Independent Evaluation of a Probe-Based Method to Estimate Annual Average Daily Traffic Volume

FINAL – September 11, 2021

Publication No. FHWA-PL-21-032

September 2021

U.S. Department of Transportation
Federal Highway Administration

NOTICE

QUALITY ASSURANCE STATEMENT

Technical Report Documentation Page

1. Report No. FHWA-PL-21-032	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle INDEPENDENT EVALUATION OF A PROBE-BASED METHOD TO ESTIMATE ANNUAL AVERAGE DAILY TRAFFIC VOLUME		5. Report Date September 2021 6. Performing Organization Code
7. Author(s) Ioannis “Yianni” Tsapakis, Shawn Turner, Pete Koeneman, and Paul Anderson		8. Performing Organization Report No.
9. Performing Organization Name and Address Texas A&M Transportation Institute, College Station, TX 77843-3135 Cambridge Systematics, Inc. (Prime Contractor)		10. Work Unit No. (TRAIS) 11. Contract or Grant No. 693JJ319D000007 (Master)
12. Sponsoring Agency Name and Address U.S. Department of Transportation Federal Highway Administration Office of Highway Policy Information 1200 New Jersey Ave SE Washington, DC 20590		13. Type of Report and Period Covered Final Report: September 2019 - September 2021 14. Sponsoring Agency Code

15. Supplementary Notes

Project performed in cooperation with the U.S. Department of Transportation (DOT), Federal Highway Administration and a Technical Advisory Committee (TAC) consisting of Alaska DOT & Public Facilities, California DOT, Colorado DOT, Georgia DOT, Idaho DOT, Illinois DOT, Maryland DOT, Minnesota DOT, Nebraska Dept. of Roads, New Jersey DOT, North Carolina DOT, North Dakota DOT, Ohio DOT, Oregon DOT, Pennsylvania DOT, South Carolina DOT, Texas DOT, and Virginia DOT.
Steven Jessberger (Task Manager)

16. Abstract

This report documents an independent evaluation of annual average daily traffic (AADT) estimates developed from traffic probes by StreetLight Data, Inc (StL). The Texas A&M Transportation Institute (TTI) researchers performed a systematic process to identify accurate permanent continuous counter sites maintained by public agencies and private entities that could be used as benchmark sites against which to evaluate these AADT estimates. Ultimately, though, TTI could only use 215 bidirectional (552 directional) continuous count sites owned and maintained by state departments of transportation.

TTI researchers used several evaluation measures and several statistical hypothesis tests to assess the accuracy and precision of StreetLight Data’s AADT estimates. These multiple evaluation measures and hypothesis tests produced mixed results. Some statistical tests provided contradictory or counterintuitive results while other statistical hypothesis tests indicated that StL’s AADT estimates are statistically the same as the benchmark AADT values. TTI used professional judgment to weight and consider each evaluation result, and concluded that StL’s AADT estimates are valid for traffic monitoring use on roads with bidirectional AADT of 5,000 or greater vehicles per day. Because the evaluation results were mixed and the validity decision was not clear cut, TTI recommends phased implementation that includes two elements: 1) a one-to-three year transition period (DOTs still collecting short duration counts but comparing to StL AADT estimates) should be used to confirm TTI's professional judgment based on these mixed evaluation results; 2) a pilot implementation program initiated by FHWA that includes up to five early adopter state DOTs, to monitor and coordinate findings in early-adopting states.

17. Key Word Annual average daily traffic, AADT, independent evaluation, probe-based traffic monitoring, traffic count accuracy, traffic count precision		18. Distribution Statement No restrictions.	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 79	22. Price

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

TABLE OF CONTENTS

TABLE OF CONTENTS ... ii
LIST OF FIGURES ... iii
LIST OF TABLES ... iv
LIST OF ABBREVIATIONS ... vi
ACKNOWLEDGMENTS ... vii
INTRODUCTION ... 1
  Objective of the Overall Pooled Fund Study ... 1
  Objective and Organization of this Report ... 2
METHODS ... 3
  Define Independent Evaluation Process ... 3
  Identify Possible Evaluation Sites ... 4
  Send Full List of Possible Evaluation Sites to StL ... 8
  Reviewing and Processing Data from Possible Evaluation Sites ... 13
  Check Technology Accuracy at Possible Evaluation Sites ... 14
  Integrate StL's AADT Estimates with Benchmark Data at Final Evaluation Sites ... 19
  Conduct Analysis ... 19
RESULTS ... 29
  Exploratory Data Analysis ... 29
  Metrics and Summary Descriptive Statistics ... 38
  Statistical Testing Methods ... 43
CONCLUSIONS AND RECOMMENDATIONS ... 50
  Conclusions on Methods to Evaluate Non-Traditional Sources of AADT Estimates ... 50
  Conclusions on the Validity of StL AADT Estimates ... 50
  Recommendations for Next Steps ... 53
APPENDIX ... 54
  Exploratory Data Analysis ... 54
  Metrics and Summary Descriptive Statistics ... 58
  Statistical Testing Methods ... 65

LIST OF FIGURES

Figure 1. FHWA Targets for AADT Evaluation Sites.......... 5
Figure 2. Example: Translating a Centerline-Based Bi-Direction Evaluation Site to Two Directional OSM Segments.......... 9
Figure 3. Example: Translating a Centerline-Based Bi-Direction Evaluation Site to a Bi-Directional OSM Segment.......... 10
Figure 4. Nine Precision Targets Plotted Against AADT.......... 22
Figure 5. Scatterplot of TCEs Against Observed AADT for Six States (CA, ME, MN, NJ, OR, TX).......... 23
Figure 6. Non-Linear Trendlines Representing the Upper and Lower Precision Bounds.......... 24
Figure 7. Scatterplot (Observed AADT vs TCE) and Precision Bounds (Red Dotted Lines) – 286 Directional AADT Records in Texas.......... 25
Figure 8. Scatterplot (Observed AADT vs. Estimated AADT) for 286 Directional AADT Records in Texas.. 27
Figure 9. Scatterplot: Bidirectional AADT Data from Six States (CA, ME, MN, NJ, OR, TX).......... 31
Figure 10. Scatterplot: Directional AADT Data from Six States (CA, ME, MN, NJ, OR, TX).......... 32
Figure 11. Histogram of Bidirectional AADT Difference (States: CA, ME, MN, NJ, OR, TX).......... 33
Figure 12. Normal Q-Q Plot of Bidirectional AADT Difference (=Estimated AADT – Observed AADT) Data (States: CA, ME, MN, NJ, OR, TX).......... 33
Figure 13. TCE Frequency Histogram – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX).......... 34
Figure 14. Normal Q-Q Plot of Bidirectional TCE Data (States: CA, ME, MN, NJ, OR, TX).......... 35
Figure 15. Scatterplot: TCE vs. Observed AADT – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX).......... 36
Figure 16. Boxplots: TCE by Volume Group – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX) 37
Figure 17. Boxplots: TCE by Volume Group and State – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX).......... 37
Figure 18. Example of Precise But Overestimated AADT Estimates.......... 45
Figure 19. Histogram of Directional AADT Difference (States: CA, ME, MN, NJ, OR, TX).......... 54
Figure 20. Normal Q-Q Plot of Directional AADT Difference (=Estimated AADT – Observed AADT) Data (States: CA, ME, MN, NJ, OR, TX).......... 55
Figure 21. TCE Frequency Histogram – Directional AADT Data (States: CA, ME, MN, NJ, OR, TX).......... 55
Figure 22. Normal Q-Q Plot of Directional TCE Data (States: CA, ME, MN, NJ, OR, TX).......... 56
Figure 23. Scatterplot: TCE vs. Observed AADT – Directional AADT Data (States: CA, ME, MN, NJ, OR, TX).......... 56
Figure 24. Boxplots: TCE by Volume Group – Directional AADT Data (States: CA, ME, MN, NJ, OR, TX)... 57
Figure 25. Boxplot: TCE by Volume Group and State – Directional AADT Data (States: CA, ME, MN, NJ, OR, TX).......... 57

LIST OF TABLES

Table 1. States Selected by TTI in Which to Evaluate StL AADT Estimates ... 5
Table 2. Distribution of Possible Evaluation Sites in TTI's Initial Request to StL ... 12
Table 3. Accuracy Assessment of Doppler Radar for Use as Evaluation Sites ... 15
Table 4. Accuracy Assessment of Inductance Loops in California for Use as Evaluation Sites ... 16
Table 5. Accuracy Assessment of Side Fire Radar in Houston, Texas for Use as Evaluation Sites ... 17
Table 6. Accuracy Assessment of Video Machine Vision in Dallas, Texas for Use as Evaluation Sites ... 18
Table 7. Example of Summary Results for 286 Directional AADT Records in Texas ... 26
Table 8. Descriptive Statistics for All TTI Evaluation Sites (6 states) ... 29
Table 9. Basic Statistics of Observed AADT and Estimated AADT by State and AADT Directionality ... 30
Table 10. Normality Test Results (Target Variable = AADT Difference) (States: CA, ME, MN, NJ, OR, TX) ... 34
Table 11. Normality Test Results (Target Variable = TCE) by State ... 35
Table 12. Summary Statistics by Volume Group – Bidirectional AADT (States: CA, ME, MN, NJ, OR, TX) ... 39
Table 13. Summary Statistics by Volume Group – Directional AADT (States: CA, ME, MN, NJ, OR, TX) ... 39
Table 14. Summary Statistics by State – Bidirectional AADT (States: CA, ME, MN, NJ, OR, TX) ... 40
Table 15. Summary Statistics by State – Directional AADT (States: CA, ME, MN, NJ, OR, TX) ... 40
Table 16. Summary Statistics by State and Volume Group – Bidirectional AADT (States: CA, ME, MN, NJ, OR, TX) ... 41
Table 17. Summary Statistics by State and Volume Group – Directional AADT (States: CA, ME, MN, NJ, OR, TX) ... 42
Table 18. Summary Results by Four Volume Groups – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX) ... 46
Table 19. Summary Results by Four Volume Groups – Directional AADT Data (States: CA, ME, MN, NJ, OR, TX) ... 47
Table 20. Summary Results by Ten Volume Groups – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX) ... 47
Table 21. Summary Results by Ten Volume Groups – Directional AADT Data (States: CA, ME, MN, NJ, OR, TX) ... 48
Table 22. Summary Results by State – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX) ... 48
Table 23. Summary Results by State – Directional AADT Data (States: CA, ME, MN, NJ, OR, TX) ... 49
Table 24. Summary Statistics for Ten Volume Groups – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX) ... 58
Table 25. Summary Statistics for Ten Volume Groups – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX) ... 59
Table 26. Summary Statistics by Rural/Urban Designation – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX) ... 59
Table 27. Summary Statistics by Rural/Urban Designation – Directional AADT Data (States: CA, ME, MN, NJ, OR, TX) ... 60
Table 28. Summary Statistics by Roadway Functional Class – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX) ... 60
Table 29. Summary Statistics by Roadway Functional Class – Directional AADT Data (States: CA, ME, MN, NJ, OR, TX) ... 60
Table 30. Summary Statistics by Roadway Functional Class and Rural/Urban Designation – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX) ... 61
Table 31. Summary Statistics by Roadway Functional Class and Rural/Urban Designation – Directional AADT Data (States: CA, ME, MN, NJ, OR, TX) ... 62

Table 32. Summary Statistics by State and Rural/Urban Designation – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX) 63
Table 33. Summary Statistics by State and Rural/Urban Designation – Directional AADT Data (States: CA, ME, MN, NJ, OR, TX) 64
Table 34. Summary Results by State and Volume Group – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX) 65
Table 35. Summary Results by State and Volume Group – Directional AADT Data (States: CA, ME, MN, NJ, OR, TX) 66
Table 36. Summary Results by Rural/Urban Designation – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX) 67
Table 37. Summary Results by Rural/Urban Designation – Directional AADT Data (States: CA, ME, MN, NJ, OR, TX) 67
Table 38. Summary Results by State and Rural/Urban Designation – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX) 67
Table 39. Summary Results by State and Rural/Urban Designation – Directional AADT Data (States: CA, ME, MN, NJ, OR, TX) 68
Table 40. Summary Results by Roadway Functional Class – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX) 68
Table 41. Summary Results by Roadway Functional Class – Directional AADT Data (States: CA, ME, MN, NJ, OR, TX) 69
Table 42. Summary Results by Roadway Functional Class and Rural/Urban Code – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX) 69
Table 43. Summary Results by Roadway Functional Class and Rural/Urban Code – Directional AADT Data (States: CA, ME, MN, NJ, OR, TX) 70

LIST OF ABBREVIATIONS

AADT	Annual average daily traffic
ADT	Average daily traffic
CCS	Continuous count station
DOT	Department of Transportation
EDA	Exploratory data analysis
FHWA	Federal Highway Administration
GPS	Global positioning system
GIS	Geographic information system
HPMS	Highway Performance Monitoring System
ITS	Intelligent transportation system
LBS	Location-based service
LRS	Linear referencing system
MAPE	Mean absolute percent error
MedAPE	Median absolute percent error
NPS	National Park Service
NREL	National Renewable Energy Laboratory
NRMSE	Normalized root mean square error
OSM	Open Street Map
Signed Rank	S-R
StL	StreetLight Data, Inc.
TMAS	Travel Monitoring Analysis System
TTI	Texas A&M Transportation Institute
VMT	Vehicle-miles of travel
VPD	Vehicles per day

ACKNOWLEDGMENTS

The authors acknowledge this project's Technical Advisory Committee members, who provided feedback and comments throughout this research project:

- Scott Vockeroth, Alaska DOT & Public Facilities
- Afrid Sarker, California DOT
- Steve Abeyta, Colorado DOT
- Steven Jessberger, FHWA
- Eric Conklin, Georgia DOT
- John Phillips, Idaho DOT
- William (Bill) Morgan, Illinois DOT
- Carole Delion, Maryland DOT
- Gene Hicks, Minnesota DOT
- David Schoenmaker, Nebraska Dept. of Roads
- Chris Zajac, New Jersey DOT
- Kent Taylor, North Carolina DOT
- Terry Woehl, North Dakota DOT
- Dave Gardner, Ohio DOT
- Josh Roll, Oregon DOT
- Greg Dunmire, Pennsylvania DOT
- Todd Anderson, South Carolina DOT
- Chris Didear/Laura Dablain, Texas DOT
- Hamlin Williams, Virginia DOT

INTRODUCTION

For decades, motor vehicle traffic volumes have been monitored by government agencies that install data collection equipment on or near the roadways of interest. The traffic monitoring technology has evolved over the decades, from pneumatic air hoses to electric inductance, piezoelectric, magnetic, radar, infrared, video, etc. However, traffic volumes have always been collected by designated staff who must visit each roadway link and install either permanent or portable traffic data collection equipment.

Probe-based methods of collecting traffic data have been tested since the 1990's as an alternative to traditional methods of installing specialized data collection equipment. These probe-based methods mainly rely on travelers' mobile devices within the traffic stream to act as probes and report various data about the traffic stream, like current location, heading, and speed. The increasing ubiquity of smartphones with global positioning systems (GPS) that provide location-based services (LBS) has enabled many revolutionary features, like the ability to see real-time traffic speeds or how busy a local restaurant/store is relative to normal conditions (i.e., Google Traffic and Google Popular Times).

In the past few years, several universities and data analytics companies have developed algorithms to estimate total traffic volumes from a sample of traffic probes. These algorithms typically use multiple inputs to expand the sample of traffic probes to the total estimated traffic volumes. Continuous traffic count stations (CCSs) are used to calibrate and validate these algorithms. Other non-traffic data, like census, land use, and weather may also be used to fine-tune the probe sample expansion. These probe-based algorithms are being offered as a safer and more cost-effective approach to traffic monitoring, instead of sending designated staff to visit roadway sites and install traffic monitoring equipment.

Objective of the Overall Pooled Fund Study

The Federal Highway Administration (FHWA) recognized the need to evaluate these non-traditional, probe-based methods of traffic monitoring, and organized a pooled fund study for that explicit purpose. FHWA initiated Pooled Fund Study Number TPF-5(384) (https://pooledfund.org/Details/Study/636), and 18 state departments of transportation (DOTs) joined in contributing to this pooled fund study.

From the study documents, the objective of this overall study is to:

> *"…develop and deploy methods and approaches to obtain vehicle volume and classification data with passively collected data. Volume data refers to the annual average daily traffic (AADT) for all vehicles (both passenger and trucks) covering all roadway functional classes by traffic link or finer levels of segmentation with emphasis on functional classes of minor arterials, collectors, and local roads. Volume data on high volume urban interstates is also highly desired as there is a greater risk for collecting this data in these environments because maintenance of traffic is more expensive and these activities can disrupt normal traffic patterns."*

The FHWA selected Streetlight Data, Inc. (StL) as the data provider who will develop algorithms and provide traffic volume estimates. The FHWA selected two independent evaluators to assess the accuracy and precision of StL's traffic volume estimates, the Texas A&M Transportation Institute (TTI) (as a subcontractor to Cambridge Systematics, Inc.) and the National Renewable Energy Laboratory (NREL).

Objective and Organization of this Report

The objective of this report is to document the independent evaluation conducted by TTI. Within this report, TTI documents the evaluation sites and methods used, the results, and the conclusions reached by their independent evaluation.

The rest of this report is organized as follows:

- **METHODS**: Describes the statistical comparison methods used to evaluate the accuracy and precision of StL's AADT estimates.

- **RESULTS**: Describes the results of TTI's evaluation of StL's AADT estimates.

- **CONCLUSIONS AND RECOMMENDATIONS**: Summarizes the conclusions about the suitability and technology-readiness of StL's probe-based traffic monitoring methods to estimate AADT and provides recommendations about next steps.

- **APPENDIX**: Contains various technical information that was deemed too detailed for inclusion in the main body of this report.

METHODS

This chapter documents the statistical comparison methods that TTI used to evaluate the accuracy and precision of StL's 2019 AADT estimates. The chapter is divided into the following sections, which were also the major tasks in the independent evaluation process:

- Define Independent Evaluation Process
- Identify Possible Evaluation Sites
- Send Full List of Possible Evaluation Sites to StL
- Check Technology Accuracy at Possible Evaluation Sites
- Integrate StL's AADT Estimates with Benchmark Data at Final Evaluation Sites
- Conduct Statistical Analysis

Define Independent Evaluation Process

There are several possible ways to evaluate the accuracy and precision of AADT estimates. In early project discussions, StL indicated that they were conducting a self-evaluation of their AADT estimates using repeated cross-validation, a method that uses a repeated process of dividing available data into separate training and testing datasets.[1]

Therefore, TTI chose a traditional "blind" evaluation to compare StL 2019 AADT estimates to trusted count data at various locations. The evaluation is considered "blind" because StL presumably does not know the true AADT value at each of the trusted count data locations.

Although simple in theory, a blind evaluation of AADT estimates at a national scale is challenging for two main reasons:

1. **Historical AADT estimates are publicly available (i.e., not blind)**. Most state DOTs make their own AADT estimates publicly available through their website or a geographic information system (GIS) clearinghouse. State DOTs also report their AADT estimates to FHWA in the Highway Performance Monitoring System (HPMS), which is then made publicly available (https://www.fhwa.dot.gov/policyinformation/hpms/shapefiles.cfm). Therefore, anyone with internet access has access to nationwide AADT estimates for past years, and historical AADT estimates are not "blind".

2. **The number of high-quality AADT estimates are limited nationally and many are also publicly available**. The most accurate AADT estimates typically come from a limited number of permanent continuous monitoring stations (currently about 6,000) installed and maintained to a high standard by state DOTs. Because this traffic data from permanent stations is so valuable, state DOTs use it widely and make it publicly available, and some even stream the traffic counts in near real-time (e.g., https://www.scdot.org/travel/travel-trafficdata.aspx).

TTI and FHWA took several steps to minimize the possibility that StL would know the true value of 2019 AADT estimates at designated evaluation locations.

---

[1] More information on repeated cross-validation can be found in Burman 1989 (https://iri.columbia.edu/~tippett/cv_papers/Burman1989.pdf).

1. **Delay public release of 2019 AADT data**. The FHWA Office of Highway Policy Information delayed the public release of 2019 traffic count data in their Travel Monitoring Analysis System, or TMAS (https://www.fhwa.dot.gov/policyinformation/tables/tmasdata/) until after StL had provided 2019 AADT estimates.

2. **Request StL to withhold certain sites where they already had access to AADT data**. StL had MS2 (a traffic data integration software company) as a subcontractor to provide easy digital access to traffic data in three states for AADT model development (i.e., New Hampshire, North Carolina, and Texas). Therefore, TTI and FHWA asked StL to certify that they would withhold several hundred sites for TTI and NREL independent testing. Later in the evaluation, TTI verified that the results from these "virtually blind" test locations were comparable to the truly "blind" locations where StL did not have access to 2019 AADT data.

3. **Identify states where StL did not have easy access to AADT data through MS2**. As an extra precaution to the above step, TTI also identified several state DOTs that did not use MS2's traffic monitoring software to ensure that StL did not have easy access to 2019 AADT data in mid- to late-2020. Therefore, several state DOTs not using the MS2 traffic data system were given extra consideration when identifying states in which to test.

4. **Identify non-traditional monitoring sites where StL is unlikely to have access.** These non-traditional sites included a wide variety of sources: traffic signal control sensors, toll facilities, ITS or operations-based sensors, national park entrance monitoring, and even dynamic speed limit feedback signs installed at some school zone boundaries. TTI recognized that traffic counts from non-traditional sources would have to go through a quality and accuracy assurance process before they were deemed trusted sources. This process is discussed in the next two sections.

5. **Request AADT estimates at extra sites that won't be tested to minimize any manual StL intervention**. At the beginning of the project, FHWA gave TTI a target of 5,000 road segments at which AADT was to be evaluated, to be distributed across different traffic volume ranges, road functional classes, geographic areas, and climate zones. Therefore, TTI planned to send StL about 15,000 road segments for which StL was to provide AADT estimates, but TTI would only test 5,000 of those 15,000 road segments. Therefore, StL would not know exactly which segments were the actual test segments, and the quantity of segments (in the tens of thousands with TTI and NREL test sites combined) is too large for manual intervention by StL (i.e., the AADT estimates are output automatically by StL algorithm, without manual StL intervention).

Identify Possible Evaluation Sites

Once the blind evaluation structure was established in the early stages of the study, TTI started to identify possible evaluation sites that met several desired criteria (see Figure 1):

- At least five states;
- At least three distinct geographic areas;
- At least two distinct climate zones; and,
- All seven FHWA road functional classes.

Functional Class	State 1	State 2	State 3	State 4	State 5
	(At least 5 states to include a minimum of 3 distinct geographic areas and at least 2 distinct climate zones)				
Interstate					
Other Freeways & Expressways					
Other Principal Arterials					
Minor Arterials					
Major Collectors					
Minor Collectors					
Local Roads					

Include a minimum of 6 different local agency-owned roadways

Minimum of 5,000 unique roadway segments distributed among all matrix cells

Figure 1. FHWA Targets for AADT Evaluation Sites

FHWA established a target for TTI of 5,000 trusted evaluation sites among at least five states. However, most state DOTs only maintain several hundred CCSs. It was obvious that TTI would have to identify non-traditional traffic monitoring locations in these five states if the goal of 5,000 trusted evaluation sites was to be achieved. Therefore, the possible availability of non-traditional traffic monitoring locations also became a criterion in selecting the preferred states in which to evaluate StL AADT estimates.

After considering several of the required and desired criteria described above, TTI identified six states as shown in Table 1, with their corresponding attributes. Note that each evaluator was required to select five states, nonetheless, TTI added a sixth state to the list (Table 1) to increase the chances of finding a higher number of possible ground truth sites.

Table 1. States Selected by TTI in Which to Evaluate StL AADT Estimates

State	Geographic Area	Climate Zone	Other Important Criteria
California	West Coast	Varied	Many ITS sensors, not MS2 state
Maine	New England	Cold/snow/ice	Not MS2 state
Minnesota	Upper Midwest	Cold/snow/ice	Not MS2 state, ITS sensors
New Jersey	Mid-Atlantic	Moderate, some snow/ice	Many DOT & ITS sensors
Oregon	Pacific Northwest	Coastal, damp	ITS sensors
Texas	South Central	Varied, hot, coastal	Many ITS sensors

Since NREL was conducting a parallel, second independent evaluation of StL AADT estimates, FHWA requested that TTI and NREL select a single overlap state, where both TTI and NREL would each conduct their own evaluation. After some discussion, TTI and NREL selected California as the overlap state to be included in both TTI's and NREL's independent evaluation.

Once these six states were identified, TTI pursued various options for non-traditional traffic monitoring locations in these states for evaluating StL AADT estimates. The previous section outlined these non-traditional sources, and the following subsections provide more detailed discussion about the consideration of these non-traditional sources as possible evaluation sites.

Traffic signal control sensors

Most traffic signals use sensors for motor vehicle presence detection, to adapt the signal timing to better reflect real-time traffic demands. Some newer traffic signal systems can count motor vehicles (in addition to simple presence detection) traveling through a signalized intersection. Some of these newer traffic signal systems also can save the motor vehicle counts from the sensors.

TTI pursued the availability of 2019 traffic count data from traffic signal control sensors as possible evaluation sites for StL AADT estimates. TTI identified two specific equipment manufacturers that offer the ability to count motor vehicles and store those counts in a database. In back-and-forth discussions with these companies, TTI concluded that this was not a feasible option for two reasons: 1) the number of installed systems with these counting and storage features in 2020 was very low; and 2) the number of installed systems with complete count data for 2019 (the evaluation year for StL AADT estimates) was even lower (many of their systems were installed some time in 2019).

Toll facilities

Toll road authorities charge a user fee for each vehicle traveling on their roads. Therefore, for revenue and accounting purposes, toll road authorities should have a very accurate count of motor vehicle traffic at specified locations on all toll roads under their jurisdiction.

