Non-Traditional Methods to Obtain Annual Average Daily Traffic (AADT)

Publication No. FHWA-PL-21-030

September 2021

U.S. Department of Transportation
Federal Highway Administration

Notice

Quality Assurance Statement

Technical Report Documentation Page

1. Report No. FHWA-PL-21-030	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Non-Traditional Methods to Obtain Annual Average Daily Traffic (AADT)		5. Report Date September 2021 6. Performing Organization Code
7. Author(s) Laura Schewel, Sean Co, Christy Willoughby, Louis Yan, Nancy Clarke, Jon Wergin		8. Performing Organization Report No.
9. Performing Organization Name and Address StreetLight Data 677 Harrison Street San Francisco, CA 94107		10. Work Unit No. (TRAIS) 11. Contract or Grant No. 693JJ319C000015
12. Sponsoring Agency Name and Address Federal Highway Administration Office of Highway Policy Information 1200 New Jersey Avenue SE Washington, DC 20590		13. Type of Report and Period Covered 14. Sponsoring Agency Code

15. Supplementary Notes

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

Steven Jessberger (Task Manager)

16. Abstract

The use of passive data from location-based smartphone applications (LBS) and Global Positioning Services (GPS) to collect Annual Average Daily Traffic (AADT) has the potential to greatly reduce costs to State Department of Transportations (DOTs) and Metropolitan Planning Organizations (MPOs) and expand the coverage of up-to-date counts. This report evaluates the technical and statistical validity of traffic data derived from these sources using machine learning methods. Validity was determined by comparison to 4255 permanent counters, and a survey of recent publications about accuracy expectations.

The document covers the input data and the development of the machine learning models and model validation. The results include the error by road volume, roadway and regional characteristics compared to typical estimation. The effects of reduced trip sample, ping rate, spatial accuracy and reference counters were also tested. The applicability of Probe Data was tested for other factors including, day of week, month of year, directional and ramp AADT, work zones ADT, K and D factors, peak hour truck data, special events or unusual weather and AADT by vehicle type.

17. Key Words AADT, probe data, volume, Big Data, LBS, GPS, Machine Learning, MADT, Volume, AADTT, K-Factor, D-Factor, AAHT		18. Distribution Statement No restrictions. This document is available to the public.	
19. Security Classif.(of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 151	22. Price

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

Preface

We would like to acknowledge the following Technical Advisory Committee members for their contributions, support and technical guidance during this project.

Scott Vockeroth	Alaska DOT
Afrid Sarker	Caltrans
Steve Abeyta	Colorado DOT
Eric Conklin	Georgia DOT
John Phillips	Idaho DOT
William (Bill) Morgan	Illinois DOT
Lisa Shemer	Maryland DOT
Gene Hicks	Minnesota DOT
David Schoenmaker	Nebraska DOT
Chris Zajac	New Jersey DOT
Kent Taylor	North Carolina DOT
Terry Woehl	North Dakota DOT
Josh Roll	Oregon DOT
Sandie Mapel	Ohio DOT
Greg Dunmire	Pennsylvania DOT
Todd Anderson	South Carolina DOT
Chris Didear	Texas DOT
Hamlin Williams	Virginia DOT

Non-Traditional Methods to Obtain Annual Average Daily Traffic (AADT)

Table of Contents

Abstract 3

Executive Summary 3

Chapter 1: Introduction 4

Chapter 2: Data 7

Permanent Traffic Counter Data 7

Probe Data - LBS and GPS Trips 12

LBS Data 12

GPS Data from Commercial Trucks 13

Creation of Sample Trips 13

Outlier Detection and Diagnostics for Probe Data 15

Predictive Features for Estimating Traffic Volume 16

Chapter 3: Estimating AADT Using Probe Data Inputs and Machine Learning 17

Methods for Estimating AADT with Probe Data 17

Candidate Machine Learning Models 17

Model Validation Metrics 19

Results 21

Correlation of Probe Data Trip Sample Count With AADT 21

Model Comparison for Estimating AADT with Probe Data 24

National Probe Data AADT Model for Estimating AADT From Probe Data 27

National Probe Data AADT Model Error by Road Volume 28

Model Error Across Roadway and Regional Characteristics 36

Conclusions 40

Chapter 4: Impact of Probe Data Characteristics on the Traffic Volume Estimation 41

Effects of Time Periods Sampled on AADT Model Accuracy 41
Forecasting AADT Mid-Year 44
Effects of Reduced Trip Sample on Probe Data AADT Model Accuracy 49
Effects of Reduced Ping Rate and Spatial Accuracy on MADT Model Accuracy - Cellular Data Simulation 52
Impact of Reduced Reference Counters on AADT estimation 56
One-State Calibration 59
Conclusions: Estimation of AADT With Probe Data 60
Chapter 5: Estimation of MADT using Probe Data 62
Method 62
Results 63
Chapter 6: Estimation of Daily Traffic Using Probe Data - Including Special Events 66
Methods 66
Summary of Results 66
Conclusions 72
Chapter 7: Estimation of Hourly Traffic Using Probe Data 73
Summary of Results 73
Conclusions 78
Chapter 8: Traffic Items 79
Vehicle Classification AADTs 79
Introduction 79
Methods 79
Results 80
Estimation of AADT for PV Vehicles 80
Estimation of AADT for SU Trucks and Buses 81
Estimation of AADT for CU Trucks 82
Conclusions 83
Estimation of Seasonal Factors 83
Methods 83
Estimation of Month of Year Factors 84
Estimation of Day of Week Factors 88
Estimation of Time of Day Factors 91
Estimation of K Factor 100

Methods 100
Accuracy and Precision of K-factor Estimation 102
Estimation of D Factor 103
Methods 104
Summary of Results: 105
Conclusions 107
Estimation of Percent Peak CU and SU Trucks 107
Methods 107
Estimation of Percent Peak Combination-Unit (CU) Trucks 108
Conclusions 110
Chapter 9: Conclusions 111
References 114
Appendix 116

Non-Traditional Methods to Obtain Annual Average Daily Traffic (AADT)

Abstract

The use of passive data from location-based smartphone applications (LBS) and Global Positioning Services (GPS) to collect Annual Average Daily Traffic (AADT) has the potential to greatly reduce costs to State Department of Transportations (DOTs) and Metropolitan Planning Organizations (MPOs) and expand the coverage of up-to-date counts. This report evaluates the technical and statistical validity of traffic data derived from these sources using machine learning methods. Validity was determined by comparison to 4255 permanent counters, and a survey of recent publications about accuracy expectations.

The document covers the input data and the development of the machine learning models and model validation. The results include the error by road volume, roadway and regional characteristics compared to typical estimation. The effects of reduced trip sample, ping rate, spatial accuracy and reference counters were also tested. The applicability of Probe Data was tested for other factors including, day of week, month of year, directional and ramp AADT, work zones ADT, K and D factors, peak hour truck data, special events or unusual weather and AADT by vehicle type.
Key words: AADT, probe data, volume, Big Data, LBS, GPS, Machine Learning, MADT, Volume, AADTT, K-Factor, D-Factor, AAHT

Executive Summary

Traffic volume information, especially Annual Average Daily Traffic (AADT), serves as the basis of many transportation applications, highway planning, roadway geometry and pavement design, safety analysis, congestion management, and policy decision-making. This report is the first of its kind to broadly examine the validity of using Big Data from location-based smartphone applications (LBS) and Global Positioning Services (GPS) data to estimate AADT.

This research is part of FHWA's two-part initiative to evaluate the technical and statistical validity of traffic data derived from passively collected data from various sources. The research covered in this report is a response to the pooled fund solicitation, "Non-Traditional Methods to Obtain Annual Average Daily Traffic (AADT)." The second part, "Validation of Non-Traditional Methods to Obtain Annual Average Daily Traffic (AADT), conducted by another team, will further validate the findings from this research.

The items for evaluation include the following:

1. Evaluate the estimation of AADT and compare the precision and accuracy against the AADT data derived from (a) the traditional permanent counter approach and (b) factoring up methods based on 48-hr counts;
2. Evaluate the effect of impoverished Probe Data on the traffic volume estimation accuracy;
3. Evaluate the applicability of future AADT estimation using Probe Data;
4. Evaluate the applicability of Probe Data for estimating other traffic volume information, including:
    a. Factors such as day of week (DOW) and month of year (MOY) - Chapter 8;
    b. Directional AADT – volume only - Chapter 3;
    c. Ramp AADT - Chapter 3;
    d. Work Zone ADTs - Chapter 6;
    e. K Factors (Design Hour Volumes) and or D Factors (Percentage of traffic moving in the peak travel direction) - Chapter 8;
    f. Peak hour truck data - Chapter 8
    g. Special events or unusual weather conditions ADT - Chapter 6;
    h. Methods to deploy the new AADT data - Chapter 4;
    i. AADT by vehicle type – 3 classification groups (FHWA classes:1-3, 4-7, and 8-13) - Chapter 8;

The results for AADT estimations across road classifications and locations consistently out-performed same-year short term counters for roads over 2,000 AADT. For roads between 500 and 2000 AADT, results were mixed. Breakdown of MAPE by state for 500-2,000 AADT roads reveled that Probe Data outperformed same-year two-day expansion for all but five states. Using Big Data to estimate Monthly Average Daily Traffic (MADT) and Special event estimates successfully captured atypical events. For directional, ramp, and AADT by vehicle type (personal car and truck), the errors were better than or similar to same-year two-day expansion benchmark metrics. K- and D-factor estimations had relatively high error for extreme values. More work and research are needed for strategies to model these values with higher accuracy using Probe Data, specifically researching finding new reference roads with extreme values that could be used to better calibrate a model.

Chapter 1: Introduction

Traffic volume information, especially the Annual Average Daily Traffic (AADT), serves as the basis of many transportation applications, highway planning, roadway geometry and pavement design, safety analysis, congestion management, and policy decision-making. To collect the traffic volumes, the most accurate and reliable way is through the pavement-embedded sensors, such as loop detectors, along with roadside-based radars and other permanent, fixed-point installed detection systems. However, such traditional detection systems are expensive to install, operate, and maintain. Therefore, implementing continuous collection of traffic volume information on every single road segment of a vast spatio-temporal transportation network is impractical.

To illustrate how difficult, it is for public agencies to obtain traffic volumes from permanent counters, the count type and methodology for reporting AADT was analyzed for the Minnesota Department of Transportation in 2019. States report AADT data with temporary counters as well as permanent counters. The AADT values can be reported using temporary counters that are typically short duration 48-hour portable counts that are then expanded to compute annual counts. Due to budget or staffing constraints, states may use temporary counters collected 2-3 years ago, counters collected 4-6 years ago, or may have no count but instead extrapolate from similar locations. Minnesota's statistics are given as an example – though states vary significantly in their split between count methods. As shown in Table 1 below, only 1% of the total mileage of roadways in Minnesota where some counts were collected was conducted by a permanent counter. Overall, no counts were the most common (57%), though all of these were concentrated on roads believed to be fewer than 1,000 AADT, and often are off the federal aid roadway system. Temporary counts that were collected 2-4 years ago were the next most common with 23% of the mileage covered by these types of counts.

Table 1: Percentage of roadway mileage by count type for the Minnesota Department of Transportation in 2019.

Road Size Bin	Permanent counter	Same-year temporary	Temporary 2-4 years prior	Temporary 4-12 years prior	No counts
A: 0-500	0%	5%	17%	12%	67%
B: 500-1,999	1%	15%	30%	2%	52%
C: 2,000 - 4,999	2%	36%	60%	2%	0%
D: 5,000 - 9,999	4%	39%	55%	2%	0%
E: 10,000 - 19,999	4%	33%	59%	3%	0%
F: 20,000 - 34,999	11%	31%	53%	4%	0%
G: 35,000 - 54,999	17%	42%	37%	4%	0%
H: 55,000 - 84,999	24%	66%	9%	1%	0%
I: 85,000 - 12,499	16%	78%	6%	0%	0%
J: 125,000+	27%	67%	6%	0%	0%
Total	**1%**	**10%**	**23%**	**9%**	**57%**

While 10+ years ago, data from mobile sources such as cell phone towers was found to be insufficient for volume estimation, in recent years, mobile location and communication technologies, including some passive data options for such sources, have shown more promise in this area. These will be referred to as probe data in this report. Note that data from probes alone is not sufficient to estimate volume – thus, when the term "Probe Data" is used in this report to refer to the overall approach, it implies the use of probe sources in conjunction with other contextual sources of data, such as permanent counters for calibration, weather data, etc. Probe sources have been explored in many transportation fields, including travel pattern characterization, traffic volume and vehicle miles travelled (VMT) estimation, multi-mode classification, and origin-destination travel activities recognition. Probe Data sources include location-based service (LBS) activities where the location and timestamp are collected from travelers' smartphone applications, Global Positioning Services (GPS), Automatic Vehicle Locators (AVL) electronic logs, and cellular communication devices. As Probe Data originates from smartphone or vehicle devices without relying on any stationary sensors, it can be collected on any road segments where vehicles travel and thus has potential to provide traffic volume information in a vast spatial-temporal road network with significantly reduced cost and labor intensity.

Recent research has shown some successful applications of using Probe Data on traffic volume estimation on unmeasured road segments. Among the limited studies, Chang and Cheon (2019) proposed a methodology of AADT estimation using GPS probe vehicle data with two components: first, a k-surveyed site selection method was performed to group the permanent counters and surrounding unmeasured sites with provided GPS data by assuming that the occupancy rate of the vehicle GPS device had a local pattern; then the relationship between the true AADT and GPS counts was determined through an expansion process based on a locally weighted power curve method. In (Sekuła et al, 2017), applications of leveraging GPS probe vehicle data, permanent counter data, and Neural Network models on the estimation of historical hourly traffic volumes were investigated in the Maryland highway network. In this method, the Neural Network model was trained to learn the relation between the ground-truth traffic volume and probe vehicle sample count, along with contextual features, including probe vehicle speeds, weather, infrastructure, temporal information, and volume profiles. Through testing, the proposed method yielded 24% more accurate estimation than the commonly-used volume profiling method (Schrank et al., 2015), which converts AADT estimates from the Highway Performance Monitoring System into hourly volume estimates based on historical speed profiles or other factors. Later in (Hou et al, 2018), some tree-based ensemble learning models, including Random Forest, Gradient Boosting, and Extreme Gradient Boosting, were applied to learn the relationship between the ground-truth traffic volume and GPS sample data and resulted in a promising estimation accuracy with the median absolute percent error of 16%.

However, there are many challenges waiting to be resolved such as (1) evaluating the effect of different quantity and quality of Probe Data on the estimation accuracy, (2) characterizing model error across the dimension of road volume, with emphasis on performance on smaller volume roads and (3) exploring the applicability of leveraging Probe Data on the estimation of various types of traffic volume information beyond AADT in a national spatial-temporal network.

In this report, a comprehensive study will be conducted to evaluate the technical and statistical validity of Probe Data on the traffic volume estimation. More specifically, the evaluation of Probe Data includes the following aspects:

1) Evaluate the estimation of AADT and compare the precision and accuracy against the AADT data derived from (a) the traditional permanent counter approach and (b) factoring up methods based on 48-hr counts;
2) Evaluate the effect of impoverished Probe Data on the traffic volume estimation accuracy;
3) Evaluate the applicability of future AADT estimation using Probe Data;
4) Evaluate the applicability of Probe Data for estimating other traffic volume information, including:
    a) Factors such as day of week (DOW) and month of year (MOY) - Chapter 8;
    b) Directional AADT – volume only - Chapter 3;
    c) Ramp AADT - Chapter 3;
    d) Work Zone ADTs - Chapter 6;
    e) K Factors (Design Hour Volumes) and or D Factors (Percentage of traffic moving in the peak travel direction) - Chapter 8;
    f) Peak hour truck data - Chapter 8
    g) Special events or unusual weather conditions ADT - Chapter 6;
    h) Methods to deploy the new AADT data - Chapter 4;
    i) AADT by vehicle type – 3 classification groups (FHWA classes:1-3, 4-7, and 8-13) - Chapter 8;

Chapter 2: Data

In this chapter, data sources that were used for traffic volume estimation throughout this report are introduced. The data sources include hourly volume counts collected from permanent traffic monitoring stations, a type of Probe data – Location Based Services (LBS) and global positioning system (GPS) (combined cars and trucks) trips, and contextual data; such as weather, road infrastructure, and socio-economic factors. In addition, the method for Probe Data outlier detection and diagnostics is presented. Finally, the predicted features derived from the data sources for traffic volume estimation are introduced.

Permanent Traffic Counter Data

In this report, permanent traffic monitoring stations in the US were selected as the study sites to provide ground-truth traffic volume information. The stations were located across 48 states excluding Alaska and Hawaii. The data was sourced from three providers: the FHWA Traffic Monitoring Analysis System (TMAS), MS2 cloud-based traffic data management system, and state DOTs, where the data type is the hourly traffic volume on each travel direction. Collaboration with the FHWA TMAS substantially increased the available pool of unique locations and road types across the United States available for training and evaluating a Probe Data model, compared to prior AADT models by StreetLight. FHWA TMAS additionally provided the functional class and area type information. A subset of the FHWA TAMS stations also contained hourly counts by vehicle class. A summary of the data sources is shown in Table A1.

Table A1. Summary of the utilized permanent traffic counter data sources

Data provider	Station type	Year	States
FHWA TMAS	Permanent	2011 - 2019	Alabama, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, Florida, Georgia, Idaho, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Montana, Nebraska, Nevada, New Jersey, New Mexico, New York, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Utah, Vermont, Virginia, Washington, West Virginia, Wisconsin, Wyoming
FHWA TMAS	Permanent with Vehicle Class	2018 - 2019	Alabama, Arizona, Arkansas, Colorado, Connecticut, Florida, Georgia, Idaho, Illinois, Indiana, Iowa, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Montana, Nevada, New Mexico, New York, Ohio, Oklahoma, Pennsylvania, Rhode Island, South Carolina, South Dakota, Vermont, Virginia, Washington, West Virginia, Wisconsin, Wyoming
MS2	Permanent	2018 - 2019	New Hampshire, North Carolina, Texas
State DOT	Permanent	2018 - 2019	Virginia

The quality of a fit model is tightly tied to the quality of the input training data. A subset of permanent counters were selected for generation and testing of 2019 AADT values according to the following strict criteria below:

1. At least 80% of hourly volume records through the entire year were observed;
2. In a single day, if over one-third of the hourly volume records have zero counts or over 80% of the hourly volume records have counts below 5, all the records for that day should be discarded;
3. At least 10 of 12 Monthly Average Daily Traffic (MADT) values were captured, and the missing MADTs can be filled in by the corresponding MADTs in 2018;
4. If the percentage change of MADTs between two consecutive months is over 50% or the percentage difference of MADTs between two travel directions is over 100%, the questionable stations were marked and manually diagnosed to ensure that the data is qualified for the AADT calculation.

As a note, the selection criteria described above are not standard methods, and don’t follow the guidelines set out in the Traffic Monitoring Guide (*FHWA, 2016*). For the permanent counters satisfying the above criteria, the AADT values were calculated by the FHWA AADT method in *Traffic Monitoring Guide*:

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

$$AADT = \frac{\sum_{m=1}^{12} \quad d_m * MADT_{HP_m}}{\sum_{m=1}^{12} \quad d_m},$$

where

VOL_{ihjm} = total traffic volume for i^{th} occurrence of the h^{th} hour of day with 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, ...,n_{hjm}) for which traffic volume is available;

h = hour of the day (h = 1,...,24);

j = day of the week (j = 1,...,7);

m = month (m = 1,...,12);

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

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

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

From the counter selection criteria, 1,922 of 6,813 permanent counter stations were removed. In other words, 28% of permanent counters used for AADT didn't capture a significant portion of hours in the year (15+%), which we deemed insufficient for machine learning calibration purposes. AADT calculations used in the *TMG* are robust to giving meaningful AADT values despite occasional data gaps in counters. However, few counters report all hours of the day for all hours of the year, and for this reason this data is not referred to as "ground truth." The quality issues inherent even in permanent counters are important to keep in mind when evaluating the practical benefits and shortcomings of adopting Probe Data techniques, and in future validation work.

A portion of the valid stations were withheld from this report team by the FHWA evaluation team to validate the accuracy and precision of traffic volume estimation methods, and the data from the withheld stations was not used in any model development and self-validation. Specifically, the procedures for withholding stations are:

1. The evaluation team (TTI) has withheld 50% of stations from Maine, Maryland, New Jersey, North Dakota, and Oregon;
2. North Carolina and Texas directly sent lists of stations to be withheld and those were eliminated from the counter database;
3. Among the remaining states, the withholding procedures are:
    a. Ranked by the AADT values, the highest and lowest t 1% of the entire stations are retained, respectively, as non-removable stations;
    b. For the other 98% of stations, 15% of them are withheld from each state by the stratified random sampling method, where the strata are the functional class. For those sites without functional class provided, 15% of stations are withheld by the simple random sampling method.

Among the remaining 4,891 permanent counter stations, 636 of them were withheld for the evaluation team (NREL), leaving 4,255 permanent count stations left as the final study stations for traffic volume estimation throughout this report. The distribution of the stations is shown in Figure A1. By combining the site locations with the road infrastructure and census features, which will be introduced later in this section, the characteristics of the stations are described as follows:

- 3,395 stations have two travel directions; 860 have one travel direction;
- 2,246 stations are in urban or metropolitan transportation commission (MPO) areas and 2,009 are in rural or non-MPO areas;
- 2,077 are on freeway segments and 2,178 are on arterial or local roads;
- 129 stations are on ramps;
- 628 are on road segments shared with bus, rail or subway routes;
- The summary of sites by FHWA functional class is shown in Table A2;
- The summary of permanent and class stations by states is shown in Table A3.

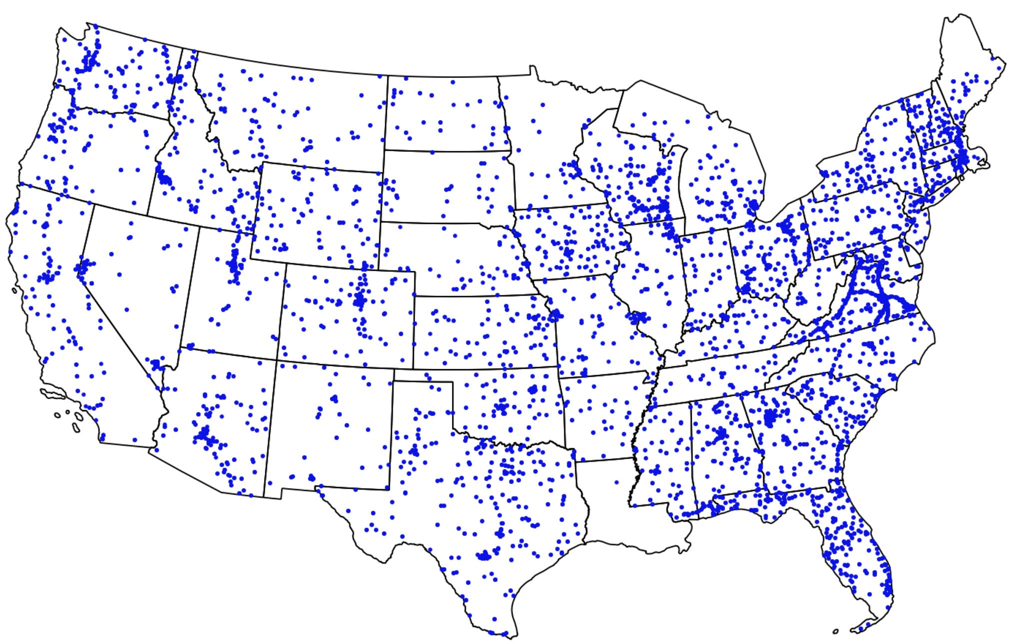


Figure A1. Map of locations of the 4,255 selected permanent counter stations obtained from the FHWA TMAS.

Table A2. Summary of stations *by FHWA functional class*

Functional Class	Number of stations	Number of unidirectional counters
1	1,510	2,408
2	324	599
3	1,419	2,718
4	636	1,235
5	310	591
6	26	47
7	14	25

Table A3. Summary of permanent (and class) traffic stations by state

State	Number of unidirectional counters	State	Number of unidirectional counters
Alabama	279 (16)	Nebraska	72
Arizona	203 (182)	Nevada	130 (26)
Arkansas	54 (48)	New Hampshire	86
California	192	New Jersey	29
Colorado	173 (68)	New Mexico	82 (24)
Connecticut	30 (6)	New York	225 (106)
Delaware	14	North Carolina	136
Florida	417 (398)	North Dakota	53
Georgia	343 (291)	Ohio	259 (152)
Idaho	304 (52)	Oklahoma	148 (148)
Illinois	122 (86)	Oregon	145
Indiana	77 (68)	Pennsylvania	159 (94)
Iowa	198 (90)	Rhode Island	52 (17)
Kansas	158	South Carolina	200 (187)

State	Number of unidirectional counters	State	Number of unidirectional counters
Kentucky	90	South Dakota	66 (29)
Louisiana	1	Tennessee	54
Maine	49	Texas	487
Maryland	58	Utah	155
Massachusetts	81 (41)	Vermont	78 (6)
Michigan	180 (72)	Virginia	699 (309)
Minnesota	55 (53)	Washington	238 (217)
Mississippi	132 (128)	West Virginia	52 (20)
Missouri	126 (46)	Wisconsin	315 (90)
Montana	148 (109)	Wyoming	208 (157)

Probe Data - LBS and GPS Trips

First, this section will describe why access to two different Probe Data sources is uniquely beneficial for AADT estimation. Then, it will describe how StreetLight Data used two data sources to derive sample trips, which are used as the input to drive models of traffic volume, and other traffic factors.

An important note - the work for this report was performed in 2019 and 2020. Probe Data is a fast-evolving field and data options may evolve in subsequent years. Thus, it is more important to understand the useful characteristics of the data sources, and not simply the name/classification of the sources used at the time of this writing.

LBS Data

LBS data can be processed into personal travel patterns at a comprehensive scale. Its high spatial precision and regular ping rate allow for capturing trips as well as activity patterns (i.e., home and work locations), trip purpose, and demographics. This makes it an ideal alternative to data derived from cellular towers, which also has a large sample size but unfortunately lacks spatial precision and pings infrequently (challenges with infrequent pings are discussed in Chapter 4).

In most circumstances, LBS data suppliers provide pieces of software (called SDKs) to developers of mobile smartphone apps to facilitate LBS. These smartphone apps include couponing, dating, weather, tourism, productivity, locating nearby services (i.e., finding the closest restaurants, banks, or gas stations), and many more apps, all of which utilize their users' location in the physical world as part of their value. The apps used in this work required user opt-in, and collect de-identified user locations when they are operating in the foreground. In addition, these apps may collect anonymous user locations when operating in the background. This “background” data collection occurs when the device is moving. LBS data may be collected using WiFi proximity, a-GPS and several other technologies, which don’t require cell coverage. Additionally, most data that StreetLight uses have better than 25-meter spatial precision. StreetLight only licenses data from suppliers where device owners have “opted-in” to location tracking.

GPS Data from Commercial Trucks

Navigation-GPS data has a smaller sample size than LBS data, but it allows users the ability to differentiate commercial truck trips from personal vehicle trips. This makes navigation-GPS data ideal for commercial travel pattern analyses.

Navigation-GPS data suppliers provide data that comes from commercial fleet navigation systems, navigation-GPS devices in personal vehicles, and turn-by-turn navigation smartphone apps. Segmented analytics for medium-duty and heavy-duty commercial trucks are available. For commercial trucks, if the vehicle’s on-board fleet management system is within the data supplier’s partner system, the supplier will collect a ping every one to three minutes whenever the vehicle is on, even if the driver is not actively using navigation.

For personal vehicles, if the vehicle is the data supplier’s partner system and has a navigation console, the supplier will collect a “ping” whenever the vehicle is on, even if the driver is not actively using the navigation system. This provides a very complete picture of vehicles’ travel patterns and certainty that the trips are in vehicles.

Creation of Sample Trips

The following section contains an overview of the fundamental methodology that StreetLight Data uses to create trips from Probe Data.

Step 1 – ETL (Extract Transform and Load)

First, data is pulled in bulk batches from suppliers' secure cloud environments. This can occur daily, weekly, or monthly, depending on the supplier. StreetLight does not access, use, or acquire personal information from suppliers. StreetLight requires all suppliers to provide data that has been sufficiently de-identified prior to ingestion by StreetLight. StreetLight continues to protect supplier data throughout the product creation and distribution lifecycle in accordance with the FTC 3-Part Test for De-Identification. The ETL process not only pulls the data from one environment securely to another, but also eliminates corrupted or spurious points, reorganizes data, and indexes it for faster retrieval and more efficient storage.

Step 2 – Data Cleaning and Quality Assurance

After the ETL process, several automated, rigorous quality assurance tests are run to establish key parameters of the data. To give a few examples, tests were conducted to:

- Verify that the volume of data has not changed unexpectedly,
- Ensure the data is properly geolocated,
- Confirm the data shares similar patterns to the previous batch of data from that particular supplier.

In addition, a visual and manual review of key statistics is performed for each data set. If anomalies or flaws are found, the data are reviewed by StreetLight in detail. Any concerns are escalated to suppliers for further discussion. If the anomaly is deemed explainable, the data is incorporated. If not, the error is corrected in conjunction with suppliers.

Step 3 – Create Trips and Activities

For any type of data supply, the next step is to group the data into key patterns. For example, for navigation-GPS data, a series of data points whose first time-stamp is early in the morning, travels at reasonable speeds for a number of minutes, and then stands still for several minutes, could be grouped into a probable “trip.” For LBS data, a similar approach is followed. However, since LBS data continues to ping while the device is at the destination, there are often clusters of pings in close proximity at the beginnings and ends of trips that are eliminated.

Step 4 – Contextualize

Next, other “contextual” data sets are integrated in order to add richness and improve accuracy of the mobile data. These include road networks and information like speed limits and directionality, land use data, parcel data, and census data, and more.

For example, a “trip” from a navigation-GPS or LBS device is a series of connected dots. If the traveler turns a corner but the device is only pinging every ten seconds, then that intersection might be “missed” when all the device’s pings are connected to form a complete trip. StreetLight utilizes road network information including speed limits and directionality, and performs map matching, to “lock” the trip to the road network. This “locking” process ensures that the complete route of the vehicle is represented, even though discrepancies in ping frequency may occur.

Step 5 – More Quality Assurance

After patterns and context are established, additional automatic quality assurance tests are conducted to flag patterns that appear suspicious or unusual. For example, if a trip appears to start at 50 miles per hour in the middle of a four-lane highway, that start is flagged as “bad.” Flagged trips and activities are not deleted from databases altogether, but they are filtered out from queries and metrics.

Step 6 – Normalize

Next, the data is normalized along several different parameters to account for monthly variation on the underlying sample size.

For LBS devices, population-level normalization is performed for each month of data. For each census block, StreetLight measures the number of devices in that sample that appear to live there, and makes a ratio to the total population that are reported to live there according to U.S. Census Data (and related sources, such as the American Community Survey). A device from a census block that has 1,000 residents and 200 StreetLight devices will be scaled differently everywhere in comparison to a device from a census block that has 1,000 residents and 500 StreetLight devices. Thus, the LBS data is normalized to adjust for any population sampling bias. It is not yet “expanded” to estimate the actual flow of travel.

Step 7 – Store Clean Data in Secure Data Repository

After being made into patterns, checked for quality assurance, normalized, and contextualized, the data is stored in a proprietary format. This enables extremely efficient responses to queries. By the time the data reaches this step, it takes up less than 5% of the initial space of the data before ETL. However, no information has been lost, and contextual richness has been added.

Step 8 – Aggregate in Response to Queries

Zones are created, corresponding to desired analysis locations. In this case, zones will represent each individual permanent counter site across the U.S. Queries then pull trips that intersect each counter zone in the specified direction of travel across a defined time period. Results always describe aggregate behavior, never the behavior of individuals.

Outlier Detection and Diagnostics for Probe Data

Some problematic stations were identified due to unusual or missing road network information in OSM (Open Streetmap 2020), leading to incorrect “locking” of trips. Such stations could become outliers when modelling traffic volume estimation, impact the estimation accuracy, and thus were removed in advance. To detect the potential outliers extreme values or changes in trip penetration, which typically signal an issue with creation of Probe Data trips, are evaluated. The following criteria were applied at each permanent station:

1. The monthly average daily trip sample count should be greater than or equal to 2;
2. The monthly penetration rates (sample count divided by permanent counter traffic volume) should fall between 0.1% and 15%;
3. The percentage change of penetration rates between two consecutive months should be lower than 80%;
4. The percentage difference of penetration rates between two opposite travel directions should be lower than 60%.

For those potential outliers detected through the above criteria, a second-round of manual diagnosis was performed by inspecting the raw LBS and GPS trip trajectories on the road map network and removing the confirmed outliers from the dataset.

Predictive Features for Estimating Traffic Volume

Sample counts derived from Probe Data serve as key features for traffic volume estimation. Additional annual contextual features, such as road infrastructure, local weather, and socio-economic information, also play important roles to enhance the predictive relationship between Probe Data sample counts and traffic volume. In the following, the breadth of predictive features considered primarily for the AADT model are listed, as well as other types of traffic volume estimation.

Sample counts derived from LBS trip repository:

- Annual total LBS trip counts;
- Standard deviation, minimum, maximum, and percentage change of monthly LBS trip counts;
- LBS trip counts on weekdays and weekends;
- LBS trip counts in peak and off-peak periods;
- Scaled LBS trip counts (LBS trip counts weighted by local census population factors).

Sample counts derived from GPS trip repository:

- Annual total GPS trip counts;
- GPS trip counts of personal vehicles;
- GPS trip counts of commercial (light, medium, and heavy-duty) vehicles;
- Standard deviation, minimum, maximum, and percentage change of monthly GPS trip counts;
- GPS trip counts on weekdays and weekends;
- GPS trip counts in peak and off-peak periods.

Road infrastructure characteristics:

- Speed limit, number of lanes, highway types, and tags of ramps and freeway-to-freeway connectors, derived from the OpenStreetMap (OSM) road network;
- Multi-mode features such as whether a road segment is shared with bus, railway, or subway routes, derived from OSM;
- Walkability features such as the density index of the road network, gathered from the Environmental Protection Agency's Smart Location Database.

Census and land-use factors:

- Socio-economic factors, including the block group population, housing units, employment, household and per capita incomes, derived from the United States Census Bureau's database;
- Urban/rural and Metropolitan Planning Organization (MPO)/non-MPO indicators;
- US region indicators (South-Gulf, North-Central, West, South-Atlantic, North-East).

Weather features:

- Temperature, precipitation, snow depth, wind speed, visibility, rain and snow rates. These features are extracted in forms of annual average, standard deviation, minimum, maximum, and seasonal average, and were derived from the NOAA National Climatic Data Center.

Among the large pool of candidate predictive features, a final set of features was chosen to fit a predictive traffic volume estimation model based on a variety of feature selection methods. First, the candidate features with low variance (lower than 0.05 or 0.10) were removed from the pool. Three feature selection methods to comprehensively rank feature importance and select final features for each model:

- Univariate feature selection: the method measures the predictiveness of each feature to the traffic volume based on the strength of correlation and quantification of mutual information;
- Feature importance: for ensemble tree-based models, there are inherent mechanisms to compute and rank impurity-based feature importance when the algorithm builds a decision tree.
- Recursive feature elimination: this method selects features by recursively considering smaller and smaller sets of features. First, the estimator is trained on the initial set of features and the importance of each feature is obtained through either of the above two methods.

The least important features were pruned from the current set of features. That procedure is recursively repeated on the pruned set until the desired number of features to select was eventually reached. Ultimately, the type of feature selection method to be used is both dependent on the model being created, and also the modeler's own judgement of what they find produces the best results for their dataset.

Chapter 3: Estimating AADT Using Probe Data Inputs and Machine Learning

Estimation of annual average daily traffic (AADT) is an essential benchmark of road volume, and serves many purposes for transportation applications. Probe Data presents a novel mechanism by which to create estimation of AADT for every road segment across a country. This section explores the model error of a National AADT model fit with Probe Data inputs, using reference permanency counter 2019 AADTs from across the United States. Model performance is evaluated across road characteristics, with a special emphasis on performance for small roads, where traffic volume via traditional methods (permanent and temporary counts) are sometimes absent.

Methods for Estimating AADT with Probe Data

Candidate Machine Learning Models

Throughout this project, the performance of a variety of machine learning methods were evaluated for predicting 2019 AADT, as well as the other related items in this report. 2019 was chosen as the most recent complete year during the time of the study. The quality of Probe Data samples is constantly changing. As Probe Data sample quality and quantity improve over time, the most recent year will be more indicative of future results than past years. A high-level flow diagram of using Probe Data and machine learning models to estimate AADT is shown in Figure 1. First, the machine learning model is trained to learn the relationship between the AADT derived from permanent counts and Probe Data, along with contextual features influencing the AADT. Next the hyperparameters of the model are tuned through cross-validation to enhance the model performance and avoid overfitting. Finally, the model is applied to new stations with the input Probe Data and contextual features and produces the estimated AADT.

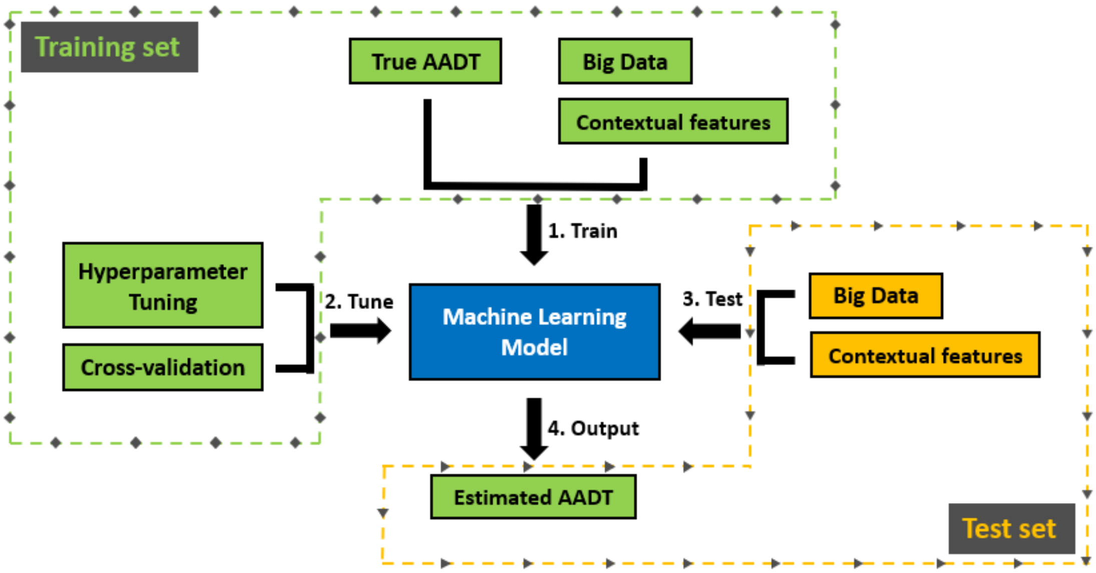


Figure A1. Flow diagram of AADT estimation using Probe Data and machine learning.

Among the candidate machine learning models, Linear Regression was considered first. There is a strong signal between LBS trips and AADT (see later discussion). This suggests that with a basic Excel model, one could create a model with a sample of trips and calibration counters. However, a goal of the current project is to optimize the National AADT model to obtain the best performance possible. To do this, additional features beyond the total count of LBS sample trips are needed to improve model performance. Thus, multivariate regression and regularized linear regression were also explored as candidate models.

In contrast to multivariate linear regression, tree based models are often used when there is no obvious linear relationship between the features and the label. Although there is a linear relationship between StreetLight trips and AADT, model fit can be improved if predictors that have a non-linear relationship with the data are included. All tree based models are based on decision trees. A decision tree comprises a set of nodes and leaves. Each node represents a feature split and each leaf represents a predicted value (for regression) or a class label (for classification). A decision tree is built by recursively splitting each feature in a way that minimizes the cost function. The cost function is the sum of squared residuals for a regression problem, and the Gini index for a classification problem.

Decision trees are widely used in machine learning because they are simple to understand, interpret and visualize. However, they are unstable and prone to overfitting. Hence, other machine learning methods that are based on decision trees are often used as alternatives. Two of those methods used in this report are Random Forest and Gradient Boosting.

Random Forest works by creating hundreds of bootstrapped samples from the training data, building a decision tree from each of these bootstrapped samples, then averaging their outputs (for regression) or tallying the votes (for classification) for the final decision.

These hundreds of trees are what make Random Forests much more effective and less prone to overfitting than individual decision trees. Just like Random Forest, Gradient Boosting works by constructing multiple decision trees from bootstrapped samples. However, unlike Random Forest which builds decision trees simultaneously, Gradient Boosting builds one decision tree at a time and improves on the performance from the previous tree. In some applications, Gradient Boosting can out-perform Random Forest but with some caveats: gradient boosting takes more time to run, and it may overfit the training data if the predictive features are not well generalized.

Model Validation Metrics

In order to report on error, the stratified k(10) fold cross validation was used to generate test sets for metrics. The individual counter stations were randomly divided into ten groups, stratified by AADT grouping. For each fold, a given model was trained on a 9 of the folds, and the tenth fold was used to generate error metrics for the model. Cross validation was used because it allows for every counter to be tested, so that the model performance can be evaluated across as many unique types of roads and regions throughout the United States as possible. The method of assigning subsets of the data to training and test folds, as well as the ratio between the training and validation data impact how representative the metrics from cross validation are. The current method stratified by AADT grouping to ensure that each fold was trained and tested on locations representing all AADT volume groupings, as model error is tightly tied to road volume. Folds were randomly assigned by counter in order to maximize the spatial distribution observed by the training folds, as penetration rates are influenced by local geography. Overfitting of metrics may occur if training and validation sites are in close proximity (such as across a divided highway). As detailed in Chapter 2, a separate hold out validation set of permanent counters across the United States was reserved for testing the final trained National Probe Data AADT Model independent of this study, and results will be discussed in a future report.

A range of metrics to describe error in the model estimations are used in this study. To describing the general relationship between the Probe Data AADT estimate and the permanent counter value, the Pearson correlation is used. When the measure approaches 1, it means a stronger relationship with lower variance. The fitted line of the bivariate relation between permanent counter and estimated AADT through cross validation is also checked to measure the bias of the model. When the fitted line is close to the 45-degree line, it means no bias.

For all reports, permanent counter stations are grouped by AADT, and the results are given within those groupings (labeled road size bin). Error increases with smaller roads, and groupings by road volume allow for more visibility into where the errors of the model lie. For summary metrics to describe the general error across an AADT grouping, the mean absolute percent error (MAPE, see equation below), and normalized root mean squared error (NRMSE, see equation below) are given. These metrics are helpful to give an idea of what an expected error is for most stations. Both metrics are included because they are both commonly used in literature, and highlight different aspects of model performance. MAPE describes errors well on small roads, and in contrast, NRMSE penalizes large errors, making it more sensitive to the accuracy of AADT estimation on high-volume roads. For bar charts to visualize MAPE, the bootstrapped confidence intervals of the standard error of the mean were visualized with the python library seaborn in order to allow statistical comparison between groups.

$$MAPE = \left(\frac{1}{n}\right) * \sum_{i=1}^{n} \quad 100 * \left|\frac{AADT_{Estimate(i)} - AADT_{Permanent\ counter(i)}}{AADT_{Permanent\ counter(i)}}\right|$$

$$NRMSE = 100 * \frac{\sqrt{\left(\frac{1}{n}\right) * \left(AADT_{Permanent\ counter(i)} - AADT_{Estimate(i)}\right)^2}}{\left(\frac{1}{n} * \sum_{i=1}^{n} \quad \left(AADT_{Permanent\ counter(i)}\right)\right)}$$

Although the summary of errors is useful for getting a sense of model performance, describing the percentile spread of error across the test set is also be illustrative of where error lies. The 68th percentile absolute error is given, as it represents one standard deviation of the mean within a standard bell curve, and thus is a useful descriptor of the typical error across the road segments. As MAPE and NRMSE may be sensitive to outliers, the 68th percentile absolute error can provide more visibility into expected 'typical' error. The 95th percentile absolute error is provided to measure the spread of errors across a broader array of sites. Finally, the median percentage error is given to give an indication of bias in the model. Values close to 0 suggest that the model has low bias, while positive values would indicate overestimation, and negative values indicate systematic underestimation. For AADT model metrics with sufficient sample size, the 95th percentile error was also calculated, which is labeled '95% TCE Error Range (%)'. For this metric, percent error and the log of AADT was fit to a quantile regression. From this fit, the larger absolute value between the 2.5th and 9.75th error range is reported.

When evaluating model accuracy, it is useful to consider how well Probe Data AADT estimation compares to current estimation approaches. One widely used approach is to collect a two-day temporary count from a link, and expand that count to estimate AADT using calibration of nearby permanent counters ("short term count expansion"). Research performed by Battelle for FHWA[1] (Krile et al, 2016) included exhaustive study of expected error from short term count expansion. Table A4 shows the computed errors form short term expansion in their study. These errors to serve as a point of comparison for Probe Data AADT model error in this paper (Table A4). It should be noted that Table A4 represents a conservative case of current AADT factorization methods, which assumes short term counts are available on a link for the year AADT is being estimated, and that reference counts from permanent counters from the same roadway classification type and year are available for the region. This is not the case for some links across the United States. For example, as shown in Chapter 1, Table 1, less than 5% of locations <500 AADT and less than 15% of locations 500-2000 AADT have such same year temporary counts available in Minnesota.

