

Leveraging Probe Data for HPMS Volume Submissions

A guide for using probe-based volume: Lessons learned from the Eastern Transportation Coalition Transportation Data Marketplace validation program

Final Report

10/17/2026 Rev 27

Prepared by: Stanley Young and Zachary Vander Laan

The report provides practical guidance for the use of probe-based volume data in HPMS submittals, aligned with current federal standards.

Acknowledgements

The project team would like to thank the TDM Leadership Team for their review and feedback, the vendors that volunteered their time and expertise, and various experts that reviewed the report and offered feedback and insight.

Table of Contents

Acknowledgements **2**

List of Acronyms **4**

Executive Summary **5**

1.0 Introduction **7**

1.1 The History of Probe-Based Volume Estimates **7**

1.2 Organization of the Guidebook **8**

2.0 Background **10**

2.1 Volume Inferences from Probe-Based Travel Time and Speed Data ~2012 . **10**

2.2 Initial Research in Viability and Requirements of Probe-Data to Estimate Volumes 2015-2018 ..... **111**

2.3 TETC Hurricane Proof of Concept Demonstration - 2020 ..... **114**

2.4 FHWA Consideration of Non-Traditional Methods for Reporting AADT **115**

2.5 The Eastern Transportation Coalition Transportation Data Marketplace – 2022 ..... **116**

3.0 Summarized Case Studies of Industry Capacity ..... Error! Bookmark not defined.

3.1 Early Exercises **10**

3.2 North Carolina Volume Validation (Nov 2023) **118**

3.3 Georgia Volume Validation (2024) **119**

3.4 Massachusetts Volume Validation (2025) **119**

3.5 Minnesota DOT Case Study (2021) **20**

4.0 Guidelines for Qualifying Probe-Based Volume Estimates for HPMS ..... **2**Error! Bookmark not defined.

4.1 Planning **22**

4.2 Testing Preparation **26**

4.3 Evaluation of Vendor Data **28**

4.4 Ongoing Monitoring and Re-Evaluation **30**

5.0 Issues for Future Attention **32**

APPENDIX A: Error Metrics used in Probe-Based Volume Estimates Validation A-1

APPENDIX B: Summary of FHWA Method **B-1**

APPENDIX C: North Carolina Volume Validation (2023) **C-1**

APPENDIX D: Georgia Volume Validation (2024) **D-1**

APPENDIX E: Massachusetts Volume Validation (2025) **E-1**

APPENDIX F: Minnesota DOT Case Study (2021) **F-1**

List of Acronyms

AADT – Annual Average Daily Traffic
ADT – Average Daily Traffic
AHDT – Average Hourly Daily Traffic
CCS – Continuous Count Station
CVD – Connected Vehicle Data
DOT – Department of Transportation
FDE – Fundamental Data Elements
FDOT – Florida Department of Transportation
FHWA – Federal Highway Administration
HPMS – Highway Performance Monitoring System
HSIP – Highway Safety Improvement Program
MAP-21 – Moving Ahead for Progress in the 21st Century Act
MCOMP - Multistate Corridor Operations and Management Program
MIRE – Model Inventory of Roadway Elements
NFAS – Non-Federal Aid System
REVEAL - Real-Time Volume Estimates from Probe Traffic Data
STC – Short Term Counter
TDM – Transportation Data Marketplace
TETC – The Eastern Transportation Coalition
TSMO - Transportation Systems Management and Operations

Executive Summary

This guidebook provides practical guidance for using emerging probe-based volume estimates for reporting Average Annual Daily Traffic (AADT) measures within the Highway Performance Monitoring System (HPMS) — as well as other applications within a department of transportation (DOT) in need of network-wide traffic volume measures. This project was inspired by the need to meet federal reporting guidelines to expand volume measures in the form of AADT to the entire roadway system, commonly referred to as the MIRE reporting requirements that are scheduled to go into effect in 2026.

Meeting these heightened reporting requirements using traditional continuous count stations (typically inductive loop) and temporary counts (typically tube-based counts for 24 to 48 hours) to collect the base data would greatly increase the fiscal burden on DOTs. Probe-based volume estimates provide a path to meet these reporting requirements through a method that is significantly less expensive, while also providing a rich data set for other DOT planning and operations purposes.

This report provides a thorough backdrop of the technology and policies behind the utilization of probe-based volume estimates. It traces the development of the technology and provides a practical guide for harnessing it for use in HPMS reporting and other applications. Specifically, this SCOOP initiative seeks to accomplish the following objectives:

- **Review Federal Requirements and Industry Capabilities:** This guide provides both an analysis and a practical interpretation of USDOT's established accuracy thresholds and statistical testing protocols for probe-based volume data for use in HPMS reporting. These protocols were used in the TETC's validation efforts from 2022 to the present, reflecting industry capabilities to meet these standards. Validation activities hosted in North Carolina, Georgia, and Massachusetts from 2022 through 2026 provide a benchmark for industry capability to provide scalable volume measures with an emphasis on AADT reporting for HPMS purposes.
- **Engage Stakeholders and Coordinate with USDOT:** This guide encompasses input from Coalition member agencies to understand implementation challenges and opportunities. Much of this engagement was through the TDM Technical Advisory Committee, as well as direct briefings and exchanges with member agencies that hosted validation activities from 2022 through 2026. This also includes coordination with the USDOT Office of Highway Policy Information to ensure alignment with federal expectations and policy direction.
- **Document Case Studies:** This guide encompasses two types of case studies. The first is a summary of the three major validation activities conducted by the TETC TDM validation program spanning North Carolina, Georgia, and Massachusetts, alongside the lessons learned with respect to assessing probe-based volume accuracy when compared to agency counts. The second type features examples of states that have submitted probe-based volume data for HPMS reporting purposes. Combined, these case studies highlight lessons learned, methodologies used, and outcomes achieved.

- **Develop Guidance and Provide Technical Assistance:** The findings and lessons learned from the above are synthesized into a comprehensive “How-to” procedural guide for state DOTs, detailing recommended practices for assessing probe data for use in HPMS reporting. This report and the corresponding workshop are intended to share guidance, demonstrate methodologies, and support agency readiness.
- **Identify Issues for Future Attention:** While assembling this guide, assessing industry maturity, and reviewing case studies, several issues for future attention were identified that require additional analysis, research and development, and testing. A summary of these issues is provided in this last section.

This report is the final deliverable providing guidance to Coalition members: a 'How-to' guide for state DOTs detailing best practices for acquiring, testing, and utilizing probe-based volume estimates in HPMS reporting as well as other similar applications. For the sake of brevity, much of the technical detail, particularly the detail associated with probe-based volume background and case studies, are provided in appendices. The material is summarized in the body of the report, and the reader referenced to the appendices for more in depth review.

1.0 Introduction

The Eastern Transportation Coalition (TETC) has been highly active in developing probe-based traffic volume estimation technology since the early 2010s. Since 2022, TETC has evaluated vendor volume estimation products in North Carolina, Georgia, and, most recently, Massachusetts. These evaluations focused on accuracy, consistency, and the vendors' ability to meet published federal quality thresholds for HPMS reporting. Each validation effort has deepened the understanding of industry capacity to deliver accurate volume estimates, as well as the accuracy assessment process recommended by the USDOT.

1.1 The History of Probe-Based Volume Estimates

Around 2010, the use of probe-based travel time and speed data began scaling across multiple states. During this time, transportation systems management and operations (TSMO) professionals noticed that industry data feeds were revealing large-scale shifts in travel behavior during major weather events, such as blizzards and hurricanes. By combining speed and travel time data with accompanying confidence scores available in probe-based travel time and speed data, traffic professionals could infer roadway traffic volumes to a baseline level of accuracy.

At the time, confidence scores used three categorical values to indicate whether sufficient probe data was available to accurately discern travel times and speeds. A confidence score of '10' meant the reported speed or travel time was derived assuming free-flow conditions. This inferred that traffic was too low to generate probe data and vehicular volume was insufficient to impede traffic flow. A score of '30' indicated sufficient probe data, meaning travel time and speed estimates were based entirely on observed vehicle probes. A score of '20' represented a hybrid, relying on a combination of inferred conditions and a limited number of probe observations.

This ability to estimate basic traffic volume thresholds from confidence scores and reported speeds sparked the idea of formally estimating volumes using probe vehicle data. Although probe data penetration rates were estimated at 2% or lower at the time, the volume of source data was steadily growing and doubling roughly every 18 months. In 2013, TETC proposed a research project via the Multistate Corridor Operations and Management Program (MCOMP) to determine the feasibility and potential accuracy of probe-based volume estimation. This research grant helped establish the foundational statistical science—rooted in emerging artificial intelligence and machine learning technologies—that underpins many of today's volume estimation techniques.