TTI pursued the availability of permanent continuous traffic counts on various toll roads in the six test states. After extended back-and-forth communication with several toll road authorities, TTI concluded that toll road test sites were not feasible due to lack of responsiveness and very low number of test sites in proportion to total test sites needed to meet FHWA targets.

ITS or operations-based sensors

Many city, regional and state transportation departments have installed traffic sensors in the past two to three decades for real-time traffic information, management, and operations. DOTs typically install these operations-based sensors on high-volume roads in urban areas that are typically congested; however, some DOTs have installed these sensors on important intercity or interstate routes in rural areas, although typically at less frequent intervals than in urban areas.

In the late 1990s and early 2000s, FHWA encouraged these agencies to save this valuable real-time traffic data for historical uses in planning, through the Archived Data User Service (ADUS) program. Therefore, in 2020, most agencies do save real-time traffic data from ITS-based sensors. However, these agencies primarily use these sensors for traffic speed measurement, and numerous studies have documented the lack of sensor calibration for accurate traffic counting. Also, data completeness has been documented as a known issue for ITS-based traffic data.

With these cautions in mind, TTI identified hundreds of ITS sensors in California, Minnesota, New Jersey, Oregon, and Texas that could possibly serve as evaluation sites pending further review. More details on the consideration of ITS sensors as possible evaluation sites are included in the next major section.

National Park Service (NPS) visitation monitoring stations

The NPS is mandated by Congress to report visitation to national parks on an annual basis. At large parks with high visitation, NPS uses permanent inductance loop detectors to monitor entering motor vehicles at selected entrance stations to estimate total park visitors.

TTI contacted the traffic monitoring group within the NPS and pursued the availability of 2019 traffic count data from NPS roads. The NPS was in the process of refurbishing and conducting quality assurance procedures at selected monitoring stations and offered 33 sites where trusted traffic count data was available for the entire year of 2019. However, only 2 of those 33 sites were in the 6 states that TTI had chosen for evaluation (see Table 1). Therefore, TTI did not further consider these NPS sites in the evaluation.

Dynamic speed limit feedback signs

Several companies manufacture a dynamic speed limit feedback sign for speed management purposes. A dynamic speed limit feedback sign is designed to measure the speed of approaching vehicles with doppler radar and then posts the measured speed below the speed limit sign. If the measured speed exceeds the posted speed limit by a designated threshold, then the sign will flash the measured speed or provide other warning signs to the approaching driver to slow down. These signs are often used at entrance to school zones, for traffic calming purposes during school hours. These signs are also often used on lower functional class roadways, which would help to significantly bolster the number of possible test sites on collector and minor arterial streets.

TTI and NREL contacted a particular sign manufacturer that had advertised the availability of historical traffic count data (i.e., in 2019, only one of the seven known manufacturers had traffic count data available) from their dynamic speed limit feedback signs. After several discussions, TTI was able to identify a significant number of sign locations that could possibly serve as evaluation sites pending further review. More details on the consideration of these doppler radar sensors as possible evaluation sites are included in the next major section.

Short duration count sites

While identifying possible evaluation sites, it became clear that TTI would not be able to meet the FHWA target of 5,000 unique roadway segments distributed among all matrix cells (Figure 1). In particular, TTI could not identify enough permanent monitoring sites in lower functional classes. After extended discussion with FHWA and NREL, the consensus among all three groups was for TTI and NREL to identify short duration count sites as possible evaluation sites to meet the FHWA targets set in Table 1.

There were several concerns about using short duration count sites to evaluate StL AADT estimates. Most importantly, AADT estimates from short duration counts can have error from several sources: 1) data collection from portable equipment; 2) assignment of a site to a factor group; and 3) factoring short duration count to an annual estimate. These possible sources of error could prevent a short duration count site from being a “ground truth,” and if large differences exist between these short duration AADT estimates and StL AADT estimates, TTI would not conclusively know if the error was in the short duration AADT estimate or the StL estimate.

To mitigate two of these sources of error (i.e., the assignment and factoring error), TTI considered the comparison of the unfactored daily traffic estimates from short duration count sites to daily traffic estimates on the exact same days from StL. Essentially, the measured average daily traffic (ADT) from a state DOT's short duration count would be used to evaluate an ADT estimate from StL. TTI also considered this approach to be less than desirable because the differences in this comparison would be for estimated error in ADT, not AADT. Random positive and negative errors in ADT could cancel each other over the course of a year, resulting in AADT error that is less than average ADT error.

In TTI's initial submittal of evaluation sites to StL in September 2020 (see next section for more detail), TTI did request AADT and ADT values from several thousand short duration count sites. However, because of several lingering concerns, FHWA, TTI and NREL decided in early 2021 to <u>NOT</u> include any short duration count sites in the evaluation of StL estimates. Therefore, tables in the next section include short duration count sites for accuracy and completeness, but these short duration count sites were not analyzed and are not included in the Results chapter.

Send Full List of Possible Evaluation Sites to StL

By mid-2020, TTI had identified about 5,000 evaluation sites (including short duration count sites) at which StL AADT estimates could <u>possibly</u> be evaluated. To ensure that the permanent sites being considered as evaluation sites were very accurate (and could serve as a ground truth), FHWA, TTI and NREL decided to conduct technology validation tests at selected permanent sites (discussed in detail in the next section). However, the timing of these technology validation spot checks experienced large delays due to the pandemic, and therefore TTI proceeded with submitting a full list of possible evaluation sites to StL in September 2020. Once the technology validation results were available, TTI could make final decisions about which possible evaluation sites could be used in the final evaluation analysis. The following paragraphs describe this process of submitting possible evaluation sites to StL in more detail.

<u>Use of Filler Sites</u>

As mentioned earlier, TTI requested AADT estimates at extra sites that would not be tested (i.e., filler sites) to minimize any manual StL intervention during generation of AADT estimates. Therefore, TTI tripled the number of evaluation sites from the target of 5,000 to nearly 15,000 road segments for eventual submittal to StL. TTI kept internal records of which sites were evaluation sites and which sites were filler sites. When StL returned AADT estimates for the full list of about 15,000 road segments, TTI simply removed the filler sites from the evaluation process.

<u>Use of OSM Network to Identify Possible Evaluation Sites</u>

There was extensive discussion by all groups (i.e., FHWA, StL, TTI, NREL) about how to accurately and unambiguously identify the evaluation sites where AADT estimates were desired for evaluation. Most state DOTs use their own linear referencing system (LRS) in GIS to identify their count sites. In some cases (e.g., ITS or doppler radar sites), the possible evaluation sites were also referenced by a latitude-longitude pair. In other cases, the possible evaluation sites were referenced using a route number and distance from cross-street or landmark (e.g., US 66 2 miles east of Main Street).

However, StL uses OSM as their base network for AADT estimates. Therefore, there could be ambiguity with complex road geometry if TTI provided DOT LRS points or latitude-longitude pairs to StL, and these location references were translated to OSM incorrectly by StL. To minimize any possible error and ambiguity with location references for possible evaluation sites, TTI translated all location references for possible evaluation sites to OSM, for direct consumption by StL in their AADT estimation model. Figure 2 and Figure 3 show examples of how TTI translated the various DOT location references to OSM location references in GIS.

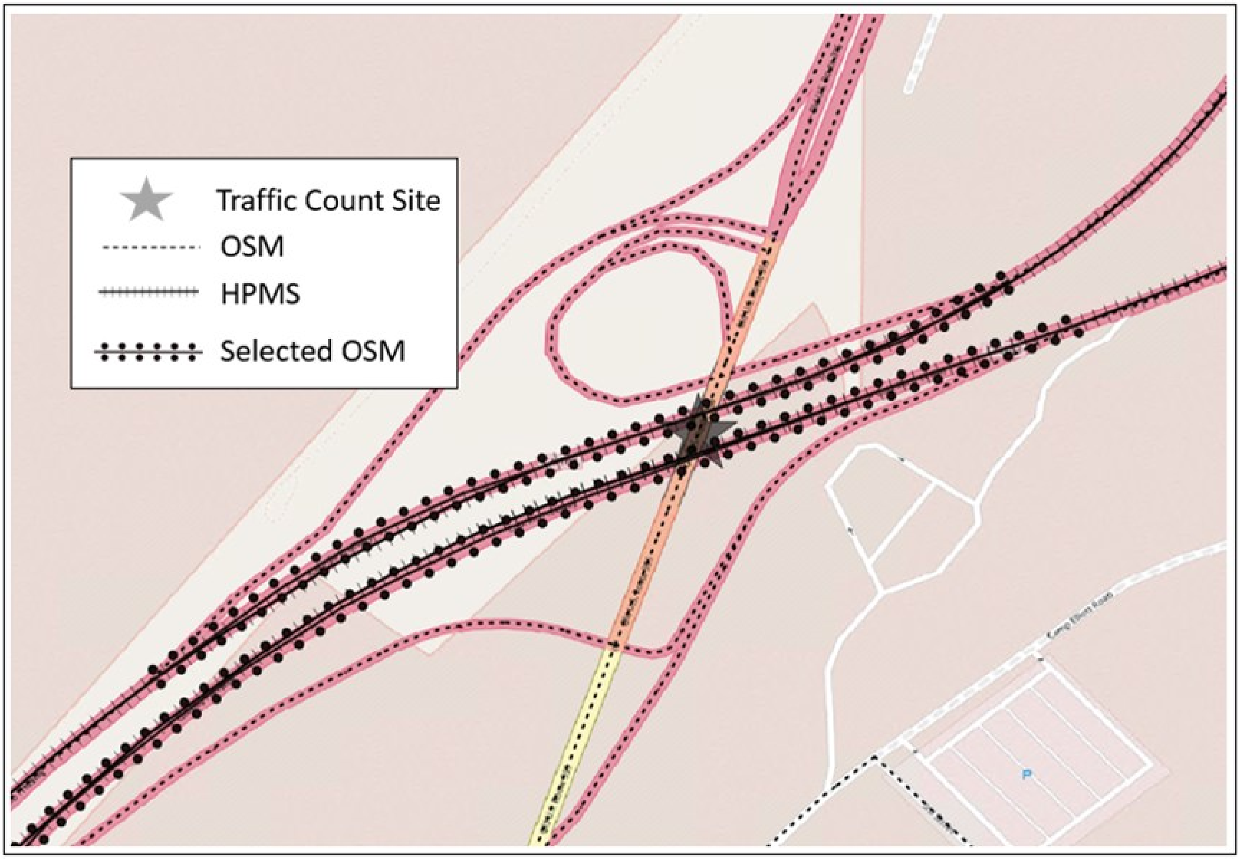

Figure 2. Example: Translating a Centerline-Based Bi-Direction Evaluation Site to Two Directional OSM Segments

Figure 3. Example: Translating a Centerline-Based Bi-Direction Evaluation Site to a Bi-Directional OSM Segment

One of the challenges that TTI faced in translating location references is that OSM is a directional roadway network (i.e., each travel direction is represented by a single unique segment), whereas in some state DOT LRSs a single segment (e.g., centerline) may represent both directions of travel. Further, a single pair of geographical coordinates is typically provided for each CCS, regardless of whether the road (where the CCS is located) is divided or undivided. In other cases (like ITS monitoring sites), the DOT location references were to a specific travel direction at the site, and the other travel direction was monitored some distance upstream or downstream.

Because of these location referencing challenges, TTI identified each unique directional OSM road segment (also known as link) that was associated with each evaluation site. At some evaluation sites (like traditional DOT CCSs), there were two directional OSM road segments identified as being associated with that evaluation site. At other evaluation sites (typically ITS sites), only a single directional OSM road segment was associated with that evaluation site. TTI then provided the full list of directional OSM road segments to StL for automated generation of AADT estimates in their models.

Use of a Two-Phased OSM Segment Request to StL

Because of possible errors in the automated process of matching possible evaluation sites to OSM segments, TTI used a two-phased request to StL:

1. In the initial request, TTI sent a full list of possible evaluation sites referenced to specific directional OSM segments. StL then returned AADT estimates for each of the requested directional OSM segments. TTI then reviewed these AADT estimates as compared to the benchmark values, identified segments with large differences in AADT, and then double-checked to ensure that TTI had matched to the correct OSM segment.
2. In some cases, TTI's semi-automated map-matching process identified the incorrect directional OSM segment, and the mismatch was obvious when reviewing large differences in AADT. In these cases, a closer review by TTI identified a corrected directional OSM segment. Then, in a second request/phase, TTI sent to StL a much smaller list of corrected directional OSM segments.

TTI's initial request to StL was made on September 7, 2020 and included 11,815 directional OSM segments (see Table 2). StL provided AADT estimates to TTI for these OSM road segments on October 28, 2020. TTI reviewed the initial comparisons, performed quality assurance in November and early December, and provided a second request to StL on December 14, 2020. This second request had 140 segments that were corrected from TTI's first request, and StL provided AADT estimates for the segments on this second request on December 23, 2020.

StL had been working on improvements and enhancements to their AADT model in late December 2020 and January 2021 and requested FHWA permission to submit revised AADT estimates to TTI in late January 2021. FHWA agreed to accept revised AADT estimates, and StL delivered revised AADT estimates on February 7, 2021.

Table 2. Distribution of Possible Evaluation Sites in TTI's Initial Request to StL

State	Source of Possible Benchmark Counts	Number of Sites (1 site=1 travel direction)
California	California DOT CCS	513
	ITS	305
	Doppler Radar	205
	National Park Service	2
	Short Duration Counts	79
	State Subtotal	**1,104**
Maine	Maine DOT CCS	67
	Doppler Radar	22
	Short Duration Counts	1,763
	State Subtotal	**1,852**
Minnesota	Minnesota DOT CCS	147
	Doppler Radar	54
	Short Duration Counts	3,911
	State Subtotal	**4,112**
New Jersey	New Jersey DOT CCS	187
	ITS	91
	Doppler Radar	151
	Short Duration Counts	1,077
	State Subtotal	**1,506**
Oregon	Oregon DOT CCS	185
	ITS	19
	Doppler Radar	12
	State Subtotal	**216**
Texas	Texas DOT CCS	639
	ITS	1,827
	Short Duration Counts	218
	State Subtotal	**3,025**
All States Combined	Permanent Installations	4,767
	Short Duration Counts	7,048
	All States Combined Total	**11,815**

Note: Although requested from StL, short duration counts were not used to evaluate StL AADT estimates.

Reviewing and Processing Data from Possible Evaluation Sites

TTI obtained data for possible evaluation sites from the sources listed earlier in this section. These data generally consisted of hourly volumes for all of 2019, although the formats and processing needs varied by source. In terms of format, some datasets provided hourly volumes by lane, others contained hourly counts by direction, and a small number of sites provided only an hourly bidirectional count.

Some datasets (such as TMAS) were provided in raw form and TTI performed both quality control and completeness checks, as described below. Others, such as ITS datasets, had already gone through the data provider's QC process and were only checked for completeness. In some datasets, hours with incomplete data were missing, while in other cases the data provider replaced incomplete or rejected data with imputed values so that all hours were present. Datasets with imputed values, such as ITS sites, typically included a field that could be used to distinguish observed from imputed values.

Quality Control

Quality control on the TMAS dataset was performed according to the following FHWA-recommended criteria:

1. 7 or more consecutive hours with 0 volume
2. 100-fold change in volume from one hour to the next
3. Hour with volume of 50 or greater adjacent to an hour with 0 volume
4. Hourly volume of greater than 3000 vehicles per lane
5. Directional split greater than 60-40 at the daily level

Any records meeting at least one of these conditions were flagged and removed.

Completeness

Each possible evaluation site was checked for data completeness. Before performing this check, all imputed and partially observed hours were removed.

Benchmark AADTs were calculated using the current FHWA method, which is described below. To apply this method, it was necessary to have at least one observation for each hour in each day of the week in each of the 12 months in 2019. Any directional site that did not meet this data availability condition was removed from further consideration. The completeness check was performed separately for each direction, so it was possible to remove just one direction of a bidirectional site (e.g., due to a sensor malfunction).

AADT Calculation

TTI calculated AADT for all of the evaluation sites using FHWA's method for averaging hourly counts (Federal Highway Administration 2016b, pp. 1-7). The formulas for this method are reproduced below.

First, MADT values are calculated using the following expression:

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

Then, the MADTs are aggregated to obtain AADT using the following expression:

$$AADT = \frac{\sum_{m=1}^{12} d_m * MADT_m}{\sum_{m=1}^{12} d_m} \quad (2)$$

Where:

$AADT$ = Annual Average Daily Traffic

$MADT_m$ = Monthly Average Daily Traffic in month m

VOL_{ihjm} = total traffic volume for the *i*th occurrence of the *h*th hour of day within *j*th day of week during the *m*th month.

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

h = hour of the day ($h = 1, \ldots, 24$) or other temporal interval

j = day of the week ($j = 1, \ldots, 7$)

m = month (m$= 1, \ldots, 12$)

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 or 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 particular year

Check Technology Accuracy at Possible Evaluation Sites

As indicated earlier in this section, TTI was considering using AADT values from non-traditional traffic monitoring sites (e.g., ITS and doppler radar) to evaluate AADT estimates. Because these non-traditional sites were primarily used to measure speeds and not volumes, TTI was concerned about whether these non-traditional sites were calibrated for accurate vehicle counting.

Therefore, TTI and NREL jointly conducted technology validation tests at selected non-traditional sites that were being considered as possible evaluation sites. These technology validation spot checks suffered delay due to the pandemic conditions in mid-2020, and therefore TTI proceeded with submitting a full list of possible evaluation sites to StL in September 2020. Once the technology validation results were available in late 2020 and early 2021, TTI could make final decisions about which possible evaluation sites could be used in the final evaluation analysis.

The following subsections describe the technology validation tests that were used to ensure that evaluation sites were generating accurate AADT values for StL AADT estimate validation.

Establishing Video Analytics as Benchmark

To assess the accuracy of non-traditional permanent monitoring sites, TTI had to first identify a portable counter equipment technology that had been established as accurate enough to serve as a benchmark. This portable counter benchmark could then be moved around to a representative sample of non-traditional monitoring sites being considered as possible evaluation sites.

TTI conducted a state-of-the-practice review by searching available literature and conferring with numerous traffic monitoring staff at state DOTs. TTI's review identified one model of portable video collection equipment that, when combined with automated machine vision, provided very accurate (e.g., typically less than 2-3 percent average error when no occlusion is present) traffic counts.

Once a specific portable video collection model had been identified, NREL contracted with a commercial traffic data collector to use this specific model of video collection equipment to collect 48 hours of traffic counts at a representative sample of non-traditional sites. TTI then used these video-based counts to determine the accuracy at specific technology types in several different locations.

Evaluating Doppler Radar Sites

Early in the project, TTI identified 837 possible evaluation sites among the six states that used doppler radar in dynamic speed limit feedback signs. These doppler radar sites were typically located near school zones and/or on lower functional class sites and could be a valuable addition to the ensure representative evaluation sites, especially on lower volume roadways. Therefore, TTI and NREL conducted technology accuracy tests at several doppler radar sites.

NREL evaluated almost all of the doppler radar sites and their findings are included in NREL's final report to FHWA. TTI did evaluate count accuracy in each direction at two doppler radar sites in Houston, Texas. Each of these two doppler radar sites had two travel lanes in each direction. The results are shown in Table 3 and indicate that the doppler radar installations at these two sites significantly undercount traffic (by 21 percent). The cause for undercounting was unknown, but TTI thought it was likely due to occlusion. The NREL findings were similar and indicated that the doppler radar sites were not accurate enough to be used as evaluation sites. The doppler radar sites were then excluded from further evaluation.

Table 3. Accuracy Assessment of Doppler Radar for Use as Evaluation Sites

Site Name and Travel Direction	Day #	Daily Count from Doppler Radar	Daily Count from Video Benchmark	Error in Doppler Radar
Hammerly EB	1	6,587	8,146	-19%
Hammerly EB	2	6,646	8,146	-18%
Hammerly WB	1	5,842	8,189	-29%
Hammerly WB	2	5,621	8,123	-31%
Kempwood EB	1	4,570	5,473	-16%
Hammerly EB	2	4,688	5,633	-17%
Kempwood WB	1	4,572	5,636	-19%
Kempwood WB	2	4,610	5,707	-19%
All Tested Sites	-	-	-	**-21%** (Average daily count error)

Evaluating ITS-based Sites

Early in the project, TTI identified 2,242 possible evaluation sites among the six states that used ITS-based sensors. The ITS-based sensors included three basic technology types: inductance loops, video machine vision, and side fire radar. Most of these ITS-based sites are in Texas and are located on high-volume congested roadways, typically Interstate roadways and freeways. Because of the much greater total number of ITS-based sensors that could serve as possible evaluation sites, TTI and NREL conducted technology accuracy tests at numerous ITS-based sites to represent all three technologies in use.

Inductance Loops in Orange County, California

TTI evaluated the accuracy of two inductance loop detector sites in Orange County, California over a 48-hour period. The results are shown in Table 4 and indicate mixed results, with one site consistently within 3 percent of the benchmark and the other site being 6 to 8 percent higher than the video benchmark. The average count error across all tested sites was 5 percent, just on the cusp of what is considered acceptable as a benchmark to evaluate StL AADT estimates.

Table 4. Accuracy Assessment of Inductance Loops in California for Use as Evaluation Sites

Site Name and Travel Direction	Date	Count Subtotal from Inductance Loop	Daily Count from Video Benchmark	Error in Doppler Radar
I-5 SB at Chapman	5/5/21	80,827	76,250	+6%
I-5 SB at Chapman	5/6/21	136,975	128,383	+7%
I-5 SB at Chapman	5/7/21	47,773	44,078	+8%
CA 22 WB at Valley	5/5/21	35,355	34,371	+3%
CA 22 WB at Valley	5/6/21	60,192	58,663	+3%
CA 22 WB at Valley	5/7/21	8,551	8,376	+2%
All Tested Sites	-	-	-	**+5%** (Average count error)

Side Fire Radar in Houston, Texas

TTI evaluated the accuracy of 15 side fire radar sites in Houston, Texas over a 48-hour period. The results are shown in Table 5 and show fairly good results, with about half of the errors being within 3 percent of the video benchmark. The mean absolute count error was 5 percent, just on the cusp of what is considered acceptable as a benchmark to evaluate StL AADT estimates. The worst-performing side fire radar sensors had error greater than 10 percent, which is not acceptable for StL evaluation purposes.

Table 5. Accuracy Assessment of Side Fire Radar in Houston, Texas for Use as Evaluation Sites

Site Name and Travel Direction	Date	Count Subtotal from Side Fire Radar	Daily Count from Video Benchmark	Error in Side Fire Radar
Beltway 8 EB @ Mesa	10/7/2020	27,278	29,968	-9%
	10/8/2020	27,474	30,489	-10%
Beltway 8 WB @ Mesa	10/7/2020	30,673	33,372	-8%
	10/8/2020	32,642	35,639	-8%
I-10 EB @ UP Railroad	10/6/2020	116,937	119,119	-2%
	10/7/2020	118,626	120,959	-2%
I-10 WB @ Dairy-Ashford	9/30/2020	107,975	115,411	-6%
	10/1/2020	104,494	117,813	-11%
I-10 WB @ UP Railroad	10/6/2020	117,651	120,000	-2%
	10/7/2020	122,464	124,347	-2%
I-45 NB @ Fuqua	9/29/2020	77,962	83,499	-7%
	9/30/2020	79,660	88,063	-10%
I-45 SB @ Fuqua	9/29/2020	77,888	85,078	-8%
	9/30/2020	79,528	86,709	-8%
I-45 SB @ W Mount Houston	10/6/2020	94,698	104,535	-9%
	10/7/2020	95,062	105,230	-10%
I-610 NB @ Beechnut	10/6/2020	84,776	81,895	+4%
	10/7/2020	86,299	83,874	+3%
TX 288 NB @ TX 6	9/29/2020	36,427	36,828	-1%
	9/30/2020	37,083	37,270	-1%
TX 288 SB @ TX 6	9/29/2020	33,719	35,143	-4%
	9/30/2020	36,521	37,068	-1%
TX 99 NB @ Cinco Ranch	9/29/2020	40,259	40,355	0%
	9/30/2020	41,141	40,948	0%
TX 99 SB @ Cinco Ranch	9/29/2020	41,680	42,308	-1%
	9/30/2020	43,048	43,698	-1%
US 90 EB @ Runneburg Rd	10/7/2020	10,414	9,296	+12%
	10/8/2020	10,706	9,448	+13%
US 90 WB @ Runneburg Rd	10/7/2020	9,669	9,416	+3%
	10/8/2020	12,965	12,496	+4%
All Tested Sites	-	-	-	**5%** (Mean absolute count error)

Video Machine Vision in Texas

TTI evaluated the accuracy of 8 side fire radar sites in Houston, Texas over a 48-hour period. The results are shown in Table 6 and show fair to poor results, with the mean absolute count error at 14 percent. The two best daily counts were at 3 and 5 percent error; however, all other daily counts had errors greater than 5 percent. Therefore, TTI concluded that these video-based sensors were not accurate enough for StL evaluation purposes.