Table A4: Bias and absolute percentage error from short term counter expansion in estimating AADT across ten volume-based categories of road size for a same-year 48-hour count. This assumes short term counts are available on a link for the year AADT is being estimated, and that reference counts from permanent counters from the same roadway classification type and year are available for the region. This is not the case for some links across the United States, especially for small roads (see introduction). Data from Krile et al (2015).

Road Size Bin	Median Error (Bias) %	95% TCE Error Range (%)	68th Percentile Absolute Error (%)	95th Percentile Absolute Error (%)	NRMSE	MAPE
B 500 - 1,999	-0.01	34.2	11.7	26.9	12.9	10.0
C: 2000 - 4,999	2.2	30.8	10.8	33.6	17.2	10.4
D: 5,000 - 9,999	3.3	28.5	9.68	28.1	14.0	9.2
E: 10,000 - 19,999	1.4	26.7	9.2	27.9	12.9	8.9
F: 20,000 - 34,999	0.9	25.7	8.31	24.3	13.3	8.1
G: 35,000 - 54,999	0.4	24.8	8.35	19.3	9.8	7.2
H: 55,000 - 84,999	-0.3	24.1	6.07	14.5	7.2	5.3
I: 85,000 - 124,999	0	23.5	4.83	14.7	6.8	4.6
J: > 125,000+	3.0	23.3	7.06	17.7	10.0	6.2

Note – the median error bias floats up and down in a way that reflects sample distribution. When expressing targets compared to two-day counts we have simplified the median bias targets. For discussion of this see Task 3 Report.

Results

Correlation of Probe Data Trip Sample Count With AADT

It’s important to also evaluate the effectiveness of the raw signal (LBS or GPS) compared to a machine learning model. Trips derived from LBS and GPS (personal and commercial) data alone may be strong indicators of traffic volume. However different data providers and paths for data procurement lead to significant variation in data characteristics, either spatial variation or temporal variation that could impact the quality of a raw data signal. Thus, simply naming a data source “LBS” or “GPS” is less meaningful than understanding it’s characteristics. The LBS and GPS data used in this study has a typical ping rate of every 1-3 minutes, and the spatial accuracy of the LBS data typically ranged between 5-20 meters. There is a strong linear relationship between LBS (Figure A2a) or GPS counts (Figure A2b) in the sources described in this report, and 2019 AADT values at permanent counters. Total counts of LBS trips achieve a Pearson correlation score of 0.96, while GPS trips have a score of 0.89 (p-value < 0.001).

Figure A2a: Correlation between LBS trips and 2019 AADTs from permanent counters

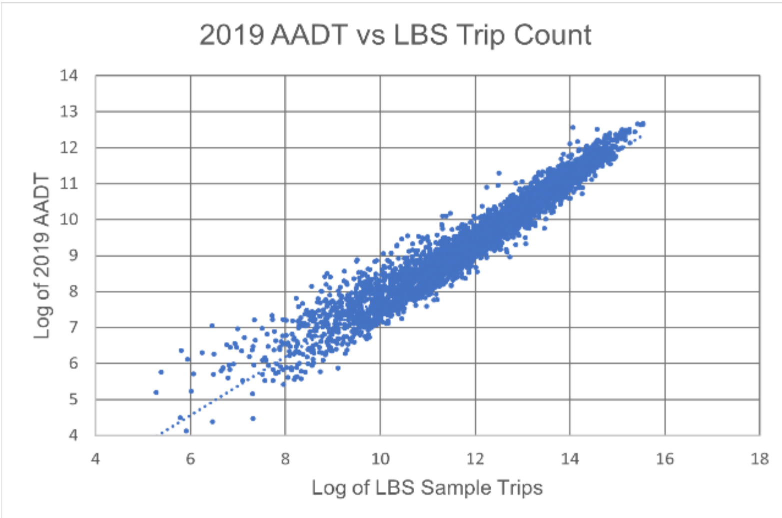
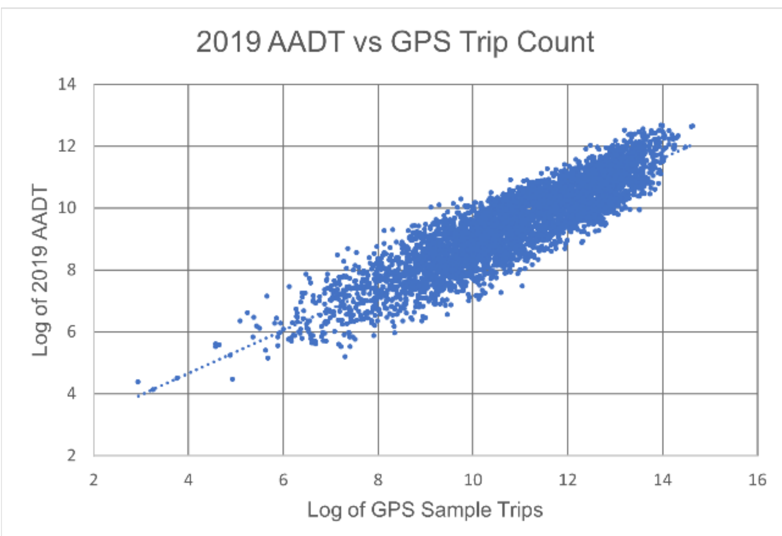


Figure A2b. Correlation between GPS trips (commercial and personal) and 2019 AADTs from permanent counters



However, it's clear from the correlation plots that the GPS (commercial and personal) data has much more variation than LBS data, which is clustered more consistently across the 45-degree line. A histogram of sample penetration comparing the total Probe Data sample trips at each counter (Figure A3). The GPS penetration is more skewed toward lower penetration and has a greater spread of penetration from counters. This emphasizes that a simple 1:1 relationship between GPS trip count and counter AADT would not work for all locations. The following table (Table A5) expands upon this, further detailing error and performance when the LBS and GPS data alone are compared to the output of a National Probe Data AADT Model fit to a Gradient Boosting model, which will be described in more detail later. It becomes clear that LBS data alone performs similarly to a fit machine learning model, but GPS data alone shows more significant variation and error.

Figure A3. Histogram depicting the variance in sample trip penetration observed between two different Probe Data sources for sample trips: LBS (blue) vs GPS (commercial and personal, striped). This emphasizes that a simple expansion relationship between to estimate AADT from GPS trip count and counter AADT would not work for all locations.

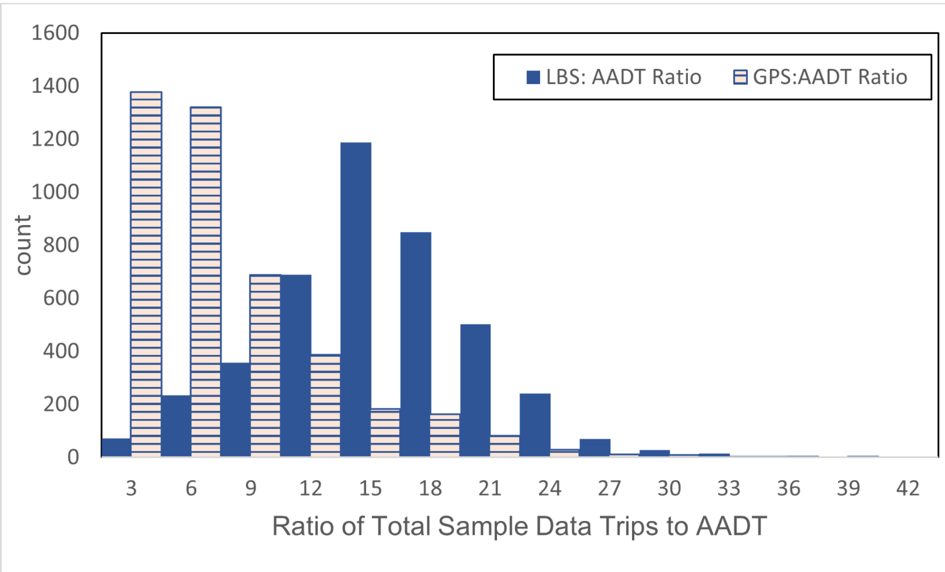
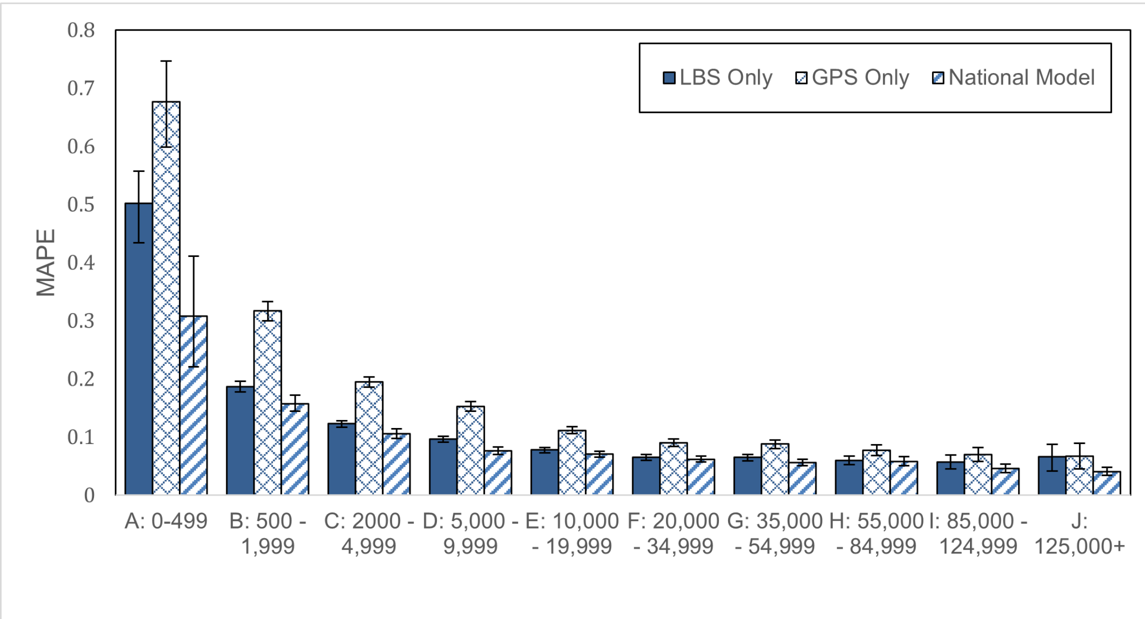


Table A5: Errors compared across Probe Data AADT models trained with either LBS source data only, GPS (commercial and personal) source data only, or both LBS and GPS data sources (National Probe Data AADT Model).

Group	**Median Bias (%)**	**95% TCE Error Range (%)**	**68th Abs. Percentile (%)**	**95th Abs. Percentile (%)**	**MAPE**	**NRMSE**
LBS Only	1.06	49.54	11.84	37.81	11.94	15.68
GPS Only	1.39	84.08	18.10	59.86	18.24	19.75
National Probe Data Model (GPS + LBS)	0.5	25.45	8.2	24.96	8.11	13.36

When these results are further broken out by road size, the differences between the raw data sources and the National Probe Data AADT become more evident (Figure A4). LBS data alone is closely aligned with the error in the National Probe Data AADT Model until road sizes decrease, specifically when they fall below 500 AADT. GPS data alone performs pretty well on very high-volume roads but quickly deteriorates in quality relative to the other methods once road volume decrease. To improve predictability on small roads, the GPS data source could be mixed with other factors. The GPS data source has a much smaller sample size relative to the LBS data utilized in this analysis (Figure A3), which may contribute to higher error on small roads. Additionally, because GPS data is tied to newer automobiles, it is possible that the traces may be more biased toward specific demographics, and thus the total GPS trips need to be fit to other contextual factors to improve predictability. In summary, for smaller roads, model performance benefits from blending both GPS and LBS data source. The trade-off between gains in accuracy with financial cost of acquiring both GPS and LBS data are among the various factors to consider in adapting a model framework.

Figure A4: MAPE across roads of different sizes for LBS only, GPS (personal and commercial) only and the National Probe Data AADT Model. Error bars represent the standard error of the mean of absolute percentage errors.



Model Comparison for Estimating AADT with Probe Data

Although the sample trips derived from LBS and GPS (personal and commercial) data alone have a strong predictive relationship with AADT, the relationship can be greatly improved by incorporating other predictive features in a model for 2019 AADT. There are lots of choices for models that can take a collection of predictors to estimate a continuous predictor, like AADT. See methods for detail on additional predictive features considered. The model algorithms that were tested included Multivariate Linear Regression (LR), Elastic Net Regression (EN), Support Vector Regression (SVR), Random Forest (RF), Gradient Boosting (GB) and Extreme Gradient Boosting (XGB). Over aggregate methods, the tree-based models (RF, GB and XGB) outperformed other regression techniques (LR, EN, SVR) (Table A6).

Table A6: Comparison of Model Errors of 2019 AADT Across Six Different Algorithms. LR (multivariate linear regression), EN (elastic net regression), SVR (support vector regression), RF (random forest), GB (gradient boosting), XGB (extreme gradient boosting), and the Final Probe Data model, which was created from two separate XGB models.

Group	Median Bias (%)	95% TCE Error Range (%)	68th Abs. Percentile (%)	95th Abs. Percentile (%)	MAPE	NRMSE
LR	0.23	131.07	27.80	129.18	36.32	24.05
EN	3.20	104.76	20.40	66.02	23.44	25.06
SVR	-0.02	51.69	12.08	42.09	12.75	18.59
RF	1.31	43.18	11.35	34.16	11.04	19.50
GB	0.67	42.45	10.76	34.99	10.87	15.34
XGB	0.94	30.26	10.91	33.87	10.98	15.68
Final Probe Data Model	0.5	25.45	8.28	24.96	8.11	13.36

Model error was broken down across road sizes (Figure A5). Large roads (above 20,000 AADT) have similar MAPE. However, as seen in Table A6, the distribution of absolute error is substantially greater with Linear Regression (compare the 95th percentile absolute error with Gradient Boosting: 129% compared to 35%). The higher error for linear regression becomes even more substantial as average road volume decreases.

Figure A5: Comparison of Mean Absolute Percent Error for 2019 AADT estimation between different machine learning algorithms, across varying Road Volumes . The real difference in performance comes with the lower volume roads, where the Linear Regression and Elastic Net algorithms have substantially higher error than other techniques. Error bars represent the standard error of the mean.

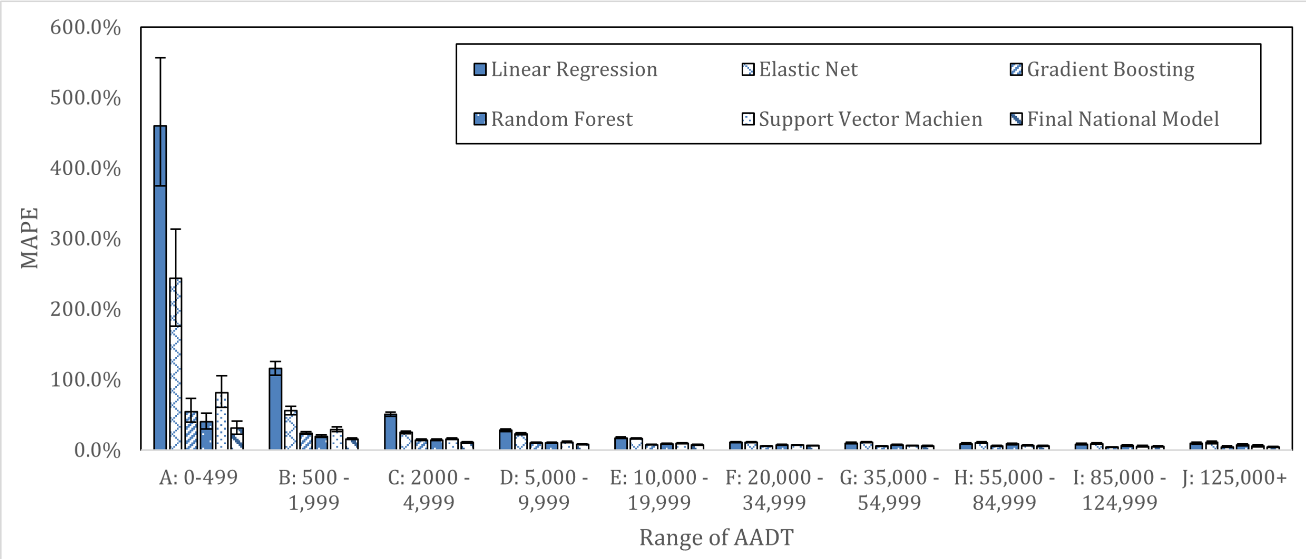


Table A7: Comparison of LBS Probe Data penetration (sample Probe Data trips / permanent counter trip count) across AADT range.

AADT Range	Mean LBS Penetration	Standard Deviation
A: <= 499	0.02	0.01
B: 500 - 1,999	0.03	0.01
C: 2000 - 4,999	0.03	0.02
D-J: 5,000 +	0.04	0.01

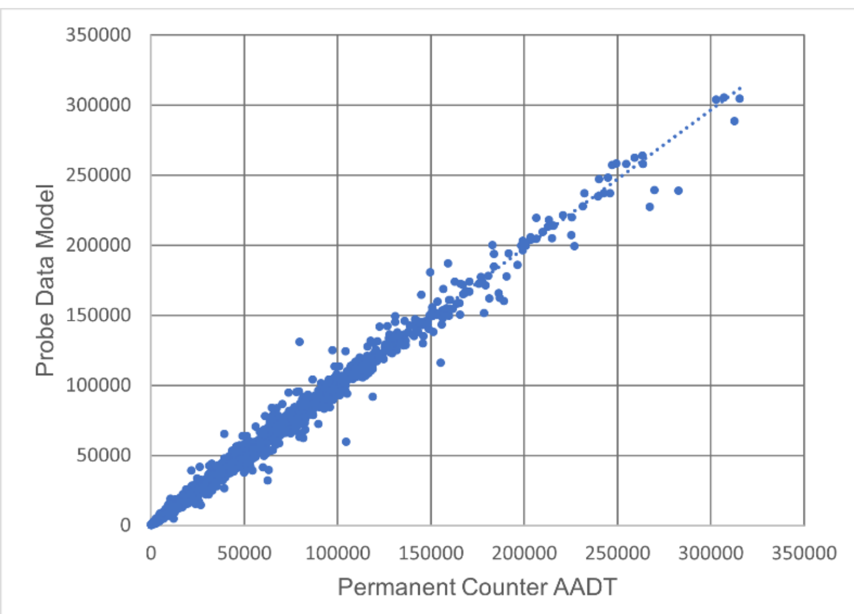
Penetration rate of Probe Data tends to increase across road size (Table A7), but the spread of penetration rate can also fluctuate. Linear regression-derived models (including Elastic Net) may not be well suited to handle the variability of sample penetration across road volume size. Support Vector Regression (SVR) performs well in general, but does not do as well on lower or higher volume roads. As one goal was to model small roads as well as possible with Probe Data, results suggest that tree-based methods are the best option to meet that goal. All three tree-based models are quite similar in their performance, as seen in the plots of MAPE (Figure A5). If you compare the distribution of errors (Table A6), Gradient Boosting and Extreme Gradient boosting have less spread of errors, and lower NRMSE as compared to the Random Forest algorithm. The Gradient Boosting algorithm adds additional complexity to the Random Forest algorithm by fitting errors as the model is built, which can further boost model performance. Extreme Gradient Boosting has faster running time than traditional Gradient Boosting algorithms, but model error was similar, so Extreme Gradient Boosting model was selected as the final model for 2019 AADT estimation in this report.

In order to further optimize model performance for small roads, results from two different Extreme Gradient Boosting models was selected as the final model ('Final Probe Data Model' in Table A6). For this approach, the full data set of calibration sites, and contextual features, were fit to an Extreme Gradient Boosting algorithm. A second Extreme Gradient Boosting model was trained only sites with lower volume (AADT under 10,000). For testing, sites were assigned to the second 'low volume' model based on their total LBS count. This forced the model to minimize errors for small roads, and allowing for tuning of features specifically for lower volume roads. This two-part model will be discussed in more detail, and referred to as the 'National Probe Data AADT Model'. For later chapters (Chapters 4 and beyond), a single Extreme Gradient Boosting model fit to all sites was used for subsequent analysis and comparisons because it adds less complexity to analysis.

National Probe Data AADT Model for Estimating AADT From Probe Data

The national AADT model achieves a Pearson correlation of 0.99 (p-value <0.001), suggesting a very strong relationship between the Probe Data model and permanent counter measured 2019 AADT. As you can see from a correlation plot (Figure A6), most counter values lie close to the line of best fit, with few outliers. Simple visual comparison with Figure A2a demonstrates how both blending GPS (personal and commercial) and LBS trip input, and adding in contextual features drastically improved model error. It should not be assumed that features that are of top importance in the National Probe Data AADT Model for 2019 will always be the most important features. For example, in years with more snow storms or hurricanes, weather features may be more important for local accuracy than other years.

Figure A6: Correlation plot between the actual AADTs and the National Probe Data AADT Model's estimated AADTs



National Probe Data AADT Model Error by Road Volume

AADT estimation was evaluated for both directional AADT (Table A9a, Figure A7), and total AADT for both directions on roads (Table A9b) across three road groupings, classified by average traffic volume. Bi-directional AADT were created by adding up directional estimates, resulting in a bi-directional AADTs for each road. In general, error increases as AADT decreases. Krile et al. (2015) also noted a trend of higher error on smaller roads for short term count expansion, and noted this may be due to higher variability in daily traffic for smaller roads. For all methods, small roads present an additional complication, as the lower traffic means fewer trips in a sample, which naturally leads to higher estimation error.

When modeling 2019 AADT, we found the model performed best when we modeled at the individual permanent counter level for each unique direction of traffic, especially for lower and mid-volume roads. One benefit of this approach may be that using unidirectional counters for modeling nearly doubles the training sample size and adds more smaller AADT values. This may allow the model to learn the relationship between the true AADT and probe data better on low-volume roads due to more examples. To generate a total road AADT for bidirectional roads, we summed the estimated AADT for each direction. This approach had lower error than creating a separate model for bidirectional AADT. When comparing error by AADT volume, directional AADTs for small roads (500 - 4,999) have slightly less error than bidirectional AADTs (Table 9b) of the same average traffic volume. Compare, for example, an MAPE for 10.42 for directional AADTs vs an MAPE of 12.38 for total AADT for "low" volume roads. In contrast, the MAPEs for medium and high roads are very similar between directional and total AADT estimates.

Table A9a: National Probe Data AADT Model Error for Directional AADTs Across 3 Groupings by Road Volume.

Road Size Bin	Median Bias (%)	95% TCE Error Range (%)	68th Abs. Percentile (%)	95th Abs. Percentile (%)	MAPE	NRMSE
500-4,999 (low)	1.72	39.52	11.56	29.31	10.42	13.63
5,000-54,999 (medium)	0.1	24.14	7.88	20.15	7.11	11.44
55,000+ (high)	-1.18	16.67	6.49	19.36	6.13	9.76

Table A9b: National Probe Data AADT Model Error for Road Bidirectional AADTs Across 3 Groupings by Road Volume

Road Size Bin	Median Bias (%)	95% TCE Error Range (%)	68th Abs. Percentile (%)	95th Abs. Percentile (%)	MAPE	NRMSE
500-4,999 (low)	3.48	40.4	13.53	33.36	12.38	15.92
5,000-54,999 (medium)	0.2	24.06	7.33	19.47	6.75	10.24
55,000+ (high)	-0.31	16.42	5.04	16.28	5.22	8.78

Krile et al. (2015) simulated model errors from same-year 48-hour temporary count (SY-TC) expansion. We calculated the errors from their study and have presented them in table A9c, to allow comparison within the presented 3 groupings by road volume AADT. There are a few differences between the two studies. The current report examined 2019 AADT alone, while the Krile et al (2015) study looked at estimated AADTs from 2000 to 2012. The current study involved one 2019 AADT estimate for each of 4,255 stations, while the Krile et al (2015) study utilized 206 total stations, with multiple comparisons within each station (hundreds of pairs of 48 hour count to yearly AADTs for each station).

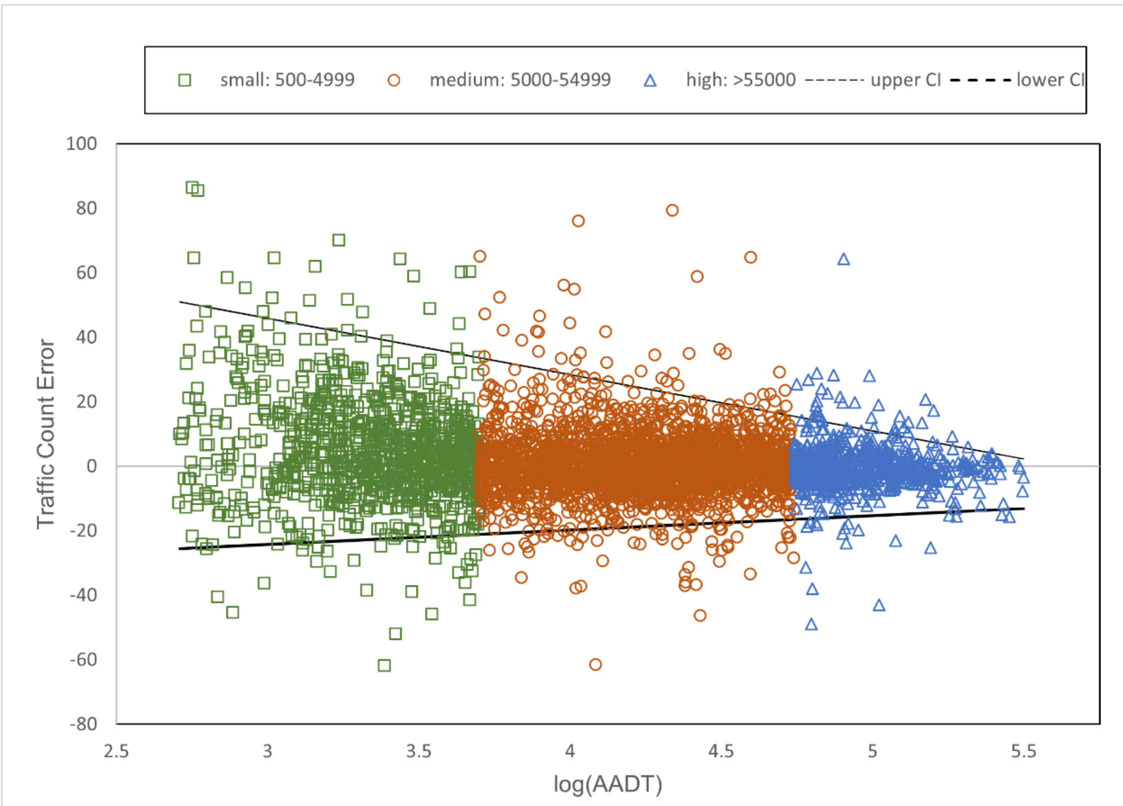
The TMG recommends a different temporary count cycle (such as a 3-year cycle) for various roads. Not all states are able to meet these thresholds, and even when they are, it means that many counts are out of date or derived from “similar” roads (as opposed to actual counts). The tables below only compare results to temporary counts taken in the same year. Comparison to counts from temporary counts from old years, or counts derived from models or extrapolation by road type should necessarily be less accurate than same year temporary counts.

Table A9c - Comparison of National Probe Data AADT Model for 2019 AADT compared to 48-hour same year temporary count expansion (SY-TC) as represented by Krile et al (2015). Places where Probe Data exceeds same year two-day count performance are in green with heavy outlines. Places where Probe Data performs within 2 percent of SY-TC are highlighted green and bold. Places where Probe Data performs within 5 percent of SY-TC are in italicized yellow. Sample size (n) refers to the number of sites evaluated.

	Method (n)	**Median Bias (%)**	**95% TCE Error Range (%)**	**68th Abs. Percentile (%)**	**95th Abs. Percentile (%)**	**MAPE**	**NRMSE**
500-4,999 (low)	SY-TC (77)	-0.2	34.0	11.2	30.0	10.3	18.0
500-4,999 (low)	Probe Data, Bidi (1,003)	*3.4*	40.4	*13.5*	*33.4*	**12.4**	**15.9**
500-4,999 (low)	Probe Data, Directional (2,746)	1.7	39.52	**11.6**	**29.3**	**10.4**	**13.6**
5,000-54,999 (medium)	*SY-TC* (103)	1.1	28	9.11	26.36	8.7	14.3
5,000-54,999 (medium)	Probe Data, Bidi (2,524)	**0.2**	**24.0**	**7.3**	**19.4**	**6.8**	**10.2**
5,000-54,999 (medium)	Probe Data, Directional (4,074)	**0.1**	**24.1**	**7.9**	**20.1**	**7.1**	**11.4**
55,000+ (high)	*SY-TC* (25)	1.5	24.0	5.9	15.6	5.3	9.5
55,000+ (high)	Probe Data, Bidi (655)	**-0.3**	**16.4**	**5.0**	**16.3**	**5.2**	**8.8**
55,000+ (high)	Probe Data, Directional (517)	**-1.2**	**16.7**	**6.5**	*19.36*	**6.1**	**9.8**

For medium and large roads, the National Probe Data AADT Model performs competitively with 48-hour same-year short term counts across most accuracy metrics. Error is compared between short term counters and the National Probe Data AADT Model with finer resolution across road sides in the following section, which provided more visibility and better comparison between errors across road volume (Table A11).

Figure A7: Quantile regression plot of the National AADT model error for directional AADTs across three groupings of road volume by 2019 AADT.



Within the ‘small’ road category (AADTs under 5000), there is a more dramatic trend of increasing error with smaller roads. To better describe model error for small roads, roads were further segmented by AADT across 10 road size categories for both directional (Table A10a) and bidirectional AADTs (Table A10b). In both versions of the metrics, errors are fairly steady across different categories of road size, until roads become smaller than AADTs of 5,000. At that point, there is increasing error with each increment of smaller roads (bins A, B, C).

Table A10a: National AADT model errors for directional AADTs across 10 groupings of road volume

Road Size Bin	Total Counters	Median Bias (%)	95% TCE Error Range (%)	68th Abs. Percentile (%)	95th Abs. Percentile (%)	MAPE	NRMSE
A: 0-499	275	14.21	51.47	27.37	66.93	25.98	32.01
B: 500-1,999	1240	3.66	41.93	15.33	34.4	13.03	16.96
C: 2,000-4,999	1504	0.48	34.73	9.76	26.92	8.72	12.35
D: 5,000-9,999	1258	0.42	28.74	9.13	22	8.06	11.05
E: 10,000-19,999	1363	0.26	24	7.85	21.36	7.25	10.32
F: 20,000-34,999	937	-0.13	19.7	6.62	16.79	5.94	9.27
G: 35,000-54,999	516	0.46	16.62	6.95	20.35	6.69	9.92
H: 55,000-84,999	381	-1.22	15.77	6.21	19.06	5.95	8.98
I: 85,000-124,999	105	-0.57	14.89	6.05	15.04	5.19	7
J: 125,000+	33	-1.18	14.28	6.67	19.85	6.52	10.51

Table A10b: National AADT model errors for bi-directional stations across 10 groupings of road volume

Road Size Bin	Stations	Median Bias (%)	95% TCE Error Range (%)	68th Abs. Percentile (%)	95th Abs. Percentile (%)	MAPE	NRMSE
A: 0-499	51	21.3	54.06	32.97	84.22	30.82	42.58
B: 500-1,999	387	5.87	43.81	18.97	41.61	15.76	19.97
C: 2,000-4,999	616	2.56	36.78	12.2	29.19	10.59	14.64
D: 5,000-9,999	608	0.33	30.81	8.19	24.12	7.68	11.56
E: 10,000-19,999	763	0.22	25.55	7.83	19.23	7.09	9.53
F: 20,000-34,999	716	0.05	21.14	6.66	18.2	6.19	9.31
G: 35,000-54,999	437	0.36	17.19	6.26	15.97	5.64	8.24
H: 55,000-84,999	330	-0.05	16.05	5.69	18.76	5.8	9.35
I: 85,000-124,999	182	-0.08	15.25	4.97	12.05	4.59	6.88
J: 125,000+	143	-0.88	14.38	4.15	13.77	4.1	6.44

The strengths and weaknesses of Probe Data to model AADT may be highlighted by comparing model errors with those from a 48-hour same year short-term count expansion (Table 11a below). These errors were generated from a study by Krile et al., (2015). It is noted that these model errors do not address additional expected error from performance of data collection equipment. For road size bin A (sites with AADTs under 500), there was only one site for comparison with short term count methods, which is not statistically sufficient for error comparisons.

The TMG recommends a different temporary count cycle (such as a 3-year cycle) for various roads. Not all states are able to meet these thresholds, and even when they are, it means that many counts are out of date or derived from "similar" roads (as opposed to actual counts). The tables below only compare results to temporary counts taken in the same year. Comparison to counts from temporary counts from old years, or counts derived from models or extrapolation by road type should necessarily be less accurate than same year temporary counts.

Table A11a - Comparison of National Probe Data AADT Model for 2019 AADT errors for bidirectional and directional roads compared to temporary same year count expansion (SY-TC, as represented by Krile et al., 2015). Places where Probe Data performs within 2 percent of SY-TC are highlighted green and bold. Places where Probe Data performs within 5 percent of SY-TC are in italics and yellow. Places where Probe Data exceeds SY-TC performance are in green with heavy outlines. Values have been reduced to one place after the decimal for legibility. Sample size refers to the number of sites evaluated.

Road Size Bin	Source	Sample Size (n)	Median Bias (%)	95% TCE Error Range (%)	68th Abs. Pcntile (%)	95th Abs. Pcntile (%)	MAPE	NRMSE
A: 0-499	SY-TC*	1	-	-	-	-	-	-
A: 0-499	Probe Data,bidi	51	**21.3**	**54.1**	**33.0**	**84.2**	**30.8**	**42.5**
A: 0-499	Probe Data, directional	275	**14.2**	**51.5**	**27.3**	**66.9**	**26.0**	**32.0**
B: 500-1,999	SY-TC	33	-0.1	34.2	11.7	26.9	10.1	13.0
B: 500-1,999	Probe Data, bidi	387	5.9	43.8	18.9	41.6	15.8	19.9
B: 500-1,999	Probe Data, directional	1240	**3.7**	42.0	*15.3*	34.4	*13.0*	*16.9*
C: 2,000-4,999	SY-TC	44	2.3	30.8	10.8	33.6	10.5	17.3
C: 2,000-4,999	Probe Data, bidi	616	**2.6**	36.8	**12.2**	**29.2**	**10.6**	**14.6**
C: 2,000-4,999	Probe Data,directional	1504	**0.5**	*34.7*	**9.8**	**26.9**	**8.7**	**12.4**
D: 5,000-9,999	SY-TC	47	3.2	28.5	9.7	28.1	9.3	14

Road Size Bin	Source	Sample Size (n)	Median Bias (%)	95% TCE Error Range (%)	68th Abs. Pcntile (%)	95th Abs. Pcntile (%)	MAPE	NRMSE
D: 5,000-9,999	Probe Data,bidi	608	**0.3**	*30.8*	**8.2**	**24.1**	**7.6**	**11.6**
D: 5,000-9,999	Probe Data,directional	1258	**0.4**	**28.7**	**9.1**	**22.0**	**8.0**	**11.1**
E: 10,000-19,999	SY-TC	23	**1.3**	**26.7**	**9.2**	**27.9**	**9.0**	**12.9**
E: 10,000-19,999	Probe Data, bidi	763	**0.2**	**25.5**	**7.8**	**19.2**	**7.0**	**9.5**
E: 10,000-19,999	Probe Data,directional	1363	**0.3**	**24**	**7.8**	**21.3**	**7.2**	**10.3**
F: 20,000-34,999	SY-TC	20	0.9	25.7	8.3	24.3	8.2	13.3
F: 20,000-34,999	Probe Data,bidi	716	**0.1**	**21.1**	**6.6**	**18.2**	**6.1**	**9.3**
F: 20,000-34,999	Probe Data,directional	937	**-0.1**	**19.7**	**6.6**	**16.7**	**5.9**	**9.2**
G: 35,000-54,999	SY-TC	13	0.5	24.8	8.4	19.3	7.3	9.8
G: 35,000-54,999	Probe Data,bidi	437	**0.3**	**17.1**	**6.2**	**15.9**	**5.6**	**8.2**
G: 35,000-54,999	Probe Data,directional	516	**0.4**	**16.6**	**6.9**	**20.3**	**6.6**	**9.9**
H: 55,000-84,999	SY-TC	10	2.2	24.1	6.1	14.4	5.3	7.3
H: 55,000-84,999	Probe Data, bidi	330	**-0.3**	**16.1**	**5.7**	*18.8*	**5.8**	*9.4*
H: 55,000-84,999	Probe Data,directional	381	**-1.2**	**15.8**	**6.2**	*19.1*	**5.9**	*8.9*
I: 85,000-124,999	SY-TC	8	0	23.5	4.8	14.7	4.7	6.8
I: 85,000-124,999	Probe Data, bidi	182	**-0.1**	**15.3**	**4.9**	**12.0**	**4.6**	**6.9**
I: 85,000-124,999	Probe Data,directional	105	**-0.6**	**14.9**	*6.0*	**15.0**	**5.2**	**7**
J: 125,000+	SY-TC	7	3.1	23.3	7.1	17.7	6.2	10
J: 125,000+	Probe Data, bidi	143	**-0.9**	**14.4**	**4.1**	**13.8**	**4.1**	**6.4**
J: 125,000+	Probe Data,directional	33	**-1.2**	**14.2**	**6.6**	*19.8*	**6.5**	**10.5**

*Krile et al only had one station in this bin, which is not sufficient for statistical purposes. Thus, no "TC" or temporary count comparison is known or available.

The Probe Data National AADT Model is in line with same-year 48-hour expansion methods for roads above 2,000 AADT. Although error is higher for smaller roads, for AADTs between 500 and 2,000, the MAPE remains at or under 15% for this category. As an example, if the true AADT of a road was 1000 cars, the model prediction would on average predict between 850 and 1150.

In addition, there is variability in how well the National Probe Data AADT model performs state by state, notably for smaller volume roads. Table S4 showcases the variation in how error for small roads (AADT < 5000) from the Probe Data model compares to expected error from a typical same year count. State to state variability in error is likely due to a wide range of variables, including availability of reference permanent counts within each state and nearby states of similar road type and size, the available data for the Probe Data model, the weather effects in travel to name just a few.

Table S4. Variability in National Probe Data Model MAPE for small roads (AADT < 5000) by state. Performance is grouped by how MAPE for small road sites within the state compare to a typical same year count (Typ-SC), with expected MAPE of ~ 18% (see Chapter 3 Table A9c for reference). n = total unique permanent sites within the state with AADT < 5000. MAPE was used due to the small number of small road sites within most states.

Category (MAPE)	State (n)
Exceeds MAPE for Typ-SC (<18%)	Alabama(18), Arizona(27), Arkansas(27), California(13), Colorado(23), Connecticut(1), Florida (33), Georgia(22), Illinois(10), Indiana (11), Iowa(47), Kansas(40), Kentucky(16), Louisiana(1), Maine(26), Maryland(3), Massachusetts(3), Michigan(16), Minnesota(6), Mississippi (16), Missouri(37), Nebraska(16), New Hampshire (5), New Jersey(1), New Mexico(18), Nevada(15), New York(27), North Carolina(12), North Dakota(16), Ohio(16), Oklahoma (20), Oregon(23), Pennsylvania(18), South Dakota(16), South Carolina(24), Texas(64), Utah(16), Vermont(17), Virginia(70), Washington(26), Wisconsin(26), West Virginia (7), Wyoming(66)
Meets Typ-SC	Delaware(1), Idaho(44), Montana(45)
Worse than Typ-SC (MAPE > 18%)	Tennessee(11), Rhode Island(1)

Probe Data model bias is very comparable to 48-hour same year short term expansion, except for bidirectional estimates on small roads (AADT < 2000), where the National AADT Model tends to overestimate. We believe this overestimation is due to the distribution of road volumes across permanent counters in the training sample and can be further improved by local calibration efforts to bring in more data from smaller roads. This is bolstered by the fact that in some states, the small road estimation accuracy exceeds that of same-year, appropriately factored temporary-count expansion. AADT values cannot go below zero, and agencies place permanent counters at locations where they expect measurable traffic. These two factors create a skew toward higher AADTs within lower volume bins (bins A & B above), which creates slight overestimation in the fit models.

In addition to evaluating the aggregate error (MAPE, NRMSE), one can evaluate the spread of error in the sample with the 68th and 95th absolute percentile errors. As compared to 48-hour same-year factoring methods, the probe data model has comparable (within 2 percentage points) or less error for road groups with bidirectional AADTs over 5000 (groups D-J) for the 68th percentile error. When the model error is expanded to the 95th percentile error, the National AADT maintains lower error as compared to short term expansion for roads above 2,000 AADT. The one exception is for category H, where the error shows a small increase (18.76% error, vs 14.44%). There may be a few ‘abnormal’ locations with bias the model was not able to tackle in this category. Additional predictive features to attack these unique locations, or a model with specific regional tuning may further improve fit for these locations. Regardless, the distributed errors, even toward the tail, on large roads maintain within the error ranges of 48-hour same year short term expansion. This suggests a strong application for Probe Data to model AADT. This model generated accurate estimates for locations without use of any short-term counter information for calibration or training, and suggests a great approach for creating useful volume estimates for roads where permanent or short-term counts are more expensive, or otherwise not feasible or likely.

As discussed in Chapter 1, we asked Minnesota DOT to give an idea of how often the short-term same year expansion method in Table A9c is used. For roads under 2,000 AADT, over half of roadways lacked any direct counts at all, and for roads under 55,000 AADT, less than half had a short-term count from the same year (as per TMG current recommended practice, rotating every few years between locations). Thus, to give an idea of how the National Probe Data AADT Model error compares to a more typical estimation (Typ-SC) arising from an older count, we compare errors in Tables A9c and A11a. No comprehensive or complete data source was found to describe the accuracy of typical techniques such as ‘similar’ segments, akin to the Krile paper for same year two-day expansions. Therefore, indicative reports were combined with expert input to create the estimate errors.

It's ultimately up to departments what an acceptable level of error for small roads are, and will depend on the goals for the road. The National AADT Model errors are small enough that the model is unlikely to classify a small road, say 500 AADT as a substantially larger road, such as 2000 AADT. There will likely always be more error for the smallest of road classes, where statistical sampling has an impact on the results. However, there is a lot of potential for improving model performance on these smaller roads. This may include incorporating more local calibration points such as from nearby short-term counts (which was outside the scope of this study), blending additional Probe Data and other Big Data sources (such as signal cameras), additional feature engineering from LBS and GPS trips to derive more predictive signal, or use of ensembling models, which may provide more accurate estimations by averaging the results of multiple models.

Model Error Across Roadway and Regional Characteristics

Of all the states in the analysis, some states have fairly low error across all metrics, especially as compared to errors from traditional short-term expansion methods. Example states include Arizona, Colorado, Connecticut, Delaware, Idaho, Iowa, Kentucky, Maryland, Massachusetts, Michigan, Minnesota, Missouri, New Hampshire, North Dakota, Ohio, Oregon, South Dakota, Texas and Wisconsin. The following three tables show the model performance for three example states:

Maryland (n=58), Ohio (n=259), and Texas (n=487). These states were selected because they contained a large number of permanent counters across a variety of road sizes. More research is needed into what factors create the best setup for a region to have good volume estimation by Probe Data. Evaluation at the predictive features in the National AADT Model (Figure A7b demonstrates that essential characteristics are Probe Data sample trips that are both well spatially distributed across a region and well represented across demographics. The better statistically representative sample of trips you can have from a Probe Data source, the better the sample will generalize to the population. But, in order to compensate for some inevitable bias from the statistical sample, adequate local calibration data, be it accurate adjustments for bias by census data, or calibration with local counter volume data are also essential.

Table A9: Directional AADT model error for the state of Maryland by road size (n=58)

Road Size Bin	Counters (n)	Median Bias (%)	95% TCE Error Range (%)	68th Abs Prcentile (%)	95th Abs. Prcentile (%)	MAPE	NRMSE
500-4,999 (low)	10	1.4	21.75	6.04	10.58	5.28	4.18
5,000-54,999 (medium)	33	1	14.62	7.04	13.45	5.46	8.7
55,000+ (high)	14	-1.21	16.68	3.54	5.05	2.58	3.61

Table A10: Directional AADT model error for the state of Ohio by road size (n=259)

Road Size Bin	Counters (n)	Median Bias (%)	95% TCE Error Range (%)	68th Abs Percentile (%)	95th Abs. Percentile (%)	MAPE	NRMSE
500-4,999 (low)	70	3.91	29.27	11.8	19.84	8.89	11.08
5,000-54,999 (medium)	174	2.33	23.62	6.6	17.49	6.31	8.85
55,000+ (high)	15	2.68	19.13	5.01	14.17	4.74	6.75

Table A11: Directional AADT model for the state of Texas by road size (n=487)

Road Size Bin	Counters (n)	Median Bias (%)	95% TCE Error Range (%)	68th Abs. Percentile (%)	95th Abs. Percentile (%)	MAPE	NRMSE
500-4,999 (low)	211	0.64	33.46	13.78	33.07	12.59	13.54
5,000-54,999 (medium)	239	0.30	17.98	7.95	19.90	7.52	10.12
55,000+ (high)	21	-2.66	11.15	4.91	7.92	4.18	4.58

The National Probe Data AADT model error was evaluated across a variety of road and regional characteristics. First, errors were evaluated across the seven different functional classes (Table A6). Functional class information was available for 3,298 roads through the HPMS. Model error is more tightly correlated with AADT than functional class divisions. Classes 3-7 contain the smallest AADTs. There is a spread of errors across functional classification size, which we feel is due to the variance in AADTs within each class, rather than characteristics of the functional class itself. This is most apparent in the errors for the 95th Absolute Percentile, which is mainly driven by the error of small AADTs within each functional class.

Table A6: National AADT model error metrics comparison by FHWA road classification

Functional Class	AADT range (25th-75th percentile)	Stations	Median (Bias)	95% TCE Error Range (%)	68th Absolute Percentile (%)	95th Absolute Percentile (%)	MAPE	NRMSE
1	[25,039-83,477]	956	-0.07	17.94	5.81	18.26	5.96	10.11
2	[25,756-77,207]	271	1.09	18.35	7.14	18.95	6.38	12.27
3	[4,361-19,605]	1229	0.65	30.5	8.83	22.68	7.85	12.73
4	[2,197-9,025]	538	2.25	36.13	11.56	31.76	11.2	14.42
5	[1,192-4,262]	269	3.03	42.2	15.36	38.95	14.48	19.21
6	[537-2,790]	23	5.72	47.5	13.2	50.06	15.37	11.41
7	[3,162-9,666]	12	2.99	35.95	21.23	50.54	18.78	39.53

With GPS (personal and commercial) and LBS as data sources, availability of pings may be influenced by geography. The ability to collect LBS data is also tied to what apps are running in the background. Some apps may be more popular in specific regions of the US. When LBS or GPS penetration is plotted across the US, regional variation is apparent, as compared to the Regions or Climate Zones

Second, model performance was compared across five different regions in the United States. It was found that the errors across different regions on the whole are comparable with one another (Table A7). There is a trend of lower absolute and squared error among stations in the South Gulf, South Atlantic, and North Central regions as compared to the national error. However, model error is tightly correlated with road volume, and the small changes in model error across regions is likely tightly tied with differences in road volume across the training set. Inclusion of the region, latitude and longitude in the National AADT Model largely allowed the Extreme Gradient Boosting algorithm to tune the input Probe Data sample trips by region as needed.

Table A7: National AADT Model Error metrics comparison by region for bidirectional AADTs

Region	AADT range (25th-75th percentile)	Stations	Median Bias (%)	95% TCE Error Range (%)	68th Absolute Percentil e (%)	95th Absolute Percentil e (%)	MAP E	NRMSE
North Central	[3971-30866]	926	1.03	27.9	7.92	21.19	7.31	13.26
North East	[5163-39062]	449	0.84	26.63	8.99	25.24	7.92	14.93
South Atlantic	[8250-50418]	1202	0.12	22.51	7.48	25.12	7.46	11.48
South Gulf	[502-42970]	629	1.29	26.3	8.36	21.98	8.08	10.31
West	[3378-34494]	1027	0.23	27.99	9.39	29.05	9.68	16.03
All Regions	[5034-36416]	4233	0.5	25.45	8.28	24.96.	8.11	13.36

Model error across different climate zone regions, as defined by the Köppen-Geiger climate classification (Table A8, spatially visualized in Figure A8). Florida was the only state in climate zone A (equatorial), so the superior performance of climate zone A may be equally attributed to the model performing well for the state of Florida itself. Climate zone B has a shift toward lower AADTs across the climate zone, which may also explain the slight trend toward higher error for that climate. The National AADT Model uses counter latitude, longitude, and annual weather in the AADT estimation. This allows the model to account for potential differences based on climate. Alaska and Hawaii were not included in this report to speed up the process but all techniques discussed can be applied to those states as well.

Figure A8. Map of the main climate zones within the contiguous United States, as defined by the Köppen-Geiger climate classification. Figure adapted from Kotekk et al., 2006. Zones are defined as A (equatorial), B (arid), C (warm temperature), D (snow) and E (polar, rare in US).

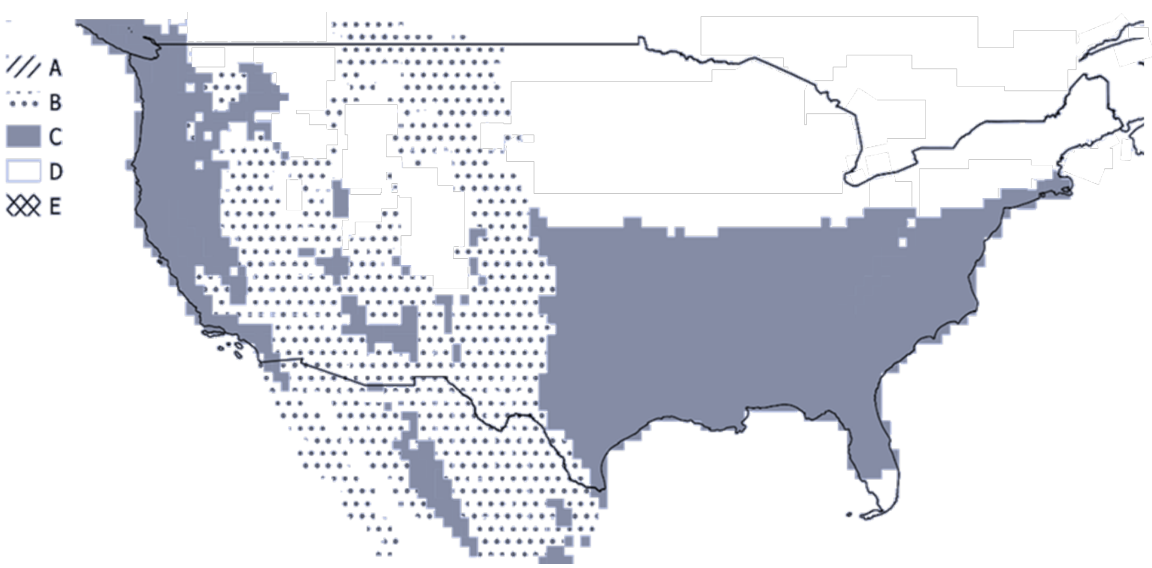


Table A8: National AADT model error metrics comparison by climate zone

Climate Zone	AADT Range [25th- 75th percentile]	Median Bias (%)	95% TCE Error Range (%)	68th Absolute Percentil e (%)	95th Absolute Percentile (%)	MAPE	NRMSE
A (equatorial)	[25,411-112,513]	-0.71	17.44	8.66	13.71	5.98	11.21
B (arid)	[2,545-23,086]	1.49	30.74	9.51	26.9	9.44	16.18
C (warm temperature)	[7,095- 42,576]	0.47	23.53	7.94	24.92	7.89	12.86
D (snow)	[3,681-28,745]	0.25	28.68	8.41	24.45	8.1	12.96

The majority of the calibration stations were along roadways, model error was also compared between ramps and non-ramp roads (Table A8). The total number of ramps in the sample was low (125), and model error metrics are sensitive to the total number of locations observed. To draw appropriate comparisons between model performance on ramps and non-ramp roadways, permanent counter stations were randomly sub sampled, and used as a test set for describing model error. This process was repeated 10 times to get an averaged estimated metric for non-ramp roadways. Ramp roadways perform comparatively to non-roadways across road volumes.

Table A8: Comparison of National AADT model error between ramp and non-ramp roadways.

Road Size Bin	Ramp	Stations	Median (Bias)	68th % Absolute Percentile	95th % Absolute Percentile	MAPE	NRMSE
500 - 4,999 (low)	No	57	8.6	27	66.7	24.5	29
500 - 4,999 (low)	Yes	57	13.66	26.19	60.3	35.21	23.23
5,000-54,999 (medium)	No	65	0.5	13.2	30.5	11.4	16.1
5,000-54,999 (medium)	Yes	65	-1.43	11.35	29.47	22.04	10.85
55000+ (high)	No	3	1.8	10.5	15.3	9.4	10.3
55000+ (high)	Yes	3	-11.12	13.47	16.99	15.65	9.67

Conclusions

The results of the Probe Data National AADT Model are presented as a representation of expected error from Probe Data AADT estimates. The results should be seen as representative of the baseline results that could be achieved, but by no means are the upper limit of the potential of Probe Data as an estimation source. As state-level breakdowns suggest, a state may find that their region can be modeled by Probe Data with an even stronger relationship than depicted in this report. Probe Data matches or outperforms short term expansion counts for small, medium, and large roads. Probe Data estimates can out-perform short term expansion methods for extremely small roads in areas with more local small AADT calibration sites. Future research is needed to investigate how additional tuning of models for unique regions, and how enrichment of the model by short term counters could even further boost model performance.

Chapter 4: Impact of Probe Data Characteristics on the Traffic Volume Estimation

In the previous chapter, the Probe Data National AADT Model has shown successful use of a full-year of Probe Data to estimate AADT for roads across the lower 48 states. However, such results are built upon the provided LBS and GPS data with the best currently available quality and quantity. The impact of different characteristics of LBS and GPS trips on the traffic volume estimation still remains unknown and needs to be explored.

In this chapter, the impact of the impoverished LBS and GPS trips on the accuracy, precision, and stability of the traffic volume estimation and prediction will be discussed. More specifically, the impoverishment procedures include the size of sampling devices, ping rate, spatial accuracy, and the number of available months. In addition, the impact of impoverished permanent counters for calibration on the AADT estimation accuracy will be studied.

Effects of Time Periods Sampled on AADT Model Accuracy

The National Probe Data AADT Model utilizes LBS trips from all 12 months in 2019, however it may be possible to create a model for AADT from a smaller window of time. Probe Data providers may not supply 12 months in a calendar year, or budget constraints might prevent the procurement of a full year's worth of data, which makes it important to understand the necessity of observations across time for estimating AADT. Additionally, Probe Data quality may be variable across time, producing different penetration rates or biases month to month, that may impact the quality of a model Probe Data providers frequently sell data in monthly intervals, and so in this section the impact of availability of Probe Data across months on model accuracy was explored. Finally, localities may want an “early” estimation of roads’ AADT before the year actually ends (discussed in the subsequent section), with appropriate caveats for unusual end of year events disrupting estimates.

Our 12-month National Probe Data AADT model included reference data from 4,232 permanent counter stations from 48 states across the U.S. Data was aggregated across all months in 2019 and compared to 2019 AADT values reported by the permanent counters. See section ‘AADT Estimation’ for details on building, training, and validating the National AADT Model. The same methods and dataset were used for this work. For comparisons in this chapter, the Extreme Gradient Boosted model fit to all calibration sites was used as reference, as opposed to blending results with a second additional model fit to only small roads. This was chosen to minimize complexity of analysis. In order to test the dependency of AADT model accuracy on the number of unique calendar months of Probe Data sample trips observed, subsets of months from 2019 were selected for training and testing a national 2019 AADT model. Subsets of 9, 6, 3 and 1 month were randomly generated across all possible month combinations in 2019. For single month tests, all 12 months were tested individually. As an example, a model for 2019 AADT was trained on January 2019 Probe Data only, and tested for accuracy of predicting 2019 AADT. This was repeated for every month of the year, to describe how predictive AADT models based off of single months across the year could be. For the remaining subsets of 3 ,6 and 9-month selections, 30% of all possible month combinations were selected. For example, there are 220 different combinations of 3-month pairs that can be derived from 12 calendar months. Instead of testing all 220 combinations without replacement, 30%, or 66 out of the 220 combinations were tested. Results were averaged across all possible monthly combinations.

Summary of Results

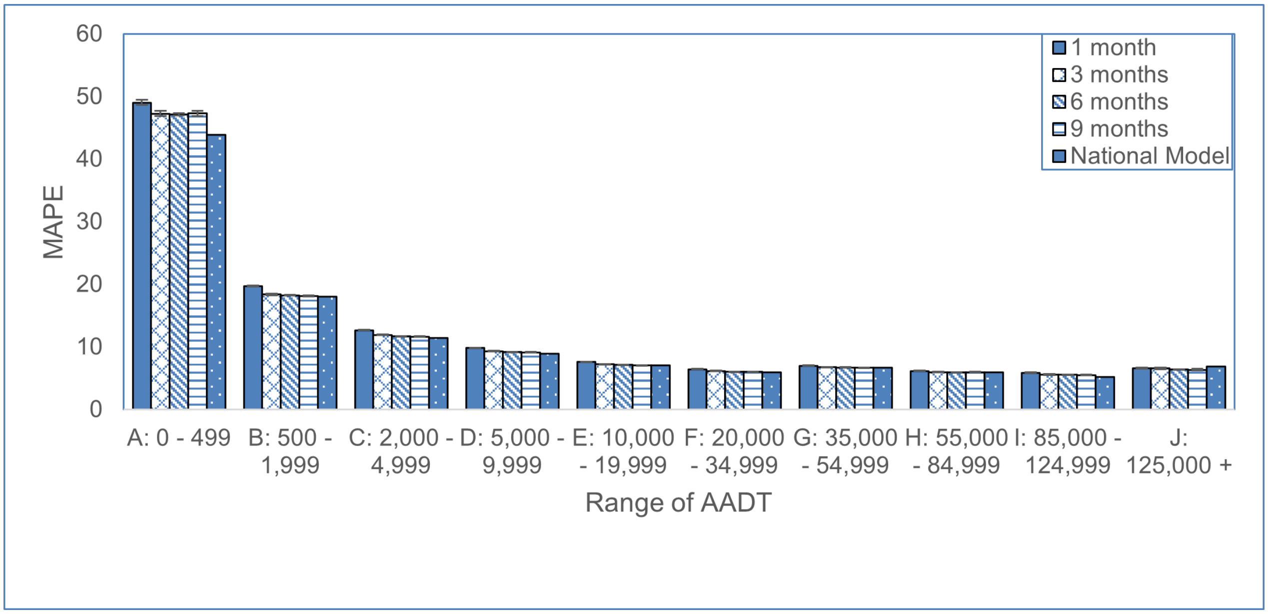
Error metrics for the 2019 AADT models trained on Probe Data inputs from either 9, 6, 3- or 1-month combinations across 2019 were compared to the National AADT Model, which was trained on a full year of 2019 Probe Data (Table T1). All error metrics show a trend of increasing error as models are trained on decreasing windows of time. However, looking at the error in the 95th Absolute percentile, the error from building a model from only 3 to 9 months of data is fairly consistent. This suggests that with just a handful of months, one can reasonably estimate AADT. The difference in model error between 12 months and a single month is also not as dramatic as one might anticipate. These results may be partly because the range of AADT volume within a year at a location (+/- 40%, for example) is smaller than the differences in the range of AADT across all of the locations in the sample (100-100,000). Additionally, the LBS and GPS (personal and commercial) sources used for this study result in a fairly large collection of monthly trips and the sampling rate is consistent across months, allowing for fairly accurate model creation from only a few months of data.

Table T1: Comparison of error between 2019 AADT models trained with different combinations of 2019 months

Group	Median Bias (%)	95% TCE Error Range (%)	68th Abs. Percentile (%)	95th Abs. Percentile (%)	MAPE	NRMSE
1 month	0.91	51.60	12.10	38.84	12.18	16.00
3 months	0.93	48.23	11.46	36.93	11.55	15.63
6 months	0.94	47.34	11.33	36.26	11.42	15.50
9 months	0.95	46.96	11.27	36.01	11.36	15.41
National Probe Data Model	0.94	30.26	10.91	33.87	10.98	15.68

As results are broken down by road volume (Figure T1), the benefit of having a full 12 months' worth of data becomes more evident. The error differences become more significant as the AADT decreases. If only one month of data is available, then the model does not perform as well, particularly for small roads. However, in order to get the most accurate estimates possible, particularly for small roads, a full year of Probe Data from the target AADT year may be much preferable.

Figure T1: Comparison of MAPE for 2019 AADT estimation across models fit with varying total 2019 months of data. Comparisons are shown across different groupings by road volume to account for changes as road volume decreases. Error bars indicate the standard error of the mean across each station within each group.



Conclusions

Overall, the quality of AADT estimates on smaller volume roads will be most impacted by a reduction in data months. It is suggested that at least 3 months of data should be used to accurately estimate AADT values among different sizes of roads. But for areas with lots of small volume roads and where high accuracy targets may be needed, a full year of sample data will significantly improve accuracy for AADT estimation.

When selecting a subset of months for AADT estimation, it’s difficult to know which particular months will be preferable. Depending on the available data source, some months may be better predictors than others, and what makes one month “better” than another may also vary regionally. Considering potential biases due to tourism, student populations, special events and weather, local knowledge may help to determine which calendar months that are most representative. Data vendors may also have fluctuating samples due to the integration of new apps or devices, which can change the stability and quality of the Probe Data source month over month.

Ultimately, the approach to selecting a preferred subset of months should involve consideration of what’s most representative of a given study area. Examining historical volume trends can also be a valuable tool, specifically looking at whether a calendar month or subset of months has also been predictive in prior years.

Further research on this topic could involve sub-selecting days of the week to see which are most effective at predicting AADT. For example, how would a model perform if only Mondays were made available, or weekends, etc. This daily format is not typical of the way vendors aggregate and supply Probe Data, and thus was not explored as part of this investigation, but different permutations of the data could be further analyzed. Or course, using a subset of months will reduce the availability of other metrics such as MADT.

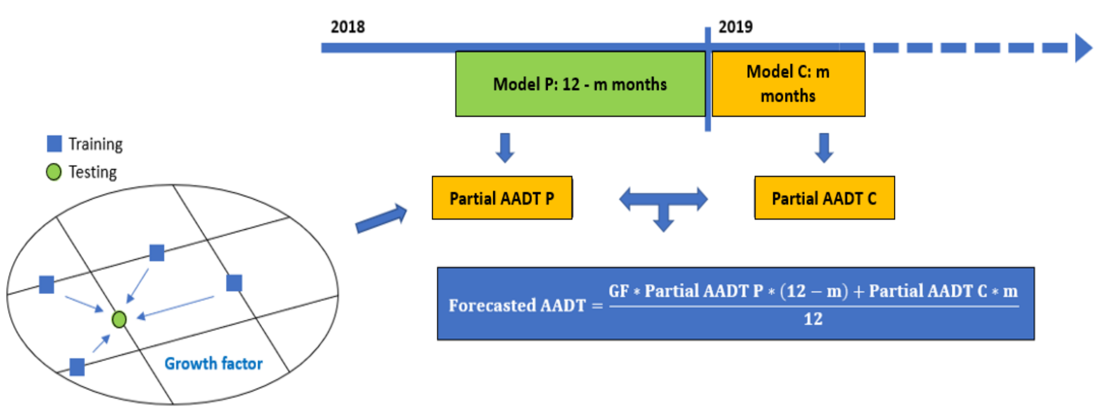
Forecasting AADT Mid-Year

If data months can be reduced to estimate AADT, it begs the question whether a small subset of months aggregated from early in a calendar year will be sufficient in predicting AADT for the remainder of the year. Traditionally, Probe Data has been used to estimate AADT after the full calendar year has elapsed, so that data can be aggregated from all 12 months. If models are able to predict AADT with a smaller subset of months, then it may be possible to estimate AADT at a given location before the year has elapsed.

The goal of this section is to imagine the accuracy of forecasting 2019 AADT with only an early month subset given. To assess the strength of Probe Data to forecast AADT, two approaches are tested. The first is to gather a rolling 12 months of data to predict AADT for the remainder of the year. This will be referred to as the “rolling year” approach. The second is to utilize the given years’ AADT in order to forecast for the remainder of the year. This will be referred to as the “existing year” approach. Each approach is described in further detail below.

For the first, rolling year approach, consider the scenario where one is trying to forecast AADT for 2019. Thus far in 2019 there are *m* sequential months’ worth of data available. In order to aggregate 12 months’ worth of data, the most recent *12 - m* months are considered, such that some months are aggregated from the prior year, i.e. 2018. The AADT forecasting method is designed in Figure F1.

Figure F1: Graphic representation of the “rolling year” forecasting approach for predicting AADT



First, two independent machine learning models are developed to perform the “partial “AADT estimation in 2018 and 2019, respectively.

Model P (Partial AADT P) is trained by the 12 - m months of partial AADT in the previous year and applied to estimate the partial AADT in the previous year (here the previous year refers to 2018).

Model C (Partial AADT C) is trained by the m months of partial AADT in the current year and applied to estimate the partial AADT in the current year (here the current year refers to 2019).

First, the growth factor of each individual permanent counter for the *12 - m* months' period is calculated based on the historical MADT values. Note that the historical MADT values are preferred over the historical AADTs to best capture the growth factor during those months in the previous year. However, if there is no way to get historical MADTs, the historical AADT values can be accepted.

Second, the growth factor of the test zone is calculated by averaging the growth factors of surrounding permanent counters with the same highway type and urban/rural indicator. In terms of the scale of the surrounding area, it could be the state-based level or a buffer with a specific radius. If there is no available permanent counter nearby, \a default state-based or region-based growth factor was used instead. In terms of averaging, either a simple average or weighted average by the distance was used.

The forecasted AADT in the current year is calculated by the weighted average of two partial AADT estimates. The function can be written as follows:

$$\textit{Forecasted AADT} = (GF * \textit{Partial AADT P} * (12 - m) + \textit{Partial AADT C} * m) / 12$$

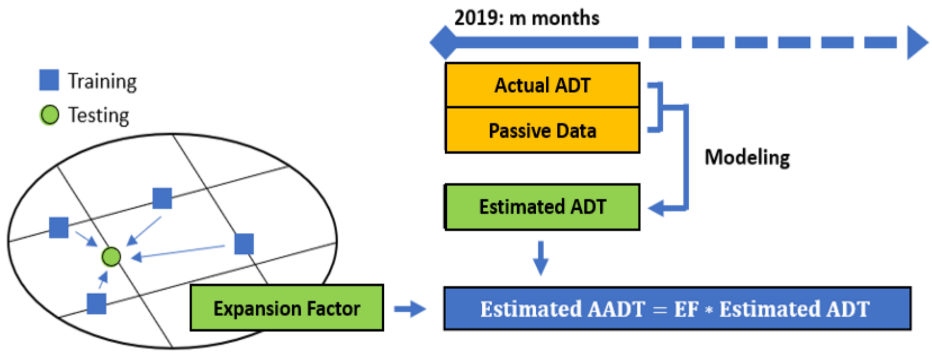
The following scenarios are tested in order to forecast AADT for 2019:

- M = 0, Previous 12 - m = 12: data is not available for 2019
- M = 3, Previous 12 - m = 9: data is available for the first quarter of 2019
- M = 6, Previous 12 - m = 6: data is available for the first two quarters of 2019
- M = 9, Previous 12 - m = 3: data is available for the first three quarters of 2019

The expectation is that as *m* increases, there will be lower errors through cross-validation. However, such expectation is not guaranteed due to the consideration of model errors.

For the second, “existing year” approach, assume a desire to predict AADT in 2019 with only *m* months currently available. Similar to the “rolling year” approach, a model is built to estimate AADT based on *m* months in the current year (2019). In the “rolling year” approach, the growth factor can be considered as a longitudinal factor to inflate the previous year’s ADT estimates to the current year’s. In the “existing year” approach, the expansion factor can be considered as a lateral factor to adjust the current year’s ADT estimates to the full year’s. See Figure F2 for a diagram of model design.

Figure F2: Graphic representation of the existing year forecasting approach



To calculate the expansion factor:

*Estimated AADT = EF * Estimated ADT*

Test cases will explore the following four scenarios:

- M = 1: one month of data is available for 2019
- M = 3: three months of data is available for 2019
- M = 6: six months of data is available for 2019
- M = 9: nine months of data is available for 2019

Summary of Results

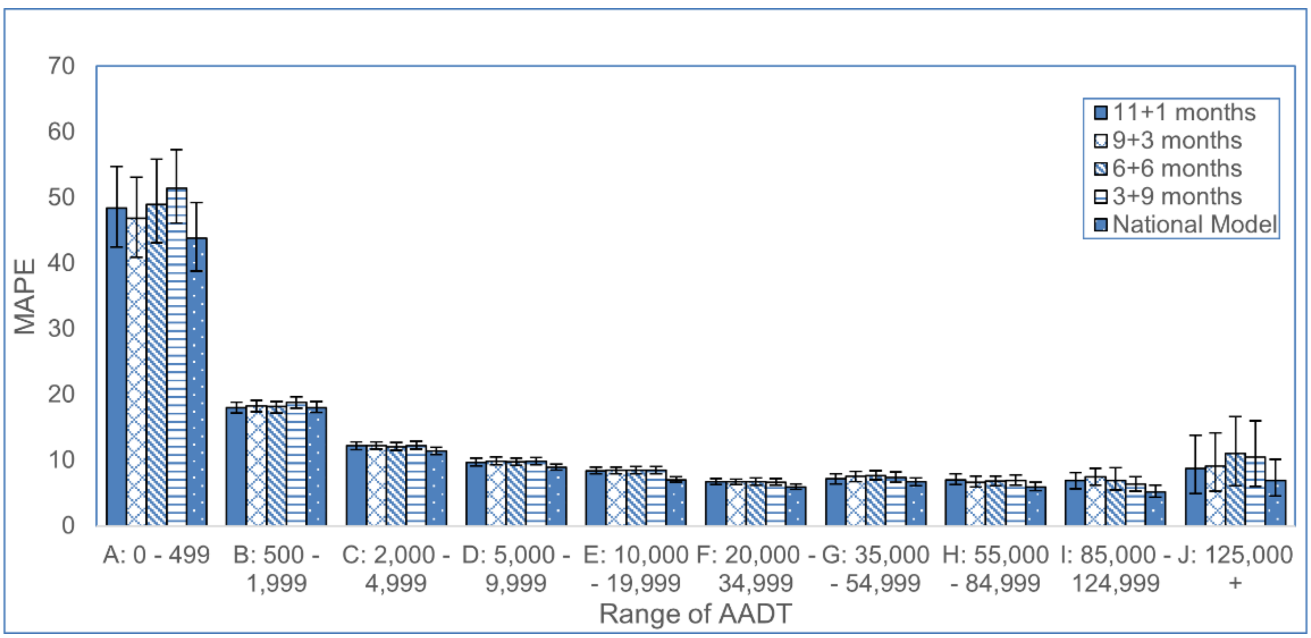
The predictive models for AADT using a rolling method perform remarkably similar to the National AADT model, which uses a full year of 2019 AADT data, and served as the benchmark (Table F1). Intuitively, the more months of training data borrowed from 2018, the higher the model error. When MAPE is compared across different sizes of roads (Figure F2), the error is fairly consistent across all road sizes, though small roads are most sensitive. It is expected that were this repeated for an abnormal year, such as 2020, or in an area experiencing much change, the results may be very different. But for a ‘typical’ year or in an area without much change, these results suggest mixing current year with previous year information can be predictive of future AADT.

Table F1: Errors compared across month combinations for the "rolling year" approach to forecasting.

Group	Median Bias (%)	95% TCE Error Range (%)	68th Abs. Percentile (%)	95th Abs. Percentile (%)	MAPE	NRMSE
11 + 1 months*	1.42	48.41	12.13	37.59	12.21	18.06
9 + 3 months*	1.38	46.18	11.67	35.46	11.60	17.08
6 + 6 months*	1.37	44.98	11.00	34.55	11.06	16.83
3 + 9 months*	1.22	44.92	10.98	35.70	11.03	14.98
National Probe Data Model	0.94	30.26	10.91	33.87	10.98	15.68

(*Note that the label of 'm + n months' represents the Probe Data in use is collected from the last m months in 2018 and the first n months in 2019.)

Figure F1: Comparison of AADT MAPE across road volumes for AADT forecasting using the rolling month approach. Error bars represent the 95th confidence interval of the mean.

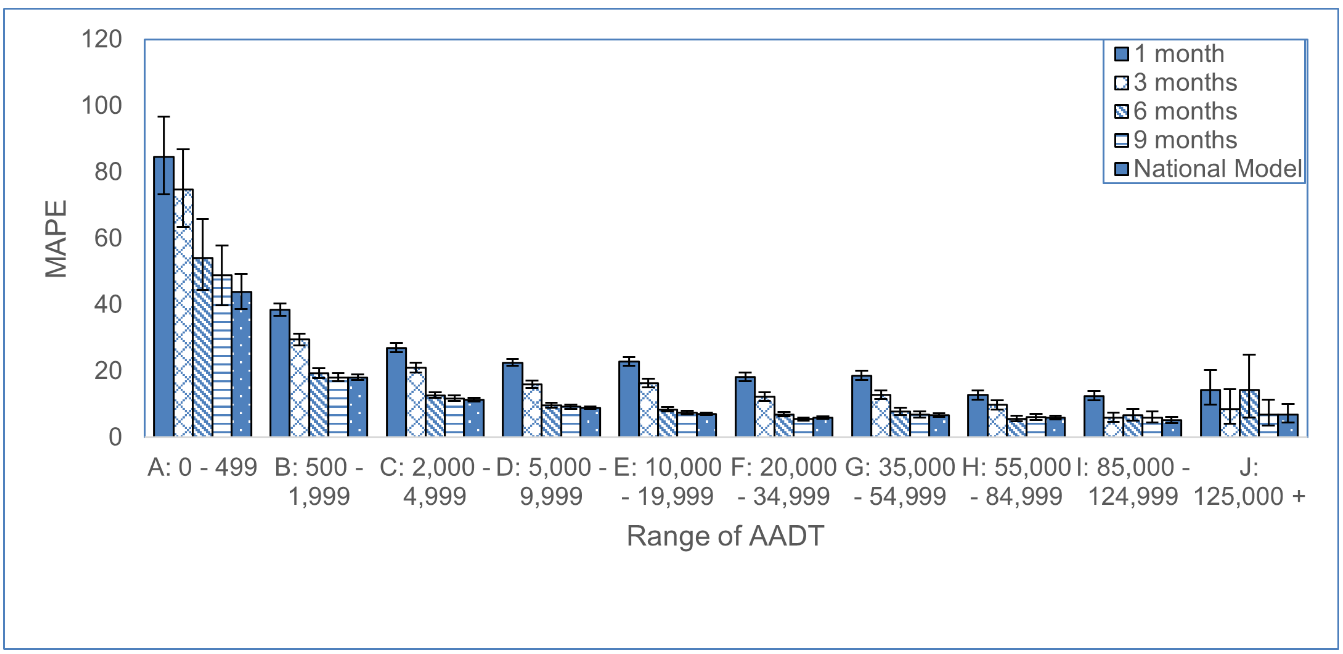


How early can AADT be forecasted without leveraging historical data from the prior year? In contrast to the "rolling months" approach, errors are much higher for forecasts based on one month (January), or 3 months (January, February, March) (Table F2). However, at six months the error moves into realms that may be acceptable for many applications. This trend is fairly consistent across all ranges of road volume (Figure F2).

Table F2: Errors compared across month combinations for the "existing year" approach to forecasting.

Group	Median Bias (%)	95% TCE Error Range (%)	68th Abs. Percentile (%)	95th Abs. Percentile (%)	MAPE	NRMSE
First 1 month	5.67	76.49	17.83	54.61	17.72	23.73
First 3 months	2.84	63.53	15.26	48.25	15.24	20.62
First 6 months	1.43	50.48	12.22	37.67	12.14	17.85
First 9 months	1.21	46.63	11.49	36.21	11.51	15.24
National Probe Data Model	0.94	30.26	10.91	33.87	10.98	15.68

Figure F2: Comparison of AADT MAPE across road volumes for AADT forecasting using the existing month approach. Error bars represent standard error of the mean.



Conclusions

The research shows that AADT can be forecasted before the years' end, assuming that the current year in the area of interest will be reasonably similar to historical trends (for example, not 5106 experiencing a pandemic, a closing of a major local employer, etc.). The results found much lower error in forecasting if historical trends and prior year data could be included. However, if training data is not available for the prior year, Probe Data can be used to forecast AADT as early as 6 months into the calendar year, and achieve comparable results to a model built off of a full year.

The key takeaway here is that more data available across time will lead to higher quality forecasting for the remainder of the year. If only January 2019 data is available to forecast the remainder of the year, estimates will be far more accurate if that one month of data from January 2019 is supplemented with the prior 11 months of data from 2018.
One major point of consideration is that models assume a typical year when forecasting. Therefore, models will be unable to forecast AADT if there are unforeseen events, such as regional natural disasters, or global pandemics. This is a significant caveat, that is especially relevant to any effort to forecast AADT estimates in 2020.

Effects of Reduced Trip Sample on Probe Data AADT Model Accuracy

When creating a modeled AADT estimation from Probe Data sources, a representative sample of trips is key to estimate AADT. As examples, for vehicle-derived traces, a provider may only collect data from certain makes of automobiles or only for more recent years (thus skewing the sample to higher income individuals). For LBS data, providers typically run software to collect location information on an array of applications built for smartphones. While this approach has less bias in terms of demographics, some apps do not collect data all days of the year leading to other types of sample gaps. To generalize: it is important to have the largest trip sample possible (as opposed to a high number of “pings” or “devices”) and that the sample be as unbiased as possible. This section explores how varying trip penetration (ratio of sampled trips to observed trips) impacts estimated AADT, assuming no large changes in trip bias as sample is reduced.

To model the variance in trip penetration expected between different data providers, subsampled from unique devices for both GPS and LBS Data Sources, and analyzed the resulting generated trips.

Impoverishing LBS samples by device is different from impoverishing the LBS data samples by trips. Devices correspond to smartphones, and may be related to app usage depending on the vendor, and thus can contribute a single trip, or thousands of trips to the sample over time. Reducing the number of devices included in the sample, as opposed to trips, better reflects potential variation in the field of data vendors. Some data vendors may have a very large number of devices included in their sample, while others might have fewer devices. The number of trips generated from each device may be independent of whether the device sample is small or large.

In order to randomly reduce the sample of devices, subsets of devices were selected from the base data source (shown in earlier sections to be the LBS source). First, 66% of the total device sample from each month of the calendar year was randomly selected, then 33% followed by 11%. By reducing the device sample, trip penetration rates also decreased. The resulting range of reduced sample trip penetration is described in Table I1. Penetration rates are calculated as the average daily Probe Data trip sample at a counter location divided by the AADT reported at the same counter.

Table I1: Distribution of sample trip penetration by different device sampling groups

Trip penetration rate (sampled trip count divided by traffic volume count)

Avg Monthly Devices	10th percentile	25th percentile	50th percentile	75th percentile
28 million (100%)	2.76%	4.00%	5.31%	6.69%
18 million (66%)	1.81%	2.66%	3.54%	4.46%
9 million (33%)	0.89%	1.32%	1.76%	2.23%
3 million (11%)	0.27%	0.43%	0.58%	0.74%

Specifically, the combined LBS and GPS (commercial and personal) data source used to build the Probe Data National AADT model had a median sample penetration of 5.31%. When the total device IDs sampled was reduced to 11% of total, the resulting trip sample penetration dropped to a median of 0.58%. This means that when the sample was reduced to about 3 million devices a month, for a road with AADT of 1000, only observe 5 trips on an average day would be observed. Thus, although trips were impoverished by device IDs, there was a corresponding reduction in sampled trips.

Our full reference data of 4,232 permanent counter stations from the lower 48 states across the U.S. were aggregated for this analysis. Data was aggregated across all months in 2019 and compared to 2019 AADT values reported by the permanent counters. See section 'AADT Estimation' for details on building, training, and validating the Probe Data National AADT model. For comparisons in this chapter, the Extreme Gradient Boosted model fit to all calibration sites was used as reference, as opposed blending results with the additional model fit to only small roads. This was chosen to minimize complexity of analysis.

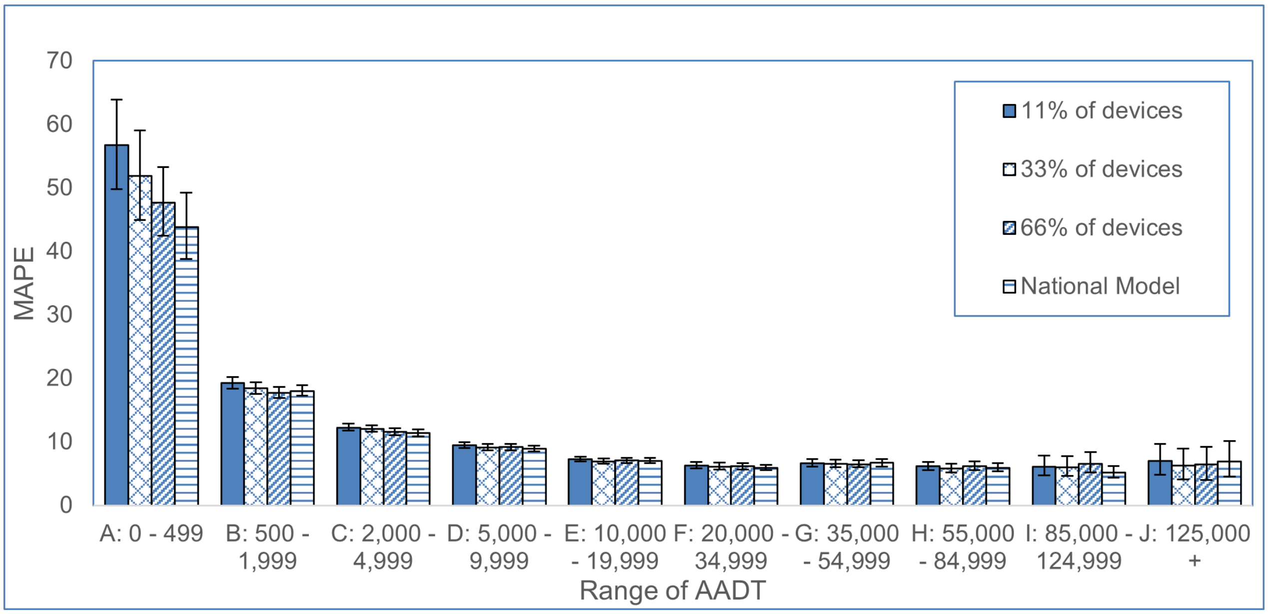
Probe Data National AADT Model Error Comparison Across Impoverished Trips

While the overall model results are robust to impoverishment, (Table id1), when results are isolated by road size, it becomes evident that smaller roads are impacted more severely by the reductions in trip samples (Figure I2). Reducing the sample trips has the largest impact on small roads, which is expected, as a sample trip is less likely to be observed on a given day.

Table id1: Error results broken down by device impoverishment

Group	Median Bias (%)	95% TCE Error Range (%)	68th Abs. Percentile (%)	95th Abs. Percentile (%)	MAPE	NRMSE
100% of Devices	0.94	30.26	10.91	33.87	10.98	15.68
66% of Devices	0.96	47.64	11.34	35.67	11.31	16.20
33% of Devices	0.85	48.43	11.25	37.15	11.63	15.78
11% of Devices	0.87	51.95	11.88	38.87	12.14	16.46

FigureI2: MAPE for 2019 AADT estimation across reductions in sample of unique device IDs across AADT groupings. Error bars represent standard error of the mean.



For large roads, reductions in the device sample may not have a dramatic impact on the quality of AADT estimates assuming good ML techniques. The device sample was reduced to 11% of the original size (average of 2.8 million devices monthly), with little change in model performance. For large roads, this suggests that one can get fairly accurate estimates of road volume with a lower proportion of devices, and by extension, sampled trips. It is likely that if this experiment were run to predict MADT, or daily or hourly volumes, that the smaller sample size would have a much more dramatic impact on volume estimation.

In contrast to larger roads, accuracy on smaller roads is more sensitive to a reduction in device sample. In order to maintain high accuracy on small roads, the largest sample of devices possible is necessary, and one that is representative across factors such as geography, time of day, purpose, etc. Only with such representative sampling can the model fully take advantage of other factors (weather, speed, urban v rural etc.) to create an accurate estimate. When the sample trips are reduced to only a few devices, it impedes the model's ability to adjust for the noise and bias in this Probe Data -- or any other source of data.

This report shows that sample trip penetrations under 1% tend to result in substantially worse model error for small roads, as compared to traditional factoring methods. Further work is needed to generate more detailed targets on necessary trip penetration across varying road sizes in order to generate the best estimations of road volume. Finally, how representative each trip is may vary from supplier to supplier. Different suppliers may have bias in the type of individuals they are able to gather trips from - such socioeconomic status or travel behavior (for a broader characterization, see StreetLight, 2020). For example, a supplier may track trips via a weather application installed only on iPhones. This may be vulnerable to biasing toward trip collection during unusual weather patterns, and individuals of a specific demographic. Thus, when choosing a supplier for trip samples, one needs to pay attention not just to the total number of unique individuals sampled, and resulting total trips, but one also needs to perform checks to ensure that the sample trips have even coverage both spatially and across time.

Effects of Reduced Ping Rate and Spatial Accuracy on MADT Model Accuracy - Cellular Data Simulation

There are a variety of methods by which Probe Data can collect a sample of trips from travelers, including from LBS sources (mobile apps), GPS systems, or from cellular devices, where location is inferred by nearby cellular towers. The main signal of these sources consisted of pings, which record the time, location in terms of a latitude and longitude, and a unique ID to connect the information to a unique traveler. Probe Data sources have a broad range of accuracy and precision for the location data, depending on both the mechanisms by which location is derived and also by provider, for frequently the location data is collected. This section seeks to explore how spatial and temporal precision of pings impacts the ability to create a model for monthly average daily traffic.

Method

To describe how temporal and spatial precision impact AADT estimation, LBS data was reduced both its temporal and spatial precision in order to mimic the variance in quality that exists among different sources of Probe Data, and between different data providers. to the goal was to mimic temporal and spatial precision of data collected from cellular networks and low frequency-pinging apps (such as who use spatial features for advertising only).

Probe Data trips derived from LBS sources are composed of pings that record the time and location information collected from US smartphone devices. The pings are normally located around the road segments with spatial errors. To transform the discrete pings into trips, the pings are stitched into lines and the trip locking algorithm maps the line geometries onto appropriate OSM road segments. Impoverished trips were constructed from a set of LBS trips collected from October 2019. October was chosen as a representative month since during a typical year, travel is assumed to be representative of school and work trips. Note that it was found that the MADT model performs with similar accuracy for the 2019 AADT model.

The baseline ping rate of the LBS source is 1-3 minutes, and the baseline spatial accuracy of these pings is 5-20 meters. This is referred to this as Tier 1, and for these experiments it represents the current ‘best in class’ Probe Data for deriving Probe Data trips (Table I1). Impoverished trips are created by reducing the ping rate and spatial accuracy from the baseline to different tiers. The impoverished pings were then mapped onto the OSM road segments through the same trip locking algorithm to generate synthetic trips. Five total combinations of trips with varying combinations of spatial and temporal precision were compared: PRt1 + SAt2, PRt1 + SAt3, PRt2 + SAt1, PRt2 + SAt3, and PRt1 + SAt1 as the baseline.

Table I1. Different tiers of reduced ping rate and spatial accuracy.

Tier number	**Ping rate (PR)**	**Spatial accuracy (SA)**
1	1 - 3 minutes (PRt1)	5 - 20 meters (SAt1)
2	5 - 15 minutes (PRt2)	50 - 500 meters (SAt2)
3		500 - 1000 meters (SAt3)

*Note that this scenario was not considered because the range of ping rates between 15 and 60 minutes is not realistic among any prevailing Probe Data sources.

Impact of Spatial and Temporal Impoverishment of Pings on MADT Model Accuracy

The descriptive analysis of different combinations of impoverished trips is shown in Table 2, where the correlation measures the linear relationship between the permanent counter MADT and impoverished LBS trips, and the penetration instability is measured by the coefficient of variation of the penetration rate. Among the statistics, both correlation and penetration instability represent the quality of impoverished trips, and both the sample count and penetration rate represent the quantity of impoverished trips. As can be seen, (1) when only reducing the ping rate or spatial accuracy to Tier 2, the quantity is reduced but the quality is not affected; (2) the quality is significantly affected when reducing the spatial accuracy to Tier 3; (3) however, holding the spatial accuracy at Tier 3, both the quality and quantity are improved when reducing the ping rate to Tier 2.

Table 2. Descriptive analysis of impoverished trips with reduced ping rate and spatial accuracy.