Building on this, TETC funded a 2020 project to demonstrate these methods at scale across five southeastern states. This pilot utilized emerging connected vehicle data from vehicle original equipment manufacturers (OEMs), marketed through Wejo. The demonstration provided definitive proof that probe data could successfully monitor traffic volumes during major events like hurricanes, though it also highlighted significant big data management challenges that still needed to be addressed.

More recently, new Model Inventory of Roadway Elements (MIRE) reporting requirements for DOTs prompted discussions on alternative ways to estimate Annual Average Daily Traffic (AADT) for federal reporting. This led to a USDOT pooled-fund

initiative investigating 'non-traditional methods' for AADT assessment, ultimately spurring the development of new industry products tailored for planning-level volume measures. In 2022, the Coalition included probe-based volume measures in its Transportation Data Marketplace (TDM) request for proposals, resulting in over five firms currently offering planning-level volume estimates derived from vehicle probe data. Since then, the TDM validation program has continually assessed the quality of this vendor-supplied data through annual exercises hosted by various Coalition states.

This guidebook culminates and summarizes the lessons learned throughout the multi-year development of probe-based volume estimates. It examines the industry's current capacity to provide planning-level volume data for the Highway Performance Monitoring System (HPMS) and offers guidance on procuring, quality-assessing, and utilizing probe-based volume estimates for HPMS and other planning applications.

1.2 Organization of the Guidebook

This guidebook is organized into four sections which can be accessed and reviewed independently.

- **Background** - The report begins by tracing the key advancements contributing to the development of probe-based volume estimates both for non-real-time applications (referred to as planning-level volume estimates) and real-time transportation system management and operations (TSMO) applications—the latter of which has yet to be realized. This provides a backdrop, informing the reader of the basic technology, lexicon, and key concepts. Readers familiar with probe-based volume estimates may skip this section. Those not familiar with the history of probe-based volume estimates are encouraged to review this section to become familiar with the timeline and key technical issues.
- **Case Studies** - The guidebook then provides evidence from validation activities by the Coalition regarding current industry capacity to report AADT and other volume metrics such as average daily traffic (ADT), average hourly daily traffic (AHDT), daily volume estimates, and hourly volume estimates. In short, the validation results show gradual improvement and maturity of industry volume products in providing accurate probe-based volume data. The most recent validation (that of Massachusetts) reveals that the majority of vendors tested (three out of five) either passed or very nearly passed all seven quality tests advocated by the USDOT. However, the validation also revealed several issues in the use and integration of probe-based volume estimates that need to be well understood to achieve intended, quality results. In addition to TETC Volume validation studies, other state initiatives to use probe-based volume estimates for HPMS reporting are also summarized. This section highlights the findings of each case study, with a more in-depth summary provided in appendices. The finding from these case studies are later referenced in the procedural guidance section.
- **Federal Guidelines** - A review of the federal guidelines for the acceptable use of non-traditional methods of volume estimation is provided. The root of these guidelines are statistical tests comparing vendor-provided estimates against the expected accuracy of temporary volume counts. Understanding the thought and

reasoning behind these federal guidelines adds further insight into the state of the industry's ability to provide acceptable volume estimates.

- **Procedural Guidance for Validating Probe-Based Volume Estimates for HPMS** - This section provides procedural guidance for any agency to qualify industry volume estimates for HPMS MIRE reporting purposes, as well as for understanding the overall quality of probe-based data and its use in various applications. The procedural guidelines were adapted from the TETC validation procedures refined since 2022 through yearly validation exercises.

As a result of the findings over the last three years of validation testing and close coordination with industry and USDOT, the Coalition made the following general observations regarding the state of the industry in early 2026, prior to authoring this guidance document:

- Volume estimations for planning measures such as AADT, ADT, and AHDT are available from multiple vendors, and the quality is approaching a high degree of maturity based on TETC validation results. These products address the market need for additional AADT volume measures that span the entirety of the roadway network, assisting DOTs in fulfilling the MIRE reporting requirements which continue to drive investment by the industry.
- While volume estimation technology has matured (incorporating various modern AI techniques), several issues impact data accuracy beyond the fidelity of the vendor-provided data. These additional sources of error need to be managed in order to achieve quality results for AADT reporting as well as other applications of volumes estimates. These additional sources of error include location accuracy of existing sensors, location referencing (i.e., conflation) that may introduce errors, and emerging video-based methods of volume sensing whose accuracy is not fully understood. In short, many sources of error can impact the efficacy of using probe-based volume estimates for various agency purposes.
- The process to qualify non-traditional methods of volume estimation is conceptually straightforward, but its implementation can be complex. This report provides a thorough review of the concepts and recommendations for implementation when qualifying probe-based volume estimates for HPMS MIRE reporting purposes.
- The probe-based volume estimate market, and its underlying data science technology, is mature enough to begin using for HPMS reporting and other applications requiring scalable volume estimates; HOWEVER, error can be introduced at several points independent of the vendor’s core process or technology, and these error sources need to be well understood and mitigated in order to obtain the best possible results.

This guidebook provides the Coalition’s best recommendations for states seeking to leverage probe-based volumes for HPMS and other purposes. Though emphasis is placed on AADT measures, other measures such as ADT, AHDT, Daily, and Hourly volumes are addressed, both as products to leverage from the transportation market and as conduits to provide insight into the quality of industry volume products.

2.0 Background

The development of probe-based volume data began in the early 2010s, paralleling the rapid maturation of probe-based travel time and speed data. Although initial vehicle probe penetration rates hovered around just 2%, data observations doubled roughly every 18 months as the number of probe vehicles and their reporting frequencies increased. During this period, Transportation Systems Management and Operations (TSMO) practitioners began leveraging travel time and speed data for real-time applications—such as changeable message signs, congestion awareness, and incident detection—demonstrating that network-wide traffic monitoring was feasible without physical sensors.

Probe data also proved highly effective for monitoring the impacts of major weather events, like blizzards and hurricanes, across multiple states. The use of travel time and speed data quickly expanded beyond real-time operations into performance metrics and broader planning initiatives. The data's scalability, combined with the ability to crudely infer volumes using fundamental speed-flow relationships and data confidence scores, provided the first indications that probe data could eventually produce accurate roadway volume estimates.

2.1 Volume Inferences from Probe-Based Travel Time and Speed Data ~2012

The Transportation Data Marketplace (TDM) requires vendors to include a confidence score alongside travel time and speed measures. This score indicates whether estimates are based on sufficient real-time probe data (high confidence) or inferred from historical free-flow conditions due to low traffic volumes (low confidence). Initially, scores were categorical (e.g., 10, 20, 30), while later TDM phases adopted linear scales (e.g., 0–1 or 0–100). Across all variations, the core concept remains the same: higher scores indicate a higher volume of observed probes.

Under normal conditions, confidence scores mirror daily traffic patterns, remaining high during the day and dropping overnight (as shown in **Figure 2.1.1**).

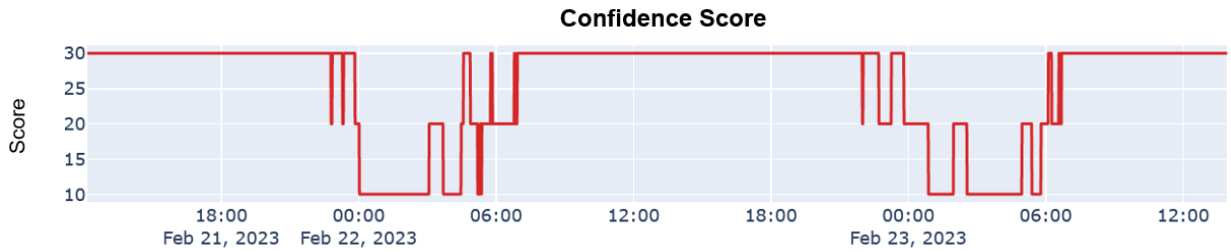

Figure 2.1.1: A 24-hour plot of confidence scores that accompanies travel time and speed data within the TDM.

However, practitioners soon realized these scores could serve as a pseudo-volume indicator during major disruptions. During severe winter storms or hurricanes, plummeting traffic volumes consistently resulted in dropping confidence scores. This correlation allowed agencies to monitor the operational impact and recovery of the

roadway network on a massive scale. An example of this behavior during a winter weather event in Maryland is shown in **Figure 2.1.2**.

