ACCURACY OF PROBE-BASED ANNUAL AVERAGE DAILY TRAFFIC (AADT) ESTIMATES IN BORDER REGIONS

by

Ioannis Tsapakis, Ph.D.
Associate Research Scientist

Lorenzo Cornejo
Assistant Transportation Researcher

Alfredo Sánchez, P.E.
Associate Research Engineer

Project performed by
Center for International Intelligent Transportation Research

In cooperation with
StreetLight Data Inc.

CIITR Project: Accuracy of Probe-Based Annual Average Daily Traffic (AADT) Estimates in Border Regions

January 2020

Report prepared by
Center for International Intelligent Transportation Research
Texas A&M Transportation Institute
4050 Rio Bravo, Suite 151
El Paso, Texas 79902

TEXAS A&M TRANSPORTATION INSTITUTE
The Texas A&M University System
College Station, Texas 77843-3135

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

DISCLAIMER AND ACKNOWLEDGMENTS

This research was performed by the Center for International Intelligent Transportation Research, a part of the Texas A&M Transportation Institute (TTI), in cooperation with StreetLight Data Inc. The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein.

The research team would like to thank the following individuals for providing TTI with traffic volume data that were used for evaluation purposes: Erika Garcia (Cameron County), Margie Montez (City of Del Rio), Maricela Gonzales (City of Donna), Anna De La Garza (City of Eagle Pass), Paul Stresow (City of El Paso), Eddie Romero (City of El Paso), Yvette Limon (City of Laredo), Juan Olaguibel (City of McAllen), Luis A. Bazán (City of Pharr), and Fred Brouwen (City of Pharr).

The authors would like to thank StreetLight Data for providing different types of data for evaluation purposes. In particular, Ms. Laura Schewel (chief executive officer) and Mr. Matthew Pettit (solution engineer) were especially helpful in providing access to StreetLight data and technical support throughout the project.

EXECUTIVE SUMMARY

The traffic on the U.S. side of international border crossings is monitored by traditional detection systems. Even though some of these systems can characterize traffic and determine freight movements, they tend to be expensive to purchase, install, operate, and maintain. Considering their limited budgets, transportation agencies in border regions try to find alternative ways to reduce data collection costs without sacrificing the accuracy of traffic volume data and estimates. In addition, they try to use common performance measures to monitor traffic operations consistently on both sides of border crossings.

Over the last few years, there has been an increasing interest in exploring whether passively collected data from mobile devices that are already in the traffic stream can be used accurately and confidently to estimate traffic volumes and other types of data. Examples of these devices include smartphones, personal and commercial navigation devices, and fleet monitoring systems. This study examined the accuracy of probe-based annual average daily traffic (AADT) estimates in two study areas: (a) at Texas-Mexico border crossings, and (b) on counted Texas roadways that are in proximity to the Mexican borders.

In this project, StreetLight Data Inc., a third-party data vendor, provided the Texas A&M Transportation Institute (TTI) with unscaled and uncalibrated mobile device count data for commercial and privately owned vehicles, as well as probe-based AADT estimates for several locations in the two study areas. For each study area, TTI determined the penetration rate of mobile devices and compared StreetLight Data AADT estimates against traffic volume data collected by various state and local agencies in Texas. The main research questions and the corresponding findings from this study are summarized below:

- **What is the penetration rate of mobile devices in the two study areas?**
  - The average penetration rate in the first study area (i.e., at border crossings) was 1.06 percent.
  - The average penetration rate in the second study area (i.e., at count locations within the three border Texas Department of Transportation [TxDOT] districts) was 0.86 percent. The analysis revealed that the penetration rate on rural roads was higher than that on urban roads in all three TxDOT districts examined in the study. This trend was also observed within all functional classes except functional class 7 (local roads).
- **What is the penetration rate of mobile devices used in commercial vehicles versus privately owned vehicles?**
  - The penetration rate of global positioning system (GPS) commercial vehicle trips was 8.7 percent, followed by the penetration rate of location-based services (LBS) privately owned vehicle trips (0.85 percent). The lowest penetration rate was 0.03 percent and was observed for GPS privately owned vehicle trips.
- **What is the anticipated accuracy of probe-based AADT estimates in the two study areas?**
  - In the first study area, the mean absolute percent error (MAPE) that was estimated using data from 10 ports of entry (POEs) was 33.0 percent. StreetLight Data 2018 AADT estimates were lower than the observed AADT values at nine POEs. In the second study area, the MAPE was 50 percent, which is lower than the corresponding

MAPE (61 percent) reported in a 2017 study that evaluated 2015 AADT estimates developed by StreetLight Data.