Group	Correlation	Sample count average	Penetration average	Penetration instability
PR (1-3min) + SA (5-20meters)	0.96	27,474	3.63%	0.389
(1-3min) + (50 - 500 meters)	0.94	19,934	2.70%	0.49
(1-3min) + (500 - 1000 meters)	0.71	6,846	0.88%	1.88
(5 - 15 minutes) +(5 - 20 meters)	0.96	16,467	2.17%	0.39
(5 - 15 minutes) + (500 - 1000 meters)	0.90	10,030	0.89%	1.12

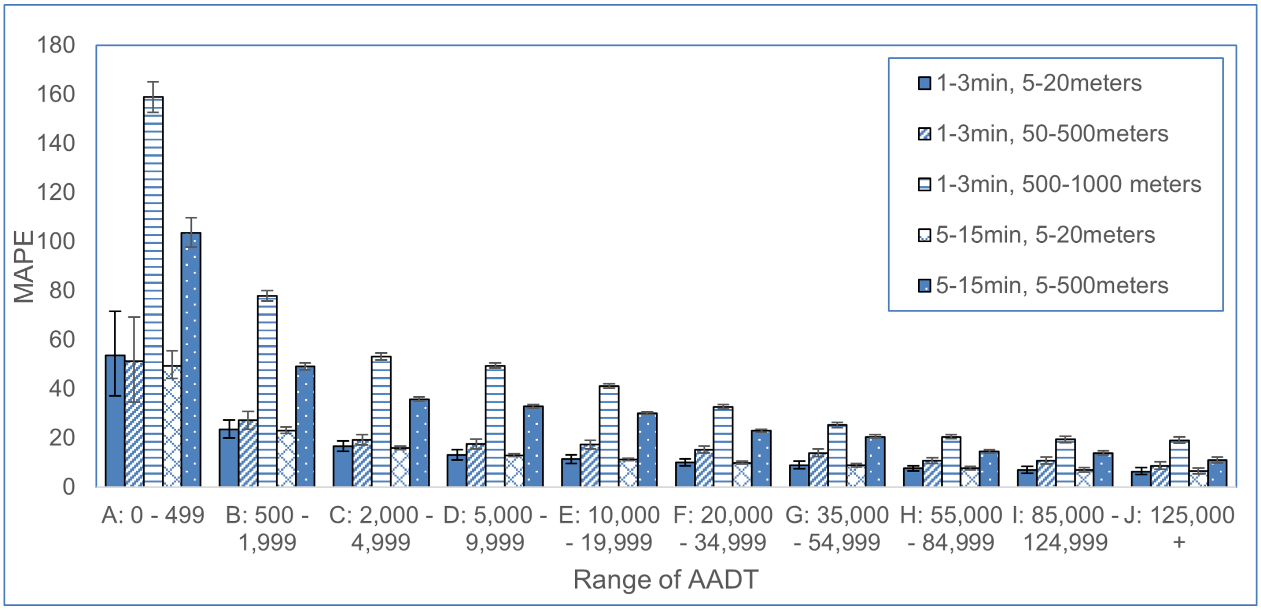
These impoverished trips were then used as the primary input to fit a Random Forest model for 2019 MADT. The same features and calibration set of permanent counters were used as with the National AADT model. Through the 10-fold stratified cross-validation, the validation accuracy and precision are shown in Table 3. Similar to the descriptive statistics, the validation errors increase as the spatial accuracy reduces, and the model fails to provide valid MADT estimates when the spatial accuracy reduces to Tier 3. Interestingly, when the ping rate reduces from Tier 1 to Tier 2, the metrics are not significantly affected. This implies an important finding for implementation - that increasing the precision and ping rate (which some suppliers may offer at significantly higher cost) will probably not improve AADT estimation accuracy.

Table 3. Validation metrics of impoverished trips with reduced ping rate and spatial accuracy.

Group	Median Bias (%)	95% TCE Error Range (%)	68th Abs. Percentile (%)	95th Abs. Percentile (%)	MAPE	NRMSE
(1-3min) + (5 -20meters)	2.23	57.31	16.09	44.71	14.93	19.61
(1-3min) + (50 - 500 meters)	1.86	78.99	20.57	58.54	19.86	28.31
(1-3min) + (500 - 1000 meters)	5.36	263.64	41.90	185.19	50.93	57.41
(5 - 15 minutes) +(5 - 20 meters)	2.18	57.92	16.35	46.81	15.33	19.78
(5 - 15 minutes) + (500 - 1000 meters)	4.06	147.85	34.32	109.94	34.24	39.00

The MAPE along with the 95% confidence interval of absolute percentage errors is shown in Figure 1. It can be clearly seen that even though reducing the ping rate or spatial accuracy to Tier 2, the model can still provide promising results across all different sizes of roads. Surprisingly, when spatial precision was at the lowest tier (precision of 500-1000 meters), Probe Data sourced trips were more predictive of MADT when they pinged less often (Tier 2, 5-15 minutes, purple bar) than baseline (Tier 1, 1-3 minutes, green bar). The current locking algorithms naturally have difficulty assigning pings to a likely road when the spatial accuracy is low. Therefore, the locking algorithm had an easier time assigning a trip to the road when it only had to reconcile a few very imprecise pings. When the locking algorithm is given very imprecise frequent pings, it creates nonsensical trips that correspondingly do not predict MADT trends well along roads.

Figure 1. MAPE and 95% CI of absolute percentage errors among different sizes of roads. Error bars represent the 95th confidence interval of the mean.



As spatial accuracy had the largest impact of Probe Data model quality, it was further investigated if the effects were unique to urban or rural areas (Table 4). As visible across metrics, sites from both urban and rural areas across the board show similar model error, with a trend toward slightly higher error in urban areas. This may be due to increased complexity in road networks, which could result in more trips being routed incorrectly across the denser roads, given noisier location data.

Table 4. Comparison of Probe Data AADT model error between urban and rural sites when spatial accuracy is reduced to 50-500 meters (holding ping rate at every 1-3 min). This spatial accuracy may be considered representative of data from cell-towers.

Group	**Median Bias (%)**	**95% TCE Error Range (%)**	**68th Abs Percentile (%)**	**95th Abs. Percentile (%)**	**MAPE**	**NRMSE**
Urban	2.31	80.65	21.78	60.54	20.57	25.47
Rural	1.48	76.99	19.6	56.39	19.22	23.99

Conclusion and discussion

Probe Data trips remained highly predictive of MADT even when ping rate was substantially reduced. Reducing the ping rate from 1 - 3 to 5 - 20 minutes or reducing the spatial accuracy from 5 - 20 to 50 - 500 meters mainly impacted the quantity of LBS trips collected from the road segments, not the quality. Thus, the model tested can still provide promising MADT estimation accuracy and precision across any sizes of roads. This is likely attributed to the strength of the locking algorithm, which is able to stitch together a most likely trip path with very little information. Thus, reducing ping frequency essentially reduces the observed sample trip penetration, but not necessarily the quality of the trips themselves.

Among the dimensions of spatial accuracy, when spatial accuracy was reduced to 500 - 1000 meters, most of the trips failed to be locked to the correct OSM road segments, and MADT model error concurrently had high error. Based on this current study, use of Probe Data sources with low spatial accuracy (such as data derived from cellular networks) is not recommended. Interestingly, holding the spatial accuracy at the worst scenario (500 - 1000 meters), a reduced ping rate from 1 - 3 to 5 - 20 minutes resulted in an improved quality of trips, even though the model still fails to provide promising results. The reason is that, for the LBS trips with large spatial errors, smaller ping rate would make the erroneous pings less clustered, provide clearer trajectories, and thus improve the possibility of accurate trip locking.

Although reasonable results were achieved with low ping rate, a limitation of this study may be that most of the study permanent counter stations are located on highways or primary roads in near sparse road networks. Thus, when pings are sampled infrequently, or sampled infrequently with low precision, the algorithm was still able to assign a trip to the correct road path. However, it's unclear if these results would hold in a dense urban area. In a region where many trip distances may be very short, such as with local errands, there may be bias in trips kept after impoverishment, which would have impacted results more. Similarly, in an urban area there may be more possible routes between two pings, and thus impoverishment of ping rate may be more likely to cause error in trip routing to the wrong route than was observed in the study sites. The impact of spatial and temporal accuracy on trip prediction needs to be further explored with an emphasis on dense urban areas.

This implies an important finding for implementation - that increasing the precision and ping rate (which some suppliers may offer at significantly higher cost) will probably not improve AADT estimation accuracy.

Impact of Reduced Reference Counters on AADT estimation

When estimating AADT across the U.S. using Probe Data, it is assumed that the quantity and quality of permanent counters for calibration will impact the estimation accuracy. The number of stations determines the quantity, while the geographic distribution of the counters determines the quality, assuming the traffic volumes provided by the permanent counters are close to 100% accurate.

The following study demonstrates how the Probe Data National AADT Model estimates are affected by a reduction in the quantity and quality of permanent counters available to train a model. This aims to determine the minimum required quantity and quality of permanent counters in order to maintain the acceptable AADT estimation. For this analysis, 4,232 permanent counter stations from 48 states across the U.S. were aggregated. Data was aggregated across all months in 2019 and compared to 2019 AADT values reported by the permanent counters, as in prior sections. Although greatest performance for AADT for small roads may be achieved by utilizing two separately fit models, for this analysis, a single Probe Data model was fit to the Extreme Gradient Boosting algorithm from all calibration sites, and used for comparisons (as described in Chapter 3). This was done to minimize complexity of analysis.

The first test examines what happens to the quality of AADT estimates from the National AADT Model as the total quantity of permanent counters are reduced. The National AADT Model was fit to either 10%, 20%, 33%, 50%, 80% or 90% of the total 4,232 unique AADT locations. All of the metrics are evaluated using k (10) folds cross validation, which means that the National AADT Model was trained on random folds of 90% of the total number of reference permanent counters. Thus, in this reports, the 90% impoverishment group represents the 'baseline' model for comparison. Equal impoverishment was checked across states and regions. This will help determine how much counter sample is required in order to achieve reliable AADT estimates.

The second test examines what happens to the quality of AADT estimates when counters are only selected from specific regions within the U.S. To do this, the continental U.S. was divided into five regions representing the West, North Central, South Gulf, South Atlantic and Northeast (Figure R1).

Figure R1. Five Defined Regions of the United States

Regional combinations were tested to determine whether regional biases or differences in trip penetration rates would impact the quality of the model. This will help determine how geographically dispersed the sample of permanent counters must be in order to achieve reliable AADT estimates.

Summary of Results

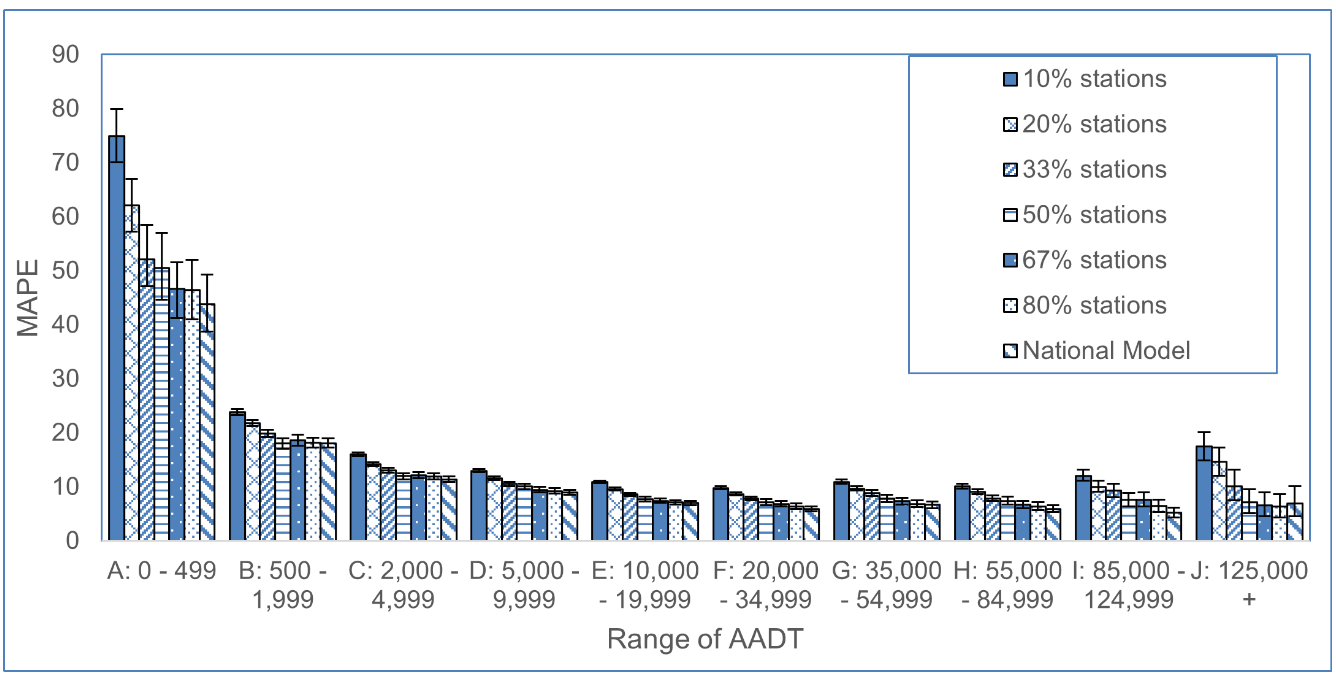
Table S1 illustrates the performance of the AADT model as the number of available stations are diminished incrementally. As expected, NRMSE and MAPE values degrade as the number of stations is reduced. Figure S2 illustrates how errors also relate to road size. As roads become smaller, the impact of station impoverishment is more significant.

Table S1: Comparison of National Probe Data AADT Model AADT error across different groups of station impoverishment. The National Probe Data AADT Model was evaluated on 90% of total stations as a result of 10-fold cross validation.

Group	Median Bias (%)	95% TCE Error Range (%)	68th Abs. Percentile (%)	95th Abs. Percentile (%)	MAPE	NRMSE
National Probe Data Model (90%)	0.94	30.26	10.91	33.87	10.98	15.68
80% of stations	1.03	47.44	11.41	36.92	11.52	16.25
67% of stations	0.89	48.2	11.82	38.53	11.85	17.00
50% of stations	1.23	48.6	12.08	38.88	12.13	17.75
33% of stations	1.19	51.23	13.29	41.07	13.14	21.02
20% of stations	1.38	59.54	14.51	45.72	14.63	24.32
10% of stations	1.61	65.02	16.54	49.00	16.57	27.05

As expected, NRMSE and MAPE values degrade as the number of stations is reduced. The image below illustrates how errors also relate to road size. As roads become smaller, the impact of station impoverishment is more significant.

Figure S2: Comparison of National Probe Data AADT Model AADT error across different groups of station impoverishment by road volume. Error bars represent standard error of the mean.



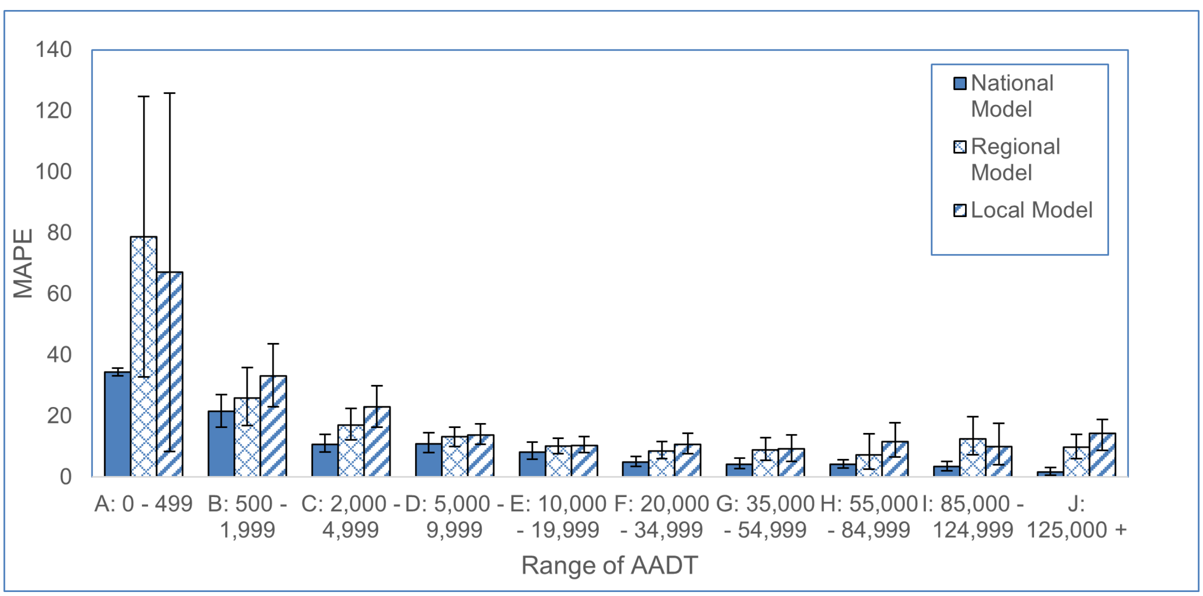
One-State Calibration

The results suggest that the spread of data (permanent sites) across a region is essentially the best estimate. However, if a location has a rich reservoir of permanent data, could a good model be created from local data alone? To test this proof of concept, two additional models were fit to estimate AADT for Texas. One model was fit with only permanent counters from Texas (n = 245), and the second was only fit with permanent counters from the south gulf region (n = 625). Texas was chosen for this representative study because the state maintains a large number of permanent counters (245 stations) across a wide range of AADT volumes. Methods for collecting features, fitting the extreme gradient boosted model, and testing error were the same as with the National AADT model. The results of AADT estimations for Texas for the National AADT model compared are described in Table S3, and a depiction of how error varied across road volume is visualized in Figure S3.

Table S3. Comparison of Texas AADT estimation error between three different models trained on different sets of permanent counter groups. Texas AADT was tested on a model fit with either Texas counters (state model), counters from the south gulf states (regional model), or the full national permanent counter set.

Group	Median (Bias)	68th Percentile Absolute Percent Error	95th Percentile Absolute Percent Error	MAPE	NRMSE
National Probe Data Model	0.16	9.65	34.64	9.8	8.21
Regional Model (south gulf)	-1.33	15.92	42.77	15.5	26.78
State model (Texas)	0.97	18.79	48.46	16.65	28.23

Figure S4: Comparison of AADT error for estimation of 2019 AADT in Texas for models trained with permanent counters within the state, region or nation across road volume



A National AADT model, trained from counters across the entire United States resulted in substantially less error than regional models fit on either only Texas permanent counters, or permanent counters from Texas' region. These results suggest that although local calibration information is useful for fitting an AADT model, a wide breadth of example calibration locations is also important for achieving accurate AADT estimations.

Conclusion

In summary, it should be noted that impoverishing the permanent counter data by regions appears to have a more significant impact on results than random proportional impoverishment. Therefore, in modeling AADT with Probe Data, it may be more beneficial to have regional diversity and fewer counter locations as opposed to a lot of permanent counters in a single region within the U.S.

If one is estimating AADT for a counter within a region in which broad calibration data exists, results will likely be strong. But when picking reference stations for calibration an AADT model, it is recommended that the locations be spread out to cover as many different road types and regions as possible to ensure the sample is representative and not biased toward a specific region or road type. In practical terms, if a state has the choice of investing in more local permanent stations or using other states' pre-existing permanent counter data for calibration, using the other states' permanent counter location will be more useful (and probably far less expensive).

Conclusions: Estimation of AADT With Probe Data

This report presents results of a National Probe Data AADT Model for 2019. This model paradigm can produce volume estimates for any road across the United States. When compared to errors from short term count expansion, AADT estimates from Probe Data are competitive for road volumes above 2000 AADT and below 500 AADT. For smaller roads overall, AADT are within useful bounds for most locations (as represented by MAPE), even if the model has higher error than short term expansion methods (as evaluated by the 95th TCE error range). Many states are Probe Data competitive with accuracy achieved by same-year, appropriately factored temporary count expansion when evaluated by MAPE (Table S4). For most roadways across the United States, particularly small and rural roads, there are no permanent counters nor short term counts available at the precise location for deriving volume estimates. For these roadways especially, Probe Data presents a powerful tool for obtaining useful volume estimations.

For any model, the most important variable is that the sample data being used is representative of the population. So, if there is uneven sampling of trips, such as by bias for certain regions of demographics by the Probe Data provider, that will impact model error. However, if the model inputs remain representative, you can decimate both the percentage of trips sampled (see impoverishment of trips section), and the time period over which the data is trained on fairly substantially (see impoverishment of months), and still achieve a usable model of AADT.

AADT models based on Probe Data also have a high potential to estimate even beyond reported in this study. The AADT model in the current study was based on yearly counts of Probe Data trips. However, it's possible that a bottom-up based model, where AADT is calculated from estimates of

ADT or MADT may have even less error. Additionally, a subset of states had much less error. For states with a robust collection of reference short term and permanent counters, a specific model fit to their region would likely have even lower error than found in this study. More research is needed to further explore how Probe Data AADT error can be further improved with regionally tuned models. Further research is also needed to further characterize improvements in model accuracy from a hybrid approach, which may incorporate a blend of short-term count strategies with Probe Data, which was outside the scope of this study.

And, of course, Probe Data offers the ability to append additional characteristics, not just monthly and daily variation across the year (as discussed in the next section) but also demographics, trip purpose, vehicle type and other factors which may enhance the application of the AADT collection process.

Chapter 5: Estimation of MADT using Probe Data

The National AADT Model, as described earlier, can also be expanded to account for unique months of the year in order to generate an estimate of monthly average daily traffic (MADT). Although MADT could be derived from the month of day factors discussed later, with Probe Data MADT can also be estimated directly.

Method

MADT estimates were generated using the same predictive and labeled data as the National AADT method. For a subset of permanent counters, 2020 monthly volumes were collected from MS2, a company with software that stores and processes states' count data, and other online sources. MADT models were also fit for early 2020 months, in order to evaluate model performance during highly atypical travel patterns, in light of the COVID-19 pandemic, which substantially altered travel patterns across the United States in spring of 2020. The predictive features considered mirrored the National AADT model, but they were fit to the monthly average daily traffic, rather than the yearly average. A separate model for each month was fit to a Random Forest algorithm. MADT was calculated for each month from all permanent counter locations based on the equation below:

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

where

$MADT_m$ = monthly average daily traffic for month m

VOL_{ihjm} = total traffic volume for 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 ($i=1,...,n_{hjm}$) for which traffic volume is available

h = hour of the day ($h=1,2,...,24$) - or other temporal interval

j = day of the week ($j=1,2,...,7$)

m = month ($m=1,...,12$)

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

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

Results

Table 1 depicts the errors of the MADT model, by road size for all months in 2019. As with the National AADT model, errors increase as the road size decreases. Errors across all metrics are higher than the National AADT model. However, for roads over 2000 AADT, MAPE remains under 20%.

Table 1: MADT model error by road size

Road Size Bin	Median Bias (%)	95% TCE Error Range (%)	68th Abs Percentile (%)	95th Abs. Percentile (%)	MAPE	NRMSE
A: 0 – 499	25.10	153.29	46.21	123.91	39.25	58.79
B: 500 - 1,999	5.89	103.96	25.67	78.94	24.85	36.40
C: 2,000 - 4,999	-1.48	46.49	20.33	42.20	16.80	21.47
D: 5,000 - 9,999	-1.26	40.76	17.82	38.50	14.77	19.56
E: 10,000 - 19,999	-1.31	37.77	15.94	34.84	13.52	18.24
F: 20,000 - 34,999	-1.97	31.40	13.89	29.16	12.13	18.21
G: 35,000 - 54,999	0.30	33.73	12.56	29.15	10.61	14.06
H: 55,000 - 84,999	-0.97	25.00	11.43	23.51	9.44	12.57
I: 85,000 - 124,999	0.62	29.52	11.54	25.16	9.91	14.62
J: 125,000 +	-2.08	21.26	10.57	20.96	8.68	11.87

In addition to understanding performance on roads of various sizes, it’s important to consider the performance of the model across months of the year. The goal is that the model would perform well across all calendar months, without monthly or seasonal biases. The following table (Table 2) illustrates those results for all months in 2019.The 95% percentile absolute error trends higher for summer months (June, July, August). As the Probe Data AADT model is a single national model made to support all regions, this may be due to variance in summer seasonality across locations that can be improved with a regionally calibrated model.

Table 2: MADT error by month in 2019

Month Number (2019)	Median Bias (%)	95% TCE Error Range (%)	68th Abs. Percentile (%)	95th Abs. Percentile (%)	MAPE	NRMSE
1	-10.50	36.09	16.69	33.35	14.33	24.69
2	-14.23	40.49	19.70	36.34	16.85	28.16
3	-10.69	36.74	16.89	33.06	14.49	22.44
4	-0.76	36.69	11.07	30.81	10.44	16.44
5	2.49	42.03	11.31	33.43	10.99	17.82
6	7.89	55.11	15.75	41.65	14.95	22.26
7	13.06	67.09	21.19	52.79	19.48	23.80
8	11.09	61.54	19.17	49.80	17.60	24.16
9	8.15	52.48	15.93	41.21	14.84	21.43
10	-0.60	36.39	11.34	31.40	10.55	15.17
11	-4.04	32.30	12.47	29.93	11.11	17.80
12	-8.30	36.04	15.36	33.24	13.26	20.51

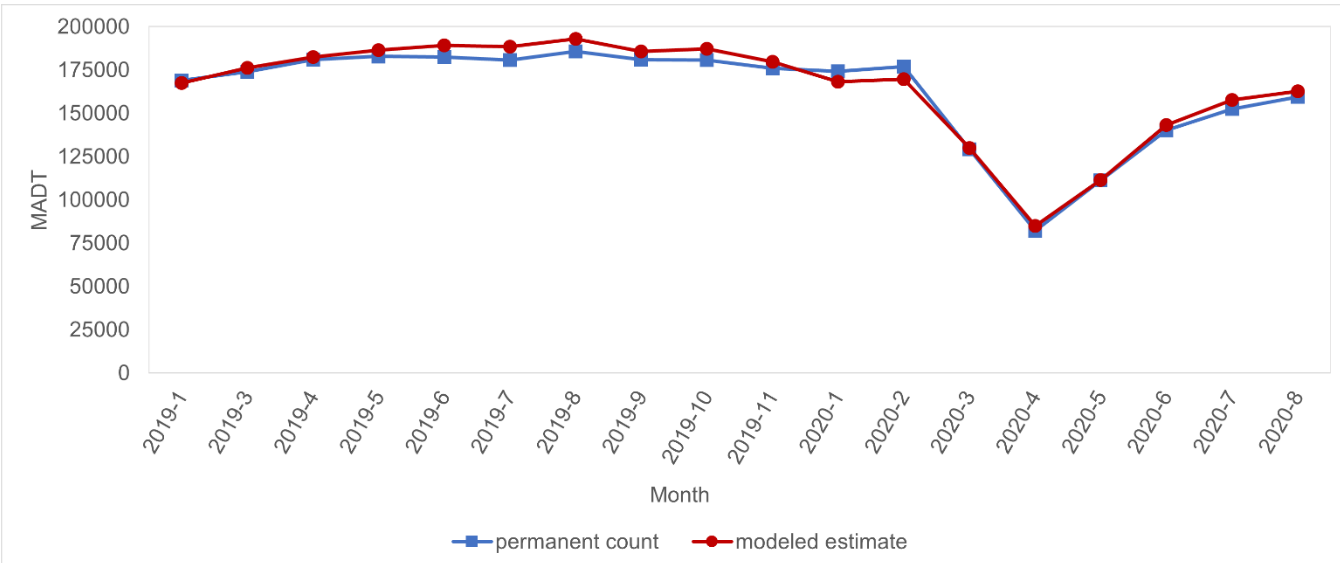
As mentioned earlier, the MADT model can be used to estimate traffic beyond the 2019 year, and could theoretically be updated on a monthly basis, or as frequently as data vendors supply updates. It should be noted that the model is able to monthly capture trends, even in unusual months. For example, in spring 2020 traffic patterns dramatically shifted (and primarily decreased) due to the COVID-19 pandemic. As seen in Table 3 below, these months still reflect fluctuation and range of errors consistent with “normal” months like those in 2019.

Table 3: MADT error by month in spring 2020. April and May were marked by severe reductions in road volume across the united states in response to the covid-19 pandemic.

Month Number (2020)	**Median Bias (%)**	**95% TCE Error Range (%)**	**68th Abs Percentile (%)**	**95th Abs. Percentile (%)**	**MAPE**	**NRMSE**
3	-2.41	35.74	14.21	33.05	12.64	20.29
4	4.69	55.83	17.18	43.24	15.85	26.47
5	5.07	61.11	17.51	46.92	16.70	26.79

A time series example at one location over time can be a useful way to visualize how the MADT model captures trends over time. Figure 1 shows MADT estimates over time for a representative permanent counter station in Massachusetts. Both the counter and the model reflect the steep decline in vehicle trips at the facility between March and May 2020.

Figure: Time series of MADT compared to estimated MADT for a counter facility from January 2019 through August 2020.



These results are promising, and indicate the power of Probe Data for vehicle estimates on a monthly basis, even in unusual circumstances. Probe Data captures dramatic changes in traffic volume, without the need for having any specific equipment installed ahead of time to measure changes. This is a very unique benefit of Probe Data, as compared to short term counters. Especially for locations with atypical trends, Probe Data will outperform short term counter expansion methods for MADT. Future research is needed to determine if further improvements can be obtained with custom models built for unique regions, such as cities or states.

Chapter 6: Estimation of Daily Traffic Using Probe Data - Including Special Events

In addition to modeling average annual daily traffic, there is value in exploring whether average daily traffic can also be estimated using Probe Data. A value of Probe Data, over short-term counts, is that observations of trips are taken 24 hours a day, 365 days a year, and thus have the potential to be sensitive to unusual events and trends such as holidays, special events, work zone periods and more. A model to predict ADT was created for the months of March, June, August and October 2019. These months were chosen are representative months across the seasons in the year. In addition to estimating ADT for the selected time period, another important component and goal was to estimate traffic during special events, or periods of construction or unusual weather.

Methods

Permanent counters for obtaining daily traffic volumes were obtained from the FHWA Traffic Monitoring Analysis System and MS2, as described in methods for the National AADT model. In selecting permanent counters for training, it was required that each counter used had a daily observation rate of at least 80% across the four selected calendar months (this method differs from the FHWA TMG recommendations.) Given the ADT model is designed to estimate daily traffic across a subset of training locations, the sample size of the training data increases 123 times (daily for the four selected months). This makes the model tuning more time consuming.

In total, 19 features were selected for the ADT model. In addition to standard features utilized for the base AADT model, some additional features were explored. Considering the strong oscillation (high penetration variance) of daily LBS and GPS samples, the preceding and succeeding daily and day of week counts were captured to improve the estimation accuracy. Among the time-related features, day of year and day of week were also selected.

After testing different models, a Random Forest model was chosen due to its computational efficiency and strength on lower volume facilities. Metrics were collected using k(10) folds stratified cross validation, using the same methods as detailed in the National AADT model.

Summary of Results

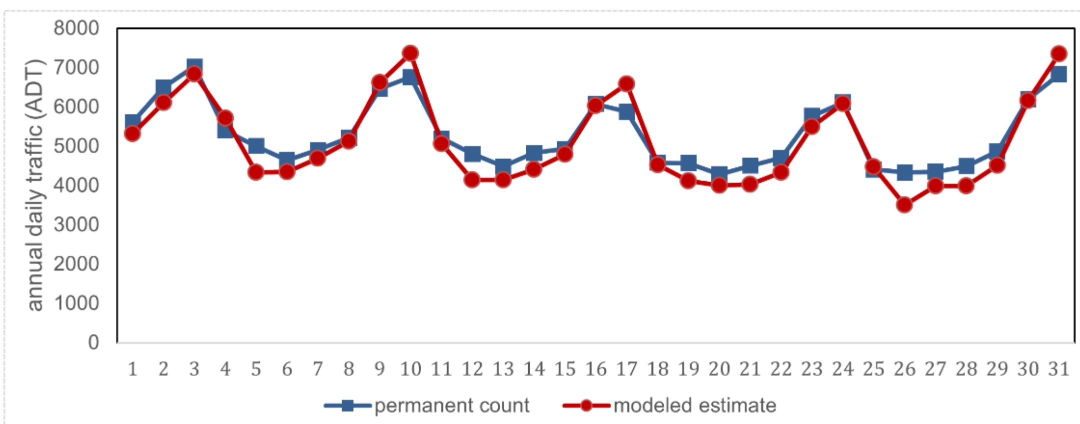
Table 1 details the model error from an ADT model. As observed with the National AADT model, high volume facilities have less error and bias than those on lower volume facilities.

Table 1: Comparison of ADT model error across AADT volume classes

Road Size Bin	Median Bias (%)	95% TCE Error Range (%)	68th Abs. Percentile (%)	95th Abs. Percentile (%)	MAPE	NRMSE
A: 0 – 499	39.4	149.62	57.69	120.82	48.96	60.37
B: 500 - 1,999	11.2	75.99	28.78	63.85	24.31	33.06
C: 2,000 - 4,999	3.28	50.81	20.61	44.74	17.26	24.13
D: 5,000 - 9,999	2.03	39.53	16.43	36.29	13.86	19.79
E: 10,000 - 19,999	1.27	32.85	13.23	30.1	11.23	15.91
F: 20,000 - 34,999	-0.2	27.51	10.64	25.72	9.25	13.2
G: 35,000 - 54,999	0.3	26.96	10.49	25.37	9.05	12.3
H: 55,000 - 84,999	-0.88	23.07	9.32	22.5	8.01	10.82
I: 85,000 - 124,999	-2.48	22.65	9.58	21.35	8.08	10.41
J: 125,000 +	-3.22	20.47	10.14	20.12	8.17	11.15

Figure 1 depicts the resulting daily volume estimations for a single road, as an example of a typical result on a medium sized road. Note the close alignment of daily trends throughout the days of week. The daily model also captured the seasonal changes of lower volume in March and October as compared to June.

Figure 1. Time series of ADT model estimations for an example permanent counter as compared to daily counts for selected months in 2019. Directional AADT is 4,116. Horizontal lines mark a break in the time series to a new month: March, July, August, October 2019.



Daily trends may vary significantly from the norm due to special events related to weather, construction or holidays. In order to evaluate the model's ability to estimate special events, it was necessary to know when and where special events occurred. Fifty-five sites across both Texas and Nebraska reported special events across the four-month data period. This resulted in 330 special event dates and 7,311 normal dates available for comparison. Table 2 contrasts the error in daily volume estimates between normal and tagged unique days. Results are promising, with errors for special event days are very close to errors for normal traffic days. Roads are evaluated by high, medium and low categories due to the limited number of special event days in total.

Table 2: Comparison of ADT model error between special events and normal traffic days across low, medium and high-volume roads.

Road Size Bin	Data type	Count (days)	MAPE	NRMSE
Low (0 - 4,999)	Special events	56	51.58	35.38
Low (0 - 4,999)	Normal traffic	2604	35.21	27.89
Medium (5,000 - 54,999)	Special events	245	9.21	15.45
Medium (5,000 - 54,999)	Normal traffic	4047	11.94	16.65
High (55,000 +)	Special events	29	15.47	22.74
High (55,000 +)	Normal traffic	660	14.52	21.29

In order to give more color to the metrics above, Figure 2 illustrates the daily permanent counter data compared to the model's estimated ADT for an example event in Nebraska. Permanent counter data is represented in blue (circles). Estimated ADT for normal traffic days is represented in orange (stars) and the estimated ADT for special events (related to weather events and construction) are highlighted in green (squares). Comparison of traffic volumes between the permanent counter and the Probe Data modeled is detailed in table 3. Note the model correctly dips as the road traffic stopped at the advent of a blizzard on March 13th and 14th, with surprisingly low percent error between the permanent counter and model. Interestingly, for this location, ADT is systemically slightly underestimated, including for the rebound in traffic following the blizzard. But as is apparently in the shape of the timeseries graph (Figure 2), the Probe Data model does capture day of week time trends well.

Figure 2: Time series of ADT model estimates for a single Nebraska station with reported blizzard on March 13 and 14 with post-blizzard bump on March 15. Horizontal lines mark the days directly before the blizzard, and after the post-blizzard rebound. Model estimates (red with circles) are compared to daily counts from a permanent counter (blue with squares). Location represents a medium size road, with an AADT of 22,846.

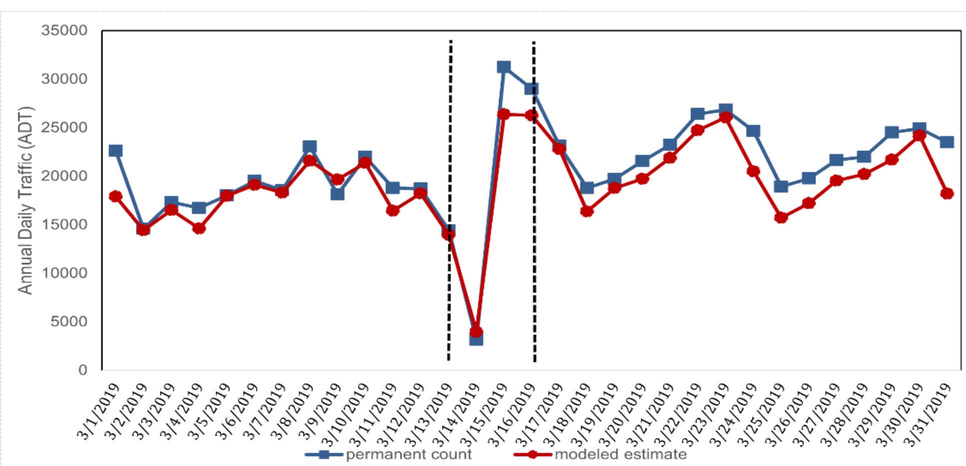


Table 3. Probe Data modeled estimates for daily traffic volumes for site 000054, surrounding a blizzard in 2019.

Event	Day	ADT (Permanent Count)	Model Estimate	Percent Error
None	2019-03-11	18782	15955	-15.05%
none	2019-03-12	18708	16487	-11.87%
blizzard	2019-03-13	14442	14695	1.75%
blizzard	2019-03-14	3149	3600	14.32%
post-blizzard	2019-03-15	31238	24963	-20.09%
none	2019-03-16	28990	24776	-14.54%
none	2019-03-17	23128	22402	-3.14%

A similar location in Nebraska also reported a dip in volume due to a blizzard, and was followed by a flooding detour for March 20 - Dec 31 (Figure 3), which routed more traffic onto the road. Modeled estimates for the blizzard, and first few days of the flooding detour are detailed in Table 4. Percent error spikes for the first day of the blizzard event (224%). However, the total traffic volume for this day is very small (313 cars). If expressed as percent change from the prior day, March 14th had a 91% percent drop in traffic, while the Probe Data ADT model predicts a large drop as well (71%). The Probe Data model is sensitive to this large plummet in traffic outside of normal patterns, but at such small volumes of traffic, there is an overestimation. This overestimation for very small traffic volumes is also present in the Probe Data National AADT model, as discussed in detail in Chapter 3. If you compare ADT estimates between early March and the rest of the year, you see the model consistently predicts substantially higher traffic during the detour period.

Figure 3: Time series of ADT model estimations for a single Nebraska station with reported blizzard on March 13 and 14 followed by a flooding detour on Mar 20 - Dec 31.. Horizontal dashed lines mark the day before and prior to the blizzard event, and the beginning and end of the flooding detour. Model estimates are presented as orange dashed lines with star points, and are compared to 'est_adt_se') traffic daily counts from a permanent counter (solid blue lines with circles). Site AADT is 5,861.

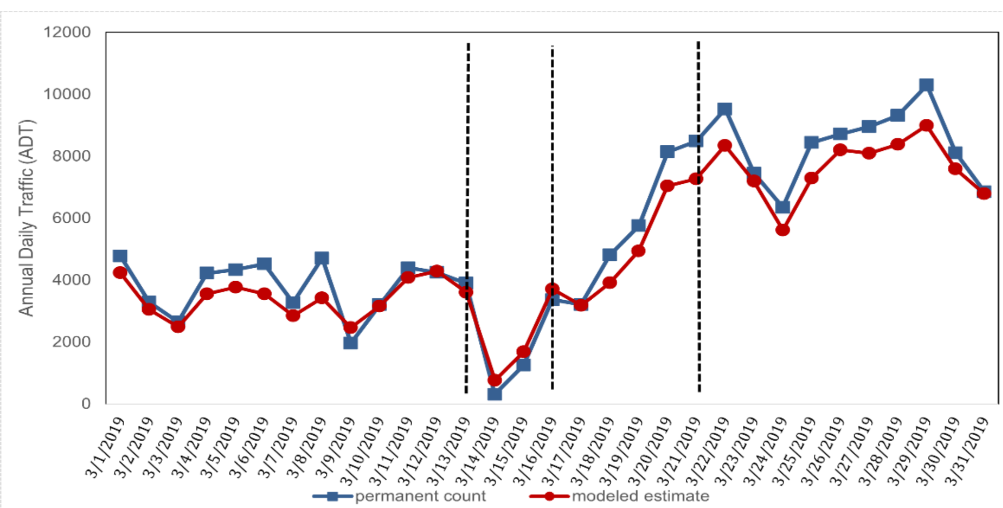


Table 4. Probe Data modeled estimates for daily traffic volumes for site 000059, with both a reported blizzard and subsequent flooding detour. Example error for the first few days of the flooding detour is provided as a representative detail.

Event	Day	ADT (Permanent Count)	Model Estimate	Percent Error
none	2019-03-13	3902	3627	-7.04%
blizzard	2019-03-14	313	1017	224.77%
blizzard	2019-03-15	1254	1599	27.55%
none	2019-03-16	3367	3547	5.35%
none	2019-03-17	3211	3141	-2.18%
none	2019-03-18	4808	3970	-17.42%
none	2019-03-19	5755	4796	-16.67%
flooding detour	2019-03-20	8143	7231	-11.20%
flooding detour	2019-03-21	8488	7024	-17.25%
flooding detour	2019-03-22	9525	7744	-18.70%
flooding detour	2019-03-23	7454	7036	-5.61%
flooding detour	2019-03-24	6353	5361	-15.62%

Table 5. Comparison of error between permanent counts and Probe Data ADT model for a typical or a-typical road conditions for site 000059.

Condition	Count (days)	NRMSE	MAPE
typical day	17	15.74	12.42
blizzard or detour	106	9	8.41

The selected time period of March, June, August and October 2019 did not include any federal holidays resulting in significant changes in expected travel behavior, however it did include Columbus Day or Indigenous People's Day which occurred on Monday October 14th, 2019. This pattern may be slightly different from the Monday norm across the month of October. Figure 4 highlights an example site in Georgia, which had a small decrease in typical Monday traffic on the holiday. The decrease on Columbus Day itself is accurately captured by the Probe ADT model. However, the traffic volume the Monday after Columbus Day is slightly underestimated (overly influenced by the decrease the prior Monday). For this size of road, the change in traffic from a small event such as Columbus Day is within the range of estimation error for the model. This is a nice representative illustration of how the ability of an ADT model to robustly represent changes in traffic pattern is be related to both the site traffic volume (and therefore expected model error based on the road volume), and the magnitude of change in traffic the model needs to detect. Larger volume roads with smaller expected error can robustly model more nuanced changes in daily traffic, while the Probe Data ADT model may only be able to track large events for locations with very small road volumes

Figure 4: Time series of ADT volume estimations for a single Georgia counter location with Columbus Day activity on October 14, 2019. Vertical dashed lines indicate Sundays to aid in comparing across unique Mondays. The week before, of and after Columbus Day are presented as a time series. Model estimates are presented as the red line with circles, and compared to counts from a reference permanent counter (blue line with squares). Site AADT is 57,929.

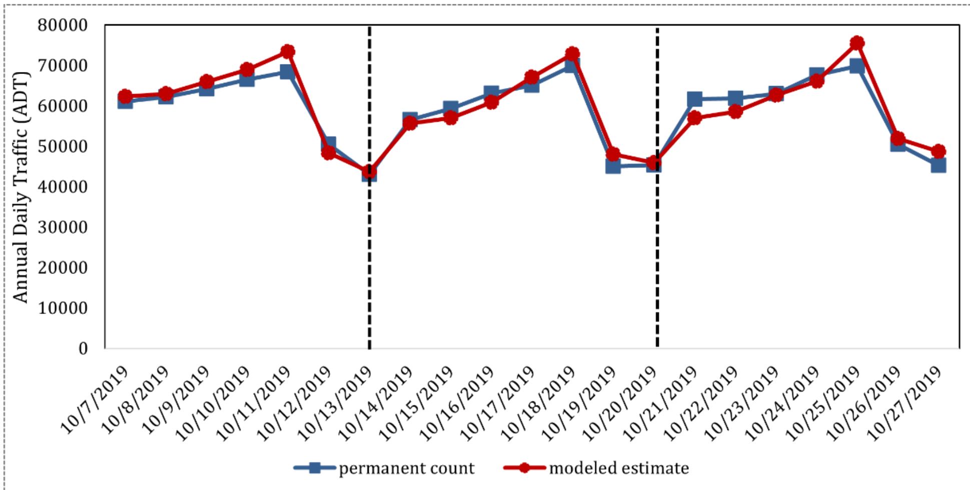


Table 6. Daily estimated traffic volume for the weeks of, prior to and post Columbus Day for site 000256 in Georgia. The location saw a slight decrease in typical Monday traffic over the holiday.