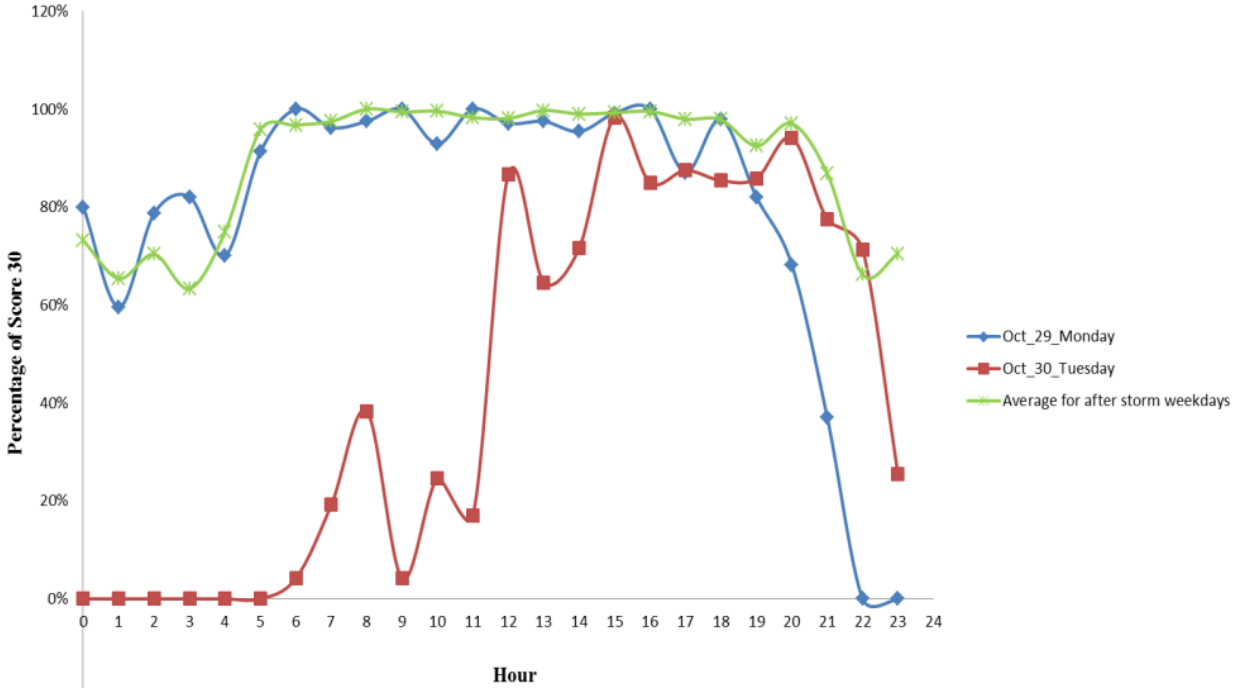

Figure 1.1.2**:** Example of the behavior of the confidence score during a major winter event on I-68 in Maryland[1].

The realization that dropping confidence scores correlated directly with drops in traffic volume prompted researchers to investigate whether probe data could accurately estimate traffic volumes under all conditions. This insight initiated the foundational research into dedicated probe-based volume estimation.

2.2 Initial Research in Viability and Requirements of Probe-Data to Estimate Volumes 2015-2018

One of the earliest research initiatives was funded by a 2013 MCOMP grant awarded in 2015. At that time, industry probe data providers began reporting the number of probe observations within a given reporting period. As probe samples increased, the accuracy of the inferred volume naturally increased. Although probe penetration rates remained low (<15%) and highly variable, a clear path to estimating volumes began to emerge.

The MCOMP-funded research led to a 2018 ITS World Congress publication[2] characterizing the feasibility of this process. The paper explored using commercial probe data combined with validated continuous counter data to generate accurate, ubiquitous volume measures across the highway network. This approach aimed to

[1] https://cattworks.org/wp-content/uploads/2015/08/winter-weather-analysis-elham-edit.pdf

[2] Young, Stanley E., Kaveh Sadabadi, Przemysław Sekuła, Yi Hou, and Denise Markow. 2018. "Estimating Highway Volumes Using Vehicle Probe Data – Proof of Concept: Preprint." Golden, CO: National Renewable Energy Laboratory https://www.nrel.gov/docs/fy18osti/70938.pdf

overcome the limitations of sparse, location-specific Annual Average Daily Traffic (AADT) measures derived from the Highway Performance Monitoring System (HPMS). The study examined fundamental accuracy limitations and presented early proof-of-concept results showing that calibrated probe data could successfully meet user expectations for accuracy.

While initially focused on real-time Transportation Systems Management and Operations (TSMO) applications, this research laid the data science foundation for estimating all volume metrics, from AADT to real-time data. A survey of Coalition Members identified key TSMO use cases benefiting from real-time volume data:

- **Incident and Event Management:** Real-time volume flows enhance situational awareness and provide an objective understanding of regional impacts during major events.
- **Road Closure and Event Clearance Monitoring:** Flow estimates resolve ambiguities in probe-based data, distinguishing between closed roads, extreme congestion, and low travel demand during severe weather.
- **Performance Metrics:** Combining volumes and speeds allows for accurate calculation of real-time user delay costs for management reporting.
- **Diversion Route Analysis:** Flow and turning movement data help track traffic diversion during incidents and predict impacts on adjoining roads.
- **Traveler Information Efficacy:** Volume data can evaluate the effectiveness of traveler information, such as changeable message signs (CMS).

Beyond TSMO, the survey identified planning-level applications focused on typical, day-to-day volume measurements, including:

- **Capacity Utilization:** Volume heat maps help locate under-utilized capacity and monitor traffic diversion to prevent secondary route congestion.
- **Signal Timing:** Reliable volume data during both typical and atypical conditions informs effective signal timing adjustments.
- **Statewide Scaling:** The planning community showed strong interest in using non-real-time probe data to scale traditional measures like AADT statewide.
- **Truck Volumes:** There was distinct interest in measuring volumes for medium and heavy-duty truck classes.

This early research established that the accuracy of probe-based volume estimates depends on the probe penetration rate (sampling percentage), time aggregation, and base traffic volume. Using a simple scaling method based solely on observed probe volumes, the study demonstrated that accuracy improves with higher penetration rates, larger time aggregations, and greater traffic volumes (as shown in Table 2.2.1: Estimated Accuracy of Real-Time Volume – lower percentages indicated higher accuracy).

Table 2.2.1: Estimated Accuracy of Real-Time Volume

	Time Aggregation (minutes)	Volume (vehicles/hour)						
		100	500	1000	1500	2000	3000	5000
5% Sampling	15	87%	39%	28%	23%	19%	16%	12%
	30	62%	28%	19%	16%	14%	11%	9%
	60	44%	19%	14%	11%	10%	8%	6%
10% Sampling	15	60%	27%	19%	15%	13%	11%	8%
	30	42%	19%	13%	11%	9%	8%	6%
	60	30%	13%	9%	8%	7%	5%	4%
25% Sampling	15	35%	15%	11%	9%	8%	6%	5%
	30	24%	11%	8%	6%	5%	4%	3%
	60	17%	8%	5%	4%	4%	3%	2%
50% Sampling	15	20%	9%	6%	5%	4%	4%	3%
	30	14%	6%	4%	4%	3%	3%	2%
	60	10%	4%	3%	3%	2%	2%	1%

An initial proof-of-concept using available probe data yielded the following results:

- **Data Scope:** Probe trace data from 12 freeway locations in the Baltimore-Washington region were compared to continuous count data. Penetration rates ranged from 0.18% to 0.72% (median 0.57%).
- **Traffic Characteristics:** Average hourly observed samples ranged from 22.3 to 62.3 vehicles. The approximate daily traffic on these freeways ranged from 40,000 to 110,000 vehicles per day.
- **Model Inputs:** A machine learning model incorporated supplemental data including speed, road characteristics, incident reports, time of day, day of week, and weather conditions.
- **Performance:** The model achieved R^2 values from 0.61 to 0.94 (median 0.82) and a Mean Absolute Percentage Error (MAPE) from 14% to 48% (median 27%). Despite the low 0.57% penetration rate, incorporating additional attributes allowed the model to perform well across both typical and atypical traffic patterns, highlighting the viability of probe data for planning-level volume estimation.
- **Probe Data Value:** A baseline model run *without* probe samples saw performance drop (median R^2 of 0.73, median MAPE of 37%). This proved that while probe counts are critical for accurate predictions, supplemental attributes also significantly improve planning-level volume estimates.

This proof-of-concept catalyzed further research across diverse datasets using advanced machine learning and AI techniques. The resulting technical papers and presentations established the feasibility and practical value of probe-based volume estimates, ultimately sparking the development of modern industry products.

2.3 TETC Hurricane Proof of Concept Demonstration - 2020

In 2020, TETC launched an ambitious project to demonstrate that emerging connected vehicle data (CVD) could effectively monitor traffic flows across a multi-state region. Unlike fleet telematics, CVD is relayed directly from the vehicle to the original equipment manufacturer (OEM) via real-time wireless communication. The primary objective was to prove that de-identified CVD could support various transportation operations and planning applications at scale.

CVD from automakers first entered the market in 2019, following the widespread integration of broadband communications and GPS into vehicle electronics around 2014. This technological shift provided sufficient data to test CVD for volume estimation. In 2020, the Coalition partnered with data provider Wejo to conduct a hurricane pilot study[3] that addressed several key questions:

- **Viability:** Could global, OEM-sourced CVD serve as a reliable, real-time traffic data source during the Southeast US hurricane season?
- **Processing & Visualization:** Could this massive data feed be processed and visualized in real-time, and what operational challenges would arise?
- **Volume Estimation:** Could CVD generate real-time traffic volume estimates to monitor broad, multi-state hurricane evacuations?
- **Actionable Feedback:** Could this data provide operators with meaningful insights into travel patterns related to preparedness, evacuations, and sheltering in place?

Conducted during the 2020 hurricane season across five southeastern states, the pilot visualized regional CVD as individual vehicles in a live stream, offering profound intuitive value. The dataset was unprecedented in size, generating over 75 billion data points monthly and totaling over 230 billion data points over the three-month study period.

To estimate volumes during these extreme weather events, the study successfully employed a simple scaling factor based on roadway class, time of day, day of week, and state.

Key Insights:

- **Viable Base Dataset:** Even with a penetration rate of just ~2%, CV data proved viable for creating value-added, real-time information like volume estimates and Origin-Destination (O&D) patterns. The value of this dataset will continually increase as newer connected vehicles replace older models in the general fleet.
- **Big Data Challenges:** Managing the sheer size and velocity of CV data at scale is highly challenging. It necessitates efficient processing architectures and streamlined calibration, calculation, and conflation techniques.

[3] https://tetcoalition.org/wp-content/uploads/2021/03/MT2008_Wejo_HurricaneReport_2021.pdf

- **Visualization Value:** Mapping individual vehicle movements provides exceptional intuitive context for transportation data analysis.
- **Estimation Methodology:** While simple scaling factors provided a functional baseline for estimating traffic volumes during major storms, incorporating additional variables and advanced machine learning will be necessary to further improve accuracy.

Ultimately, the 2020 proof of concept successfully demonstrated both the immense potential and the significant data management challenges of utilizing large-scale probe vehicle data. It proved that highly dynamic, granular traffic data could be generated from these sources, outlining a viable technical path for the future of probe-based volume estimation.

2.4 FHWA Consideration of Non-Traditional Methods for Reporting AADT

In 2016, following the MAP-21 legislation, the Highway Safety Improvement Program (HSIP) Final Rule mandated that states have "access to AADT for all public paved roads by 2026[4]." A March 2020 FHWA publication summarized this requirement: State Departments of Transportation (DOTs) must report AADT annually to the Highway Performance Monitoring System (HPMS) for all Federal-aid facilities. Furthermore, the 2016 HSIP Final Rule requires states to access AADT and other Model Inventory of Roadway Elements (MIRE) Fundamental Data Elements (FDE) for all public paved roads—including non-Federal aid-system (NFAS) roads—by 2026[5].

Integrating these FDEs into safety analyses helps agencies identify problem areas and apply appropriate countermeasures, ultimately improving safety. The NFAS includes roads functionally classified as: minor collectors in rural areas (6R), local roads in rural areas (7R), and local roads in urban areas (7U). According to the 2016 FHWA Highway Statistics, these classes account for over 75% of the total US roadway mileage. This massive expansion in AADT reporting requirements forced states to consider expanding their volume count systems. Using traditional methods would require deploying significantly more temporary and permanent count stations to monitor lower-volume roads, which would drastically increase the costs of gathering HPMS data.

To address this challenge—commonly referred to as the MIRE 2026 requirements—and recognizing the potential of probe-based data, the FHWA sponsored a multi-state pooled-fund study: "Exploring Non-Traditional Methods to Obtain Vehicle Volume and Class Data." Running from 2018 through 2021, the study included 19 state DOTs and evaluated how accurately probe-based data could estimate AADT for the upcoming reporting requirements. The study produced a series of reports accessible at https://www.fhwa.dot.gov/policyinformation/travel_monitoring/pubs/aadtnt/.

[4] Federal Highway Administration, Highway Safety Improvement Program Final Rule, Docket No. FHWA-2013-0019. Federal Register, Vo. 81, No. 50, March 15, 2016. https://www.federalregister.gov/documents/2016/03/15/2016-05190/highway-safety-improvement-program

[5] https://safety.fhwa.dot.gov/rsdp/downloads/fhwasa20064.pdf

This FHWA initiative focused on AADT as a planning-level measure and was instrumental for several reasons:

- **Market Validation:** Previously, non-sensor volume estimation was mostly an academic pursuit because vendors doubted the market demand. The FHWA study, combined with impending 2026 MIRE requirements and budget implications, proved market demand and spurred commercial investment.
- **Applicability of Existing Research:** Estimating an annual average (AADT) is technically less demanding than real-time traffic tracking. The core datasets and research models developed for operational tracking (such as the MCOMP study) proved highly applicable to planning-level AADT estimation.
- **Established Accuracy Benchmarks:** The FHWA developed accuracy standards for non-traditional AADT estimates, establishing that they should be at least as accurate as factored 48-hour counts. This provided clear "goalposts" and validation methods for the industry.
- **Industry Catalyst:** The MAP-21 rulemaking and the resulting 2026 MIRE requirements created an urgent need for scalable, cost-effective volume estimation, directly driving industry investment in probe-based technologies.

Ultimately, the FHWA's initiative was a significant catalyst that prompted the industry to develop commercial probe-based volume estimation products.

2.5 The Eastern Transportation Coalition Transportation Data Marketplace – 2022

The Eastern Transportation Coalition (TETC) launched the initial Transportation Data Marketplace (TDM) in 2008. Member agencies pooled resources to create a corridor-wide traffic monitoring system that remains the largest of its kind in the US. Initially, the TDM provided only travel time and speed measures through a single vendor. This groundbreaking procurement introduced specifications-based accuracy requirements for probe-derived data, a validation program, performance penalties, confidence indicators for low-probe periods, and common licensing terms allowing states to share data with adjoining and sub-jurisdictions.

In 2014, Phase II of the TDM introduced a multi-vendor approach, allowing member states to select from three pre-approved probe data vendors for travel time and speed. In 2022, Phase III retained the core tenets of the previous phases while expanding to include five new data categories in addition to travel time and speed:

- Traffic volume measures derived from probe data
- Origin-destination data
- Waypoint data (including connected vehicle and location-based service data)
- Freight data
- Conflation data and services

The volume data specifications within the TDM covered planning-level measures (AADT, ADT, and AHDT), high time-granular measures (daily and hourly traffic), and

real-time data feeds for TSMO applications. Multiple vendors were awarded contracts and successfully delivered planning and granular volume estimates, though none were able to provide real-time volume products.

TMD Phase III set a new benchmark for procuring planning-level volume data. Since 2023, the TDM's validation program has continuously evaluated the quality of these volume data feeds, beginning with an initial study in North Carolina (2023), followed by a comprehensive study in Georgia (2024), a validation effort in Massachusetts (2025), and a study in Maryland currently in progress (2026). Driven by the impending MIRE reporting requirements, Coalition members are primarily interested in planning applications. Consequently, validation efforts have heavily emphasized testing AADT data against FHWA accuracy guidelines for all public roadways, while also assessing ADT, AHDT, daily, and hourly volumes. The results of these validation studies are reviewed in Case Studies.

Commercial Validation of Probe-Based Volume Estimates (2024)

Further supporting these marketplace advancements, Kia Eisinga and Stefan Lorkowski of TomTom (a TDM source data supplier) presented "Network-Wide Traffic Volume Estimation Based on Probe Vehicle Data"[6] at the January 2024 Transportation Research Board Annual Meeting. The paper proposed estimating network-wide traffic volumes by combining large-scale probe data with stationary detector data to model probe penetration rates at the road level. While the technical approach built upon earlier academic research, the paper is notable because it validated these foundational methods using a massive commercial dataset.

Using probe data from the Netherlands with penetration rates of 20–30% for motorways and 10–25% for other roads, the authors modeled planning-level volume measures based on road class, time of day, and lane counts. The model achieved a mean absolute percentage error (MAPE) of 25% for two-minute intervals on motorways, which improved to 7% for daily estimates. For non-motorways, maximum accuracy was roughly 12% MAPE over longer intervals. These findings confirmed the established relationships between volume estimation accuracy, probe penetration rates, time aggregation, and base volumes.

Collectively, these milestones—from early research to the TDM validation program and large-scale commercial validation—illustrate the significant contributions that have established current industry capacity for probe-based volume estimation.

[6] Eisinga, K., & Lorkowski, S. (2025). Network-Wide Traffic Volume Estimation Based on Probe Vehicle Data. Transportation Research Record, 0(0). https://doi.org/10.1177/03611981241289408

3.0 Summarized Case Studies of Industry Capacity

Activities within the TETC Transportation Data Marketplace (TDM) validation program reflect the ongoing assessment of industry capacity since the TETC initiated the current phase of the TDM in 2022. The TETC conducted a series of validation activities spanning North Carolina, Georgia, and Massachusetts to assess the quality of probe-based volume estimates for planning applications, with particular emphasis on Average Annual Daily Traffic (AADT).

This section summarizes these validation activities sequentially, providing a narrative of the continuing evolution of industry maturity and the practical lessons learned for effectively leveraging probe-based volumes within a DOT. *Note: Detailed statistical breakdowns, scatter plots, and error distribution analyses for each case study can be found in the Appendices.*

3.1 Early Exercises

Prior to conducting formal volume accuracy validations, the TETC validation team conducted a literature review[7] and a preliminary validation exercise[8] to test methods, metrics, and logistics. A critical outcome of these early exercises was the implementation of the CATTWorks Geo-referencing Protocol (CWGP) [9]. Requiring each vendor to submit results in a common geo-referencing format allowed the TETC to efficiently evaluate traffic data accuracy across multiple vendors without extensive, error-prone conflation activities.

3.2 North Carolina Volume Validation (Nov 2023)

The 2023 North Carolina validation was the first full attempt to quantify industry capacity to deliver volume estimates at scale. The study evaluated data from four TDM vendors across 263 continuous count stations (CCS), focusing on hourly counts, Average Hourly Daily Traffic (AHDT), Average Daily Traffic (ADT), and AADT.

Key Takeaways:

- **Proven Feasibility:** All four vendors successfully delivered volume data across a majority of the CCS locations in a standardized format, marking a major milestone for the industry's scalable capabilities.
- **Standardized Evaluation:** The exercise successfully established a standard set of visualizations, metrics, and benchmarks (incorporating FHWA evaluation methods) that would be used in all future validations.
- **Early Product Maturity:** Accuracy was quantifiable and varied moderately between vendors and across volume ranges. While performance was mixed across urban and rural aggregations, the viability of using probe data for typical volume patterns was confirmed.

[7] Traffic Volume Validation - Literature Review and Recommendations – Nov 2022, https://tetcoalition.org/wp-content/uploads/2015/02/TDM-Val-1-Report-Volume-Nov2022.pdf
[8] TDM Validation Activity: Volume & Travel Time Study – April 2023 https://tetcoalition.org/wp-content/uploads/2015/02/TDM-Val-02-Report-FINAL-V2.4.pdf
[9] https://tetcoalition.org/wp-content/uploads/2015/02/2021-ETC-TDM-CWGP_V1.10-FINAL.pdf

(For full methodology, data distribution, and detailed vendor accuracy results, see Appendix C).

3.3 Georgia Volume Validation (2024)

Building on the North Carolina study, the 2024 Georgia validation expanded the assessment to include a fifth vendor and integrated temporary, short-term count (STC) sites alongside CCS locations. This study encompassed 963 directional validation locations across all functional road classes (1-7) and introduced daily volumes as a fifth time aggregation.

Key Takeaways:

- **Measurable Industry Progression:** Vendor products showed noticeable improvement from the previous year. Three of the five vendors demonstrated the ability to capture average day-of-week and time-of-day patterns (AHDT) and were on the verge of meeting FHWA accuracy benchmarks for AADT.
- **Detection of Additional Error Sources:** As vendor modeling improved, other sources of error became apparent. The validation exposed that conflation/georeferencing misinterpretations and errors within the agency's own reference data heavily impacted accuracy metrics.
- **Importance of Training Data:** Vendor performance was notably stronger on higher functional class roads (where continuous count training data is abundant) and less accurate at temporary count sites (predominantly lower-volume roads where training data is sparse).

(For detailed statistical metrics, urban vs. rural performance breakdowns, and FHWA test results, see Appendix D).

3.4 Massachusetts Volume Validation (2025)

In 2025, TETC partnered with MassDOT to evaluate vendor volume estimates at over 1,000 evaluation locations. This validation introduced new complexities by incorporating toll road facilities and video camera counting technology deployed at signalized intersections.

Key Takeaways:

- **Counter Technology and Context Matter:** The diversity of MassDOT’s count sites revealed that performance varies by counter type. Traditional inductive loops and Automated Electronic Tolling (AET) sites yielded better results than video camera sites. Errors at camera locations were generally higher, likely due to a combination of technological limitations and the complexity of urban intersection environments.
- **Refining the FHWA Test Implementation:** The validation team adopted updated FHWA guidance, focusing exclusively on bi-directional volumes and aggregating the precision test across all volume bins. Under these updated guidelines (and excluding complex camera locations), two vendors successfully passed all components of the FHWA statistical test.

- **Non-Vendor Errors Must Be Managed:** The study underscored the necessity of robust data cleaning. Miscoded agency sites, ambiguous location referencing near ramps, and occlusion in camera feeds created massive outliers that distorted initial vendor performance metrics.

(For a comprehensive breakdown of performance by counter type, site complexity analyses, and finalized FHWA methodology results, see Appendix E).

3.5 Minnesota DOT Case Study (2021)

Independent of the TETC validations, the Minnesota DOT (MnDOT) initiated a multi-year implementation plan starting in 2021 to qualify third-party probe data for HPMS submission.

Key Takeaways:

- **High-Volume Viability:** MnDOT's initial 2021 analysis of roadways exceeding 20,000 AADT showed strong alignment between probe estimates and traditional counts. After applying their acceptable tolerance criteria, the Mean Absolute Percent Error (MAPE) reached 5.4%, prompting MnDOT to recommend mobile source data for high-volume roadways to improve personnel safety.
- **Lower-Volume Challenges:** A subsequent 2023 retest expanding into the 10,000–19,999 AADT range revealed significant performance drops. Probe estimates failed to meet acceptable parameters in these mid-to-lower volume ranges, highlighting the ongoing challenge of relying on probe data where vehicle samples are less dense.

(For MnDOT's complete testing plan parameters and detailed tolerance criteria results, see Appendix F).

4.0 Guidelines for Qualifying Probe-Based Volume Estimates for HPMS

This section provides practical guidelines for acquiring, evaluating, and integrating probe-based volume estimates for agency use. The Eastern Transportation Coalition (TETC) synthesized these guidelines from cumulative lessons learned across earlier research, Transportation Data Marketplace (TDM) validation activities, USDOT pooled-fund studies into non-traditional volume estimation, and peer state DOT initiatives.

The guidance is organized chronologically across four core phases:

1. **Planning** (2–3 months prior to evaluation)
2. **Testing Preparation** (1 month prior to evaluation through data collection)
3. **Execution of Evaluation** (post-collection statistical testing and assessment)
4. **Ongoing Monitoring and Re-evaluation** (long-term data quality maintenance)

This procedural roadmap provides a step-by-step approach for qualifying probe-based volume estimates for Highway Performance Monitoring System (HPMS) submittals and broader departmental workflows. Supporting technical details, statistical equations, and historical case studies are referenced throughout and detailed in the report's appendices.

- **Planning:** Covers foundational tasks undertaken two to three months before initiating a validation activity. Key steps include chartering a steering committee, identifying candidate vendors, establishing evaluation parameters, scoping required resources, adopting a standardized georeferencing protocol, and anticipating potential non-vendor error sources.
- **Testing Preparation:** Details the operational procedures executed within one month of scheduled testing. This phase encompasses defining evaluation timeframes, packaging calibration data for vendors, performing quality assurance checks, and establishing secure data delivery mechanisms.
- **Evaluation:** Commences once reference ground-truth data and vendor estimates are assembled for side-by-side comparison. This stage outlines statistical analyses, diagnostic visualizations, outlier screening protocols, and the formal application of the Federal Highway Administration (FHWA) acceptance criteria.
- **Ongoing Monitoring and Re-evaluation:** Outlines strategies for periodically assessing outsourced volume data over time, accounting for continuous evolutions in vendor algorithms, probe sampling penetration, and underlying road network baselines.

While these procedures are tailored specifically to probe-based volume data for HPMS reporting, the overarching principles apply to any outsourced transportation data procurement.

4.1 Planning

The planning phase establishes the organizational governance, technical scope, and evaluation protocols. Planning activities should begin at least two to three months prior to the scheduled evaluation window.

4.1.1 Identify a Project Champion

A dedicated project champion within the agency must lead the initiative from inception through completion. The champion should develop a strong working knowledge of probe-based volume technologies, previous validation studies, and federal standards. They will serve as the primary liaison between the internal steering committee, the technical evaluation team, participating vendors, and FHWA division representatives.

4.1.2 Assemble a Steering Committee

Form a multi-disciplinary steering committee composed of agency leaders and stakeholders who will rely on or benefit from probe-based volume data. Recommended committee representation includes:

- **HPMS and Traffic Monitoring Staff:** Personnel responsible for collecting, processing, factoring, and reporting volume data (e.g., AADT) for federal submittals.
- **Cross-Departmental Stakeholders:** Representatives from operations (TSMO), performance management, roadway design, access management, safety, and planning.
- **FHWA Division Office:** Regional technical liaisons to ensure alignment with federal reporting expectations.
- **Peer Agency Partners:** Representatives from adjacent states or regional planning organizations (MPOs) pursuing similar data initiatives.