Table 6. Accuracy Assessment of Video Machine Vision in Dallas, Texas for Use as Evaluation Sites

Site Name and Travel Direction	Date	Count Subtotal from ITS Video	Daily Count from Video Benchmark	Error in ITS Video
I-35 SB @ US 380 Upstream	9/9/2020	43,951	28,965	52%
	9/10/2020	38,027	32,749	16%
I-35 SB @ US 380 Downstream	9/9/2020	35,111	28,965	21%
	9/10/2020	31,184	32,749	-5%
I-35E NB @ Inwood Upstream	9/9/2020	112,657	94,495	19%
	9/10/2020	105,584	100,584	5%
I-35E NB @ Inwood Downstream	9/9/2020	112,150	94,495	19%
	9/10/2020	103,566	100,584	3%
I-35E SB @ Inwood Upstream	9/9/2020	102,928	95,731	8%
	9/10/2020	94,977	106,788	-11%
I-35E SB @ Inwood Downstream	9/9/2020	106,503	95,731	11%
	9/10/2020	99,024	106,788	-7%
US 75 NB @ Meadow	9/9/2020	101,571	100,920	+1%
	9/10/2020	87,201	116,346	-25%
US 75 SB @ Meadow	9/9/2020	97,471	91,991	+6%
	9/10/2020	86,143	102,496	-16%
All Tested Sites	-	-	-	**14%** (Mean absolute count error)

<u>Removal of Possible Evaluation Sites that were Deemed Not Usable</u>

Based on the various technology accuracy assessments shown in Table 3 through Table 6, TTI decided NOT TO INCLUDE any of the doppler radar sites NOR any of the ITS-based sites. Simply put, the results were mixed at the sample of sites, and TTI concluded that we could not trust that these non-traditional monitoring sites to provide highly accurate benchmark traffic counts for evaluating StL AADT estimates.

Integrate StL's AADT Estimates with Benchmark Data at Final Evaluation Sites

In the last step before statistical analysis, TTI integrated the StL AADT estimates with other site metadata associated with each evaluation site. The result was a combined database that had the following attributes:

- State name
- Unique evaluation site identifier
- Roadway functional class at evaluation site
- Rural or urban designation
- Code for single travel direction or bi-directional
- Cardinal direction for single travel direction
- Latitude-longitude pair at evaluation site
- StL AADT estimate
- TTI-calculated AADT value from benchmark data
- 4-category volume group based on benchmark AADT (aggregate volume categories)
- 10-category volume group based on benchmark AADT (granular volume categories)

This database contains three types of evaluation sites:

- Training sites – These sites were used by StL to TRAIN their AADT estimation models.
- Testing sites – These sites were used by StL to TEST their AADT estimation models.
- Withheld sites – These sites were NOT used by StL.

This database then served as the foundation to perform the analyses described in the following sections. For completeness, TTI evaluated all three types of sites; however, this report shows only the results obtained from the evaluation of testing and withheld sites, which were treated as one group – these sites were NOT used by StL to develop their models.

Conduct Analysis

TTI conducted several analyses to understand the data, reveal general trends, and determine both the accuracy and precision of the 2019 AADT estimates provided by StL. The analyses included the following:

a) Conducted an exploratory data analysis (EDA).
b) Calculated a series of metrics and summary descriptive statistics that have been widely used in similar types of data evaluations.
c) Applied the following three statistical testing methods:
    - The method described in Appendix B of the FHWA document "*Guidelines for Obtaining AADT Estimates from Non-Traditional Sources*".[2] This method uses accuracy and precision targets developed from historical traffic data. For simplicity, this method is referred to hereinafter as the TPF-5(384) method.
    - A modified version of the TPF-5(384) method. This version was developed by TTI using the precision targets of the TPF-5(384) method.
    - The Wilcoxon signed-rank (S-R) hypothesis test.

[2] FHWA (2021), *Guidelines for Obtaining AADT Estimates from Non-Traditional Sources*. Unpublished document as of 6/7/2021.

These analyses are described in the three subsections that follow. The last subsection, called "Aggregation and Presentation of Results", describes how TTI sliced the data and aggregated the results at various levels using different variables (e.g., by state, volume group, functional class, rural/urban code, etc.).

Exploratory Data Analysis

TTI conducted an EDA that involved developing and reviewing frequency histograms, boxplots, scatterplots, and descriptive statistics. TTI also performed two normality tests, the Kolmogorov-Smirnov test and the Shapiro-Wilk test, to determine whether the errors are normally distributed. The results from the EDA helped the research team to develop a better understanding of the data, the underlying distributions, the relationship between the observed and the estimated AADT values, as well as to identify possible outliers and determine appropriate statistical methods for further examination. Like the metrics described in the previous section, the EDA and the normality tests were conducted multiple times for different groups of evaluation sites as explained in the last subsection "Aggregation and Presentation of Results".

Metrics and Summary Descriptive Statistics

TTI calculated a series of metrics and statistics that have been widely used in the past to quantify the accuracy of various traffic estimates, including AADT. For completeness, at the beginning of the analysis TTI calculated more than 20 metrics. The analysis revealed that some metrics result in similar trends and ultimately lead to similar conclusions. To avoid potential redundancy and keep this report concise, TTI selected to present the results for the following metrics that capture the most important trends and findings from the analysis:

- Mean and median algebraic difference (AD). The AD was initially calculated as the difference between the estimated AADT and the observed AADT at each evaluation site *i* as follows:

  $$AD_i\ (vehicles) = AADT_{Estimated,i} - AADT_{Observed,i} \tag{3}$$

  After calculating the ADs for all sites, the mean AD and the median AD were calculated for different groups of sites (e.g., by state, by volume group, by functional class, etc.), as explained in subsection "Aggregation and Presentation of Results".

- Mean and median traffic count error (TCE). The TCE, also known as percent error (PE), was initially calculated for each pair of observed AADT and estimated AADT at evaluation site *i* as follows:

  $$TCE_i\ or\ PE_i\ (\%) = \frac{AADT_{Estimated,i} - AADT_{Observed,i}}{AADT_{Observed,i}} \times 100 \tag{4}$$

  After calculating the TCEs for all sites, the mean TCE and the median TCE were calculated for different groups of sites, as explained in subsection "Aggregation and Presentation of Results".

- Mean and median absolute percent error (APE). The APE was initially calculated for each pair of observed AADT and estimated AADT at evaluation site *i* as follows:

  $$APE_i\ (\%) = \frac{\left|AADT_{Estimated,i} - AADT_{Observed,i}\right|}{AADT_{Observed,i}} \times 100 \tag{5}$$

After calculating the APEs for all sites, the mean APE (MAPE) and the median APE (MedAPE) were calculated for different groups of sites, as explained in subsection "Aggregation and Presentation of Results".

- Normalized root mean square error (NRMSE). The NRMSE was calculated for various groups of evaluation sites, as follows:

$$NRMSE = 100 * \frac{\sqrt{\frac{1}{n} * \sum_{i=1}^{n} \left(AADT_{Estimated,i} - AADT_{Observed,i}\right)^2}}{\frac{1}{n} * \sum_{i=1}^{n} \left(AADT_{Observed,i}\right)} \quad (6)$$

- Coefficient of determination (R^2). The R^2 was calculated between the observed AADT values and the estimated AADT values for each group of evaluation sites described in subsection "Aggregation and Presentation of Results".

- 68% and 95% TCE ranges. The 68% TCE range is the distance between the 16th and the 84th percentiles of the TCE. The 95% TCE range is the distance between the 2.5th and the 97.5th TCE percentiles.

Where:

SD_i	= signed difference for site *i*.
$AADT_{Estimated,\ i}$	= StLD AADT estimate for site *i*.
$AADT_{Observed,\ i}$	= observed AADT at site *i*.
TCE_i	= traffic count error for site *i*.
PE_i	= percent error for site *i*.
APE_i	= absolute percent error for site *i*.
NRMSE	= normalized root mean square error.

Statistical Testing Methods

TPF-5(384) Approach

Pooled-fund study TPF-5(384) developed a document titled *Guidelines for Obtaining AADT Estimates from Non-Traditional Sources*".[2] Appendix B of this document describes a method that can be used to determine whether the accuracy and precision of alternative AADT estimates meet certain (accuracy and precision) targets. The targets were developed based on data quality assessments conducted in the 2015 pooled-fund study TPF-5(292) *Assessing Roadway Traffic Count Duration and Frequency Impacts on AADT Estimations*[3]. TPF-5(292) determined the accuracy and precision of AADT data derived from sample short term counts that were extracted from permanent sites. The accuracy and precision targets were developed based on a 95% confidence interval (the significance level is 0.05) to judge the acceptability of alternative AADT methods. The TPF-5(384) method is appropriate for non-parametric data.

[3] FHWA (2015), Assessing Roadway Traffic Count Duration and Frequency Impacts on AADT Estimations. http://dev.www.pooledfund.org/Details/Study/534

Modified TPF-5(384) Approach

A modified TPF-5(384) method was developed by TTI based on the nine precision targets calculated in pooled-fund study TPF-5(292) and provided in FHWA document *Guidelines for Obtaining AADT Estimates from Non-Traditional Sources*.[2] The nine precision targets were developed for nine volume groups, respectively, as shown in Figure 4.

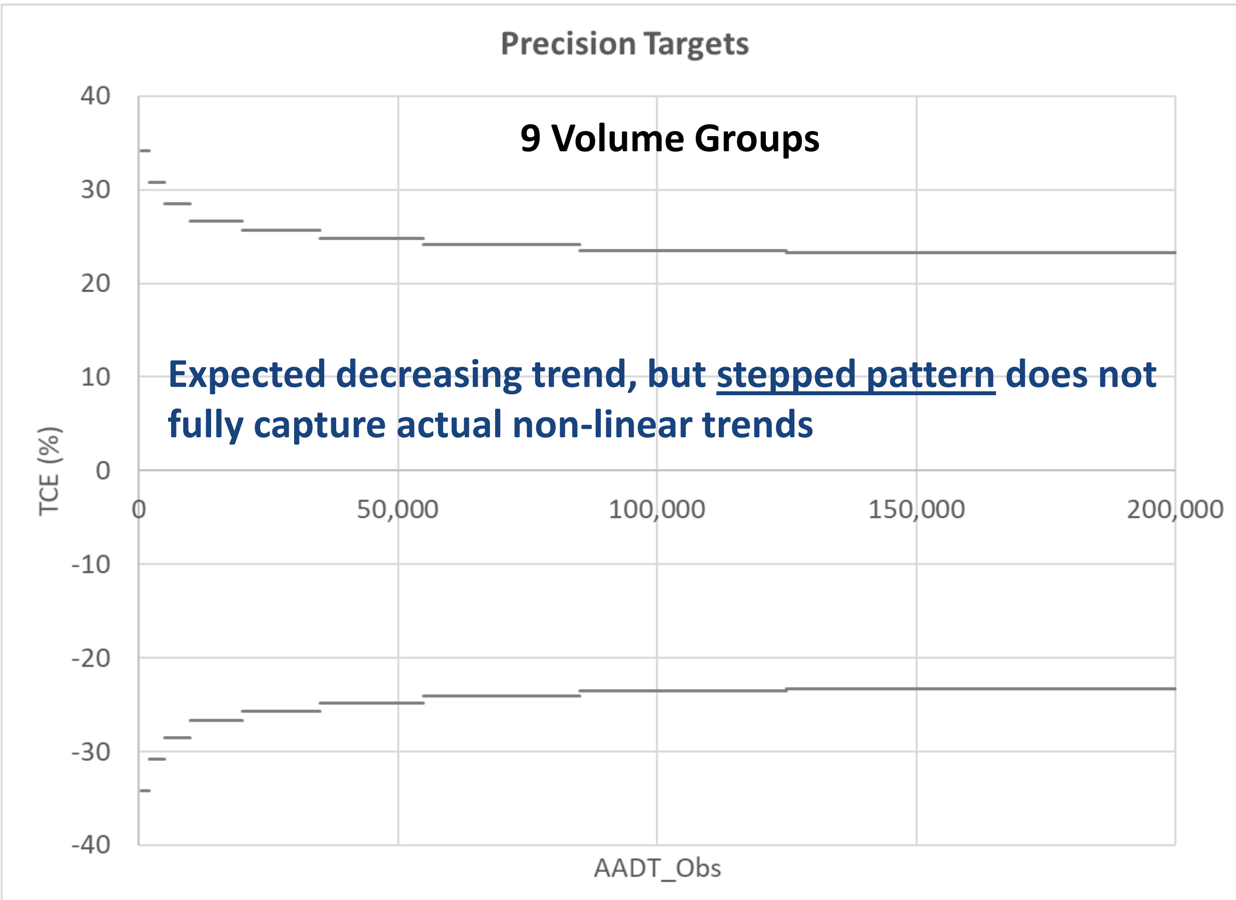

Figure 4. Nine Precision Targets Plotted Against AADT

In general, the vertical distance of the precision targets tends to decrease as AADT increases. This decreasing trend captures the actual trend that is observed in practice. Figure 5 shows the TCEs calculated using observed AADT and estimated AADT data from the six states (CA, ME, MN, NJ, OR, TX) that TTI selected at the beginning of the project. The two red dotted non-linear lines capture the general trends of the TCEs that tend to be less variable as AADT increases.

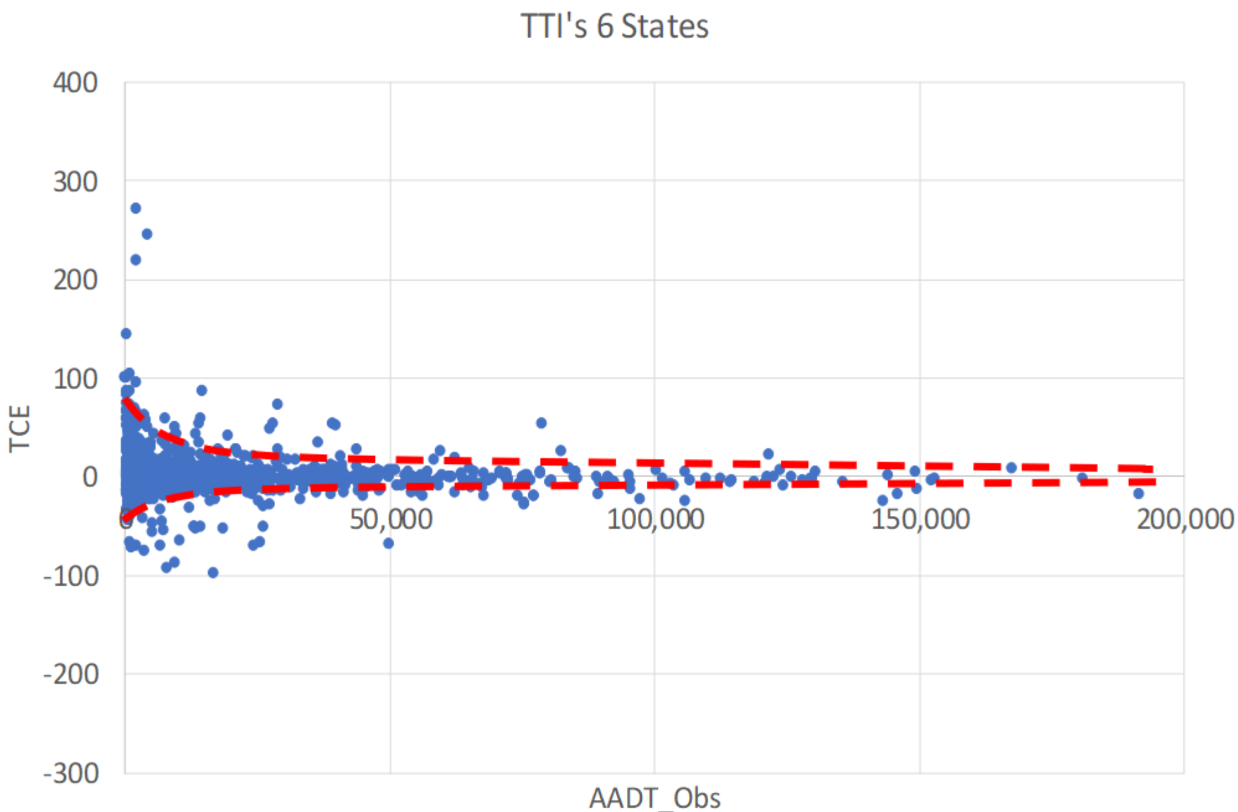

Figure 5. Scatterplot of TCEs Against Observed AADT for Six States (CA, ME, MN, NJ, OR, TX)

One caveat of the precision targets shown in Figure 4 is that they follow a stepped pattern, as opposed to the observed non-linear pattern that is illustrated in Figure 5. To address this shortcoming, TTI modified the TPF-5(384) method as follows:

- Extracted the upper midpoint and lower midpoint of each volume group (Figure 6). The two midpoints correspond to the upper and lower precision targets (of each volume group), respectively.
- Fitted two non-linear trendlines that represent the upper and lower precision bounds. The two trendlines are shown in Figure 7 along with their equation and R^2 value.
- Calculated the TCEs for any given group of evaluation sites (e.g., for a specific state) using equation 4.
- Determined both the number and the percent of points that fall within the 2 precision bounds.

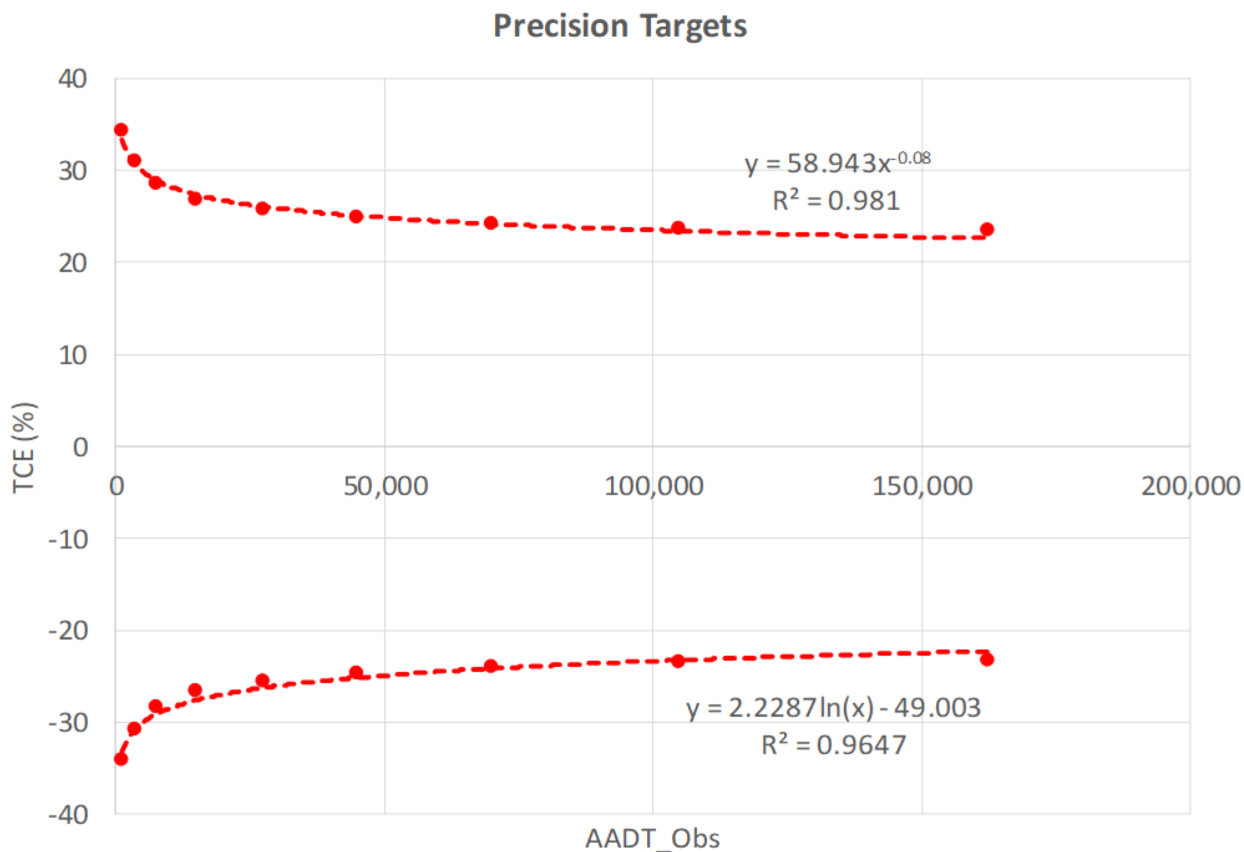

Figure 6. Non-Linear Trendlines Representing the Upper and Lower Precision Bounds

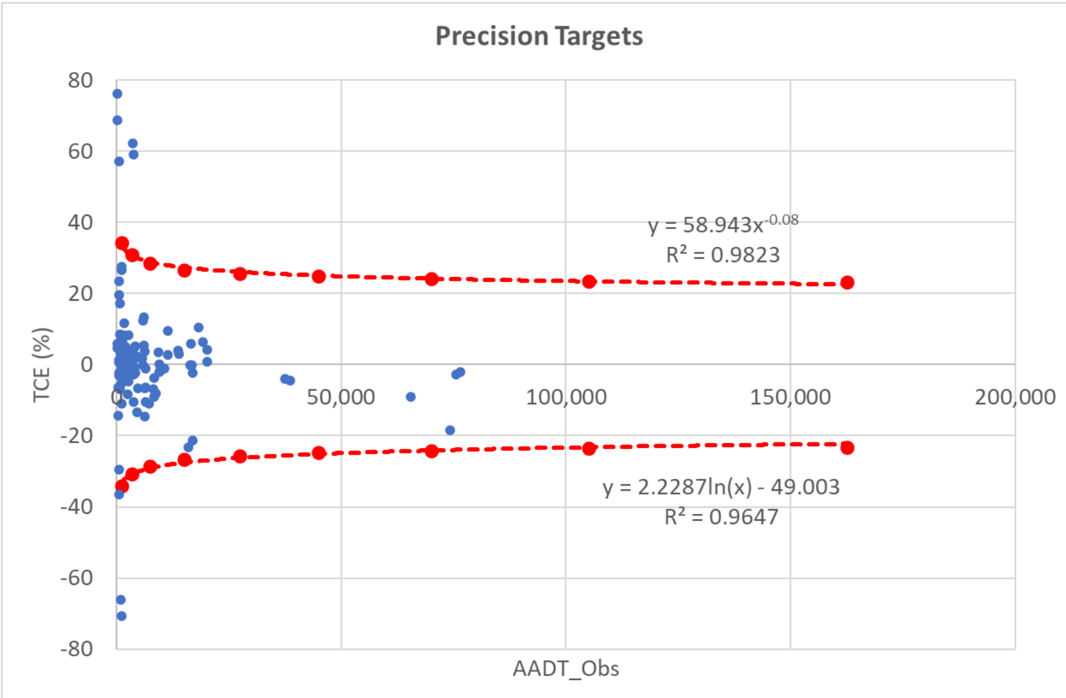

Figure 7. Scatterplot (Observed AADT vs TCE) and Precision Bounds (Red Dotted Lines) – 286 Directional AADT Records in Texas

TTI developed an Excel file that automates this analysis. The required data inputs are the observed AADT and estimated AADT values for a given group of sites. After users enter or simply copy/paste AADT data in columns A (observed AADT) and B (estimated AADT), the file automatically produces the following:

- A summary (Table 7) that shows the results of the analysis for four and ten volume groups. These groups correspond to the volume groups described in FHWA document "*Guidelines for Obtaining AADT Estimates from Non-Traditional Sources*".[2] Table 7 shows an example of a summary table that was produced using directional AADT data from Texas. In the last column, the cells highlighted in thick borderline indicate that the percent of points that fall within the 2 precision bounds is higher than 95%, which was selected in this example as a threshold. The file allows users to change this threshold, as needed. If the threshold is modified, all the results are automatically updated.
- A scatterplot between the observed AADT and the calculated TCEs (Figure 7). The same directional AADT data from Texas are shown in Figure 7.
- A scatterplot between the observed AADT and the estimated AADT (Figure 8). The same directional AADT data from Texas are shown in Figure 8.