- **Where do probe-based AADT estimates tend to be more accurate and why?**
    - In general, the AADT accuracy gradually improved from low-volume roads to high-volume roads. The AADT estimates were higher than TxDOT AADT values within the two lower traffic volume ranges (401–5,000 and 5,001–10,000 vehicles per day), but this trend was reversed in the case of higher-volume roads (10,001–20,000, 20,001–50,000, >50,000 vehicles/day). Overall, the AADT estimates for urban roads were more accurate (MAPE=47 percent) than those for rural roads (MAPE=63 percent).

As the use of mobile devices continues to increase and data providers continue to enhance their traffic volume prediction methods, the accuracy of AADT estimates is expected to improve. For example, the 2017 AADT estimates used in this project resulted in lower errors than those reported in a published report that evaluated the accuracy of 2015 AADT estimates. Future evaluations of probe-based AADT estimates are needed using data from different states and regions that have diverse traffic, geometric, demographic, socioeconomic, and weather characteristics.

The findings of this research may apply to other border regions across the United States. This presents an opportunity for technology transfer. In addition, obtaining AADT estimates from probe data could provide a common measure to assess the performance of traffic operations on both sides of border crossings. It could also yield time and cost savings for transportation agencies that either do not collect traffic volume data (e.g. Mexican agencies) or deploy expensive traffic equipment. It could also reduce safety risks to employees and contractors who typically go out in the field to install sensor devices in and on roadways. Further, it can assist agencies in meeting new federal requirements according to which States must have access to a series of Model Inventory Roadway Elements - Fundamental Data Elements, including AADT, for all public paved roads by 2026.

CHAPTER 1: INTRODUCTION

Traffic congestion remains one of the main problems at U.S.-Mexico border crossings. At some ports of entry (POEs), traffic queuing has worsened over time. The increasing number of personal and commercial vehicles, in conjunction with more thorough inspections, particularly in the southbound direction of some POEs, makes the problem more acute. The observed queues of vehicles are much longer now compared to just a few years ago. Further, the absence of traffic monitoring equipment, and thus traffic volume data, on Mexican roads is another obstacle that transportation agencies must overcome as they try to understand existing problems and find appropriate strategies to alleviate traffic congestion and improve traffic operations.

The traffic on the U.S. side of international border crossings is monitored by traditional detection systems. Even though some of these systems can characterize the traffic and determine freight movements, they tend to be expensive to purchase, install, operate, and maintain. Considering their limited budgets, transportation agencies in border regions try to find alternative ways to reduce data collection costs without sacrificing the accuracy of traffic volume data and estimates. In addition, they try to use common performance measures to monitor traffic operations consistently on both sides of border crossings. Over the last few years, there has been an increasing interest in exploring whether passively collected data from mobile devices that are already in the traffic stream can be used accurately and confidently to estimate traffic volumes and other types of data (e.g., origin-destination data). Examples of these devices include smartphones, personal and commercial navigation devices, and fleet monitoring systems.

1.1 GOAL AND RESEARCH QUESTIONS

The goal of this project was to examine the accuracy of annual average daily traffic (AADT) estimates developed using passively collected and other types of non-traffic data (e.g., socioeconomic and demographic data) for two study areas: (a) at Texas-Mexico crossings, and (b) on U.S. roads that are in proximity to the Mexican borders. The main research questions were:

- What is the penetration rate (PR), also known as capture or sample rate, of mobile devices in the two study areas?
- What is the penetration rate of mobile devices used in commercial vehicles versus privately owned vehicles?
- What is the anticipated accuracy of probe-based AADT estimates in the two study areas?
- Where do probe-based AADT estimates tend to be more accurate and why?

To address these questions, the researchers compared different types of passively collected data against traffic volume data collected by various state and local agencies in Texas. In this study, the Texas A&M Transportation Institute (TTI) evaluated passively collected data provided by StreetLight Data Inc. (https://www.streetlightdata.com/), a third-party data vendor. In particular, StreetLight Data provided TTI with unscaled and uncalibrated global positioning system (GPS) and location-based services (LBS) trip (raw count) data for commercial and privately owned vehicles, as well as probe-based AADT estimates for several locations in the two study areas.

1.2 ORGANIZATION OF THE REPORT

The remaining chapters of this report include the following:

- **Chapter 2: Study Data**. This chapter describes the study network and the different types of data processed and used in the analysis.
- **Chapter 3: Evaluation of Passively Collected Data**. This chapter presents the analysis performed to determine the penetration rate of uncalibrated probe-based trips and the accuracy of probe-based AADT estimates developed by StreetLight Data. The results are presented separately for the two study areas.
- **Chapter 4: Conclusions**. This chapter presents the main findings and conclusions drawn from this study.

CHAPTER 2: STUDY DATA

2.1 STUDY AREA

The main focus area of this study included 28 POEs along the Texas-Mexico border. However, for completeness, TTI examined as a secondary study area more than 4,000 roadway locations within the three border Texas Department of Transportation (TxDOT) districts (El Paso, Laredo, and Pharr Districts), where permanent and short-duration traffic data have been collected by TxDOT. Figure 1 shows the two study areas: the POEs and the traffic count locations within the three districts of interest.

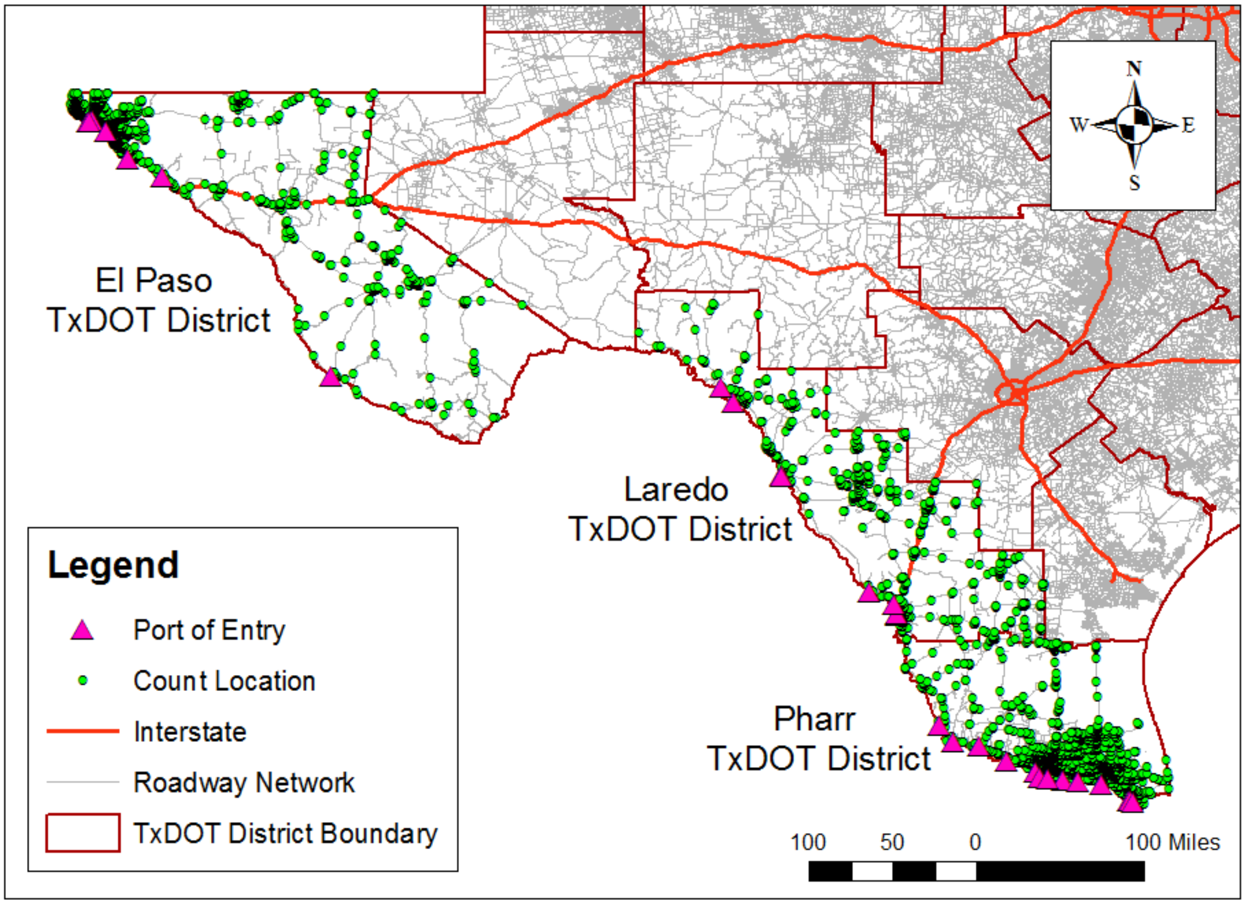

Figure 1. Study Area: (a) Ports of Entry, and (b) Count Locations in Three Border TxDOT Districts.

For clarity, the evaluation of passively collected data (see Chapter 3) was separately performed for each study area. Table 1 lists the 28 study POEs and their basic characteristics, including the TxDOT district, type of traffic permitted (i.e., commercial and privately owned vehicle traffic), and direction of travel (i.e., northbound and/or southbound traffic).

Table 1. Study Ports of Entry.

No.	Port of Entry Description	TxDOT District[a]	Traffic Type[b]	Direction of Travel
1	Brownsville B&M Bridge	PHR	POV	NB SB
2	Brownsville Gateway International Bridge	PHR	POV	NB SB
3	Brownsville Veterans International Bridge	PHR	COM&POV	NB SB
4	Del Rio–Ciudad Acuna International Bridge	LRD	COM&POV	NB SB
5	Donna Rio Bravo International Bridge	PHR	POV	NB SB
6	Eagle Pass Bridge I	LRD	POV	NB SB
7	Eagle Pass Camino Real International Bridge (Bridge II)	LRD	COM&POV	NB SB
8	El Paso Bridge of the Americas (BOTA)	ELP	COM&POV	NB SB
9	El Paso del Norte Bridge	ELP	POV	NB
10	El Paso Stanton Bridge	ELP	POV	SB
11	El Paso Ysleta-Zaragoza Bridge	ELP	COM&POV	NB SB
12	Fort Hancock El Porvenir Bridge	ELP	POV	NB SB
13	Free Trade Bridge	PHR	COM&POV	NB SB
14	Lake Amistad Dam Crossing	LRD	POV	NB SB
15	Lake Falcon Dam Crossing	PHR	POV	NB SB
16	Laredo Gateway to the Americas Bridge (Laredo Bridge I)	LRD	COM[c]	SB
17	Laredo Juarez-Lincoln Bridge (Laredo Bridge II)	LRD	COM&POV	NB SB
18	Laredo World Trade Bridge	LRD	COM	NB SB
19	Laredo-Colombia Solidarity Bridge	LRD	COM&POV	NB SB
20	Los Ebanos Ferry	PHR	POV	NB SB
21	McAllen Anzalduas International Bridge	PHR	POV	NB SB
22	McAllen-Hidalgo-Reynosa Bridge	PHR	POV	NB SB
23	Pharr-Reynosa Bridge	PHR	COM&POV	NB SB
24	Presidio Bridge	ELP	COM&POV	NB SB
25	Progreso Bridge	PHR	COM&POV	NB SB
26	Rio Grande City–Camargo Bridge	PHR	COM&POV	NB SB
27	Roma–Ciudad Miguel Aleman Bridge	PHR	COM&POV	NB SB
28	Tornillo-Guadalupe Bridge	ELP	COM&POV	NB SB

[a] ELP = El Paso, LRD = Laredo, PHR = Pharr.
[b] COM = commercial vehicles, POV = privately owned vehicles.
[c] The bridge was closed to personal vehicular traffic in 2017 and reopened in April 2018.

2.2 STUDY DATA

TTI processed and analyzed two types of data. The first dataset included actual traffic volume data that were used in this study for comparison purposes. The second dataset included passively collected data that were evaluated, as described in Chapter 3. The following sections describe the two types of data along with the corresponding source agencies.

2.2.1 Traffic Volume Data

For the main study focus, the POEs, TTI requested and obtained traffic volume data from the following agencies:

- **U.S. Customs and Border Protection (CBP)**. TTI requested northbound traffic volumes from CBP officials via email. CBP captures the number of and time that vehicles arrive at primary inspection booths at all land POEs. This information is captured by the license plate number of each vehicle. Researchers received monthly northbound POV and COM traffic volumes for each POE. Table 2 shows the POV and COM traffic volumes at each POE by direction of travel.
- **Local public agencies**. TTI requested traffic volume data from Cameron County and the cities of Del Rio, Donna, Eagle Pass, El Paso, Laredo, McAllen-Hidalgo, and Pharr. These agencies provided TTI with traffic volumes for the southbound direction of 15 POEs (Table 2).

For completeness, TTI requested traffic volume data for the years 2016 and 2017; however, only the 2017 data were used in the analysis because StreetLight Data provided passively collected data for only year 2017. Because some of the aforementioned agencies provided disaggregated traffic volumes at the daily or monthly levels, researchers calculated the AADT at each POE using the American Association of State Highway and Transportation Officials (AASHTO) formula (*1*), which is also recommended in the *Traffic Monitoring Guide* (*2*):

$$AADT = \frac{1}{7}\sum_{i=1}^{7}\left[\frac{1}{12}\sum_{j=1}^{12}\left(\frac{1}{n}\sum_{k=1}^{n} Volume_{i,j,k}\right)\right] \qquad (1)$$

Where:

Volume = daily traffic volume for day *k*, of day-of-week *i*, and month *j*.
i = day of the week (1, 2, …, 7).
j = month of the year (1, 2, …, 12).
k = 1 when the day is the first occurrence of that day of the week in a month, and 4 when it is the fourth day of the week.
n = the number of days of that day of the week during that month (usually between one and five, depending on the amount of missing data).

Table 2 provides directional POV and COM traffic volumes at the study POEs. Note that the table is not exhaustive and may not include all southbound traffic volumes that various local agencies may collect. It only includes southbound volume data provided by a certain number of agencies that TTI contacted during this research project.

Table 2. Traffic Volumes at Ports of Entry.

No	Port of Entry Description	2017 Traffic Volume Data							
		Northbound Traffic				Southbound Traffic			
		AADT (POV)	AADT (COM)	AADT (POV+COM)	Source Agency	AADT (POV)	AADT (COM)	AADT (POV+COM)	Source Agency
1	Brownsville B&M Bridge	4,567		4,567	CBP				
2	Brownsville Gateway International Bridge	3,540		3,540	CBP				
3	Brownsville Veterans International Bridge	3,842	554	4,396	CBP	4,265	607	4,872	Cameron County
4	Del Rio-Ciudad Acuna Intl. Bridge	4,247	203	4,450	CBP	4,371	188	4,559	City of Del Rio
5	Donna Rio Bravo International Bridge	1,794		1,794	CBP	1,552		1,552	City of Donna
6	Eagle Pass Bridge I	3,246		3,246	CBP	3,709	1	3,710	City of Eagle Pass
7	Eagle Pass Camino Real International Bridge (Bridge II)	4,021	464	4,485	CBP	3,518	454	3,972	City of Eagle Pass
8	El Paso Bridge of the Americas (BOTA)	10,885	554	11,439	CBP				
9	El Paso del Norte Bridge	8,214		8,214	CBP				
10	El Paso Stanton Bridge					3,310		3,310	City of El Paso
11	El Paso Ysleta-Zaragoza Bridge	10,625	1,395	12,020	CBP	7,749	1,307	9,056	City of El Paso
12	Fort Hancock El Porvenir Bridge	215		215	CBP				
13	Free Trade Bridge	1,392	70	1,462	CBP	1,086	45	1,131	Cameron County
14	Lake Amistad Dam Crossing	160		160	CBP				
15	Lake Falcon Dam Crossing	269		269	CBP				
16	Laredo Gateway to the Americas Bridge (Laredo Bridge I)					2,456		2,456	City of Laredo
17	Laredo Juarez-Lincoln Bridge (Laredo Bridge II)	12,866		12,866	CBP	9,883	106	9,989	City of Laredo
18	Laredo World Trade Bridge		4,555	4,555	CBP		4,628	4,628	City of Laredo
19	Laredo-Colombia Solidarity Bridge	902	1,361	2,263	CBP	229	826	1,055	City of Laredo
20	Los Ebanos Ferry	81		81	CBP				
21	McAllen Anzalduas International Bridge	2,829		2,829	CBP	2,412	4	2,416	City of McAllen
22	McAllen-Hidalgo-Reynosa Bridge	6,095		6,095	CBP	7,685	72	7,757	City of McAllen
23	Pharr-Reynosa Bridge	3,133	1,677	4,810	CBP	2,237	1,590	3,827	City of Pharr
24	Presidio Bridge	1,893	24	1,917	CBP				
25	Progreso Bridge	1,542	144	1,686	CBP				
26	Rio Grande City-Camargo Bridge	1,018	103	1,121	CBP				
27	Roma-Ciudad Miguel Aleman Bridge	1,933	21	1,954	CBP				
28	Tornillo-Guadalupe Bridge	681		681	CBP				

For the second study area (i.e., count locations within three border TxDOT districts), TTI extracted traffic volume data from TxDOT's MS2 web platform (*3*). MS2 contains data from continuous count stations and short-duration counts. MS2 hosts TxDOT's Statewide Traffic Analysis and Reporting System (STARS II) database (*4*). STARS II is a data analysis and reporting database with detailed traffic data and statistics. It contains traffic data that TxDOT submits to the Federal Highway Administration (FHWA) for Highway Performance Monitoring System reporting purposes. TxDOT continuously updates STARS II with new traffic data as they become available.

TTI initially extracted both permanent and short-term traffic volume data for more than 6,000 roadway locations within the three districts of interest. After downloading the data, the researchers processed and filtered out counts that were missing at least one of the following attributes: station ID, latitude, longitude, rural/urban designation, roadway functional class, and count type (i.e., permanent or short term). In addition, TTI excluded counts that had an AADT between 0–400 vehicles/day (vpd). The reason for excluding these low-volume roads is because StreetLight Data did not produce 2017 AADT estimates below 400 vpd. The final clean dataset contained 4,643 records, of which 35 were permanent stations and the remaining 4,608 were short-term counts. In the case of permanent stations, TTI calculated the AADT using Equation 1.

Table 3 shows descriptive statistics of permanent and short-term traffic volumes by rural/urban designation and roadway functional class for all three TxDOT districts. Table 4, Table 5, and Table 6 show the same descriptive statistics for the El Paso, Laredo, and Pharr Districts, respectively.

Table 3. Descriptive Statistics of Traffic Volume Data for All Three TxDOT Districts.

Rural/ Urban	***Functional Class***	***Number of Sites***			***Average AADT***			***St. Deviation AADT***		
		Perm.	***Short Term***	***All***	***Perm.***	***Short Term***	***All***	***Perm.***	***Short Term***	***All***
Rural	1) Interstate	4	10	14	18,975	17,151	17,672	3,231	9,611	8,191
	2) Principal Arterial—Freeways	2	—	2	4,873	—	4,873	1,042	—	1,042
	3) Principal Arterial—Other	11	216	227	12,664	6,097	6,415	18,693	5,042	6,453
	4) Minor Arterial	3	157	160	6,083	3,625	3,671	5,276	3,958	3,979
	5) Major Collector	—	332	332	—	1,690	1,690	—	1,561	1,561
	6) Minor Collector	—	43	43	—	1,044	1,044	—	570	570
	7) Local	—	8	8	—	2,020	2,020	—	2,002	2,002
	Total Rural	**20**	**766**	**786**	**12,160**	**3,498**	**3,719**	**14,487**	**4,320**	**5,013**
Urban	1) Interstate	3	29	32	91,999	50,636	54,513	76,842	51,862	54,409
	2) Principal Arterial—Freeways	2	17	19	36,544	33,435	33,762	35,132	36,783	35,668
	3) Principal Arterial—Other	8	1,044	1,052	19,962	18,081	18,096	13,121	9,987	10,008
	4) Minor Arterial	1	798	799	15,080	8,681	8,689	NA	6,069	6,069
	5) Major Collector	—	1,219	1,219	—	5,707	5,707	—	5,404	5,404
	6) Minor Collector	—	45	45	—	3,137	3,137	—	2,412	2,412
	7) