information can also be a risk, as availability of these data can change with government regulation and company data policy changes.

A. PREPARING TO PURCHASE AADT ESTIMATES

Before starting the purchase of non-traditional AADT estimates from the private sector, your agency needs to ensure that it has considered and fully understands the extent and quality of information that will be available compared to the agency's needs, data access and integration, and the process for negotiating, procuring, obtaining, and getting support from the applicable vendor. Each of these areas of consideration are highlighted by a few key ideas. These ideas are then associated with a series of Questions, A.1 through A.7, as referenced in the text.

Extent and Quality of Information Being Purchased

- Determine what data items are required and how they will be used. If the data are for the Federal-aid highway program, the FHWA Division office should be consulted. Much of the guidance in this document is aligned with the Federal-aid highway – Highway Performance Monitoring System (HPMS) segment AADT data, but other transportation parameters such as class-based AADT, peak volumes, and average speeds could also be considered important. (Question A.1)
- Identify the latency and frequency of data delivery that you will require, understand the vendor's capabilities relative to your needs, and consider implications if there are variances. (Question A.2)
- Establish the accuracy and precision standards that are expected of the data to be provided. Once data are received from the vendor, these standards may include your agency's existing data quality standards for legacy data, but if specific expectations are required for the vendor, these will need to be negotiated at the time of purchase/renewal. (Question A.3)

Data Access and Integration

- Consider your staff who currently obtain, process, and validate this information. Will these people be able to access these new data, will they require any training, will this new data source change their labor expectations, and how many licenses or installations of software will be necessary? (Question A.4)
- Identify the frequency and process by which the data will be accessed, especially as related to data currently being used. (Questions A.5 and A.6)
- For AADT and other traffic data parameters with geocoding, an important consideration is how the geocoding, as provided by the vendor, is reconciled with the internal data it is meant to enhance or replace (e.g., HPMS segment data). Performing this reconciliation may require use of vendor support, so the nature, extent, and responsiveness of this support should be defined. If internal staff will perform this function, their training, expertise, and level of effort should be considered. (Questions A.5 and A.6)
- Define the form of these data (e.g., *.CSV files, ASCII text files) and what processes will be required to integrate them with existing agency data systems. (Questions A.5 and A.6)

Negotiating, Procuring, and Obtaining Data and Support (Question A.7)

- What kind of pricing models are available for obtaining AADT estimates?
- How much opportunity is there to negotiate the pricing?
- What is the procurement process?
- How long is the time from purchase of AADT estimates to delivery of data?
- Are sample data available to evaluate the AADT estimates and experiment with it?
- What is the nature, extent, and responsiveness of vendor support both to integrate the data into the agency's systems and assess initial quality, and then to address subsequent issues as ongoing data delivery occurs.

A.1) What data do you wish to obtain and how do you intend to use the data?

When answering question A.1, your agency should know where the data will be used and should understand FHWA Division and HQ Office of Highway Policy Information guidance on data utilization procedures and steps. FHWA will share their knowledge and expertise and may also coordinate interstate knowledge exchange. Some examples of data uses are listed below.

1. Meet very specific project analysis needs (e.g., validate/calibrate link volumes for traffic demand models; determine existing traffic volumes for design purposes; estimate turning movements or other data required for operational analysis; perform safety analysis on specific road segments).
2. Replace and/or complement existing short-term traffic counts, allowing the agency to reduce the number of short-term counts they perform.
3. Provide AADT values on road segments for which recent short-term counts are not available to supply vehicle miles traveled (VMT) estimates as part of PM3 reporting.
4. Conduct trend analysis for a given segment of roadway and determination of temporal factors.

The data being purchased may be needed for more than one of these reasons. The purchasing agency then needs to understand how the "controlling" requirements from these uses affect its purchasing decisions.

If the plan is to significantly reduce or replace the current data collection program, then your agency may need to ensure that the successful vendor can provide the variables needed by users that would otherwise be provided by the short-term count data (e.g., time of day volumes, truck volumes, percent peak trucks and/or directional factors). The need for traffic volume variables besides AADT should be discussed with the intended users of the data and that list of variables must be included in the request for proposal (RFP) to vendors.

Preparing RFP to vendors:

1. Define your project, scope, and budget.
2. Provide background and introductory information.
3. Describe the services you're looking for.
4. Set the requirements that need to be met.
5. Detail your selection criteria and timelines.

A.2) How often will the data be delivered?

If data delivery is required only for HPMS reporting, the latency of approximately six months after the end of the calendar year and a single delivery will suffice. If data to be purchased will be utilized for other operational activities, determining the timing of data deliveries (e.g., daily, monthly, quarterly) is important. Additionally, each time period for which a data delivery is needed will have a delay from when that time period ended. This delay should be defined by the vendor so the agency can compare against the similar latency of its legacy data to determine if this presents any challenges.

In addition to ensuring that your agency receives data it needs in a timely fashion, this question helps an agency understand the level of automation required to accept the data. The more frequently data will be transferred to the agency, the more automated the data acceptance and upload process needs to be. For example, in a once-a-year process, agency staff can download a compressed file from a secure server, after having the file name and server password noted in a phone call. Such an approach would not be feasible if data were being downloaded daily.

A.3) What data quality requirements must the data meet and how will these be determined?

After establishing the content and timing of data required, it is critical that the agency defines the quality requirements of the data it requires from the private sector. If the data are being purchased to replace all, a significant part of, or a limited number of the agency's current short-term count program, then the AADT estimates should meet or have better accuracy and precision levels as those of the current annualized short-term counts. Accuracy is defined as the "average" amount that the candidate estimates for a set of sites vary from the "true" values, while precision is the variability of these estimates. Accuracy and precision may be calculated in several valid ways.

Table 1 provides guidelines for the accuracy and precision of AADT estimates computed from short-term count programs as determined by research for this project, updating and enhancing work from a previous FHWA pooled fund study.[2] The values from this table are presented in terms of Traffic Count Error (TCE), as a percent, where TCE is the difference in count (e.g., AADT) between a candidate traffic estimate (e.g., probe data) and a "true" AADT determined from a continuous count. When applied to a number of sites, the TCE values can be arranged from smallest to largest. The middle value (median) represents an estimate of how much the TCE is different, on average, from the true value. The range (difference) between the 2.5th percentile estimate and 97.5th percentile estimate represents an estimate of how much the TCE may vary, with 95% probability, from site to site. The values indicate, for instance, that short-term counts factored to AADT for roadways in the 5,000-54,999 AADT range should have a median TCE (i.e., the middle value for a large number of sites), with 95 percent probability, that is no more than ± 2.0 percent from the true AADT, when evaluated with a sample of 1,000 sites. For the same volume range, the upper 95 percent confidence bound on the 95th percentile of absolute TCE (i.e., Mean Absolute Percent Error, or MAPE) is 10.5 percent with factored short term counts. These accuracy limits are indexed to the number of sample sites and can be determined for a particular sample by using the equations in the table. Development of these standards is provided in Appendix B.

For precision, the standards for sites with AADT in the 5,000-54,999 range show that with 95 percent probability, 95 percent of short-term count TCEs are expected to be within 33.6 percent

[2] Assessing Roadway Traffic Count Duration and Frequency Impacts on Annual Average Daily Traffic Estimation Assessing AADT Accuracy Issues Related to Short-Term Count Durations, by R. Krile, F. Todt, and J. Schroeder. FHWA Publication No. FHWA-PL-16-008, October 2015

below the true AADT to 33.6 percent above the true AADT. This paper includes a method for using these limits that is adjusted to the sample size of number of sites in the evaluation, so only a single set of standards are needed. This procedure and the development of these limits is discussed further in Appendix B.

Table 1. Accuracy and Precision Limits Based on Historical Short-Term Counts Factored to AADT When Compared to Continuous Count Station Reference Data

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

†Roadways with AADT less than 500 have no pre-defined standards.

Accuracy assessment with less than 25 sites is not advised as the uncertainty will be too large. For sites with AADT < 500, there was insufficient data in this research dataset to establish limits, but these would be expected to be at least as variable as those shown in Table 1 for sites with AADT of 500-4,999, so these limits could be used conservatively for such sites. Alternatively, the methods used to develop these limits (as documented in Appendix B), could be applied to a dataset with these low AADTs to produce such limits.

Data validation may be desired to include sites or locations for which no continuous count data are available, but for which portable counts are available. The guidance for this scenario is included in Appendix B under the section, Application of Historical Accuracy and Precision Experienced When Using Short-Term, 48-Hour Portable Counts for Validating Alternative AADT Estimates. This section includes different approaches, but one of them includes a set of limits analogous to those of Table 1 where the user will compare a candidate AADT estimate to that of a portable count factored to AADT at that same site. Such factored portable counts have more uncertainty than that of continuous count site values, but the limits have been appropriately relaxed to reflect this, and the statistical properties of the limits are comparable to those of Table 1. This procedure is very useful since states often have many times more portable count sites, and these may cover a larger range of highway functional classes, than the limited number of continuous count sites. Additionally, the portable counts are less likely to have been directly incorporated into the statistical models of the probe-based data provider, and hence are more likely to be fully independent results.

The Table 1 standards may be stricter than the true TCE characteristics for any specific agency and time period. For instance, it is known that use of continuous traffic counters produces a more accurate AADT than that of temporary counts, but even the continuous count AADT values are subject to some error. This error was not directly considered in the development of the Table 1 ranges. Additionally, the methodology of subsetting continuous count station (CCS) data to simulate short-term counts and applying corresponding factoring to estimate AADT was based on a relatively small sample of sites (i.e., 206). This was especially an issue with low-volume count sites, and FHWA suppressed a standard for sites with AADT below 500 as only one test site would have contributed to the estimate. Accuracy and precision standards derived from a direct comparison of true portable counts compared to a CCS at the same location would likely be wider than those shown in Table 1 due to the additional variability in portable counter equipment, algorithms, and calibrations as compared to the more robust CCS installations.

For official federal reporting, such as required AADT on highways submitted through the HPMS, FHWA has standards for how these estimates are performed with short-term counts and has accepted the observed level of uncertainty through systematic bias or random variability like that of Table 1. When agencies consider alternative technologies to produce the same data, FHWA will consider these data acceptable assuming they have met standards like those provided in Table 1. Agencies may use other values than Table 1 if supported by their own research when reviewed and accepted by FHWA.

In addition to the accuracy and precision of AADT estimates by non-traditional measures, agencies often include other data quality measures. These may include trend evaluations or outlier checks. The new measures should be subject to the same data quality evaluation as the data they are to replace. If the data are to supplement other reporting, similar data quality procedures should be applied, though the thresholds for acceptance or rejection may be modified or defined differently. For instance, AADT on very low-volume roadways may not be well suited for data quality checks on percent changes over time. The specifics of the application of these data quality standards are provided in Section C and in Appendix B.

If the data are being purchased only to meet the needs for a specific project, the accuracy requirements for that project should be used to drive the accuracy acceptance testing of vendor's data. If the data will be used only to provide volume estimates in places your agency currently is not able to develop AADT values of known quality, or if the AADT values will be supplied for local roads that have very low volumes, then the agency may need to adjust the quality levels shown in Table 1 (although Table 1 is the preferred outcome for all AADT estimates.) Agencies will need to develop their own acceptance criteria that meets agency and FHWA (when reported to FHWA) needs for these data. A suggested way the agency could generate acceptable tolerances for data to replace the estimates currently used for these "uncounted" locations would be to perform an analysis of the error currently present in these estimates by counting a sample of these locations, expanding those data to AADT following their standard procedures, and comparing the "true" value with the estimates currently being used. The acceptance criteria would then be data equal to or more accurate than the current estimates shown in Table 1.

Determining the uses for the data gives the agency the opportunity to have an open discussion among its staff about traffic data needs across the agency. The accuracy requirements for the data being purchased (most likely the values in Table 1) are the most important outcome of those discussions.

A key aspect of purchasing private sector data is acceptance testing those data to ensure that they meet the specifications identified in response to question A.1. Agencies can perform data quality analysis through a wide range of approaches including but not limited to:

- utilize independent third-party certification if such entities exist,
- perform validation of the vendor's data accuracy using either agency staff or outside assistance (e.g., consulting firm/university) by comparing data submitted by the vendor against a validation dataset (see Appendix B for a complete example),
- trust the vendor's data quality report.

Appendix B describes a rigorous method for performing validation tests. A cross validation[3] analysis which combines the ideas in the bullets provides an additional approach that is common to the review of "big data" machine learning analyses.

The first approach is the least expensive to the agency, while not relying entirely on the vendor. It assumes that tests performed by an independent agency or institution such as a university (not the vendor) for some other state or agency, under a well-defined protocol, can indicate the accuracy that the vendor will deliver for this purchase. The disadvantage of this approach is that the accuracy achieved elsewhere may or may not be similar to that achieved for your agency, as accuracy achieved by any given vendor will depend not just on its technical approach, but a variety of factors such as the size of the dataset available to calibrate the vendor's models/algorithms and the characteristics of the state (e.g., population demographics, penetration rates of specific cellular providers, density of cellphone towers).

The second option is most specific or tailored to the agency's needs. However, it requires that your agency have ground truth volume estimates and have the resources to perform analyses with those data. This can be difficult, in that many states post their permanent count data online and those data are thus readily available to potential vendors for their use in calibration activities. If used by the vendor, those data are no longer independent data sources that can be used for conventional testing.

The only real difference between using your agency staff and hiring outside assistance is whether your agency staff have the time and expertise to perform the validation work. If not, then hiring outside assistance is necessary. Appendix B includes the preferred method for computing independent accuracy statistics.

If the second option is selected, your agency must provide ground truth data against which to measure the accuracy of the vendor's data. In some cases, vendors will also wish to calibrate its algorithms by using AADT values from ground truth datasets that are like those needed for validation testing. The data provided to vendors for model calibration purposes should not be used in the accuracy validation tests. Thus, it is important for the agency to have access to a relatively large pool of ground truth AADT statistics that can be used for the accuracy testing process when those tests will be performed on the state-specific data being provided by the

[3] https://en.wikipedia.org/wiki/Cross-validation_(statistics)

successful vendor.

The current best sources of ground truth data are well calibrated, properly working, continuously operating traffic count locations that have gone through some data cleaning or quality assurance review. Your agency may need to allocate its available count locations between these two functions (calibrating vendor algorithms and validating the results of that calibration effort). It is important to note that CCSs from other agencies (such as other states with similar traffic patterns or other metropolitan planning organizations [MPOs] in your state) may be able to be used for these purposes. The research project, Exploring Non-Traditional Methods to Obtain Vehicle Volume and Class Data, Pooled Fund project TPF-5(384)[4], evaluated the technical and statistical validity of traffic data derived from non-traditional data sources using machine learning models. The project compared AADT estimated from machine learning models with data collected from 4,255 permanent counters and has documented AADT estimation errors by road volume, roadway, and regional characteristics. Overall, results indicated that AADT estimates from non-traditional data sources, with some limitations, consistently out-performed same-year temporary counts for roads over 2,000 AADT. The results were mixed for roads between 500 and 2,000 AADT. For roads under 500 AADT, there were not enough reliable sources to calculate the error of same-year temporary count.

The third option is easy and inexpensive, but is not recommended, as it provides no confirmation that the validation report provided by the vendor applies to the purchased data in the given state, as some probe-based models have worked well in some states, but poorly in others.

If the approach discussed in Appendix B is not possible due to lack of data, this guide recommends using a combination of the bullets above. The alternative to Appendix B is to require the vendor to share the results of their cross validation process to verify the accuracy of the vendor's estimates. Cross validation is typically performed by the vendor as part of developing their model and provides estimates of each model's accuracy. The technique involves splitting the available data into "training" and "validation" datasets. The model is developed/calibrated on the training dataset and the resulting model is used to predict the data in the validation dataset, with the difference between the model outcomes and the validation dataset being a good measure of the error in the model. This process (split the dataset / build model / check errors) is performed multiple times, with different datapoints being included in the training and validation datasets for each pass through this process. A summary of all error estimates from these tests provides an estimate of the performance of the model.

Tests of cross validation should be specifically oriented to the calibration of the vendor's model to the state purchasing access to the data.

Having an independent group audit the vendor's cross validation process, confirm the validity of the model outputs, and report on the error rates achieved by the offered model is a good approach for limiting both the cost of the independent verification process and the need for independent data sources for performing that verification. Use of an independent group for performing the audit both allows the agency to access individuals with skillsets the agency might not otherwise have (e.g., machine learning calibration and testing experience), and ensures that the audit is viewed as being independent by both the vendor and the purchasing agency. This

[4] https://www.pooledfund.org/Details/Study/636.

approach does require the vendor to provide access to their model calibration process and results to the independent auditor. This can be performed under nondisclosure agreements (NDAs), so long as the results of that audit can be provided to the agency requesting that audit. The independent auditor must follow approved validation methods and must report in detail to the hiring agency any issues identified in the vendor's technical process, so that the resultant accuracy test results are clearly understood.

A.4) Who will use these data?

Because AADT statistics are a key input to many important analytical tasks, they are the foundation of the highway data program. AADTs are heavily used by the HPMS, your agency's safety management system, pavement management system, roadway design process, performance reporting system, safe routes to school, railroad crossing, bridge inventory reporting, overall planning process for the agency, and a wide variety of analytical tasks and tools. Understanding the analyses for which the purchased data will be used also defines who will be using those data and the technical procedures needed to access those data.

The questions below, along with the checklist, will help determine the data users.

- Do all agency staff need to access these data, or will only a small subset of agency personnel working on a specific project need access?
- Do these data need to be shared outside of the agency, for example with MPO, city, or county staff?
- Does the agency need to be able to share the data with the public?

The answers to these questions will define how the data will need to be accessed and stored, as well as the needed data rights. Examining the following four bullet points will help your agency formulate your RFPs as well as evaluate vendor responses.

- Data rights – Who has ownership, usage, publication, and distribution rights to the data?
- Access to the data – Is your agency provided with a copy of the data or is access to the data on a case-by-case basis via a web-platform? If a copy is provided, via what file format (e.g., CSV, shapefile, database file, etc.) and what location referencing system will be used (open street map (OSM) or the FHWA/HPMS - ARNOLD)?
- Archiving the data – In the event a copy of the data is not provided, how will the data purchased be accessed in the event the web-platform usage right is terminated at some point in the future?[5]
- Data integration – How will your agency integrate the provided traffic data into its current software systems? This includes the need to conflate purchased data to the location referencing system used by your agency's software systems.
- Versioning control – Each dataset needs to have a model version and date of extract so that the agency can properly reference changes in the future.

A final issue that must be identified in the RFP is whether there are contractual limitations on who can see and use the volume data being purchased. Many data providers contractually limit

[5] It is recommended that the agency ensure that they get to maintain, in perpetuity, access to any data used for federal reporting purposes, to meet various Freedom of Information Act or open government regulations.

who can use their data to create a market for additional sales of that data to other agencies or companies. **Thus, if the data must be released to the public, the RFP used to purchase those data needs to be explicit that this sharing is required.** This could raise the cost of the purchase. In addition, if either outside agency staff or the public will be accessing the data, the agency needs to understand how that access will be supplied.

A.5) How do the data need to be delivered from the vendor?

Understanding the answers to questions A.1 and A.2 helps define the parameters that determine the most cost-effective mechanism for obtaining the purchased data from the vendor. If agency staff will be accessing vendor data from the vendor's platform, then including a live demonstration of that portal in the bid process may be important to allow your staff to determine how easily they can find and download the data they need. If your agency intends to purchase large (e.g., statewide) quantities of data for use within the existing analytical processes the agency already performs, it is important to work with your agency IT staff to define how they wish the data to be delivered.

Short-term counts are location specific. Probe-based volume estimates are typically road-segment specific. Thus, a key aspect of the data processing task associated with integrating purchased data with agency systems is linking the purchased data to the correct roadway segments being used by your agency. This process (commonly called "conflation") may or may not be complex depending on the location referencing systems currently used by both your agency and the vendor. At a minimum, your agency will need to understand the conflation process and be able to provide prospective vendors with information on the roadway segmentation of your highway system. Vendors may or may not be able to directly match your roadway segmentation precisely. Understanding the resources required from your agency to assign the purchased volume data into your required location referencing system is an important factor when selecting between vendors.

There are typically three ways in which large network data are transferred between data systems:

1. The agency delivers a basemap to the vendor to populate with volume data.
2. The vendor provides the agency with a standard basemap (e.g., Traffic Message Channel - TMC, Open StreetMap - OSM, or FHWA's ARNOLD) and the agency must translate/conflate the data from that mapping system into the format desired by the agency.
3. The agency can provide geodetic locations and direction (heading) for which volume data are needed and the agency assigns those geolocated datapoints to the appropriate roadway location (e.g., road segment or route and milepost.)

A.6) How will the data be uploaded to the agency's corporate data system?

The answer to question A.1 describes which data (in addition to AADT) are required from the vendor, whether the purchased data need to reside inside the agency's corporate data structure, and whether data items other than AADT will be part of the data purchase. The answer to question A.3 describes how the data will be obtained from the vendor and matched with your agency's location referencing system.

This question (A.6) starts with the answers from A.1 and A.3 and recommends that your agency's IT staff (or technology and data team) identify the tasks and software required to upload the data being obtained into the corporate data system. This includes their need to understand what data will be uploaded (e.g., extract version, directional volumes, definitions of direction of travel, whether hourly volumes will be provided (average daily traffic (ADT) values for specific days and locations, monthly average daily traffic) and how those variables map to the data currently stored in your agency's corporate data system, as the data you purchase need to be accessible by your existing analytical tools without having to make changes to those systems.

This information allows your IT staff to build conversion, quality assurance, and data upload scripts that check the received data and import valid data from the vendor into the corporate data system.

Finally, a key part of this task is to develop and implement data quality checks for the outside data loaded into the corporate data system. This will likely to be a joint task of the central traffic office, which will oversee developing the acceptance testing rules, and the IT staff who need to code those rules into the data acceptance and upload process.

A.7) Negotiating, Procuring, and Obtaining Data and Support

Purchasing non-traditional AADT estimates from third-party vendors can be a complicated process that includes technical, legal, and institutional issues. Overall, it entails issuing a RFP to collect bids from qualified vendors followed by issuing a purchase order (PO) when the agency is ready to purchase the AADT estimates. The purchase process can be generalized as shown in Figure 1, although this process can change from state to state.

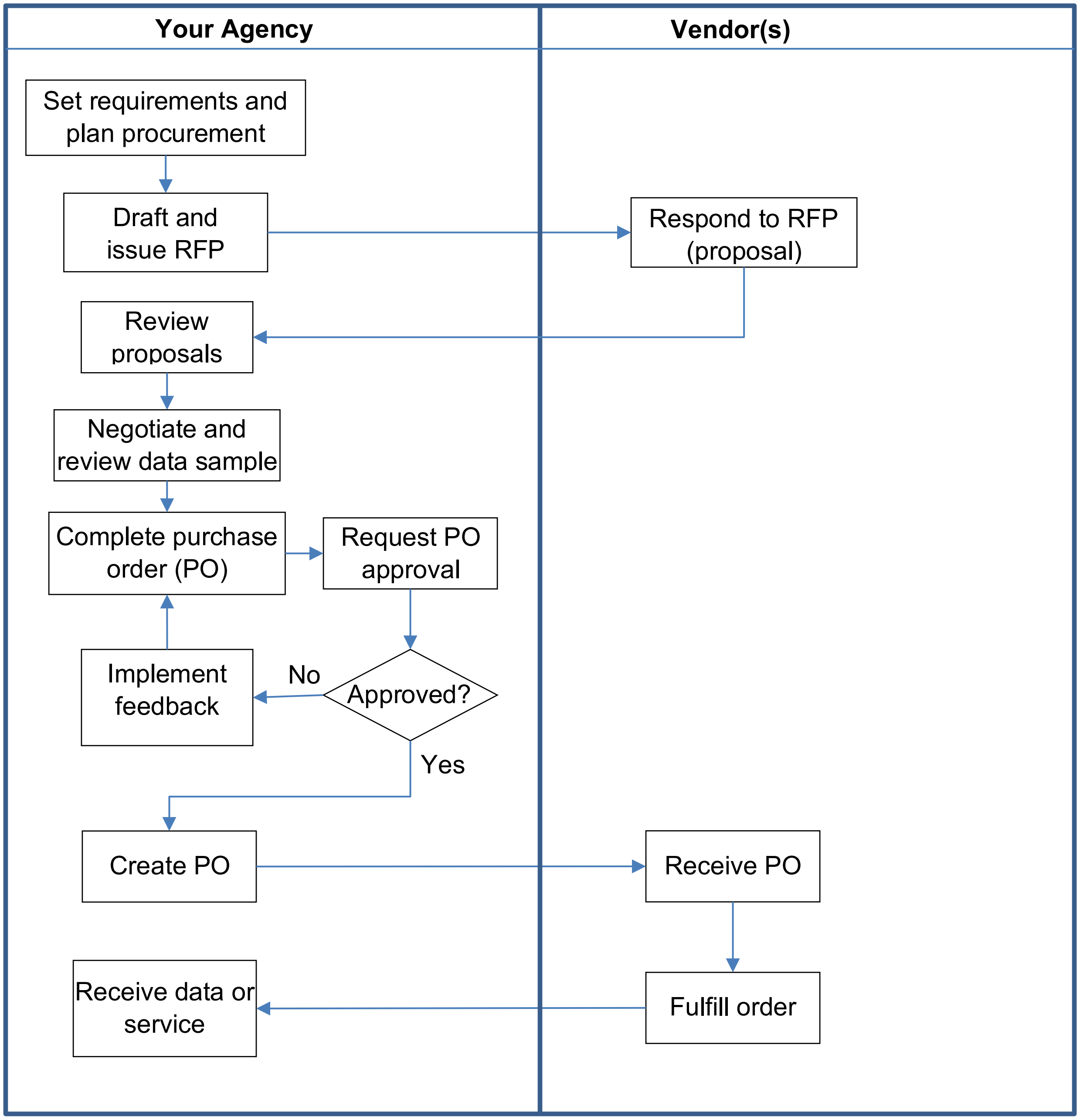

Figure 1. Generalized process of AADT purchase

B. DATA REQUIRED TO COMPARE BIDS

B.1) What is the purchase price?

Every vendor's response to the agency's RFP will include a price. Understanding both the total cost of the vendor's bid and any ongoing expenses is key to understanding the vendor's full cost. For example, will new software licenses be required? What will be the cost of additional licenses if they are needed?

The agency also needs to consider the costs it will impose on itself in terms of the work required by agency staff or consultants to assist in calibrating the vendor's algorithms, validating the obtained data, and (potentially) refining existing agency software systems to import the purchased data into the current corporate data structure.

B.2) Who owns the data you have purchased or licensed?

Determining whether the agency "owns" a copy of the data or "rents" the data will help determine the agency's ability to share those data with other agencies or the public. The structure of the "rental" license also helps determine what will happen if the agency changes data vendors at some point in the future. Will your agency retain a copy of the data indefinitely, as well as the right to use it? (This is highly recommended!) Or do they lose access and rights to use to the data they have been using? This is particularly important if the data are stored on a vendor's server and accessed only remotely. In such a case, the concern is an audit issue, in which access to data from "last year's vendor" may be needed to confirm the data used in an analysis, but that vendor may no longer be under contract. Also important is the ability of your agency to share that data. Can your agency further distribute the data to public partners? Can they be released to the public? Are there restrictions to the degree they can be released? Many companies restrict access to their data to maintain a market for further sales. Such restrictions can be problematic when they conflict with state public records laws which require publicly used data to be available to all members of the public.

B.3) What is included in that price?

The answers to this question are used to compare the value obtained from competing vendors. Value is determined by both the price and what is obtained for that price. The secondary questions in the checklist (Appendix A) are designed to help your agency identify the attributes offered in a bid. Some of these items will be "required" in the bid. Others may be offered by vendors as part of their data package. Your agency will need to determine the value of "extra" data provided by different bidders, as well as the significance of data "missing" from the list of required or desired traffic variables.

As part of this consideration, your agency will need <u>estimates</u> of the accuracy of each traffic variable being provided. These may be provided by the vendor as part of the bid. These estimates are used for the initial valuation of the bids. Your agency can then decide to validate those statistics as it sees fit.

Traffic statistics other than AADT (e.g., SU AADT, CU AADT, D factor, K factors, or other statistics) may or may not be of sufficient accuracy to meet the needs of the agency, but determining the value of these secondary statistics is important when determining the relative value of a vendor's proposal. (How accurate are they? Do they meet specific needs of the agency? Do they reduce either other data collection costs or safety concerns to the agency?)

The value of additional traffic variables included in a bid, but not part of the minimum requirement in the RFP, will need to be determined as part of each bid review.

Other considerations your agency should include in its valuation of each vendor's bid include the following:

- How many years of data are included in the bid?
- Who can use the data being licensed?
- How many people may access the software and/or data?
- What is the geographic and roadway system extent included in the bid?
- What type of vendor support is provided as part of the bid? How does the vendor propose to handle outcomes in which quality assurance checks identify counts that are illogical or potentially in error?
- Does your agency have a choice in terms of the location referencing system the vendor uses to supply the data?

Your agency will have to determine the value of these additional bid attributes. For example, choosing one referencing system option over another may greatly reduce the cost of uploading the data into your corporate data system, and thus provide considerable value.

B.4) How accurate are the provided data?

This is a vital statistic for comparison between vendors. The checklist assumes that AADT is the primary statistic being supplied by the vendor. The accuracy of this statistic from each vendor is therefore of paramount importance, but the inclusion and accuracy of other traffic statistics should also be considered. A good set of error statistics to request from the vendors are the cross-validation results from each vendor that best approximate the data that will be provided for your state. (So, if the vendor has already produced estimates for your state, then use those accuracies. If they have not, chose a nearby state with similar traffic patterns.) Independent measurements of AADT accuracy provided by independent sources (e.g., other state agencies, independent testing organizations) can be used during the initial vendor selection process. However, these values should be validated for your state (see Checklist C).

Purchased data needs to meet the accuracy criteria being set by the agency and described in the answer to A.1. However, the agency should consider the relative importance of data accuracy levels better than the requirement, as well as the accuracy and importance of the other traffic statistics provided as part of the bid.

A key aspect of the RFP and purchase agreement is to include how accuracy will be tested, the expense to the vendor, and what will happen if the results of those tests demonstrate that the results for the selected vendor fall wholly or partially outside of the bid specifications. For example, how will your agency treat the bid of a vendor that produces acceptable accuracy validation test results for roads with AADTs of above 5,000 AADT but not below 5,000? Additionally, data quality may vary over time and some consideration should be provided for what recourse the agency will have if data for a particular time period is unacceptable. These consequences should be spelled out in the RFP evaluation process.

B.5) What data are required to calibrate the data for delivery to your agency?

Most vendors will need access to accurately functioning, continuously operating traffic count data to calibrate their models/algorithms. Access to these data will likely be needed not just for the initial calibration but also for future updates to that calibration process, as expected changes to the input data the vendor uses (e.g., penetration rates for various cellphone providers, use of the apps that provide location-based service (LBS) data, and which apps are included in the vendor's data feed) mean that vendors will need to periodically update their models/algorithms.

If data are purchased for more than one delivery (e.g., the bid will supply data not just for the past three years but also for the next three years, or monthly AADT updates will be supplied for the next 24 months), then the vendor will likely want to re-calibrate its algorithms multiple times during the data delivery period. This means that the vendor will need updated calibration data. Each vendor may require that a different set of data be provided to calibrate its algorithm to the state purchasing data. The cost (and ability) of providing those data should be compared between vendor bids. If the agency is unable to supply the data some vendors require to calibrate their estimates it then becomes necessary to learn 1) how the lack of those data will change the expected level of AADT accuracy, and 2) whether that will change the cost of the vendor's bid.

Vendors may need data for a specific number of calibrated, permanent, continuously operating count sites for each geographic area and/or roadway classification requested by the state or included in the vendor specifications. The vendor may need multiple years of data for each of these locations. In addition to permanent count data, vendors may also ask for other datasets to be provided in support of its calibration or algorithm development process. For example, does the vendor need just the AADT value for each site? Daily ADT values? Hourly records? The cost of producing these data must also be included in the cost of each bid.

B.6) How will the data be delivered?

See Question A.5 for each potential vendor.

If agency staff will be accessing vendor data from the vendor's platform, then inclusion of a live demonstration of that platform in the bid process is important to help your agency determine whether a vendor's system is markedly better or worse than that of competing vendors. That is, how intuitive is the interface, so that agency staff can quickly and efficiently obtain the data they need?

The mechanism used by different vendors to deliver data may increase or decrease agency costs for obtaining data. These costs should be considered in selecting the best value to be obtained from the competing vendors.

B.7) What is required to enter the data into your agency's corporate data system?

See Question A.6 for each potential vendor.

Questions B.3 and B.6 will provide the information required by agency IT staff to determine the cost of incorporating the purchased data from each vendor into the agency's corporate data system. These costs should be considered in addition to the base price of each vendor's asking price to determine the true cost of each bid. It is imperative that IT staff familiar with the current corporate data system be included in the review and comparison of competing bids to determine the agency costs of obtaining and ingesting any traffic data purchased.

B.8) What is the earliest date for when AADT values can be delivered?

The answer to this question determines whether the vendor can deliver the purchased data within the timeframe required to meet user needs. Vendors that cannot deliver data before the required deadline should not be considered.

C. TESTING THE QUALITY OF THE DATA

Because different vendors may have both different data sources and different mathematical approaches, the accuracy and precision of different vendors' AADT values will be different. Before your agency pays for data, it is vital that the vendor shows that its data and algorithm generate accurate (i.e., unbiased) and precise AADT values for your road system. Therefore, agencies need to be confident that a vendor's data are as accurate as it claims. This checklist will help an agency make that determination.

C.1) Is your agency willing to accept independent 3rd party testing or audits of a vendor's cross validation work?

See the answer to question A.3.

If tests previously performed by an independent agency (not the vendor) for some other state or agency, under a well-defined protocol, indicate the accuracy that the vendor will deliver for this purchase, third-party testing can be a good way to save money on the data purchase. The advantage of making this assumption is that it will save time and money for the agency. One mechanism for doing this is to have an independent audit of the vendor's cross-validation results obtained during their model development. This involves having an independent third-party review the cross-validation steps performed and the outcome of those steps to ensure that the model has been calibrated appropriately and is accurately reporting the expected error bounds using best practices for machine learning techniques.

The disadvantage of this approach is that the accuracy achieved elsewhere may or may not be like the accuracy that will be achieved for your agency, as accuracy achieved by any given vendor will depend not just on the technical approach it uses, but also on the size of the base data sample that is available from your jurisdiction. Sample sizes can differ from one state to another, as well as within different regions of a state.

Using existing testing results will likely be an acceptable choice if the previous tests were performed recently and in a similar environment (e.g., two states with a similar urban/rural profile, or two MPOs in the same state), as the underlying relationships between the available data and traffic volumes will likely be consistent across those similar locations.

A previous test may not be a good estimate of accuracy for your agency when the underlying relationships between probe data points and traffic volume have likely changed between that test and your roadway system. For example, if the existing tests were performed in a heavily populated area, and your agency needs data for rural, mountainous environments, then new validation tests will likely be required. This is because there may have been changes in the relationship between cellphone reports and traffic volumes because of the changes in cell service providers, cellphone usage rates, the cellphone applications being used, and the level of cell tower coverage.

C.2) What is needed from the vendor to calibrate their AADT estimates for your state's conditions?

See the response to question B.5.

Some vendors will wish to locally calibrate their algorithms to your agency's specific conditions. To do that, they will need local ground truth data which with to train their models. As part of each vendor’s proposal, it is important to learn what each vendor requires to locally calibrate its algorithm.

Calibration typically requires a specific number of calibrated, permanent, CCSs for each geographic area and/or roadway classification requested by the state or included in the vendor specifications. The vendor may need multiple years of data for each of these locations. Does the vendor need just the AADT value for each site? Does it need directional volumes or total volumes in both directions? Does it need hourly records or just daily volumes?

An important question for the vendor is whether ADT measurements—as opposed to AADT measurements from continuous counters—would be used in the calibration process. ADT is readily available from many short-term counts taken around the state. These measurements are more accurate than the AADT values computed from a short-term count, because of the error inherent in the temporal adjustment process. Unfortunately, this practice of calibrating with ADT has typically not been the case as the vehicle probe to traffic volume estimation process typically works more accurately at the annual level. Therefore, vendors may prefer a smaller number of accurate AADT values over a larger number of ADT values.

C.3) Does your agency have sufficient continuous count locations on the roads where they are needed?

Given the requirements of the vendor(s) selected for testing, your agency needs to determine whether it has sufficient count locations to give to the vendor for calibration. In making this determination, also remember that your agency likely needs to reserve enough of these same counters for validation testing. If the agency does not have enough count locations, it will be necessary to identify other count locations that can be used for these tests. Other sources of these counters may include other agencies (e.g., cities), other transportation agencies (e.g., toll authorities), or even other divisions within your agency (e.g., truck weigh station operations [where counts at those stations extend across the entire roadway] or intersection counts).

The ideal testing scenario is to withhold some continuous count locations from the vendors’ internal model calibration process so that an independent test can be performed after the model has run. This limits the chance that the model can be manipulated to look more accurate and precise than it is by just assuring that its results match the continuous count locations built into the model. Importantly, though, this risk is likely not able to be completely controlled. As discussed earlier, agencies generally make their continuous count data publicly available. Even if some count data were shielded from being directly provided to a vendor, data from previous years for the same site might have already been incorporated into the model, giving the vendor some information. Additionally, much of the count data are obtained and processed by commercial vendors and the provider may have a business relationship with these providers to obtain access. Since this risk likely cannot be entirely removed, an agency should use other criteria to consider this risk. Three considerations for this include:

1. Only consider vendors whose products include reasonable transparency and who have demonstrated appropriate data validation results in the past, such as for another agency or for sites that you know could not have been included in the model (e.g., a new continuous count station).
2. Beware of validation data that is too good to be true such as AADT estimates that are too close to the calibration site data provided.

3. Beware of vendors who claim they can provide any and all desired data and claim they have the quality level required. Currently available products are based on the integration of location-based services and/or GPS data, with demographic and geographic factors. They are only accurate and precise in estimating traffic volume if they have a high enough penetration rate, or number of measurements as a proportion of the overall population of vehicles traveling through a site. Without a comprehensive source of LBS and GPS data for the whole country, inputs are likely to include many different sources. Consequently, there should be some low fidelity estimates, especially in low-volume areas.

If CCS data can be withheld from the model calibration, this is preferred, but the agency should balance the extent of such validation counts against the lost benefit the model would get from having these inputs. The most important aspect of the withheld validation data is not the sheer amount, but that it includes a variety of road types and volumes, functional classifications, vehicle class distributions, and geography.

C.4) Collect and store the independent ground truth data needed to compare with the vendor data.

After providing the calibration data to the vendor, your agency should also tell the vendor the locations for which volume estimates will be provided. Your agency is then responsible for developing accurate ground truth traffic volume estimates (the recommended process for AADT calculation is found on page 3-68 of the 2022 TMG, previous research has shown this to be more accurate and significantly less biased than the AASHTO AADT method) from those locations to compare against the vendor's data. This includes making sure the equipment collecting the ground truth data is well calibrated and functioning correctly.

C.5) Obtain and store data from the vendor.

Once the vendor has completed its calibration, it will provide your agency with its estimates at the agreed upon locations.

C.6) Combine datasets, matching AADT locations from both datasets to ensure the correct one-to-one comparison between vendor data and ground truth AADT value.

Your agency (or selected consultants) will then perform a series of one-to-one comparisons. This process is detailed in Appendix B. There will be three types of ground truth scenarios that will be encountered in general. Each will have recommended approaches as detailed in Appendix B:

1. Ground truth is a CCS. Each set of such data should be characterized with respect to whether it was known to have been provided to and included in the vendor's product or whether it was a site set aside specifically for validation. Executing the testing on both sets of data separately will be beneficial, as well as a combined analysis assuming no consistency issues are found.
2. Ground truth is a short-term, portable count factored to AADT by the agency's corresponding seasonal and day of week factors. If the portable count is from a previous year, annual factoring will also be applied.
3. There is no site-specific continuous or portable count available. This situation may apply

if a vendor specifies that it can provide estimates of quality the agency finds acceptable even though the agency has no direct ground truth data.

C.7) Follow the accuracy testing procedures that are shown in Appendix B or perform an audit of the cross-validation process and outputs used for model development and testing for your state.

Appendix B shows the recommended process for independent computation of the accuracy of a vendor's data. It requires many permanent count locations that were not used by the vendor for model calibration. If an insufficient number of independent sites are available for testing, the recommended approach is to audit the vendor's cross validation results from that vendor's model development for your state.

D. UNDERSTANDING THE DATA TO BE LICENSED

Transportation agencies should have a good understanding of the data they are planning to purchase. The typical systems engineering approach can provide a good understanding of the data as summarized in Figure 2.

Step	Intended Outcome
1) Initiate	Define the problem, have a clear motivation, and identify intended results that will bring value to the agency.
2) Concept and Requirements	Verify the proposed solutions are theoretically and analytically valid, and meet the requirements set.
3) Discovery and Education	Evaluate alternatives and gauge their feasibility and practicality. Also, conduct planning for a low-risk execution.
4) Proof of Concept and Prototype	Verify the performance of the concept through a relatively quick, low-cost, and non-resource intensive demonstration.
5) Pilot	Conduct a small-scale demonstration to test the quality, performance, reliability, and value of proposed data product to gauge its overall utility.
6) Implement	Integrate the new data product with the agency's existing data system for maximizing the benefits of the new data.
7) Feedback and Evaluation	Create a supportive learning environment and continuous improvement.
8) Sustain	Strengthen the agency's data infrastructure by monitoring the process and refreshing the data products. Also, find new use cases for the data product and identify improvement opportunities.

Figure 2. Procedure to understanding and utilizing the data to be licensed

To have confidence in the quality of the AADT and other traffic volume estimates that your agency is licensing, it is important that your agency understand the basis for those estimates and have confidence that those estimates will be valid across the relevant set of roads. Obtaining this confidence starts with understanding three aspects about the way those estimates are produced:

- What are the input variables used to produce those AADT estimates, and any other volume statistics being purchased?

- What are the basics of the analytical process used to convert those input variables into traffic volume estimates?
- What is the demonstrated accuracy of those volume estimates, and specifically the accuracy of the model they are using to provide the estimates being purchased?

D.1) What are the input variables used to produce the AADT estimates and other volume statistics being licensed?

Understanding the basic inputs will give your agency an overview of the factors that can directly affect the estimation of the provided volumes—and conversely, the factors that can NOT directly affect those estimates (although they may indirectly affect them). Start with, "What is the basic data source used to produce the estimate? Is it mobile phone location data, GPS-based navigation apps, LBS data, and/or Fleet-management system data?" This information will allow your agency to ask questions about how the analytical process accounts for potential biases that these data sources might introduce and to test for those outcomes.

For example, if the primary data source is counts of smartphones, then there is a potential for underestimating vehicles driven by older drivers, as this population is less likely to carry smartphones or use many apps that provide LBS data. This suggests that when examining the accuracy of the data, attention should be paid to determining whether geographic areas with high levels of older drivers or tourists have biased AADT estimates relative to other areas in the state. Similarly, data from connected cars come from newer-model cars which are more likely to be driven by high-income individuals. AADTs on roads with heavy transit use, where many users are using their smartphones in each bus, may be overestimated relative to roads with no transit service.

A key aspect of learning about the inputs a vendor uses to estimate AADT is the sample size associated with that input source. For AADT, the unit of sample size is the number trips. At a high level, the relevant sample size should be the number of distinct, observable vehicle movements during the time being studied at the specific count location in question. Different vendors may capture this in different ways, such as by making "trips" (e.g., "on average we capture 8 percent of trips at any given count location") or by linking two "activities" and inferring a trip uses the relevant segment. Obviously, the larger and more consistent the fraction of traffic being observed, the better the opportunity that the resulting AADT values will be accurate.

Another useful, but not definitive, input is the approximate fraction of unique smartphones or connected vehicles in operation that are being captured by the data supplier's data source (e.g., "Our suppliers have around ~25 percent of the smartphone market in your state.") The higher the penetration of phones or connected vehicles, the more opportunity to correct for demographic bias. However, some vendors may not have this number due to privacy protections. In all cases, the share of smartphones/vehicles will be much larger than the share of trips (the actual unit of sample for AADT) because the smartphones in a vendor system will not collect every (or even most) movement taken by each phone throughout the year.

An additional consideration in the sample size discussion is how frequently a given device reports its location, which is commonly referred to as the "ping rate" for that device. The more frequently a device reports its location, especially in urban areas, the better the ability to identify the road on which that device is located, and the better the accuracy of determining whether the device is in a car, truck, bus, or other motor vehicle, or whether the device is being carried by someone walking or biking. Frequent location reports also allow for better estimation of other traffic statistics, such as turning movement percentages and speeds. Ping rate is less important in rural areas. A review of the impact on increased ping rate on accuracy of volume indicates that ping rates that range from every 5-15 minutes enables equally as accurate volume estimations as ping rates of every 1-3 minutes. In other words, for volume estimation, ping rate at the resolution of minutes is sufficient for map matching. See *Non-Traditional Methods to Obtain Annual Average Daily Traffic (AADT)*[6] for the detailed findings associated with this topic.

However, a final area for consideration is the need to understand the spatial accuracy of those pings. The more accurately the device is located, the more accurately the device is assigned to a specific road segment, and the better the resulting AADT estimates (and all other derived metrics) will be. Thus, agencies should request not only the ping rate of the devices reporting their positions, but the accuracy of those devices, with more accurate devices (± 20 meters) being roughly twice as accurate as devices with spatial errors over 500 meters, and more than 25 percent more accurate than devices with spatial errors of between 50 and 500 meters.

Having access to sample size information associated with specific AADT statistics provides considerable insight into the accuracy of those estimates. However, some vendors consider sample size to be key, proprietary business information and often do not share it. While this is understandable, having access to these data is extremely helpful to an agency trying to understand the reliability of their purchased data and guide them in day-to-day implementation. The ideal scenario is for the vendor to share trip sample size on each analysis run (e.g., the trip sample size used for the AADT 2019 estimate on this road segment was X).

In addition, vendors using modern machine learning or other techniques for analysis can calculate expected error ranges for any given road segment estimate. For example, they may be able to estimate an AADT for a certain segment of 8,200 with a 95th percentile accuracy range of 8,000 to 8,400. Another road segment with a best estimate of 8,200 might have a much broader 95th percentile range, say one of 7,200 to 9,200. These individualized accuracy statistics can help staff understand and implement specific AADT road segment estimates.

In addition to the details associated with the primary data that serve as the basis for traffic count estimates, vendors may well use other supporting data sources to adjust their initial estimates.

Other variables that might explain variations in travel behavior (e.g., the presence of major snowstorms or other events) may also be used by a vendor's algorithms to adjust its AADT estimates to account for activities that affect the basic relationships between the number of observed devices and the number of vehicles using the roadway.

D.2) What is the basic mathematical (theoretical) approach used?

It is important that your agency have a basic understanding of the algorithmic process being used to convert the input variables into the AADT statistics. Such an understanding will improve your agency's ability to quality assurance check the purchased data, as well as describe to decision makers and the public how the traffic volume estimates are being generated to maintain their confidence in the values being used.

[6] Schewel, L., Co, S., Willoughby, C., Yan, L., Clarke, N., & Wergin, J., (2021). *Non-Traditional Methods to Obtain Annual Average Daily Traffic (AADT)* (No. FHWA-PL-21-030). United States. Department of Transportation. Federal Highway Administration. Office of Highway Policy Information.

Unfortunately, many of the current techniques being used involve complex machine learning approaches that identify, classify, and apply relationships between variables. Common approaches include various types of regression, neural networks, gradient boosting, and random forest techniques. These terms and techniques may not be well understood by the transportation engineering staff tasked with checking and using the resulting traffic estimates so an external panel may be needed to validate the subject data.

Consequently, each vendor should be required to include a description of its overall analytical process that can be understood by agency staff. This does not mean that the vendor must supply specific equations to the agency or specific software code that performs the AADT estimation, as these are rightly the vendor's proprietary information. However, the vendor should be able to provide a clear description of the analytical process and the nature of the relationships being used to compute their volume estimates. Included in this description should be the vendor's plan for identifying when and how the algorithmic relationships are updated over time, and information about how that update process affects the year-to-year trends the vendor provides.

For example, if the vendor uses LBS data to compute AADT values, then how does it adapt that algorithm over time when the smartphone apps that generate the base data change over time? In the realm of LBS, the use of some apps that supply data grow over time, while others decline. This changes the relationship between the number of location points obtained and traffic volumes on the road. Similarly, the business environment for apps causes data from new apps to be added to the data feed while others are removed. Thus, most vendors of traffic volume data frequently update their algorithms to account for these changing relationships. Understanding this update process and its implications for the traffic volume data being purchased is important.

While it is not possible when licensing data to understand how a specific vendor's algorithms will change over the course of the agreement, it is important for the agency to understand when the algorithms are changing, how they are changing, and the effect those algorithmic changes have on expected data accuracy.

D.3) Can this explanation be released to the public?

It is important that this "simplified description" of how traffic volumes being used by your agency, including the imbedded privacy protection measures, can be made publicly available. It is also critical, because it supports a transparent government, and because this is a significant change in technology, that the public needs to be able to understand if they are to have confidence in the traffic volume data being used to design and operate their road system.

D.4) Are the input data sources consistent from year to year, or do they change over time? And if the input sources change over time, what activities are undertaken to maintain consistent trends over time?

As noted in D.2, many of the data sources used by vendors change over time. This can be because of changes in the apps used to deliver location data points, or the usage of those apps, or even the companies that operate the cellphone network. For example, what happens when two cellphone companies merge, and data on phone locations from one of those companies is the basis for the AADT values? Does the coverage grow due to access to a larger set of cellphones, or does the traffic volume vendor have to sign a new data access agreement with a different cellphone data vendor?

Because these changes occur over time, the actual algorithms (or algorithmic coefficients) used to compute traffic volume also change over time. These changes are risks and have the potential to cause discontinuities in traffic volume estimates on some roads. (Note that these same discontinuities can occur with traditional short-term counting techniques. In this latter case, they typically occur either because of undetected equipment error or when the short-term count observes traffic associated with an unusual event, for example, a traffic diversion due to a construction event, that was not accounted for when the count was scheduled.)

Your agency should be aware of the potential for these discontinuities (risks) and should work with the selected vendor to understand 1) the degree to which the vendor will proactively identify and mitigate the effects of these inconsistencies, 2) the frequency with which your agency should perform those inconsistency checks, and 3) the appropriate response to identified inconsistencies.

D.5) Are there known limitations in the AADT estimations? And if there are known limitations in the system, how does the vendor address those limitations?

Your agency should work with the successful vendor and other organizations that use the vendor's data to gain an understanding of any limitations in the purchased data. Historically, the accuracy of AADT estimates (in terms of expected percentage error) declines as traffic volume increases. This is likely true for some vendors' data simply because modest changes in absolute volume on low-volume roads result in large changes in percentage error. At the same time, even when a modest fraction of vehicles using a low-volume road are vehicle probes, there may still be a very low volume of probes observed on specific low-volume roads. This will produce lower accuracy in the conversion of probe volumes to AADT or other traffic statistics because minor biases in which vehicles are or are not using those low-volume roads will result in errors in the AADT estimation process.

In addition to this known limitation, machine learning techniques are subject to bias, in large part because these models are only as good as the data with which they are trained (calibrated). Many times, biases occur because some types of roads are under-represented in the training dataset. Biases can be geographic or associated with specific types of roads. For example, vehicle probe datasets often struggle to differentiate travel on contiguous roadways, such as general purpose (GP) and HOV lanes, or on roadways that are vertically stacked (e.g., an express lane roadway that is in a tunnel underneath a GP facility). In these cases, the probe data cannot provide reliable independent HOV/HOT and GP volumes, although they may well provide good AADT values for the combined roadway. The same limitation is typically true if lane-specific volumes are needed for operational analysis on a multi-lane facility.

Where limitations exist, your agency may need to perform traditional traffic counting to supplement the vehicle probe data.

D.6) What data are used to calibrate the system or enable the "machine learning"? How much data (number of locations) are required to perform that calibration? (i.e., X% of calibration locations are on roads of < 5,000 AADT)?

Answers to previous questions (see B.5 and C.2) will affect your agency's understanding of the calibration requirements associated with the successful vendor. It is important to understand that your agency will need to maintain the appropriate calibration and validation dataset into the future. As noted in D.2 and elsewhere, most traffic volume data vendors continually update their estimation algorithms. This is a common outcome of the use of machine learning techniques. In this type of analytical model, the output forecast is not a static function of the inputs but is dynamically redefined for each new input data stream over time. In addition, as described in D.5, analyses performed over time may determine specific limitations in the current process due to limitations in the calibration dataset.

To resolve these issues, your agency must work with the vendor to supply and improve the ground truth datasets used to calibrate the vendor's process. Calibration data may also come from nearby, similar jurisdictions. The more improved those calibration datasets, the better the resulting AADT values. Thus, the agency should expect to work in concert with its selected vendor to both maintain and improve the calibration (training) datasets over time and to partner with similar jurisdictions to share calibration and validation data. The agency should also work with their traffic count database vendor to ensure that, when using the validation method shown in Appendix B, that subsets of the permanent count database are reserved for the validation task and not shared with vehicle probe-based AADT vendors.

OTHER IMPORTANT CONSIDERATIONS

This section of the guidelines raises other issues that your agency should consider as it approaches the purchase of volume data based on vehicle probe technologies.

Consider a phased approach to the integration of private sector data.

The shift toward vehicle probe-based traffic volume estimates and away from traditional short-term counts is still a very new concept. The result is that some unexpected outcomes may occur. These can be both positive and negative. For this reason, agencies may find value in a phased approach to the purchase and adoption of these new data sources with consistent verification of the data source over time. For example, an agency might wish to purchase a one-year license to a subset of the data while also funding a research project to compare those data to the current traffic volume estimates, while also allowing IT staff to explore the true IT costs associated with adopting that dataset within the corporate data system.

The research project should follow steps outlined in Checklist steps C.4 through C.7 and described in Appendix B.

APPENDIX A: CHECKLISTS

A. PREPARING TO PURCHASE AADT ESTIMATES

A.1) How do you intend to use the data?

______Meet specific project needs? Accuracy requirements for those project analyses
______Replace existing short-term counts (Use accuracy Table 2 to test data)
If this is the intention, what additional traffic variables need to be included in the purchase, to replace data no longer being collected by short-term counts?
______Provide AADT values where reliable counts are not affordable (Use Table 2 for an accuracy allowance)
______Others (explain)

Table 2. Accuracy and Precision Limits Based on Historical Short-Term Counts Factored to AADT When Compared to Continuous Count Station Reference Data

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

A.2) Who will use these data?

______A small set of staff working on a specific project – *Number of seat licenses required?*
______The entire agency - *Must be able to input AADT data to the existing agency data system*
______Outside agencies we work with (e.g., MPOs, cities, counties) - *Access to the data needs to be provided to those staff, either through the agency's corporate data system (see above) or through additional licenses*
______The general public – *Agency must be allowed to release the data. There must be a mechanism for releasing the data. Is that the existing agency's corporate data system?*

A.3) How do the data need to be delivered from the vendor?

______Via the vendor's web site (visual delivery of data at a location, downloading of selected data via link as Excel or CSV files) – *appropriate for MB of data*.
______Large CSV file transmitted via secure file transfer protocol – *appropriate for GB of data*
______API used to extract data based on software queries – *appropriate for real-time delivery of data, or automated downloading of frequent dataset updates.*
______Others

A.4) How often will the data be delivered?

______Annually – (*AADTs are provided as part of the "end of year" data processing*)
______Monthly – (*AADTs are routinely updated by the vendor to reflect changing travel conditions*)
______On demand – (*based on a selection/query to the vendor's web site*)
______In real time – (*the vendor provides volume data for operational use, in addition to AADTs, so data is delivered routinely through an API.*)

A.5) How will the data be uploaded to the agency's corporate data system?

(Required only if the answer to questions 1 and 2 indicate that the data need to be uploaded into the current corporate data system. This question requires discussion with the agency IT staff and an understanding of how the AADT values will be stored – either as location-specific values or as segment values within the agency system and requires an understanding of how data can be uploaded to, stored in, and accessed within the corporate data system currently used by agency staff.)

____

____

A.6) How will the accuracy of the data be determined?

______Based on independent 3rd party testing – (*this assumes tests in other geographic areas will produce similar results in your state*)
______Performing your own validation testing – (*This is best done by using data from well-calibrated, permanent, continuously counted locations as ground truth for comparison*)

B. DATA REQUIRED TO COMPARE BIDS

B.1) What is the purchase price?

What is the pricing model? (for example: by number of concurrent users allowed, by mile of roadway, by population, by internal vs. external access rights)

____

B.2) Who owns the data you have purchased or licensed?

______Does your agency retain a copy of the data indefinitely, as well as the right to use it, after the conclusion of the contract period?
______Can your agency share those data (or a subset of the data), with other public agencies?
______Can the data be shared with the public?
______Are there conditions associated with sharing the data (*e.g., no live data can be shared, but data archived for over a week can be shared*)?
______Are there limitations on what programs and projects the data can be used for?

B.3) What is included in that price?

______Extent of licenses included
______Number and frequency of data deliveries
______Number of years of data provided (historical and future years)
______Amount of roadway covered in the data (*e.g., state routes only, city & county roads, local roads?*)
______Access to data allowed to other agencies and the public
______If changes in vendor modeling approach or input data availability change during the contract, and better estimates result from those changes, are new estimates of previously supplied data included in the bid price?
______Does the vendor supply a sample size value with each estimate? What about a confidence interval?
______What other traffic statistics are included in the price (if any)?

______AADT by truck class? (which classes?)
______MADT
______Average Hourly Volumes?
______Average Weekday Volumes?
______ADT volumes at specific sites?
______Turning movements?
______K?
______D?
______T?
______Design Hour Percent Trucks?
______Other ______________________

B.4) How accurate are the provided data?

______What accuracy does the vendor claim?
______Is it possible/desirable to use the results from accuracy tests performed by independent 3rd parties?
______Perform state specific tests (*see Testing Guidelines – Appendix B and Table 2 above*)

B.5) What data are required to calibrate the data for delivery to your agency?

(vendor response required, response can be "no state-specific calibration is needed")

______Number of calibration sites required
    ______permanently operating, continuous count locations (AADT values)
    ______short-term count locations (ADT values)
    ______other
______Distribution of those sites across roadway volume classes or geographic areas

____

____

B.6) How will the data be delivered?

(How the data are delivered may add costs to the agency, based on the work required by agency staff to make those data accessible to others in the agency.)

____

____

B.7) What is required to enter the data into your agency's corporate data system?

(This question is not required if data are only obtained by staff via direct download of data or data files from a vendor's web site).

____

____

B.8) What is the earliest date for when AADT values can be delivered?

____

Are these values for the past 12 months or for a calendar year? *(Also see the answer to the question in (B.2), how often will the data be delivered? If data are delivered monthly, are new AADT values provided, and are they for a rolling 12-month period, or updates of the previous calendar year?)*

C. TESTING THE QUALITY OF THE DATA

C.1) Is your agency willing to accept independent 3rd party testing or audits of a vendor's cross validation work?

______This allows initial vendor selection (subject to validation review)

C.2) What is needed from the vendor to calibrate their AADT estimates for your state's conditions?

(Repeated from B.5)
______Number of calibration sites required
    ______permanently operating, continuous count locations (AADT values)
    ______short-term count locations (ADT values)
    ______other
______Distribution of those sites across roadway volume classes or geographic areas
______Can data be purchased and tested for accuracy without local calibration?
______Can ADT measurements be used instead of AADT values for ground truth testing?

C.3) Does your agency have sufficient CCS locations on the roads where they are needed?

(Note: If the vendor requires data for calibration, sites given to the vendor for calibration purposes should NOT be used as part of the dataset used for accuracy validation testing.)
______Geographical coverage?
______By volume or factor group?
______Can they obtain those data from other sources in your agency or other agencies? (*e.g., high-definition traffic signal system data, WIM enforcement count locations, toll road payment statistics, etc.*)

C.4) Collect and store the independent ground truth data needed to compare with the vendor data.

(This likely requires providing the vendor with calibration data from your agency.)

C.5) Obtain and store data from the vendor.

______Tell the vendor the locations for which independent AADT estimates are available, so that the vendor can provide their AADT values for those locations.
______Arrange to obtain and store the vendor data.

C.6) Combine datasets, matching AADT locations from both datasets to ensure the correct one-to-one comparison between vendor data and ground truth AADT value.

C.7) Follow accuracy testing procedures that are shown in Appendix B or audit the vendor's cross validation analyses used for model development and testing for your state.

D. UNDERSTANDING THE DATA BEING PURCHASED

D.1) What are the input variables used to produce the AADT estimates and other volume statistics being purchased?

______In addition to cellphone and other passive, mobile data, what other sources of data are required to obtain the required inputs? (*e.g., census data, weather data, employment statistics, economic indicators, etc.*)

______Do any of these data need to be contributed by the agency?

______If the primary input data come from cellphone location reporting or location-based services (LBS), what is the average sample trip rate (# of sample trips compared to actual trips at any given location) and "ping rate" (frequency with which each device reports its location) for those location reports? (*Note that there can be multiple sources of data in each vendor's input data stream, and each of those sources might have a different ping rate. The goal is to understand the approximate fraction of devices with each ping rate.)*

Table 3. Input Data and Characteristics to AADT Model

Type of Data	Overall Sample Size of Trips	Ping Rate	Spatial Accuracy of Data Points
Cellphone-derived trips			
Location-based services data-derived trips			
WiFi/Bluetooth location-derived trips			
Fleet management-derived trips			
Other ____________________ ________________________ ________________________			

Table 4. Supplemental Data for AADT Models

Type of Data	Provided By?	Data Description
Census Data		
Weather Data		
Employment Data		
Construction Activity		
GPD		
Other (1)		
Other (2)		

D.2) What is the basic mathematical (theoretical) approach used?

(The vendor should provide a written description that can be understood by a transportation professional that may not have a degree in modern data science techniques.)

______Is there an identifiable, and direct relationship between the input data variables and the output values (AADT), or is the relationship mostly a “black box”?

D.3) Can this explanation be released to the public?

(Or are there trade secret issues with the release of the proprietary approach to AADT estimation?)

______Yes/No

D.4) Are the input data sources consistent from year to year, or do they change over time? And if the input source data change over time, what activities are undertaken to maintain consistent trends over time?

______Yes/No

For example, if the data source is cellphone location records, does the data supplier of those records ever change? If the data source is primarily location-based services data, do the applications which provide those data points change over time?

____

____

D.5) Are there known limitations in the AADT estimations? And if there are known limitations in the system, how does the vendor address those limitations?

______For example, do the estimates lose reliability on

- very low-volume roads?
- roads that are open only on a seasonal basis?
- volumes are not available on adjacent roadways (for example, it is not possible to reliably differentiate volumes on adjacent general purpose and HOV/HOT lanes)
- volume estimates in areas without cellphone coverage, or where such coverage is very low quality

____

____

____

D.6) What data are used to calibrate the system? How much data (number of locations) are required to perform that calibration and are there distribution requirements (i.e., X% of calibration locations are on roads of <5,000 AADT)?

(See question B.5 and C.2)

____

____

APPENDIX B: HOW TO DETERMINE AADT ACCURACY AND PRECISION

BASIC CONCEPTS

This section provides a discussion of accuracy, precision, and uncertainty, which are key concepts used to understand the quality and reliability of the traffic data your agency is purchasing. No data should be purchased without a thorough understanding of the accuracy, precision, and uncertainty associated with those data.

Your data vendor should provide your agency with information about the accuracy and precision of the data they are providing, as well as the corresponding uncertainties associated with their AADT estimates. Directions provided later in this Appendix describe how to independently verify a vendor's accuracy and precision as well as how to determine whether it conforms to uncertainty standards.

Accuracy describes how close the estimated AADT values provided by the vendor are expected to be to the ground truth AADT for a given location (e.g., the expected error). Ground truth AADT is the value computed from a well-calibrated and fully functioning continuous permanent count station.

Precision is a measure of statistical variability.[7] It describes the distribution of the computed differences between the estimated and true AADT measurements taken across several sites.

Specific measures of accuracy and precision are discussed in the following section. Whenever a measure of accuracy or precision is made for AADT, and that measure is based on a sample of sites, as opposed to the entire population, the resulting measure is subject to uncertainty. A common way to express that uncertainty is to include a confidence interval with the estimate. A "confidence interval" is a statistical technique used for describing the upper and lower bounds within which the ground truth value for an estimate is expected to reside, given an estimate of that ground truth. The confidence interval is associated with a "confidence level", which is a percentage that represents the likelihood that an interval of the type constructed will bracket the true value. When used for vehicle-probe based AADT estimates, the confidence interval typically takes the form:

"estimated AADT measure (X) with a confidence interval (±Y) and confidence level (Z percent)"

meaning:

"the AADT measure estimated by the term "X" and expanded to an interval between the values of X-Y (the lower bound) and X+Y (the upper bound), should bracket the ground truth AADT measure with probability Z percent." Note that the AADT measure could be the actual AADT itself, or as detailed throughout this section, it will most often be an accuracy or precision measure associated with an AADT or a set of AADTs.

[7] https://en.wikipedia.org/wiki/Accuracy_and_precision, extracted December 12, 2020

HISTORICAL ACCURACY AND PRECISION EXPERIENCED WHEN USING SHORT-TERM, 48-HOUR PORTABLE COUNTS FOR ESTIMATING AADT

In 2015, a pool-fund transportation research project titled "Assessing Roadway Traffic Count Duration and Frequency Impacts on AADT Estimations TPF-5(292)" led by FHWA and supported by seven states researched and assessed AADT accuracy and precision related to using short-term counts and AADT estimations.[8, 9]

In the above study, full-year hourly data (365 days of complete 24-hour volumes) from 206 site and year combinations (i.e., complete sites) were obtained that collectively represented 9 functional classifications, 32 different states, and years from 2000 through 2012. These 206 complete sites had 48-hour counts extracted (done through 1,000s of iterative calculations) systematically from the beginning of the year to the end of the year. Each of these 48-hour counts was expanded to an estimated AADT based on the use of adjustment factors for the days of week and month of year when the count was taken. These factors were based on the functional classification of the road being counted. The factors were computed based on using the 206 sites, as well as an additional 346 incomplete sites (i.e., where a full year of data was not available.) To mirror short-term count practices more closely, only those AADT estimates that began and ended on a Monday through Thursday and which did not span any Federal holiday were included in this preliminary analysis.

ACCURACY

The AADT estimates from the 48-hour-based short-term counts were compared to the known true AADT, since each of the 206 sites had data for all hours for every day of the year. This error is termed the Traffic Count Error (TCE), and is calculated as:

$$TCE = AADT_{error} = 100^{*}(AADT_{48\text{-}hour} - AADT_{true})/AADT_{true}$$

The TCEs (dependent) were collectively fit to a quantile regression model with base 10 logarithm of $AADT_{true}$ as the predictor (explanatory). From the quantile regression model, the 50th percentile (median) of the TCEs response was estimated as a function of AADT. The model also produced 95th percentile confidence bounds for the TCE median error, which are shown in Table 5. Note that there were an insufficient number of sites in the study dataset that had AADT values under 500 for accuracy and error tolerances to be determined for that volume range, so the lowest AADT volume range in the results below is for AADT starting at 500.

[8] All AADT values were calculated using the AASHTO methodology (see page 1-6 of the 2016 Traffic Monitoring Guide at https://www.fhwa.dot.gov/policyinformation/tmguide/), though a subsequent improved FHWA method has been approved. The impact of the AADT calculation methodology is expected to be very small here.

[9] Jessberger S, Krile R, Schroeder J, Todt F, Feng J, "Improved Annual Average Daily Traffic (AADT) Estimation Processes", Transportation Research Record: Journal of the Transportation Research Board, No. 2593, TRB 16-2477, 2016.

Table 5. Reference accuracy for same year, 48-hour short-term counts for AADT estimates, AASHTO AADT Calculation and M-R Excluding Holidays

AADT Volume Range	Sample Sites (n)	Lower 95% Confidence Bound for TCE Median Error (%)	Upper 95% Confidence Bound for TCE Median Error (%)	Minimally 95% Probability, TCE Median Error (Bias) (%) (margin of error)
0-500	1	*Unknown*	*Unknown*	*Unknown*
500 – 1,999	33	-2.0	-1.4	± 2.0
2,000 – 4,999	44	-1.1	-0.7	± 1.1
5,000 – 9,999	47	-0.4	0.0	± 0.4
10,000 – 19,999	23	0.3	0.6	± 0.6
20,000 – 34,999	20	0.8	1.2	± 1.2
35,000 – 54,999	13	1.2	1.6	± 1.6
55,000 – 84,999	10	1.5	2.0	± 2.0
85,000 – 124,999	8	1.8	2.4	± 2.4
125,000+	7	2.0	2.6	± 2.6

The last column in Table 5, titled “Minimally 95% Probability, TCE Median Error (Bias) (%) (margin of error)” shows that the TCE median error is within ± 2.0 percent of its true value at least 95 percent of the time when the AADT volume range is more than 500 but less than 1,999. The results of the 2015 Pooled Fund study[10] showed that AADT estimation from 48-hour counts under the conditions outlined tended to produce median estimates (for any eligible day in a year) that were systematically below the true AADT for sites with lower AADT and systematically above the true AADT for sites with higher AADT. The statistical model was evaluated at several specific AADT values corresponding to the midpoints of similar volume ranges from HPMS as documented in the 2016 TRR paper[11], except for sites in the 125,000+ volume range, which were evaluated at 125,000. The sites in the study ranged in true AADT from 500 to 269,418. From these results, FHWA selected the largest absolute percentile error for the median as the reference accuracy value. An alternative methodology for AADT estimation can be labeled equivalently accurate to short-term counts if the median error it produces for any particular time period falls within the identified bounds shown in Table 5.

The TCE median error provides an important measure of accuracy, but other accuracy measures have also been proposed for this evaluation. These include:

Mean Absolute Percent Error (MAPE): The previous definition of TCE can be modified to always be positive.

$ATCE = \text{Absolute } AADT_{error} = 100^{*}|AADT_{48\text{-hour}} - AADT_{true}|/AADT_{true}$

[10] Krile, R., Feng, J., and Schroeder, J. (2015). Assessing Roadway Traffic Count Duration and Frequency Impacts on Annual Average Daily Traffic Estimation: Assessing Accuracy Issues with Current Known Methods in AADT Estimation from Continuous Traffic Monitoring Data. Federal Highway Administration. FHWA-PL-15-008.

[11] Jessberger, S., Krile, R., Schroeder, J., Todt, F., & Feng, J. (2016). Improved annual average daily traffic estimation processes. *Transportation Research Record*, *2593*(1), 103-109.

where the "|" sign denotes the absolute error.

When the ATCE statistics are averaged, smaller values are still indicative of better accuracy, although getting an average very near zero as was the case with the median TCE becomes unlikely. As an advantage over the median TCE, data showing large positive and negative differences from the reference will reflect a poor result which can be masked with the median TCE. The MAPE penalizes overcounts more than undercounts as an undercount of a particular number of vehicles will always produce a lower ATCE than the same overcount since the maximum for this statistic with an undercount is 100% while the maximum is unbounded.

This statistic is sensitive to outliers since it is a mean and some may prefer to use the Median Absolute Percent Error instead for this reason. Finally, just as with the median percent error, it is sensitive to the magnitude of the true AADT and small differences in counts for a low-volume location will produce large percentage errors.

Median, 68^{th} and 95^{th} Percentile Absolute Percent Errors: As noted above, the median absolute percent error may be preferred over the MAPE because it uses a median of the absolute errors and will be less sensitive to outliers. For this 2015 evaluation study, only the MAPE was included as a central distribution statistic, but the 68^{th} and 95^{th} percentile absolute percent errors were tabulated. These are just the same absolute percent errors (ATCE) but evaluated for a dataset as the 68^{th} and 95^{th} percent largest values. From an accuracy perspective, this helps identify how far away more extreme errors tend to be. As with MAPE, smaller values indicate better accuracy. The 95^{th} percentile absolute error will always be larger than the 68^{th} percentile, which in turn is larger than the median.

Normalized Root Mean Squared Error (NRMSE): This accuracy metric starts with squared differences between the measures.

$TCE^2 = AADT_{error}$ Squared $= (AADT_{48\text{-}hour} - AADT_{true})^2$

These are then converted to the NRMSE metric as:

$NRMSE = 100 * \sqrt{(TCE^2/n)} / Average(AADT_{true})$

This results in an accuracy metric that is always positive and values closer to zero indicate better accuracy. The NRMSE gives a relatively high weight to large errors because the errors are squared before they are averaged. This means the NRMSE should be more useful when large errors are particularly undesirable. Also, the NRMSE is directly related to the variance associated with the frequency distribution of error magnitudes (i.e., an increase in the variance results in an increase in the NRMSE). Larger errors are penalized more in NRMSE. This can be appropriate for some cases, e.g., if being off by 10 units is more than twice as bad as being off by 5 units. Also, the NRMSE is sensitive to outliers.

All of these additional accuracy measures have been evaluated using the same 2015 study data. In the case of the 68^{th} and 95^{th} percentile absolute errors, they are produced in a similar manner to the median reference values in Table 5. After fitting a quantile regression model to the data as a function of the base 10 logarithm of the reference AADTs, estimated 68^{th} and 95^{th} percentiles are evaluated instead of the 50^{th} percentile in the median evaluation. As with the median model, though, they are evaluated at the same midpoints of all the volume ranges except the highest range. To estimate the uncertainty of these estimates, an upper, one-sided 95% confidence bound is provided, unlike the median analysis which featured a two-sided interval. This is due to the lack of interest in a lower bound for an upper percentile.

For the MAPE analysis, all the error data for each volume range were analyzed as a single group and the MAPE for all the sites was estimated along with its upper, one-sided 95% confidence bound. Therefore, the regression relationship between MAPE and AADT was not part of this estimation as it was with the median error, 68th, and 95th percentile absolute errors. The NRMSE estimate was evaluated in the same manner as the MAPE. Due to lack of a simple uncertainty estimator for NRMSE, the reference values are simply the point estimates. Table 6 provides these additional accuracy measures by AADT volume range using the same 2015 Study data. The original median error is included in the final column of Table 6 for comparison.

Table 6. Additional reference accuracy statistics for same year, 48-hour short-term counts for AADT estimates, AASHTO AADT Calculation and M-R Excluding Holidays

AADT Volume Range	Sample Sites (n)	68th Percentile Absolute Error (Upper 95% CB) (%)	95th Percentile Absolute Error (Upper 95% CB) (%)	NRMSE	MAPE (Upper 95% CB)	Minimally 95% Probability, TCE Median Error (Bias) (%) (margin of error)
0-500	1	*Unknown*	*Unknown*	*Unknown*	*Unknown*	*Unknown*
500 – 1,999	33	12.1	32.0	12.2	9.8	± 2.0
2,000 – 4,999	44	10.7	28.5	15.8	10.6	± 1.1
5,000 – 9,999	47	9.7	25.9	13.8	9.3	± 0.4
10,000 – 19,999	23	8.8	23.6	12.6	9.1	± 0.6
20,000 – 34,999	20	8.0	21.8	14.1	8.8	± 1.2
35,000 – 54,999	13	7.5	20.5	9.1	7.3	± 1.6
55,000 – 84,999	10	6.9	19.1	6.1	4.8	± 2.0
85,000 – 124,999	8	6.4	18.0	6.0	4.4	± 2.4
125,000+	7	6.2	17.5	9.3	6.1	± 2.6

DEVELOPING A SAMPLE-SIZE DEPENDENT AND IMPROVED SET OF ACCURACY REFERENCE LIMITS

The guidance in this document substantially updates the previous work to provide a practical method for applying accuracy standards like those from Table 6. Several important changes are made to develop these newer standards:

- The AADT calculations are updated to reflect the currently recommended TMG approach.
- The short-term, 48-hour portable counts are retained with factoring to AADT, but limits are expanded to include all days of the year, not just Monday-Thursday, excluding federal holidays.
- The previously developed limits are meaningful to understand overall accuracy of portable counts, but a practical procedure to apply the limits will generally consist of a sample of sites and the limits were found to be related to the size of the sample used.
- The new methodology reduces the accuracy limits to the two most prominent accuracy measures, MAPE and Median TCE.

The methodology for determining the sample-size dependent accuracy measures follows:

A sample of N AADT estimates ($AADT_{48\text{-hour}}$) was randomly selected, with replacement, from the entire set of AADT estimates for the 206 sites across all days of the year. The sampling utilized a stratified approach (i.e., through SAS PROC SURVEYSELECT) such that the distribution of the AADTs sampled resembled that of the population (i.e., by the known true AADTs for all sites). The TCE was computed for these N estimates and used as the dependent variable in a quantile regression model with base 10 logarithm of the true AADT ($AADT_{true}$) as the predictor (explanatory) variable. From this model, the 50th percentile (median) of the TCE response was estimated as a function of true AADT. These TCE median error estimates were stored for a set of pre-defined true AADT values, and the process was repeated 10,000 times with different random samples of N AADT estimates. Following these iterations, the 2.5th and 97.5th percentiles of the stored TCE median errors were calculated as the lower and upper 95% confidence bounds, respectively.

This process was then repeated for 20 different sample sizes ranging from 25 to 1,000, evenly spaced on the base 10 logarithm scale. Once lower and upper confidence bounds for the TCE median error were calculated for each pre-defined true AADT value and each sample size, a quadratic regression equation was fitted to the confidence bounds as a function of sample size. To satisfy the model assumptions of normality and constant variance, the following regression model was fitted:

$$y = \beta_0 + \beta_1(\log_{10} N) + \beta_2 (\log_{10} N)^2 + \epsilon$$

where y is either the lower 95% confidence bound or the upper 95% confidence bound, β_i are the regression coefficients, N is the sample size ranging from 25 to 1,000, and ϵ is the random error assumed to follow a normal distribution. A separate regression model was fitted to the lower confidence bounds (Table 7) and to the upper confidence bounds (Table 8) at each of the volume ranges. In addition to the regression equations in Table 7 and Table 8 which can be used to identify limits for a particular sample size, N, within the range of 25 to 1000, the two tables provide values from the regression equations for the five specific sample sizes of 25, 50, 100, 200, and 1000.

For evaluations with more than 1000 sites, the true uncertainty associated with factored portable counts might be expected to continue to shrink. However, there are other uncertainties in the application of these limits, so an evaluation with a larger number of sites than 1000 could reasonably use the limits for N=1000. The risk here is that an alternative big data method might then end up being evaluated against easier limits than are strictly applicable, but the limited width of the N=1000 intervals would still ensure the alternative methodology estimates of AADT were very close to true AADTs from the reference CCSs.

Table 7. Sample Size Dependent TCE Median Error Lower Bounds for same year, 48-hour short-term counts for all days of the year factored by functional class factors to FHWA-based AADT estimates

AADT Volume Range	Lower 95% Confidence Bound for TCE Median Error (%), by Sample Size					
	Regression Equation	N=25	N=50	N=100	N=200	N≥1000
0-500	$-27.31+15.39(\log N)-2.33(\log N)^2$	-10.3	-7.9	-5.8	-4.2	-2.1
500-1,999	$-20.98+11.94(\log N)-1.82(\log N)^2$	-7.8	-5.9	-4.4	-3.1	-1.5
2,000-4,999	$-14.93+8.71(\log N)-1.33(\log N)^2$	-5.4	-4.0	-2.8	-1.9	-0.8
5,000-9,999	$-12.66+7.80(\log N)-1.23(\log N)^2$	-4.2	-3.0	-2.0	-1.2	-0.3
10,000-19,999	$-13.79+9.16(\log N)-1.54(\log N)^2$	-4.0	-2.7	-1.6	-0.9	-0.2
20,000-34,999	$-16.63+11.34(\log N)-1.95(\log N)^2$	-4.6	-3.0	-1.7	-0.9	-0.2
35,000-54,999	$-19.48+13.27(\log N)-2.27(\log N)^2$	-5.4	-3.5	-2.0	-1.0	-0.1
55,000-84,999	$-23.16+15.83(\log N)-2.72(\log N)^2$	-6.3	-4.1	-2.4	-1.1	-0.1
85,000-124,999	$-26.63+18.16(\log N)-3.12(\log N)^2$	-7.3	-4.8	-2.8	-1.4	-0.2
125,000 +	$-28.13+19.19(\log N)-3.31(\log N)^2$	-7.8	-5.1	-3.0	-1.5	-0.3

Log = Base 10 logarithm

Table 8. Sample Size Dependent TCE Median Error Upper Bounds for same year, 48-hour short-term counts for all days of the year factored by functional class factors to FHWA-based AADT estimates

AADT Volume Range	Upper 95% Confidence Bound for TCE Median Error (%), by Sample Size					
	Regression Equation	N=25	N=50	N=100	N=200	N≥1000
0-500	$33.59-20.82(\log N)+3.40(\log N)^2$	11.1	8.0	5.6	3.7	1.7
500-1,999	$26.04-15.97(\log N)+2.60(\log N)^2$	8.8	6.4	4.5	3.1	1.5
2,000-4,999	$18.89-11.40(\log N)+1.85(\log N)^2$	6.6	4.9	3.5	2.5	1.3
5,000-9,999	$15.26-9.05(\log N)+1.46(\log N)^2$	5.5	4.1	3.0	2.2	1.2
10,000-19,999	$14.96-8.97(\log N)+1.49(\log N)^2$	5.3	4.0	3.0	2.2	1.5
20,000-34,999	$16.85-10.17(\log N)+1.71(\log N)^2$	6.0	4.5	3.4	2.5	1.7
35,000-54,999	$19.38-11.71(\log N)+1.97(\log N)^2$	6.9	5.2	3.8	2.9	2.0
55,000-84,999	$22.08-13.22(\log N)+2.20(\log N)^2$	7.9	6.0	4.4	3.3	2.2
85,000-124,999	$25.00-14.97(\log N)+2.49(\log N)^2$	8.9	6.8	5.0	3.7	2.5
125,000 +	$26.15-15.59(\log N)+2.59(\log N)^2$	9.4	7.1	5.3	4.0	2.7

Log = Base 10 logarithm

Using the regression relationships in Table 7 and Table 8, confidence bounds for the TCE median error can be computed for any sample size between 25 and 1,000 for each AADT volume range. Since these bounds are not symmetric about zero, confidence bounds with at least 95% probability can be obtained as plus or minus the larger of the lower and upper bounds, in absolute value.

As an example, the lower confidence bound for median TCE of sites in the 500-1,999 AADT range based on a sample size of 50 is -5.9 percent (Table 7), while the upper bound is 6.4 percent (Table 8). A conservative, symmetric error bound for an evaluation with N=50 would be ± 6.4 percent. This value would be analogous to the ± 2.0 percent value shown in Table 6 for sites with AADT in the 500-1,999 range. While several factors impact these new limits, the primary driver is the sample size. The legacy limits in Table 5 and Table 6 correctly characterize error in 48-hour short-term counts factored to AADT, but as practical acceptance limits for an alternative set of measurements, they are only appropriate if the comparison sample is of a large size.

FHWA also considered that the change of confidence bounds across AADT volume ranges was smooth enough that the number of AADT volume ranges might reasonably be reduced. Accordingly, the following simplifications were made:

- AADT Volume Range 0-500 – No limits are established as source data with only 1 site is insufficient to establish uncertainty limit.
- AADT Volume Range 500-4,999 – These collectively are categorized as low volume sites and the acceptance limit is the largest of the limits for the 500-1,999 and 2,000-4,999 volume ranges.
- AADT Volume Range 5,000-54,999 – These collectively are categorized as medium volume sites and the acceptance limit is the largest of the limits for the 5,000-9,999, 10,000-19,999, 20,000-34,999, and 35,000-54,999 volume ranges.
- AADT Volume Range 55,000+ – These collectively are categorized as high-volume sites and the acceptance limit is the largest of the limits for the 55,000-84,999, 85,000-124,999, and 125,000+ volume ranges.

Following this simplification, the limits for each range were also fit to a quadratic regression relationship as a function of sample size. The limits can therefore be evaluated using a regression equation as shown in Table 9. Table 9 also shows the specific values for limits at the five sample sizes of 25, 50, 100, 200, and 1000. Table 9 limits clearly show tighter acceptance limits as the number of comparison sites increases.

Table 9. Sample Size Dependent TCE Median Error as a Function of Sample Size, N, for same year, 48-hour short-term counts for all days of the year factored by functional class factors to FHWA-based AADT estimates

AADT Volume Range	Minimally 95% Probability, TCE Median Error (Bias) (%)					
	$25 \le N \le 1000$	N=25	N=50	N=100	N=200	N≥1000
0 – 500	*Unknown*					
500 – 4,999 (low)	$25.65-15.64(\log N)+2.54(\log N)^2$	±8.7	±6.4	±4.5	±3.1	±1.6
5,000 – 54,999 (medium)	$19.38-11.71(\log N)+1.97(\log N)^2$	±6.9	±5.2	±3.8	±2.9	±2.0
55,000 + (high)	$26.15-15.59(\log N)+2.59(\log N)^2$	±9.4	±7.1	±5.3	±4.0	±2.7

Log = Base 10 logarithm

The process for computing a sample-size-dependent upper 95% confidence bound for the MAPE is analogous to that for the median TCE, with the exception that a mean regression model is fitted to the ATCE values as opposed to a median quantile regression model fitted to the TCE values. In addition, only a one-sided upper 95% confidence bound is needed for MAPE, so the upper 95th quantile of the estimated mean ATCEs was used, as opposed to the 2.5th and 97.5th percentiles for the median TCEs. Results are shown in Table 10.

Table 10. Sample Size Dependent MAPE Upper Bounds for same year, 48-hour short-term counts for all days of the year factored by functional class factors to FHWA-based AADT estimates

AADT Volume Range	One-Sided Upper 95% Confidence Bound for MAPE, by Sample Size					
	Regression Equation	N=25	N=50	N=100	N=200	N≥1000
0-500	$33.31-11.12(\log N)+1.61(\log N)^2$	20.9	19.1	17.5	16.2	14.4
500-1,999	$27.75-8.60(\log N)+1.23(\log N)^2$	18.1	16.7	15.5	14.5	13.0
2,000-4,999	$22.72-6.65(\log N)+0.97(\log N)^2$	15.3	14.2	13.3	12.6	11.5
5,000-9,999	$20.46-6.23(\log N)+0.97(\log N)^2$	13.6	12.7	11.9	11.3	10.5
10,000-19,999	$20.31-7.02(\log N)+1.16(\log N)^2$	12.8	11.7	10.9	10.3	9.7
20,000-34,999	$21.62-8.57(\log N)+1.46(\log N)^2$	12.5	11.3	10.3	9.6	9.1
35,000-54,999	$23.17-10.02(\log N)+1.72(\log N)^2$	12.5	11.1	10.0	9.2	8.6
55,000-84,999	$24.87-11.42(\log N)+1.95(\log N)^2$	12.7	11.1	9.8	8.9	8.2
85,000-124,999	$26.34-12.61(\log N)+2.14(\log N)^2$	12.9	11.1	9.7	8.7	7.8
125,000 +	$26.88-13.02(\log N)+2.20(\log N)^2$	13.0	11.1	9.6	8.6	7.6

Log = Base 10 logarithm

The same sample size dependent adjustments were made to prospective limits for MAPE as were done for the TCE median error. The resulting equations and results are provided in Table 11. The MAPE measure shows tightening accuracy limits with increases in sample sizes as did the TCE median error. The MAPE limits also show a trend of tighter accuracy limits for higher-volume sites.

Table 11. Sample Size Dependent MAPE as a Function of Sample Size, N, for same year, 48-hour short-term counts for all days of the year factored by functional class factors to FHWA-based AADT estimates

AADT Volume Range	Minimally 95% Probability, MAPE Upper 95% CB					
	25 ≤ N ≤ 1000	N=25	N=50	N=100	N=200	N≥1000
0 – 500	*Unknown*					
500 – 4,999 (low)	$27.75-8.60(\log N)+1.23(\log N)^2$	18.1	16.7	15.5	14.5	13.0
5,000 – 54,999 (medium)	$20.46-6.23(\log N)+0.97(\log N)^2$	13.6	12.7	11.9	11.3	10.5
55,000 + (high)	$26.05-12.41(\log N)+2.15(\log N)^2$	12.9	11.2	9.8	8.9	8.2

PRECISION

The measures in Tables 5 to 11 assess the accuracy of AADT estimation. The other critical parameter to assess data quality adequacy is precision. From the quantile regression model of TCE as a function of base 10 logarithm of AADT previously discussed, population tolerance intervals were also estimated. The tolerance interval is a range of values expected to bracket a fraction “p” of a population, where p in this study was 0.95, or 95 percent. To capture 95 percent of the population, estimates were obtained from the regression model for the 2.5th and 97.5th percentiles. As was the case with the regression fit of the median, there is uncertainty in the model fits of the 2.5th and 97.5th percentiles. Consequently, the lower, one-sided 95 percent confidence bound on the 2.5th percentile and the upper, one-sided 95 percent confidence bound on the 97.5th percentiles are estimated. These two bounds form an interval within which 95 percent of the TCE population is expected to be contained with at least 95 percent confidence. When comparing AADT estimation products, those that have a narrower range for the TCE distribution are preferrable to those with a wider range, assuming the accuracy (or bias) is comparable. The range within which 95 percent of the population is expected to be contained reflects the precision. Other precision measures might also be applicable, such as the standard deviation of the TCEs, the range between the minimum and maximum error values, or the interquartile range (difference between the estimated population 75th and 25th percentiles (i.e., middle 50 percent of the distribution)). In all cases, smaller is better.

Table 12 shows the reference precision bounds determined from the data in the 2015 research. These limits were produced with AASHTO-based AADT calculations, functional classification factoring, and restricted to 48-hour counts falling within Monday through Thursday excluding any federal holidays. As with the median TCE accuracy limits, the upper and lower precision limit bounds are not symmetric. FHWA elected to select the larger of the absolute values of the upper and lower precision limits, which results in a more tolerant set of limits. Table 12 values are the precision limits that compare to the accuracy limits shown in Table 5 and Table 6.

Table 12. Reference Precision for same-year, 48-hour short-term counts for AADT Estimates, AASHTO AADT Calculation and M-R Excluding Holidays

AADT Volume Range	Lower 95% Confidence Bound for the TCE 2.5th Percentile Error (%)	Upper 95% Confidence Bound for the TCE 97.5th Percentile Error (%)	Minimally 95%Probability, 95% TCE Population Error Range (%)
0-500	*Unknown*	*Unknown*	*Unknown*
500 – 1,999	-31.6	34.2	+/-34.2
2,000 – 4,999	-26.8	30.8	+/-30.8
5,000 – 9,999	-23.3	28.5	+/-28.5
10,000 – 19,999	-20.2	26.7	+/-26.7
20,000 – 34,999	-17.7	25.7	+/-25.7
35,000 – 54,999	-15.7	24.8	+/-24.8
55,000 – 84,999	-13.9	24.1	+/-24.1
85,000 – 124,999	-12.3	23.5	+/-23.5
125,000+	-11.6	23.3	+/-23.3

DEVELOPING AN IMPROVED SET OF PRECISION REFERENCE LIMITS

The precision limits from Table 12 are improved in this research. The same set of 2015 research data was used with a quantile regression fit of 48-hour, factored AADTs, but the analysis included the following changes:

- The AADT calculations are updated to reflect the currently recommended TMG approach. This applies both to the AADT calculations and to the development of seasonal and day of week factors.
- The approach of short-term, 48-hour portable counts is retained with functional classification factoring to AADT, but limits are expanded to include all days of the year, not just Monday-Thursday, excluding federal holidays.

Unlike the Table 5 and Table 6 accuracy limits, which were improved by directly including sample size considerations, the improved precision limits retain only a single set of values. The impact of sample size is determined through the evaluation methodology. This evaluation methodology is provided in the upcoming section, "HOW TO ASSESS ADEQUACY OF SAMPLE PRECISION AGAINST TABLE 14 LIMITS." The improved reference limits for precision are shown in Table 13.