Table 7. Example of Summary Results for 286 Directional AADT Records in Texas

Volume Group	Min	Max	N	N within Bounds	% within Bounds
10 Volume Groups					
1) 0–499	-	499	14	9	64%
2) 500–1,999	500	1,999	56	51	91%
3) 2,000–4,999	2,000	4,999	61	59	**97%**
4) 5,000–9,999	5,000	9,999	61	61	**100%**
5) 10,000–19,999	10,000	19,999	50	48	**96%**
6) 20,000–34,999	20,000	34,999	24	24	**100%**
7) 35,000–54,999	35,000	54,999	7	7	**100%**
8) 55,000–84,999	55,000	84,999	12	12	**100%**
9) 85,000–124,999	85,000	124,999	1	1	**100%**
10) 125,000+	125,000	1,000,000	0	0	Not Appl.
4 Volume Groups					
1) 0–499	-	499	14	9	64%
2) 500–4,999	500	4,999	117	110	94%
3) 5,000–54,999	5,000	54,999	142	140	**99%**
4) 55,000+	55,000	1,000,000	13	13	**100%**
Total			**286**	**272**	**95%**

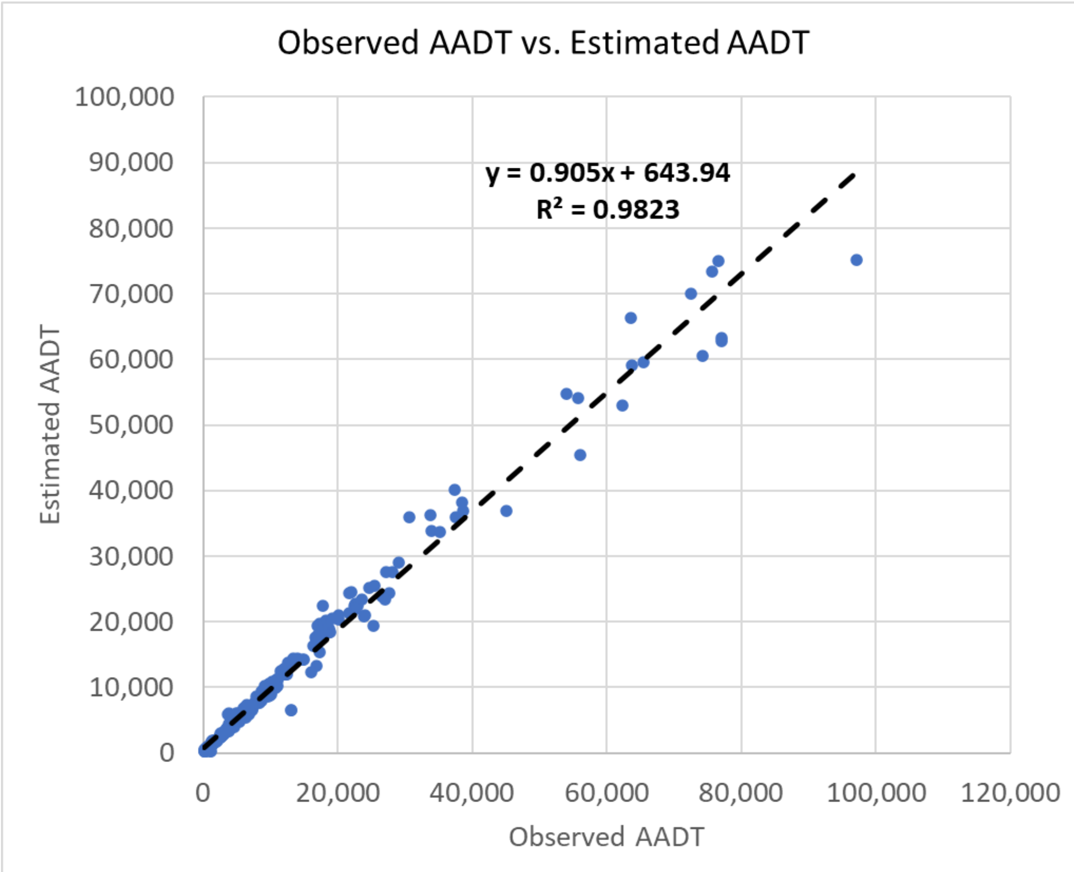

Figure 8. Scatterplot (Observed AADT vs. Estimated AADT) for 286 Directional AADT Records in Texas

Wilcoxon Signed-Rank Test

The Wilcoxon signed-rank test is a statistical hypothesis test appropriate for non-parametric data. It is used to compare two related matched samples (the estimated AADT and the observed AADT) and assess whether their population mean ranks are statistically different. The Wilcoxon signed-rank test considers information about both the sign and the magnitude of the differences between pairs (i.e., estimated AADT and observed AADT) and uses the standard normal distributed z-value to test of significance. One of the weaknesses of the Wilcoxon test is that it has little power in the case of small sample sizes. In fact, when the sample size is five or less, the Wilcoxon test will always result in a p value greater than 0.05, regardless of how far the sample median is from the hypothetical median.

Aggregation and Presentation of Results

All the analyses described in the previous sections were conducted by slicing the data and producing results at various aggregation levels using the variables described below:

- All sites from all states treated as one group.
- By state: CA, ME, MN, NJ, OR, TX.
- By rural/urban designation.
- By roadway functional class.
- By the following four volume groups:

    - 0–499 vpd
    - 500–4,999 vpd
    - 5,000–54,999 vpd
    - ≥55,000 vpd
- By the following ten volume groups:
    - 0–499 vpd
    - 500–1,999 vpd
    - 2,000–4,999 vpd
    - 5,000–9,999 vpd
    - 10,000–19,999 vpd
    - 20,000–34,999 vpd
    - 35,000–54,999 vpd
    - 55,000–84,999 vpd
    - 85,000–124,999 vpd
    - ≥125,000 vpd
- By state and the four volume groups listed above.
- By state and rural/urban designation.
- By roadway functional class and rural/urban designation.

One summary table was generated for each (first-level) bullet point listed above. In addition, each table was separately generated using bidirectional AADTs and directional AADTs:

a) <u>Bidirectional AADTs</u>. Each bidirectional AADT captures the total traffic volume in both directions of travel. A site can have only one bidirectional AADT value.
b) <u>Directional AADTs</u>. Each directional AADT captures the traffic volume in a single direction of travel. A site can have up to two directional AADT values.

RESULTS

This chapter presents the most important results of the analyses described in the previous chapter. The chapter includes analysis results from the six states (CA, ME, MN, NJ, OR, TX) that TTI selected at the beginning of the project. For consistency with the three types of analyses described in the previous chapter, the results are separately presented in the following subsections:

a) Exploratory data analysis
b) Metrics and summary descriptive statistics
c) Statistical testing methods

Exploratory Data Analysis

TTI initially explored the two variables of interest, the observed AADT and the estimated AADT (from StL). Table 8 shows descriptive statistics calculated using all TTI evaluation sites from all six states. It is worth noting that the skewness statistic is high[4] and the median AADT is significantly smaller than the mean AADT. These findings indicate that the AADT distributions are highly skewed to the right.

Table 8. Descriptive Statistics for All TTI Evaluation Sites (6 states)

Descriptive Statistic		Bidirectional AADT		Directional AADT		Bidirectional & Directional AADT	
		Observed AADT	Estimated AADT	Observed AADT	Estimated AADT	Observed AADT	Estimated AADT
N (Sample Size)		215	215	552	552	767	767
Mean		23,490	23,244	13,494	13,520	16,296	16,246
Std. Error of Mean		2,469.9	2,414.3	811.0	804.2	919.0	903.4
Median		10,075	10,427	6,171	6,026	7,035	6,601
Mode		762	399	297	201	297	201
Std. Deviation		36,216.5	35,401.3	19,054.1	18,894.8	25,451.4	25,020.6
Skewness		2.630	2.602	2.265	2.253	3.139	3.084
Std. Error of Skewness		0.166	0.166	0.104	0.104	0.088	0.088
Kurtosis		6.676	6.560	4.648	4.760	12.016	11.629
Std. Error of Kurtosis		0.330	0.330	0.208	0.208	0.176	0.176
Range		191,228	183,753	102,764	104,969	191,347	183,954
Minimum		237	399	118	198	118	198
Maximum		191,465	184,152	102,882	105,167	191,465	184,152
Percentiles	25	2,893	3,317	2,038	2,017	2,427	2,470
	50	10,075	10,427	6,171	6,026	7,035	6,601
	75	24,702	24,537	14,907	15,231	17,268	17,717

[4] As a rule of thumb if the skewness is:

- Between -0.5 and 0.5, the data are considered to be symmetrical.
- Between -1 and -0.5 or between 0.5 and 1, the data are moderately skewed.
- Less than -1 or greater than 1, the data are highly skewed.

Table 9 shows the mean, median, and standard deviation of the observed and the estimated AADT by state. In general, the mean and median AADT vary from one state to another. Note that in California and Maine, the directional AADTs are higher than the bidirectional AADTs. This can be explained by the fact that only one (high) directional AADT value was available for some evaluation sites.

Table 9. Basic Statistics of Observed AADT and Estimated AADT by State and AADT Directionality

State	AADT Directionality	Sample Size	Mean		Median		Stand. Deviation	
			Obs. AADT	Est. AADT	Obs. AADT	Est. AADT	Obs. AADT	Est. AADT
California	Bidirectional	9	19,562	23,762	14,122	14,407	14,240.4	20,827.1
	Directional	26	25,448	27,938	14,551	18,895	27,362.9	28,313.6
Maine	Bidirectional	5	5,683	5,558	2,787	2,680	5,886.9	5,396.3
	Directional	36	4,952	4,788	3,917	3,665	6,178.8	5,667.7
Minnesota	Bidirectional	4	7,009	5,139	7,104	4,280	3,240.0	2,254.2
	Directional	8	3,505	2,728	3,549	2,431	1,499.6	967.2
New Jersey	Bidirectional	4	30,285	23,225	24,225	23,342	13,034.3	5,402.0
	Directional	46	24,603	25,297	15,927	15,207	20,728.2	21,548.5
Oregon	Bidirectional	71	25,993	26,903	10,190	10,473	42,376.6	43,637.1
	Directional	150	14,669	15,031	5,322	5,447	22,346.5	22,726.2
Texas	Bidirectional	122	23,370	22,395	9,348	9,522	35,119.8	32,435.0
	Directional	286	11,359	10,924	6,041	5,939	15,884.1	14,504.8
Total	**Bidirectional**	**215**	**23,490**	**23,244**	**10,075**	**10,427**	**36,216.5**	**35,401.3**
	Directional	**552**	**13,494**	**13,520**	**6,171**	**6,026**	**19,054.1**	**18,894.8**

Figure 9 and Figure 10 show scatterplots combined with frequency histograms for bidirectional and directional AADT data, respectively. In general, there is a high number of evaluation sites in the low volume groups. This number gradually decreases as AADT increases. The R^2 in both scatterplots is high (≈0.97) suggesting a strong positive relationship between the estimated AADT and the observed AADT. Another observation is that the concentration of points is high in low-volume roads and tends to decrease along the horizontal axis as one moves from low- to high-volume roads.

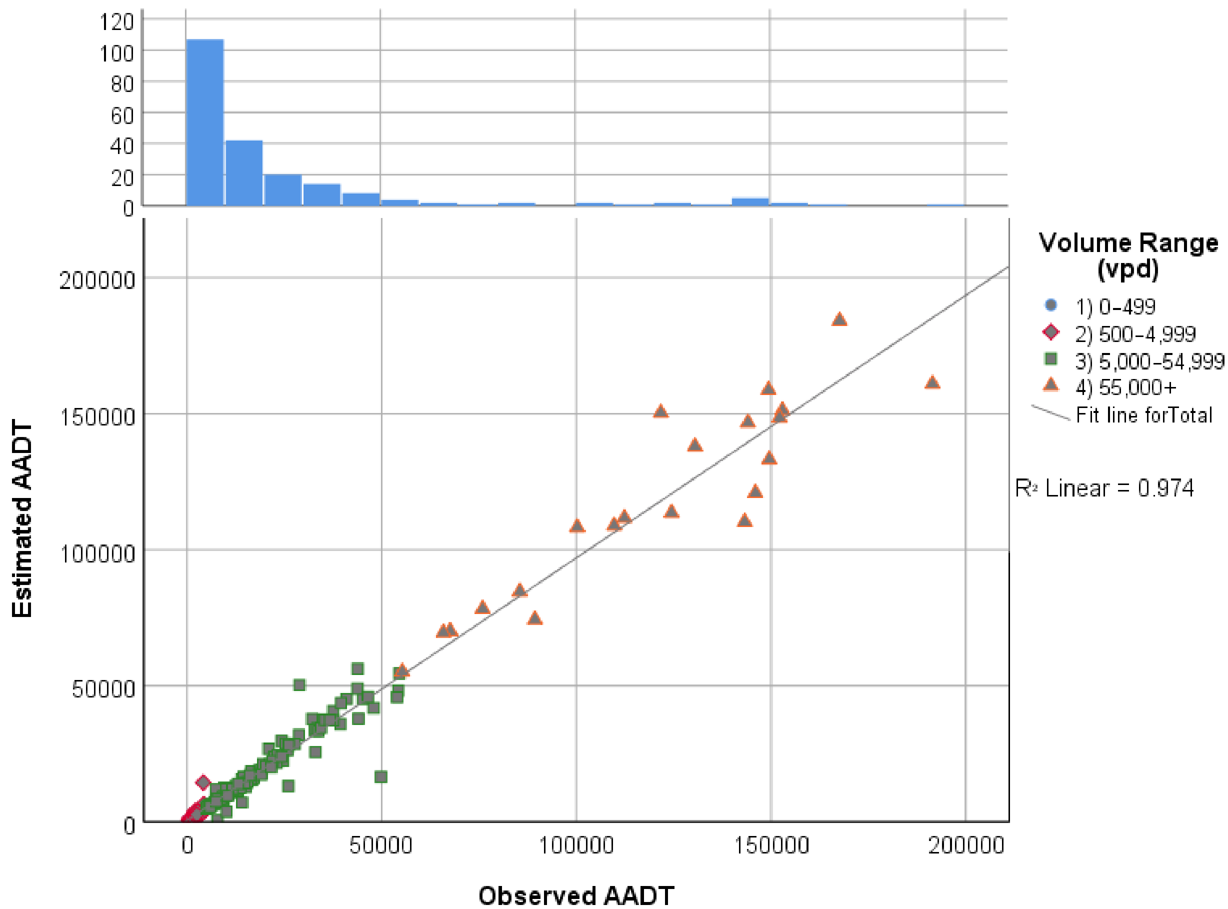

Figure 9. Scatterplot: Bidirectional AADT Data from Six States (CA, ME, MN, NJ, OR, TX)

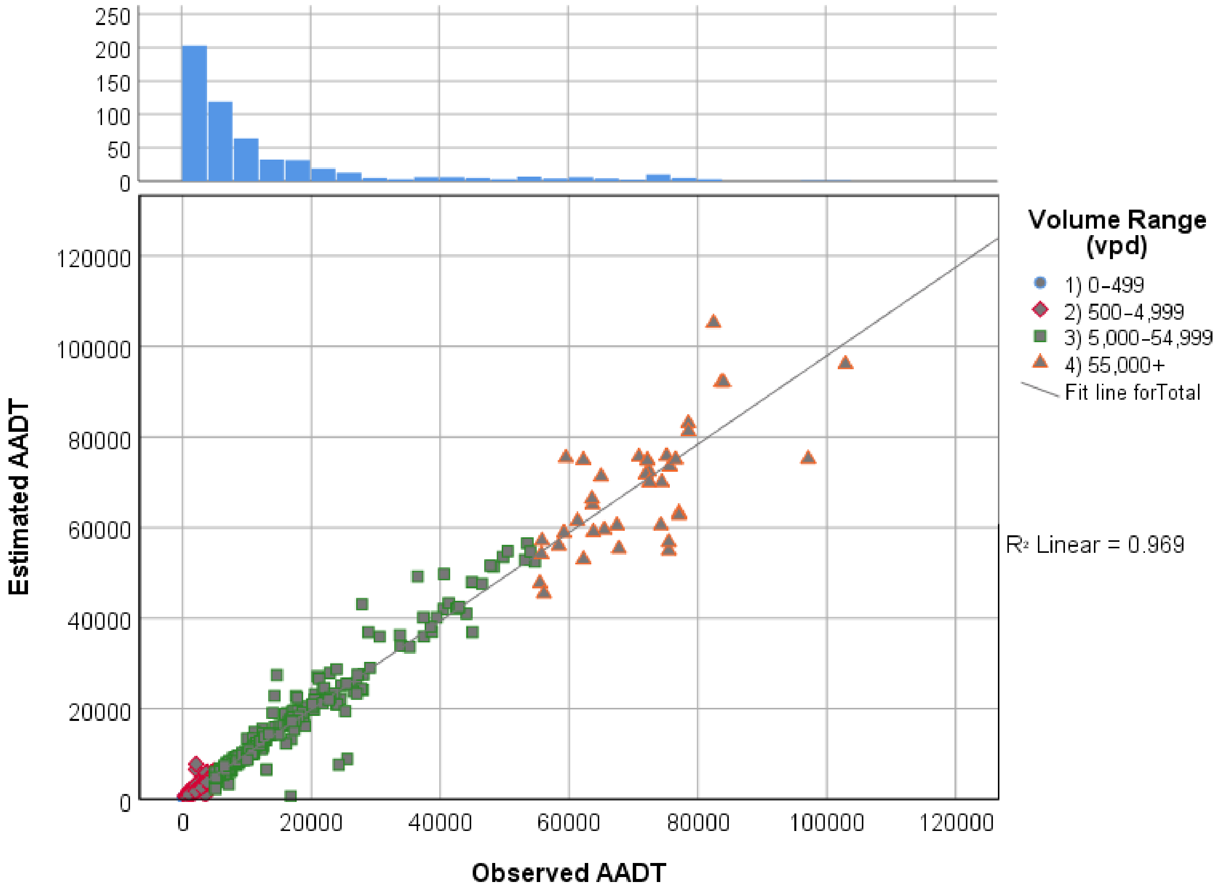

Figure 10. Scatterplot: Directional AADT Data from Six States (CA, ME, MN, NJ, OR, TX)

Figure 11 shows the frequency histogram of the algebraic AADT difference (=estimated AADT – observed AADT) for bidirectional data (Appendix A includes the corresponding histogram and other relevant charts developed using directional AADT data). The normal probability plot or quantile-quantile (Q-Q) plot shown in Figure 12 reveals that the algebraic AADT difference is not normally distributed because the points do not lie on or are not close to the straight diagonal line. This finding was confirmed by two normality tests, the Kolmogorov-Smirnov test and the Shapiro-Wilk test, performed at 95% confidence (Table 10). According to the results of these tests one can reject the null hypothesis (p value<0.05). In other words, there is not enough statistical evidence to support that the data are normally distributed.

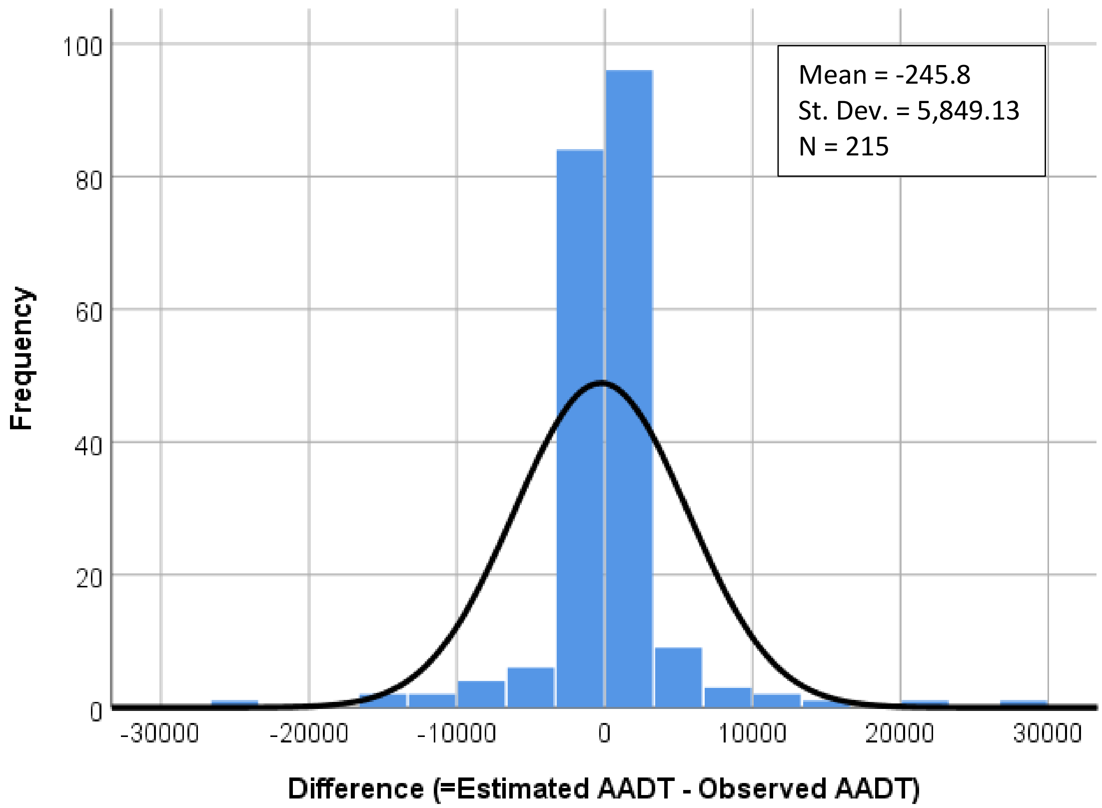

Figure 11. Histogram of Bidirectional AADT Difference (States: CA, ME, MN, NJ, OR, TX)

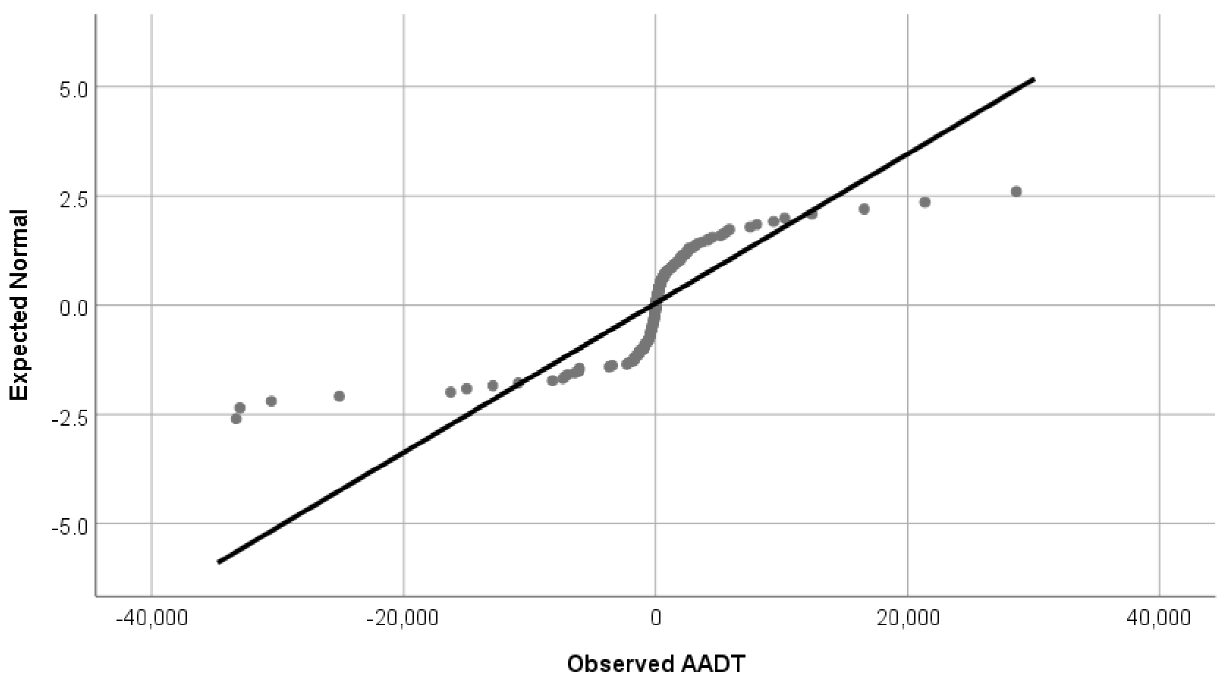

Figure 12. Normal Q-Q Plot of Bidirectional AADT Difference (=Estimated AADT – Observed AADT) Data (States: CA, ME, MN, NJ, OR, TX)

Table 10. Normality Test Results (Target Variable = AADT Difference) (States: CA, ME, MN, NJ, OR, TX)

Target Variable	AADT Directionality	Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	df	p value	Statistic	df	p value
AADT Difference	Bidirectional	0.303	215	0.000	0.587	215	0.000
	Directional	0.283	552	0.000	0.603	552	0.000

TTI repeated similar analyses using the TCE as the target variable. Figure 13 shows the TCE frequency histogram developed using bidirectional data and Figure 14 shows the corresponding Q-Q plot (see Appendix A for the corresponding plots developed for directional AADT data). Table 11 shows the results of the two normality tests conducted at 95% confidence by state using bidirectional and directional AADT data. The shaded cells highlighted with a thick borderline indicate that we cannot reject the null hypothesis (p value>0.05); however, the sample sizes are less than 10 and therefore we cannot draw safe conclusions – larger sample sizes are needed. In all other cases, the results suggest that the TCE data are not normally distributed. It is worth stating that the possible range of TCE is [-100, ∞) and therefore high positive TCE values may contribute to the non-normality of the TCE data.