Event	date_num	ATR (permanent count)	Model Estimate	Percent Error
Monday prior	2019-10-07	61054	62284	2.01%
	2019-10-08	62176	62957	1.26%
	2019-10-09	64239	65982	2.71%
	2019-10-10	66546	68911	3.55%
	2019-10-11	68413	73346	7.21%
	2019-10-12	50500	48412	-4.13%
	2019-10-13	42999	43718	1.67%
Columbus Day	2019-10-14	56500	55717	-1.39%
	2019-10-15	59313	56946	-3.99%
	2019-10-16	63127	60904	-3.52%
	2019-10-17	65104	67113	3.09%
	2019-10-18	69924	72844	4.18%
	2019-10-19	45062	48029	6.59%
	2019-10-20	45368	45964	1.31%
Monday post	2019-10-21	61703	57023	-7.59%
	2019-10-22	61845	58540	-5.34%
	2019-10-23	62971	62648	-0.51%
	2019-10-24	67614	66118	-2.21%
	2019-10-25	69803	75461	8.11%
	2019-10-26	50471	51956	2.94%
	2019-10-27	45236	48641	7.53%

Conclusions

Overall, the ADT estimates trend very well with real world values. The model's ability to capture daily variation prompted by special event days is particularly valuable and emphasizes the strength and reliability of Probe Data for vehicular estimates. The fact that the model trended well with special events caused by both weather, construction/detours, and holidays further builds confidence in the model's ability to estimate daily values.

The strength of the daily model is likely due to the quality of the Probe Data data sample, which is robust enough to react to traffic changes of various sizes, both increases and decreases. For low-volume roads, the quality of the estimates get slightly affected, which is reasonable due to the instability of Probe Data trips on lower volume roads. As highlighted before, ADT highlights a very strong value of Probe Data. In the absence of physical sensors to record traffic, using Probe Data, the model was still able to capture atypical events and unusual trends.

Chapter 7: Estimation of Hourly Traffic Using Probe Data

In addition to estimating annual and daily traffic, there's a question of whether Probe Data can be used to estimate hourly traffic volumes. To explore this further, a model was created in order to estimate hourly traffic across a specific period of time. Based on available training data, Probe Data sourced from LBS and GPS (personal and commercial) pings was used to estimate hourly volumes across a representative two-week period from October 7th through October 20th in 2019. This period also includes a holiday, Columbus Day or Indigenous People's Day which occurred on Monday October 14th. A small time period was created due to the computational time required to train such a large annual dataset. From the collection of permanent counters for 2019 obtained from TMAS and MS2, only counters with 100% hourly observation rates across all the 14 days are selected when training the model.

In total, 18 features were selected for the ADT model. In addition to standard features utilized for the National AADT model, some additional features were explored. Considering the strong oscillation (high penetration variance) of daily LBS and GPS samples, the preceding and succeeding hourly and hour of day counts are captured to improve the estimation accuracy. Among the time-related features, day of week and hour were also selected.

After testing different models, a Random Forest model was chosen due to its computational efficiency and strength on lower volume facilities.

Summary of Results

For roads with AADT above 5,000, as MAPE lies below 20% (Table 1). For smaller roads under 5000 AADT, error increases as volume decreases.

Table 1: Error metrics for hourly traffic volume estimations by road size

Road Size Bin	Median Bias (%)	95% TCE Error Range (%)	68th Abs Percentile (%)	95th Abs. Percentile (%)	MAPE	NRMSE
A: 0 – 499	114.62	305.88	142.48	257.93	125.54	133.25
B: 500 - 1,999	30.85	129.18	48.31	109.57	40.64	44.43
C: 2,000 - 4,999	6.03	56.91	15.57	45.68	17.78	23.07
D: 5,000 - 9,999	2.78	40.48	10.48	32.74	14.36	18.55
E: 10,000 - 19,999	1.59	33.28	8.03	27.25	11.76	16.01
F: 20,000 - 34,999	0.08	27.79	5.54	22.5	9.8	13.4
G: 35,000 - 54,999	0.33	26.87	5.14	20.74	9.17	12.61
H: 55,000 - 84,999	-1.77	28.7	2.91	17.72	8.82	12.35
I: 85,000 - 124,999	-3.26	21.95	0.45	13.18	7.89	10.39
J: 125,000 +	-3.46	18.58	0.19	12.79	7.52	10.05

Figures 1, 2, and 3 below plot the hourly estimated volume (orange stars) for three example roads, in comparison to the counts from permanent counters (blue circles). Large and medium volume roads show consistent hourly trends across days, while the low volume road shows much less consistent results. For roads under 2000 AADT, the Probe Data trips are much “noisier”, and for smaller roads, specific hours of day may not even have any sampled trips at all. When roads are small enough, the model is unable to recreate the observed traffic patterns during the day. It may be of note that our example small road (Figure 3) lacks a strong daily pattern of traffic observed in the medium and high traffic roads too. There may be inherent randomness in the traffic for small roads that additionally makes translating the small Probe Data trip sample into a volume estimate.

Figure 1: Time series for one week of hourly volume model estimates on a large volume road (AADT 138,035, road size bin J) for Monday through Sunday. Modeled hourly estimates (red circles) are compared to permanent counts (blue squares). Note the consistent near overlap of values across hour of day, and between weekend and weekdays. Dashed lines indicate midnight to aid in reference of hour of day.

Line graph showing the hourly volume in Wisconsin site 400004 based on the time of day for a permanent counter and a modeled estimate for Monday through Sunday from October 7th through October 13th. There are vertical lines at the midnight hours. The permanent counter and a modeled estimate are basically equal throughout the entire graph. The graph has a repeating pattern of volume through the hours for Monday through Friday. It increases in the morning up to about 11000 then decreases mid-day down to about 6500 then increases back up to about 10000 then decreases back down to about 2000 at midnight. The Saturday and Sunday volume increase and decrease in a bell shape starting at around 2000 increasing slowly up to a peak at about 9500 a little later then mid-day then decreases down to about 3000 at midnight.

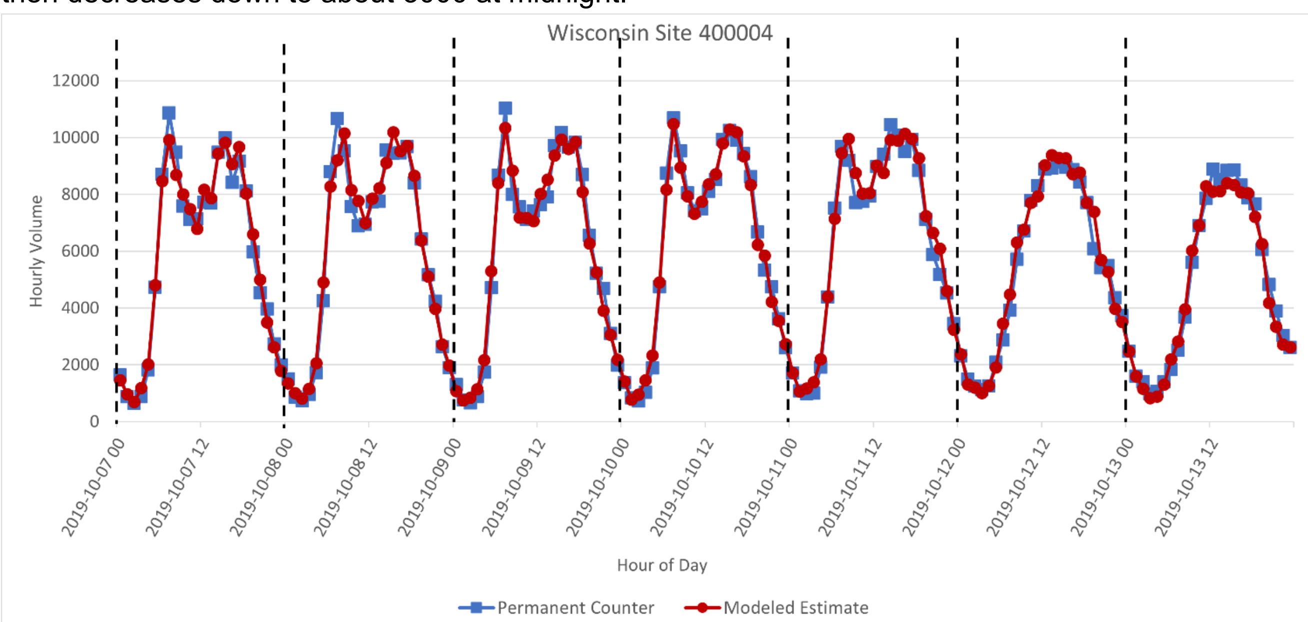


Table 2: Hourly volume estimates for one example day in the timeseries depicted in Figure 1, for a large volume road (AADT 138,035, road size bin J) in Wisconsin. For a list of all modeled estimates for this example counter, see the Appendix, Figure 1.

Date	Hour	Reference Volume (permanent count)	Modeled Estimate	Percent Error
2019-10-07	1	1635	1447	-11.47%
2019-10-07	2	866	960	10.88%
2019-10-07	3	645	684	6.03%
2019-10-07	4	867	1173	35.32%
2019-10-07	5	1815	2010	10.74%
2019-10-07	6	4731	4806	1.58%
2019-10-07	7	8702	8456	-2.83%
2019-10-07	8	10860	9912	-8.73%
2019-10-07	9	9478	8679	-8.43%
2019-10-07	10	7585	7997	5.43%
2019-10-07	11	7117	7482	5.12%
2019-10-07	12	7131	6781	-4.91%
2019-10-07	13	7721	8156	5.63%
2019-10-07	14	7689	7868	2.32%
2019-10-07	15	9487	9437	-0.53%
2019-10-07	16	9984	9817	-1.67%
2019-10-07	17	8412	9058	7.68%
2019-10-07	18	9167	9662	5.40%
2019-10-07	19	8110	8029	-1.00%
2019-10-07	20	5970	6592	10.42%
2019-10-07	21	4539	4992	9.99%
2019-10-07	22	3957	3494	-11.69%
2019-10-07	23	2734	2613	-4.43%
2019-10-07	24	1986	1785	-10.11%

Figure 2: Time series of hourly volume model estimates on a medium volume road (AADT 6688, road size bin D) for Monday through Sunday. Modeled hourly estimates (red circles) are compared to permanent counts (blue squares). Note the consistent near overlap of values across hour of day, and between weekend and weekdays. Dashed lines indicate midnight to aid in reference of hour of day.

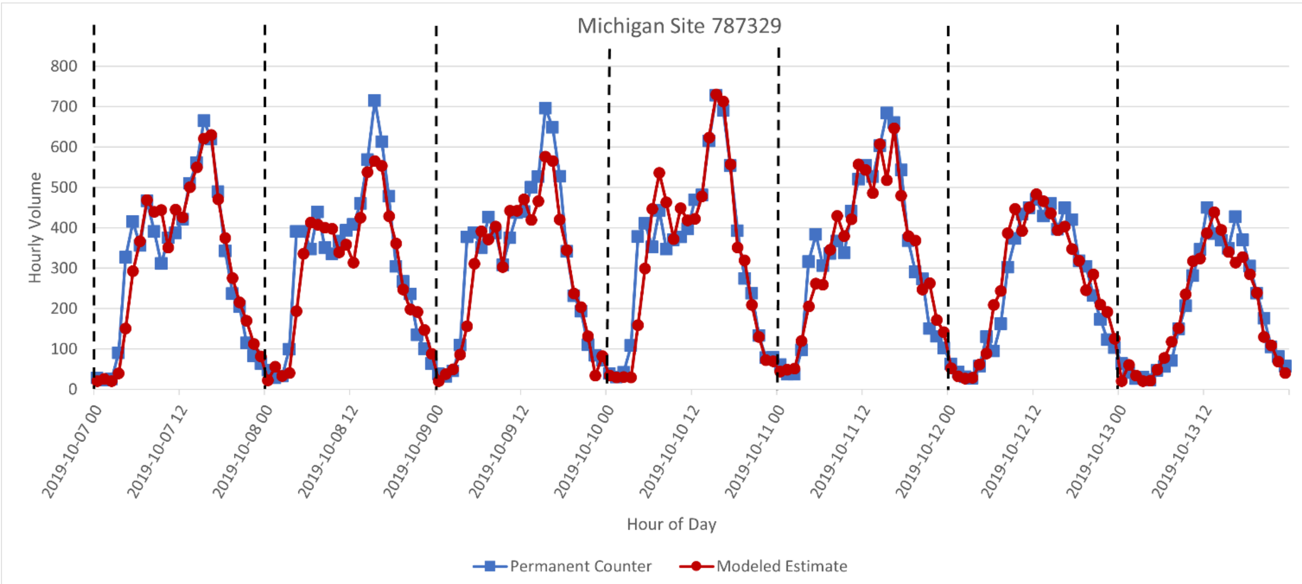


Table 3: Hourly volume estimates for one example day for the timeseries depicted in Figure 2, for a medium volume *road (AADT 6688, road size bin D)* in Michigan. For a list of all modeled estimates for this example counter, see the Appendix, Figure 2.

Date	Hour	Reference Volume (permanent count)	Modeled Estimate	Percent Error
2019-10-07	1	28	21	-25.81%
2019-10-07	2	23	26	13.33%
2019-10-07	3	25	20	-19.78%
2019-10-07	4	90	40	-55.85%
2019-10-07	5	327	150	-54.00%
2019-10-07	6	415	293	-29.40%
2019-10-07	7	356	366	2.85%
2019-10-07	8	466	468	0.48%
2019-10-07	9	391	440	12.42%
2019-10-07	10	311	444	42.77%
2019-10-07	11	375	351	-6.50%
2019-10-07	12	387	445	14.97%
2019-10-07	13	421	426	1.19%
2019-10-07	14	510	500	-1.88%
2019-10-07	15	561	550	-1.98%
2019-10-07	16	665	620	-6.73%

Date	Hour	Reference Volume (permanent count)	Modeled Estimate	Percent Error
2019-10-07	17	619	630	1.80%
2019-10-07	18	489	470	-3.88%
2019-10-07	19	342	374	9.43%
2019-10-07	20	237	276	16.29%
2019-10-07	21	205	215	4.91%
2019-10-07	22	115	170	47.67%
2019-10-07	23	83	113	35.80%
2019-10-07	24	64	81	26.41%

Figure 3: Time series of hourly volume model estimates on a low volume road (AADT 901, road size bin B) for Monday through Sunday. Modeled hourly estimates (red circles) are compared to permanent counts (blue squares). Note the consistent near overlap of values across hour of day, and between weekend and weekdays. Dashed lines indicate midnight to aid in reference of hour of day.

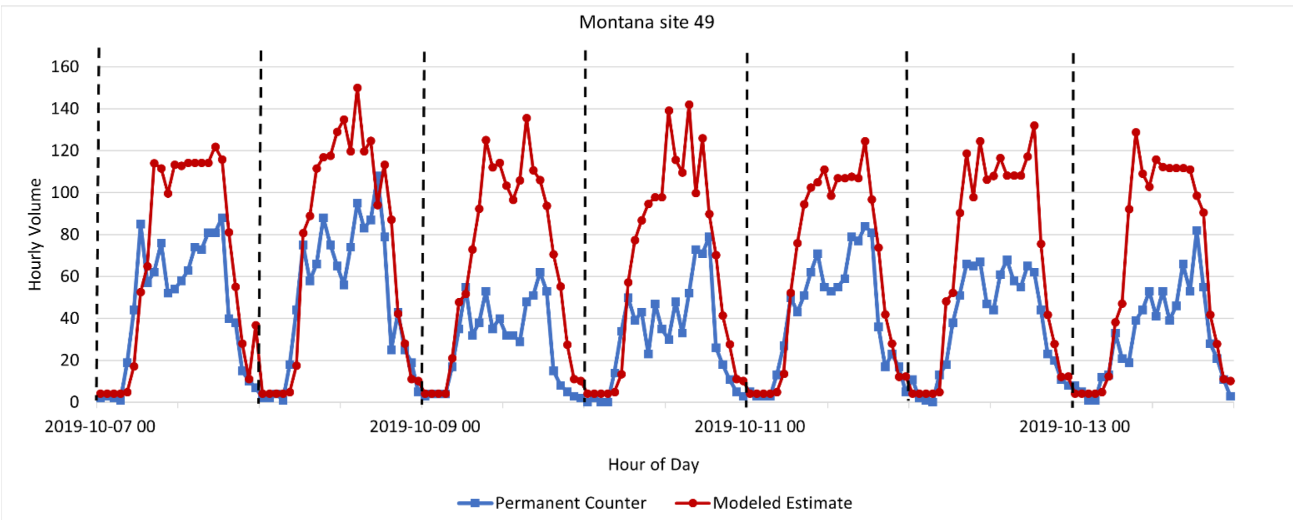


Table 4: Hourly volume estimates for one example day in the timeseries depicted in Figure 3, for a large volume *road (AADT 901, road size bin B)* in Montana. For a list of all modeled estimates for this example counter, see the Appendix, Figure 3.

Date	**Hour**	**Reference Volume (permanent count)**	**Modeled Estimate**	**Percent Error**
2019-10-07	1	2	4	110.57%
2019-10-07	2	3	4	40.38%
2019-10-07	3	2	4	110.57%
2019-10-07	4	1	4	321.13%
2019-10-07	5	19	5	-74.21%
2019-10-07	6	44	17	-60.87%
2019-10-07	7	85	53	-38.11%

Date	Hour	Reference Volume (permanent count)	Modeled Estimate	Percent Error
2019-10-07	8	57	65	13.93%
2019-10-07	9	62	114	83.84%
2019-10-07	10	76	112	46.72%
2019-10-07	11	52	100	91.70%
2019-10-07	12	54	113	109.95%
2019-10-07	13	58	113	94.61%
2019-10-07	14	63	114	81.41%
2019-10-07	15	74	114	54.44%
2019-10-07	16	73	114	56.56%
2019-10-07	17	81	114	41.10%
2019-10-07	18	81	122	50.57%
2019-10-07	19	88	116	31.72%
2019-10-07	20	40	81	102.98%
2019-10-07	21	38	55	45.08%
2019-10-07	22	15	28	87.40%
2019-10-07	23	10	11	11.66%
2019-10-07	24	7	37	424.29%

Conclusions

For large and medium volume roads, an hourly model was able to provide volume estimations for unique hours on specific days. This suggests that Probe Data could be useful for predicting changes in traffic at a high resolution. However, the hourly model struggles in comparison to annual and daily models on lower volume roads. While robust when aggregated at the daily level, the Probe Data signal may not be strong enough to capture hourly trends in locations for locations with low AADTs. Consider an example low volume low volume road location (AADT of 1000) with a sample trip penetration of 5%. An average of 2 sample trips would be observed over any given hour. Statistically, this Probe Data trip sample is a very small number to create a statistical inference from. Given these challenges, further research could be done for ways to improve hourly estimation for low volume locations. For example, a hybrid modeling approach, with a linear model for low volume roads may improve performance, as well as additional network-based features, which could include matched hourly counts from nearby roads. On the other hand, the application of exact hour-of-the-year estimation may not be often necessary for individual small roads. Finally, more research is needed to understand how the Probe Data model error for hourly estimation model compares to estimates from short term expansion methods. More research is needed to further evaluate the benefit the Probe Data may have, in contrast to current methods, in describing hourly trends for specific days at specific locations.

Chapter 8: Traffic Items

Probe Data can be leveraged to get average volume estimates, from a yearly average down to hourly resolution for large roads. The following section explored the ability of Probe Data to also model other traffic related items, including AADT by vehicle classification, seasonal factors, and factors based off of design hours (K-factor, D-factor, percent peak for trucks)

Vehicle Classification AADTs

Introduction

While it is important to estimate the overall AADT for any location, being able to estimate the AADT for different vehicle classes is extremely helpful for road design,safety, and is a requirement for FHWA reporting. This section will discuss how Probe Data can be used to provide AADT estimates for three different vehicle classes. They are:

- Pessenger- Vehicles (PV): this corresponds to the FHWA Classification categories 1 to 3 (MC, PV and LT).
- Single Unit (SU) trucks and buses: this corresponds to the FHWA categories 4 to 7.
- Combination Unit (CU) trucks: this corresponds to the FHWA categories 8 to 13.

Methods

The data that are used for this analysis comes from the TMAS classification dataset. This dataset provided traffic volume at the hourly level for different vehicle classes for each counter. With that, AADTs were calculated the three vehicle classes of interest. It should be emphasized that the analysis for this part of the project is at the counter level (uni-directional), as opposed to the site level which is bidirectional.

Using the labelled data calculated from the permanent counts and Probe Data sample trips, three different machine learning models were developed to estimate the AADTs for the three vehicle classes. For all models, the data was fit with a gradient boosting algorithm. Predictive features considered were the same as for the National AADT model with some additions. It was hypothesized that trucks may drive longer distances on average than a personal car. So the average trip length, standard deviation of trip length, average trip duration and standard deviation of trip duration were added as additional features derived from the LBS trip dataset.

It should be noted that model performance was evaluated by road size categories, which are usually defined using the overall AADT for all vehicle classes for a particular location. However, a high-volume road can have widely ranging proportions of SU and CU trucks volume, sometimes none at all because SU and CU trucks are not allowed on specific roads. Hence, reporting the errors in the estimation of AADT for the SU and CU trucks in this case under high volume road is misleading because it would lead to very large errors due to negligible truck volume. Subsequently, in each of the following subsections, road size categories are broken down by the AADT of that particular vehicle class.

Results

Estimation of AADT for PV Vehicles

The accuracy and precision of PV vehicle volume estimation was analyzed across breakdowns by total PV AADT (Table 2). The Pearson-correlation coefficient between permanent counts and the PV AADT model was very strong at 0.99 (p-value <0.001, Figure 1). As PV vehicles make up the majority of the traffic volume on the roads, the estimation of AADT for PV vehicles is very similar, with slightly lower error, than to that of the total road AADT. As with the Probe Data national AADT model, in general there is higher error on roads with lower traffic volume.

Figure 1: Comparison of estimated PV AADT between permanent counters and Probe Data model

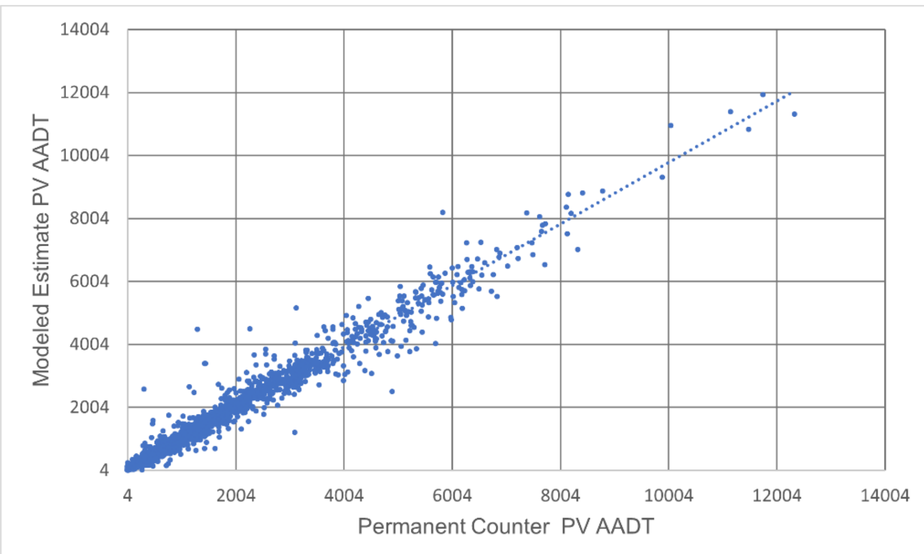


Table 2: Model errors for the AADT estimation of the PV s by across PV AADT groupings

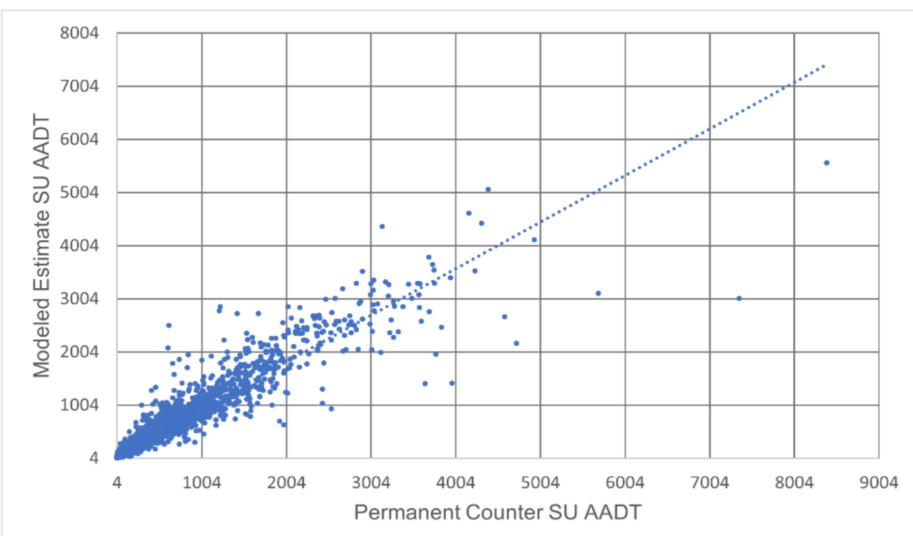
Road Size Bin*	Median Bias (%)	95% TCE Error Range (%)	68th Absolute Percentile (%)	95th Absolute Percentile (%)	MAPE (%)	NRMSE (%)
A: 0-499	14.51	208.95	39.26	167.84	42.72	56.45
B: 500-1,999	2.90	48.29	14.54	38.62	13.98	20.30
C: 2,000-4,999	-0.16	36.68	10.56	27.46	9.89	15.81
D: 5,000-9,999	0.26	26.61	9.44	24.51	8.75	12.77
E: 10,000-19,999	0.83	26.72	7.91	22.69	7.49	11.00

Road Size Bin*	Median Bias (%)	95% TCE Error Range (%)	68th Absolute Percentile (%)	95th Absolute Percentile (%)	MAPE (%)	NRMSE (%)
F: 20,000-34,999	-0.90	19.81	7.19	18.98	6.90	11.05
G: 35,000-54,999	0.23	22.29	6.78	18.13	6.56	9.14
H: 55,000-84,999	-1.06	21.22	7.57	18.52	7.57	12.47
IJ: 85,000+	1.51	29.89	7.56	20.86	8.42	11.31

*Note that the "Road Size Bin" column represents the annual average daily volume of CUs in the permanent counter vehicle class data.

Estimation of AADT for SU Trucks and Buses

SU vehicles comprise a more ambiguous class of vehicles as compared to CU trucks and PV vehicles. There was a strong and significant correlation with estimates from permanent counts(Figure 3. When considering variable feature importance, a spread of variables were used by the model, including total LBS sample trips, GPS (commercial) sample trip counts, and latitude/longitude information, which highlights that there was not as strong of a single predictor for SU AADT, as compared with PV (Figure 6) or SU AADT (Figure 2). *Figure 3: Comparison of estimated SU AADT between permanent counters and Probe Data model*



As compared to PV AADTs, the total distribution of traffic volume for SU vehicles across all permanent stations was lower, suggesting fewer SU Trucks and Buses overall on roads as a percentage of total traffic. Estimation error for SU vehicles trends slightly higher across all road volumes than for PV AADTs, which may be related to the lack of a single strong predictive feature.

Across groupings of road size, errors were very similar between SU vehicle AADT and total road AADT.

Table 3: Model errors for the AADT estimation of SU trucks and buses by SU vehicle AADT road size bins

Road Size Bin*	Median Bias (%)	95% TCE Error Range (%)	68th Absolute Percentile (%)	95th Absolute Percentile (%)	MAPE (%)	NRMSE (%)
A: 0-499	6.73	141.94	30.26	99.46	32.34	42.81
B: 500-1,999	-2.11	64.72	21.51	51.42	19.54	28.54
C: 2,000+	-6.92	58.82	16.79	47.68	16.44	28.62

*Note that the “Road Size Bin” column represents the annual average daily volume of CUs in permanent station data.

Estimation of AADT for CU Trucks

Estimated CU Truck AADT had a strong relationship with counts derived from permanent counters, with a Pearson-r correlation of 0.99 (p-value < 0.001, Figure 5). Further breakdown of errors by CU truck AADT road size (Table 4) confirms lower error as compared to the national AADT model and AADT for SU vehicles (Table 4).

Figure 5: Comparison of estimated SU AADT between permanent counters and Probe Data model

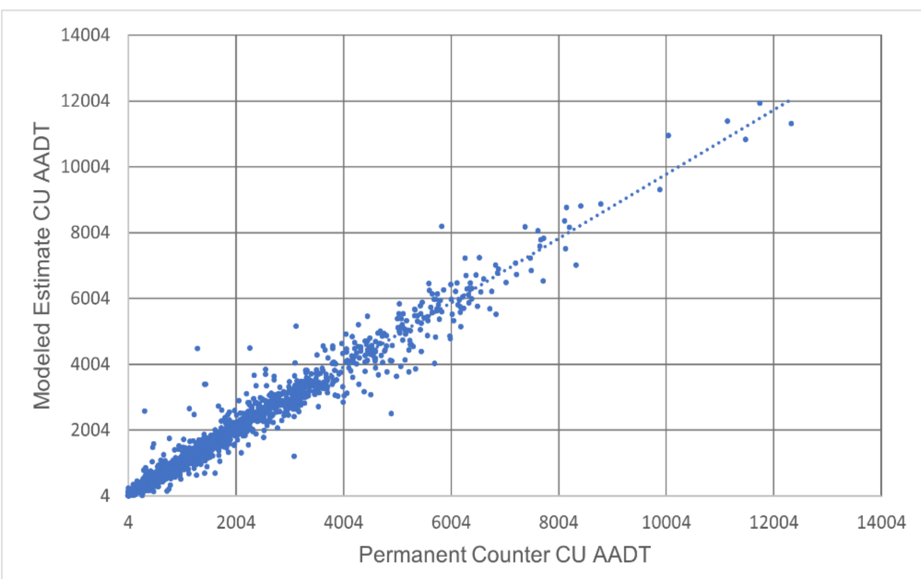


Table 4: Model errors for the AADT estimation of CU trucks by CU truck AADT road size bins

Road Size Bin*	Median Bias (%)	95% TCE Error Range (%)	68th Absolute Percentile (%)	95th Absolute Percentile (%)	MAPE (%)	NRMSE (%)
A: 0-499	2.93	141.08	28.59	104.45	34.36	51.75
B: 500-1,999	-0.39	48.50	14.27	38.61	13.79	23.89
C: 2,000-4,999	-1.17	26.56	9.44	25.14	8.79	13.06
D: 5,000+	-1.55	27.30	7.97	20.41	7.46	9.84

*Note that the "Road Size Bin" column represents the annual average daily volume of CUs in permanent station data.

There was a strong predictive relationship between CU Truck AADT and the count of GPS commercial heavy sample trips at each permanent counter location (Figure 6), which likely accounts for the strong performance of this model on small roads, as compared to SU vehicle AADT. The count of total GPS sample medium commercial trips was also utilized by the model, as the GPS sample tagged classifications of 'heavy' and 'medium' from the given GPS data source did not completely overlap with the FHWA vehicle classification scheme.

Conclusions

When model error for AADT for specific vehicle classification (PV, SU, CU) was evaluated by groupings of road volume, model errors were comparable with the National AADT model. However, the AADT estimation for PV cars and CU trucks had lower errors and stronger relationship to class-specific AADTs derived from permanent counts, compared to SU trucks. Use of data explicitly derived from medium and heavy-duty commercial trucks is important to accurately estimate CU volumes, and less important for SU and PV volume estimation. The National AADT and SU AADT models, in contrast, rely heavily on a mix of features from both GPS (commercial, personal), LBS, and additional contextual data to achieve optimal model results.

Estimation of Seasonal Factors

Methods

Travel patterns change over time and are affected by seasonal, monthly, day of week, and time of day factors. In order to model the expected seasonality of a location, seasonal factors are calculated using continuous count data to properly annualize short duration counts. This section explores the ability of Probe Data to provide estimates for Month of Year (MOY), Day of Week (DOW), and Time of Day (TOD) factors, which are computed as follows:

$$M_j = \frac{AADT}{MADT_j} \text{ and } D_i = \frac{AADT}{ADT_i} \text{ and } T_k = \frac{AADT}{AHDT_k}$$

where:

M_j is the monthly factor for month j of the year
D_i is the day of the week factor for day i of the week
T_k is the time of day factor for k hour of the day

The months, days and hours with higher average daily traffic would have lower MOD, DOW and TOD factors and vice versa. The DOW factors can be calculated for each of the seven days (seven factors) or for weekdays and weekends separately (two factors).

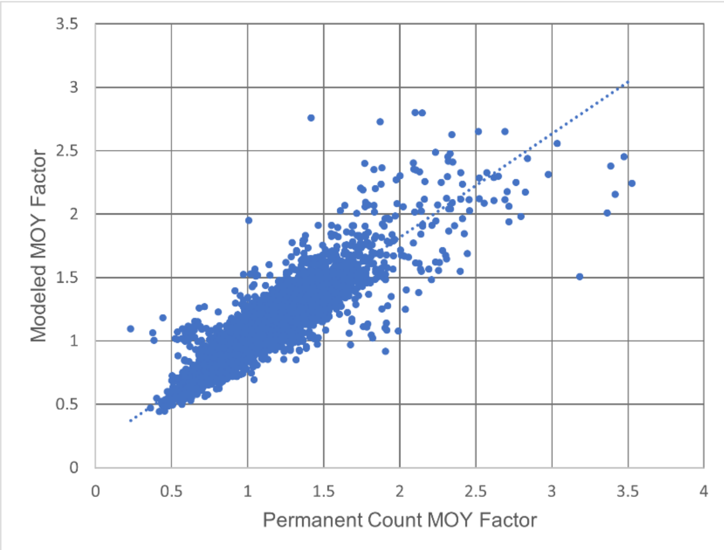
The full reference data of 4,232 permanent counter stations from the lower 48 states across the U.S. were used to train the MOD and DOW factor models. Because the TOD factor models were generated separately for trucks and cars, a subset of data from permanent counter stations was used for training and evaluated the TOD factor model (See Vehicle Classification AADTs for more info). Reference MOD, DOW and TOD factors were generated using the equations above from hourly counts in order to fit a model and test model performance. Sample trip counts at each counter location derived from LBS and GPS data were aggregated across combinations of hour of day, day of week, and month to generate counts for predictive features for fitting the model. Additional contextual predictive features were also considered, as described in Chapter 1. While other models were also considered, Gradient Boosting was chosen to fit the predictive features for its optimal performance. K(10) folds cross validation was used to evaluate performance of each generated model. See Chapters 2 and 3 for more details on building, training, and validating the Probe Data National AADT model.

Estimation of Month of Year Factors

As in the Probe Data National AADT Model, a mix of LBS, GPS, and external factors are key for successful MOD factors. In contrast, road and weather conditions are not strong predictors of the MOY trends. In other words, the rate of trips at a location in the sample -- even if impacted by weather -- was more indicative of changes in volume than the weather on its own.

Figure 2plots the correlation between the permanent-counter-derived and estimated MOY factors for the entire data set. The Pearson correlation between model estimate and permanent counter value is 0.91 ($p<0.001$), suggesting a strong predictive relationship.

Figure 2: Correlation plot between the actual and estimated MOY factors



The following tables illustrate the performance of the model in the estimation of MOD factors, broken down by road size categories (Table 1) and by month (Table 2). In Table 1, it can be seen that the median bias for all road size categories is small across the board. For other metrics, errors are larger for the 0-499 road size category but this is expected because MOY factors are less stable for low volume roads.

Table 1: MOY model error metrics broken down by road size categorization.

Road Size Bin	Median Bias (%)	95% TCE Error Range (%)	68th Absolute Percentile (%)	95th Absolute Percentile (%)	MAPE (%)	NRMSE (%)
0-499	1.05	28.40	9.91	24.51	8.60	13.17
500-4,999	0.20	15.13	4.90	14.41	4.68	8.76
5,000-54,999	0.12	7.74	2.73	7.63	2.67	4.63
55,000+	0.04	6.60	2.02	6.29	2.15	4.08

Examination of model performance by month (Table 2) demonstrates that the model performs exceptionally well in estimating the MOY factors with negligible median bias for all 12 months. The winter months of December, January, and February tend to have slightly higher errors than other months but it’s not substantial. This may be due to higher variance across regions in winter months.

Table 2: MOY model error metrics broken down by calendar month

Month	Median Bias (%)	95% TCE Error Range (%)	68th Absolute Percentile (%)	95th Absolute Percentile (%)	MAPE (%)	NRMSE (%)
1	-0.17	14.06	3.82	12.63	4.04	7.37
2	-0.03	16.41	3.96	13.98	4.38	8.43
3	0.20	11.34	3.14	10.75	3.17	5.89
4	0.17	10.24	2.69	9.38	2.85	6.33
5	0.15	8.89	2.60	8.34	2.61	4.75
6	-0.01	9.80	2.86	9.08	2.83	4.78
7	0.18	10.07	3.06	9.44	3.03	5.15
8	0.12	10.16	2.79	9.17	2.84	4.43
9	-0.18	9.35	2.98	8.94	2.83	4.43
10	0.44	10.20	2.96	9.39	2.87	4.78
11	0.33	10.12	3.10	9.34	3.03	5.66
12	0.23	11.93	3.56	11.57	3.72	7.36

Not only does the model perform well for all road size categories and months, it also estimates the MOY factors very well across the states. Below are results from three example states, North Dakota with a small number of stations (Figure 3a, n=24), Illinois with a bigger sample size (Figure 3b, n=59) and the large state of Ohio (Figure 3c, n=125). In each plot, the blue and orange bands show the respective 95% confidence intervals for the actual and estimated MOY factors, while the dashed lines represent the actual and estimated mean factors for that state. When the blue and orange bands overlap, it indicates that there is no significant difference between the actual and estimated MOY factors. That is the pattern seen for most states in the dataset. In other words, the Probe Data model can capture the MOY factors exceptionally well for most states.

Figure 3a: Time series of MOY factors estimation compared to true values for North Dakota. The blue solid and red grid bands show the range of values within one standard deviation of the mean between MOY factors as derived from permanent counts (red circle, 'actual_value') and from Probe Data (blue square, 'estimated_value').

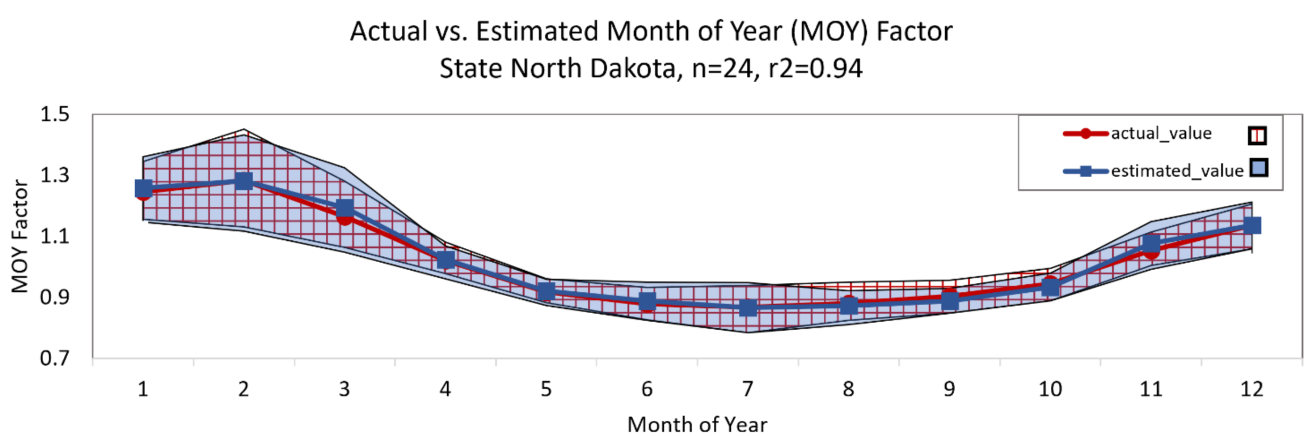


Figure 3b: Time series of MOY factors estimation compared to true values for Illinois. The blue solid and red grid bands show the range of values within one standard deviation of the mean between MOY factors as derived from permanent counts (red circle, 'actual_value') and from Probe Data (blue square, 'estimated_value').

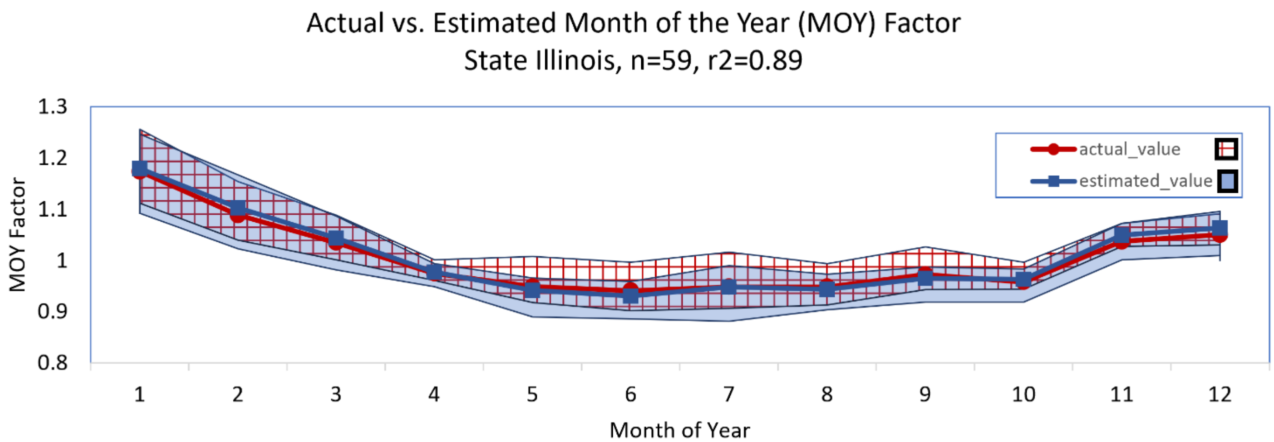
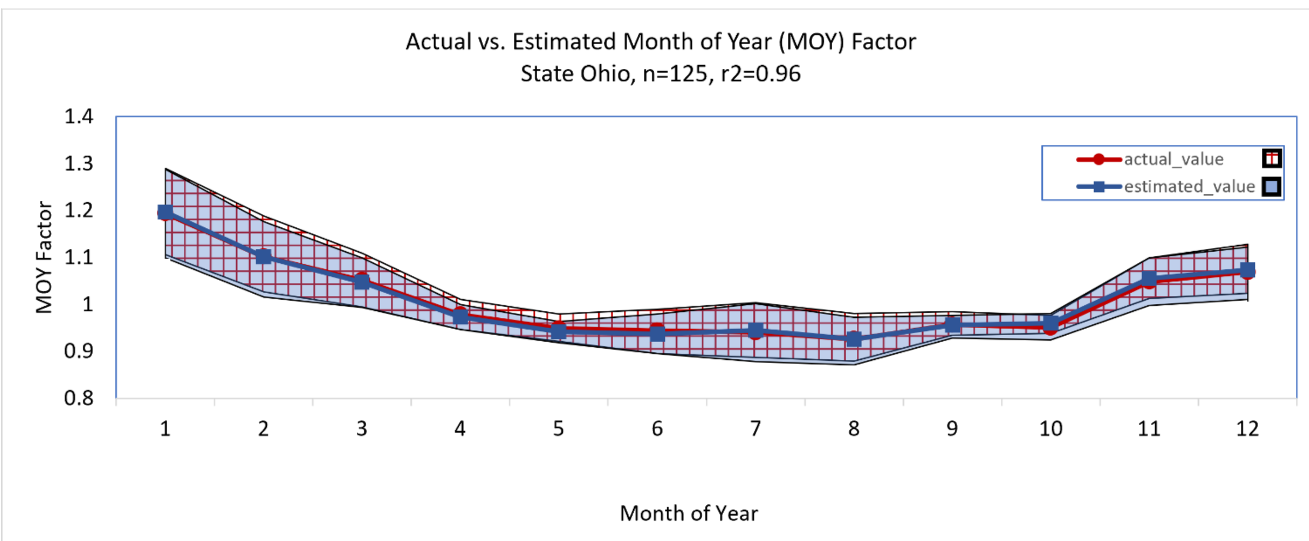


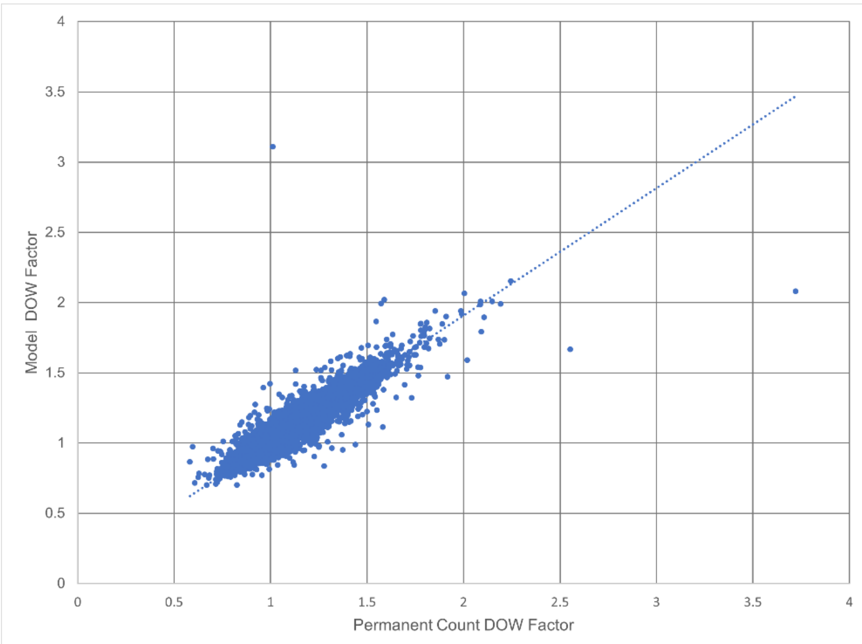
Figure 3c: Time series of MOY factors estimation compared to true values for Ohio. The blue solid and red grid bands show the range of values within one standard deviation of the mean between MOY factors as derived from permanent counts (red circle, 'actual_value') and from Probe Data (blue square, 'estimated_value').



Estimation of Day of Week Factors

The model also estimates the DOW factors very well, with an achieved R-square value of 0.95 (Figure 2).

Figure 2: Correlation plot between the actual and estimated DOW factors



The model performance is again broken down by all road size categories (Table 1) and the seven DOW factors (Table 2). Table 1 shows a similar pattern to Table 1 in the MOY factors subsection, as the median bias is small across the board and the errors for the low volume roads (road size 0-499) are higher than those for higher volume ones. This is expected due to the fluctuation in DOW factors for smaller roads.

Table 1: DOW factor model error metrics comparison across road size categorization

Road Size Bin	Median Bias (%)	95% TCE Error Range (%)	68th Absolute Percentile (%)	95th Absolute Percentile (%)	MAPE (%)	NRMSE (%)
0-499	1.85	27.16	8.17	22.90	8.00	15.89
500-4,999	0.10	11.40	4.03	10.64	3.67	5.69
5,000-54,999	0.11	6.40	2.24	6.28	2.08	3.64
55,000+	0.08	5.25	1.71	5.17	1.65	2.80

Table 2 describes model error across seven days of week. Errors are small across the board, Saturday and Sunday produce higher errors than other days. This is expected as traffic volumes on Saturdays and Sundays may have more variance than other days due to the absence of consistent work commute patterns.

Table 2: DOW factor model error metrics comparison across day of week

Day of Week	Median Bias (%)	95% TCE Error Range (%)	68th Absolute Percentile (%)	95th Absolute Percentile (%)	MAPE (%)	NRMSE (%)
Sunday	0.52	13.79	4.87	12.11	4.38	6.51
Monday	0.30	5.93	1.98	5.60	1.85	2.89
Tuesday	-0.09	7.07	2.21	6.41	2.09	3.24
Wednesday	-0.01	6.53	2.22	6.06	2.04	3.10
Thursday	0.35	5.65	1.96	5.29	1.90	4.80
Friday	0.20	7.01	1.65	5.47	1.71	2.98
Saturday	-0.68	10.14	4.12	9.41	3.54	4.87

In addition to summary metrics, the performance of DOW factors at the state level were also broken down, and show examples for three example states with a variety of reference counters: Massachusetts (Figure 3a), Alabama (Figure 3b), and Florida (Figure 3c).

Figure 3a: Comparison of DOW factors estimation for all roads in Massachusetts. The blue solid and red grid bands show the range of values within one standard deviation of the mean between DOW factors as derived from permanent counts (red circle, 'actual_value') and from Probe Data (blue square, 'estimated_value').

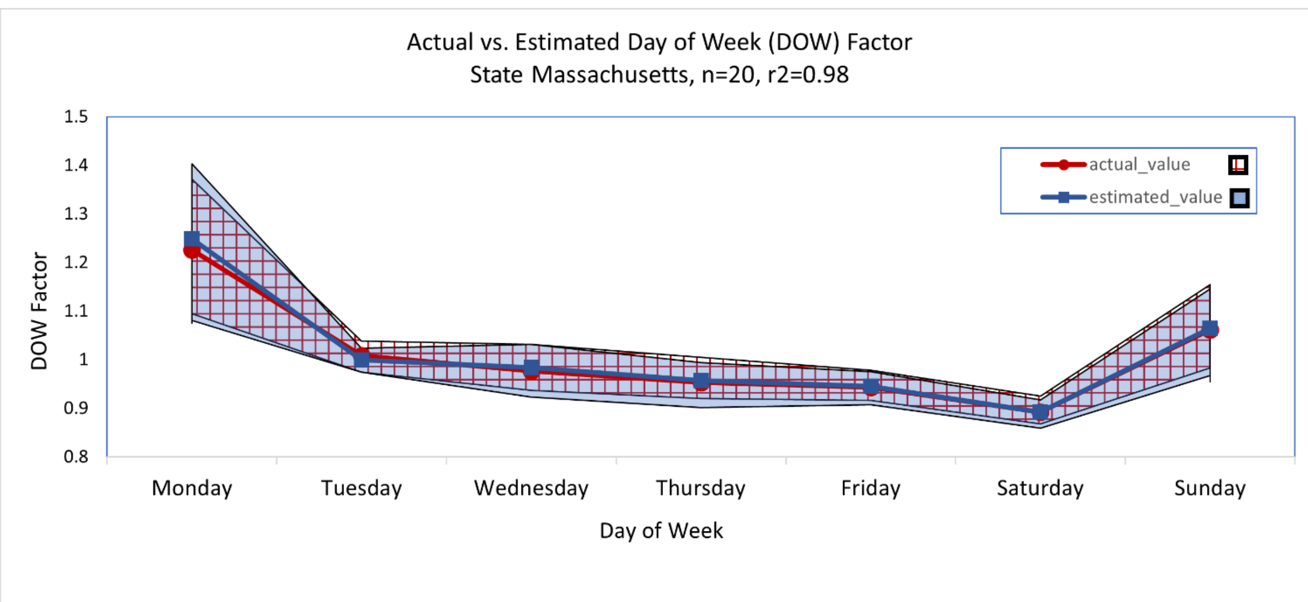


Figure 3b: Comparison of DOW factors estimation for all roads in Alabama. The blue solid and red grid bands show the range of values within one standard deviation of the mean between DOW factors as derived from permanent counts (red circle, 'actual_value') and from Probe Data (blue square, 'estimated_value').

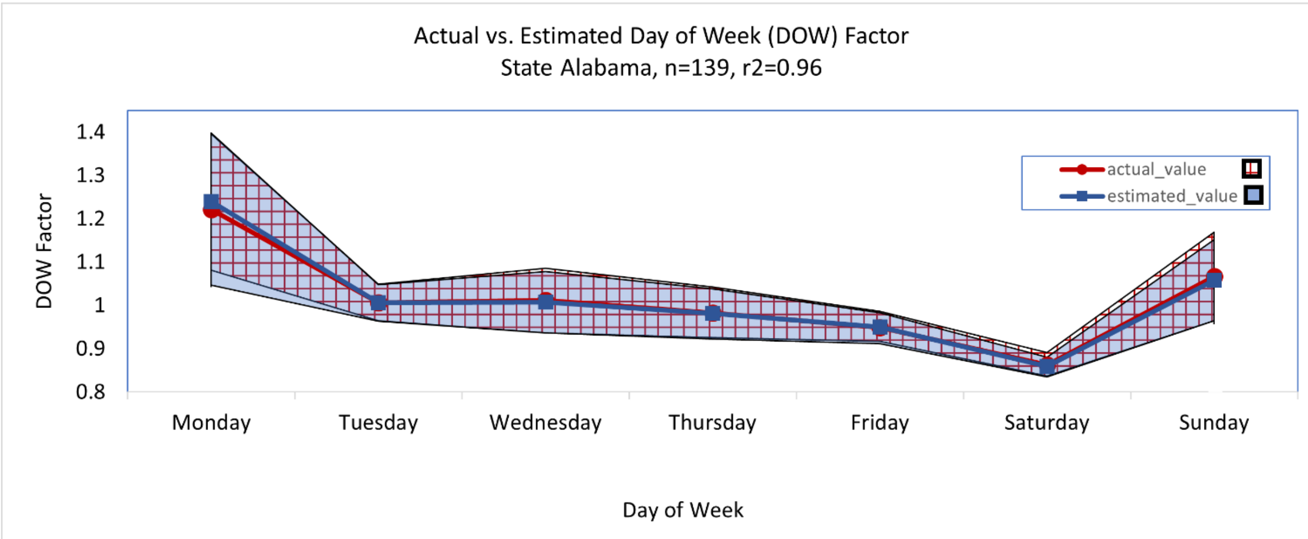
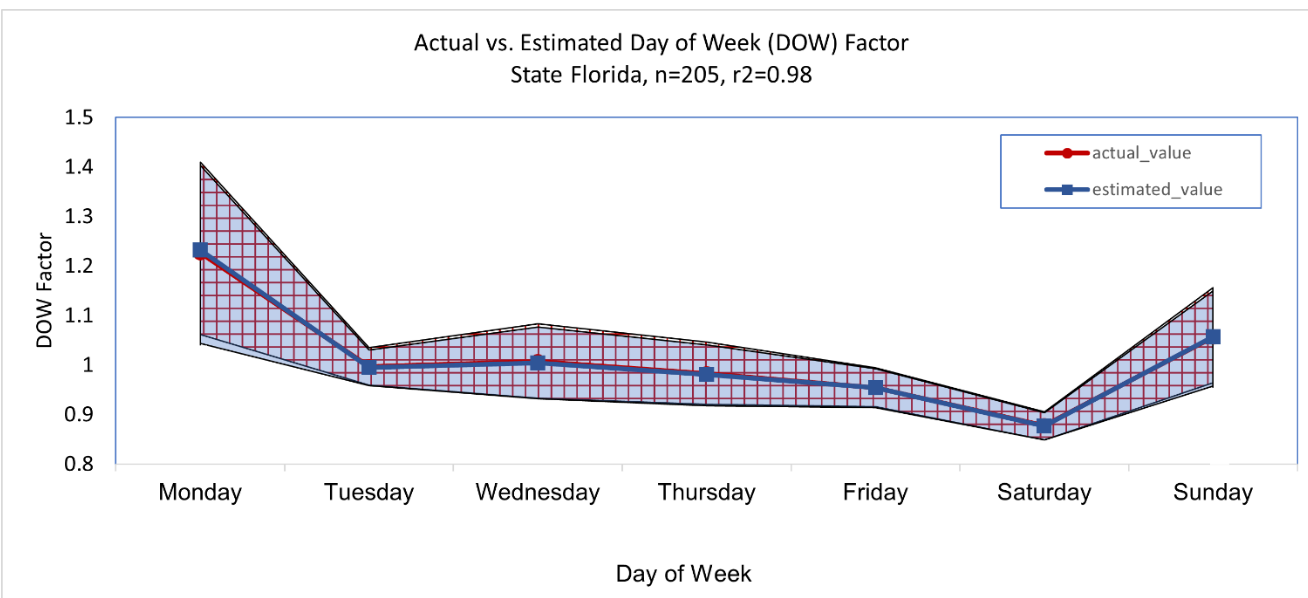


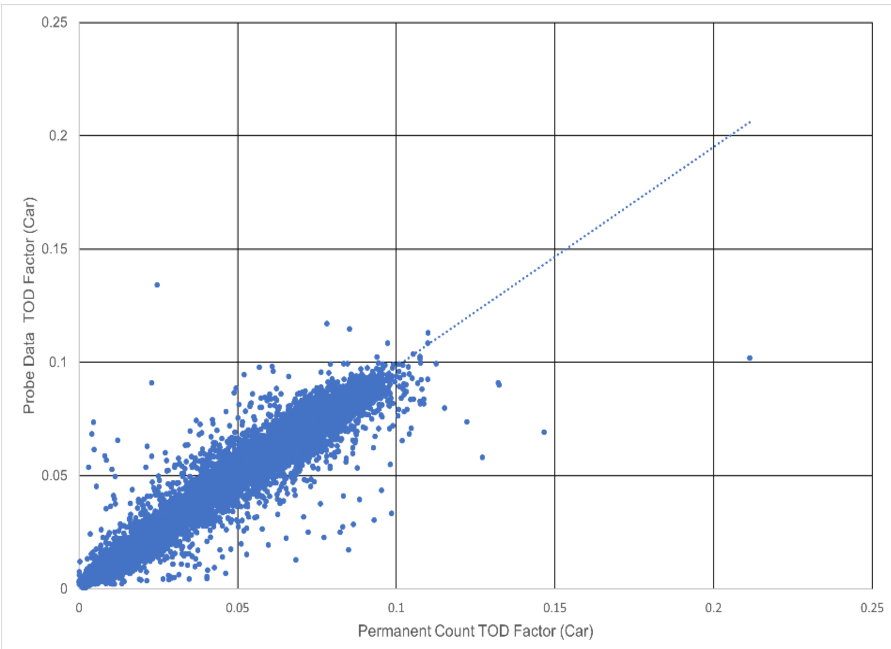
Figure 3c: Comparison of DOW factors estimation for all roads in Florida. The blue solid and red grid bands show the range of values within one standard deviation of the mean between DOW factors as derived from permanent counts (red circle, 'actual_value') and from Probe Data (blue square, 'estimated_value').



Estimation of Time of Day Factors

Time of Day or TOD factors are the ratio of an hourly traffic average divided by the sum of hourly averages for that day on a given roadway segment. Factors are meant to estimate weekday hourly volumes. These factors are often computed separately by vehicle class. For this analysis, TOD estimates will be broken out by cars (vehicle class 1-3) and trucks (vehicle class 4-13), as per AASHTO recommendations.

Figure 2: Correlation plot between the permanent-counter derived and Probe Data-estimated TOD factors for cars



Errors across all road size categories (Table 1) and across the separate 24 TOD factors (Table 2) are described in the tables below. Table 1 shows a similar pattern to Table 1 in the MOY and DOW factors subsections, as the median bias is small across the board and the errors for the low volume roads (road size 0-499) are higher than those for higher volume ones. This is expected due to the fluctuation in TOD factors for smaller roads.

Table 1: TOD factor for cars model error metrics broken down by road size categorization

Road Size Bin	Median Bias (%)	95% TCE Error Range (%)	68th Absolute Percentile (%)	95th Absolute Percentile (%)
0-499	-0.17	209.63	22.17	100.89
500-4,999	0.85	75.05	13.92	52.51
5,000-54,999	0.49	47.99	8.79	36.28
55,000+	0.27	28.14	6.76	23.65

Table 2 indicates that errors are much lower during day time hours, and particularly strong between the important morning peak periods of 7am and 8pm. This is expected since hourly traffic volumes will have more variance during late evening and early morning hours due to smaller sample sizes and less consistent trends, especially relative to peak morning and evening travel patterns.

Table 2: TOD factor model error metrics broken down by hour of day

Hour of Day	Median Bias (%)	95% TCE Error Range (%)	68th Absolute Percentile (%)	95th Absolute Percentile (%)
1	0.42	63.57	18.8	51.08
2	4	90.14	22.83	69.53
3	9.74	123.35	29.88	88.82
4	3.18	107.65	29.38	76.49
5	5.16	122.61	29.69	88.36
6	1.53	103.93	22	67.29
7	-0.15	61.43	14.76	43.82
8	0.02	36.44	10.33	28.88
9	0.41	25.77	7.3	20.97
10	0.2	18.79	6.17	16.41
11	0.1	16.84	6.21	14.86
12	0.18	15.21	5.7	13.91
13	0.21	13.09	5.25	12.61
14	0.33	13.14	5.19	12.59
15	-0.09	14.51	5.02	13.4
16	-0.16	15.34	4.65	12.77
17	-0.17	12.81	4.49	12.67
18	-0.17	17.48	5.54	15.7
19	0.17	20.13	6.04	18.15
20	0.44	23.57	7.39	20.79
21	0.6	26.56	8.2	22.45
22	0.83	31.06	9.77	26.17
23	1.27	43.42	12.04	33.75
24	2.61	54.64	15.2	42.79

In addition to summary metrics, it’s useful to visualize performance of TOD factors across roads of different sizes. Figures 3a, 3b, 3c, 3d shows the plots of actual vs estimated values for the TOD factors across the four road size categories based on AADT. The shading within the plots highlight the range of values within one standard deviation of the mean, for both the TOD factors and the TOD factor estimates by hour. For roads with AADTs greater than 500, the bands almost completely overlap between estimated and true TOD factor. In the smallest road size category (0-499 AADT) the model slightly under predicts the AM peak. As seen in the wide range of the standard deviation of the TOD factors around that time, the signal has high variation across all observed locations and is more difficult to estimate.

Figure 3a: Comparison of car TOD factor estimation vs permanent counter-derived value across all hours of day for very small volume roads (AADT 0-499). The blue solid and red grid bands within the plot highlights the range of values within one standard deviation of the mean, for TOD factors derived from permanent counts (red circle, tod_factor_car) and from Probe Data (blue square, estimated_value), by hour.

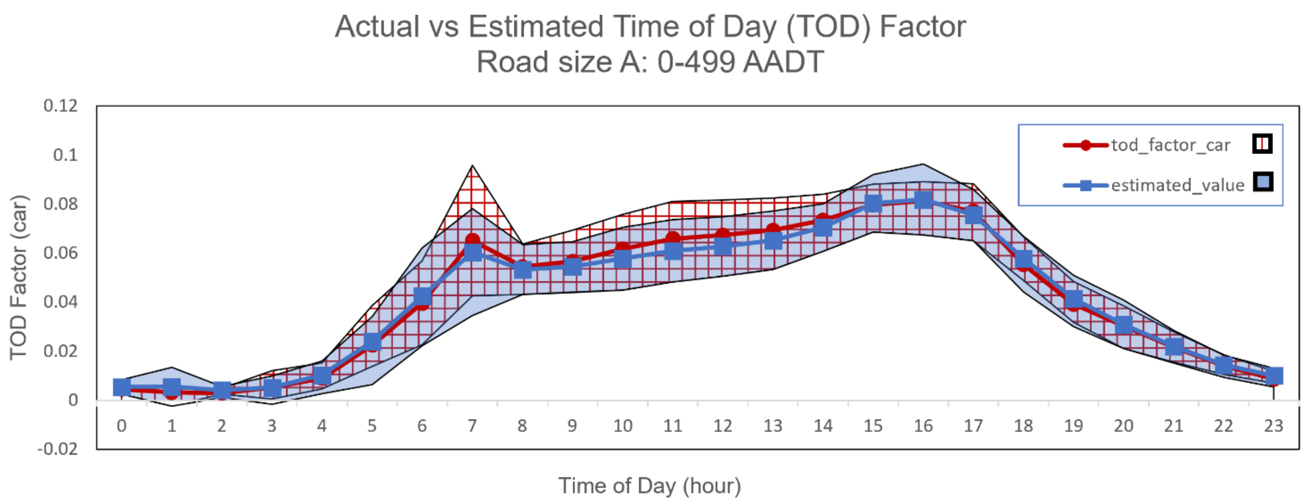


Figure 3b: Comparison of car TOD factor estimation vs permanent count-derived value across all hours of day for small volume roads (AADT 500-4999). The blue solid and red grid bands within the plot highlights the range of values within one standard deviation of the mean, for TOD factors derived from permanent counts (red circle, tod_factor_car) and from Probe Data (blue square, estimated_value), by hour.

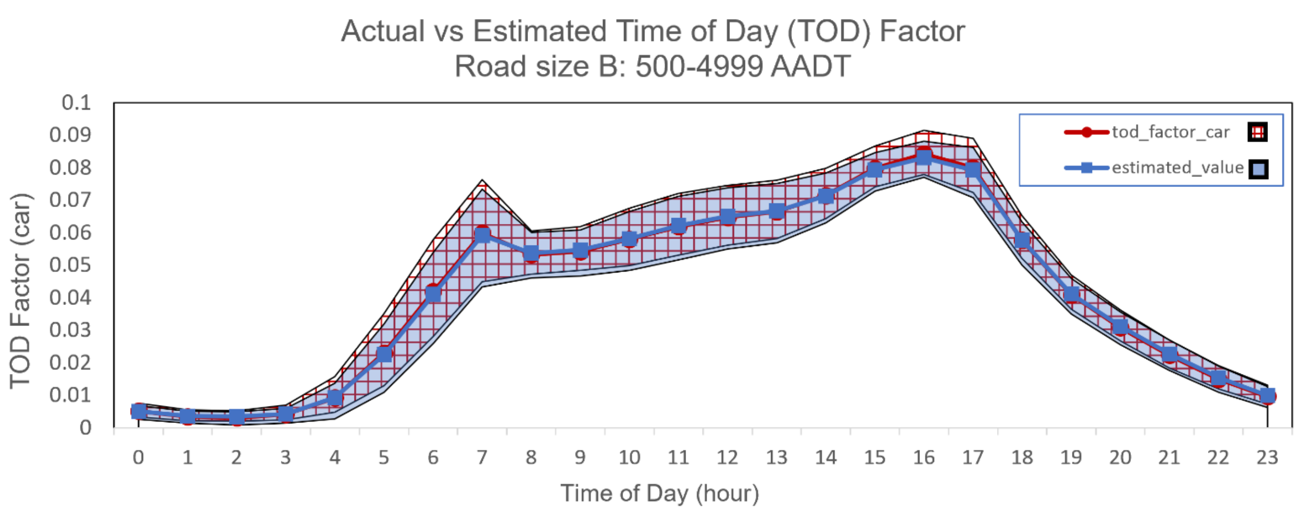


Figure 3c: Comparison of car TOD factor estimation vs permanent count-derived value across all hours of day for medium volume roads (AADT 5000-54999). The blue solid and red grid bands within the plot highlights the range of values within one standard deviation of the mean, for TOD factors derived from permanent counts (red circle, tod_factor_car) and from Probe Data (blue square, estimated_value), by hour.

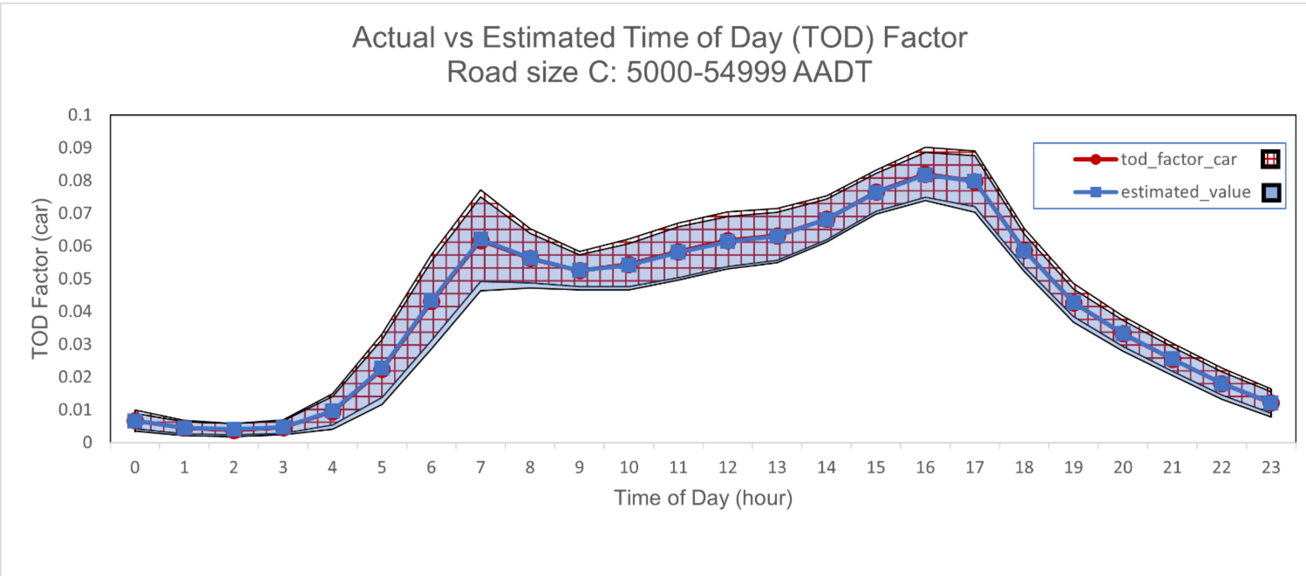


Figure 3d: Comparison of car TOD factor estimation vs true value across all hours of day for high volume roads (AADT 55000+). The blue solid and red grid bands within the plot highlights the range of values within one standard deviation of the mean for TOD factors derived from permanent counts (red circle, tod_factor_car) and from Probe Data (blue square, estimated_value), by hour.

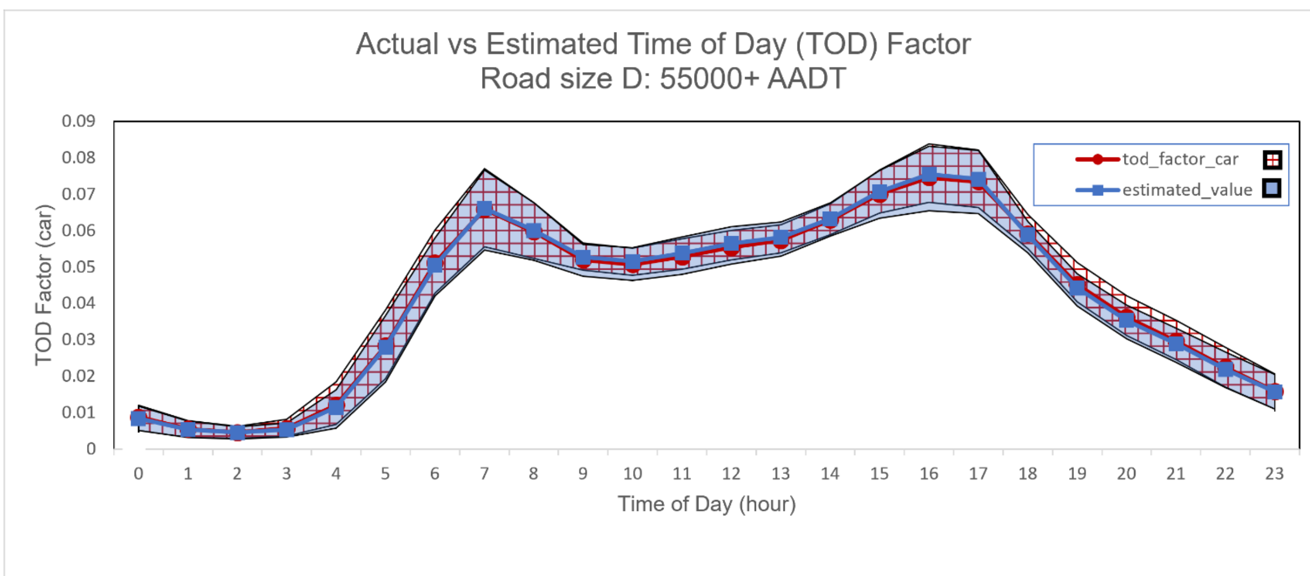
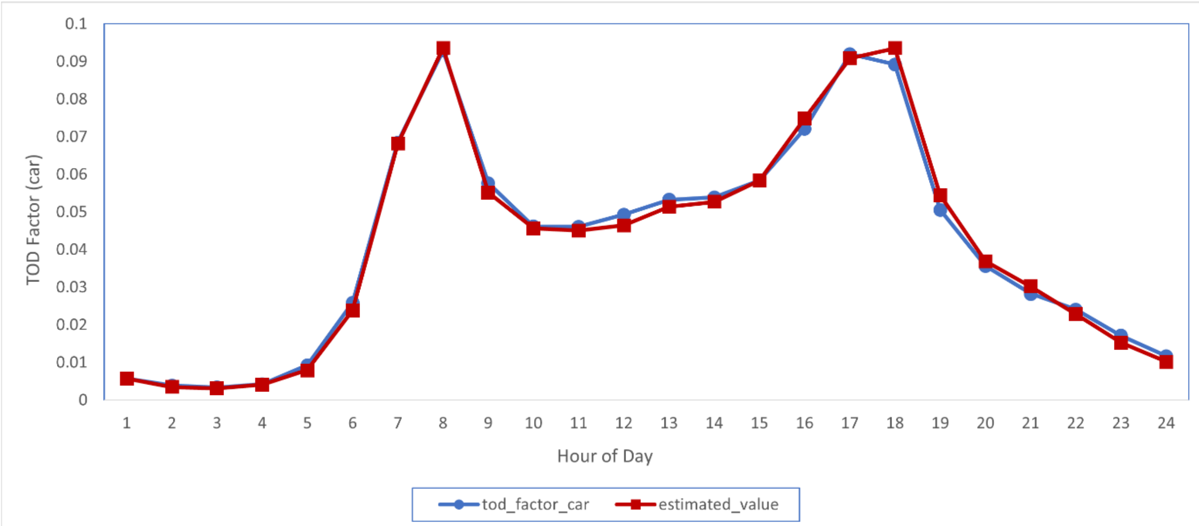


Figure 4 depicts a time series of the time of day factors for one example station in Alabama. You can see the close relationship between permanent count-derived and estimated values.

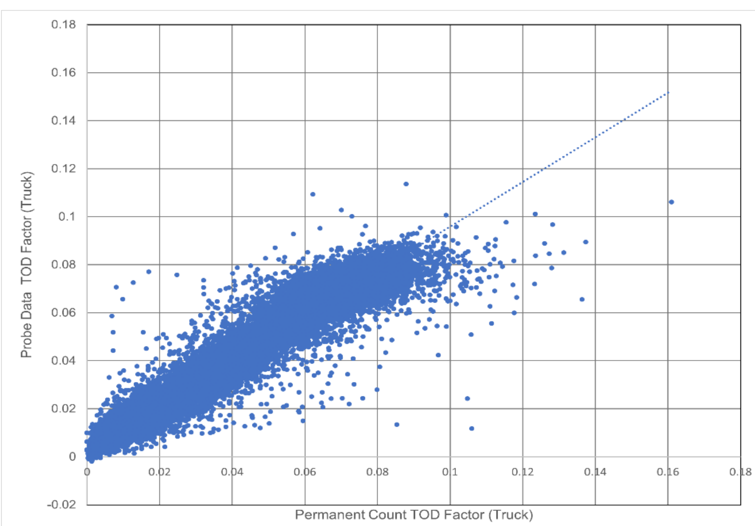
Figure 4: Time series of car TOD factor estimation (square, 'estimated_value') for an individual station, compared to TOD factor as derived from a permanent count (circle, 'tod_factor_car')



In addition to TOD factors for cars, a separate model was created in order to estimate TOD factors for trucks. Unsurprisingly, inclusion of data from a source explicitly containing heavy and medium-duty vehicles proved critical.

Figure 6 contains the correlation plot between the permanent count-derived and estimated TOD factors for trucks. A Pearson-r correlation of 0.96 (p value < 0.001) suggests that Probe Data can present a strong predictor for TOD factors for trucks, as well.

Figure 6: Correlation plot between the actual and estimated TOD factors for trucks



The truck model performance is again broken down by all road size categories (Table 3) and the 24 TOD factors (Table 4). Road size categories are assigned based on the total volume of trucks on the road, not the total volume of vehicles. As a result, there are no results for the 55,000+ category.

Table 3 shows a similar pattern as other road size breakdowns, as the median bias is small across the board and the errors for the low volume roads (road size 0-499) are higher than those for higher volume ones. This is expected due to the fluctuation in TOD factors for smaller roads.

Table 3: Comparison of model error for truck TOD factor across road size categories

Road Size Bin*	Median Bias (%)	95% TCE Error Range (%)	68th Absolute Percentile (%)	95th Absolute Percentile (%)
0-499	1.8	202.66	26.51	134.51
500-4,999	0.68	73.04	15.34	51.54
5,000-54,999	1.09	26.35	10.87	25.89

*Note that the "Road Size Bin" column represents the annual average daily volume of CUs in permanent station data.

Errors are slightly higher for trucks than for cars. This is expected due to the fact that trucks make up a smaller proportion of vehicles on the road and are thus likely to have a more variable sample size. Table 4 indicates that errors are much lower during day time hours, and particularly low in the mid-day hours. As with the TOD model for cars, this is expected since hourly traffic volumes will have more variance during late evening/early morning hours due to smaller sample sizes and less consistent trends.

Table 4: Model error for truck TOD factor estimation across hour of day

Hour of Day	Median Bias (%)	95% TCE Error Range (%)	68th Absolute Percentile (%)	95th Absolute Percentile (%)
1	3.41	192.68	34.49	141.42
2	3.11	247.15	34.73	164.37
3	2.6	249.04	31.83	162.69
4	2.06	260.95	28.88	143.82
5	2.18	194.95	27.63	105.19
6	1.43	103.56	20.3	61.67
7	0.48	64.6	17.06	44.43
8	2.44	41.33	14.12	34.42
9	1.13	29.91	12.25	26.81
10	0.91	26.44	10.36	22.49
11	0.68	24.72	9.09	21.75
12	0.31	20.6	8.55	19.19
13	0.57	22.05	7.92	19.08
14	0.16	20.69	7.68	18.77
15	1.06	20.31	8.22	19.52

Hour of Day	Median Bias (%)	95% TCE Error Range (%)	68th Absolute Percentile (%)	95th Absolute Percentile (%)
16	1.72	23.95	8.72	23.99
17	0.47	36.65	10.52	29.44
18	-0.36	51.67	13.79	40.27
19	-0.42	69.2	15.36	50.19
20	-0.05	91.71	19.93	65.7
21	-0.32	112.57	23.92	79.53
22	1.43	126.33	27.05	89.55
23	3.06	143.16	29.56	105.28
24	1.77	183.1	31.78	125

In addition to summary metrics, it's useful to visualize performance of TOD factors across roads of different sizes. Figures 7a, 7b, 7c show the plots of actual vs estimated values for truck TOD factors across three road size categories displayed earlier. The shading within the plots highlight the range of values within one standard deviation of the mean, for both the true and estimated TOD factors by hour. It can be seen that Probe Data can capture the DOW trends very well, especially across higher volume roads. In the smallest road size category (0-499 vehicles) there is increased variation in the TOD factors, making it more difficult for the model to estimate. For roads with AADTs under 500, roads with a strong evening peak tended to be slightly underpredicted (Figure 7a). Across all road volumes, the orange band of estimated values tends to be slightly narrower than the true TOD factors, suggesting that the model may not predict locations that strongly deviate from the average well.

Figure 7a: Time series of estimated TOD factor for trucks compared to permanent count-derived TOD factors for road size A. The blue solid and red grid bands within the plot highlights the range of values within one standard deviation of the mean, for both the TOD factors derived from permanent counts (red circle, 'trod_factor_truck') and estimated TOD factors (blue square, estimated_value) by hour.

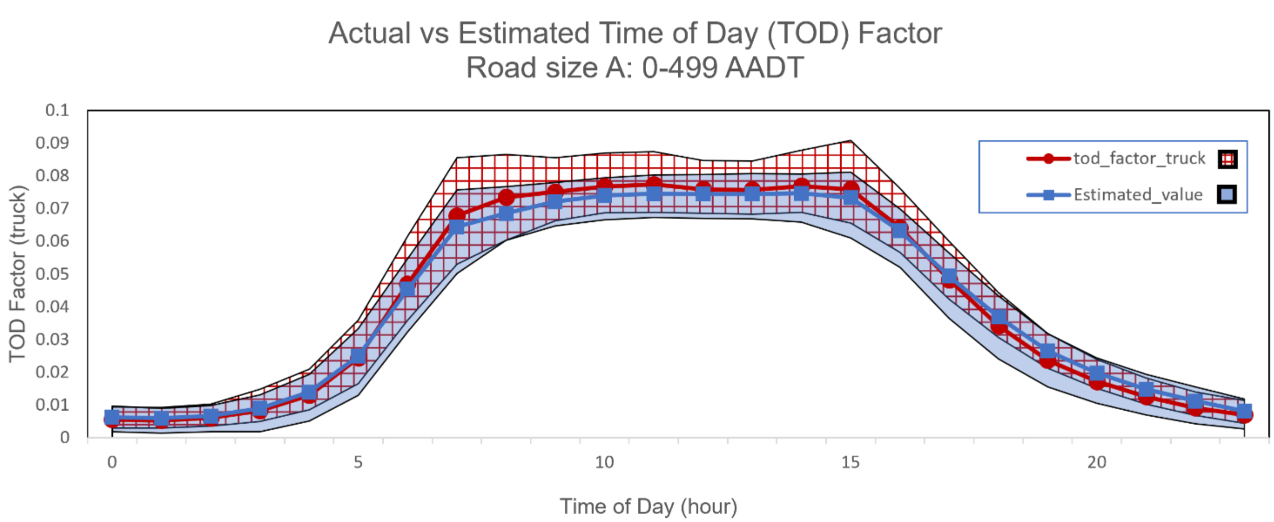


Figure 7b: Time series of estimated TOD factor for trucks compared to permanent count-derived *TOD factors for road size B. The blue solid and red grid bands within the plot highlights the range of values within one standard deviation of the mean, for both the TOD factors derived from permanent counts (red circle, 'trod_factor_truck') and estimated TOD factors (blue square, estimated_value) by hour.*

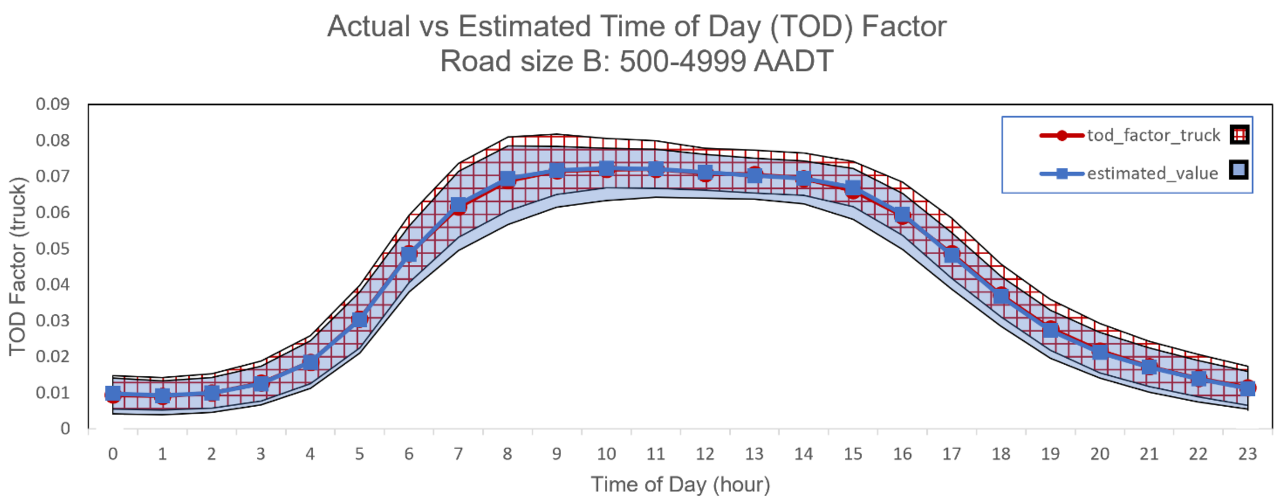
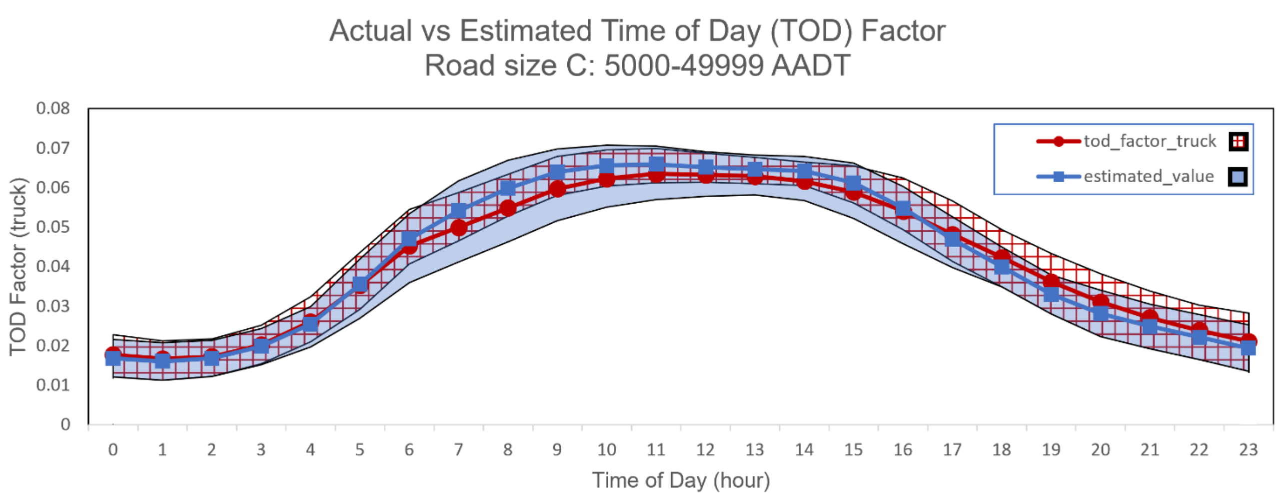
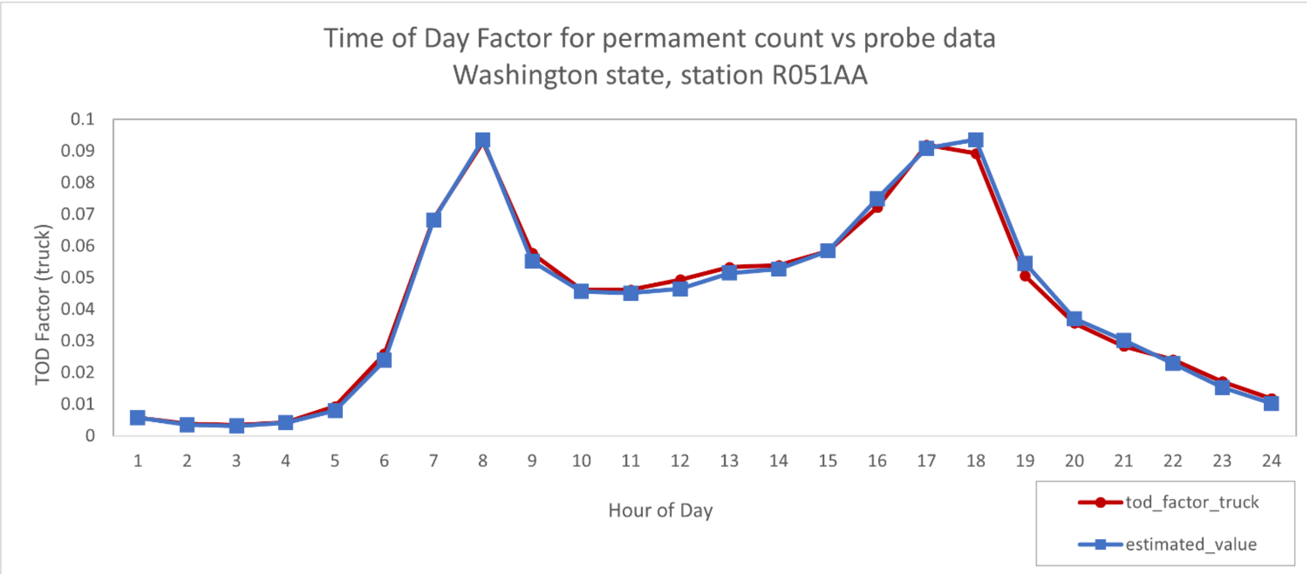


Figure 7c: Time series of estimated TOD factor for trucks compared to permanent count-derived *TOD factors for road size C. The blue solid and red grid bands within the plot highlights the range of values within one standard deviation of the mean, for both the TOD factors derived from permanent counts (red circle, 'trod_factor_truck') and estimated TOD factors (blue square, estimated_value) by hour.*



Individual locations can also be visualized in order to evaluate the performance of the TOD model for trucks. As seen in the figure below, even when analyzing an individual station, the model performs well, capturing the weekday hourly curve.

Figure 8: Time series of truck TOD factor estimation (square, 'estimated_value') for an individual station, compared to TOD factor as derived from a permanent count (circle, 'tod_factor_truck')



Probe Data can be used to model MOY, DOW factors for all road sizes, months and days, but also for different states. This is extremely helpful for DOTs to annualize short term duration counts when there are no permanent counters nearby, with similar characteristics, to provide the data for the calculation of the MOY and DOW factors.

Calculation of factors with Probe Data has many of the same caveats as annualization from short term duration counts. Both methods have higher error with smaller roads, and are sensitive to the availability of nearby permanent counters for calibration. However, an advantage of Probe Data is that unlike annualization from short term duration counts (FHWA, 2018), factors derived from Probe Data can be applicable to atypical traffic patterns. With Probe Data, sampled trips are available at the exact location of interest, and thus models are sensitive to and can estimate volume given abnormal traffic patterns that may be unique to the road (such as a road detour, or spike in traffic due to an event). See sections on 'MADT Estimation', as well as "ADT" and "Hourly Volume" for more detail and examples.

This section specifically sought to investigate if Probe Data could be used to generate the HOD, DOW, and MOY factors. However, if the goal of those factors is to generate estimates of MADT or AADT, those estimates can be directly derived from a Probe Data model itself, rather than backing out from MOY, DOW factors. See sections on 'MADT', 'ADT' and 'Hourly volume' for more detail on model performance.

Estimation of K Factor

K-factor is expressed as the Design Hour Volume (DHV) as a proportion of AADT for a given facility. The DHV is frequently selected as the 30th highest hourly volume of the year.

$$K\text{-factor} = \frac{30th\ highest\ volume}{AADT} \times 100$$

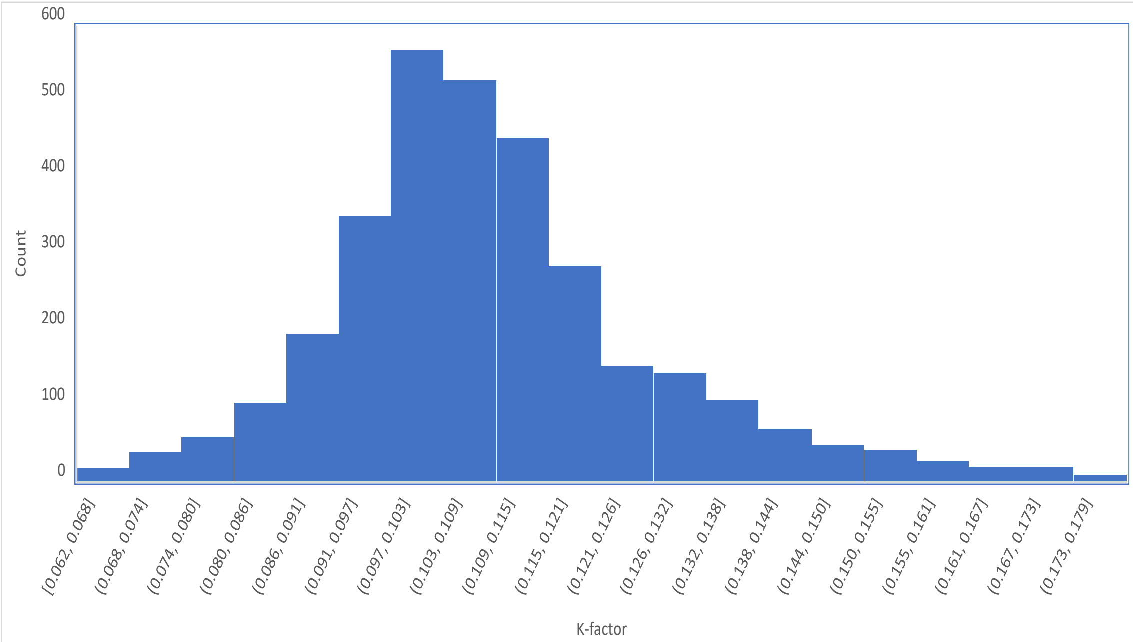
The purpose of the DHV is to balance the desire to provide an adequate level of service (LOS) for the peak hour traffic volume with proposing a design in which the highway capacity would only be utilized for a few hours of the year. K-factor is used in pavement design, geometric design (e.g., number of lanes needed), capacity analysis, estimation of volume-to capacity ratios and levels of service, functional classification of roads, and analysis of traffic operations (e.g., effect of lane closures).

K-factor values usually range from 7-18% depending on whether a facility is in an urban, suburban or rural area. K-factors generally decrease as AADT increases or as developmental density increases. Therefore, K-factors are often lowest for urban facilities, and highest for recreational facilities, followed by rural facilities (Thomason, 2020). K-factors are also influenced by traffic flow patterns, road geometry and the location of the facility.

Methods

For this evaluation, Probe Data was used to model K-factor values, specifically K-30, which represents the K-factor based on the 30th highest volume of the year. From the collection of permanent counters (see Methods for National Probe Data AADT Model), 3,310 permanent counters met the criteria for calculating K-Factor for 2019. The U.S. Counters were required to be bi-directional and have a high daily observation rate, counting at least 60% of 365 days in the year. Research indicates that K-factors are expected to range from 7 to 18%, thus outlier K-factors were removed from the sample due to concerns about reporting errors. A 20% buffer was added to the expected range, thus facilities with K-factors below 5% or above 22% were excluded. The distribution of K-factor values among the training data set of permanent counters is detailed in table K1. The median K-factor was 10, with the range from the 10th to 90th percentile over all counters was 8 to 13. As seen in the distribution, the values were skewed toward smaller K-factors, with very few example stations in the higher tail (K-factors over 13).

Table K1. Distribution of K-Factor30 values of the 3,310 permanent counters used to generate an estimated K-factor.



In order to model the K-30 value, predictive features were collected for each facility. For each station, the same contextual features were used as in the National AADT Model: historical weather information, population statistics as derived from the census, information about the roadway, as derived from Open Street Maps, and trip counts derived from LBS and GPS data sources (see section on National Probe Data AADT Model I for more detail).

LBS and GPS trip counts were aggregated at each location in a variety of methods to see how the optimal signal of K-factor might be derived. The 30th highest hour for 2019 was taken, but additionally, the trip count at each road was aggregated into hourly counts by month, quarter and year. Raw trip count from Probe Data can be noisy, but aggregation and smoothing of the sample over time can smooth out the noise, allowing trends to be captured. All aggregated counts were then divided by the annual average daily Probe Data trip count for the facility.

These aggregated counts of LBS and GPS trips were labeled as such:

- **Def:** No scaling (by definition): the ratio of 30th highest hourly Probe Data trip count out of 8,760 hours in a year, over the annual average daily Probe Data trip count
- **MWH:** K-factor derived from monthly-weekly-hourly time series of aggregated Probe Data trip count
- **QWH:** K-factor derived from quarterly-weekly-hourly time series
- **YWH**: K-factor derived from yearly-weekly-hourly time series

Of the gathered reference values from permanent counters, 90% of roads had K-factors under 13 (Table K1). This is of note, because it means there were only a very few number of locations that could be used to fit a model for high K-factors. The predictive features, as described above, were fit to a Random Forest model, which was chosen for its accuracy and computational efficiency. Cross validation was used to estimate model error and derive accuracy metrics. See methods for the National Probe Data AADT Model for more detail.

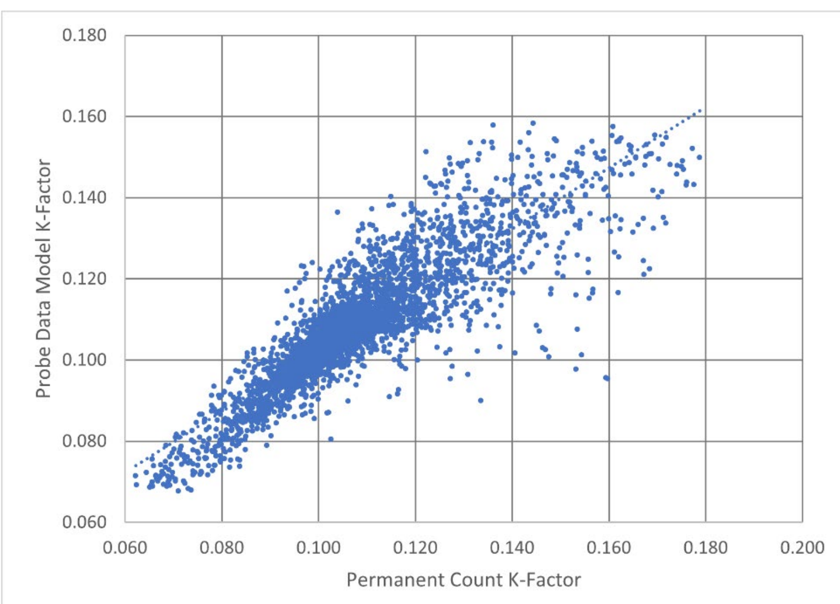
Table K1. Breakdown of K-factor 30 values over the 3,310 permanent counters used to create a K-factor 30 model

Percentile	K-factor 30 Value
10^{th}	8
25^{th}	9
50^{th}	10
75^{th}	11
90^{th}	13

Accuracy and Precision of K-factor Estimation

There was a strong correlation between the aggregated “K-factor” derived from LBS counts, and K-factor as derived from permanent counters; this was the strongest input to the fitted Probe Data model. This suggests that a simple K-factor model derived only from aggregated LBS counts may be feasible, if lower accuracy, particularly for smaller roads, is permitted. For the fitted Gradient Boosted model, there was a strong correlation between the estimated K-Factor and K-factor as derived from permanent counters (Figure K1). There is a trend toward underestimation for lower K-factor values.

Figure K1: Correlation between estimated K-factor vs permanent count derived K-factor values.



When evaluating model error across road volume, K-factor estimation errors increased as road size decreased (Table K2). Overall, there is higher variance as K-30 becomes larger, thus as K-factor increases, the error increases. It should be noted that as K factors increase, roads also become more rural with smaller sample sizes (Thomason, 2020). These facilities are inherently difficult to model with Probe Data or any sampling technique, because of the low traffic sample size. Additionally, in the reference counter set for fitting the Probe Data model, only 10% of locations had K-factor values above 13 (Table K2). The efficacy of the model to fit higher K-factor values was also constrained by the fact that there were few observed locations with high K-factor values to learn from. Future studies are needed to further tune a model and investigate if more precise estimates for higher K values are possible with Probe Data.

Table K2: Comparison of K-factor model estimation error by road AADT

Road Size Bin	Total sites	Median Bias (%)	68th Abs Percentile (%)	95th Abs. Percentile (%)
NA: <= 500	36	0.77	9.56	16.33
small: 500 – 4999	753	1.68	8.25	16.43
medium: 5000-54999	1946	0.9	5.54	14.55
high: > 55000	506	0.04	5.57	14.55

Our work was in no means comprehensive of all the possible approaches by which Probe Data could estimate K-factor. One method was used as a proof of concept for how well Probe Data estimation could be for deriving traffic factors. As an example of an alternative modeling approach, an hourly volume model could be used. This model is expensive to build in terms of computational power and working hours, but could serve as a basis for an improved model for K-Factor with Probe Data. Volume for every hour of the year could first be generated at a location, and K-factor could then be derived from the hourly modeled estimates. This method, for example, may achieve lower error than the method outlined in this paper. These are presented as just one of the many refinements that could be possible to improve the model presented here.

Along with general alternative methods that could be further explored to model K-factor with LBS data, more focused work is needed to improve model accuracy for high K-factors. While the model could estimate lower K-factors with minimal error, it struggled to estimate K-factors above 0.175. In order to improve this result going forward, more training data would be required for facilities with high K-factors. Incorporating additional features that are more sensitive to seasonal changes in the road may also improve performance.

Estimation of D Factor

D-factor is expressed as the Directional Design Hour Volume (DDHV) as a proportion of AADT in the peak hour (design hour) in the predominant direction of flow for a given facility. DDHV is determined from field measurements on the facility under consideration or on parallel and similar facilities. It is given by multiplying AADT by K-factor and D-factor: DDHV = AADT * K * D.

The DDHV is frequently selected as the 30th highest hourly volume of the year. The D-factor is usually expressed as a percentage, and represents the directional distribution of hourly traffic volumes. Below is the definition of D-30.

$$D\text{ - }Factor = \frac{30th\ highest\ volume\ in\ direction}{Volume\ in\ both\ directions}\ x\ 100$$

Directional distribution of traffic may significantly affect the level of service (LOS) of a facility. Therefore, D-factor plays an important role in highway design by considering the directional split of traffic, especially for two-lane rural highways. For this evaluation, Probe Data was used to model D-factor values, specifically D-30, which would represent the D-factor based on the 30th highest volume of the year.

Methods

To create a model for D-factor (30), 3,325 permanent counters from the FHWA TMAS traffic volume data from 2019 were used to fit a model. Counter data were required to be bi-directional, and have a daily observation rate of at least 60%. The resulting permanent counters were heavily skewed toward low D-factors (Figure D1), with 90% percent of the values below 62 (Table D1). Examples of higher D-factor for building a model were limited, which impacts See National AADT methods for more detail on reference permanent counters, and constructed LBS, GPS (personal and commercial) trips from Probe Data inputs.

Figure D1. Histogram of D-Factor 30 values over the 3,325 permanent counters used to create a D-Factor 30 model.

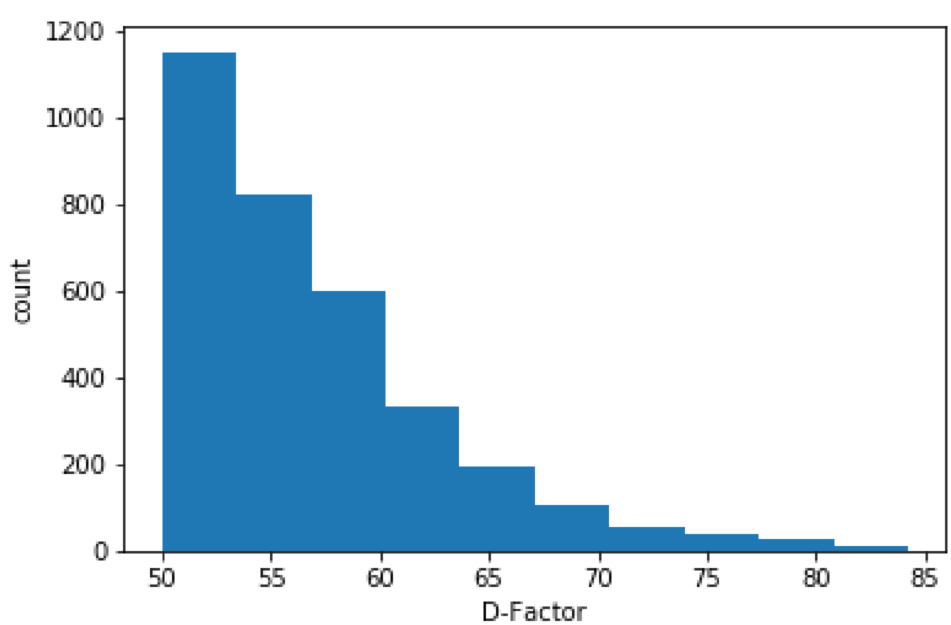


Table D1. Percentile breakdown of D-factor 30 values over the 3,325 permanent counters used to create a D-factor 30 model

Percentile	D-Factor 30 Value
10th	8
25th	52
50th	55
75th	6
90th	62

In order to model the D-factor value, predictive features were collected for each facility. Features included the LBS and GPS (personal and commercial) data collected for the road. The raw trip counts at each location were aggregated in a variety of ways, in order to examine how to best pick up the intended signal the D-factor was designed to represent. Total hourly LBS trip counts for each day of week were aggregated across months, quarters and years.

The following scenarios were tested:

- **Def:** No scaling (by definition): trip count proportion in the 30th highest hourly trip counts out of 8,760 (Percentile = 1 - 30/8760 = 99.66%) travelling in the peak direction.
- **MWH:** Scaled into monthly-weekly-hourly time series
- **QWH:** Scaled into quarterly-weekly-hourly time series
- **YWH**: Scaled into yearly-weekly-hourly time series

Substantial filtering of the LBS data was done, so that only time periods with strong signal for both directions among roads were used. Contextual features were also included for each facility. These contextual features take into consideration the location of the facility, including local weather conditions and census data. Model fit could be substantially improved if the time series was aggregated at the quarterly, instead of yearly level.

Summary of Results:

In modeling the D-factor, three approaches were explored.

- **LBS:** The raw D-factor computed as the percent of LBS trips in the ‘majority’ direction during the ‘busiest’ aggregated hour was used.
- **Linear Fit**: The new features were adjusted for a linear fit.
- **Gradient Boosting**: This is a hybrid model where the new engineered features were combined with the original K/D factor features.

After evaluation, the gradient boosting model showed the most favorable results. LBS data alone showed a strong basic signal, but did not perform as well as gradient boosting, while the linear fit improved the correlation, but lost the sensitivity to pick up high D-factor roads.

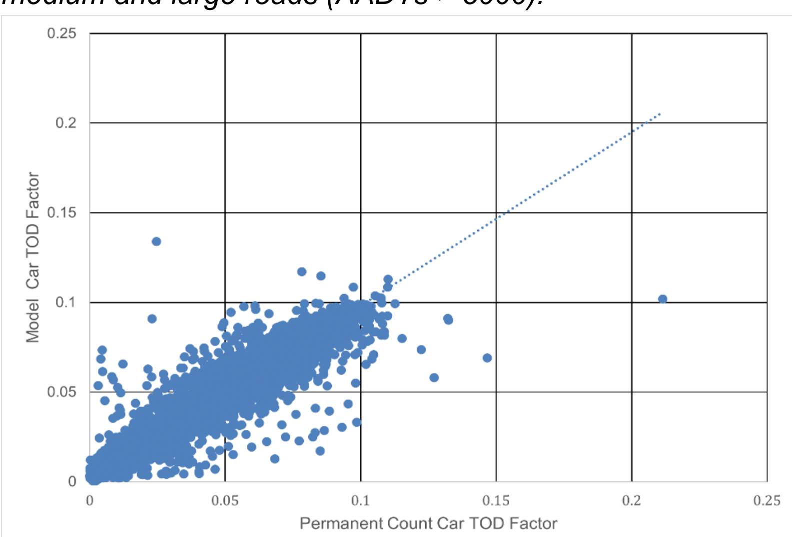
Cross validation results showed errors increasing as road sizes decreased (Table 1).

Table 1: Modeled D-factor absolute percent error across a range of volume road sizes

Road Size Bin	Stations	Median (bias)	68th Abs Percentile (%)	95th Abs Percentile (%)
<= 500	38	6.22	14.51	27.51
500 – 4999 (small)	811	2.04	10.4	20.05
5000-54999 (medium)	1961	1.17	7.09	14.94
> 55000 (large)	520	0.99	5.42	12.43

The Pearson-r correlation aggregated LBS data to true D-30 shows a statistically significant Pearson correlation of 0.51 (p-value < 0.001). This suggests that a predictor for D-Factor can be created from Probe Data alone, without the use of additional contextual factors to improve the model. Fitting the LBS ratio data to a gradient boosted model improves the Pearson correlation across all sites slightly, to 0.54 (p< 0.001). If the LBS is filtered for locations where there are sample trips in both directions for the full day of week, hour of day time series at the quarterly aggregation level, the Pearson correlation increases to 0.68 (p-value <=0.001). However, this filter excludes all roads with AADT of under 7000 AADT (H, I, J). The performance of the D-Factor Probe Data model is presented in Figure 2 for all medium and large roads (AADTs > 5000). As apparent in the scatterplot, D-Factors near 50 were never predicted, and high values of D-factor were systematically under predicted as well. This finding emphasizes that a sufficient sample trip size for both directions on a road is needed to predict D-Factor on lower volume roads.

Figure 2: Correlation between Probe Data modeled D-factor and D-factor (permanent counter) for medium and large roads (AADTs > 5000).



Conclusions

Probe Data can be used to infer a D-factor estimate, especially for medium and larger roads. However, the errors and model fit have high error as compared to K-factor. A main limitation of this study was the lack of training data for large D-factors (usually found on smaller roads). In thedataset used for this study, half of the counters had D-factors from a range of 52 to 60 (25th to 50th percentile). Only 10% of the data had D-factors over 65. The lack of reference data to create and train a model makes it difficult to create and evaluate a model to correctly predict the larger values (D-factors above 65). Additionally, D-factor involves predicting a very small range (50 percent to 80 percent), which means that a model estimates can only have a very small error (less than 30 percent) to pick up on differences.

An accurate calculation of D-factor relies on a robust Probe Data trip sample at the hourly level for both directions of traffic. Additionally, D-factor itself may have some instability as a metric, making it harder to estimate with Probe Data. There are potentially other metrics that Probe Data may be better positioned to address which would still get at the goal of a D-factor, such as comparing direct estimations of hourly volume on a road. See the section on ‘Hourly Volume’ for more detail on the potential for Probe Data to model hourly volumes.

Our approach sought to model a D-factor by looking at aggregated Probe Data trips across a whole year, in order to assess how easily Probe Data might generate a factor. However, there are many possible creative ways to model D-factor not explored in this report which could potentially substantially improve estimates. On the extreme end of complexity, Probe Data can generate a fairly robust model of hourly traffic volumes. If the hourly volume model was extended to every hour of the year, a D-factor estimate could be generated from an hourly volume model directly. Additionally, there are a multitude of ways to improve upon the existing aggregation method to pick up more nuances in seasonality changes across time to improve results. Further research is needed into potential methods to extract a D-factor signal from Probe Data beyond this report. This examination by no means exhausted the potential of Probe Data to create estimates of D-factor.

Estimation of Percent Peak CU and SU Trucks

The previous section discussed how to use Probe Data to estimate the AADTs for different vehicle classes, including for those of Single Unit (SU) and Combination Unit (CU) trucks. This section will focus on how Percent Peak Trucks values can be estimated to help with road and pavement design.

Methods

In order to train and compare a model for percent peak CU and SU, hourly volumes from 2019 for the TMAS permanent counts with vehicle class dataset were used (as described for AADT by vehicle classification). This resulted in 1,843 locations for training and model validation. Percent Peak CU and Percent Peak CU were calculated using the following standard methods:

The Percent Peak CU value is the number of CU trucks (FHWA classes 8-13) during the peak hour divided by the AADT.

$$\text{Percent Peak CU Trucks} = \frac{Peak\ Hour\ CU\ Trucks}{AADT}\ x\ 100$$

Similarly, the Percent Peak SU value is the number of SU trucks (FHWA classes 4-7) during the peak hour divided by the AADT.

$$\text{Percent Peak SU Trucks} = \frac{Peak\ Hour\ SU\ Trucks}{AADT}\ x\ 100$$

The peak hour uses the hour with the 30th highest volume from the entire year.

Two different extreme gradient boosted models were fit from the TMAS data set, one to predict Percent Peak CU Trucks, and one to predict Percent Peak SU Trucks. K(10) folds cross validation was used to both train and test the model, to ensure that the model performance for each location was properly tested. Gradient boosted models may result in negative predictive values for sites that are substantially outside the observed range of training values, due to the fit regression lines in creation of the models. Less than 1% of Percent Peak SU trucks modeled estimates were negative values, and 2.7% of estimated values for Percent Peak CU trucks were negative. All negative results were transformed into estimates of 0, which is a more meaningful value,

For predictive features, factors derived from LBS sample trips, GPS sample trips, and contextual features were considered, as discussed in the Probe Data methods section. As with the predictive model for K factor, LBS and GPS trip data at each location were also aggregated into the average counts for each hour of day over the year, and included as predictive features in the model as well.

Estimation of Percent Peak Combination-Unit (CU) Trucks

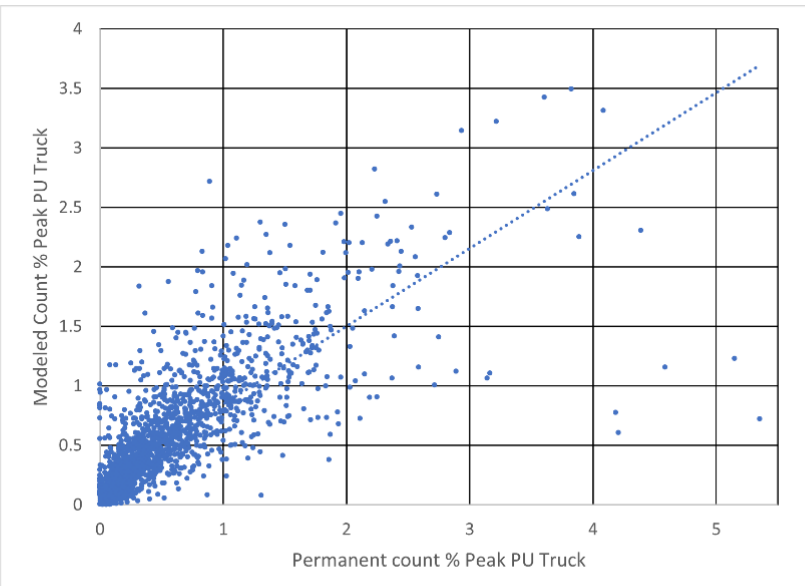
Percent peak CU Truck represents the percentage of traffic composed of CU trucks within the peak hour, normalized to AADT. As a result, this value is typically a small percentage: in the dataset, 83.3% of the stations have less than 1% peak CU trucks. Probe Data relies on sampling of the population to infer estimation, but for Percent Peak Truck calculations, a very small window of time, and a very small percentage of total trips on a road segment have to be estimated accurately within a very small margin. In order to simplify the description of the model's effectiveness, the problem was framed as a classification problem. Can we predict a road as having a high Percent Peak CU Truck, which is greater than 1% (and represented by 17% of the data)?

The model results are described in Table 1. In the dataset, 83.3% of the stations have less than 1% peak CU trucks, and of those stations, 94.7% of them are successfully predicted have low percent peak CU trucks. Of the remaining 16.7% of stations with 1% and higher peak CU trucks, 67.2% were classified correctly. The Person correlation score of the model estimates vs true peak is 0.78 (p-value <0.001), showing a strong, statistically significant correlation (Figure 1).

Table 1: Cross-tabulation result for model accuracy for Percent Peak CU Trucks. Values represent the percentage of time the Percent Peak CU Truck model was correctly able to label a site as having a low (<1%) value or high value (1% or greater).

Actual Range	Predicted Range < 1% (low)	Predicted Range 1% + (high)	Proportion
< 1% (low)	0.947	0.053	0.83
1% + (high)	0.328	0.672	0.16

Figure 1. Scatterplot comparing Percent Peak CU Truck modeled estimated to true values. Modeled percent estimates are presented rather than the classification groupings as presented in Table 1. Post-processing was done to convert negative estimates to zero.



Estimation of Percent Peak Single-Unit (SU) Trucks and Buses

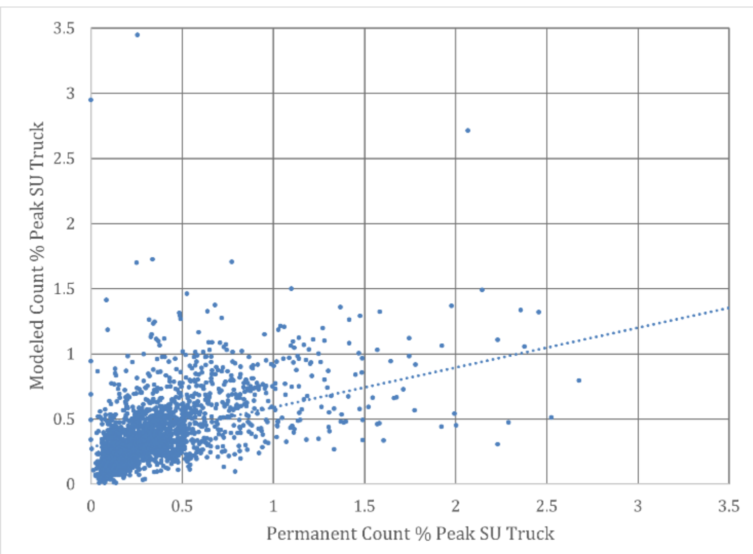
As with Percent Peak CU Truck, percent Peak SU Truck values in the dataset also encompass a very narrow range, with most values 93.8% of the stations have less than 1% peak SU trucks. Of stations with “low” Peak CU Truck percentages, 98.0% were correctly predicted to have low percent peak SU trucks (Table 2). Of the remaining 6.2% of stations with 1% and higher peak SU trucks, 25.2% were classified correctly.

Table 2: Cross-tabulation result for model accuracy for peak percent SU trucks. Values represent the percentage of time the Percent Peak SU Truck model was correctly able to label a site as having a low (<1%) value or high value (1% or greater).

Actual Range	Predicted Range < 1%	Predicted Range 1% +	Proportion
< 1%	0.980	0.020	0.938
1% +	0.748	0.252	0.062

As compared to Combination Trucks, the error on Single-Unit trucks is greater. The Pearson correlation between modeled estimates and true values is 0.44 (p-value <0.001), showing a significant relationship, but dramatically lower than the Pearson correlation for predicting Percent Peak CU truck. The higher spread in model error, particularly for higher values of Percent Peak SU is apparently in a scatterplot (Figure 3) of modeled vs true values.

Figure 3. Scatterplot comparing Percent Peak SU Truck modeled estimated to true values. Modeled percent estimates are presented rather than the classification groupings as presented in Table 1. Post-processing was done to convert negative estimates to zero.



Conclusions

Percent Peak SU and CU Truck values encompass a very narrow range, and most values are under 1%. Thus, these values demand a model with a very small window of error, since the range of possible values is so small. Model error was lower for predictions of Peak Percent CU as compared to SU. This mirrors or findings that modeled AADT for CU trucks also had lower error than for SU Trucks. Single-Unit trucks present a more ambiguous class that is not easily teased apart from the various data sources.

A model fit with Probe Data input derived from LBS and GPS data sources was able to successfully classify stations as having a low or high percent peak SU, CU Truck percentage at rates substantially better than chance, and thus for some use cases, a Probe Data derived model for Percent Peak SU and CU truck may be useful.

Chapter 9: Conclusions

The following study evaluated the validity of using Probe Data to estimate annual average daily traffic (AADT), as well as related estimations of traffic volume and derived factors. This evaluation was completed using a set of 4,255 permanent counter stations with data for 2019 traffic volume in the United States from the Federal Highway Administration Travel Monitoring System, individual State DOTs, and MS2 cloud-based traffic data management system. Traffic volumes and other factors were derived from these sources, and the resulting estimates were compared to modeled estimates fit using Probe Data sources.

The current study presented the following conclusions about estimation for traffic volume using Probe Data inputs:

1. **AADT** estimations were evaluated across a wide range of roadway classifications and locations throughout all 48 states of the contiguous United States (Chapter 3). Model error comparison between Probe Data estimates and same-year short term counter factorization showed that
    a. Probe Data estimates consistently out-performed same-year temporary counts for roads over 2,000 AADT.
    b. For roads between 500 and 2000 AADT, results were mixed: Probe Data outperformed same-year two-day expansion on MAPE for all but 5 states, on some metrics across all states, and the traditional method outperformed Probe Data for other metrics. Both methods produced usable results for most applications in this small road range, but individual states may wish to look at their local results for decisions in implementation. In addition, addition of more very-low volume roads with permanent counters to the calibration set available nationally may improve results for all.
    c. There is no reliable source to calculate the error of same year temporary count expansion for roads under 500, so it is unclear whether Probe Data or traditional methods are better.
    d. Because less than 9% of road miles on <5000 AADT roads (at least in the state of Minnesota, and per TMG guidance of multi-year rotation of counts) actually use a same-year temporary expansion, a note was added that the bar for comparing to places with no opportunity for a same year count should be different (less stringent)
    e. Probe Data was able to predict AADT for 2019 with acceptable accuracy with only ~6 months of data (though more months included do improve results). This implies the potential to make preliminary estimates of AADT mid-year for typical years.
    f. Probe Data model error is likely related to both the sample trip size of the Probe Data inputs, which naturally decreases as AADT volume decreases, and to the lower number of permanent counter calibration sites available for low-volume roads.

2. **Directional AADT** volume estimations from Probe Data were comparable to total road AADT estimations, and in general had less error than bi-directional estimates.
3. **Ramp AADT** showed comparable errors to non-ramp roads when evaluated by road volume categorizations.
4. **AADT by vehicle type** for three classes of vehicles was evaluated: personal vehicles (PV) (FHWA classes 1-3), single-unit (SU) vehicles (FHWA classes 4-7), and combination-unit (CU) vehicles (FHWA classes 8-13). AADT models maintain similar error as the National Probe Data AADT Model when errors are compared against the same vehicle-type specific bin of average volume. Predictions are strongest for PV and CU vehicles. The National AADT model relies heavily on a mix of LBS and GPS (personal and commercial) trips, suggesting that both the mix of different vehicle classes on the road, and inherent biases in different data sources all contribute to complexities that an AADT model must address.
5. **Monthly average daily traffic (MADT**) **estimation, average daily traffic (ADT) for specific days, and hourly volume** estimations were evaluated with Probe Data sample trips. Model error is correlated with total sample Probe Data trips, so smaller windows of time (monthly vs hourly), and smaller total road volumes (500 vs 5000 AADT) had higher error. There is no recognized source of comparison data (such as the Krile report) for MADT and ADT extrapolated from temporary counts. However, the Probe Data models captured normal and abnormal trends, and maintained consistent error as the National AADT model, suggesting that these estimated volumes can be used broadly.
6. **Special events,** explored in the finer resolution traffic volume estimates, were able to capture atypical trends, including changes in traffic due to **weather**, **road detours**, or local closing of businesses due to the COVID-19 pandemic in early 2020. Model estimates for atypical traffic volumes had the same range of error as typical days, indicating that Probe Data can be used to accurately measure volume during special events.
7. Probe Data was able to capture not only AADT, but also reflect seasonal changes over month of year (MOY), day of week (DOW), or time of day (TOD), as demonstrated through estimation of **MOY, DOW** and **TOD factors**.
8. **K-factor** and **D-facto**r estimations using Probe Data had relatively high errors for extreme values. More work and research are needed for strategies to model these values with higher accuracy using Probe Data, specifically researching finding new reference roads with extreme values that could be used to calibrate a model. But additionally, K and D factor estimation are measurements of a very small window of time: traffic volumes over one specific hour over an entire year. Probe Data creates inferences based on samples of trips, and thus there will likely be higher error when trying to estimate smaller windows of time. There may be other metrics correlated to K and D factors and serving the same use in traffic engineering that Probe Data is better situated to model.

There are a variety of technical aspects to consider with respect to modeling traffic volumes with Probe Data. First, is the availability of reference traffic volume counts to train and evaluate a Probe Data model. Continuous counts from permanent counters provide the best standard of comparison. However, such sensors still commonly malfunction, so use of these as a reference still requires strict quality assurance of the hourly volume data, before incorporating into a model. As detailed in Chapter 4, the most important factor is that the reference counters are well distributed across different types of roads, and a spread of urban and rural areas. Given a good distribution, the total number of counters could be reduced substantially with little impact on model accuracy. This report demonstrated with the example of Texas that even if a state has a large number of permanent counters, adding in reference counters from nearby states improves model error. The best results come from a full national model with the broadest set of permanent counters, despite additional counters being far from Texas. In practical terms, if a state has the choice of investing in more local permanent stations or using other states' pre-existing permanent counter data for calibration, using the other states' permanent counter locations may be more useful (and probably far less expensive). **It also implies that sharing of permanent counter data across many classes of roads between states will be to the benefit of all. In particular, sharing of permanent counter data from low volume roads could help drastically reduce model error on such roads for all.**

Certain characteristics of the Probe Data sample are also crucial for a successful model of traffic volume. As explored in Chapter 4, the most important characteristic is that the Probe Data sample be as representative of the population as possible. Blending different data sources with different biases, as done in the Prove Data National AADT model, is a good technique to capture a representative sample of traffic volumes along all types of roadways across the United States. One should consider the stability of trip penetration across a region of interest (and across the time of interest) when considering a potential Probe Data source. In addition, if truck counts are required, a data source known to be from trucks is critical.

Although representativeness of the population is crucial, additional Probe Data characteristics such as ping frequency and spatial accuracy should be considered. Volume estimation results only start to degrade with spatial accuracy above 500 meters and ping frequency above 15 minutes. This implies that the model is robust to some degradation of raw data quality, assuming a strong routing algorithm, and also that increased spatial precision and ping frequency (say to 1x/second and 2 meters) will not improve volume estimation results.

Given reference traffic volumes, and input Probe Data, there is still substantial work involved to generate an accurate traffic volume estimation. For AADT, there is a strong linear relationship between Probe Data and permanent counts. Depending on the tolerance of error, directionally helpful estimates may be possible with tools available in simple tools like excel. However, in order to make estimated AADT errors within the range of short-term count expansion, a simple linear model is not sufficient. Additional contextual information, such as census demographics, environmental, and roadway characteristics, are needed, as well as a machine learning model in order to obtain the best estimate. Use of multiple models to best fit different conditions, such as small vs large roads, may also be useful, depending on the model choice. The most important aspect here is not the specific algorithm used (Random Forest vs support vector machines vs neural networks vs ensemble learning), but that the model inputs be well generalized and that the predictor data is clean and representative.

Volume estimations from Probe Data have some clear advantages as compared to short term factor methods. With a National Probe Data AADT Model, volume estimations can be generated for any link of road at any given time point. Thus, MADT, hourly curves, truck counts, demographics of travelers, and more can be appended to AADTs with little additional investment. In the event of extreme events, such as a large flood, snowstorm, quarantine, work zone, or evacuation, the Probe Data trip sample will capture the traffic changes very well. In contrast, temporary count expansion requires a physical sensor to have been deployed ahead of time to capture such trends. Most importantly, most roads throughout the United states do not currently have a temporary or same-year permanent counter recording traffic counts for the current year, especially small roads. Probe Data volume estimation techniques provide a potentially powerful mechanism to improve estimation in an ongoing programmatic basis.

There are many additional areas to be explored that can offer improved results. For example, more research is needed to detail optimal placement of permanent counters across the country on a variety of road types to enhance model estimations. There is also a potential for blending short term counts and regional weighting of local counts to further improve estimation.

In light of the potential tools that Probe Data may provide to transportation professionals, there are many ways that Public and Private agencies can collaborate to utilize Probe Data. These include coordination of placement and distribution of reference permanent counters, to enhance the ability to develop models from all road types and regions throughout the entire US. Private and Public agencies should also work together in collaboration to redefine metrics, with a new eye toward what is really needed, and what may be achieved more cheaply in light of new technologies. For example, the Probe Data model had high error for Percent Peak CU, which involves modeling the percentage of trucks on a road for one specific hour of the entire year. This is a small percentage over a narrow window of time, which is not optimized for a model which is based on statistical sampling. Beyond Percent Factor CU, other metrics could similarly be re-draw in light of Probe Data techniques, to facilitate Probe Data's ability to robustly address transportation design needs across a wide variety of use cases.

The work for this whitepaper was performed in 2019 and 2020. Probe Data is a fast-evolving field and data options may evolve in subsequent years. Thus, it is more important to understand the useful characteristics of the data sources, and not simply the name/classification of the sources used at the time of this writing.

References

Minnesota Department of Transportation (2005-2019). [HPMS and Historical AADT Data] [Unpublished Raw Data].

Open Street Map. https://wiki.openstreetmap.org/wiki/About_OpenStreetMap&sa=D&ust=1607016036667000&usg=AOvVaw1j3zPOQIfu4icTa9tnKDHM. Accessed 2 December 2020.

Krile, R., Todt, F., Schroeder, J. (2015). Assessing Roadway Traffic Count Duration and Frequency Impacts on Annual Average Daily Traffic Estimation (FHWA-PL-16-008). United States. Federal Highway Administration.

Chang, H. H., & Cheon, S. H. (2019). The potential use of big vehicle GPS data for estimations of annual average daily traffic for unmeasured road segments. *Transportation*, *46*(3), 1011-1032.

FHWA. (2016). *Traffic Monitoring Guide*. U.S. Department of Transportation. Washington, D.C.

FHWA. (2018). *Traffic Data Computation Method*. U.S. Department of Transportation. Washington, D.C.

Thomason, Camille (2020). *Roadway Design Manual.* Texas Department of Transportation. United States.

Hou, Y., Young, S. E., Dimri, A., & Cohn, N. (2018). *Network Scale Ubiquitous Volume Estimation Using Tree-Based Ensemble Learning Methods* (No. NREL/CP-5400-70896). National Renewable Energy Lab. (NREL), Golden, CO (United States).

Kottek, M., J. Grieser, C. Beck, B. Rudolf, and F. Rubel, 2006: World Map of the Köppen-Geiger climate classification updated. *Meteorol. Z.*, **15**, 259-263. DOI: 10.1127/0941-2948/2006/0130. http://koeppen-geiger.vu-wien.ac.at/present.htm Accessed 12 January 2020.

Krile, R., Todt, F., Schroeder, J., & Jessberger, S. (2016). Assessing roadway traffic count duration and frequency impacts on annual average daily traffic estimation: assessing accuracy issues related to annual factoring (No. FHWA-PL-16-012). United States. Federal Highway Administration.

Schrank, D., Eisele, B., & Lomax, T. (2012). TTI's 2012 urban mobility report. *Texas A&M Transportation Institute. The Texas A&M University System*, *4*.

Sekuła, P., Marković, N., Laan, Z. V., & Sadabadi, K. F. (2017). Estimating historical hourly traffic volumes via machine learning and vehicle probe data: A Maryland case study. *arXiv preprint arXiv:1711.00721*.

StreetLight Data (2020). Larger and more representative samples: how big data can support equitable transportaiton analytics & decisions. Version 1.0. https://www.streetlightdata.com/big-data-supports-environmental-justice-in-transportation

Appendix

Table 1: Bi-directional Probe Data hourly volume model estimates for a two-week time series in October, as graphed in Chapter 7 Figure 1. Site is a representative large road in Wisconsin, with an AADT of 138,035, road size bin J. Model estimates are compared to volume from a reference permanent counter.