Table 13. Reference Precision for same-year, 48-hour short-term counts for AADT Estimates, FHWA AADT Method and Factoring Based on All Days of Year

AADT Volume Range	Lower 95% Confidence Bound for the TCE 2.5th Percentile Error (%)	Upper 95% Confidence Bound for the TCE 97.5th Percentile Error (%)	Minimally 95%Probability, 95% TCE Population Error Range (%)
0-500	*Unknown*	*Unknown*	*Unknown*
500 – 1,999	-30.4	43.4	±43.4
2,000 – 4,999	-27.9	37.8	±37.8
5,000 – 9,999	-26.1	33.6	±33.6
10,000 – 19,999	-24.5	29.9	±29.9
20,000 – 34,999	-23.2	26.7	±26.7
35,000 – 54,999	-22.1	24.2	±24.2
55,000 – 84,999	-21.2	22.0	±22.0
85,000 – 124,999	-20.4	19.9	±20.4
125,000+	-20.0	19.1	±20.0

The previous results from Tables 9, 11, and 13 are consolidated in Table 14 to show the reference values for the expected accuracy and precision of the current AADT values used by state agencies, when those AADT values are based off factored, short-term counts. A table footnote clarifies that sites with AADT less than 500 have no defined standard as insufficient data were available in the research dataset to report reliable values.

Table 14. Accuracy and Precision Limits Based on Historical Short-Term Counts Factored to AADT When Compared to Continuous Count Station Reference Data

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

†Roadways with AADT less than 500 have no pre-defined standards.

The Table 14 standards for median error and MAPE are a function of the number of continuous count comparison sites. If an agency is evaluating the limits for a different number of sites than shown in the table, it may do so by interpolating between the limits. Alternatively, Table 14a allows the limits for a particular number of sites, N, to be calculated by equation. Note that evaluations should include at least 25 reference sites, but generally the largest number of sites attainable is desired. Additional context for sample size selection is provided below. Statistical uncertainty for traditional factored counts for less than 25 sites will likely be higher than the values shown here, so the values for N=25 could be considered conservative references to use in such a case. However, evaluating a count program with so few reference sites risks missing important sources of variability and is not generally advisable.

Table 14a. Equation Form of Accuracy and Precision Limits Based on Historical Short-Term Counts Factored to AADT When Compared to Continuous Count Station Reference Data

AADT Volume Range†	Sample Size for Evaluation	Minimally 95% Probability, TCE Median Error (Bias) (%)	Minimally 95% Probability, MAPE Upper 95% CB (%)	Minimally 95% Probability, 95% TCE Population Error Range (%)
500 – 4,999 (low)	By Equation ($25 \le N \le 1000$)	$25.65-15.64(\log_{10}N)+2.54(\log_{10}N)^2$	$27.75-8.60(\log_{10}N)+1.23(\log_{10}N)^2$	± 43.4
5,000 – 54,999 (medium)	By Equation ($25 \le N \le 1000$)	$19.38-11.71(\log_{10}N)+1.97(\log_{10}N)^2$	$20.46-6.23(\log_{10}N)+0.97(\log_{10}N)^2$	± 33.6
55,000 + (high)	By Equation ($25 \le N \le 1000$)	$26.15-15.59(\log_{10}N)+2.59(\log_{10}N)^2$	$26.05-12.41(\log_{10}N)+2.15(\log_{10}N)^2$	± 22.0

†Roadways with AADT less than 500 have no pre-defined standards.

The accuracy and precision values in Table 14 or Table 14a are provided as a tool for agencies to use in their validation of data accuracy and precision. However, the specific values provided are integrally tied to the historic research effort in 2015 and may not be fully representative of a specific agency's data quality. For instance, an agency may be using 24- or 72-hour counts for their short-term counts rather than 48-hour counts; they may treat weekend and holiday counting differently; or they may have a different mix of urban and rural count sites as were used in this evaluation. The quantile regression-based approach used for this effort could easily be replaced by single estimates within ranges of sites of a particular type. For that matter, even the basic structure of setting limits by volume ranges could be changed, as some agencies might prefer to have accuracy and precision metrics tied to site characteristics (e.g., urban/rural, functional classification). The selection of limits by volume ranges was a deliberate choice since it was observed in the legacy research that accuracy and precision measures differed by volume ranges so a single standard would not be appropriate for all sites.

Finally, it is important to note that the simulation procedure employed to obtain these limits supposes that the accuracy and precision associated with AADT estimation from a sub-sample of a CCS is the same as that of a portable counter. For a variety of reasons, this may not be the case. Generally, the limits provided here should be expected to be a fair representation of the degree of error that exists in estimating true AADT from short-term portable counts. Therefore, these limits may be used to fairly assess accuracy and precision of an alternative technology without unduly penalizing such technology for the same uncertainty that exists with the reference technology of factored portable counts. These limits may be modified over time as more research is completed. Alternately, agencies may choose to produce their own limits using data specific to their locations/geographic area and consistent with other assumptions, so long as they are documented and statistically defensible.

APPLICATION OF HISTORICAL ACCURACY AND PRECISION EXPERIENCED WHEN USING SHORT-TERM, 48-HOUR PORTABLE COUNTS FOR VALIDATING ALTERNATIVE AADT ESTIMATES

The first time an agency wishes to utilize non-traditional AADT estimates, they should perform a validation of the accuracy and precision of these estimates. To do so, a sample of CCSs are obtained from that agency. It is important for this initial validation exercise that these CCSs have complete (or nearly complete) data for the year; and that they are well-calibrated and reliable estimates. These will form the sample of ground truth AADT observations. These counts may or may not have previously been available by the vendor in the development of their model. It is preferrable that as many data sites as possible not have previously been provided, but as discussed earlier in this document, the public nature of transportation agencies' data volume counts makes it unlikely that zero data overlap can be avoided.

In addition to having as many sites as possible, the sites should be as representative as possible of the subsequent estimates that are expected to be provided by the vendor. For instance, the mix of sites by functional classification and volume ranges should be similar. If there are limitations, such as a lack of low-volume roadways, these should be documented so that the conclusions of the validation exercise will reflect this limitation. The geographic coordinates and locations for these sites are provided to the vendor, and the vendor should produce a set of estimates using their own algorithm. The agency should expect this process will require some coordination with the vendor to reconcile conflation or other geographic mapping issues so that the roadways selected by the state are the same as those understood by the vendor.

The outcome of the first part of this analysis is a "paired" dataset, where each site or location has the continuous-count-based AADT, which will serve as the ground truth data, and the corresponding AADT estimated for that same site from the vendor.

The importance of establishing the accuracy and precision of short-term portable counts from the previous section is that the AADT estimates from non-traditional sources are being evaluated to replace or augment similar counts in the future. The continuous counts are used in the validation process since they are the best source of ground truth counts. If a vendor's data passes the validation testing, it is evidence that they can produce estimates of equal or better accuracy and precision as short-term portable counts. It should not lead to a conclusion that their data is equivalent to continuous counts. The continuous count data should continue to be collected as before. It is also important that these data are often part of the vendor's model development and internal validation, so they should be retained. If a vendor's traffic counting product were to mature to be potentially of the same quality as current continuous count sites, a separate evaluation and validation process would be required.

The validation process outlined here is applicable for an initial qualification of a vendor's product but may also be used as a periodic revalidation. For instance, a yearly re-evaluation may be included as a requirement for the vendor, or an ad hoc re-evaluation if some estimates fail other agency data quality checks, or there is an important change in the vendor's algorithm (e.g., losing access to an important LBS or GPS data stream). It is also important to recognize that technical challenges may result in a validation conclusion that some model estimates are acceptable to use, while others are not. This has been the case in some states where AADT estimation for low-volume roadways by the non-traditional methods cannot demonstrate equivalent accuracy and precision to portable counts, while higher-volume roadways can meet the validation requirements. In this case, the agency and vendor may agree to a contract based only on the validated counts. The parties may wish to perform a re-validation if and when the non-traditional data product matures to the extent that it can meet these requirements for the previously unvalidated roadways.

TRAFFIC DATA ACCURACY AND PRECISION EXPECTATION

Based on the above historical data analysis and traffic data professionals' familiarity with permanent and portable counting, data professionals should consider data quality from any other source with the reference values in Table 14, or other such tables as developed by the agency. Ideally, no data quality dimension should perform below the reference values shown in Table 14. However, this may be an iterative evaluation, in which the following steps apply:

Step 1: Preliminary evaluation of data accuracy and precision from the full set of paired observations of the data provider's AADT estimates and corresponding ground truth estimates using the continuous count data. If all sites produce accuracy measures at or below the limits specified in Table 14, or other such limits developed by the agency, or these limits as suitably modified for known limitations, the vendor will have passed the validation and the agency can accept their estimates in lieu of, or in addition to, portable count results at substantively similar sites.

> **Step 1a:** Meeting accuracy limits means that the accuracy data measures (TCE median error (bias), MAPE) as provided by the vendor fall within the limits specified in Table 14. Depending on the vendor and state, either or both of the accuracy data measures may be used, or even another equivalent measure if limits have been established (e.g., NRMSE, 68th or 95th percentile absolute error, median absolute percent error). The agency, in partnership with the vendor, and with FHWA approval if applicable, should agree on the measure(s) that must pass prior to the evaluation.
>
> **Step 1b:** Meeting precision limits requires additional consideration, which is provided in the next section.

Step 2: If some limit(s) are not met, there are possible remediation steps that may be considered.

> **Step 2a:** The data may be reviewed to determine if any of the estimates provided by the vendor as candidate data and/or the agency as the ground truth could be erroneous. While all such validation data should previously have passed respective internal quality checks (especially that of the ground truth data), there could be extenuating conditions that lead some data to be determined to be invalid. However, truth-in-data principles should apply, and no data should be removed simply because they lead to undesirable outcomes.
>
> **Step 2b:** The data review in Step 2a may find certain data that are highly influential in leading to the failure. This happened historically with one vendor when a significant source of continuous count data from toll sites was found to not be reflected well in their model estimates. If the vendor is able to adjust their algorithms to reconcile these failures, a new set of estimates may be permitted, and evaluation started again such that the data quality requirements can be met. It is important that the remediation by the vendor is not to simply

force its previous results to match for sites that were previously known to fail. For instance, it would be beneficial for the agency to find other previously unused sites of this kind to add to the validation in the second round.

Step 2c: The vendor and the agency may agree to only a partial validation. Recent experience has shown that non-traditional data are less successful in replicating AADT estimates with acceptable accuracy and precision on low-volume roadways. If this is the case, and high-volume roadways do exhibit acceptable accuracy and precision, the vendor may be validated for just higher-volume roadways. Then, as detailed above, the vendor may attempt later to expand the range of sites for which they can provide data if their internal algorithms and/or data sources improve.

HOW TO ASSESS ADEQUACY OF SAMPLE PRECISION AGAINST TABLE 14 LIMITS

The requirements for precision limits are a little different from those of the accuracy limits. Each set of data or volume range produces a single estimate for the accuracy measures and then these measures are compared to the limit to determine acceptance or failure. For the precision limits, bounds are provided that should bracket 95 percent of the population, with 95 percent confidence, if the non-traditional data source has similar precision to that of the historic portable count data.

To evaluate the precision limits, an acceptance sampling approach is proposed[12]. Given a sample, *n*, of sites for which the validation is to be performed, the number of vendor's estimates will be tabulated that fail to fall within the precision limits of Table 14. These could be tabulated against only a single volume range, or alternatively could be tabulated against the entire validation dataset so long as each individual AADT estimate is compared against its respective volume range limits. The number of failures is then compared to an acceptance number, *c*, and if it exceeds this count, the test does not pass. This is what Douglas refers to as a single sample plan for attributes.

The sample plan is based off binomial sampling statistics with the following characteristics:

There is a true fraction of the vendors population, *p*, that falls outside the 95 percent population acceptance limits as determined from 48-hour portable counts. Ideally the vendor's fraction *p* would be no more than five percent. Due to the nature of taking a sample of data from the population and determining acceptance or failure of a population against a set of limits, there will be variability in how many test samples fall outside the acceptance limits. When a test with *n* items is performed where a result of *c* or less failures conclude the vendor's precision is comparable to that of the reference, there are two risks in the test that can be related by the following equations:

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

Where α is the probability that the test fails to accept when the true percentage is below p_1.

[12] Montgomery, Douglas, "Introduction to Statistical Quality Control, Second Edition," Wiley, 1985, p. 565.

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

Where β is the probability that the test passes when the true percentage is above p_2.

Because of the discrete nature of the two preceding equations and depending on which of the six variables of interest, p_1, p_2, n, c, α, and β, are specified, solutions are limited. In this evaluation, the sample size, n, is likely to be a natural consequence of the data validation and not able to be controlled. The selection of α and β are usually to obtain respective risks no higher than five percent.

Finally, the selection of p_1 and p_2 can pose challenges:

From the perspective of the vendor, a value for $p_1 = 0.05$ is preferred to limit the chance the vendor data sample fails the precision test even though 95 percent or more of its true population does fall within the precision limits. Alternately, though, this forces $p_2 > 0.05$ (and possibly much bigger) so that it becomes possible a vendor's product population will fall outside the limits much more frequently than five percent and really should have been rejected but will not be.

From the perspective of the agency, a value for $p_2 = 0.05$ is preferred to limit the chance the vendor data sample passes the precision test even though less than 95 percent or more of its true population falls within the precision limits. Alternately, though, this forces $p_1 < 0.05$ (and possibly much lower) so that it becomes possible a vendor's product population has well more than 95 percent within the limits but could still end up failing the test.

For this application, Table 15 provides a number of acceptance sampling plans that balance these two risks, so that the vendor accepts some risk that more than 95 percent of its population TCE errors really do fall within the precision limits but a particular sample may still fail, and the agency accepts some risk that truly less than 95 percent of the vendors population of TCE errors falls within the precision limits but a particular sample may still pass acceptance.

Table 15. Acceptance Sampling Plans for Attributes with assumed p_1, p_2, n, c, α, and β

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

Example: A vendor produces a sample of AADTs to match those of the continuous traffic count values used to perform a validation test. There are 200 total AADT values. For each of the 200 values, a TCE is calculated relative to the corresponding ground truth values. Each TCE is compared to the minimally 95% probability, 95% TCE Population Error Range (%) from Table 14 for its ground truth volume range. For those with reference 500-4,999, each TCE is counted as a failure if it is not within the interval (-43.4,+43.4). Similarly, the TCEs for sites with reference volume 5,000-54,999 are evaluated against the interval (-33.6,+33.6), and the TCEs for sites with reference volume 55,000+ are evaluated against the interval (-22,+22), Suppose the numbers of failures are 1, 3, and 2 respectively. Table 15 is then consulted with a sample size of n=200, and it identifies a single sample plan with c=7. Since the total number of failures was 6 and $6 \le 7$, the sample passes validation for precision. Specifically, the sample plan had no more than a five percent risk (α) of failing to pass if the vendor's true population within the interval limits was 100*(1-0.02), or 98 percent or more. For the agency, the sample plan had no more than a five percent risk (β) of passing if the vendor's true population within the interval limits was 100*(1-0.065), or 93.5 percent. Note that the actual estimate of the number of failures, 6/200, corresponds to 3 percent, or 97 percent observed to have fallen within the limits.

CONSIDERATIONS OF MULTIPLE COMPARISONS

To validate that estimates exhibit acceptable accuracy and precision, a vendor may need to pass as many as nine tests for the nine identified volume ranges times the two accuracy limits, for a total of 18 accuracy tests. Additionally, the vendor must pass the precision limit test. Since the limits for each of these tests are associated with a probability, some tests may produce a passing result even if the true accuracy or precision from the vendor is not within the identified limits. Conversely, some tests may produce a failing result even if the true accuracy or precision from the vendor is within the identified limits.

To be fair in this validation process because of the latter risk, the agency may elect to not mandate a "zero failure" policy. For instance, if 16 of the 18 volume range tests for accuracy pass, and the ones that do not are both close to the limit, and do not exhibit a particular pattern (e.g., they are all tests within a particular volume range), the agency may exercise some discretion and still declare the validation as passing. Conversely, due to the first risk, acceptance of all tests should not be considered the end of validation. For instance, if a new data source or site for ground truth data became available, a confirmatory test might be performed off the full validation cycle. Also, as detailed previously, due to the risk of a vendor manipulating their product to match the reference site volumes (whether provided by the agency or obtained in some other way), data quality that appears too good to be true should also be examined critically. One evaluation technique that has been used is to determine which continuous count site volumes have likely been available to the vendor, and then to perform the validation excluding these sites. If the results are similar whether these sites are included or excluded, it diminishes the risk that the vendor's product is specifically targeted to match the reference sites.

CONSIDERATIONS OF SAMPLE SIZE

A natural question to ask for this process is what sample size is required to complete a validation. Unfortunately, no simple answer exists for this question due to the large number of comparisons that will be made and to lack of a priori knowledge regarding the distribution of errors in the candidate data population. However, some insights can be provided from the data used to develop the proposed limits. As discussed below, the validations documented to date have utilized several hundred sites to get acceptable conclusions.

In the case of the precision estimates, the design of the sampling plan helps to answer the question of whether there is an adequate sample size. If the sample size is adequately large, a suitable plan can be found that does not overly penalize either the agency or the vendor in risk of an incorrect decision being made from sample data. For this, the p_1 and p_2 variables are the critical indicators. If these two values can be selected to bracket the desired population containment level of 95% (or correspondingly a failure rate close to 5%), while simultaneously having an available acceptance number, c, and being able to get both α and β to levels of 0.05 or below, the sample size will have been adequate.

To quantify the relative meaning of the Table 15 sample plans, an analysis was conducted on the data used to develop the acceptance limits. The p_1 and p_2 probabilities of Table 15 were obtained by scaling the original distribution of TCEs (i.e., multiplying each value by the same number) without any translation (i.e., positive or negative bias). Then the relative scaling was compared between each sampling plan and that of the n=1000/c=50 sample plan as the reference. Table 16 shows the results of this analysis. With a sample plan of n=25/c=0 for the acceptance test, the distribution of TCEs would have to be shrunk to 37 percent of their relative contraction for the n=1000/c=50 plan to achieve the same five percent risk of false failure. This more compressed data for n=25/c=0 corresponds to the p_1 of 0.002 where 99.8% of the TCE population now falls within the unscaled 95 percent population limits, whereas the p_1 of 0.04 for n=1000/c=50 meant that 96 percent of the TCE population fell within the unscaled 95 percent population limits. The vendor may achieve acceptance with a sample plan as small as n=25/c=0, but its relative precision would likely have to be much better than that of factored, portable counts.

With a sample plan of n=25/c=0 for the acceptance test, the distribution of TCEs would have to be expanded 133 percent beyond that of the n=1000/c=50 acceptance test to limit to the same five percent risk of false acceptance. This more expanded data for n=25/c=0 corresponds to the p_2 of 0.113 where 88.7 percent of the TCE population now falls within the unscaled 95 percent population limits, whereas the p_2 of 0.063 for n=1000/c=50 meant that 93.7 percent of the TCE population fell within the unscaled 95 percent population limits. The agency may utilize a sample plan as small as n=25/c=0, but it risks allowing probe data vendors with considerably more variable data to still have a reasonable chance of falsely passing acceptance.

The preceding analysis is based on the TCE data used to determine the limits as well as the modified limits themselves. As such, these results are only strictly applicable to the source data, and values would differ for other source data. However, they help inform the selection of a preferred sample size for the acceptance test as the largest value the vendor and agency can reasonably assess.

Table 16. Total Count Error Scaling Factors to Attain Same Five Percent False Acceptance or False Failure as Maximum Sample Plan of *n*=1000/*c*=50

n	c	TCE Scaling to Get Same Risk of False Failure (0.05) as that of *n*=1000/*c*=50	p_1	TCE Scaling to Get Same Risk of False Acceptance (0.05) as that of *n*=1000/*c*=50	p_2
25	0	37%	0.002	133%	0.113
50	1	56%	0.007	118%	0.091
80	2	62%	0.01	109%	0.077
100	4	78%	0.02	118%	0.089
200	7	78%	0.02	101%	0.065
400	16	87%	0.027	98%	0.060
1000	50	100%	0.04	100%	0.063

In the case of accuracy measures, the Table 14 limits are adjusted for sample size. The standards are wider for smaller sample sizes where more uncertainty is expected and narrower for larger sample sizes where less uncertainty is expected. This controls the risk of false failure to a comparable value (nominally 5 percent) regardless of the sample size selected. The converse risk of the agency falsely accepting data which is truly not comparable in accuracy to factored, portable counts was also considered with the original data. This evaluation can assist the agency in deciding whether the risk of falsely accepting non-comparable data is worth the additional sample sizes.

The methodology for this accuracy evaluation consisted of applying both positive and negative translations (shifts) to the base TCE data and finding how far each direction had to move until such shifted TCE data would only erroneously “pass” acceptance 2.5 percent of the time. The translations outside the 2.5 percent pass regions for positive and negative shifts would collectively represent a five percent probability of false acceptance. The evaluation does not consider the impact of scaling. It also is specific to the TCE data evaluated and the limits of Table 14. The sample sizes evaluated are based on a representative sample of sites across the range of AADTs and the determination of acceptance is against a weighted Median Error or MAPE. For example, If 1/3 of sites in a sample had AADT of 500-4,999 and corresponding median TCE error of ± 8.7 percent, 1/2 had AADT of 5,000 – 54,999 with corresponding median TCE error of ± 6.9 percent, and 1/6 had AADT of 55,000+ with corresponding median TCE error of ± 9.4 percent, the sample median TCE would be compared to a limit of ± [(1/3)*8.7 + (1/2)*6.9 + (1/6)*9.4] = ± 7.9 percent.

Table 17 shows the boundaries outside which false acceptance risk is below five percent. Note for the TCE median error measure of accuracy, an evaluation based on N=25 will not fall below five percent probability of falsely accepting until the vendor's true population median error is more than 13 percent low or more than 12 percent high relative to that of factored portable counts. This degree of risk is cut nearly in half with a sample of size N=100, where the vendor's chances of false acceptance fall below five percent when their true population median error is more than 7.1 percent low or more than 6.1 percent high relative to that of factored portable counts.

The effect sizes that result in a low chance of false acceptance in MAPE are higher than those of the median error. The MAPE is less sensitive as a measure of accuracy to shifts. However, the same pattern of better sensitivity for larger sample sizes still applies to MAPE.

Table 17. Additive Factor Interval for TCE Median Error and MAPE Outside of Which False Acceptance Rate Decreases Below 5 Percent

N	Additive Factor Interval for TCE Median Error to Be No More Than 5 Percent False Acceptance Rate	Additive Factor Interval for MAPE to Be No More Than 5 Percent False Acceptance Rate
25	≥ +12.0% or ≤ -13.0%	≥ +17.5% or ≤ -18.5%
50	≥ +8.7% or ≤ -9.8%	≥ +14.4% or ≤ -15.6%
100	≥ +6.1% or ≤ -7.1%	≥ +12.1% or ≤ -12.9%
200	≥ +4.3% or ≤ -5.2%	≥ +10.1% or ≤ -11.1%
1000	≥ +2.2% or ≤ -3.2%	≥ +7.0% or ≤ -8.0%

Beyond the effect size results of Table 16 and Table 17, the sample size for evaluation should also be selected to be adequately representative of volume ranges, types of sites, geography, and other factors so the results can be considered appropriate to extend to the full population of similar sites.

For reference, the research in this pooled fund project utilized over 800 sites to evaluate the probe-based AADT product. The state of Minnesota has evaluated the same vendor and obtained acceptance to use its data based on a validation effort with about 440 sites.[13] Hence, values in the hundreds are currently the norm.

EXAMPLE VALIDATION DEMONSTRATION

Under the Pooled Fund project TPF-5(384), Non-Traditional Methods to Obtain Annual Average Daily Traffic (AADT) Evaluation and Analysis, the product of one vendor was evaluated by several different contractor teams with respect to its accuracy and precision for estimating AADT of a large national sample of continuous count sites. After some cleaning and preparation, a dataset of 827 sites was prepared with paired AADT estimates from the vendor

[13] Turner, Shawn, Evaluation of Streetlight Data's Traffic Count Estimates from Mobile Device Data, Research Project 2020-30, November 2020, accessed via internet at https://mdl.mndot.gov/items/202030

and the corresponding ground truth observations of the continuous counts. A total of 827 traffic monitoring sites were used in the study. AADT from each site were measured by two methods with one from the vendor's big data procedure and the other one from the traditional count method. These two AADTs formed a pair of measurements. In the previous evaluations, large differences between the toll volumes estimated by the vendor and from the ground truth data were revealed. These 184 sites have been removed from this evaluation, leaving an overall evaluation of 643 sites.

This example uses the acceptance criteria documented in Table 14, and then calculates the accuracy and precision statistics as detailed above and compares the outcomes to the acceptance limits to establish a pass or fail criteria. Note that accuracy and precision estimates are provided for the eight sample sites with volumes under 500, though there are not tabulated acceptance limits for these.

Table 18. Observed Accuracy by Bias from 643 Sites using StreetLight data

AADT Volume Range	Sample Sites (n)	Acceptance Limit of TCE Median Error (Bias) (%)	StreetLight Observed TCE Median Estimate (Bias) (95% CI)	Acceptance Outcome
0-500	8	*Unknown*	46.7 (6.3,102.5)	N/A
500 – 4,999	147	± 3.7	2.0 (-1.5, 6.0)	**Pass**
5,000 – 54,999	376	± 2.3	-0.3 (-0.8, 1.9)	**Pass**
55,000+	112	± 5.1	0.3 (-0.9, 1.4)	**Pass**

For the median error, the probe-based data passed the acceptance test in all volume ranges for which there was a standard. For instance, the median bias for the 147 sites with AADT between 500 and 4,999 was 2.0 percent, which falls within the acceptance limit (calculated by the applicable equation in Table 14) of ± 3.7 percent. For this evaluation, 95 percent confidence intervals were calculated for each TCE median estimate. For sites in the 500-4,999 range, the 95 percent confidence interval for the 2.0 percent TCE median bias was (-1.5, 6.0). This interval is not directly used in the acceptance determination, but it does provide additional context. The median estimate and its entire 95 percent confidence interval are entirely within the acceptance limits for the 5,000-54,999 and 55,000+ volume ranges, but even though the median bias of 2.0 percent is accepted against a limit of (-3.7, +3.7), its upper uncertainty bounds do extend outside the limit range.

Table 19. Observed Accuracy by MAPE from 643 sites using StreetLight data

AADT Volume Range	Sample Sites (n)	Acceptance Limit of Mean Absolute Percent Error (MAPE) (%)	StreetLight Observed MAPE (Upper 95% CB) (%)	Acceptance Outcome
0-500	8	*Unknown*	46.7 (85.3)	*N/A*
500 – 4,999	147	14.9	12.5 (14.9)	**Pass**
5,000 – 54,999	376	10.8	6.5 (7.2)	**Pass**
55,000+	112	9.6	5.8 (7.2)	**Pass**

The MAPE accuracy levels meet the acceptance limits for all three volume ranges for which there was a standard. For instance, the MAPE estimate for the 147 sites with AADT between 500 and 4,999 was 12.5 percent, which falls within the acceptance limit (calculated by the applicable equation in Table 14) of 14.9 percent. An upper 95 percent confidence bound for the observed MAPE is also provided in Table 19. Although not used directly in the acceptance limit, the fact that this upper bound for the AADT volume ranges of 5,000-54,999 and 55,000+ also falls within the acceptance limit (i.e., 7.2 < 10.8 and 7.2 < 9.6, respectively) provides additional evidence that the probe-based data estimates are no poorer in accuracy estimation than may be expected from 48-hour, short-term counts factored to AADT.

The evaluation of precision was not as favorable. Using a sampling plan with the 635 sites (from the 643 total sites, the 8 sites with volume below 500 cannot be included as they have no acceptance standard), an acceptance sample number of 33 sites falling outside their respective 95 percent population limits is chosen from the principles discussed in the section, "HOW TO ASSESS ADEQUACY OF SAMPLE PRECISION AGAINST TABLE 14 LIMITS." This sample plan affords good protection (α=0.05) against a vendor's true population proportion within the limits of 1-p_1 = 1 - 0.04=96 percent or more erroneously failing. It also provides protection (β=0.05) against a vendor's true proportion within the limits of 1- p_2 = 1 - 0.069=93.1 percent or less erroneously passing. The detailed results are shown in Table 20, where the 53 total samples outside the limits is well above the acceptance threshold of 33, and it is concluded that the 95 percent population proportion of the vendor's estimates does not fall within the acceptance limits. Specifically, only 91.6 percent of sites fell within population limits that would be expected to contain 95 percent of the population based on 48-hour, short-term counts.

Table 20. Observed Precision by 95 Percent Population Tolerance Bounds from 643 sites using StreetLight data

AADT Volume Range	Sample Sites (n)	Minimally 95% Probability, 95% TCE Population Error Range (%)	Sites Outside Range	Acceptance Outcome
0-500	8	*Unknown*	*N/A*	N/A
500 – 1,999	61	±43.4	11	**Fail** (53>33) Acceptance Sampling Plan; n=635, Accept <= c=33 failures; α=0.05, β=0.05 against p_1=0.04, p_2=0.069
2,000 – 4,999	86	±37.8	11	
5,000 – 9,999	99	±33.6	4	
10,000 – 19,999	115	±29.9	6	
20,000 – 34,999	100	±26.7	7	
35,000 – 54,999	62	±24.2	5	
55,000 – 84,999	41	±22.0	6	
85,000 – 124,999	39	±20.4	1	
125,000+	32	±20.0	2	
Total	643		53	

The results of this example evaluation were generally positive, and the candidate technology for AADT estimation might have been considered acceptable under some circumstances. For instance, 22 of the 53 precision limit failures were for estimates below an AADT of 5,000 (Table 20). As such, the agency might consider accepting estimates for this alternative methodology for AADT values above 5,000. If the vendor improved its estimation of low-volume AADT, these might be tested later and added to the acceptable use range.

This recognition of the acceptability of StreetLight estimates from the 643 sites is only an example. Recalling that this example started with the exclusion of 184 toll sites with highly variable AADT estimates, the StreetLight product methodology would likely need to be improved to reintegrate these exclusions into the validation testing before an agency would affirm it had uncertainty comparable to or better than factored short-term counts.

VALIDATION IN CASES WITH ONLY PORTABLE COUNT DATA

The preceding methodology evaluates the accuracy and precision of non-traditional measurements when the sites to which they will be applied are comparable to similar sites with continuous count data to allow a comparison where a ground truth is known. However, some sites for which non-traditional count data are desired do not have any analogous continuous count data, but they may have had some portable counts performed. These counts could be considered to serve as ground truth values, but they are importantly very different than the continuous counts. While continuous counts are subject to some measurement error, that error is generally considered small enough that validation with these allows the estimated AADT from such counts to serve as a direct ground truth value. This is not true of portable count locations where the best case scenario, a well-calibrated count from the most recent time period, must be factored by monthly (seasonal) and day of week factors from a factor grouping, (other factors such as hour of day or axle correction factors [ACF] may be applied) and then may be as inaccurate in AADT estimation as the values derived from the 2015 research.

To allow some consideration of potential validation in these cases, one approach was completed by Texas A&M Transportation Institute for MnDOT[14]. Using the terminology of this report, Turner takes the two equations:

$TCE = AADT_{error} = 100*(AADT_{48\text{-}hour} – AADT_{true})/AADT_{true}$
$ATCE = \text{Absolute } AADT_{error} = 100*|AADT_{48\text{-}hour} – AADT_{true}|/AADT_{true}$

Recognizing that the vendor's estimate, which is denoted as $AADT_{SLD}$ plays the part of $AADT_{48\text{-}hour}$ and the reference $AADT_{true}$ is now $AADT_{48\text{-}hour}$, which is not a known value, this evaluation considers potential error ranges for it.

Using other data provided by Minnesota, this evaluation assumes a portable count site with volume in the 1,500-4,999 range is considered to have approximate AADT estimate uncertainty of ± 20% for a same-year, annualized short-term count. In the formulas above, the potential of that uncertainty is accounted for as follows:

a) If "$AADT_{SLD} > 1.2*AADT_{48\text{-}hour}$", replace with "$AADT_{SLD} - 1.2*AADT_{48\text{-}hour}$". Within the uncertainty bounds, this is largest error that could be unequivocally associated with SLD.
b) If "$1.2*AADT_{48\text{-}hour} > AADT_{SLD} > 0.8*AADT_{48\text{-}hour}$", replace with 0. Within the uncertainty bounds, it is not possible to unequivocally associate any error to SLD.
c) If "$AADT_{SLD} < 0.8*AADT_{48\text{-}hour}$", replace with "$AADT_{SLD} - 0.8*AADT_{48\text{-}hour}$". Within the uncertainty bounds, this is largest (absolute) error that could be unequivocally associated with SLD.

After making these replacements, the calculation of accuracy estimates for bias, MAPE, NRMSE, or absolute 68th and 95th percentile errors proceed as for the continuous count reference case. This approach is understandable, but it should be noted that it will be very favorable to the vendor since it selects the minimum possible vendor error in every case.

Alternative Methodology Proposal for Using Portable Counts as a Reference

Alternatively, a different set of reference error bounds can be generated using a simulation approach like that described in the section "HISTORICAL ACCURACY AND PRECISION EXPERIENCED WHEN USING SHORT-TERM, 48-HOUR PORTABLE COUNTS FOR ESTIMATING AADT", but where the assumed ground truth value is based on data from a portable count site.

For accuracy as measured by TCE median error or MAPE, the approach for calculating 95% confidence bounds was identical to that described above when comparing to a true AADT from a CCS, with a small change to the TCE and ACTE formulas. Rather than using the $AADT_{true}$ as the ground truth reference value, another randomly selected $AADT_{48\text{-}hour}$ estimate from the same site and year was selected as the ground truth. Thus, the TCE and ACTE were computed as

$TCE = AADT_{error} = 100*(AADT_1 – AADT_2)/AADT_2$

and

[14] Turner, Shawn, "Evaluation of StreetLight Data's Traffic Count Estimates from Mobile Device Data", Minnesota Department of Transportation Research Project 2020-30, November 2020.

ATCE=Absolute $AADT_{error} = 100*|AADT_1 – AADT_2|/AADT_2$

where $AADT_1$ is the estimated AADT from a short-term count and $AADT_2$ is another estimated AADT from a short-term count at the same site to serve as the assumed "ground truth." Each $AADT_1$ is paired with a random $AADT_2$ from the same site, so even if multiple counts are sampled from the same site, a difference reference AADT may be used for each count. For the quantile regression models, the true AADT from the continuous count data was still used as the predictor variable, rather than $AADT_2$, to eliminate variability in the explanatory variable. All other steps in determining the sample size dependent bounds still applied. This process yields confidence bounds for the expected accuracy (TCE median error or MAPE) when an alternative method of AADT estimation is used in comparison to an AADT estimated from a portable count site.

As shown in Table 21, the TCE median estimates with short-term counts as the ground truth are at least as large as those with true AADT as the ground truth values (from Table 9). Initial simulation results identified some sample size cases where the portable reference limits were narrower than those from the true AADT reference. This is due to the biasing effect in the limits when the set of factors used to estimate AADT from 48-hour counts does not exactly match the true AADT. This biasing effect is less pronounced in the limits with portable counts as the reference because factoring bias differences between the estimated AADT on one day and another reference day tend to offset. To prevent the counterintuitive presentation of narrower limits for the portable count reference results, Table 21 equations and bounds were modified so that the limits for the portable count scenarios are not less than those of the true AADT reference.

Table 21. Portable Count Reference Versus True AADT Reference, Sample Size Dependent TCE Median Error as a Function of Sample Size, N, for same year, 48-hour short-term counts for AADT estimates

AADT Volume Range	Reference AADT	Minimally 95% Probability, TCE Median Error (Bias)(%)					
		25 ≤ N ≤ 1000	N=25	N=50	N=100	N=200	N≥1000
500 – 4,999 (low)	Factored 48-Hour Portable Count	$36.35-22.42(\log N) +3.67(\log N)^2$	± 12.2	± 8.9	± 6.2	± 4.2	± 2.1
	True AADT	$25.65-15.64(\log N) +2.54(\log N)^2$	± 8.7	± 6.4	± 4.5	± 3.1	± 1.6
5,000 – 54,999 (medium)	Factored 48-Hour Portable Count	$29.77-19.86(\log N) +3.55(\log N)^2$	± 8.9	± 6.3	± 4.3	± 2.9	± 2.1
	True AADT	$19.38-11.71(\log N) +1.97(\log N)^2$	± 6.9	± 5.2	± 3.8	± 2.9	± 2.0
55,000 + (high)	Factored 48-Hour Portable Count	$41.31-27.41(\log N) +4.87(\log N)^2$	± 12.5	± 8.8	± 6.0	± 4.0	± 2.9
	True AADT	$26.15-15.59(\log N) +2.59(\log N)^2$	± 9.4	± 7.1	± 5.3	± 4.0	± 2.7

Log = Base 10 logarithm

The MAPE estimates with short-term counts as the ground truth values were larger than those with true AADT as the ground truth values (from Table 11) as shown in Table 22.

Table 22. Portable Count Reference Versus True AADT Reference, Sample Size Dependent MAPE as a Function of Sample Size, N, for same year, 48-hour short-term counts for AADT estimates

AADT Volume Range	Reference AADT	Minimally 95% Probability, MAPE Upper 95% CB					
		25 ≤ N ≤ 1000	N=25	N=50	N=100	N=200	N≥1000
500 – 4,999 (low)	Factored 48-Hour Portable Count	$40.00-11.67(\log N)+1.53(\log N)^2$	26.7	24.6	22.8	21.2	18.8
	True AADT	$27.75-8.60(\log N)+1.23(\log N)^2$	18.1	16.7	15.5	14.5	13.0
5,000 – 54,999 (medium)	Factored 48-Hour Portable Count	$29.83-8.93(\log N)+1.34(\log N)^2$	20.0	18.5	17.3	16.4	15.1
	True AADT	$20.46-6.23(\log N)+0.97(\log N)^2$	13.6	12.7	11.9	11.3	10.5
55,000 + (high)	Factored 48-Hour Portable Count	$38.04-18.02(\log N)+3.08(\log N)^2$	18.9	16.3	14.3	12.9	11.7
	True AADT	$26.05-12.41(\log N)+2.15(\log N)^2$	12.9	11.2	9.8	8.9	8.2

Log = Base 10 logarithm

For precision, an additional simulation was required to estimate confidence bounds for the tolerance intervals of the TCE errors when compared to an assumed "ground truth" AADT based on portable count data. First, one AADT estimate from a short-term count was randomly selected from each site to serve as the assumed true AADT ($AADT_2$) of the site. The TCE was then calculated for all estimated AADTs from each site using the randomly selected $AADT_2$ as the assumed ground truth. A quantile regression model was fitted to this set of TCEs and the 2.5^{th} and 97.5^{th} quantile estimates were obtained across the range of true AADTs; as with the accuracy approach, the true AADT from the continuous count data was still used as the predictor variable. The upper and lower quantile estimates were stored for a set of pre-defined true AADT values, and the process was repeated 10,000 times with different randomly selected AADTs to serve as the assumed ground truth for each site. Following these iterations, the 5^{th} quantile of the 2.5^{th} quantile estimates and the 95^{th} quantile of the 97.5^{th} quantile estimates were calculated as the one-sided lower and one-sided upper 95% confidence bounds, respectively. These two bounds form an interval within which 95 percent of the TCE population (compared to an assumed ground truth AADT estimate based on short-term counts) is expected to be contained with at least 95 percent confidence. The precision bounds against 48-hour portable counts as compared to those with true AADT (from Table 13) are shown in Table 23. The precision bounds with portable counts as reference are substantially larger as would be expected since the reference AADT for these comparisons is subject to much more variability than the continuous count AADTs.

Table 23. Portable Count Reference Versus True AADT Reference Precision for same-year, 48-hour short-term counts for AADT Estimates

AADT Volume Range	Reference AADT	Lower 95% Confidence Bound for the TCE 2.5th Percentile Error (%)	Upper 95% Confidence Bound for the TCE 97.5th Percentile Error (%)	Minimally 95%Probability, 95% TCE Population Error Range (%)
500 – 1,999	Factored 48-Hour Portable Count	-44.6	78.2	±78.2
	True AADT	-30.4	43.4	±43.4
2,000 – 4,999	Factored 48-Hour Portable Count	-39.8	67.3	±67.3
	True AADT	-27.9	37.8	±37.8
5,000 – 9,999	Factored 48-Hour Portable Count	-36.6	60.6	±60.6
	True AADT	-26.1	33.6	±33.6
10,000 – 19,999	Factored 48-Hour Portable Count	-34.1	56.3	±56.3
	True AADT	-24.5	29.9	±29.9
20,000 – 34,999	Factored 48-Hour Portable Count	-32.5	54.3	±54.3
	True AADT	-23.2	26.7	±26.7
35,000 – 54,999	Factored 48-Hour Portable Count	-31.6	53.5	±53.5
	True AADT	-22.1	24.2	±24.2
55,000 – 84,999	Factored 48-Hour Portable Count	-30.8	53.4	±53.4
	True AADT	-21.2	22.0	±22.0
85,000 – 124,999	Factored 48-Hour Portable Count	-30.3	53.3	±53.3
	True AADT	-20.4	19.9	±20.4
125,000+	Factored 48-Hour Portable Count	-30.0	53.2	±53.2
	True AADT	-20.0	19.1	±20.0

A set of limits for portable count-based reference AADTs is provided as Table 24, summarizing the results of Table 21, Table 22, and Table 23. If evaluating probe-based AADT values against short-term, 48-hour, factored count-based AADTs, the limits in Table 24 can be used instead of those in Table 14. This is a very significant benefit since states often have a much larger number of such portable count-based AADT estimates available than those of CCSs and the portable count-based AADT estimates often cover a more diverse set of geographic and roadway conditions, especially including lower-volume roadways. Furthermore, these portable count-based AADTs are not necessarily available to data vendors who provide the new estimates and this eliminates the concern that modeled AADT estimates are being matched to measured values rather than truly estimated from big data sources.

Table 24. Accuracy and Precision Limits Based on Historical Short-Term Counts Factored to AADT When Compared to Factored Portable Count Station Reference Data

AADT Volume Range†	Sample Size for Evaluation	Minimally 95% Probability, TCE Median Error (Bias) (%)	Minimally 95% Probability, MAPE Upper 95% CB	Minimally 95% Probability, 95% TCE Population Error Range (%)
500 – 4,999 (low)	N ≥ 1,000	± 2.1	18.8	± 78.2
	N = 200	± 4.2	21.2	
	N = 100	± 6.2	22.8	
	N = 50	± 8.9	24.6	
	N = 25	± 12.2	26.7	
5,000 – 54,999 (medium)	N ≥ 1,000	± 2.1	15.1	± 60.6
	N = 200	± 2.9	16.4	
	N = 100	± 4.3	17.3	
	N = 50	± 6.3	18.5	
	N = 25	± 8.9	20.0	
55,000 + (high)	N ≥ 1,000	± 2.9	11.7	± 53.4
	N = 200	± 4.0	12.9	
	N = 100	± 6.0	14.3	
	N = 50	± 8.8	16.3	
	N = 25	± 12.5	18.9	

†Roadways with AADT less than 500 have no pre-defined standards.

The Table 24 standards for median error and MAPE are a function of the number of continuous count comparison sites. If an agency is evaluating the limits for a different number of sites than shown in the table, it may do so by interpolating between the limits. Alternatively, Table 24a allows the limits for a particular number of sites, N, to be calculated by equation. Note that evaluations should include at least 25 reference sites. Statistical uncertainty for traditional factored counts for less than 25 sites will likely be higher than the values shown here, so the values for N=25 could be considered conservative references to use in such a case. However, evaluating a count program with so few reference sites risks missing important sources of variability and is not generally advisable. As with the continuous count reference standards, evaluation of factored portable count AADTs favors the largest number of representative sites possible. The specific effect size results of precision and accuracy measures of continuous counts as a reference (see Table 16 and Table 17) are not provided for the limits with factored portable counts as the reference, but the relative results are similar.

Table 24a. Equation Form of Accuracy and Precision Limits Based on Historical Short-Term Counts Factored to AADT When Compared to Factored Portable Count Station Reference Data

AADT Volume Range†	Sample Size for Evaluation	Minimally 95% Probability, TCE Median Error (Bias) (%)	Minimally 95% Probability, MAPE Upper 95% CB	Minimally 95% Probability, 95% TCE Population Error Range (%)
500 – 4,999 (low)	By Equation ($25 \leq N \leq 1000$)	$36.35-22.42(\log_{10}N)+3.67(\log_{10}N)^2$	$40.00-11.67(\log_{10}N)+1.53(\log_{10}N)^2$	± 78.2
5,000 – 54,999 (medium)	By Equation ($25 \leq N \leq 1000$)	$29.77-19.86(\log_{10}N)+3.55(\log_{10}N)^2$	$29.83-8.93(\log_{10}N)+1.34(\log_{10}N)^2$	± 60.6
55,000 + (high)	By Equation ($25 \leq N \leq 1000$)	$41.31-27.41(\log_{10}N)+4.87(\log_{10}N)^2$	$38.04-18.02(\log_{10}N)+3.08(\log_{10}N)^2$	± 53.4

†Roadways with AADT less than 500 have no pre-defined standards.

The reference accuracy and precision limits computed by the preceding methodologies provide a tool for agencies to use in their validation of data accuracy and precision. However, the specific calculated values are integrally tied to the historic research effort in 2015 and may not be fully representative of a specific agency's data quality. For instance, an agency may be using 24 or 72 hours counts for their short-term counts rather than 48-hour counts; they may treat weekend and holiday counting differently; or they may have a different mix of urban and rural count sites as were used in this evaluation. The quantile regression-based approach used for this effort could easily be replaced by single estimates within ranges of sites of a particular type. For that matter, even the basic structure of setting limits by volume ranges could be changed, as some agencies might prefer to have accuracy and precision metrics tied to site characteristics (e.g., urban/rural, functional classification). The selection of limits by volume ranges was a deliberate choice since it was observed in the legacy research that accuracy and precision measures differed by volume ranges so a single standard would not be appropriate for all sites.

While directly associated with portable counts, the methods of this section could also be applicable to sites with more robust counts, such as regional traffic management centers, or automated counts from traffic signal sensors, but where the calibration and data quality are not comparable to continuous traffic counters. In these cases, as long as a data uncertainty model can be measured or assumed, the methods of this section could be used to establish accuracy and precision limits for probe-based AADT estimates at these sites.

Finally, it is important to note that the simulation procedure employed to obtain these limits supposes that the accuracy and precision associated with AADT estimation from a sub-sample of a CCS is the same as that of a portable counter. For a variety of reasons, this may not be the case. Generally, the limits provided here should be expected to be a fair representation of the degree of error that exists in estimating AADT from short-term portable counts when compared to a factored portable count AADT as the reference for the same site. Therefore, these limits may be used to fairly assess accuracy and precision of an alternative technology without unduly penalizing such technology for the same uncertainty that exists with the reference technology of factored portable counts. These limits may be modified over time as more research is completed. Alternately, agencies may choose to produce their own limits using data specific to their locations/geographic area and consistent with other assumptions, so long as they are documented and statistically defensible.

VALIDATION IN CASES WITH NO COUNT DATA

The preceding methodology assumes there are some reference data to use for an evaluation with preference for ground truth data as continuous count stations, but the possibility of portable counts also being used as ground truth, after accounting for their uncertainty. In the most extreme case, a provider may propose to deliver estimates for a group of sites, such as very low-volume roads, where no portable count has been performed, or at least not in any recent time (over 10 years).

This scenario may be the final frontier for the use of non-traditional count data. From a statistical perspective, the options around validation are limited. It may be possible to develop a model of the volumes for these sites using sites in proximity (using conservation of flow) or associated by other roadway or geographic characteristics. One proposal from FHWA was to develop a conservation of mass flow model for low-volume roads based on the related network of roads for which volume is measured.

It seems reasonable to suggest that non-traditional estimates for these locations not be considered until the vendor has demonstrated accuracy and precision for all the associated roadway groups near these, where data are available for validation.

If a traditional count is unavailable in a higher-volume roadway due to operational or safety constraints, but the vendor's product should be successful on that type of roadway based on its record for other locations, an agency could certainly be justified in accepting the non-traditional count without true validation. It is expected that these cases would be limited. In other cases, such as a state with a large proportion of low-volume roadways, making that same decision would be risky, especially if the limitations of the non-traditional model, such as poor ping rates, apply directly to these locations.

APPENDIX C: BEST VALUE ASSESSMENT

OVERVIEW

In evaluating the best value assessment of bids, the evaluation must consider the specific criteria and requirements enumerated in the procurement documents. Purchase price alone should not be the governing criterion for assessing the best value for a given data product or service. Beyond the purchase price, there are any number of factors that can influence the state's decision to decide on which products or service to purchase (e.g., delivery, maintenance, and quality assessments). Although some aspects of the procurement criteria are quantitative (e.g., purchase price), many of the requirements may be qualitative in nature. Qualitative criteria may include who owns the data you have purchased or licensed, how the data will be delivered, vendor resources, and vendor past experience.

A best value assessment of bids examines both qualitative and quantitative criteria to compare the benefits the state will receive against the important factors associated with the purchase of the data products or services. Best value assessment evaluates the strengths and weaknesses, pros and cons, and risks and rewards of different bids and selects the bid that provides the best value to the state's data needs.

NOTE: Each agency has its own procedure for procurement and best value assessment. Only a generalized procedure is presented here.

PROCUREMENT STEPS

There are two types of procurement steps in the best value selection process (i.e., two-step and one-step).

- Two-step: This procurement type entails two stages. The first step is that the vendors submit their qualifications documentation. The vendors that meet the minimum level of qualifications will be invited to submit their technical and price proposals.
- One-step: In this procurement type, the vendors submit their qualifications and technical and price proposals all at the same time.

If a two-step procurement is followed, then selecting the vendor with the lowest price may make sense because that is already vetted for its qualifications. However, if a one-step procurement is followed, then it is crucial that price and non-price factors be considered.

There may be pros and cons for using either the two-step or one-step process. However, it is beneficial for the state to identify the process it intends to use early in the planning stages of the data procurement and communicate the chosen approach to all parties throughout the process.

EVALUATION FACTORS

The two most important questions asked in conducting best value assessment are:

1. **What are important factors to the state's data needs?** This is described in the specific criteria and requirements in the procurement documents. A recommended practice is to list all potential criteria and prioritize the list into three categories: (1) very important, (2) important, and (3) not important. Next, challenge why it is necessary to

expand the criteria to more than just the "very important" category criteria in the procurement decision.

2. **How important are the factors?** This is reflected in the weights given to them to identify their relative importance to the procurement objectives. In evaluating the bids for AADT estimate, the following can be considered:

   - Number of bids received
   - Difference between lowest bid prices and the agency's budget
   - Difference between low bid and other bidders
   - Spread or variation of all bids
   - Data delivery method
   - Estimates of the accuracy of each traffic variable being provided
   - Extent of the access to accurately functioning continuously operating traffic count data for vendors to calibrate their models/algorithms
   - Temporal and geographic coverage of the data
   - Data usage policy
   - Data ownership policy
   - Urgency or expected date of data delivery
   - Type of additional support services offered
   - Involvement of independent organization to conduct data validation
   - Current market conditions and workload of the offerors
   - Any other factors the contracting agency has determined to be important

SCORING OF EVALUATION FACTORS

There are two ways evaluation factors can be scored.

1. Adjectival Scoring: This is also known as merit or qualitative scoring. Proposal criteria are rated on a conceptual scale from positive to negative using adjectives to indicate the degree to which the proposal has met the evaluation factors. Adjectival scoring can be expressed in such terms as 'Excellent', 'Very Good', 'Good', 'Satisfactory', or 'Unsatisfactory'.
2. Direct Point Scoring: This is proposal criteria are scored on a numerical basis within the varying ranges assigned to each evaluation criterion and the scores are totaled for an overall score. For example, each criterion is given a score on a scale of 0 to 100. Assigning points accurately to specific criteria is crucial. Direct point scoring may be beneficial because it provides numerical scoring.

ASSIGNING A SCORE TO PURCHASE PRICE

One approach to assigning points to price is to award the maximum number of points to the reference price and the rest of the proposed prices are assigned prorated points relative to the reference price. The reference price can be the budgeted amount, or the lowest price, or the average of the proposed prices. Points are assigned to price proposals by dividing the reference price by the proposed price and multiplying the ratio with the maximum points. Assume there are three proposals with prices of $50,000, $60,000, and $65,000 and 100 points is given to the lowest price as reference price. Then, the first proposal gets 100 points ($50,000 / $50,000 x

100 = 100). The second proposal gets 83 points ($50,000 / $60,00 x 100 = 83) and the third proposal gets 77 points ($50,000 / $65,000 x 100 = 77). Note that if the average of the proposed prices is used as reference price, then points for price proposal less than the average will be greater than 100 points.

EXAMPLE OF BEST VALUE ASSESSMENT

Assume that a state has received three proposals for a data purchase effort. In addition to price, the state has three evaluation criteria. The weights assigned to criteria 1, criteria 2, criteria 3, and price are 25 percent, 15 percent, 10 percent, and 50 percent, respectively. The raw and weighted scores for each proposal are shown in Table 25. Based on the total weighed score, Proposal 1 offers the best value.

Table 25. Example of best value assessment.

Criteria		**Proposal 1**	**Proposal 2**	**Proposal 3**
Scores	Criteria 1	90	90	75
	Criteria 2	85	80	95
	Criteria 3	90	85	85
	Price	100	95	80
Weighted Scores	Criteria 1	2,250	2,250	1,875
	Criteria 2	1,275	1,200	1,425
	Criteria 3	900	850	850
	Price	5,000	4,750	4,000
Total Weighted Score		**9,425**	**9,050**	**8,150**

APPENDIX D: LITERATURE REVIEW ON AADT ESTIMATION FROM NON-TRADITIONAL DATA (BIG DATA)

INTRODUCTION

Traffic volume is an important metric for many applications in transportation engineering. To mention some, traffic volume is important for design and optimization traffic signal control, transportation project prioritization, road maintenance plans, allocation of funds, and more. Traditional methods of quantifying vehicle volume rely on manual counting, video cameras, and loop detectors. This method is limited to getting traffic volumes at specific locations. In addition, such efforts require significant labor and are expensive. To address this need, transportation agencies, researchers, and private sector companies have been exploring alternative solutions for estimating traffic volumes. With technological advancement in mobile sensors and mobile networks, probe vehicle-based big data has been particularly getting increased attention as a promising solution. Probe vehicles can record their trajectory data at high granularity, and a variety of analysis methods can be used to analyze the trajectory data and estimate numerous transportation parameters.

Big data is defined as a dataset that is too large in size and too complex in nature which needs special treatment when storing, transferring, sharing, curating, querying, and analyzing the data. The advantage of such data is that it can be collected relatively inexpensively, and it is expected to provide superior insights because of higher spatial and temporal coverage it offers, which has not been previously possible. However, the relatively low penetration rate (i.e., 2 percent to 10 percent) of the probe vehicle data remains the core challenge with respect to drawing the whole picture of travel patterns based on such big data analytics.

Although probe vehicle-based volume estimates offer an opportunity to easily determine traffic volume, there is no formal process in place to examine the accuracy of data products derived from the probe-based data in relation to the traditional traffic data collection and processing techniques. The accuracy of AADT estimates from probe-vehicles is expected to improve as more data from mobile devices is being collected and data providers continue to enhance their roadway traffic volume prediction methos.

VOLUME DATA VENDORS

There are numerous private companies (as of when this was researched) that provide volume data or related data products. The following are some of the vendors the research team identified.

- **Wejo –** Wejo sources much of its data from GM vehicles but is expanding to cover other OEMs. Wejo also has a data partnership with Geotab.
- **Iteris –** Like so many others in this list, Iteris purchases Wejo data to drive several of its products.
- **TomTom –** Based in Europe, TomTom has a variety of probe vehicle-based traffic data products.
- **INRIX –** INRIX has a comparatively large data panel of over 30 vendors across different markets, primarily a mix of connected cars and commercial vehicle fleets, but now has some mobile device data sources.

- **HERE –** Like INRIX, HERE sources data from multiple providers—many of whom used to be based on cell phone users but are now more connected vehicles.
- **Future Mobility Labs –** Future Mobility Labs (FML) is startup company established by University of Maryland researchers.
- **Streetlight Data –** The bulk of Streetlight's (SLD's) data are sourced from mobile device large location-based service (LBS) data providers.
- **Geotab –** Geotab is a telematics provider for large fleets such as commercial shipping companies, distributors, etc. and it does have a strong presence and high-quality freight movement data.
- **Replica –** Replica represents movement by combining data from three primary sources: public use population census data, proprietary locational data from telecommunications and other IT infrastructure in the region, and field observations data from customer public agencies.
- **Streetlytics –** Streetlytics is owned by Bentley Systems and sources its data from mobile phone locations, traffic counts, government surveys, and hourly speeds, points of interest, routable transportation systems, and demographic and business data.
- **Strava –** provides pedestrian movement and volume data.

SOURCES OF BIAS IN BIG TRANSPORTATION DATA AND MITIGATION STRATEGIES

Big data from mobile phones, social media, and on-board vehicle systems have inherent bias and accuracy due to representation issue, phone users with non-typical movement patterns, and data collected only when network connections are present. In addition, groups from lower-income socio-economic groups are less likely to own a smartphone or have a 3G or 4G subscription, thus producing less geocoded information for travel patterns. Such bias might distort the analysis and products developed using big data. Also, big data from mobile phone locations can obscure shorter trips in which the origin and destination are within the same cellular tower range.

Recognizing this, Griffin et al. (2020) examined both the sources of bias and approaches to mitigate them. Their approach was based on a review of published studies as well as interviews with targeted stakeholders and experts. Griffin et al. have discussed four categories of bias, which includes sampling, measurement, demographics, and aggregation. They discussed the mitigation strategies for each bias type as shown in Table 26.

Table 26. Big data bias type and mitigation strategies.

Types of Bias	Mitigation Strategy
Sampling, Coverage, or Non-response Bias	Data Fusion (add multiple providers, or weighting with Census)
Measurement Bias	Filtering unreasonable data points
Demographic Bias or Social Desirability	Weighting with traditional counts or surveys, Validation with mixed methods
Aggregation Bias	Spatial or temporal modeling, including Markovian techniques

ISSUES WITH VOLUME DATA PRODUCTS FROM PROBE-BASED DATA PROVIDERS

There are numerous issues with the volume data and related data products provided by big transportation data providers. The major issues are listed below.

- Reliance on agency sensors – these data are largely modeled and are reliant on spot volume data collection, either permanent traffic or short-term counters. Vendors ingest AADT or other ITS sensor data and fuse that with their other probe datasets to estimate volumes and create “profiles” of volumes on the roads. Agencies are excited about vendor-provided volume data and (once proven) want to stop funding continuous count stations. However, the vendors still rely on continuous count stations to feed their models.
- Format inconsistencies – these exist in terms of spatial resolution (TMCs vs. proprietary segments vs. agency linear referencing systems), temporal resolution (5-minute, 10-minute, 15-minute, hourly), profile type (weekday vs. weekend vs. each day-of-week vs. seasonal), and update frequency (true real-time vs. monthly vs. quarterly vs. annually).
- Lack of validation – transportation experts, agency data users, and others are still working on determining the best methods for validation—including how often to validate.
- Lack of maturity – some companies are still determining how to create these volume related data products despite having active orders in place.
- Variable methodologies for determining volumes – although how volume and related data products are derived are proprietary information, there seems there is difference in terms of how the estimates are derived. Some companies use trajectory, on-street parking, speed data, and historic volumes, while others leverage probe sample sizes. Incorporation of weather, incidents, and related external factors affecting traffic volume is yet to be considered for accurate estimates of volume and related products.

Other issues related to the reference volume data (ground truth data) that it may be used by the probe-based data vendors in developing commercial data products include:

- Calibrations – Ensuring continuous count stations (CCS) and portable devices used for counting traffic volumes are working as intended and accurately is critical. Per the Traffic Monitoring Guide (TMG), all traffic counting devices should be calibrated annually.
- Proper documentation – When setting up traffic counting devices, documenting how the devices were set up, what they are for, how they count traffic, maintenance history, etc. is important to accurately understand the data that they collect. This allows for reproducible data collection and analysis process.

BIG DATA-BASED AADT ESTIMATION APPROACHES AND RESULT ACCURACIES

Zhang and Chen (2020) developed a new and enhanced method for estimating AADT using probe vehicles. Their work estimated AADT for the entire state of Kentucky.[15] The authors explored probe data in two ways: 1) deriving an annual average daily probes (AADP) variable

[15] Zhang, X., & Chen, M. (2020). Enhancing statewide annual average daily traffic estimation with ubiquitous probe vehicle data. *Transportation Research Record*, *2674*(9), 649-660.

from hourly probe counts, and 2) deriving betweenness centrality (BC) variable calculated using probe speeds. Then, they developed a random forest model to predict AADT. They compared the model results that use only sociodemographic and roadway characteristics with the model results that incorporates AADP and BC with a random forest model. Their findings indicated that when AADP and BC are incorporated in the random forest model, the resulting accuracy of AADT estimates improved by 30 to 37 percent for all roads and 23 to 43 percent for highways in functional classes 5–7. For roadways with more than 53 AADP (i.e., 2.2 probes per hour), the median and the mean absolute percent errors are below 20 and 25 percent, respectively. Including AADP and BC and using the random forest model results in results with accuracy exceeds those previously reported for statewide applications. These results demonstrate the value of the probe data for enhancing AADT estimation.

A study by Hou et. al. (2018) examined the used of GPS probe vehicle trajectory data to estimate traffic volumes in conjunction with other variables like speeds, road characteristics, incident reports, weather information, and temporal information.[16] To enhance the volume predictions, a neural network was used to model the relationship between GPS traces (based on a dataset of INRIX trip records in February, June, July, and October of 2016) and real-time traffic volume at 12 ATR reference stations in Maryland. The average penetration rates of GPS probe vehicle traces at the 12 reference count locations varied from 0.18 to 0.72 percent and the median was 0.57 percent. The average hourly probe volumes at the 12 reference stations varied from 22.3 to 62.3 vehicles, and the median was 37.3 vehicles. The neural network model was trained using the data from the other eleven reference locations.

A neural network model with GPS vehicle trace as an additional input indicated that the R^2 varies from 0.61 to 0.94, with a median of 0.82. MAPE varies from 14 to 48 percent with a median of 27 percent. Without the GPS probe vehicle traces (models operated only on auxiliary data such as number of lanes, weather, time of day, day of week, etc.), the R^2 varies from 0.49 to 0.90, with a median of 0.73 compared to 0.82 with GPS traces included in the model. Similarly, the MAPE metric varies from 16 percent to 54 percent with a median of 37 percent without the GPS traces, compared to 27 percent with GPS traces. The difference in R^2 and MAPE with and without GPS trajectory data reinforces the value of probe vehicle information for improved volume estimates. Further investigation of AADT estimation error with respect to 10 percent capacity increases ranged from 3.5 percent to 13.7 percent, with an average of 9.5 percent. These results are within the preferred level of accuracy by transportation professionals, which is 10 percent of the roadway capacity.

Turner et al. (2020) evaluated the accuracy of 2019 StreetLight Data AADT estimates for Minnesota DOT comparing 442 sites with annual traffic data.[17] The accuracy results revealed that MAPE ranged from 8 percent to 10 percent for locations greater than 10,000 AADT and gradually increased to 42 percent for sites with less than 1,000 AADT. However, the researchers did mention that the accuracy of StreetLight Data has improved significantly since

[16] Hou, Y., Young, S. E., Sadabadi, K., SekuBa, P., & Markow, D. (2018). *Estimating highway volumes using vehicle probe data-proof of concept* (No. NREL/CP-5400-70938). National Renewable Energy Lab.(NREL), Golden, CO (United States).

[17] Turner, S., Tsapakis, I., and Koeneman, P. (2020). Evaluation of StreetLight Data's Traffic Count Estimates from Mobile Device DataShawn. Minnesota Department of Transportation Report 2020-30. https://mdl.mndot.gov/_flysystem/fedora/2023-01/202030.pdf

the previous 2017 evaluation (the first phase of the evaluation study), especially in moderate- to high-volume categories (i.e., more than 10,000 AADT). About 65 MnDOT permanent ATRs (CCS) were used as benchmark sites by StreetLight Data as a training dataset to calibrate the StreetLight Data AADT prediction model. Therefore, the researchers removed all the CCSs used by StreetLight Data model training and evaluated the effect on accuracy of results. They reported that removal of all 65 CCSs did not substantially affect the accuracy results, except in the lowest-volume category of less than 1,000 AADT. The result of their analysis is shown in Table 27.

The researchers recommended that MnDOT consider a phased approach to using probe-based traffic count estimates encompassing the following steps:

1. Continue to maintain MnDOT permanent counter sites
2. Start using probe-based counts for about 90 percent of the moderate- to high-volume roadways (20,000 or more AADT)
3. Continue to use traditional short-term counts at the remaining 10 percent of the moderate- to high-volume roadways as a spot check to ensure that probe-based AADT estimates remain within acceptable tolerances in the next five to ten years
4. Periodically monitor the error of AADT estimates on low- to moderate-volume roadways (less than 20,000 AADT)
5. Once acceptable error tolerances for these lower-volume categories are reached, repeat Step 2 for these lower-volume categories.

Table 27. StreetLight Data AADT Estimates as Compared to Permanent Benchmark Sites (Turner et al., 2020)

AADT Range (vehicle per day)	**Number of sites**	**Absolute Error (%)**			
		Mean	**50th percentile**	**68th percentile**	**95th percentile**
Less than 1,000	10	42%	27%	56%	116%
1,000 to 4,999	24	22%	17%	23%	69%
5,000 to 9,999	12	13%	11%	15%	29%
10,000 to 19,999	17	10%	5%	11%	30%
20,000 to 34,999	35	8%	6%	8%	25%
35,000 to 49,999	39	8%	7%	10%	18%
50,000 to 99,999	200	10%	8%	12%	27%
More than 100,000	105	8%	5%	10%	26%
Total Locations	**442**				

Roll (2019) evaluated StreetLight Data AADT estimates by comparing those estimates with AADT measures from 173 CCSs in Oregon, which had collected data for 2017.[18] MAPE ranged from 7 percent to 55 percent with error diminishing as traffic volume increased. The median error for all sites was 17 percent. For sites with more than 75,000 AADT, the median and mean MAPE were just 3 percent and 7 percent, respectively. Details of the error measurements are shown in Table 28.

Table 28. Absolute Percent Error Summary by Volume Bin (Roll 2019).

Bin	Absolute Percent Error			Count
	Median	Mean	Max	
0-1,000	25%	55%	197%	15
1,000-2,500	24%	34%	109%	15
2,500 - 5,000	32%	32%	64%	21
5,000 - 10,000	21%	24%	98%	31
10,000 - 20,000	10%	16%	47%	33
20,000-30,000	14%	22%	79%	22
30,000 - 50,000	10%	15%	44%	13
50,000 - 75,000	22%	22%	55%	9
75,000 +	3%	7%	49%	14
All Sites	17%	25%	197%	173

The researcher also compared StreetLight Data AADT estimates with short-term AADTs collected in the Bend MPO. The results indicate that accuracy of StreetLight Data was degraded. The median and mean MAPE for short-term AADTs were 32 percent and 68 percent, respectively, as shown in Table 29. However, the researcher noted the lack of precision in the short-term based ‘ground truth’ data sources due to factoring error.

Table 29. ODOT AADT Comparison Absolute Percent Error Summary by Urban Area (Roll 2019)

Functional Classification	Absolute Percent Error			# Sites
	Median	Mean	Maximum	
Other Principal Arterial	13%	22%	111%	21
Minor Arterial	21%	43%	182%	19
Collector	83%	125%	468%	26
All Sites	32%	68%	468%	66

[18] Roll, J. (2019). Evaluating StreetLight Estimates of Annual Average Daily Traffic in Oregon, Oregon Dept. of Transportation and Federal Highway Administration. Report No. OR-RD-19-11, Salem, 2019. https://www.oregon.gov/odot/Programs/ResearchDocuments/StreetlightEvaluation.pdf

Roll (2023) developed a rubric that can be used to evaluate the validity of AADT estimates provided by data vendors.[19] The rubric includes metrics such as accuracy, completeness, validity, timeliness, accessibility, and independence. Their study was focused on the analysis of INRIX AADT estimates in Oregon in the years 2019, 2020, and 2021. The accuracy analysis shows the INRIX AADT estimates perform relatively well to observed data in 2019. However, accuracy of the INRIX AADT estimates degrade quite a bit in the 2020 and 2021 evaluation periods (this could be because of the COVID-19 pandemic). The authors suggested the reason could be use of accuracy benchmarks that are derived from the likely error in short-term AADT estimates. For the other metric of completeness, validity, timeliness, and accessibility INRIX AADT estimates performed to an acceptable standard. In the last measure of data quality, which is independence, they indicated that the evaluation effort is not independent since data used in the validation was used in the development of the INRIX AADT estimation model.

Codjoe et al. (2020) conducted a study to explore non-traditional methods of obtaining vehicle volumes.[20] Their study's primary objective was to validate AADT estimates generated by StreetLight Data and Streetlytics™ in 2020 against ground-truth data provided by Louisiana Department of Transportation and Development (DOTD) and make a recommendation on whether each tool can be adopted by DOTD to provide supplemental traffic volume data for its operations. This study collected ground truth data from continuous counters from 14 permanent stations (AADT), full-month traffic volumes from 30 locations, and 24-hour daily volumes from 60 locations and compared them against corresponding data from StreetLight and Streetlytics. The study evaluates the StreetLight and Streetlytics estimates for accuracy, completeness, validity, timeliness, and accessibility. The results of their accuracy analysis are shown in Table 30.

Table 30. MAPE and percent RMSE of StreetLight and Streetlytics for permanent sites, full-month sites, and 24-hour sites (Codjoe et al., 2020).

Data Type	Mean Absolute Percentage Error (MAPE) %		Percent Root Mean Square Error (%RMSE)	
	StreetLight	Streetlytics	StreetLight	Streetlytics
Permanent Data	18.93	25.55	27.61	24.23
Full-Month Data	93.82	57.17	75.22	75.59
24-Hour Specific Data	70.43	64.33	68.7	74.79
24-Hour Typical Day Data	70.54	59.03	64	69.56
Weighted	70.04	57.68	64.41	68.58

MAPE results for the permanent stations showed StreetLight outperforming Streetlytics.

[19] Roll, J. (2023). Evaluating Third-Party Traffic Volume Data: A Case Study and Proposal for a Data Quality Evaluation Clearinghouse. Presentation at the 102nd Annual Meeting of the Transportation Research Board.

[20] Codjoe, J., Thapa, R., & Yeboah, A. S. (2020). *Exploring Non-Traditional Methods of Obtaining Vehicle Volumes* (No. FHWA/LA. 20/635). Louisiana. Department of Transportation and Development. https://www.ltrc.lsu.edu/pdf/2020/FR_635.pdf

However, MAPE results for both the full-month and 24-hour locations showed Streetlytics outperforming StreetLight. The study found that depending on which accuracy metric was used, different conclusions can be drawn in terms of which data vender provides more accurate AADT estimates. When MAPE was used as the primary metric, one could conclude that Streetlytics AADT estimates are more accurate than StreetLight. On the other hand, if %RMSE was used as the primary metric, one could conclude that StreetLight AADT estimates are more accurate than Streetlytics. The researchers mentioned that there is currently no consensus in the scientific body of knowledge on which metric is better between MAPE and %RMSE. Agencies generally choose one as a matter of preference and not as an indication of one being superior to the other. While MAPE gives a uniform assessment of the errors, %RMSE applies bigger weights to bigger errors in the data and hence may give a different result and conclusions than MAPE. Streetlytics outperforms StreetLight in estimating for roadways with volumes under 300 vpd, while StreetLight generally outperforms Streetlytics for roadways with volumes between 300 to 500 vpd. However, data from both tools were determined to be valid for use for traffic assessments. For accuracy, completeness, and validity, even though Streetlytics generally scored higher than StreetLight, both tools had acceptable score ranges.

Yang et al. (2023) presented a big-data driven framework that analyzes Mobile Device Location Data (MDLD) and estimates vehicle volume for all roadway segments.[21] First, a series of cloud-based computational algorithms was applied, including but not limited to a trip and tour identification algorithm to mine travel behavior information and a travel mode imputation model that impute multimodal trajectories from MDLD. A map matching and routing algorithm was then applied to snap and route vehicle trajectories to the roadway network. The observed vehicle counts on each roadway segment were weighted to match the VMT by county, urban/rural status, and functional classes. Further, a random forest regression model was used to calibrate the weighted vehicle volume against the AADT acquired from loop detectors. The proposed framework was implemented on the all-street network in Maryland using MDLD data for the entire year of 2019. After weighing and calibration processes with AVMT, high correlation and low RMSE values were observed between the vehicle volume estimates and the ground truth data. The researchers used random forest regression to calibrate the probe vehicle volume weighted by AVMT against the AADT to obtain the final vehicle volume. During the calibration process, a 10-fold cross-validation (CV) process was used to fine-tune the random forest regression hyperparameters with 90 percent training data. The fine-tuned models were then applied to the 10 percent testing data.

Yang et al. (2023) used the Smart Location Database (SLD), which is a nationwide geographic data resource for measuring location efficiency, as calibration variables as features in the random forest regression to calibrate the weighted vehicle volume to account for the effects of the built environment. The SLD variables used in this study include:

- TotEMP = total employment
- Pct_AO0 = percent of zero-car households
- D1A = gross residential density (housing units per acre) on unprotected land

[21] Yang, Mofeng, Weiyu Luo, Mohammad Ashoori, Jina Mahmoudi, Chenfeng Xiong, Jiawei Lu, Guangchen Zhao, Saeed Saleh Namedi, Songhua Hu, and Aliakbar Kabiri. "A Big-Data Driven Framework to Estimating Vehicle Volume based on Mobile Device Location Data." *arXiv preprint arXiv:2301.08660* (2023). https://arxiv.org/ftp/arxiv/papers/2301/2301.08660.pdf

- D1C = gross employment density (jobs per acre) on unprotected land
- D3AAO = network density in terms of facility miles of auto-oriented links per square miles
- D3B = street intersection density (weighted, auto-oriented intersections eliminated)
- D5AR = jobs within 45 minutes auto travel time, time decay (network travel time) weighted

The results of the study by Yang et al. (2023) are shown in Table 31 and Table 32. The Pearson correlation value and the Root Mean Square Error (RMSE) for the training dataset between the weighted vehicle volume and the ground truth AADT are 0.746 and 7,912, respectively. These values are improved to 0.966 and 2,996 after calibration. Similarly, for the testing set, the Pearson correlation and RMSE are improved from 0.764 and 7,548, to 0.854 and 5,701 respectively after calibration. For all link types, a good correlation (i.e., over 0.80, as denoted in italics) can be observed between the calibrated vehicle volume and the ground truth AADT, except for Local Roads and Highway Ramps in the testing set.

Table 31. Volume Calibration Results Comparison by Link Type (Yang et al., 2023)

Link Type	Training Set				Testing Set			
	Corr.		RMSE		Corr.		RMSE	
	Before	After	Before	After	Before	After	Before	After
All	0.746	*0.966*	7912	2996	0.764	*0.854*	7548	5701
Interstate Highways and Other Highways	0.752	*0.975*	20081	6559	0.712	0.775	19633	15246
Primary Roads	0.699	*0.971*	7909	2695	0.721	*0.846*	8665	6509
Secondary Roads	0.627	*0.960*	4899	1776	0.617	*0.813*	3667	2667
Tertiary Roads	0.414	*0.959*	3486	994	0.511	*0.869*	3090	1877
Local Roads	0.374	*0.944*	2474	853	0.426	0.742	1701	1083
Highway Ramps	0.242	*0.866*	10426	4722	0.182	0.402	9119	6846

Table 32. Volume Calibration Results by Urban/Rural Status (Yang et al., 2023)

Link Type	Training Set				Testing Set			
	Corr.		RMSE		Corr.		RMSE	
	Before	After	Before	After	Before	After	Before	After
All	0.746	*0.966*	7912	2996	0.764	*0.854*	7548	5701
Rural	0.769	*0.967*	3583	1442	0.727	*0.826*	4810	4075
Urban	0.738	*0.964*	8913	3363	0.764	*0.853*	8311	6179

Predicting traffic volume on low-volume roadways is challenging, and only a few studies have been conducted to address this. A recent study by Yeboah et al. (2023) focused on estimating AADT on low-volume roads (less than 500 vehicles per day) in Louisiana.[22] The objective of their study was to find practical, cost-effective, and progressive methods of estimating and classifying traffic on low-volume rural roadways. They selected 395 low-volume locations in Louisiana and extracted roadway, demographic, and socioeconomic information for each location, which included functional class, land use, number of lanes, population density, median age, median household income, and household density. Using this information, they developed two low-volume roadway AADT prediction models: linear regression and random forest regression models. The results indicated that the linear regression model had the highest predictive accuracy, with R^2 of 0.979 and root mean square error (RMSE) of 70.26. The functional class, land use, number of lanes, population density, median age, median household income, and household density were found to be significantly affecting the AADT prediction in both the linear regression and random forest regression models.

A study by Tsapakis et al. (2020) evaluated accuracy of AADT estimates from StreetLight data at 4,643 count sites, of which 35 were permanent stations and the remaining 4,608 were short-term counts.[23] All sites were in Texas. The analysis included two study areas: (a) at Texas-Mexico border crossings (ports of entries), and (b) on counted Texas roadways that are in proximity to the Mexican borders (count locations in border regions – districts of El Paso, Laredo, and Pharr). StreetLight Data provided the researchers with unscaled and uncalibrated mobile device count data for commercial and privately owned vehicles, as well as probe based AADT estimates for several locations in the two study areas. The results of this study are summarized in Table 33 and Table 34. The main findings were:

1. Penetration rates of mobile devices in the two study areas were found to be 1.06 percent at the first study area (i.e., at point of entries) and 0.86 percent at the second study area

[22] Yeboah, A. S., Codjoe, J., & Thapa, R. (2023). Estimating average daily traffic on low-volume roadways in Louisiana. *Transportation research record*, 03611981221106166.

[23] Tsapakis, I., Cornejo, L., & Sánchez, A. (2020). Accuracy of probe-based Annual Average daily Traffic (AADT) Estimates in Border Regions. *Center for International Intelligent Transportation Research*. https://static.tti.tamu.edu/tti.tamu.edu/documents/TTI-2020-1.pdf

(count location in border regions). Interestingly, the penetration rate was higher on rural areas compared to urban areas in all three TxDOT districts.

2. Commercial vehicles trips usually used a global positioning system (GPS) and the penetration rate was 8.7 percent. This is followed by the penetration rate of location-based services (LBS) privately owned vehicle trips, which was found to be 0.85 percent. For GPS-based data for privately owned vehicle trips made a penetration rate of 0.03 percent.
3. In terms of the accuracy of probe-based AADT estimates:
   - In the first study area, the AADT estimation MAPE was 33.0 percent. The mean signed difference (MSD), mean absolute difference (MAD), and average coefficient of variation (ACV) were −2,528 vehicles per day (vpd), 2,806 vpd, and 28.3 percent, respectively. Linear and non-linear fitted lines between observed AADT values and StreetLight Data AADT estimates had R^2 values of 0.8207 and 0.8572, respectively.
   - In the second study area, the MAPE was 50 percent, which is lower than the corresponding MAPE (61 percent) reported in a 2017 study that evaluated 2015 AADT estimates developed by StreetLight Data. The MSD, MAD, and ACV for the second study area were −68 vpd, 2,345 vpd, and 25 percent, respectively.
4. In terms of accuracy of the probe-based AADT estimates by facility type and direction of error:
   - The accuracy of estimated AADT values gradually improved from low-volume roads to high-volume roads. The AADT estimates were higher than TxDOT AADT values within the two lower traffic volume ranges (401–5,000 and 5,001–10,000 vpd), but this trend was reversed in the case of higher-volume roads (10,001–20,000, 20,001–50,000, >50,000 vehicles/day).
   - The AADT estimates for urban roads were more accurate (MAPE = 47 percent) than those for rural roads (MAPE = 63 percent). The MSD, MAD, MAPE, and ACV for urban roads were -126 vph, 2,548 vph, 47 percent, and 24 percent, respectively. Similarly, the MSD, MAD, MAPE, and ACV for rural roads were 214 vph, 1,349 vph, 63 percent, and 28 percent, respectively.

Table 33. Penetration Rate and Accuracy of StreetLight Data AADT Estimates by AADT Range and Rural/Urban Designation.

Functional Class	Rural/ Urban	Counts	Penetration Rate	MSD	MAD	MAPE	ACV
1	Rural	14	2.87%	-2,581	3,082	26%	16%
	Urban	32	1.59%	-7,763	10,381	31%	20%
2	Rural	2	2.84%	2,778	2,778	52%	28%
	Urban	19	0.85%	-5,658	6,605	22%	18%
3	Rural	227	1.66%	-423	2,186	39%	24%
	Urban	1,052	0.60%	-2,109	3,903	27%	18%
4	Rural	160	1.75%	379	1,423	69%	32%
	Urban	799	0.58%	-381	2,039	32%	19%
5	Rural	332	1.76%	616	752	76%	30%
	Urban	1,219	0.82%	1,511	2,341	54%	26%
6	Rural	43	1.50%	613	734	79%	35%
	Urban	45	0.39%	371	932	33%	20%
7	Rural	8	0.45%	435	811	86%	33%
	Urban	691	0.67%	772	1,070	83%	37%
Grand Total		**4,643**	**0.86%**	**-68**	**2,345**	**50%**	**25%**

Table 34. Penetration Rate and Accuracy of StreetLight Data AADT Estimates by Roadway Functional Class and Rural/Urban Designation

AADT Range (vpd)	Rural/ Urban	Counts	PR	MSD	MAD	MAPE	ACV
401-5,000	Rural	633	1.77%	753	905	72%	31%
	Urban	1,767	0.77%	949	1,145	71%	31%
5001-10,000	Rural	97	1.63%	307	1,418	21%	15%
	Urban	760	0.65%	810	2,269	31%	19%
10,001-20,000	Rural	36	1.30%	-5,027	5,220	34%	31%
	Urban	825	0.58%	-352	3,356	24%	17%
20,001-50,000	Rural	19	1.52%	-7,988	8,260	37%	33%
	Urban	482	0.58%	-4,549	6,195	22%	18%
>50,000	Rural	1	0.84%	-5,009	5,009	8%	6%
	Urban	23	0.74%	-12,816	14,164	14%	11%
Grand Total		**4,643**	**0.86%**	**-68**	**2,345**	**50%**	**25%**

Baffoe-Twum et al. (2022) conducted a meta-analysis study to identify and to evaluate the performance, the sources of error, and possible advantages and disadvantages of the techniques utilized most for estimating AADT on low-volume roads.[24] They found that many

[24] Baffoe-Twum, E., Asa, E., & Awuku, B. (2022). Estimation of annual average daily traffic (AADT) data for low-volume roads: a systematic literature review and meta-analysis. *Emerald Open Research*, *4*, 13.

methods were used, including artificial neural networks, traditional factor approach, regression methods, geographical information system-based, smoothly clipped absolute deviation (SCAD) penalty, satellite-based Imagery, travel-demand modeling method, synthetic minority oversampling technique (SMOTE), generalized linear mixed model (GLMM), kriging (Geostatistics), inverse distance weighting, natural neighbor (NN) and trend techniques, and random forest. Of all these methods, the regression method was utilized the most, followed by the artificial neural network method.

MEASUREMENT ERRORS IN GROUND TRUTH AADTS OBTAINED FROM NON-TRADITIONAL SOURCES

When evaluating the accuracy of AADT estimates obtained from non-traditional sources, the AADT estimates are usually compared with AADT values obtained from traditional sources (which are considered to be the ground truth AADTs). However, the issue is that the AADT values obtained from traditional sources can contain counting error and bias. Therefore, there is a lack of certainty in terms of what is utilized as the ground truth AADT values. Below are some studies that evaluated the accuracy of traffic counts obtained from different traditional sources.

- A study that evaluated the accuracy of pneumatic road tube counters found that although the average counting error in a daily traffic count might be near zero, the absolute error of a typical 15-minute count averaged closer to ten percent.[25] The study showed that the counting inaccuracy is being masked by the positive and negative counting errors canceling each other out.
- A study compared the vehicle counting performance of an automatic traffic counter by Sierzega (SR4 radar) and manual traffic counting at two different intersection approaches, each from about 6:00 AM to 6:00 PM.[26] At the first location, the automatic traffic counter counted 10.2 percent more vehicles than what was manually counted. At the second location, the automatic counter counted 5.7 percent more vehicles than what was manually counted.
- A study that evaluated the accuracy of the MetroCount 5600, i.e., an automatic pneumatic-tube-based traffic count device, in mixed traffic found that the volume of traffic measured by the automatic traffic counter was on average 7.6 percent lower (consistently lower) than the total count obtained from video recording playbacks.[27] The study highlighted that MetroCount 5600 failed to register the present of overtaking vehicle when both overtaking and overtaken vehicles were exactly at the detection point, leading to traffic counts lower than the actual number of vehicles that passed the detection point.
- A study that compared the performance of Icoms TMSA-SA radar traffic counter with

---

[25] McGowen P, Sanderson M. (2011). Accuracy of pneumatic road tube counters. In: Institute of Transportation Engineers Western District Annual Meeting; pp. 1-17.

[26] Puan, O. C., Nor, N. S. M., Mashros, N., & Hainin, M. R. (2019). Applicability of an automatic pneumatic–tube–based traffic counting device for collecting data under mixed traffic. In *IOP Conference Series: Earth and Environmental Science* (Vol. 365, No. 1, p. 012032). IOP Publishing.

[27] Puan, O. C., Nor, N. S. M., Mashros, N., & Hainin, M. R. (2019). Applicability of an automatic pneumatic–tube–based traffic counting device for collecting data under mixed traffic. In *IOP Conference Series: Earth and Environmental Science* (Vol. 365, No. 1, p. 012032). IOP Publishing.

loop detectors indicated that the radar counter was undercounting by 4 percent to 6 percent at a congested test site.[28] Slow moving and stopped traffic during peak hours negatively affected the counting accuracy of the radar counter. At non-congested sites, the counting error of the Icoms TMSA-SA radar traffic counter was just less than 2%.

- A study that evaluated video-based vehicle counting using YoloV3 (real-time object detection algorithm in videos) demonstrated that the overall accuracy of the vehicle counts obtained can reach more than 90%.[29]
- A study compared a variety of detector devices and technologies, including inductive loop, magnetic, pneumatic road tube, active infrared, passive infrared, microwave radar, ultrasonic, passive acoustic, and Video Image Processing.[30] The study documented the accuracy of these traffic counting technologies in different environmental and traffic conditions. Traffic count accuracy ranged from 1% to more than 10% depending on technology and device type.

As demonstrated in these studies, the nature of errors in traditional traffic counters are diverse. Some devices consistently overcount or undercount traffic while other devices exhibit a mix of both error types. The effect of counting error by vehicle class can be amplified because miscounting of vehicles affects the count in multiple classes, e.g., undercounting in one class of vehicles is always accompanied by overcounting in the other class of vehicles. Therefore, agencies must consider the uncertainties in the ground truth AADT measurements when evaluating AADT estimates from non-traditional sources.

---

[28] CT-Technologies (2020). Compare a Radar Traffic Counter to an Inductive Loop Counter. Available online at https://ct-technologyinfo.com/2020/02/10/comparing-a-radar-traffic-counter-to-an-inductive-loop-counter/, last visited in June 2023.

[29] Dai, Z., Song, H., Wang, X., Fang, Y., Yun, X., Zhang, Z., & Li, H. (2019). Video-based vehicle counting framework. *IEEE Access*, *7*, 64460-64470.

[30] Martin, P. T., Feng, Y., & Wang, X. (2003). *Detector technology evaluation* (No. MPC Report No. 03-154). Fargo, ND, USA: Mountain-Plains Consortium. Available online at https://www.ugpti.org/resources/reports/downloads/mpc03-154.pdf, last accessed in June 2023.

APPENDIX E: FINDINGS OF TECHNICAL ADVISORY COMMITTEE TPF-5(384) IN-PERSON INFORMATION EXCHANGE SESSIONS

INTRODUCTION

This report was first produced in draft form in August of 2023 and disseminated to the Technical Advisory Committee TPF-5(384), "Exploring Non-Traditional Methods to Obtain Vehicle Class and Volume Data." It was used, along with a set of PowerPoint slides, to carry out two in-person information exchange sessions in 2023; the first in Denver, Colorado, on Thursday, August 24, 2023, at Colorado Department of Transportation and the second in Columbus, Ohio, on Wednesday, September 13, 2023, at Battelle Memorial Institute. Invitees included one participant from each state in the TAC; Alaska, California, Colorado, Georgia, Idaho, Illinois, Maryland, Minnesota, Nebraska, New Jersey, North Carolina, North Dakota, Ohio, Oregon, Pennsylvania, South Carolina, Texas, and Virginia. Other participants included the Federal Highway Administration leads and the contractor instructors who also authored this document. Additional individuals from state agencies both in the TAC and not in the TAC were able to attend remotely. In total, 24 different states participated in the information exchange sessions.

The current final guidelines represent the original inputs to the process as well as edits and changes because of the in-person information exchanges. These were changes that were recognized by FHWA or the TAC members. An important objective of the information exchange sessions was to allow state agencies to identify what non-traditional data sources they have encountered and to provide their perspective on what barriers there are to the acceptance and inclusion of these data as well as to identify any success stories they have had in using them. This Appendix summarizes key findings of that discussion.

KEY FINDINGS OF 8/24/2023 AND 9/13/2023 INFORMATION EXCHANGE SESSIONS

Most states indicated agreement with the objectives of using probe-based data to supplement or replace traditional count-based estimates. They were most likely to identify Streetlight Data, INRIX, and Replica as the data sources familiar to them.

- Many TAC states indicated they are NOT currently using probe-data products and are still relying on a traditional count program.
- Some states have tried unsuccessfully to get reliable AADT data from vendors. A common challenge has been that probe data has been evaluated but has too much variability and does not accurately match known traffic volumes, especially for lower-volume roadways. The challenges around matching geocoding between states and vendor products, the non-reproducible nature of the dynamic models, and the lack of reliable HPMS measurements for class data were all identified as drawbacks to the currently available products.
- Some states have these data products available within the agency but have not used them for their traffic count programs. Commonly, traffic modelers use them. Having the products available within the agency may put pressure for the count programs to use them to justify the costs of the products.

The guidelines provided in this report were generally received positively. A significant barrier to previous evaluations was the limitation of how many CCSs were available in each state, how representative they were of the totality of roadways for which probe-based estimates might be desired, and whether the public availability of these data constituted a bias risk of the models being tuned to known AADT values. Even with the current heavy advantage of the big data models, no vendor has been fully successful in delivering estimates of similar uncertainty to the current count-based programs, although it appears that the models are improving. The guidelines in this document include an important expansion of the method to evaluate the data products using the far more prevalent portable counts factored to AADT. Such sites outnumber CCS by as much as 100 to 1 in most states, better represent all roadways, and are more likely to be unavailable to the vendor to develop their models. States will need to begin using these guidelines to determine how well the probe-based data models are improving.

The guidelines were acknowledged by the participants to be adequately robust. The detailed nature of all the issues did result in requests for a simplified step-by-step process beyond even the checklist approach provided.

U.S. Department of Transportation
Federal Highway Administration

U.S. Department of Transportation
Federal Highway Administration
Office of Highway Policy Information
1200 New Jersey Ave., SE Washington, D.C. 20590
https://www.fhwa.dot.gov/policyinformation
August 2024
Publication No. FHWA-HPL-23-018