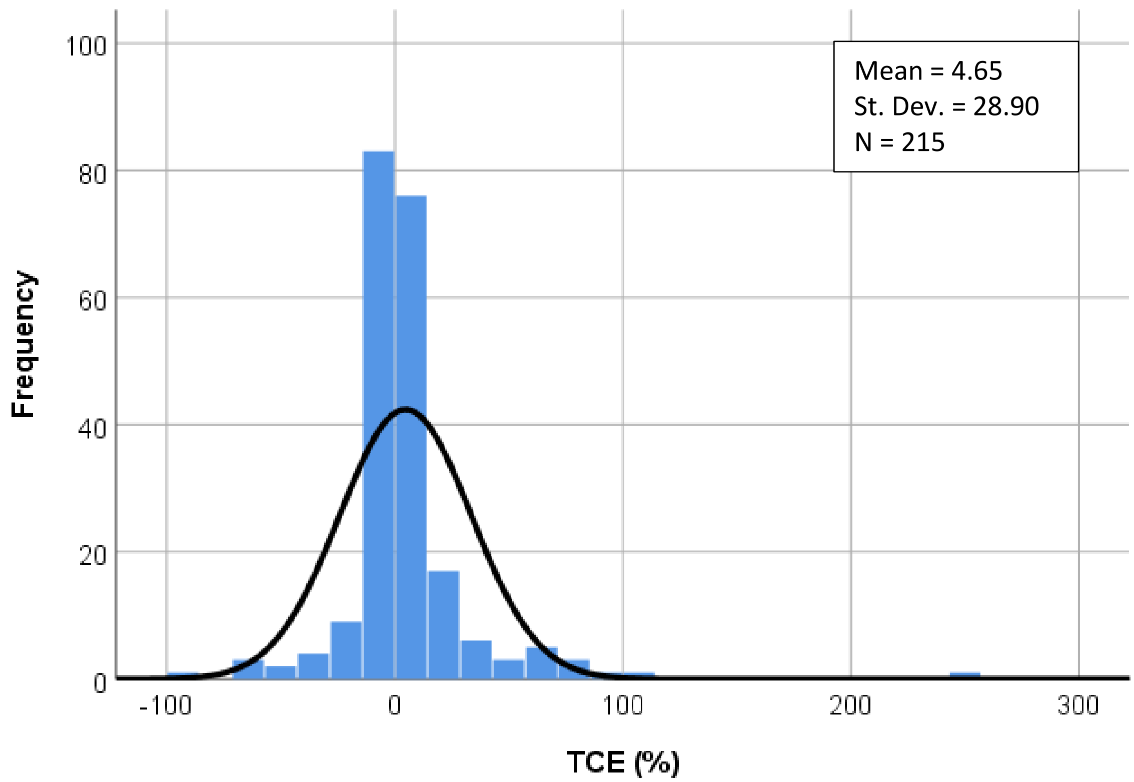

Figure 13. TCE Frequency Histogram – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX)

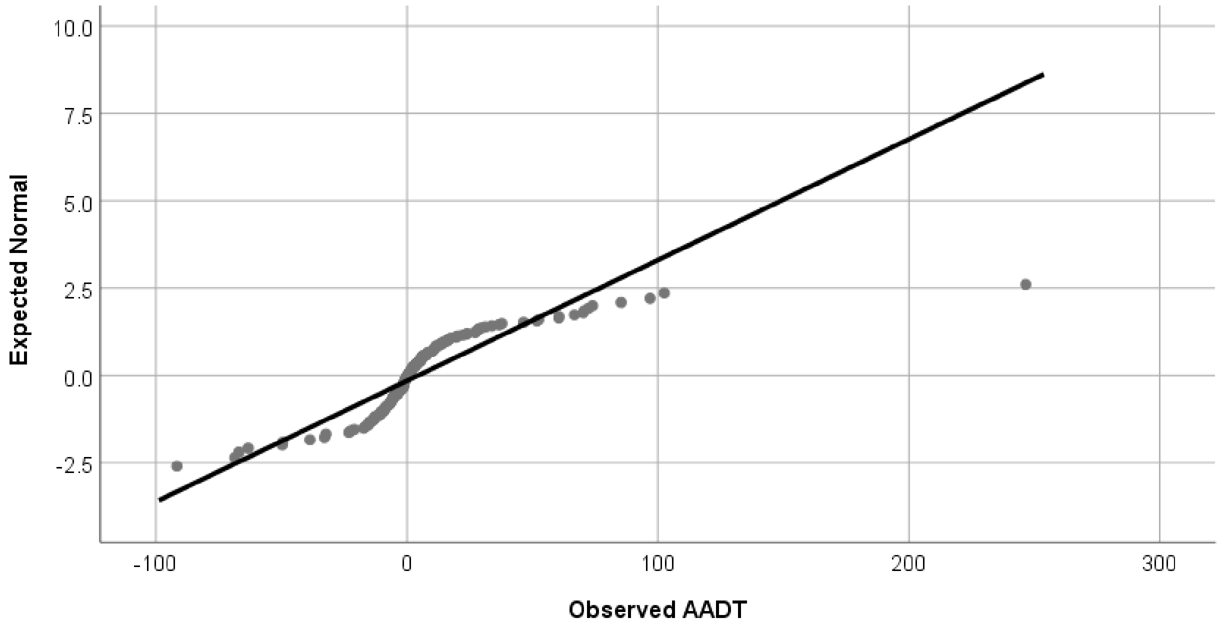

Figure 14. Normal Q-Q Plot of Bidirectional TCE Data (States: CA, ME, MN, NJ, OR, TX)

Table 11. Normality Test Results (Target Variable = TCE) by State

Target Variable	State	AADT Directionality	Kolmogorov-Smirnov			Shapiro-Wilk		
			Statistic	df	p value	Statistic	df	p value
TCE	California	Bidirectional	0.262	9	0.074	0.848	9	0.072
		Directional	0.326	26	0.000	0.658	26	0.000
	Maine	Bidirectional	0.193	5	0.200	0.948	5	0.724
		Directional	0.207	36	0.000	0.752	36	0.000
	Minnesota	Bidirectional	0.363	4	-	0.756	4	0.044
		Directional	0.342	8	0.006	0.756	8	0.009
	New Jersey	Bidirectional	0.355	4	-	0.851	4	0.229
		Directional	0.180	46	0.001	0.861	46	0.000
	Oregon	Bidirectional	0.210	71	0.000	0.875	71	0.000
		Directional	0.186	150	0.000	0.873	150	0.000
	Texas	Bidirectional	0.197	122	0.000	0.771	122	0.000
		Directional	0.174	286	0.000	0.774	286	0.000
	Total	**Bidirectional**	**0.207**	**215**	**0.000**	**0.711**	**215**	**0.000**
		Directional	**0.194**	**552**	**0.000**	**0.717**	**552**	**0.000**

Considering the fact that the TCE is the main input of the method described in FHWA document "*Guidelines for Obtaining AADT Estimates from Non-Traditional Sources*"[2], TTI employed statistical methods for non-parametric data, as explained in the previous chapter; however, other statistical methods may also be appropriate (e.g., making data transformations and then performing statistical tests for parametric data or applying bootstrapping, which is a more advanced technique to determine confidence intervals of the median of a non-parametric variable).

In addition, TTI developed various plots to examine the TCE variability in relation to the magnitude of the AADT. Figure 15 shows a scatterplot of the observed AADT (x axis) and the TCEs (y axis) calculated using bidirectional data. Figure 16 shows the results aggregated by volume group across all TTI evaluation sites, whereas in Figure 17 separate boxplots were developed by state and volume group. The circled numbers in Figure 17 indicate the volume group number that corresponds to each bar. All figures were developed using bidirectional data. Appendix A includes the corresponding graphs created for directional AADT data.

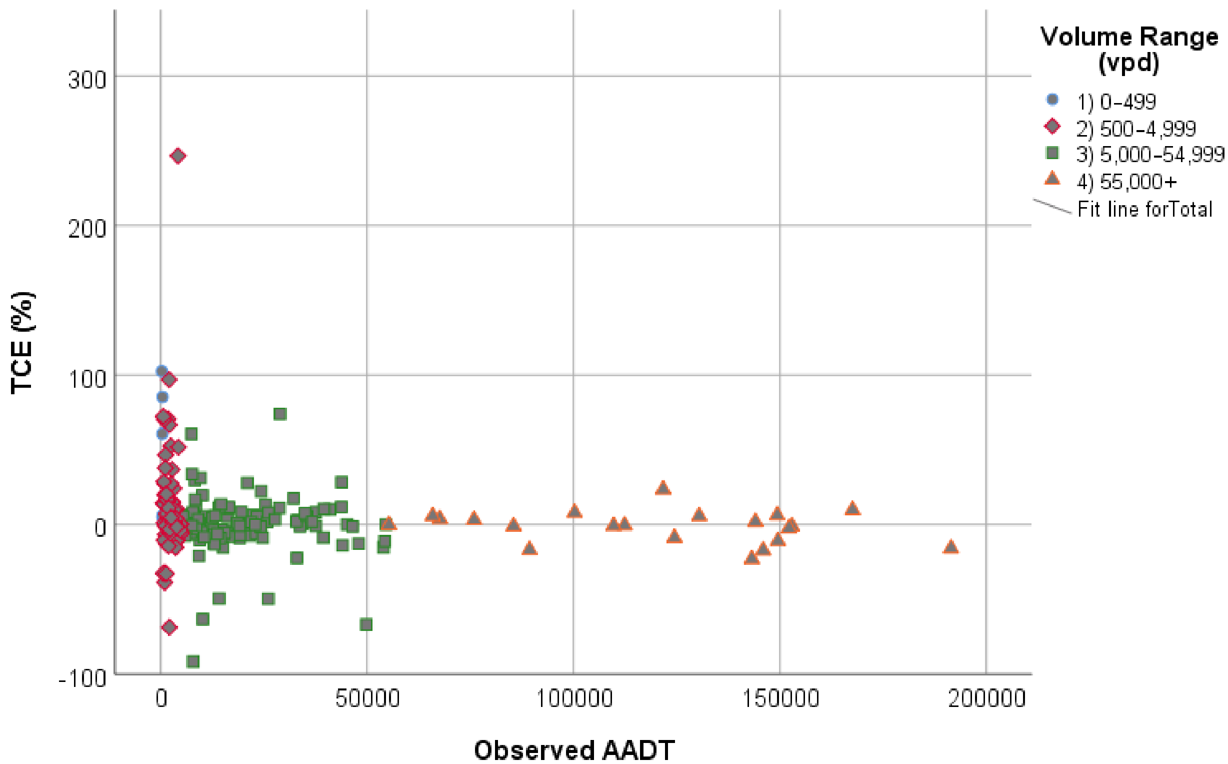

Figure 15. Scatterplot: TCE vs. Observed AADT – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX)

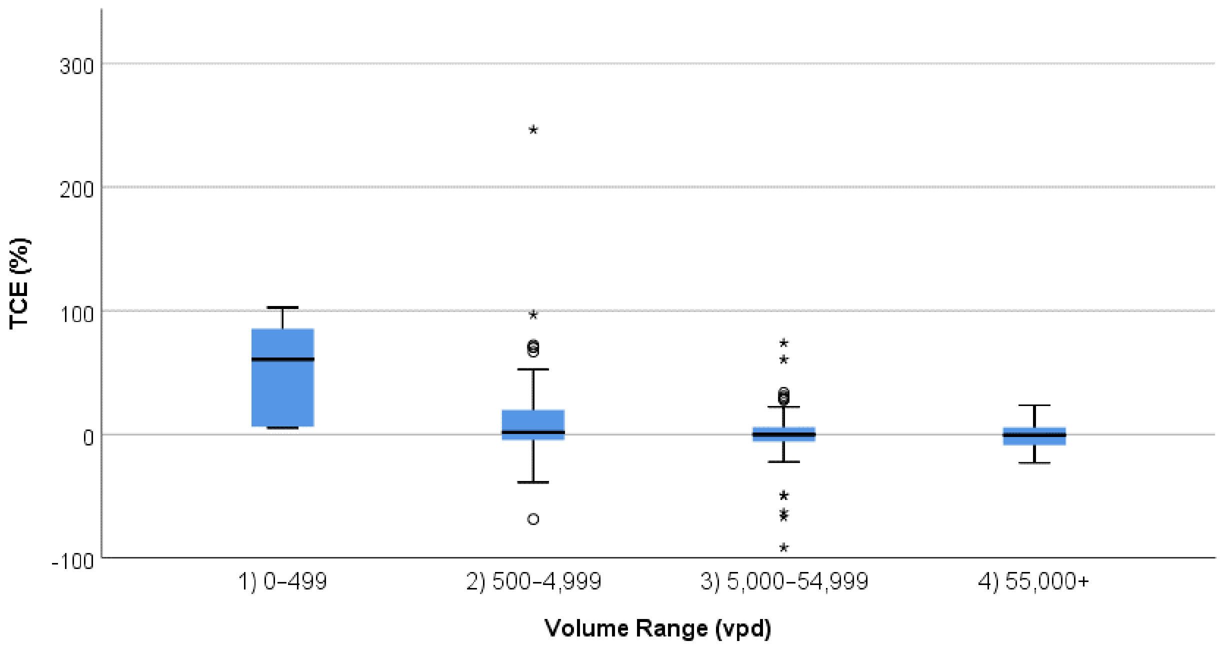

Figure 16. Boxplots: TCE by Volume Group – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX)

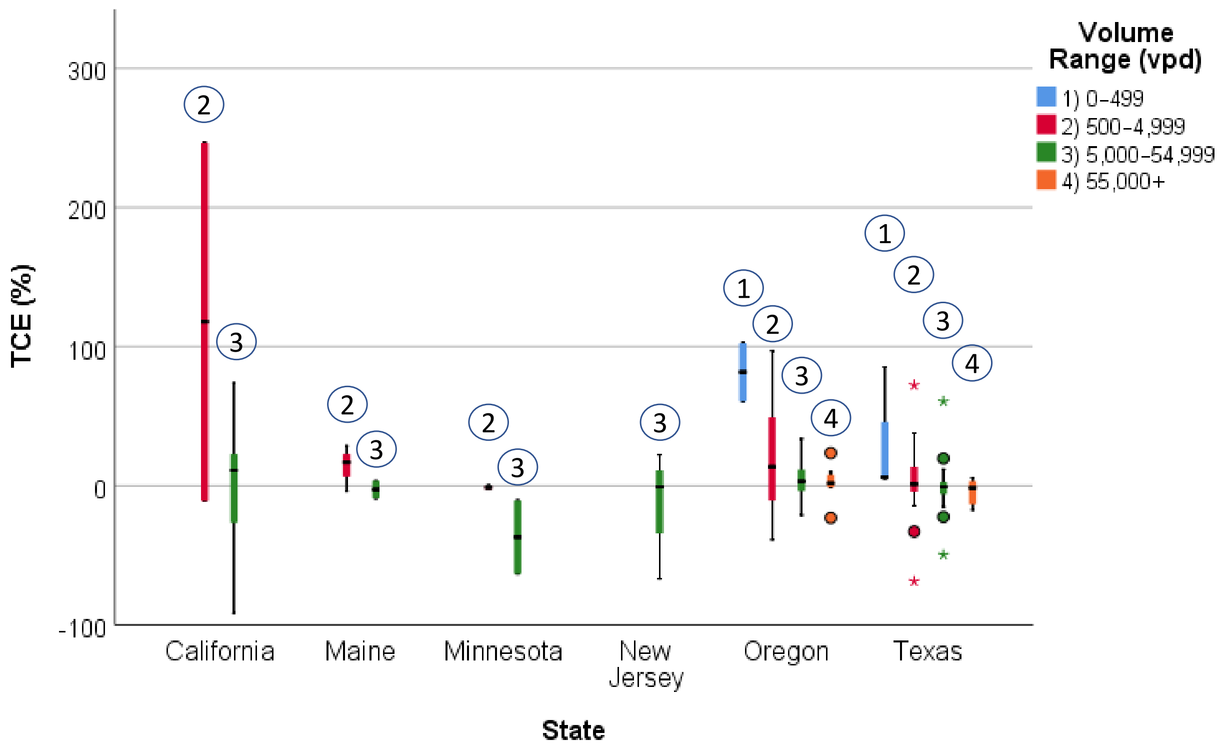

Figure 17. Boxplots: TCE by Volume Group and State – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX)

In general, the results show that the TCE variability tends to decrease as AADT increases. This decreasing trend was also captured by the precision targets shown in Figure 4 through Figure 6.

Metrics and Summary Descriptive Statistics

This section includes several summary tables that present various metrics and statistics aggregated at different levels as described below:

- Table 12 and Table 13 show the results of the analysis by four volume groups for bidirectional and directional data, respectively.
- Table 14 and Table 15 show the results of the analysis by state for bidirectional and directional data, respectively.
- Table 16 and Table 17 show the results of the analysis by state and volume group for bidirectional and directional data, respectively.

Appendix A includes additional tables that present the results by ten disaggregated volume groups; rural/urban designation; roadway functional class; functional class and rural/urban code; and state and rural/urban code.

The main findings from these tables are summarized below:

- The <u>algebraic difference</u> (=estimated AADT – observed AADT) indicates that the AADT is overestimated (i.e., positive values) within low-volume roads, while it is underestimated in higher volume groups.
- Likewise, both the mean and the median <u>TCEs</u> are relatively large and positive within the first volume group (0-499 vpd); however, they tend to be close and around zero in the remaining three volume groups making it difficult to draw general conclusions about the accuracy of AADT estimates based on this metric. Other metrics may be more useful for this purpose.
- The AADT accuracy, expressed through the <u>MAPE</u> and the <u>median APE</u>, gradually improves from low to high traffic volume roads. This can be partially attributed to the fact that the APEs tend to be high as the inputs (i.e., observed and estimated AADT) decrease – this is one of the caveats of the APE.
- The grand <u>median APEs</u> (6.6% in Table 12 and 7.3% in Table 13) are significantly smaller than the grand <u>MAPEs</u> (15.0% in Table 12 and 14.6% Table 13) confirming that the APEs are highly skewed to the right. The median APE is closer to the peak of the data distribution making it more appropriate to be used than the MAPE.
- The <u>NRMSE</u> partially addresses the limitation of the APE described above by squaring and therefore "penalizing" the large differences between the estimated AADT and the observed AADT. With a few exceptions, the NRMSE exhibits similar (but less pronounced) decreasing trends as those of the MAPE and the median APE. The errors tend to gradually decrease from low to high volume roads.
- Generally speaking, high <u>R^2</u> values were obtained in the case of large AADT ranges (e.g., 0–200,000 vpd). For example, the grand R^2 (=0.97) in Table 12 and Table 13 are higher than the R^2 values that were separately calculated for each volume group. Subdividing a dataset into small subsets (e.g., 0-499 vpd) inevitably reduces the sample size (within each subset) making the R^2 more sensitive to potential outliers or high errors. For example, the third volume group (5,000-54,999 vpd) that has a wide range and relatively large sample sizes (124 and 273 records in Table 12 and Table 13, respectively) results in high R^2, compared to the first volume group that is much narrower (0-499 vpd) and contains fewer evaluation sites (5 and 35 sites in Table 12 and Table 13, respectively).

Table 12. Summary Statistics by Volume Group – Bidirectional AADT (States: CA, ME, MN, NJ, OR, TX)

Volume Range	N	Algebraic Diff.		TCE (%)		TCE Percentile				TCE Range		APE (%)		NRMSE	R^2
		Mean	Median	Mean	Median	2.5th	16th	84th	97.5th	68%	95%	Mean	Median		
0–499	5	172.2	219.0	52.0	60.7	5.4	5.9	91.5	100.8	85.6	95.4	52.0	60.7	0.58	0.06
500–4,999	65	316.4	38.0	12.5	1.7	-35.2	-10.4	34.9	82.2	45.2	117.3	23.2	12.8	0.63	0.54
5,000–54,999	124	-136.8	-9.0	-0.4	-0.1	-49.6	-8.5	10.5	31.1	19.0	80.7	10.5	5.9	0.24	0.88
55,000+	21	-2729.3	-608.0	-1.5	-0.5	-20.1	-14.9	6.2	16.7	21.1	36.8	7.9	5.8	0.12	0.84
Total	**215**	**-245.8**	**17.0**	**4.7**	**0.2**	**-45.7**	**-8.9**	**15.7**	**71.7**	**24.7**	**117.4**	**15.0**	**6.6**	**0.25**	**0.97**

Table 13. Summary Statistics by Volume Group – Directional AADT (States: CA, ME, MN, NJ, OR, TX)

Volume Range	N	Algebraic Diff.		TCE (%)		TCE Percentile				TCE Range		APE (%)		NRMSE	R^2
		Mean	Median	Mean	Median	2.5th	16th	84th	97.5th	68%	95%	Mean	Median		
0–499	35	61.6	74.0	26.6	16.1	-35.8	-23.1	72.9	108.9	96.0	144.7	39.7	30.3	0.45	0.26
500–4,999	205	171.2	29.0	9.2	1.3	-21.0	-7.7	26.5	69.2	34.2	90.2	17.1	7.2	0.31	0.79
5,000–54,999	273	226.7	54.0	0.7	0.4	-47.1	-8.4	11.6	30.1	19.9	77.2	10.1	6.6	0.20	0.94
55,000+	39	-2169.9	-1504.0	-2.9	-2.0	-24.9	-18.3	6.7	26.8	25.0	51.7	10.0	6.8	0.14	0.50
Total	**552**	**26.3**	**31.0**	**5.2**	**1.1**	**-35.0**	**-9.0**	**16.8**	**68.8**	**25.8**	**103.8**	**14.6**	**7.3**	**0.25**	**0.97**

Table 14. Summary Statistics by State – Bidirectional AADT (States: CA, ME, MN, NJ, OR, TX)

State	N	Algebraic Diff.		TCE (%)		TCE Percentile				TCE Range		APE		NRMSE	R^2
		Mean	Median	Mean	Median	2.5th	16th	84th	97.5th	68%	95%	Mean	Median		
California	9	4199.7	3179.0	24.6	11.1	-83.2	-38.6	61.2	212.1	99.8	295.2	59.2	28.3	0.50	0.86
Maine	5	-125.0	208.0	7.3	3.6	-8.5	-5.7	21.2	27.5	26.9	36.0	12.4	9.0	0.11	0.99
Minnesota	4	-1870.5	-535.0	-19.0	-6.5	-59.3	-37.9	-1.1	0.0	36.8	59.3	19.1	6.5	0.46	0.18
New Jersey	4	-7059.8	-172.5	-11.5	-0.7	-61.9	-35.4	11.5	20.6	46.9	82.5	22.6	11.8	0.56	0.65
Oregon	71	910.9	400.0	10.7	5.8	-32.3	-8.6	29.2	77.1	37.8	109.4	18.5	11.3	0.23	0.98
Texas	122	-975.2	-19.5	0.9	-0.6	-22.2	-8.2	7.8	37.8	16.0	60.1	9.5	5.4	0.20	0.99
Total	**215**	**-245.8**	**17.0**	**4.7**	**0.2**	**-45.7**	**-8.9**	**15.7**	**71.7**	**24.7**	**117.4**	**15.0**	**6.6**	**0.25**	**0.97**

Table 15. Summary Statistics by State – Directional AADT (States: CA, ME, MN, NJ, OR, TX)

State	N	Algebraic Diff.		TCE (%)		TCE Percentile				TCE Range		APE		NRMSE	R^2
		Mean	Median	Mean	Median	2.5th	16th	84th	97.5th	68%	95%	Mean	Median		
California	26	2489.8	1503.0	25.8	5.8	-48.3	-10.3	29.4	239.8	39.8	288.1	37.4	12.2	0.26	0.95
Maine	36	-164.4	94.5	7.5	3.0	-25.0	-9.1	23.5	65.2	32.6	90.2	18.1	9.3	0.18	0.99
Minnesota	8	-776.4	-291.5	-15.8	-8.0	-54.0	-42.1	0.5	1.2	42.7	55.2	16.4	8.0	0.38	0.41
New Jersey	46	693.9	1071.5	7.4	7.4	-68.1	-5.5	29.2	62.2	34.7	130.3	20.1	13.3	0.24	0.93
Oregon	150	362.1	173.0	10.0	4.7	-32.4	-9.6	29.5	88.0	39.1	120.4	18.2	10.3	0.23	0.98
Texas	286	-434.7	-9.5	0.9	-0.2	-22.9	-8.2	8.4	43.1	16.6	66.0	9.2	4.9	0.22	0.98
Total	**552**	**26.3**	**31.0**	**5.2**	**1.1**	**-35.0**	**-9.0**	**16.8**	**68.8**	**25.8**	**103.8**	**14.6**	**7.3**	**0.25**	**0.97**

Table 16. Summary Statistics by State and Volume Group – Bidirectional AADT (States: CA, ME, MN, NJ, OR, TX)

State	Volume Range	N	Algebraic Diff.		TCE (%)		TCE Percentile				TCE Range		APE		NRMSE	R^2
			Mean	Median	Mean	Median	2.5th	16th	84th	97.5th	68%	95%	Mean	Median		
California	**500–4,999**	2	4924.5	4924.5	117.9	117.9	-4.4	30.4	205.4	240.1	175.0	244.5	128.7	128.7	1.84	1.00
California	**5,000–54,999**	7	3992.6	3179.0	-2.0	11.1	-85.3	-51.1	30.1	67.1	81.3	152.4	39.4	28.3	0.43	0.90
Maine	**500–4,999**	3	114.7	208.0	13.9	16.9	-2.8	2.8	24.9	28.1	22.1	30.9	16.5	16.9	0.12	0.99
Maine	**5,000–54,999**	2	-484.5	-484.5	-2.7	-2.7	-8.7	-7.0	1.6	3.3	8.6	12.0	6.3	6.3	0.08	1.00
Minnesota	**500–4,999**	2	-41.0	-41.0	-1.2	-1.2	-2.5	-2.1	-0.2	0.2	1.9	2.7	1.4	1.4	0.02	1.00
Minnesota	**5,000–54,999**	2	-3700.0	-3700.0	-36.8	-36.8	-61.9	-54.8	-18.9	-11.7	36.0	50.2	36.8	36.8	0.47	1.00
New Jersey	**5,000–54,999**	4	-7059.8	-172.5	-11.5	-0.7	-61.9	-35.4	11.5	20.6	46.9	82.5	22.6	11.8	0.56	0.65
Oregon	**0–499**	2	231.0	231.0	81.6	81.6	61.7	67.4	95.8	101.5	28.5	39.8	81.6	81.6	0.77	1.00
Oregon	**500–4,999**	23	400.0	112.0	17.1	13.8	-35.3	-14.1	59.4	82.4	73.5	117.7	30.6	15.2	0.41	0.78
Oregon	**5,000–54,999**	37	701.8	627.0	4.8	3.2	-15.8	-6.4	13.4	31.5	19.8	47.3	10.1	6.6	0.12	0.97
Oregon	**55,000+**	9	3227.0	2632.0	2.6	1.8	-18.6	-1.1	9.4	20.8	10.4	39.4	8.4	6.3	0.12	0.72
Texas	**0–499**	3	133.0	30.0	32.3	6.3	5.3	5.6	60.0	81.4	54.4	76.1	32.3	6.3	0.48	0.01
Texas	**500–4,999**	35	35.8	38.0	4.2	1.2	-38.2	-6.3	20.2	43.0	26.4	81.3	14.1	6.7	0.18	0.89
Texas	**5,000–54,999**	72	-476.0	-86.5	-1.1	-0.8	-16.8	-7.5	5.1	13.5	12.5	30.3	6.6	4.5	0.14	0.96
Texas	**55,000+**	12	-7196.5	-2261.5	-4.6	-1.6	-17.1	-16.1	4.2	5.7	20.3	22.8	7.6	5.7	0.12	0.94
Total		**215**	**-245.8**	**17.0**	**4.7**	**0.2**	**-45.7**	**-8.9**	**15.7**	**71.7**	**24.7**	**117.4**	**15.0**	**6.6**	**0.25**	**0.97**

Table 17. Summary Statistics by State and Volume Group – Directional AADT (States: CA, ME, MN, NJ, OR, TX)