Date	Hour of Day	Reference Volume (permanent count)	Probe Data Model Estimate	Percent Error
10/7/2019	1	1635	1447	-11.47%
10/7/2019	2	866	960	10.88%
10/7/2019	3	645	684	6.03%
10/7/2019	4	867	1173	35.32%
10/7/2019	5	1815	2010	10.74%
10/7/2019	6	4731	4806	1.58%
10/7/2019	7	8702	8456	-2.83%
10/7/2019	8	10860	9912	-8.73%
10/7/2019	9	9478	8679	-8.43%
10/7/2019	10	7585	7997	5.43%
10/7/2019	11	7117	7482	5.12%
10/7/2019	12	7131	6781	-4.91%
10/7/2019	13	7721	8156	5.63%
10/7/2019	14	7689	7868	2.32%
10/7/2019	15	9487	9437	-0.53%
10/7/2019	16	9984	9817	-1.67%
10/7/2019	17	8412	9058	7.68%
10/7/2019	18	9167	9662	5.40%
10/7/2019	19	8110	8029	-1.00%
10/7/2019	20	5970	6592	10.42%
10/7/2019	21	4539	4992	9.99%
10/7/2019	22	3957	3494	-11.69%
10/7/2019	23	2734	2613	-4.43%
10/7/2019	24	1986	1785	-10.11%
10/8/2019	1	1494	1349	-9.68%
10/8/2019	2	864	1002	16.00%
10/8/2019	3	729	810	11.11%
10/8/2019	4	955	1147	20.13%
10/8/2019	5	1724	2051	18.99%
10/8/2019	6	4250	4900	15.28%
10/8/2019	7	8791	8277	-5.85%
10/8/2019	8	10673	9199	-13.81%

Date	Hour of Day	Reference Volume (permanent count)	Probe Data Model Estimate	Percent Error
10/8/2019	9	9526	10141	6.46%
10/8/2019	10	7573	8143	7.53%
10/8/2019	11	6882	7767	12.86%
10/8/2019	12	6937	6993	0.81%
10/8/2019	13	7735	7849	1.48%
10/8/2019	14	7749	8221	6.09%
10/8/2019	15	9554	9101	-4.74%
10/8/2019	16	9444	10189	7.89%
10/8/2019	17	9459	9515	0.59%
10/8/2019	18	9669	9706	0.38%
10/8/2019	19	8395	8649	3.03%
10/8/2019	20	6420	6368	-0.81%
10/8/2019	21	5156	5100	-1.09%
10/8/2019	22	4231	3964	-6.32%
10/8/2019	23	2638	2707	2.61%
10/8/2019	24	1894	1972	4.13%
10/9/2019	1	1294	1064	-17.76%
10/9/2019	2	766	742	-3.11%
10/9/2019	3	654	837	28.06%
10/9/2019	4	879	1140	29.67%
10/9/2019	5	1727	2163	25.24%
10/9/2019	6	4713	5289	12.21%
10/9/2019	7	8680	8389	-3.35%
10/9/2019	8	11018	10331	-6.24%
10/9/2019	9	7995	8837	10.53%
10/9/2019	10	7549	7174	-4.96%
10/9/2019	11	7121	7165	0.62%
10/9/2019	12	7423	7049	-5.04%
10/9/2019	13	7638	8018	4.98%
10/9/2019	14	7915	8514	7.57%
10/9/2019	15	9719	9365	-3.65%
10/9/2019	16	10182	9924	-2.53%
10/9/2019	17	9658	9601	-0.59%
10/9/2019	18	9827	9843	0.16%
10/9/2019	19	8685	8078	-6.99%
10/9/2019	20	6553	6261	-4.46%
10/9/2019	21	5211	5256	0.87%
10/9/2019	22	4679	3897	-16.72%

Date	Hour of Day	Reference Volume (permanent count)	Probe Data Model Estimate	Percent Error
10/9/2019	23	3095	3045	-1.61%
10/9/2019	24	1985	2181	9.88%
10/10/2019	1	1365	1411	3.35%
10/10/2019	2	835	776	-7.11%
10/10/2019	3	724	944	30.43%
10/10/2019	4	1024	1449	41.46%
10/10/2019	5	1883	2321	23.28%
10/10/2019	6	4733	4901	3.56%
10/10/2019	7	8749	8156	-6.78%
10/10/2019	8	10685	10465	-2.06%
10/10/2019	9	9520	8937	-6.13%
10/10/2019	10	8046	7936	-1.37%
10/10/2019	11	7422	7317	-1.41%
10/10/2019	12	7492	7736	3.26%
10/10/2019	13	8106	8352	3.04%
10/10/2019	14	8523	8708	2.17%
10/10/2019	15	9932	9795	-1.38%
10/10/2019	16	10243	10281	0.37%
10/10/2019	17	9911	10178	2.69%
10/10/2019	18	9421	9332	-0.95%
10/10/2019	19	8619	8331	-3.35%
10/10/2019	20	6671	6227	-6.66%
10/10/2019	21	5332	5840	9.52%
10/10/2019	22	4742	4217	-11.08%
10/10/2019	23	3617	3542	-2.09%
10/10/2019	24	2588	2719	5.06%
10/11/2019	1	1703	1711	0.44%
10/11/2019	2	1079	1052	-2.54%
10/11/2019	3	977	1160	18.75%
10/11/2019	4	1001	1391	38.95%
10/11/2019	5	1912	2196	14.85%
10/11/2019	6	4379	4393	0.31%
10/11/2019	7	7513	7133	-5.06%
10/11/2019	8	9683	9441	-2.50%
10/11/2019	9	9197	9950	8.18%
10/11/2019	10	7698	8743	13.58%
10/11/2019	11	7752	8020	3.46%
10/11/2019	12	7942	8039	1.22%

Date	Hour of Day	Reference Volume (permanent count)	Probe Data Model Estimate	Percent Error
10/11/2019	13	8981	9005	0.27%
10/11/2019	14	9415	8754	-7.02%
10/11/2019	15	10449	9911	-5.15%
10/11/2019	16	10085	9878	-2.05%
10/11/2019	17	9506	10132	6.59%
10/11/2019	18	9927	9947	0.20%
10/11/2019	19	8840	9273	4.89%
10/11/2019	20	7123	7232	1.54%
10/11/2019	21	5876	6640	13.00%
10/11/2019	22	5182	6079	17.32%
10/11/2019	23	4524	4590	1.45%
10/11/2019	24	3447	3231	-6.27%
10/12/2019	1	2306	2375	2.97%
10/12/2019	2	1492	1306	-12.49%
10/12/2019	3	1242	1199	-3.42%
10/12/2019	4	1112	997	-10.33%
10/12/2019	5	1254	1263	0.73%
10/12/2019	6	2082	1899	-8.81%
10/12/2019	7	2874	3449	20.00%
10/12/2019	8	3914	4467	14.12%
10/12/2019	9	5705	6309	10.58%
10/12/2019	10	6702	6756	0.81%
10/12/2019	11	7783	7701	-1.05%
10/12/2019	12	8308	7923	-4.63%
10/12/2019	13	8881	9027	1.65%
10/12/2019	14	8924	9379	5.10%
10/12/2019	15	9128	9278	1.65%
10/12/2019	16	8950	9271	3.59%
10/12/2019	17	8860	8710	-1.69%
10/12/2019	18	8425	8763	4.01%
10/12/2019	19	7711	7697	-0.18%
10/12/2019	20	6075	7388	21.61%
10/12/2019	21	5414	5692	5.14%
10/12/2019	22	5484	5260	-4.09%
10/12/2019	23	4350	3970	-8.74%
10/12/2019	24	3732	3501	-6.18%
10/13/2019	1	2476	2466	-0.41%
10/13/2019	2	1596	1589	-0.41%

Date	Hour of Day	Reference Volume (permanent count)	Probe Data Model Estimate	Percent Error
10/13/2019	3	1412	1157	-18.06%
10/13/2019	4	985	827	-16.04%
10/13/2019	5	1060	879	-17.08%
10/13/2019	6	1396	1299	-6.95%
10/13/2019	7	1834	2190	19.43%
10/13/2019	8	2515	2816	11.98%
10/13/2019	9	3673	3955	7.68%
10/13/2019	10	5602	6015	7.38%
10/13/2019	11	6892	6906	0.20%
10/13/2019	12	7851	8282	5.49%
10/13/2019	13	8875	8096	-8.78%
10/13/2019	14	8511	8107	-4.75%
10/13/2019	15	8843	8394	-5.07%
10/13/2019	16	8858	8322	-6.05%
10/13/2019	17	8314	8061	-3.04%
10/13/2019	18	7867	8040	2.19%
10/13/2019	19	7659	7206	-5.92%
10/13/2019	20	6054	6245	3.16%
10/13/2019	21	4818	4177	-13.30%
10/13/2019	22	3893	3336	-14.32%
10/13/2019	23	3026	2714	-10.30%
10/13/2019	24	2613	2610	-0.10%
10/14/2019	1	1563	2206	41.16%
10/14/2019	2	856	1168	36.50%
10/14/2019	3	694	844	21.58%
10/14/2019	4	853	916	7.41%
10/14/2019	5	1854	1800	-2.93%
10/14/2019	6	4528	4953	9.38%
10/14/2019	7	8232	7819	-5.02%
10/14/2019	8	10613	9909	-6.63%
10/14/2019	9	9500	9136	-3.83%
10/14/2019	10	7826	7111	-9.13%
10/14/2019	11	7417	6843	-7.74%
10/14/2019	12	7552	7331	-2.93%
10/14/2019	13	8240	8354	1.38%
10/14/2019	14	7786	8851	13.68%
10/14/2019	15	8776	8711	-0.74%
10/14/2019	16	9072	9536	5.12%

Date	Hour of Day	Reference Volume (permanent count)	Probe Data Model Estimate	Percent Error
10/14/2019	17	9383	9175	-2.21%
10/14/2019	18	9327	8494	-8.93%
10/14/2019	19	8249	7289	-11.64%
10/14/2019	20	5757	6504	12.98%
10/14/2019	21	4330	4329	-0.02%
10/14/2019	22	3506	3124	-10.91%
10/14/2019	23	2812	2447	-12.99%
10/14/2019	24	2317	1903	-17.85%
10/15/2019	1	1884	1512	-19.77%
10/15/2019	2	1454	1199	-17.56%
10/15/2019	3	807	963	19.36%
10/15/2019	4	998	1429	43.14%
10/15/2019	5	1964	2022	2.97%
10/15/2019	6	4721	5168	9.46%
10/15/2019	7	8634	8363	-3.14%
10/15/2019	8	10742	10214	-4.92%
10/15/2019	9	8801	9001	2.28%
10/15/2019	10	7638	7628	-0.13%
10/15/2019	11	7364	8052	9.34%
10/15/2019	12	7478	6976	-6.72%
10/15/2019	13	7915	7753	-2.05%
10/15/2019	14	8105	8754	8.00%
10/15/2019	15	9697	9452	-2.52%
10/15/2019	16	9557	9185	-3.89%
10/15/2019	17	9439	8848	-6.26%
10/15/2019	18	9131	8613	-5.67%
10/15/2019	19	8044	7684	-4.47%
10/15/2019	20	6057	6435	6.25%
10/15/2019	21	4832	4451	-7.87%
10/15/2019	22	4128	3411	-17.37%
10/15/2019	23	3015	2713	-10.03%
10/15/2019	24	2159	2290	6.07%
10/16/2019	1	1466	1426	-2.70%
10/16/2019	2	904	952	5.37%
10/16/2019	3	752	900	19.74%
10/16/2019	4	930	1148	23.44%
10/16/2019	5	1854	2102	13.39%
10/16/2019	6	4795	5070	5.74%

Date	Hour of Day	Reference Volume (permanent count)	Probe Data Model Estimate	Percent Error
10/16/2019	7	8777	8395	-4.36%
10/16/2019	8	10991	10150	-7.65%
10/16/2019	9	8967	9339	4.15%
10/16/2019	10	7698	7796	1.28%
10/16/2019	11	7136	7242	1.49%
10/16/2019	12	7256	7182	-1.02%
10/16/2019	13	8427	9222	9.43%
10/16/2019	14	8131	8854	8.89%
10/16/2019	15	9954	9773	-1.81%
10/16/2019	16	9961	9766	-1.95%
10/16/2019	17	9712	9822	1.13%
10/16/2019	18	9819	9709	-1.12%
10/16/2019	19	8927	8943	0.18%
10/16/2019	20	6600	6382	-3.30%
10/16/2019	21	5105	4986	-2.33%
10/16/2019	22	4164	3979	-4.44%
10/16/2019	23	3012	3085	2.42%
10/16/2019	24	2074	2264	9.17%
10/17/2019	1	1380	1397	1.21%
10/17/2019	2	838	1044	24.63%
10/17/2019	3	766	1070	39.64%
10/17/2019	4	1015	1246	22.78%
10/17/2019	5	1855	2015	8.64%
10/17/2019	6	4789	4901	2.34%
10/17/2019	7	8823	8199	-7.08%
10/17/2019	8	10929	10023	-8.29%
10/17/2019	9	7145	8906	24.65%
10/17/2019	10	7917	8229	3.94%
10/17/2019	11	7274	7041	-3.21%
10/17/2019	12	7489	7310	-2.39%
10/17/2019	13	8025	8721	8.67%
10/17/2019	14	8462	8723	3.08%
10/17/2019	15	9938	9867	-0.71%
10/17/2019	16	10381	10005	-3.62%
10/17/2019	17	10003	9774	-2.29%
10/17/2019	18	10038	9338	-6.98%
10/17/2019	19	9039	8540	-5.52%
10/17/2019	20	6795	7347	8.12%

Date	Hour of Day	Reference Volume (permanent count)	Probe Data Model Estimate	Percent Error
10/17/2019	21	5680	6385	12.40%
10/17/2019	22	5252	5539	5.46%
10/17/2019	23	3829	3492	-8.80%
10/17/2019	24	2537	2341	-7.74%
10/18/2019	1	1660	2069	24.62%
10/18/2019	2	1114	1422	27.64%
10/18/2019	3	903	1147	27.02%
10/18/2019	4	1087	1252	15.16%
10/18/2019	5	1829	1953	6.79%
10/18/2019	6	4567	4572	0.11%
10/18/2019	7	8193	7612	-7.09%
10/18/2019	8	10648	10115	-5.00%
10/18/2019	9	9171	8530	-6.99%
10/18/2019	10	7910	8784	11.04%
10/18/2019	11	7891	7355	-6.80%
10/18/2019	12	8430	8022	-4.84%
10/18/2019	13	9079	8771	-3.40%
10/18/2019	14	9162	9969	8.81%
10/18/2019	15	10277	10069	-2.03%
10/18/2019	16	10470	10135	-3.20%
10/18/2019	17	10043	9363	-6.77%
10/18/2019	18	10308	10342	0.33%
10/18/2019	19	9954	10537	5.86%
10/18/2019	20	7863	8456	7.54%
10/18/2019	21	6048	6928	14.54%
10/18/2019	22	5727	6728	17.48%
10/18/2019	23	5211	5204	-0.13%
10/18/2019	24	4430	3216	-27.40%
10/19/2019	1	2559	2394	-6.44%
10/19/2019	2	1716	2063	20.23%
10/19/2019	3	1388	1369	-1.37%
10/19/2019	4	1167	1078	-7.62%
10/19/2019	5	1219	1302	6.85%
10/19/2019	6	2080	2041	-1.87%
10/19/2019	7	2978	2552	-14.31%
10/19/2019	8	4186	4645	10.96%
10/19/2019	9	5839	6590	12.86%
10/19/2019	10	6943	6784	-2.29%

Date	Hour of Day	Reference Volume (permanent count)	Probe Data Model Estimate	Percent Error
10/19/2019	11	7748	6924	-10.64%
10/19/2019	12	8550	8098	-5.29%
10/19/2019	13	9205	9097	-1.17%
10/19/2019	14	8935	8807	-1.43%
10/19/2019	15	9219	8446	-8.38%
10/19/2019	16	9497	8817	-7.15%
10/19/2019	17	9189	9286	1.05%
10/19/2019	18	9133	9286	1.68%
10/19/2019	19	8031	8544	6.38%
10/19/2019	20	7219	7984	10.59%
10/19/2019	21	5843	7310	25.12%
10/19/2019	22	5629	5502	-2.26%
10/19/2019	23	5092	5484	7.69%
10/19/2019	24	4709	5227	11.01%
10/20/2019	1	2791	3075	10.16%
10/20/2019	2	1859	2237	20.31%
10/20/2019	3	1458	1494	2.44%
10/20/2019	4	1094	1134	3.62%
10/20/2019	5	1009	1197	18.63%
10/20/2019	6	1408	1592	13.10%
10/20/2019	7	2197	2678	21.89%
10/20/2019	8	3507	4129	17.74%
10/20/2019	9	4640	5757	24.07%
10/20/2019	10	5925	6679	12.73%
10/20/2019	11	7066	7113	0.67%
10/20/2019	12	8294	7137	-13.94%
10/20/2019	13	8342	7835	-6.07%
10/20/2019	14	8006	7168	-10.46%
10/20/2019	15	8010	7715	-3.68%
10/20/2019	16	8420	7693	-8.64%
10/20/2019	17	9144	8791	-3.86%
10/20/2019	18	8585	8397	-2.19%
10/20/2019	19	8231	7946	-3.46%
10/20/2019	20	7524	6753	-10.25%
10/20/2019	21	6019	6006	-0.21%
10/20/2019	22	4320	4185	-3.12%
10/20/2019	23	3233	3069	-5.07%
10/20/2019	24	2138	2176	1.77%

Table 2: Bi-directional Probe Data Hourly volume model estimates for a two week timeseries in October, as graphed in Chapter 7 Figure 2. This site represents a medium volume road (AADT 6688, road size bin D) in Michigan. Values are compared to a reference permanent counter site.

Date	Hour	Reference Volume (permanent count)	Modeled Estimate	Percent Error
10/7/2019	1	28	21	-25.81%
10/7/2019	2	23	26	13.33%
10/7/2019	3	25	20	-19.78%
10/7/2019	4	90	40	-55.84%
10/7/2019	5	327	150	-54.00%
10/7/2019	6	415	293	-29.40%
10/7/2019	7	356	366	2.85%
10/7/2019	8	466	468	0.48%
10/7/2019	9	391	440	12.42%
10/7/2019	10	311	444	42.77%
10/7/2019	11	375	351	-6.50%
10/7/2019	12	387	445	14.97%
10/7/2019	13	421	426	1.19%
10/7/2019	14	510	500	-1.88%
10/7/2019	15	561	550	-1.98%
10/7/2019	16	665	620	-6.73%
10/7/2019	17	619	630	1.80%
10/7/2019	18	489	470	-3.88%
10/7/2019	19	342	374	9.43%
10/7/2019	20	237	276	16.29%
10/7/2019	21	205	215	4.91%
10/7/2019	22	115	170	47.67%
10/7/2019	23	83	113	35.80%
10/7/2019	24	64	81	26.41%
10/8/2019	1	47	22	-53.41%
10/8/2019	2	28	56	98.73%
10/8/2019	3	33	33	1.23%
10/8/2019	4	99	41	-58.98%
10/8/2019	5	391	193	-50.62%
10/8/2019	6	391	335	-14.24%
10/8/2019	7	347	413	18.91%
10/8/2019	8	439	407	-7.21%
10/8/2019	9	351	400	13.99%
10/8/2019	10	335	397	18.60%
10/8/2019	11	367	339	-7.54%

Date	Hour	Reference Volume (permanent count)	Modeled Estimate	Percent Error
10/8/2019	12	393	358	-8.81%
10/8/2019	13	408	313	-23.22%
10/8/2019	14	460	425	-7.65%
10/8/2019	15	568	538	-5.26%
10/8/2019	16	715	565	-20.99%
10/8/2019	17	613	554	-9.70%
10/8/2019	18	478	428	-10.38%
10/8/2019	19	304	361	18.81%
10/8/2019	20	267	248	-7.26%
10/8/2019	21	236	198	-16.07%
10/8/2019	22	135	192	41.94%
10/8/2019	23	100	147	47.31%
10/8/2019	24	64	88	37.61%
10/9/2019	1	38	21	-45.34%
10/9/2019	2	32	38	17.25%
10/9/2019	3	45	49	9.35%
10/9/2019	4	109	86	-21.53%
10/9/2019	5	377	156	-58.55%
10/9/2019	6	388	311	-19.97%
10/9/2019	7	351	391	11.39%
10/9/2019	8	426	371	-12.81%
10/9/2019	9	387	403	4.04%
10/9/2019	10	308	303	-1.71%
10/9/2019	11	375	443	18.01%
10/9/2019	12	437	442	1.25%
10/9/2019	13	440	471	6.98%
10/9/2019	14	500	420	-16.09%
10/9/2019	15	526	466	-11.45%
10/9/2019	16	696	576	-17.30%
10/9/2019	17	648	565	-12.83%
10/9/2019	18	527	421	-20.19%
10/9/2019	19	341	345	1.13%
10/9/2019	20	231	237	2.58%
10/9/2019	21	193	204	5.58%
10/9/2019	22	110	132	19.75%
10/9/2019	23	84	34	-59.98%
10/9/2019	24	72	82	14.32%
10/10/2019	1	39	34	-12.12%

Date	Hour	Reference Volume (permanent count)	Modeled Estimate	Percent Error
10/10/2019	2	30	29	-3.27%
10/10/2019	3	42	30	-28.09%
10/10/2019	4	108	30	-72.49%
10/10/2019	5	378	158	-58.10%
10/10/2019	6	411	299	-27.31%
10/10/2019	7	352	447	26.93%
10/10/2019	8	443	536	20.90%
10/10/2019	9	347	463	33.33%
10/10/2019	10	370	372	0.58%
10/10/2019	11	377	449	18.99%
10/10/2019	12	397	419	5.49%
10/10/2019	13	469	422	-9.98%
10/10/2019	14	481	478	-0.71%
10/10/2019	15	615	624	1.41%
10/10/2019	16	728	730	0.30%
10/10/2019	17	690	713	3.32%
10/10/2019	18	554	557	0.58%
10/10/2019	19	392	351	-10.53%
10/10/2019	20	274	319	16.55%
10/10/2019	21	238	209	-12.15%
10/10/2019	22	133	130	-2.63%
10/10/2019	23	78	73	-6.83%
10/10/2019	24	79	70	-12.02%
10/11/2019	1	61	44	-27.87%
10/11/2019	2	37	49	31.51%
10/11/2019	3	37	51	37.54%
10/11/2019	4	97	120	23.29%
10/11/2019	5	316	205	-34.99%
10/11/2019	6	383	262	-31.72%
10/11/2019	7	306	259	-15.46%
10/11/2019	8	336	346	2.87%
10/11/2019	9	367	429	17.00%
10/11/2019	10	338	379	12.23%
10/11/2019	11	441	421	-4.44%
10/11/2019	12	520	557	7.17%
10/11/2019	13	555	543	-2.12%
10/11/2019	14	526	485	-7.72%
10/11/2019	15	603	607	0.68%

Date	Hour	Reference Volume (permanent count)	Modeled Estimate	Percent Error
10/11/2019	16	684	517	-24.36%
10/11/2019	17	660	646	-2.15%
10/11/2019	18	543	479	-11.73%
10/11/2019	19	368	379	3.04%
10/11/2019	20	290	369	27.14%
10/11/2019	21	273	247	-9.61%
10/11/2019	22	150	263	75.36%
10/11/2019	23	131	172	30.93%
10/11/2019	24	102	141	38.63%
10/12/2019	1	63	52	-16.99%
10/12/2019	2	43	32	-25.48%
10/12/2019	3	31	27	-13.44%
10/12/2019	4	27	28	4.94%
10/12/2019	5	57	61	7.01%
10/12/2019	6	130	88	-32.06%
10/12/2019	7	95	209	120.08%
10/12/2019	8	162	243	50.15%
10/12/2019	9	302	387	28.04%
10/12/2019	10	373	446	19.67%
10/12/2019	11	433	392	-9.49%
10/12/2019	12	449	451	0.35%
10/12/2019	13	470	483	2.86%
10/12/2019	14	429	466	8.67%
10/12/2019	15	460	436	-5.20%
10/12/2019	16	396	394	-0.47%
10/12/2019	17	449	403	-10.21%
10/12/2019	18	420	347	-17.34%
10/12/2019	19	318	318	0.03%
10/12/2019	20	304	245	-19.38%
10/12/2019	21	232	285	22.73%
10/12/2019	22	173	210	21.45%
10/12/2019	23	123	192	55.87%
10/12/2019	24	103	125	21.05%
10/13/2019	1	64	21	-67.54%
10/13/2019	2	41	60	47.54%
10/13/2019	3	26	34	29.94%
10/13/2019	4	30	20	-33.15%
10/13/2019	5	23	21	-7.84%

Date	Hour	Reference Volume (permanent count)	Modeled Estimate	Percent Error
10/13/2019	6	46	49	5.65%
10/13/2019	7	56	78	39.37%
10/13/2019	8	71	118	66.40%
10/13/2019	9	149	152	1.75%
10/13/2019	10	207	236	13.80%
10/13/2019	11	281	317	12.94%
10/13/2019	12	346	324	-6.46%
10/13/2019	13	450	386	-14.19%
10/13/2019	14	385	438	13.87%
10/13/2019	15	369	395	7.00%
10/13/2019	16	349	340	-2.66%
10/13/2019	17	427	314	-26.51%
10/13/2019	18	370	327	-11.58%
10/13/2019	19	305	285	-6.65%
10/13/2019	20	238	239	0.37%
10/13/2019	21	176	130	-26.10%
10/13/2019	22	105	109	3.57%
10/13/2019	23	81	69	-15.33%
10/13/2019	24	57	40	-29.44%
10/14/2019	1	24	92	282.08%
10/14/2019	2	11	20	82.31%
10/14/2019	3	20	30	51.02%
10/14/2019	4	106	30	-71.51%
10/14/2019	5	367	204	-44.55%
10/14/2019	6	389	290	-25.33%
10/14/2019	7	311	240	-22.97%
10/14/2019	8	388	361	-6.85%
10/14/2019	9	348	455	30.85%
10/14/2019	10	355	405	14.10%
10/14/2019	11	358	393	9.76%
10/14/2019	12	369	474	28.33%
10/14/2019	13	464	555	19.60%
10/14/2019	14	473	602	27.34%
10/14/2019	15	568	547	-3.72%
10/14/2019	16	603	506	-16.14%
10/14/2019	17	637	554	-13.08%
10/14/2019	18	472	464	-1.68%
10/14/2019	19	332	311	-6.42%

Date	Hour	Reference Volume (permanent count)	Modeled Estimate	Percent Error
10/14/2019	20	205	128	-37.59%
10/14/2019	21	158	124	-21.41%
10/14/2019	22	113	88	-22.48%
10/14/2019	23	64	97	52.13%
10/14/2019	24	67	50	-24.63%
10/15/2019	1	46	34	-26.44%
10/15/2019	2	34	58	71.88%
10/15/2019	3	33	20	-39.23%
10/15/2019	4	106	126	18.69%
10/15/2019	5	375	281	-25.17%
10/15/2019	6	380	258	-32.14%
10/15/2019	7	361	301	-16.75%
10/15/2019	8	412	347	-15.89%
10/15/2019	9	372	360	-3.10%
10/15/2019	10	363	408	12.32%
10/15/2019	11	355	486	36.95%
10/15/2019	12	394	435	10.51%
10/15/2019	13	475	420	-11.62%
10/15/2019	14	492	496	0.80%
10/15/2019	15	583	696	19.37%
10/15/2019	16	702	643	-8.45%
10/15/2019	17	621	533	-14.22%
10/15/2019	18	504	387	-23.14%
10/15/2019	19	313	305	-2.71%
10/15/2019	20	238	238	-0.18%
10/15/2019	21	197	184	-6.63%
10/15/2019	22	105	111	5.33%
10/15/2019	23	85	84	-1.12%
10/15/2019	24	69	59	-14.04%
10/16/2019	1	50	34	-32.33%
10/16/2019	2	30	34	14.74%
10/16/2019	3	38	58	51.57%
10/16/2019	4	101	106	4.56%
10/16/2019	5	399	248	-37.96%
10/16/2019	6	372	315	-15.28%
10/16/2019	7	361	328	-9.18%
10/16/2019	8	402	345	-14.16%
10/16/2019	9	343	381	11.21%

Date	Hour	Reference Volume (permanent count)	Modeled Estimate	Percent Error
10/16/2019	10	312	336	7.72%
10/16/2019	11	338	374	10.62%
10/16/2019	12	372	460	23.70%
10/16/2019	13	438	490	11.91%
10/16/2019	14	450	497	10.45%
10/16/2019	15	587	628	7.06%
10/16/2019	16	654	716	9.45%
10/16/2019	17	589	612	3.97%
10/16/2019	18	465	402	-13.46%
10/16/2019	19	293	318	8.49%
10/16/2019	20	240	244	1.69%
10/16/2019	21	181	209	15.73%
10/16/2019	22	142	151	6.49%
10/16/2019	23	69	72	4.24%
10/16/2019	24	69	48	-29.86%
10/17/2019	1	43	21	-51.69%
10/17/2019	2	37	26	-30.35%
10/17/2019	3	39	30	-22.56%
10/17/2019	4	107	97	-9.68%
10/17/2019	5	391	305	-22.10%
10/17/2019	6	364	293	-19.56%
10/17/2019	7	345	340	-1.33%
10/17/2019	8	420	431	2.65%
10/17/2019	9	362	387	6.78%
10/17/2019	10	303	357	17.94%
10/17/2019	11	339	337	-0.67%
10/17/2019	12	402	450	12.03%
10/17/2019	13	471	503	6.81%
10/17/2019	14	477	478	0.24%
10/17/2019	15	555	516	-7.06%
10/17/2019	16	677	694	2.48%
10/17/2019	17	622	642	3.18%
10/17/2019	18	543	543	0.03%
10/17/2019	19	334	266	-20.32%
10/17/2019	20	259	220	-14.95%
10/17/2019	21	209	219	4.57%
10/17/2019	22	118	203	72.33%
10/17/2019	23	98	132	34.57%

Date	Hour	Reference Volume (permanent count)	Modeled Estimate	Percent Error
10/17/2019	24	68	82	20.53%
10/18/2019	1	42	22	-47.87%
10/18/2019	2	33	35	5.26%
10/18/2019	3	27	37	37.26%
10/18/2019	4	87	101	15.94%
10/18/2019	5	325	336	3.41%
10/18/2019	6	357	363	1.78%
10/18/2019	7	313	355	13.27%
10/18/2019	8	377	374	-0.68%
10/18/2019	9	396	353	-10.78%
10/18/2019	10	378	449	18.70%
10/18/2019	11	421	362	-14.08%
10/18/2019	12	457	341	-25.42%
10/18/2019	13	491	448	-8.72%
10/18/2019	14	541	489	-9.57%
10/18/2019	15	569	598	5.05%
10/18/2019	16	703	649	-7.69%
10/18/2019	17	679	753	10.87%
10/18/2019	18	568	744	30.92%
10/18/2019	19	446	462	3.60%
10/18/2019	20	299	286	-4.31%
10/18/2019	21	259	247	-4.64%
10/18/2019	22	224	216	-3.37%
10/18/2019	23	161	183	13.41%
10/18/2019	24	113	107	-5.16%
10/19/2019	1	68	45	-33.24%
10/19/2019	2	51	30	-40.78%
10/19/2019	3	25	47	88.02%
10/19/2019	4	38	20	-47.23%
10/19/2019	5	61	38	-37.03%
10/19/2019	6	83	105	26.97%
10/19/2019	7	135	212	56.79%
10/19/2019	8	174	278	59.70%
10/19/2019	9	307	253	-17.57%
10/19/2019	10	395	346	-12.49%
10/19/2019	11	437	342	-21.82%
10/19/2019	12	459	462	0.72%
10/19/2019	13	532	498	-6.46%

Date	Hour	Reference Volume (permanent count)	Modeled Estimate	Percent Error
10/19/2019	14	469	502	6.94%
10/19/2019	15	442	434	-1.86%
10/19/2019	16	437	451	3.30%
10/19/2019	17	418	472	12.98%
10/19/2019	18	478	434	-9.18%
10/19/2019	19	352	400	13.75%
10/19/2019	20	292	317	8.67%
10/19/2019	21	232	254	9.60%
10/19/2019	22	172	185	7.33%
10/19/2019	23	144	142	-1.27%
10/19/2019	24	95	122	28.56%
10/20/2019	1	51	71	39.42%
10/20/2019	2	43	48	12.48%
10/20/2019	3	16	43	167.74%
10/20/2019	4	24	55	130.15%
10/20/2019	5	29	49	70.20%
10/20/2019	6	37	78	111.14%
10/20/2019	7	70	40	-42.84%
10/20/2019	8	81	146	80.42%
10/20/2019	9	170	151	-10.96%
10/20/2019	10	217	223	2.75%
10/20/2019	11	303	296	-2.30%
10/20/2019	12	335	406	21.05%
10/20/2019	13	487	413	-15.16%
10/20/2019	14	390	480	23.19%
10/20/2019	15	398	437	9.71%
10/20/2019	16	364	415	14.03%
10/20/2019	17	382	401	4.97%
10/20/2019	18	354	414	16.95%
10/20/2019	19	323	317	-1.96%
10/20/2019	20	255	234	-8.06%
10/20/2019	21	171	151	-11.66%
10/20/2019	22	114	141	23.90%
10/20/2019	23	62	93	50.32%
10/20/2019	24	50	40	-19.56%

Table 3: Bi-directional Probe Data hourly volume model estimates for a two-week timeseries in October, as graphed in Chapter 7 Figure 3. Site is a representative small road in Montana, with an AADT of 901, road size bin B. Model estimates are compared to volume from a reference permanent counter.

Date	Hour	Reference Volume (permanent count)	Modeled Estimate	Percent Error
10/7/2019	1	2	4	110.57%
10/7/2019	2	3	4	40.38%
10/7/2019	3	2	4	110.57%
10/7/2019	4	1	4	321.13%
10/7/2019	5	19	5	-74.21%
10/7/2019	6	44	17	-60.87%
110/7/2019	7	85	53	-38.11%
10/7/2019	8	57	65	13.93%
10/7/2019	9	62	114	83.84%
10/7/2019	10	76	112	46.72%
10/7/2019	11	52	100	91.70%
10/7/2019	12	54	113	109.95%
10/7/2019	13	58	113	94.61%
10/7/2019	14	63	114	81.41%
10/7/2019	15	74	114	54.44%
10/7/2019	16	73	114	56.56%
10/7/2019	17	81	114	41.10%
10/7/2019	18	81	122	50.57%
10/7/2019	19	88	116	31.72%
10/7/2019	20	40	81	102.98%
10/7/2019	21	38	55	45.08%
10/7/2019	22	15	28	87.40%
10/7/2019	23	10	11	11.66%
10/7/2019	24	7	37	424.29%
10/8/2019	1	2	4	110.57%
10/8/2019	2	2	4	110.57%
10/8/2019	3	4	4	5.28%
10/8/2019	4	1	4	321.13%
10/8/2019	5	18	5	-72.78%
10/8/2019	6	44	18	-60.18%
10/8/2019	7	75	81	7.63%
10/8/2019	8	58	89	53.25%
10/8/2019	9	66	112	68.96%
10/8/2019	10	88	117	32.94%

Date	Hour	Reference Volume (permanent count)	Modeled Estimate	Percent Error
10/8/2019	11	75	118	56.93%
10/8/2019	12	65	129	98.42%
10/8/2019	13	56	135	141.03%
10/8/2019	14	74	120	61.85%
10/8/2019	15	95	150	58.00%
10/8/2019	16	83	120	44.30%
10/8/2019	17	87	125	43.30%
10/8/2019	18	108	94	-12.80%
10/8/2019	19	79	113	43.50%
10/8/2019	20	25	87	248.96%
10/8/2019	21	43	42	-1.74%
10/8/2019	22	25	28	12.44%
10/8/2019	23	19	11	-41.23%
10/8/2019	24	5	10	102.81%
10/9/2019	1	3	4	40.38%
10/9/2019	2	4	4	5.28%
10/9/2019	3	4	4	5.28%
10/9/2019	4	4	4	5.28%
10/9/2019	5	17	21	23.70%
10/9/2019	6	35	48	36.62%
10/9/2019	7	55	52	-5.84%
10/9/2019	8	32	73	128.02%
10/9/2019	9	38	92	142.91%
10/9/2019	10	53	125	136.05%
10/9/2019	11	35	112	220.18%
10/9/2019	12	40	114	185.68%
10/9/2019	13	32	103	223.07%
10/9/2019	14	32	97	201.99%
10/9/2019	15	29	106	265.36%
10/9/2019	16	48	136	182.53%
10/9/2019	17	51	111	117.03%
10/9/2019	18	62	106	71.06%
10/9/2019	19	53	94	76.74%
10/9/2019	20	15	71	371.19%
10/9/2019	21	8	55	591.53%
10/9/2019	22	5	27	448.43%
10/9/2019	23	3	11	272.19%
10/9/2019	24	2	10	407.01%

Date	Hour	Reference Volume (permanent count)	Modeled Estimate	Percent Error
10/10/2019	1	0	4	inf
10/10/2019	2	2	4	110.57%
10/10/2019	3	0	4	inf
10/10/2019	4	0	4	inf
10/10/2019	5	14	5	-65.00%
10/10/2019	6	34	13	-60.37%
10/10/2019	7	50	57	14.45%
10/10/2019	8	39	77	98.31%
10/10/2019	9	43	87	101.74%
10/10/2019	10	23	95	311.74%
10/10/2019	11	47	98	108.14%
10/10/2019	12	35	98	179.50%
10/10/2019	13	30	139	363.78%
10/10/2019	14	48	116	140.84%
10/10/2019	15	33	110	232.10%
10/10/2019	16	52	142	173.01%
10/10/2019	17	73	100	36.84%
10/10/2019	18	71	126	77.44%
10/10/2019	19	79	90	13.61%
10/10/2019	20	26	70	170.50%
10/10/2019	21	18	41	130.34%
10/10/2019	22	11	28	151.12%
10/10/2019	23	5	11	123.31%
10/10/2019	24	3	10	238.01%
10/11/2019	1	5	4	-15.77%
10/11/2019	2	3	4	40.38%
10/11/2019	3	3	4	40.38%
10/11/2019	4	3	4	40.38%
10/11/2019	5	13	5	-62.31%
10/11/2019	6	27	14	-49.63%
10/11/2019	7	50	52	4.67%
10/11/2019	8	43	76	76.74%
10/11/2019	9	51	94	85.09%
10/11/2019	10	62	102	65.17%
10/11/2019	11	71	105	47.78%
10/11/2019	12	55	111	101.79%
10/11/2019	13	53	99	85.99%
10/11/2019	14	55	107	94.53%

Date	Hour	Reference Volume (permanent count)	Modeled Estimate	Percent Error
10/11/2019	15	59	107	81.34%
10/11/2019	16	79	108	36.20%
10/11/2019	17	77	107	38.95%
10/11/2019	18	84	125	48.22%
10/11/2019	19	81	97	19.57%
10/11/2019	20	36	74	104.82%
10/11/2019	21	17	42	146.24%
10/11/2019	22	23	28	22.10%
10/11/2019	23	17	12	-28.14%
10/11/2019	24	5	12	147.19%
10/12/2019	1	11	4	-61.72%
10/12/2019	2	2	4	110.57%
10/12/2019	3	1	4	321.13%
10/12/2019	4	0	4	inf
10/12/2019	5	13	5	-62.31%
10/12/2019	6	18	48	167.72%
10/12/2019	7	38	52	37.72%
10/12/2019	8	51	90	77.04%
10/12/2019	9	66	119	79.76%
10/12/2019	10	65	98	50.54%
10/12/2019	11	67	125	85.96%
10/12/2019	12	47	106	125.95%
10/12/2019	13	44	108	145.57%
10/12/2019	14	61	116	90.95%
10/12/2019	15	68	108	59.21%
10/12/2019	16	58	108	86.66%
10/12/2019	17	55	108	96.84%
10/12/2019	18	65	117	80.48%
10/12/2019	19	62	132	113.10%
10/12/2019	20	44	76	71.86%
10/12/2019	21	23	42	81.27%
10/12/2019	22	20	28	39.58%
10/12/2019	23	11	12	11.05%
10/12/2019	24	8	12	54.50%
10/13/2019	1	8	4	-47.36%
10/13/2019	2	5	4	-15.77%
10/13/2019	3	1	4	321.13%
10/13/2019	4	1	4	321.13%

Date	Hour	Reference Volume (permanent count)	Modeled Estimate	Percent Error
10/13/2019	5	12	5	-59.17%
10/13/2019	6	13	12	-4.71%
10/13/2019	7	33	38	15.96%
10/13/2019	8	21	47	124.60%
10/13/2019	9	19	92	385.08 %
10/13/2019	10	39	129	230.55%
10/13/2019	11	44	109	147.97%
10/13/2019	12	53	103	93.88%
10/13/2019	13	41	116	182.41%
10/13/2019	14	53	112	111.96%
10/13/2019	15	39	112	186.75%
10/13/2019	16	46	112	143.12%
10/13/2019	17	66	112	69.45%
10/13/2019	18	53	111	109.58%
10/13/2019	19	82	99	20.18%
10/13/2019	20	55	91	64.55%
10/13/2019	21	28	42	48.90%
10/13/2019	22	21	28	33.03%
10/13/2019	23	11	11	1.51%
10/13/2019	24	3	10	238.01%
10/14/2019	1	4	4	5.28%
10/14/2019	2	3	4	40.38%
10/14/2019	3	4	4	5.28%
10/14/2019	4	0	4	inf
10/14/2019	5	19	5	-74.21%
10/14/2019	6	42	41	-1.28%
10/14/2019	7	68	68	-0.63%
10/14/2019	8	62	109	75.71%
10/14/2019	9	51	102	99.67%
10/14/2019	10	73	109	49.80%
10/14/2019	11	56	137	145.44%
10/14/2019	12	51	89	74.75%
10/14/2019	13	54	111	105.04%
10/14/2019	14	61	123	101.29%
10/14/2019	15	55	113	105.08%
10/14/2019	16	65	113	73.53%
10/14/2019	17	78	115	47.26%
10/14/2019	18	78	112	43.64%

Date	Hour	Reference Volume (permanent count)	Modeled Estimate	Percent Error
10/14/2019	19	95	103	8.38%
10/14/2019	20	43	122	183.13%
10/14/2019	21	20	42	108.46%
10/14/2019	22	25	28	11.75%
10/14/2019	23	13	11	-14.11%
10/14/2019	24	3	10	238.01%
10/15/2019	1	1	4	321.13%
10/15/2019	2	4	15	276.43%
10/15/2019	3	3	4	40.38%
10/15/2019	4	1	4	321.13%
10/15/2019	5	15	5	-67.33%
10/15/2019	6	38	14	-62.31%
10/15/2019	7	73	133	82.87%
10/15/2019	8	73	142	94.86%
10/15/2019	9	61	96	57.91%
10/15/2019	10	69	104	50.51%
10/15/2019	11	61	107	75.37%
10/15/2019	12	46	114	147.19%
10/15/2019	13	50	114	127.41%
10/15/2019	14	61	82	33.91%
10/15/2019	15	49	114	132.06%
10/15/2019	16	89	114	27.76%
10/15/2019	17	90	143	59.41%
10/15/2019	18	89	113	26.91%
10/15/2019	19	75	100	33.50%
10/15/2019	20	41	88	115.06%
10/15/2019	21	20	63	212.98%
10/15/2019	22	9	28	210.40%
10/15/2019	23	8	11	39.57%
10/15/2019	24	3	10	238.01%
10/16/2019	1	2	15	674.83%
10/16/2019	2	2	4	110.57%
10/16/2019	3	1	4	321.13%
10/16/2019	4	1	4	321.13%
10/16/2019	5	16	42	163.28%
10/16/2019	6	40	42	5.31%
10/16/2019	7	65	75	14.86%
10/16/2019	8	70	83	18.97%

Date	Hour	Reference Volume (permanent count)	Modeled Estimate	Percent Error
10/16/2019	9	66	115	73.77%
10/16/2019	10	76	113	49.29%
10/16/2019	11	59	113	90.69%
10/16/2019	12	62	140	125.98%
10/16/2019	13	67	135	100.83%
10/16/2019	14	61	101	64.80%
10/16/2019	15	65	94	45.06%
10/16/2019	16	67	119	77.58%
10/16/2019	17	83	120	44.74%
10/16/2019	18	108	114	5.61%
10/16/2019	19	86	108	26.06%
10/16/2019	20	51	90	76.01%
10/16/2019	21	27	42	56.48%
10/16/2019	22	12	28	134.24%
10/16/2019	23	12	11	-6.95%
10/16/2019	24	2	10	407.01%
10/17/2019	1	3	4	40.38%
10/17/2019	2	2	4	110.57%
10/17/2019	3	3	4	40.38%
10/17/2019	4	4	4	5.28%
10/17/2019	5	14	19	38.33%
10/17/2019	6	58	35	-40.17%
10/17/2019	7	78	74	-5.46%
10/17/2019	8	64	76	19.23%
10/17/2019	9	70	108	54.83%
10/17/2019	10	81	102	25.45%
10/17/2019	11	75	115	53.32%
10/17/2019	12	64	115	79.67%
10/17/2019	13	62	115	85.47%
10/17/2019	14	69	124	80.14%
10/17/2019	15	72	124	71.67%
10/17/2019	16	78	117	50.02%
10/17/2019	17	71	124	75.06%
10/17/2019	18	109	116	6.79%
10/17/2019	19	95	118	24.52%
10/17/2019	20	56	101	80.47%
10/17/2019	21	35	42	20.71%
10/17/2019	22	20	34	71.62%

Date	Hour	Reference Volume (permanent count)	Modeled Estimate	Percent Error
10/17/2019	23	8	11	39.57%
10/17/2019	24	5	10	102.81%
10/18/2019	1	4	4	5.28%
10/18/2019	2	3	4	40.38%
10/18/2019	3	4	4	5.28%
10/18/2019	4	3	4	40.38%
10/18/2019	5	18	5	-72.78%
10/18/2019	6	28	17	-38.01%
10/18/2019	7	61	69	13.13%
10/18/2019	8	50	76	51.45%
10/18/2019	9	76	96	26.53%
10/18/2019	10	71	122	71.47%
10/18/2019	11	61	145	137.45%
10/18/2019	12	64	106	65.74%
10/18/2019	13	59	109	84.98%
10/18/2019	14	76	123	61.62%
10/18/2019	15	85	128	51.09%
10/18/2019	16	104	167	61.04%
10/18/2019	17	98	145	47.94%
10/18/2019	18	124	122	-1.43%
10/18/2019	19	83	114	37.63%
10/18/2019	20	45	89	97.84%
10/18/2019	21	15	42	182.79%
10/18/2019	22	30	28	-5.81%
10/18/2019	23	32	13	-58.90%
10/18/2019	24	7	13	89.94%
10/19/2019	1	2	4	110.57%
10/19/2019	2	4	4	5.28%
10/19/2019	3	3	4	40.38%
10/19/2019	4	3	4	40.38%
10/19/2019	5	11	5	-55.45%
10/19/2019	6	10	17	66.34%
10/19/2019	7	40	55	38.37%
10/19/2019	8	40	72	79.03%
10/19/2019	9	52	99	90.41%
10/19/2019	10	42	109	158.41%
10/19/2019	11	59	109	85.16%
10/19/2019	12	62	115	85.79%

Date	Hour	Reference Volume (permanent count)	Modeled Estimate	Percent Error
10/19/2019	13	65	106	62.38%
10/19/2019	14	66	111	68.66%
10/19/2019	15	60	111	85.52%
10/19/2019	16	56	105	87.22%
10/19/2019	17	71	111	56.78%
10/19/2019	18	68	116	71.32%
10/19/2019	19	55	102	86.35%
10/19/2019	20	39	79	101.74%
10/19/2019	21	24	49	103.11%
10/19/2019	22	18	28	56.05%
10/19/2019	23	11	13	19.56%
10/19/2019	24	7	13	89.94%
10/20/2019	1	5	4	-15.77%
10/20/2019	2	7	4	-39.84%
10/20/2019	3	1	4	321.13%
10/20/2019	4	0	4	inf
10/20/2019	5	12	5	-59.17%
10/20/2019	6	9	15	69.99%
10/20/2019	7	27	42	56.72%
10/20/2019	8	13	96	641.58%
10/20/2019	9	15	102	581.89%
10/20/2019	10	26	108	316.66%
10/20/2019	11	45	127	181.24%
10/20/2019	12	42	115	174.30%
10/20/2019	13	59	103	75.11%
10/20/2019	14	34	112	228.73%
10/20/2019	15	41	112	172.61%
10/20/2019	16	57	112	96.09%
10/20/2019	17	75	112	49.03%
10/20/2019	18	57	122	113.21%
10/20/2019	19	56	120	114.14%
10/20/2019	20	39	75	92.94%
10/20/2019	21	25	42	67.44%
10/20/2019	22	16	28	75.65%
10/20/2019	23	6	11	86.09%
10/20/2019	24	4	10	153.51%

U.S. Department of Transportation
Federal Highway Administration

U.S. Department of Transportation
Federal Highway Administration
Office of Highway Policy Information
1200 New Jersey Ave., SE
Washington, D.C. 20590
https://www.fhwa.dot.gov/policyinformation
September 2021
Publication No. FHWA-PL-21-030