Chaired by the project champion, the steering committee should convene every two to four weeks during the planning phase and meet every four to six weeks during subsequent execution stages.

4.1.3 Designate the Evaluation Team

Evaluating vendor data accuracy requires dedicated resources and technical proficiency in data science, HPMS methodologies, and geospatial analysis. The estimated level of effort is approximately one person-year, typically distributed across a specialized team of two to four analysts.

An agency may conduct this evaluation internally or engage an independent third party (e.g., a university research center, qualified consultant, or regional coalition resource). To preserve data integrity and prevent real or perceived conflicts of interest (COI):

- Any third-party entity conducting the evaluation must demonstrate an absence of commercial ties or pre-existing business agreements with the candidate data vendors.
- If the DOT maintains ongoing commercial contracts with participating data vendors, assigning the technical evaluation to an independent research or coalition partner is strongly advised.

- The designated evaluation team reports directly to the steering committee.

4.1.4 Coordinate Closely with HPMS Staff

The primary ground-truth benchmarks used to evaluate vendor estimates originate from permanent continuous count stations (CCS) and short-term count (STC) programs typically managed by personnel associated with the HPMS reporting process. These staff members possess direct institutional knowledge regarding detector calibration histories, hardware limitations, sensor drift, and local roadway conditions.

HPMS staff must participate on the steering committee and maintain regular technical coordination with the evaluation team to ensure that baseline count data are properly screened, quality-checked, and interpreted.

4.1.5 Define and Document Agency Use Cases

While the primary objective is qualifying non-traditional AADT data to satisfy Model Inventory of Roadway Elements (MIRE) reporting requirements, probe-derived volumes can serve diverse agency needs. Under the direction of the steering committee, document all intended use cases prior to issuing vendor solicitations. Clarifying these applications ensures that evaluation scopes, data aggregation intervals, spatial coverage, and accuracy thresholds align directly with agency objectives.

4.1.6 Identify and Solicit Candidate Vendors

Engage two or more commercial vendors capable of providing probe-based volume data. The TETC validation program strongly recommends evaluating multiple vendors for several operational reasons:

- **Mitigating Supply Chain Disruptions:** Upstream probe data feeds are susceptible to market shifts, corporate restructuring, and evolving mobile operating system privacy policies. Evaluating multiple providers, each with differing data supply chains, mitigates the vulnerability of relying on a single data pipeline.
- **Isolating Error Sources:** Comparing multiple vendor estimates against agency reference data helps pinpoint the root cause of discrepancies. When all vendors deviate uniformly from agency counts, the error often stems from ground-truth anomalies or georeferencing mismatches. Conversely, when an outlier appears in only one dataset, the issue is typically vendor-specific.
- **Assessing Varied Data Sources:** Different vendors leverage distinct combinations of raw source data (e.g., connected vehicle telematics, mobile application location-based services, freight fleet tracking). Testing multiple offerings helps the agency understand how specific source feeds perform across functional road classes and other contexts of interest to the agency.
- **Preventing Vendor Lock-In:** Maintaining a competitive environment encourages ongoing product innovation and responsive technical support throughout the procurement lifecycle.

Solicit vendor participation through a Request for Information (RFI) or formal solicitation detailing the evaluation framework. Vendors must disclose their underlying data sources, processing pipelines, model calibration strategies, and contingency plans for data supply interruptions.

4.1.7 Select Evaluation Metrics, Methodologies, and Decision Criteria

The steering committee must formally adopt and document the evaluation metrics and acceptance standards prior to data submission. The TETC recommends incorporating the following components:

- **Evaluation Metrics and Visualizations:**
    - Adopt standardized statistical metrics as detailed in Appendix A, including Median Error Bias (MdE), Median Percent Error (MdPE), Mean Absolute Percent Error (MAPE), Absolute Percent Error (APE) percentiles, and Normalized Root Mean Square Error (NRMSE).
    - Generate scatter plots and error distribution charts consistent with previous North Carolina, Georgia, and Massachusetts validation protocols.
    - Implement the official FHWA evaluation methodology for non-traditional AADT data (see Appendix B).
- **Decision Criteria:**
    - **Full Acceptance:** If a vendor satisfies all FHWA accuracy and precision criteria across all volume bins, their data are deemed acceptable for statewide HPMS submission.
    - **Stratified/Conditional Acceptance:** If a vendor satisfies FHWA criteria only within specific volume bins (e.g., high- and medium-volume roadways), the agency may approve the product exclusively for those corresponding functional classifications while retaining traditional count methods elsewhere.
    - **Marginal Performance:** If a vendor narrowly misses acceptance thresholds, the steering committee may grant conditional approval paired with closer operational monitoring and an accelerated re-evaluation timeline.

4.1.8 Establish Outlier Investigation Protocols

Unusually large discrepancies between vendor estimates and reference counts must be systematically investigated before drawing conclusions about vendor accuracy. If an investigation reveals that an error is attributable to external factors—such as ground-truth sensor failure, mislabeled lane counts, or conflation errors—the affected station should be flagged and removed from the scoring dataset.

If systemic reference errors corrupted the initial calibration dataset supplied to vendors, the steering committee may, at its discretion, permit vendors to recalibrate models using corrected baseline data and resubmit estimates.

Analyses and visualizations should be segmented across meaningful roadway and operational classifications:

- Continuous Count Stations (CCS) versus Short-Term Count (STC) stations
- Functional classification (Interstates, arterials, collectors, local roads)
- Volume bins (High, Medium, Low, Very Low)
- Geographic context (Urban versus Rural)
- Sensor detection technology (Inductive loops, Automatic Electronic Tolling [AET], video analytics, radar)
- Network complexity (Segment geometry, proximity to interchange ramps, signalized intersections)

Following preliminary analysis, the evaluation team should conduct individual debriefings with participating vendors to review potential data anomalies. Vendors may provide clarifying context; however, permissions to resubmit data remain exclusively at the discretion of the steering committee.

4.1.9 Determine Confidentiality and Publication Terms

Agencies must establish confidentiality terms prior to data exchange:

- **Uncompensated/Voluntary Participation:** If vendors supply evaluation data at no cost, results should be treated confidentially or anonymized (e.g., Vendor A, Vendor B) in public documentation, reserving identifiable results for internal agency decision-makers.
- **Compensated/Procured Evaluations:** If the agency compensates vendors for providing evaluation data, the contract should explicitly retain the agency's right to publish and disseminate identified performance findings.

4.1.10 Adopt a Georeferencing Standard for Data Exchange

Both transportation agencies and private vendors maintain data in varying digital roadway network formats (such as state Linear Referencing Systems [LRS], commercial map databases, OpenStreetMap, or shared TMC networks).

TETC recommends using a geo-referencing standard for acquiring and comparing data between agency and vendor sources. TETC uses the CATT-Works Geo-referencing Protocol (CWGP) for communicating roadway locations along a road network. Other geo-referencing standards exist, such as OpenLR. In TETC validation activities, validation locations (i.e., locations where agency counts have been collected) are described using CWGP and shared with vendors to communicate where vendors should prepare volume estimates.

A georeferencing standards communicates locations using geodetic coordinates with attributes that uniquely identify the location in a map neutral fashion (i.e., latitude, longitude, heading, functional classification of roadway, number of lanes, and other attributes). This minimizes agency resources by avoiding conflating a vendor’s volume estimates from a digital map format to its own internal map representation. When using a georeferencing format, each vendor is responsible for interpreting location references into their own internal mapping system to produce volume estimates consistent with the geo-referencing protocol.

4.1.11 Convey Evaluation Plan to FHWA for Review and Approval

Before executing field tests, summarize the governance structure, technical methodology, georeferencing approach, and acceptance criteria in a formal memorandum to the regional FHWA division office. Securing early federal concurrence memorializes the validation strategy and verifies that successful outcomes will satisfy MIRE and HPMS compliance guidelines.

4.2 Testing Preparation

Testing preparation begins at least one month prior to data collection and spans the operational setup required before vendor evaluations begin.

4.2.1 Select Data Aggregation Intervals

Although Annual Average Daily Traffic (AADT) serves as the primary metric for HPMS submittals, agencies are encouraged to evaluate additional temporal aggregations—including Average Daily Traffic (ADT), Average Hourly Daily Traffic (AHDT), Daily, and Hourly volumes:

- **Enabling Blind Testing:** Because official AADT reflects a full 12-month period often published in state repositories, conducting a true blind evaluation can be challenging. Evaluating sub-annual aggregations (such as monthly ADT or hourly volumes) over a period where agency ground truth is withheld enables nire rigorous, blind validation.
- **Broadening Utility to Other Application Needs:** Assessing shorter-duration aggregations tests the data's utility for other agency applications such as signal optimization, and work-zone traffic management. (See targeted Use Cases from section 4.1.5, and assess their corresponding data aggregation needs.)
- **Diagnosing Data Anomalies:** Evaluating granular hourly or daily profiles helps determine whether AADT errors stem from baseline vendor modeling capacity or other non-vendor related causes such as conflation, georeferencing ambiguity or possibly sensor error.

4.2.2 Define the Evaluation Period

For AADT assessments, establish a 12-month evaluation window corresponding to a full calendar year (January 1 through December 31), if possible, to maintain consistency with standard federal reporting practices.

For more granular aggregations (ADT, AHDT, Daily, Hourly), designate a targeted one-to-three-month window capturing operational periods of interest (e.g., summer recreational travel surges or stable autumn commuter traffic). Count data collected during this targeted window should remain confidential until vendor data are submitted, creating a robust blind evaluation environment.

4.2.3 Assemble the Calibration Dataset

Provide participating vendors with a clean, fully quality-checked calibration dataset spanning at least 12 months prior to the evaluation window. This baseline dataset allows vendors to train and tune their estimation algorithms to local network conditions:

- Include directional volume counts formatted in accordance with the established georeferencing standard (e.g., CWGP), noting lane counts, detector technology, and roadway context.
- Restrict calibration inputs to Continuous Count Stations (CCS) that exhibit at least 75% annual data availability.
- Verify directional volume parity. If opposing directional AADT values differ by more than 5% to 10% without an identifiable geometric or operational cause, flag and exclude the station.
- Conduct a network complexity review to identify and screen ambiguous sensor placements, such as detectors located within multi-phase signalized intersections, slip ramps, or closely spaced freeway interchanges.

Note: Not all vendors will require or make use of the calibration dataset, as many already access nationally available data through the FHWA Travel Monitoring Analysis System (TMAS) and other industry data sources. Even if vendors do not require calibration data, the calibration data set provides a baseline for future reference if needed, as well as establishes the georeferencing for AADT comparison.

4.2.4 Acknowledge and Document Possible Non-Vendor Error Sources

Field validation exercises routinely uncover significant discrepancies resulting from factors outside vendor algorithms. Known sources of error that have been exposed in case studies include:

- **Conflation and Geocoding Errors:** Ambiguities in roadway location or attribution that cause vendors to map estimates to incorrect parallel facilities, collector-distributor lanes, or ramps.
- **Ground-Truth Baseline Errors:** Sensor location, calibration or other anomalies may be present within agency historical records.
- **Technology-Specific Sensor Limitations:** Counting variances inherent in complex or emerging sensor environments, such as occlusion in computer-vision camera feeds at intersections or radar detection dropout in heavy congestion.

The Steering Committee and HPMS Team may be aware of additional system specific issues that should be monitored as possible source of non-vendor error.

4.2.5 Establish Vendor Data Submission Protocols

Set a Vendor Data delivery deadline (typically four to six weeks following the close of the evaluation period) and establish a secure, access-controlled file exchange portal (e.g., SFTP or cloud repository). Require vendors to perform a test transfer prior to the final deadline to ensure seamless data delivery.

4.2.6 Assemble the Reference Dataset

Following the evaluation period, compile the ground-truth Reference Dataset across all target count locations (both permanent CCS and temporary short-term count sites). Apply rigorous quality controls:

- Calculate official reference AADT across the 12-month evaluation window using established HPMS factoring and computation standards.

- Derive benchmark ADT, AHDT, Daily, and Hourly counts across the dedicated sub-annual window.
- Format all records using the standardized georeferencing protocol, appending metadata regarding sensor technology and network geometry.
- Exclude CCS locations with less than 75% data availability and reconcile directional imbalances greater than 5% to 10%.

4.3 Evaluation of Vendor Data

Evaluation commences once vendors deliver their submittals. The evaluation team performs data formatting reviews, conducts exploratory visual analytics, isolates systemic outliers, and executes the formal FHWA statistical tests.

4.3.1 Perform Quality Assurance on Vendor Formats

Upon receipt, review all vendor files for compliance with the georeferencing standard. Verify schema definitions, field order, coordinate systems, and directional indicators. If files contain formatting anomalies, determine whether they can be resolved internally, require vendor correction, or warrant file rejection.

4.3.2 Explore Data Using Statistical Analyses and Visualizations

Conduct exploratory data evaluations on directional AADT before applying federal pass/fail criteria:

- Calculate fundamental error metrics: Median Error (MdE), Median Percent Error (MdPE), Mean Absolute Percent Error (MAPE), and error percentiles across all roadway classifications and volume tiers.
- Generate scatter plots (Vendor vs. Reference) and error distribution histograms to evaluate system-wide bias and variance.
- Segment performance across count station types (CCS vs. STC), detector technologies (loop, camera, radar, tolling gantry), functional classes, and urban versus rural environments.

4.3.3 Investigate Systemic Outliers

Using diagnostic plots, identify the most severe outliers (focusing on the top five to ten largest errors per vendor to manage analysis effort) and apply the following screening logic:

- **Multi-Vendor Agreement:** If multiple independent vendors deviate systematically in the same direction and magnitude from an agency count, the issue frequently stems from ground-truth counter failure or geocoding discrepancies.
- **Directional Balance:** Examine opposing directional estimates. If an agency count is balanced but a vendor’s directional estimates are sharply asymmetrical, the error is likely internal to the vendor's model. Conversely, if an agency count is asymmetrical, the error is likely external to the vendor.

- **Directional vs Bi-Directional Errors:** If a vendor's estimate is approximately half or double the reference volume, evaluate whether a directional processing error occurred (e.g., reporting one-way volume for a two-way roadway or vice versa).
- **Sensor/Contextual Clustering:** Determine whether large discrepancies correlate with specific sensor hardware (e.g., camera feeds experiencing occlusion) or complex intersection configurations.

Document all findings and classify outliers as vendor modeling errors, or errors not related to vendor volume models (such as georeferencing errors, or agency reference defects.)

Debrief/update Steering Committee at this stage on results of preliminary evaluation. If minimal anomalies/outliers are discovered, the evaluation results may progress to final evaluation. If substantial anomalies/outliers are discovered, additional time and resources may be needed to resolve. Vendor interaction/debriefing may be required at this stage to resolve outliers/anomalies.

4.3.4 Remove Non-Vendor Suspect Locations and Finalize Datasets

Remove count locations compromised by verified non-vendor errors from the scoring dataset. Recompute foundational statistical distributions and diagnostic charts using the cleaned reference dataset across all target temporal aggregations (AADT, ADT, AHDT, Daily, and Hourly).

4.3.5 Prepare Bi-Directional Datasets and Apply FHWA Acceptance Tests

Because the FHWA acceptance methodology evaluates bi-directional traffic, combine directional volumes at each valid count location to generate bi-directional totals:

1. **Assign Volume Bins:** Classify each reference location into its respective volume tier: Very Low (<500 AADT), Low (500–4,999 AADT), Medium (5,000–54,999 AADT), or High (AADT).
2. **Execute Accuracy Test #1 (Median Error Bias):** Compute the Median Traffic Count Error (TCE) percentage within each volume bin and verify compliance with the sample-size-adjusted tolerance interval.
3. **Execute Accuracy Test #2 (Mean Absolute Percent Error):** Calculate the MAPE within each volume tier and verify that it does not exceed the Upper 95% Confidence Bound threshold.
4. **Execute the Precision Test:** Compare individual site percentage errors against the 95% TCE Population Error Range. Sum all records falling outside the allowable error envelope and verify that the count of failing sites does not exceed the allowable limit specified in the binomial acceptance sampling plan (see Appendix B).
5. **Synthesize Findings:** Consolidate pass/fail outcomes into an agency evaluation matrix detailing performance across all volume bins.

Brief the steering committee and participating vendors on preliminary diagnostic findings.

4.3.6 Conduct Vendor Technical Debriefings

Provide each vendor with an individualized draft evaluation detailing their specific performance metrics, outlier summaries, and pass/fail statuses. Allow vendors a two-week review window to ask technical questions and submit clarifying documentation regarding anomalous results.

4.3.7 Present Findings to the Steering Committee

Convene the steering committee to review the final evaluation report. The evaluation team should deliver a detailed presentation highlighting overall industry capability, vendor-specific strengths and weaknesses, contextual performance across functional classes and other contexts of interest, and identified error mechanisms.

Based on final feedback, the Evaluation Team prepares the draft Evaluation Report.

4.3.8 Steering Committee Procurement and Utilization Decisions

Based on the final evaluation the steering committee will:

- Determine which vendors successfully satisfied the required quality thresholds.
- Authorize contract awards and establish data integration plans for annual HPMS submissions and departmental workflows as appropriate.
- Formulate formal adoption parameters (e.g., approving a vendor statewide or restricting deployment to specific volume ranges while retaining traditional short-term counts elsewhere).
- Define re-evaluation schedules:
    - **Unconditional Pass:** Re-evaluate every two to three years.
    - **Marginal or Stratified Pass:** Re-evaluate annually or every two years, pairing implementation with enhanced quality monitoring.

Finalize and convey the final evaluation report and procurement decisions to the regional FHWA division office to formalize the agency's non-traditional volume data program. Sharing technical insights and methodologies with regional peer agencies and industry organizations is also encouraged to support nationwide traffic monitoring practices.

4.4 Ongoing Monitoring and Re-Evaluation

Outsourced volume datasets require ongoing quality oversight. Upstream data supply chains evolve continuously as connected vehicle and probe-vehicle fleets penetration expand, algorithm architectures advance, and base map networks undergo maintenance.

- **Periodic Re-Evaluation Schedule:** Re-evaluate contracted vendors every two to three years. Implement shorter one-to-two-year cycles if the initial procurement was based on marginal performance or limited to specific functional tiers.
- **Ongoing Quality Control Checks:** Maintain routine spot checks and spatial sanity tests against newly collected permanent CCS and temporary count data throughout the contract period as opportunities for quality assessment.

- **Monitor Upstream Supply Chains:** Maintain direct communication with data providers regarding adjustments to upstream telematics partners, algorithm processing heuristics, or base map conflation routines that could alter reporting fidelity.
- **Mitigate Systemic Conflation Issues:** Ensure that internal agency road network updates (e.g., lane additions, re-alignments, functional reclassifications) are systematically communicated to data vendors to prevent georeferencing mismatches during data integration.
- **Track Intra-Agency Data Utilization:** Document emerging use cases across planning, safety, maintenance, and operations. Tracking broad organizational usage justifies ongoing data procurement costs and identifies specific metric requirements for future evaluations and subsequent procurements. Note that integration of vendor-provided probe-based volume estimates are subject to many of the same non-vendor error sources discussed in the evaluation process.
- **Stay Apprised of Other Agencies Evaluation:** Other states' evaluation results may apply to your agency – hastening or lengthening the needed re-evaluation. TETC validation activities are performed approximately yearly on industry volume products, and involve a majority of the industry providers.

5.0 Issues for Future Attention

During the synthesis of this guidance, the assessment of industry maturity, and the review of multi-state case studies, several unresolved technical and institutional issues emerged that warrant ongoing research, development, and policy attention. These issues span three primary areas:

1. The need for independent, blind testing of volume estimates
2. Accounting for error and technological diversity in agency reference data
3. Establishing sustainable, multi-state quality assurance and quality control (QA/QC) frameworks for procured data

---

5.1 The Need for Independent, Blind Testing of Volume Estimates

A foundational methodological challenge in validating probe-based volume models is the lack of statistical independence between training data and evaluation benchmarks. Currently, calibration data (provided to vendors to train models) and reference data (used by agencies to validate performance) often originate from the same primary pool of physical monitoring stations:

- **Ubiquitous and Historical CCS Data:** Continuous Count Stations (CCS) are fixed, publicly cataloged assets. The vast majority possess multi-year historical count records that are freely accessible through public agency web portals or the Federal Highway Administration's (FHWA) Travel Monitoring Analysis System (TMAS). Furthermore, CCS deployments are predominantly concentrated on higher-functional-class, high-volume facilities.
- **Predictability of Short-Term Counts:** Although temporary count sites (STCs) are typically surveyed on three-year cycles and are less publicly visible than permanent stations, historical counts at these repeatable locations remain accessible or inferable.
- **The Expanded Network Imperative:** Validating probe-derived estimates for Model Inventory of Roadway Elements (MIRE) compliance requires evaluating mid-to-low-volume roadways and lower functional classes (e.g., minor collectors and local roads), which constitute more than 75% of public highway mileage.

Because machine learning (ML) and artificial intelligence (AI) estimation algorithms may have ingested historical data from known validation locations during model development, evaluating vendor estimates at these same sites presents a risk of overfitting and sample contamination. To establish true statistical independence, future validation programs should pursue three complementary approaches:

5.1.1 Evaluating High-Granularity, Atypical Flow Patterns

Evaluating temporal aggregations with higher granularity—such as Average Daily Traffic (ADT), Average Hourly Daily Traffic (AHDT), Daily, and Hourly volumes—provides deeper insight into an algorithm's capability to capture true dynamic travel behavior. While models can reproduce baseline seasonal trends by memorizing historical annual averages, their ability to estimate atypical conditions (e.g., severe weather events, work-zone diversions, or holiday traffic surges) serves as a direct indicator of probe-scaling

efficacy. Demonstrating high fidelity during anomalous conditions bolsters confidence in underlying Annual Average Daily Traffic (AADT) calculations. However, not all commercial vendors offer sub-annual or hourly volume products, and formal statistical methodologies linking high-granularity precision directly to annual AADT accuracy standards remain an active area of investigation.

5.1.2 Auditing Vendor-Provided Cross-Validation Metrics

Standard ML training workflows incorporate k-fold cross-validation or out-of-bag validation, where models are iteratively trained on a subset of count stations and evaluated against withheld stations. During the 2025 Massachusetts Department of Transportation (MassDOT) validation, The Eastern Transportation Coalition (TETC) requested and received vendor-generated cross-validation performance reports. Systematically comparing vendor-reported cross-validation statistics against external ground-truth validation results provides valuable diagnostic insight into model generalization and should become standard practice in future evaluations.

5.1.3 Collecting Blind, Independent Validation Counts

The most rigorous defense against model memorization is establishing fully independent, blind validation sites. Under this protocol, agencies augment the reference dataset with temporary short-term counts (e.g., 48-hour pneumatic tube or radar counts) deployed at locations with no prior count history. Vendors are supplied only with geographic location references (e.g., via standard georeferencing coordinates) without historical volume baselines. The resulting field counts can then be factored to annual averages or directly evaluated against vendor daily estimates.

The critical necessity of blind evaluation is highlighted by an experiment conducted by the TETC validation team: using only the prior year's published AADT at permanent stations as a mock vendor submittal, this data outperformed all commercial probe-based volume products in standard benchmark testing.

5.2 Acknowledging Error and Technology Variance in Agency Reference Data

Federal statistical benchmarks for non-traditional AADT estimation were originally formulated by modeling the sampling error inherent in factoring 48-hour short-term counts relative to continuous inductive loop detectors. However, real-world baseline deployments have expanded considerably beyond traditional inductive loops:

- **Detector Technology Diversity:** Contemporary count networks increasingly incorporate non-intrusive sensor technologies, including computer-vision video analytics, and side-fire radar.
- **Site Context and Operational Environment:** As demonstrated in the Massachusetts validation study, count accuracy can vary substantially depending on physical context. Sensors positioned near interchange ramps, complex signalized intersections, or urban corridors with optical occlusion exhibit significantly higher measurement variance than open-highway loops.

These variances raise two pivotal research questions for state DOTs and the FHWA:

1. **Reference Equipment Qualification:** Should formal acceptance testing of commercial probe volumes be restricted exclusively to traditional, highly calibrated count stations (such as verified inductive loop CCS)?
2. **Re-Evaluating Accuracy Benchmarks:** Should federal baseline error distributions for short-term counts be re-benchmarked to reflect the empirical precision and operational limitations of modern sensor types and complex geometric contexts?

Addressing these questions requires systematic national research to quantify the baseline error envelopes of alternative sensor technologies and establish clear data-filtering thresholds for validation programs.

5.3 Establishing Sustainable QA/QC Programs for Outsourced Transportation Data

As public agencies transition from hardware-intensive monitoring programs to procuring third-party data services, fundamental questions arise regarding quality control governance:

- **Re-Evaluation Frequency:** Given the rapid evolution of connected vehicle data penetration, commercial data fusion techniques, and base map maintenance schedules, what is the optimal interval for recertifying commercial volume providers?
- **Transferability Across Jurisdictions:** Can rigorous validation results obtained by one state DOT serve as reciprocal qualification in neighboring jurisdictions, or do regional variations in fleet probe penetration demand state-by-state evaluations?
- **Multi-State Pooled Testing:** Could collaborative, multi-state testing frameworks reduce the substantial financial and personnel burdens placed on individual state DOTs by pooling resources and sharing validation results?
- **Standardizing Data as a Physical Commodity:** In traditional highway construction, structural commodities such as steel, aggregate, and concrete undergo mandatory, standardized testing regimes to ensure compliance before field utilization. Developing an analogous standardized quality assurance program for commercial traffic data is essential as probe-derived metrics increasingly guide capital allocation, safety evaluations, and operational decisions.

5.4 Summary

The issues outlined in this section represent the current operational challenges of probe-based traffic volume estimation. While commercial datasets have demonstrated sufficient maturity for planning-level applications and initial HPMS submittals, long-term success requires addressing ground-truth independence, refining sensor error tolerances, and instituting institutionalized data-quality governance.