State	Volume Range	N	Algebraic Diff.		TCE (%)		TCE Percentile				TCE Range		APE		NRM SE	R^2
			Mean	Median	Mea n	Median	2.5th	16th	84th	97.5t h	68%	95%	Mean	Median		
California	**500–4,999**	6	1723.3	209.5	81.0	5.4	-11.6	-6.6	230.5	266.4	237.1	278.1	86.9	9.6	1.14	0.00
California	**5,000–54,999**	16	2754.2	1851.5	10.7	11.1	-50.5	-8.6	26.7	77.2	35.3	127.7	25.2	17.5	0.25	0.92
California	**55,000+**	4	2582.0	-2722.0	3.3	-2.1	-10.1	-8.5	15.5	25.7	24.1	35.8	11.7	8.5	0.16	0.67
Maine	**0–499**	3	197.0	110.0	67.5	30.3	27.2	28.1	108.3	139.3	80.2	112.1	67.5	30.3	0.72	0.99
Maine	**500–4,999**	21	-83.1	87.0	3.8	2.0	-46.1	-8.5	20.0	51.3	28.5	97.4	16.9	10.5	0.26	0.84
Maine	**5,000–54,999**	12	-396.9	-98.0	-1.1	-1.1	-12.9	-9.9	8.6	11.5	18.5	24.4	7.8	7.9	0.12	0.99
Minnesota	**500–4,999**	6	-171.0	-47.5	-4.0	-2.6	-10.5	-10.3	0.7	1.2	11.0	11.7	4.8	3.5	0.10	0.99
Minnesota	**5,000–54,999**	2	-2592.5	-2592.5	-51.1	-51.1	-55.4	-54.2	-47.9	-46.7	6.3	8.7	51.1	51.1	0.51	1.00
New Jersey	**500–4,999**	4	1373.5	1451.5	41.4	47.3	9.0	18.8	63.5	63.6	44.7	54.6	41.4	47.3	0.42	0.45
New Jersey	**5,000–54,999**	35	803.5	1131.0	5.1	7.7	-72.7	-6.5	26.9	38.9	33.3	111.6	20.8	16.3	0.34	0.80
New Jersey	**55,000+**	7	-242.1	133.0	-0.7	0.2	-12.5	-4.5	2.6	8.6	7.1	21.2	4.5	2.3	0.06	0.87
Oregon	**0–499**	18	21.2	34.5	17.6	10.8	-39.2	-33.0	68.2	102.5	101.2	141.7	40.7	33.4	0.43	0.19
Oregon	**500–4,999**	51	346.7	224.0	19.7	14.0	-20.3	-9.1	51.7	84.6	60.8	104.9	26.1	16.6	0.29	0.87
Oregon	**5,000–54,999**	66	456.8	197.0	2.4	2.0	-14.9	-7.7	11.2	19.1	18.8	34.0	7.6	7.2	0.09	0.99
Oregon	**55,000+**	15	406.7	2701.0	0.9	3.6	-26.4	-15.4	10.0	24.5	25.4	50.9	11.2	6.8	0.14	0.25
Texas	**0–499**	14	84.6	75.5	29.4	15.9	-11.7	4.7	68.6	93.9	63.9	105.6	32.4	15.9	0.39	0.61
Texas	**500–4,999**	117	37.2	3.0	1.5	0.3	-30.2	-6.5	9.9	45.0	16.4	75.2	9.5	4.7	0.16	0.93
Texas	**5,000–54,999**	142	-214.8	-56.5	-1.4	-0.5	-22.1	-8.1	6.3	14.0	14.4	36.1	6.5	4.5	0.12	0.97
Texas	**55,000+**	13	-7643.2	-5939.0	-10.4	-9.1	-21.5	-18.6	-2.7	2.5	15.9	24.0	11.0	9.1	0.14	0.61
Total		**552**	**26.3**	**31**	**5.2**	**1.1**	**-35.0**	**-9.0**	**16.8**	**68.8**	**25.8**	**103.8**	**14.6**	**7.3**	**0.25**	**0.97**

Statistical Testing Methods

In data analysis and especially with statistical testing, there has never been a perfect methodology for a given scenario and that is probably why conclusions are stated in a probability term in addition with other conditions and constraints. Further, when multiple methods are used, there is also a possibility that conflicting information and results may arise. Consultation among engineers and statisticians should be carried out to ensure data are analyzed and results are interpreted appropriately.

Both the TPF-5(384) method and the TTI modified TPF-5(384) approach rely on historical data accuracy and precision information gained from studying AADT data through annualization of sample short-term counts extracted from permanent sites. The accuracy and precision information enabled the establishment of acceptable precisions and ranges of AADT data based on other alternative AADT methods. The rationale is that if an alternative AADT method produces AADT data with no inferior precision and accuracy then the annualization of short-term count method, such alternative methods should deem adequate. Specifically, the TPF-5(384) method specifies the accuracy and precision tolerance ranges based on a 95% confidence interval (the significance level is 0.05).

The Wilcoxon signed-rank hypothesis test was used to decipher whether differences between AADTs derived from StL and benchmark sites are statistically significant. The test offers the probability (certainty) of rejecting the so-called null hypothesis, according to which the difference between two AADT methods is due to randomness (not the methods themselves). For example, if the difference between the AADTs is 10.5% between two methods, the hypotheses testing would indicate whether the 10.5% difference is due to fundamental issues with the two methods (reject the hypothesis) or randomness (have nothing to do with the methods) (failed to reject the hypothesis). Combined with the descriptive summary statistics, this information will help to decipher whether there are real differences between the two methods. If the test fails to reject the null hypothesis that there is no statistically significant difference between the two methods, the StL method is acceptable. If the StL data also fall within the tolerances of the summary descriptive statistics, StL data is acceptable as it is. If StL data fall outside the tolerances of the summary descriptive statistics, further calibration of the StL data should be carried before accepting the data. For cases and scenarios where the null hypothesis was rejected, StL data will require further review, development, and benchmarking.

This section presents the results of the statistical testing methods described in the previous chapter. Each of the tables below includes eight columns. The first column shows the variable that was used to slice the data and aggregate the results. For example, Table 18 shows results aggregated by four volume groups. The second column shows the sample size within each subset (e.g., volume group). The remaining columns show the results of the following methods:

- TPF-5(384) method (described in FHWA document *Guidelines for Obtaining AADT Estimates from Non-Traditional Sources*): It indicates with a "Y" (yes) and an "N" (no) whether both the accuracy and the precision targets described in the aforementioned document are met. The empty cells mean that the method cannot be applied because a) there are no accuracy and precision targets (e.g., no targets are provided for the first volume group: 0-499 vpd) or b) the sample size is not adequate.
- Modified TPF-5(384) method: It provides the number and the corresponding percent of the StL AADT estimates that fall within the bounds shown in Figure 6 (see previous chapter). For example, as previously shown in Figure 7 and Table 7, of the 286 StL directional AADT estimates

provided for Texas, 272 estimates or 95% (=272/286) of all estimates fall within the target bounds depicted in Figure 7.

- Wilcoxon Signed-Rank Test: It provides a two-tailed p-value based on the z-value calculated from the StL and the benchmark AADTs. The p-value is the probability of observing the current difference or any differences larger than what is being observed between the two median AADTs from the two AADT different methods that due to random effect. In practice, any time when the p-value exceeds 5%, from a statistical standpoint, the conclusion is that the data failed to reject the null hypothesis that there are statistically significant differences between the two median AADTs. For example, a p-value of 0.92 (i.e., $p>0.05$) indicates that the probability of seeing the difference or any difference larger than what is reported due to random causes is 92%. The conclusion is that the observed difference between the two methods is not significant. In other words, we accept the null hypothesis that both samples are from the same population. On the other hand, if $p<0.05$ we reject the null hypothesis and conclude that the two samples are from different populations.

Appendix A includes additional tables that present the results by state and volume group; rural/urban code; state and rural/urban code; roadway functional class; and functional class and rural/urban code.

The main findings and lessons learned from the three statistical methods are summarized below:

- TPF-5(384) method:
  - It is appropriate for non-parametric data.
  - The accuracy targets of only one volume group (5,000-54,999 vpd) are met as shown in Table 18. In all other subsets and tables (Table 19 through Table 23), neither the accuracy nor the precision targets are met. This can be primarily attributed to the fact that the method compares the (accuracy and precision) target values for a given volume group against extreme TCE values of points (i.e., pairs of AADTs) that are located on the two tails of the data distribution. These extreme TCE values, particularly the one on the right tail, are often larger than the corresponding targets.
  - It can only be applied by volume group. In other words, it cannot be applied by slicing the data differently such as by state, functional class, rural/urban code, etc.
  - It requires a relatively large sample size, which may be difficult to obtain for any volume group. Specifically, a sample size of 93 sites per volume group is required to calculate the upper and lower precision bounds.
  - It may be difficult to understand, apply, and communicate it to others.

- TTI Modified TPF-5(384) method:
  - It is appropriate for non-parametric data and relies on the nine precision targets provided in FHWA document "*Guidelines for Obtaining AADT Estimates from Non-Traditional Sources*".
  - It is fully automated in Excel and easy to apply.
  - It addresses some of the limitations stated above. For example, it can be applied by slicing the data in many ways. In general, it produces intuitive results that are consistent with the metrics presented in the previous subsection and the Wilcoxon S-R test. High percentages (of points within the bounds) are associated with low errors (APEs and NRMSEs), and vice versa.
  - One limitation of this method is that in its current form it only captures the nine precision targets (not the accuracy targets) provided in the FHWA document stated

above. The method may need to be further improved to account for the accuracy targets as well. Though the accuracy targets may not be as critical as the precision targets, incorporating them into the method may be necessary to check for consistent overestimation or underestimation of AADT. For example, Figure 18 shows an example where all TCEs (points) fall within the two precision bounds, but they are all positive indicating that the AADT estimates are precise but consistently overestimate the actual AADTs.

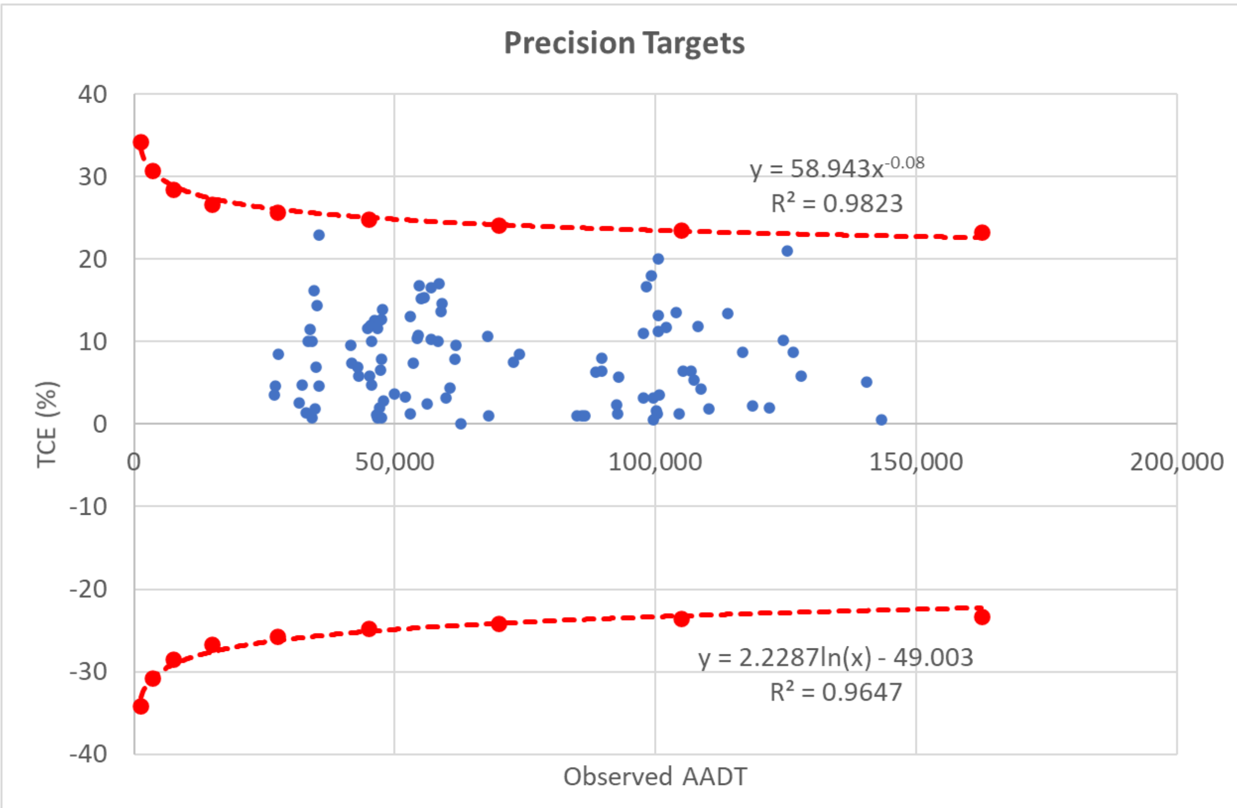

Figure 18. Example of Precise But Overestimated AADT Estimates

- Another consideration is the determination of the minimum sample size required to apply the method. For example, a minimum sample size of 30 observations is typically recommended to apply certain statistical methods.
- Another consideration relates to the determination of the minimum percent of points that need to fall within the precision bounds in order for any set of AADT estimates to be considered acceptable; however, this determination is outside the scope of this evaluation.

- Wilcoxon signed-rank test:
  - It is appropriate for non-parametric data but requires knowledge of statistical hypothesis testing and result interpretation.
  - In many cases, the results are in line with those from the alternative method and the metrics presented in the previous section.

- Like most methods, the Wilcoxon test has little power when the sample size is small. In fact, when the sample size is five or less, the Wilcoxon test results in a p value ≥ 0.05, regardless of how far the sample median is from the hypothetical median.
- The Wilcoxon test may produce misleading results when many pairs of AADT have the same value.
- Like many statistical methods, it lacks professional judgement, which may be necessary.

Table 18's p-value under the Wilcoxon signed-rank test methods that with a significance level of 5%, the hypothesis that there is no statistically difference between the two AADTs method for volume ranges for 0-499 and 500-4,999 are rejected. In other words, for the two low volume ranges, differences between the AADTs from the two methods are more than just from random causes. On the other hand, for the volume ranges of 5,000-54,999 and 55,000+, the test failed to reject the null hypothesis that there is no statistically significant difference between the two methods. In other words, the difference observed between the two AADTs is most likely due to randomness as opposed to methodological causes. When all volume groups are collapsed into one, the overall test failed to reject (p>0.05) the null hypothesis (i.e., we accept that both samples are from the same population).

Table 18. Summary Results by Four Volume Groups – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX)

Volume Range	N	TPF-5(384) Method		TTI Modified Method		Wilcoxon S-R Test	
		Accuracy Targets	Precision Targets	Points within Bounds		Z Value	p Value
				Number	Percent		
0–499	5			2	40%	-2.023	0.043
500–4,999	65	N		52	80%	-1.99	0.047
5,000–54,999	124	Y	N	112	90%	-0.188	0.851
55,000+	21	N		19	90%	-0.434	0.664
Total	**215**			**185**	**86%**	**-0.866**	**0.387**

Table 19 shows the conclusion for directional AADTs from the two methods. Based on the Wilcoxon signed-rank test p-values, only the highest volume group (55,000+) failed to reject the null hypotheses. The p-values for all other volume groups and the overall collapsed one indicate that the null-hypothesis is rejected. In other words, differences observed due to randomness cannot explain the observed differences from the two directional AADT methods adequately. Table 20's Wilcoxon signed-rank test results shows the overall directional AADT data from the two methods are statistically significantly different at a significance level of 5%.

Table 19. Summary Results by Four Volume Groups – Directional AADT Data (States: CA, ME, MN, NJ, OR, TX)

Volume Range	N	TPF-5(384) Method		TTI Modified Method		Wilcoxon S-R Test	
		Accuracy Targets	Precision Targets	Points within Bounds		Z Value	p Value
				Number	Percent		
0–499	35			22	63%	-2.588	0.010
500–4,999	205	N	N	174	85%	-3.13	0.002
5,000–54,999	273	N	N	254	93%	-1.97	0.049
55,000+	39	N		35	90%	-1.298	0.194
Total	**552**			**485**	**88%**	**-2.673**	**0.008**

Table 20 indicates that except the lowest volume range (0-499), the Wilcoxon S-R test for all other volume groups failed to reject the null hypothesis at a 5% significant level. In other words, Wilcoxon signed-rank test results indicate that there is a statistically significant difference between the median of the StL AADT estimates and the median of the ground truth AADTs.

Table 20. Summary Results by Ten Volume Groups – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX)

Volume Range	N	TPF-5(384) Method		TTI Modified Method		Wilcoxon S-R Test	
		Accuracy Targets	Precision Targets	Points within Bounds		Z Value	p Value
				Number	Percent		
0–499	5			2	40%	-2.023	0.043
500–1,999	32	N		26	81%	-1.636	0.102
2,000–4,999	33	N		26	79%	-1.253	0.210
5,000–9,999	37	N		32	86%	-0.098	0.922
10,000–19,999	42	N		40	95%	-1.282	0.200
20,000–34,999	28	N		25	89%	-1.776	0.076
35,000–54,999	17	N		15	88%	-0.450	0.653
55,000–84,999	4			4	100%	-1.461	0.144
85,000–124,999	7	N		6	86%	-0.507	0.612
125,000+	10	N		9	90%	-0.968	0.333
Total	**215**			**185**	**86%**	**-0.866**	**0.387**

According to the Wilcoxon S-R test results shown in Table 21, there are three volume groups (0-499, 500-1,999, and 10,000-19,999) where the null hypothesis is rejected (p<0.05) at a 5% significance level and therefore, we conclude that the two directional AADT medians are statistically different. In the remaining six volume groups, the results indicate that there is not enough evidence to reject the null hypothesis (i.e., the two medians are from the same population).

Table 21. Summary Results by Ten Volume Groups – Directional AADT Data (States: CA, ME, MN, NJ, OR, TX)

Volume Range	N	TPF-5(384) Method		TTI Modified Method		Wilcoxon S-R Test	
		Accuracy Targets	Precision Targets	Points within Bounds		Z Value	p Value
				Number	Percent		
0–499	35			22	63%	-2.588	0.010
500–1,999	101	N	N	84	83%	-3.005	0.003
2,000–4,999	104	N	N	90	87%	-1.68	0.093
5,000–9,999	115	N	N	111	97%	-0.818	0.414
10,000–19,999	94	N	N	85	90%	-3.002	0.003
20,000–34,999	39	N		34	87%	-0.202	0.840
35,000–54,999	25	N		24	96%	-1.655	0.098
55,000–84,999	37	N		33	89%	-0.913	0.361
85,000–124,999	2			2	100%	-1.342	0.180
125,000+	0						
Total	**552**			**485**	**88%**	**-2.673**	**0.008**

Table 22 shows the results aggregated by state based on bidirectional AADT data. Note that the TPF-5(384) method cannot be applied at the state level. With the exception of Oregon, the Wilcoxon S-R test results for the remaining five states show that the two AADT medians are from the same population. In other words, the null hypothesis cannot be rejected (p>0.05) at a 5% significance level.

Table 22. Summary Results by State – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX)

State	N	TPF-5(384) Method		TTI Modified Method		Wilcoxon S-R Test	
		Accuracy Targets	Precision Targets	Points within Bounds		Z Value	p Value
				Number	Percent		
California	9			4	44%	-1.007	0.314
Maine	5			5	100%	-0.405	0.686
Minnesota	4			3	75%	-1.461	0.144
New Jersey	4			3	75%	-0.73	0.465
Oregon	71			55	77%	-3.005	0.003
Texas	122			115	94%	-1.173	0.241
Total	**215**			**185**	**86%**	**-0.866**	**0.387**

Table 23 shows the results aggregated by state using directional AADT data. With the exception of Oregon and New Jersey, the Wilcoxon S-R test indicates the null hypothesis cannot be rejected (p>0.05) at a 5% significance level in the remaining four states (CA, ME, MN, and TX). In other words, we can conclude that the two AADT medians are from the same population.

Table 23. Summary Results by State – Directional AADT Data (States: CA, ME, MN, NJ, OR, TX)

State	N	TPF-5(384) Method		TTI Modified Method		Wilcoxon S-R Test	
		Accuracy Targets	Precision Targets	Points within Bounds		Z Value	p Value
				Number	Percent		
California	26			18	69%	-1.943	0.052
Maine	36			32	89%	-0.22	0.826
Minnesota	8			6	75%	-1.86	0.093
New Jersey	46			34	74%	-2.584	0.010
Oregon	150			123	82%	-4.042	0.000
Texas	286			272	95%	-1.681	0.093
Total	**552**			**485**	**88%**	**-2.673**	**0.008**

CONCLUSIONS AND RECOMMENDATIONS

Conclusions on Methods to Evaluate Non-Traditional Sources of AADT Estimates

TTI concludes the following about the methods used to evaluate StL AADT estimates in this study.

- TPF-5(384) method: The method produces counterintuitive results that are not in line with those obtained from other metrics and statistical methods. The results are mainly driven by the two extreme TCE values located on the two tails of the TCE data distribution. To put things in perspective, out of 24 volume groups (Table 18-Table 21, Table 34, and Table 35) where this method was applied, the accuracy (but not the precision) targets were met in only one volume group, 5,000-54,999 vpd (Table 18). The precision targets were not met in any of the 24 volume groups. Further, the method can only be applied by volume group. In other words, it cannot be applied by slicing the data differently such as by state, functional class, rural/urban code, etc. It requires a relatively large sample size, which may be difficult for a state DOT to obtain for each individual volume group.

- TTI modified TPF-5(384) method: In general, it produces intuitive results that are consistent with some metrics and the Wilcoxon S-R test. High percentages (of points within the bounds) are associated with low errors (APEs and NRMSEs), and vice versa. It also addresses some of the limitations of the sign test method described above. For example, it can be applied by slicing the data in many ways, not just by volume group. The ability to apply the method at the state level or other aggregation levels is translated into lower sample size requirements, compared to those of the sign test method. One limitation of this method is that in its current form it only captures the nine precision targets (not the accuracy targets).

- Wilcoxon signed-rank test: With a few exceptions, the results are consistent with those of the TTI modified method and some metrics. Like many statistical hypotheses testing methods, it requires both statistical and subject matter expertise for proper interpretation. The Wilcoxon test has little power when the sample size is small.

Conclusions on the Validity of StL AADT Estimates

In this study, TTI evaluated the accuracy and precision of StL's AADT estimates using a wide range of descriptive statistics, error measures, and three statistical hypothesis tests. All this evaluation was performed to answer a basic "yes or no" question: are StL's AADT estimates comparable with the benchmark AADT data derived from traditional permanent traffic counting stations? Indirectly, it answers the validity of replacing the AADT estimates obtained from short duration counts with annualization.

Unfortunately, for multiple reasons explained below, the evaluation results in the previous chapter do not provide a clear-cut answer of "yes" or "no". The FHWA and the two independent evaluation teams (i.e., TTI and NREL) did not reach consensus on a single best evaluation measure or statistical hypothesis test. Each statistical measure or test had strengths and limitations, and the only consensus among FHWA, TTI, and NREL was that multiple evaluation measures and tests should be used in concert to provide the most complete picture of the validity of StL's AADT estimates.

However, in TTI's results, the multiple evaluation measures and hypothesis tests produced mixed results. Some statistical tests provided contradictory or counterintuitive results, despite being preferred or recommended by statisticians. Other statistical hypothesis tests indicated that StL's AADT estimates are statistically the same as the benchmark AADT values. Some of the descriptive statistics (like median APE) produced good results that would point toward "yes, StL AADT estimates are valid for use" whereas other descriptive statistics (like 68th and 95th percentiles) pointed toward "no, StL AADT estimates are not valid for use."

In situations like this with mixed evaluation results, some researchers may conclude only that "more research is needed." However, significant evaluation resources were used by TTI and NREL in this effort, and TTI researchers feel compelled to provide a more definitive answer on the validity of StL AADT estimates than simply "it's too close to call, more research is needed." Therefore, TTI researchers have applied their professional judgment to interpret which results should carry more weight in making a "yes/no" validity decision.

The following paragraphs provide supporting rationale for TTI's conclusion on validity.

- Correlation (as indicated by R^2 values) between StL AADT estimates and benchmark estimates are very high overall (0.97 in Figure 9 and Figure 10). Despite the high correlation, though, low traffic volumes are biased toward overestimation (Figure 16).

- Considering the 68th and 95th percentile TCE ranges, Table 12 and Table 13 (also Table 24 and Table 25 for 10-category volume groups) indicate that both the 68th and 95th percentile values of TCE are greater than comparable short duration counts.[5] However, the comparable short duration count TCE values in this FHWA report are based on only adjustment factor error and does not include portable counter equipment error. If the short duration count TCE values included portable counter equipment error, it is possible that the TCE values in Table 12 and Table 13 would be more closely aligned.

- Considering the mean and median APE values, Table 12 and Table 13 indicate that mean APE values are slightly higher than comparable short duration counts.[5] However, the median APE values are more appropriate in this case, since the error distribution is not normal and non-parametric statistics are needed. However, median APE values are not available for comparable short duration counts.

- Considering the NRMSE, Table 12 and Table 13 indicate that NRMSE values are higher than comparable short duration counts.[6] However, the comparable short duration count NRMSE values in this FHWA report are based on only adjustment factor error and does not include portable counter equipment error. If the short duration count NRMSE values included portable counter equipment error, it is possible that the NRMSE values in Table 12 and Table 13 would be more closely aligned.

[5] See Table 5b in FHWA (2021), *Guidelines for Obtaining AADT Estimates from Non-Traditional Sources*. Unpublished document as of 6/7/2021.
[6] See Table 5b in FHWA (2021), *Guidelines for Obtaining AADT Estimates from Non-Traditional Sources*. Unpublished document as of 6/7/2021.

- Considering the statistical hypothesis tests (Table 18-Table 21), the Sign Test Method is given minimal weight in TTI's conclusions, due to the limitations discussed on page 41.

- Considering the TTI-modified TPF-5(384) method, volume groups above 5,000 AADT have 90% of all estimates within bounds, a positive result. The 500-4,999 volume group is 80%-85% of all estimates within bounds, which is slightly less than the higher volume groups. The 0-499 volume group is 40% and 63% and considered a fair-to-middling result. This result was given significant consideration in TTI's conclusions.

- Considering the Wilcoxon S-R Test, the 2 volume groups above 5,000 AADT (bidirectional) pass this test. The lower volume groups (less than 5,000 AADT directional and bidirectional) do not pass this test. This statistical hypothesis test was given significant consideration in TTI's conclusions.

- The last statistical hypothesis test, the Sign Test, passed all volume groups for bidirectional AADT, but only volume groups greater than 5,000 AADT for directional AADT. In general, this Sign Test is given minimal weight in TTI's conclusions, as some results are counterintuitive and inconsistent with other tests and evaluation measures.

TTI used professional judgment to weight and consider all of these evaluation results, acknowledging that the evaluation results are mixed, and some results are contradictory to others. TTI gave the Wilcoxon S-R Test and the TTI-modified TPF-5(384) method the most significant consideration in our conclusions.

The Wilcoxon S-R Test failed to reject the null hypothesis that there is no statistically significant difference between the two methods for bidirectional AADT data above 5,000 AADT. Therefore, TTI concluded that the differences as measured by the various summary descriptive statistics (e.g., SD, TCE, APE, RMSE) for the bidirectional AADT values is most likely from randomness and not a fundamental issue with StL's AADT estimation methods. This is a positive result for roads with bidirectional AADT above 5,000 vehicles per day.

The TTI-modified TPF-5(384) method indicated that volume groups above 5,000 AADT have 90% of all estimates within acceptance bounds. This is a positive result for roads with bidirectional AADT above 5,000 vehicles per day.

Therefore, based primarily on the Wilcoxon S-R Test and the TTI-modified TPF-5(384) method, TTI concluded that StL's AADT estimates are valid for traffic monitoring use on roads with bidirectional AADT of 5,000 or greater vehicles per day.

Recommendations for Next Steps

Because the evaluation results were mixed and the validity decision was not clear cut, TTI recommends a phased implementation that includes these two elements:

1. A one-to-three year transition period (DOTs still collecting short duration counts but comparing to StL AADT estimates) should be used to confirm TTI's professional judgment based on these mixed evaluation results. Any data produced by newly-installed CCSs should be withheld and not published anywhere, and used by DOTs only for validation purposes.

2. A pilot implementation program initiated by FHWA that includes up to five early adopter state DOTs, to monitor and coordinate findings in early-adopting states.

APPENDIX

This Appendix includes several charts and summary tables that were the result of the analysis in this project. For consistency with the chapter "Results", all the figures and tables in this appendix are organized into the following three subsections:

a) Exploratory data analysis
b) Metrics and statistics
c) Statistical methods

Exploratory Data Analysis

This subsection includes various plots developed as part of the EDA.

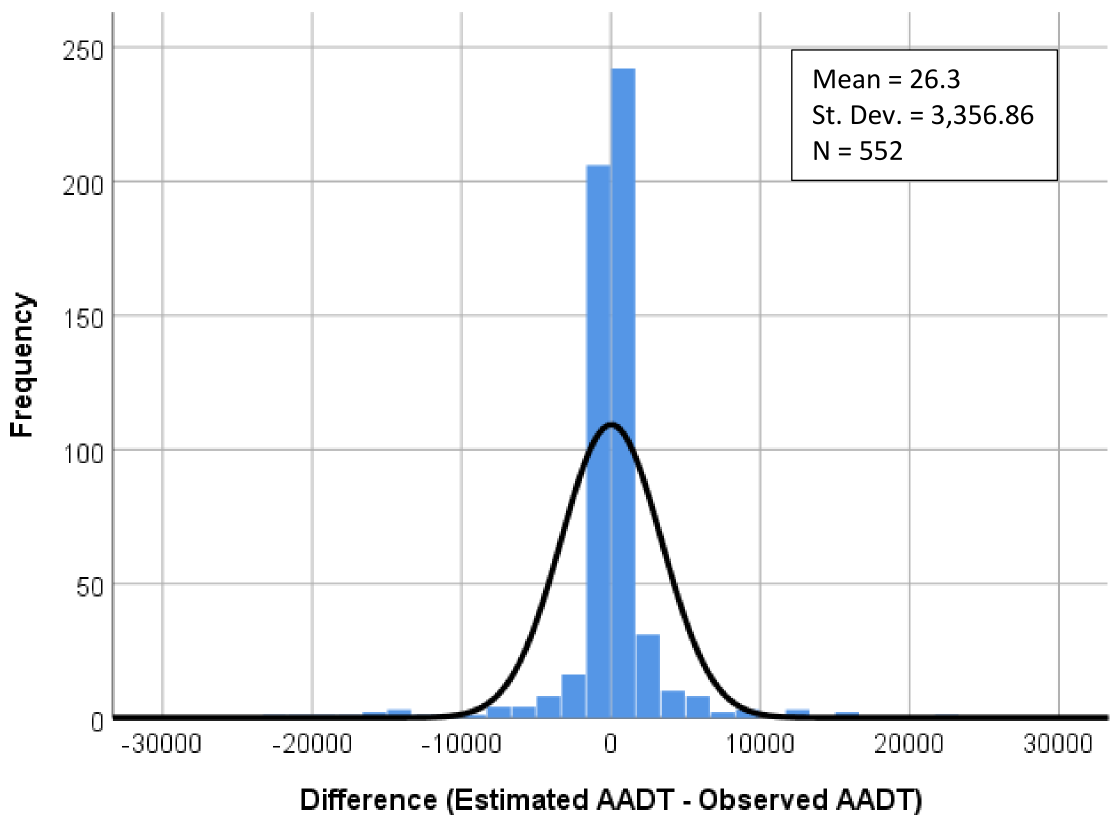

Figure 19. Histogram of Directional AADT Difference (States: CA, ME, MN, NJ, OR, TX)

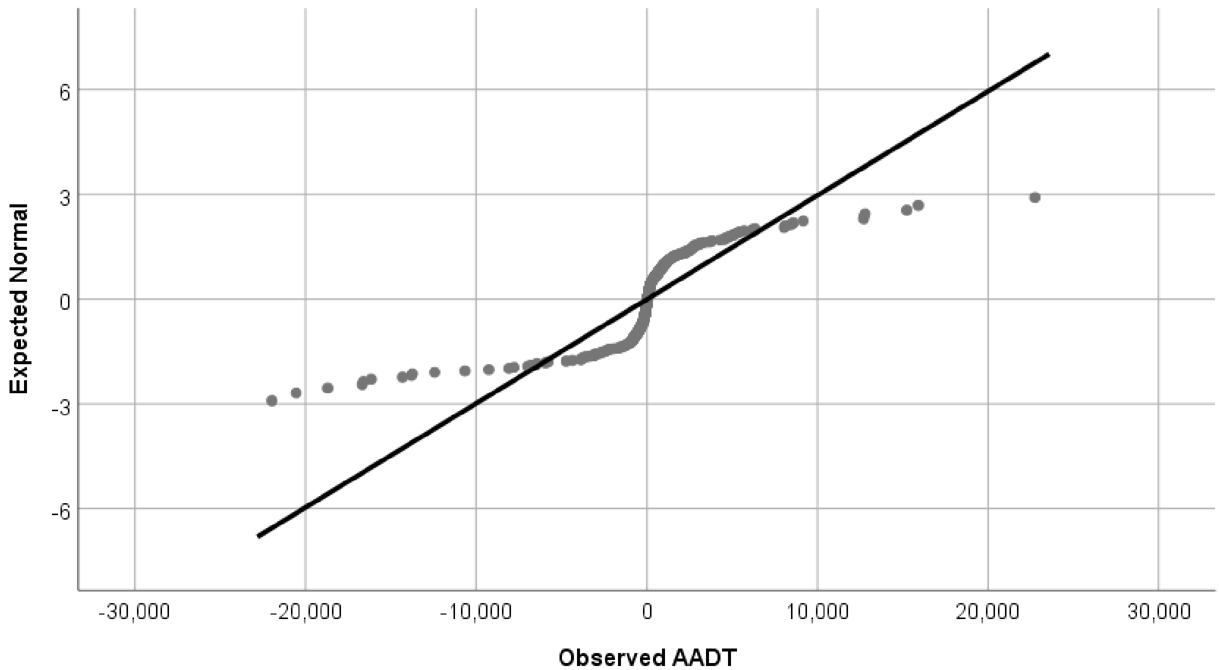

Figure 20. Normal Q-Q Plot of Directional AADT Difference (=Estimated AADT – Observed AADT) Data (States: CA, ME, MN, NJ, OR, TX)

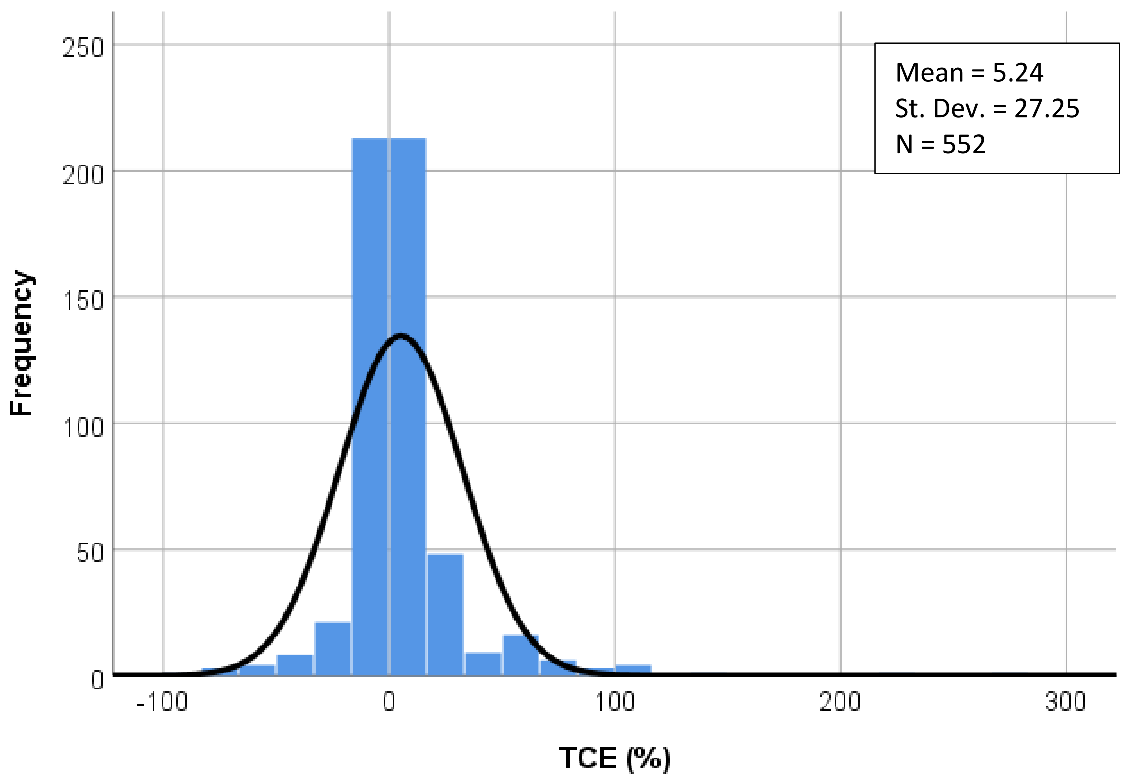

Figure 21. TCE Frequency Histogram – Directional AADT Data (States: CA, ME, MN, NJ, OR, TX)

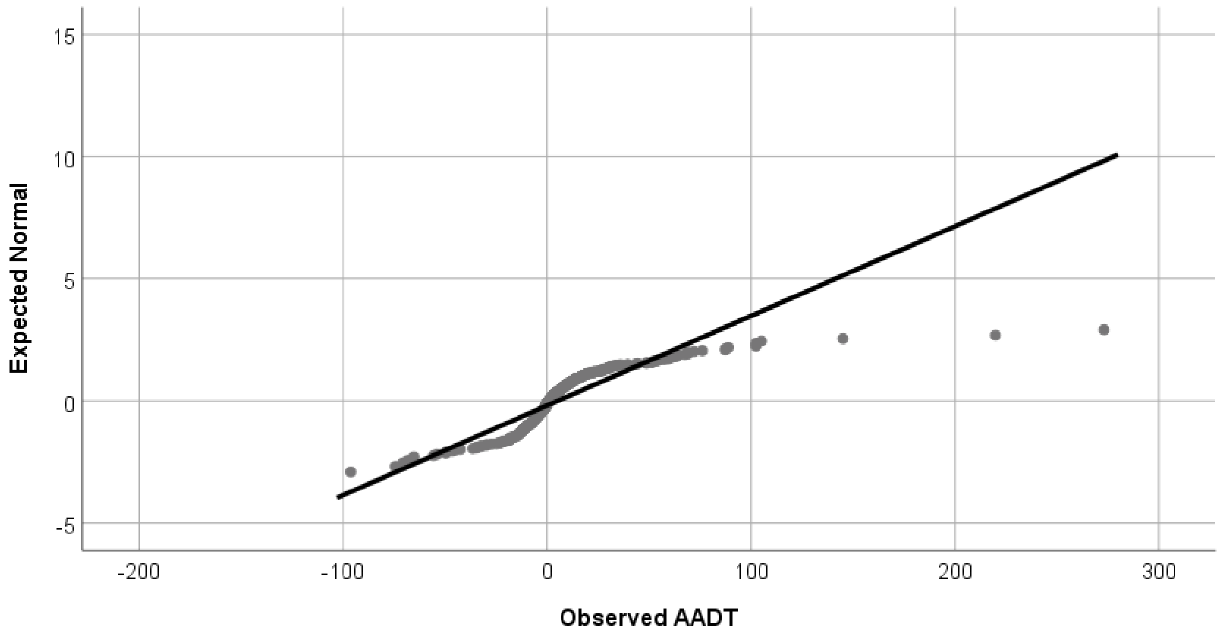

Figure 22. Normal Q-Q Plot of Directional TCE Data (States: CA, ME, MN, NJ, OR, TX)

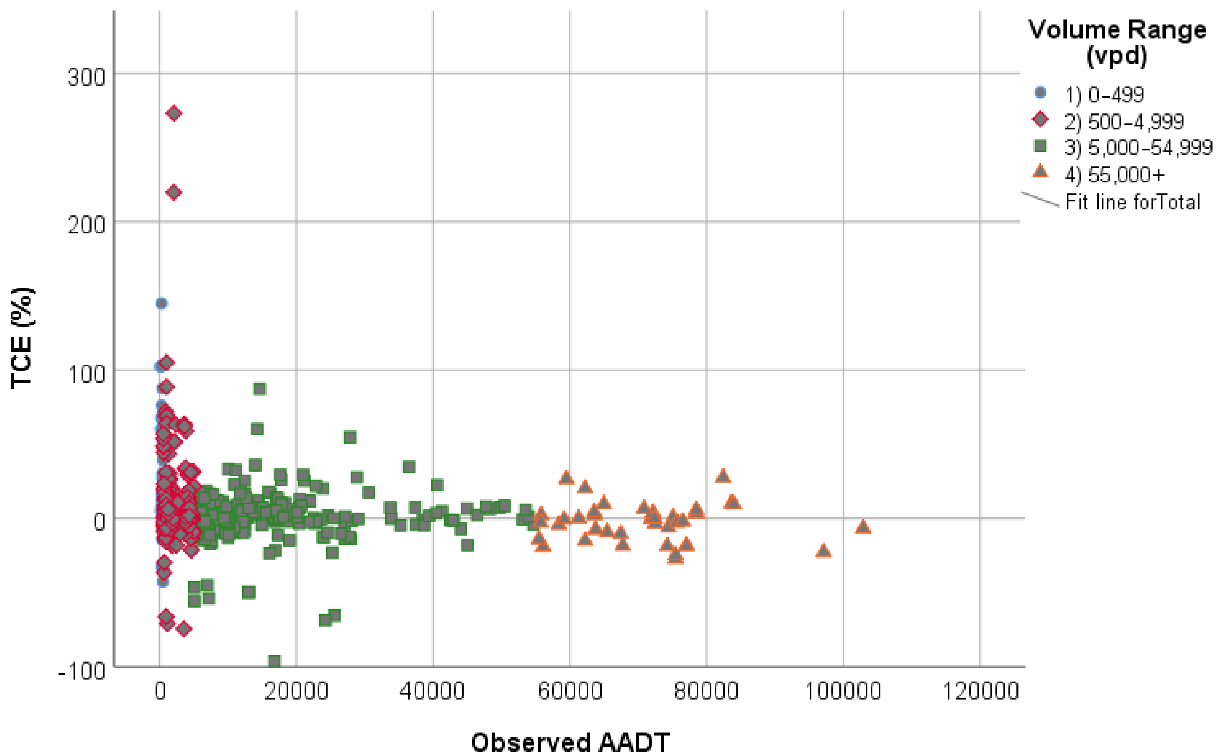

Figure 23. Scatterplot: TCE vs. Observed AADT – Directional AADT Data (States: CA, ME, MN, NJ, OR, TX)

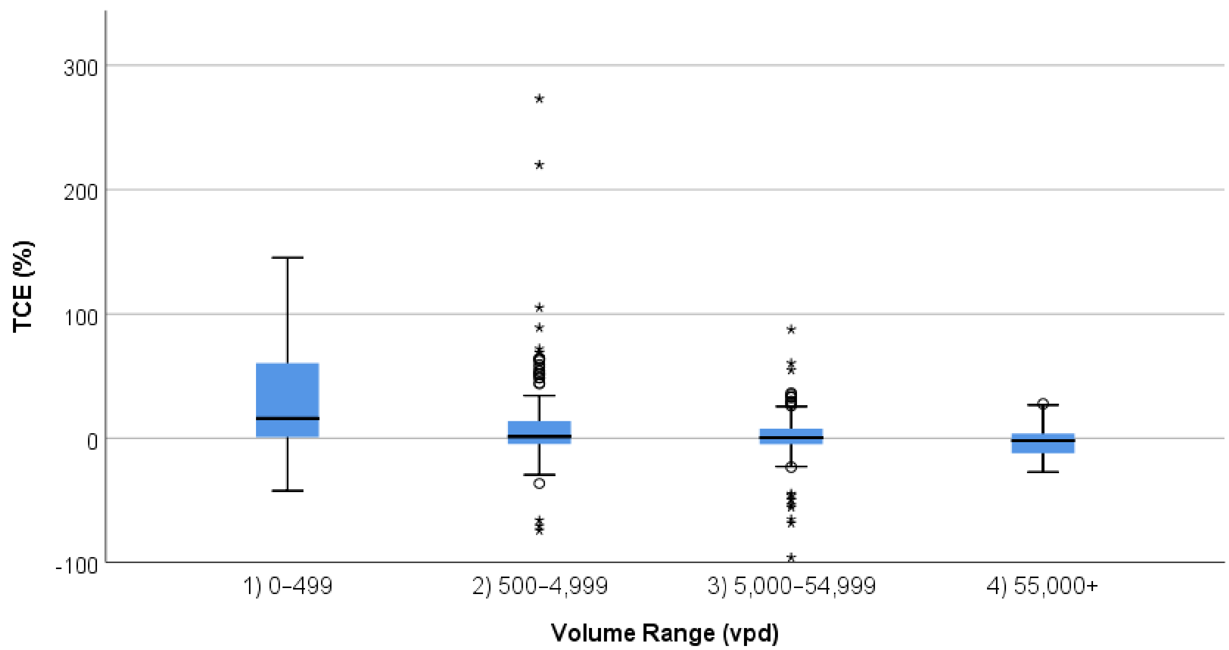

Figure 24. Boxplots: TCE by Volume Group – Directional AADT Data (States: CA, ME, MN, NJ, OR, TX)

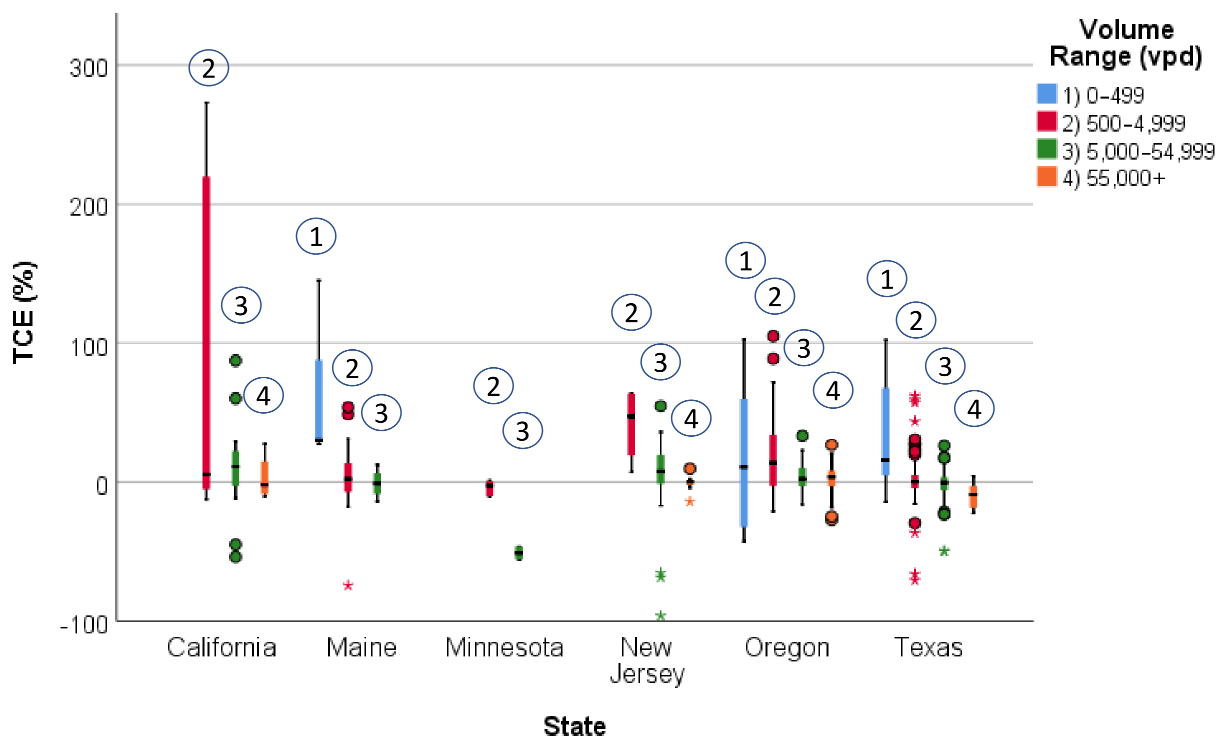

Figure 25. Boxplot: TCE by Volume Group and State – Directional AADT Data (States: CA, ME, MN, NJ, OR, TX)

Metrics and Summary Descriptive Statistics

This subsection includes various summary tables that present the metrics and summary descriptive statistics described in the Results chapter.

Table 24. Summary Statistics for Ten Volume Groups – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX)

Volume Range	N	Algebraic Diff.		TCE (%)		TCE Percentile				TCE Range		APE		NRMSE	R^2
		Mean	Median	Mean	Median	2.5th	16th	84th	97.5th	68%	95%	Mean	Median		
0–499	5	172.2	219.0	52.0	60.7	5.4	5.9	91.5	100.8	85.6	95.4	52.0	60.7	0.58	0.06
500–1,999	32	107.0	78.0	9.5	6.0	-34.1	-13.1	29.1	70.9	42.1	105.1	21.6	15.1	0.28	0.66
2,000–4,999	33	519.4	12.0	15.5	0.2	-25.9	-5.8	35.7	126.8	41.4	152.8	24.8	10.0	0.61	0.27
5,000–9,999	37	91.8	-20.0	1.2	-0.4	-28.0	-5.5	9.3	36.5	14.8	64.5	11.0	4.4	0.22	0.43
10,000–19,999	42	-435.5	-320.5	-3.5	-2.7	-48.6	-9.0	7.0	13.5	16.0	62.1	9.0	6.3	0.13	0.74
20,000–34,999	28	1078.1	564.5	3.9	1.9	-31.2	-4.4	12.6	42.9	17.0	74.1	10.9	5.6	0.20	0.46
35,000–54,999	17	-1897.2	17.0	-3.2	0.0	-46.2	-13.2	10.3	21.7	23.5	67.9	12.2	10.2	0.21	0.06
55,000–84,999	4	2113.0	2442.5	3.1	3.4	0.0	1.4	4.7	5.5	3.3	5.5	3.2	3.4	0.04	0.98
85,000–124,999	7	1200.9	-836.0	0.5	-0.8	-15.6	-9.1	8.6	21.2	17.7	36.8	8.5	8.0	0.12	0.70
125,000+	10	-7417.3	-2677.0	-4.7	-1.8	-21.7	-16.6	6.0	9.1	22.7	30.8	9.4	8.1	0.12	0.34
Total	**215**	**-245.8**	**17.0**	**4.7**	**0.2**	**-45.7**	**-8.9**	**15.7**	**71.7**	**24.7**	**117.4**	**15.0**	**6.6**	**0.25**	**0.97**

Table 25. Summary Statistics for Ten Volume Groups – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX)

Volume Range	N	Algebraic Diff.		TCE (%)		TCE Percentile				TCE Range		APE		NRMSE	R^2
		Mean	Median	Mean	Median	2.5th	16th	84th	97.5th	68%	95%	Mean	Median		
0–499	35	61.6	74.0	26.6	16.1	-35.8	-23.1	72.9	108.9	96.0	144.7	39.7	30.3	0.45	0.26
500–1,999	101	88.5	29.0	9.4	2.9	-33.0	-8.6	27.7	70.6	36.3	103.6	18.4	11.1	0.25	0.68
2,000–4,999	104	251.5	29.0	9.0	0.8	-19.2	-7.3	17.8	63.5	25.0	82.7	15.9	5.1	0.28	0.36
5,000–9,999	115	-105.6	-68.0	-1.7	-1.1	-45.1	-10.0	10.1	16.9	20.1	62.0	8.5	6.7	0.12	0.75
10,000–19,999	94	451.0	364.0	3.3	2.8	-41.0	-7.0	13.9	35.2	20.9	76.2	11.4	6.3	0.21	0.48
20,000–34,999	39	1.4	10.0	0.2	0.0	-65.4	-12.6	17.0	30.7	29.6	96.1	13.5	9.6	0.22	0.30
35,000–54,999	25	1263.3	717.0	3.0	1.8	-11.4	-4.1	7.6	27.5	11.7	38.9	6.7	4.6	0.09	0.66
55,000–84,999	37	-1509.5	-327.0	-2.3	-0.6	-25.0	-18.0	7.5	26.8	25.5	51.9	9.7	6.8	0.13	0.49
85,000–124,999	2	-14387.5	-14387.5	-14.6	-14.6	-22.2	-20.1	-9.2	-7.0	10.9	15.2	14.6	14.6	0.16	1.00
125,000+	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Total	**552**	**26.3**	**31.0**	**5.2**	**1.1**	**-35.0**	**-9.0**	**16.8**	**68.8**	**25.8**	**103.8**	**14.6**	**7.3**	**0.25**	**0.97**

Table 26. Summary Statistics by Rural/Urban Designation – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX)

Rural/Urban Code	N	Algebraic Diff.		TCE (%)		TCE Percentile				TCE Range		APE (%)		NRMSE	R^2
		Mean	Median	Mean	Median	2.5th	16th	84th	97.5th	68%	95%	Mean	Median		
Rural	137	346.2	38.0	6.4	0.9	-32.7	-8.3	20.0	73.3	28.3	106.1	15.8	7.0	0.26	0.97
Urban	78	-1285.7	-32.0	1.6	-0.2	-50.6	-10.3	10.1	31.9	20.4	82.5	13.7	6.3	0.19	0.97
Total	**215**	**-245.8**	**17.0**	**4.7**	**0.2**	**-45.7**	**-8.9**	**15.7**	**71.7**	**24.7**	**117.4**	**15.0**	**6.6**	**0.25**	**0.97**

Table 27. Summary Statistics by Rural/Urban Designation – Directional AADT Data (States: CA, ME, MN, NJ, OR, TX)

Rural/Urban Code	N	Algebraic Diff.		TCE (%)		TCE Percentile				TCE Range		APE (%)		NRMSE	R^2
		Mean	Median	Mean	Median	2.5th	16th	84th	97.5th	68%	95%	Mean	Median		
Rural	339	207.7	30.0	6.5	1.3	-31.3	-8.3	18.4	74.3	26.7	105.6	14.7	7.0	0.22	0.98
Urban	213	-262.5	39.0	3.2	0.5	-49.7	-10.9	13.8	57.7	24.7	107.4	14.4	7.4	0.20	0.96
Total	**552**	**26.3**	**31.0**	**5.2**	**1.1**	**-35.0**	**-9.0**	**16.8**	**68.8**	**25.8**	**103.8**	**14.6**	**7.3**	**0.25**	**0.97**

Table 28. Summary Statistics by Roadway Functional Class – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX)

Functional Class	N	Algebraic Diff.		TCE (%)		TCE Percentile				TCE Range		APE (%)		NRMSE	R^2
		Mean	Median	Mean	Median	2.5th	16th	84th	97.5th	68%	95%	Mean	Median		
1	50	-2255.3	-276.0	-1.6	-0.8	-17.1	-11.4	5.8	16.2	17.1	33.3	6.8	5.5	0.15	0.97
2	19	4797.7	2078.0	10.6	6.3	-5.5	-0.4	22.4	53.2	22.8	58.7	11.8	8.0	0.19	0.98
3	72	-408.2	-54.0	-0.6	-1.3	-49.5	-8.7	11.6	54.4	20.3	103.9	12.9	6.5	0.21	0.94
4	53	-170.6	38.0	10.0	1.2	-48.9	-8.8	27.2	79.8	36.1	128.6	20.1	8.4	0.76	0.67
5	21	342.3	147.0	18.7	6.3	-50.8	-2.9	68.3	99.7	71.2	150.5	32.2	20.4	0.60	0.88
Total	**215**	**-245.8**	**17.0**	**4.7**	**0.2**	**-45.7**	**-8.9**	**15.7**	**71.7**	**24.7**	**117.4**	**15.0**	**6.6**	**0.25**	**0.97**

Table 29. Summary Statistics by Roadway Functional Class – Directional AADT Data (States: CA, ME, MN, NJ, OR, TX)

Functional Class	N	Algebraic Diff.		TCE (%)		TCE Percentile				TCE Range		APE (%)		NRMSE	R^2
		Mean	Median	Mean	Median	2.5th	16th	84th	97.5th	68%	95%	Mean	Median		
1	123	-1048.1	-180.0	-1.6	-1.0	-22.4	-10.3	6.8	17.5	17.1	39.9	7.0	4.9	0.15	0.97
2	48	2161.8	805.5	9.1	5.0	-12.5	-2.2	24.1	55.5	26.3	68.1	11.8	6.2	0.19	0.96
3	207	64.7	13.0	1.5	0.5	-41.4	-9.2	12.4	53.7	21.6	95.0	11.6	6.9	0.29	0.92
4	122	63.6	43.5	10.5	3.5	-55.4	-10.6	27.3	69.2	37.9	124.6	20.5	10.7	0.60	0.88
5	52	355.9	92.5	20.4	7.4	-58.0	-7.1	61.7	104.4	68.7	162.4	32.8	20.4	0.54	0.99
Total	**552**	**26.3**	**31.0**	**5.2**	**1.1**	**-35.0**	**-9.0**	**16.8**	**68.8**	**25.8**	**103.8**	**14.6**	**7.3**	**0.25**	**0.97**

Table 30. Summary Statistics by Roadway Functional Class and Rural/Urban Designation – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX)

Functional Class	Rural/Urban Code	N	Algebraic Diff.		TCE (%)		TCE Percentile				TCE Range		APE (%)		NRMSE	R^2
			Mean	Median	Mean	Median	2.5th	16th	84th	97.5th	68%	95%	Mean	Median		
1	**Rural**	26	318.7	71.0	0.4	0.2	-13.5	-6.0	6.4	14.7	12.4	28.2	5.6	4.6	0.11	0.99
1	**Urban**	24	-5043.8	-1131.0	-3.8	-1.2	-19.7	-14.5	4.3	13.3	18.9	33.0	8.1	7.3	0.13	0.95
2	**Rural**	3	8074.3	3132.0	27.1	8.3	-0.7	1.9	53.0	70.7	51.1	71.4	27.8	8.3	0.41	0.23
2	**Urban**	16	4183.4	1868.0	7.5	6.0	-6.1	-0.1	18.7	26.2	18.8	32.3	8.8	7.1	0.17	0.99
3	**Rural**	48	-196.3	-75.5	0.6	-1.7	-37.5	-9.5	12.3	59.2	21.8	96.7	14.0	6.7	0.18	0.96
3	**Urban**	24	-832.0	-32.0	-2.9	-0.7	-49.5	-7.2	9.2	20.4	16.4	69.9	10.6	6.1	0.19	0.90
4	**Rural**	42	517.2	60.5	10.1	3.9	-14.2	-7.2	27.7	66.4	34.9	80.6	15.6	10.2	0.40	0.98
4	**Urban**	11	-2796.8	-164.0	9.5	-0.9	-66.0	-30.3	5.5	186.4	35.9	252.4	37.5	6.3	0.81	0.33
5	**Rural**	18	145.8	177.5	18.3	8.7	-53.4	-11.4	70.9	100.1	82.3	153.6	33.8	24.5	0.43	0.66
5	**Urban**	3	1521.3	117.0	21.2	4.7	-1.2	0.5	42.7	57.8	42.2	59.0	22.3	4.7	0.57	0.98
Total		**215**	**-245.8**	**17.0**	**4.7**	**0.2**	**-45.7**	**-8.9**	**15.7**	**71.7**	**24.7**	**117.4**	**15.0**	**6.6**	**0.25**	**0.97**

Table 31. Summary Statistics by Roadway Functional Class and Rural/Urban Designation – Directional AADT Data (States: CA, ME, MN, NJ, OR, TX)

Functional Class	Rural/Urban Code	N	Algebraic Diff.		TCE (%)		TCE Percentile				TCE Range		APE (%)		NRMSE	R^2
			Mean	Median	Mean	Median	2.5th	16th	84th	97.5th	68%	95%	Mean	Median		
1	**Rural**	66	129.3	-17.0	-0.1	-0.2	-13.8	-7.2	6.5	19.1	13.8	32.9	6.1	4.8	0.11	0.98
1	**Urban**	57	-2411.5	-447.0	-3.3	-1.2	-24.1	-15.0	6.8	17.3	21.8	41.3	8.0	5.9	0.14	0.93
2	**Rural**	6	4037.3	1566.0	27.0	8.4	-1.8	-0.6	65.7	84.0	66.3	85.8	27.8	8.4	0.42	0.23
2	**Urban**	42	1893.8	578.0	6.5	3.8	-12.9	-2.3	21.5	27.6	23.8	40.5	9.5	5.9	0.18	0.96
3	**Rural**	133	19.8	-7.0	2.2	-0.5	-27.3	-8.8	11.5	58.2	20.4	85.4	10.4	6.5	0.11	0.98
3	**Urban**	74	145.6	164.5	0.3	1.5	-50.5	-11.2	12.9	35.1	24.1	85.5	13.7	8.6	0.30	0.87
4	**Rural**	90	230.3	49.0	9.1	4.1	-15.3	-8.5	27.4	66.8	35.9	82.0	16.2	10.6	0.36	0.97
4	**Urban**	32	-405.4	23.5	14.7	0.7	-66.0	-10.9	17.5	231.8	28.4	297.8	32.6	10.8	0.57	0.85
5	**Rural**	44	324.9	99.5	21.6	10.9	-63.9	-7.2	69.6	104.9	76.7	168.8	35.3	23.0	0.56	0.99
5	**Urban**	8	526.5	58.5	13.9	2.6	-15.3	-2.6	52.7	61.6	55.3	76.9	19.2	4.7	0.44	0.81
Total		**552**	**26.3**	**31.0**	**5.2**	**1.1**	**-35.0**	**-9.0**	**16.8**	**68.8**	**25.8**	**103.8**	**14.6**	**7.3**	**0.25**	**0.97**

Table 32. Summary Statistics by State and Rural/Urban Designation – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX)

State	Rural/ Urban Code	N	Algebraic Diff.		TCE (%)		TCE Percentile				TCE Range		APE (%)		NRMSE	R^2
			Mean	Median	Mean	Median	2.5th	16th	84th	97.5th	68%	95%	Mean	Median		
California	**Rural**	5	5876.2	3179.0	2.19	11.05	-83.5	-39.9	44.8	69.4	84.6	152.9	43.1	28.3	0.51	0.91
California	**Urban**	4	2104.0	2574.0	52.71	6.85	-46.0	-27.5	136.6	229.4	164.1	275.4	79.3	33.5	0.43	0.70
Maine	**Rural**	5	-125.0	208.0	7.28	3.65	-8.5	-5.7	21.2	27.5	26.9	36.0	12.4	9.0	0.11	0.99
Minnesota	**Rural**	3	-352.7	-94.0	-4.23	-2.56	-10.0	-7.9	-0.6	0.1	7.2	10.1	4.4	2.6	0.09	1.00
Minnesota	**Urban**	1	-6424.0	-6424.0	-63.27	-63.27	-63.3	-63.3	-63.3	-63.3	0.0	0.0	63.3	63.3	0.63	
New Jersey	**Rural**	4	-7059.8	-172.5	-11.51	-0.72	-61.9	-35.4	11.5	20.6	46.9	82.5	22.6	11.8	0.56	0.65
Oregon	**Rural**	49	480.4	229.0	12.91	5.76	-32.5	-12.9	48.2	91.6	61.1	124.1	22.4	13.0	0.18	0.99
Oregon	**Urban**	22	1869.7	1764.0	5.66	4.02	-15.5	-3.6	13.5	28.7	17.1	44.2	9.8	7.6	0.16	0.97
Texas	**Rural**	75	-50.7	17.0	2.72	0.04	-17.8	-6.7	11.2	43.0	17.9	60.9	10.3	5.2	0.13	0.99
Texas	**Urban**	47	-2450.5	-129.0	-2.07	-0.86	-21.6	-11.1	5.7	10.1	16.8	31.7	8.2	5.8	0.17	0.98
Total		**215**	**-245.8**	**17.0**	**4.65**	**0.25**	**-45.7**	**-8.9**	**15.7**	**71.7**	**24.7**	**117.4**	**15.0**	**6.6**	**0.25**	**0.97**

Table 33. Summary Statistics by State and Rural/Urban Designation – Directional AADT Data (States: CA, ME, MN, NJ, OR, TX)

State	Rural/ Urban Code	N	Algebraic Diff.		TCE (%)		TCE Percentile				TCE Range		APE (%)		NRMSE	R^2
			Mean	Median	Mean	Median	2.5th	16th	84th	97.5th	68%	95%	Mean	Median		
California	**Rural**	15	3180.1	1516.0	16.25	6.67	-10.0	-2.5	27.8	77.9	30.3	87.9	19.2	10.3	0.22	0.95
California	**Urban**	11	1548.5	-99.0	38.78	-1.59	-51.6	-24.9	104.5	259.8	129.4	311.4	62.2	17.9	0.28	0.95
Maine	**Rural**	28	-235.9	51.5	9.01	1.97	-33.5	-9.4	29.3	83.5	38.6	117.0	20.9	10.1	0.23	0.98
Maine	**Urban**	8	85.9	235.0	2.18	3.47	-15.7	-4.2	10.2	12.1	14.4	27.8	7.9	6.8	0.08	1.00
Minnesota	**Rural**	6	-171.0	-47.5	-4.03	-2.63	-10.5	-10.3	0.7	1.2	11.0	11.7	4.8	3.5	0.10	0.99
Minnesota	**Urban**	2	-2592.5	-2592.5	-51.05	-51.05	-55.4	-54.2	-47.9	-46.7	6.3	8.7	51.1	51.1	0.51	1.00
New Jersey	**Rural**	4	910.5	-674.5	7.20	5.25	-9.4	-8.3	22.9	27.1	31.2	36.5	15.5	13.5	0.17	0.90
New Jersey	**Urban**	42	673.3	1206.0	7.38	7.38	-68.5	-3.2	30.3	63.1	33.5	131.5	20.5	13.3	0.24	0.93
Oregon	**Rural**	100	216.9	104.0	12.48	5.21	-33.5	-12.3	49.0	96.0	61.3	129.5	22.3	12.3	0.17	0.99
Oregon	**Urban**	50	652.5	916.5	4.91	4.07	-23.3	-3.6	13.6	29.7	17.2	53.0	10.2	8.4	0.17	0.96
Texas	**Rural**	186	27.0	7.5	2.48	0.43	-14.9	-6.8	10.9	48.7	17.7	63.6	9.6	5.1	0.11	0.99
Texas	**Urban**	100	-1293.4	-81.5	-2.10	-1.43	-23.2	-11.5	5.6	21.9	17.1	45.1	8.5	4.4	0.19	0.98
Total		**552**	**26.3**	**31.0**	**5.2**	**1.1**	**-35.0**	**-9.0**	**16.8**	**68.8**	**25.8**	**103.8**	**14.6**	**7.3**	**0.25**	**0.97**

Statistical Testing Methods

This subsection includes various summery tables that provide the main results of the four statistical methods described in the Results chapter.

Table 34. Summary Results by State and Volume Group – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX)

State	Volume Range	N	TPF-5(384) Method		TTI Modified Method		Wilcoxon S-R Test	
			Accuracy Targets	Precision Targets	Points within Bounds: Number	Points within Bounds: Percent	Z Value	p Value
California	**500–4,999**	2			1	50%	-0.447	0.655
California	**5,000–54,999**	7	N		3	43%	-0.676	0.499
Maine	**500–4,999**	3			3	100%	-1.069	0.285
Maine	**5,000–54,999**	2			2	100%	-0.447	0.655
Minnesota	**500–4,999**	2			2	100%	-0.447	0.655
Minnesota	**5,000–54,999**	2			1	50%	-1.342	0.180
New Jersey	**5,000–54,999**	4			3	75%	-0.73	0.465
Oregon	**0–499**	2			0	0%	-1.342	0.180
Oregon	**500–4,999**	23	N		15	65%	-2.062	0.039
Oregon	**5,000–54,999**	37	N		33	89%	-2.165	0.030
Oregon	**55,000+**	9	N		7	78%	-0.889	0.374
Texas	**0–499**	3			2	67%	-1.604	0.109
Texas	**500–4,999**	35	N		31	89%	-0.966	0.334
Texas	**5,000–54,999**	72	N		70	97%	-1.352	0.176
Texas	**55,000+**	12	N		12	100%	-1.49	0.136
Total		**215**			**185**	**86%**	**-0.866**	**0.387**

Table 35. Summary Results by State and Volume Group – Directional AADT Data (States: CA, ME, MN, NJ, OR, TX)

State	Volume Range	N	TPF-5(384) Method		TTI Modified Method		Wilcoxon S-R Test	
			Accuracy Targets	Precision Targets	Points within Bounds		Z Value	p Value
					Number	Percent		
California	**500–4,999**	6	N		4	67%	-1.363	0.173
California	**5,000–54,999**	16	N		11	69%	-1.913	0.056
California	**55,000+**	4			3	75%	0	1.000
Maine	**0–499**	3			2	67%	-1.604	0.109
Maine	**500–4,999**	21	N		18	86%	-0.678	0.498
Maine	**5,000–54,999**	12	N		12	100%	-0.941	0.347
Minnesota	**500–4,999**	6	N		6	100%	-0.943	0.345
Minnesota	**5,000–54,999**	2			0	0%	-1.342	0.180
New Jersey	**500–4,999**	4			1	25%	-1.826	0.068
New Jersey	**5,000–54,999**	35	N		26	74%	-2.342	0.019
New Jersey	**55,000+**	7	N		7	100%	-0.169	0.866
Oregon	**0–499**	18			11	61%	-0.501	0.616
Oregon	**500–4,999**	51	N		35	69%	-3.81	0.000
Oregon	**5,000–54,999**	66	N		65	98%	-2.274	0.023
Oregon	**55,000+**	15	N		12	80%	-0.625	0.532
Texas	**0–499**	14			9	64%	-2.605	0.009
Texas	**500–4,999**	117	Y	N	110	94%	-0.109	0.913
Texas	**5,000–54,999**	142	N	N	140	99%	-1.186	0.236
Texas	**55,000+**	13	N		13	100%	-2.83	0.005
Total		**552**			**485**	**88%**	**-2.673**	**0.008**

Table 36. Summary Results by Rural/Urban Designation – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX)

Rural/Urban Code	N	TPF-5(384) Method		TTI Modified Method		Wilcoxon S-R Test	
		Accuracy Targets	Precision Targets	Points within Bounds		Z Value	p Value
				Number	Percent		
Rural	137			117	85%	-1.413	0.158
Urban	78			68	87%	-0.304	0.761
Total	**215**			**185**	**86%**	**-0.866**	**0.387**

Table 37. Summary Results by Rural/Urban Designation – Directional AADT Data (States: CA, ME, MN, NJ, OR, TX)

Rural/Urban Code	N	TPF-5(384) Method		TTI Modified Method		Wilcoxon S-R Test	
		Accuracy Targets	Precision Targets	Points within Bounds		Z Value	p Value
				Number	Percent		
Rural	339			298	88%	-2.449	0.014
Urban	213			187	88%	-1.143	0.253
Total	**552**			**485**	**88%**	**-2.673**	**0.008**

Table 38. Summary Results by State and Rural/Urban Designation – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX)

State	Rural/ Urban Code	N	TPF-5(384) Method		TTI Modified Method		Wilcoxon S-R Test	
			Accuracy Targets	Precision Targets	Points within Bounds		Z Value	p Value
					Number	Percent		
California	**Rural**	5			2	40%	-0.944	0.345
California	**Urban**	4			2	50%	-0.365	0.715
Maine	**Rural**	5			5	100%	-0.405	0.686
Minnesota	**Rural**	3			3	100%	-1.069	0.285
Minnesota	**Urban**	1			0	0%		
New Jersey	**Rural**	4			3	75%	-0.73	0.465
Oregon	**Rural**	49			37	76%	-2.256	0.024
Oregon	**Urban**	22			18	82%	-1.997	0.046
Texas	**Rural**	75			70	93%	-0.129	0.897
Texas	**Urban**	47			45	96%	-1.656	0.098
Total		**215**			**185**	**86%**	**-0.866**	**0.387**

Table 39. Summary Results by State and Rural/Urban Designation – Directional AADT Data (States: CA, ME, MN, NJ, OR, TX)

State	Rural/ Urban Code	N	TPF-5(384) Method		TTI Modified Method		Wilcoxon S-R Test	
			Accuracy Targets	Precision Targets	Points within Bounds		Z Value	p Value
					Number	Percent		
California	**Rural**	15			12	80%	-2.613	0.009
California	**Urban**	11			6	55%	0.000	1.000
Maine	**Rural**	28			24	86%	-0.205	0.838
Maine	**Urban**	8			8	100%	-0.84	0.401
Minnesota	**Rural**	6			6	100%	-0.943	0.345
Minnesota	**Urban**	2			0	0%	-1.342	0.180
New Jersey	**Rural**	4			3	75%	0.000	1.000
New Jersey	**Urban**	42			31	74%	-2.807	0.005
Oregon	**Rural**	100			77	77%	-2.746	0.006
Oregon	**Urban**	50			46	92%	-2.853	0.004
Texas	**Rural**	186			176	95%	-0.276	0.783
Texas	**Urban**	100			96	96%	-2.84	0.005
Total		**552**			**485**	**88%**	**-2.673**	**0.008**

Table 40. Summary Results by Roadway Functional Class – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX)

Functional Class	N	TPF-5(384) Method		TTI Modified Method		Wilcoxon S-R Test	
		Accuracy Targets	Precision Targets	Points within Bounds		Z Value	p Value
				Number	Percent		
1	50			49	98%	-1.149	0.251
2	19			16	84%	-2.938	0.003
3	72			63	88%	-0.527	0.598
4	53			43	81%	-0.899	0.369
5	21			14	67%	-1.964	0.050
Total	**215**			**185**	**86%**	**-0.866**	**0.387**

Table 41. Summary Results by Roadway Functional Class – Directional AADT Data (States: CA, ME, MN, NJ, OR, TX)

Functional Class	N	TPF-5(384) Method		TTI Modified Method		Wilcoxon S-R Test	
		Accuracy Targets	Precision Targets	Points within Bounds		Z Value	p Value
				Number	Percent		
1	123			120	98%	-1.816	0.069
2	48			43	90%	-3.395	0.001
3	207			189	91%	-1.33	0.183
4	122			98	80%	-2.677	0.007
5	52			35	67%	-3.115	0.002
Total	**552**			**485**	**88%**	**-2.673**	**0.008**

Table 42. Summary Results by Roadway Functional Class and Rural/Urban Code – Bidirectional AADT Data (States: CA, ME, MN, NJ, OR, TX)

Functional Class	Rural/ Urban Code	N	TPF-5(384) Method		TTI Modified Method		Wilcoxon S-R Test	
			Accuracy Targets	Precision Targets	Points within Bounds		Z Value	p Value
					Number	Percent		
1	**Rural**	26			26	100%	-0.508	0.611
1	**Urban**	24			23	96%	-1.771	0.076
2	**Rural**	3			2	67%	-1.069	0.285
2	**Urban**	16			14	88%	-2.689	0.007
3	**Rural**	48			42	88%	-0.64	0.522
3	**Urban**	24			21	88%	-0.257	0.797
4	**Rural**	42			35	83%	-1.701	0.089
4	**Urban**	11			8	73%	-1.023	0.306
5	**Rural**	18			12	67%	-1.677	0.094
5	**Urban**	3			2	67%	-1.069	0.285
Total		**215**			**185**	**86%**	**-0.866**	**0.387**

Table 43. Summary Results by Roadway Functional Class and Rural/Urban Code – Directional AADT Data (States: CA, ME, MN, NJ, OR, TX)

Functional Class	Rural/ Urban Code	N	TPF-5(384) Method		TTI Modified Method		Wilcoxon S-R Test	
			Accuracy Targets	Precision Targets	Points within Bounds		Z Value	p Value
					Number	Percent		
1	**Rural**	66			65	98%	-0.118	$\boxed{0.906}$
1	**Urban**	57			55	96%	-2.233	0.026
2	**Rural**	6			4	67%	-1.572	$\boxed{0.116}$
2	**Urban**	42			39	93%	-2.97	0.003
3	**Rural**	133			125	94%	-0.243	$\boxed{0.808}$
3	**Urban**	74			64	86%	-1.425	$\boxed{0.154}$
4	**Rural**	90			75	83%	-2.719	0.007
4	**Urban**	32			23	72%	-0.841	$\boxed{0.400}$
5	**Rural**	44			29	66%	-2.993	0.003
5	**Urban**	8			6	75%	-1.12	$\boxed{0.263}$
Total		**552**			**485**	**88%**	**-2.673**	**0.008**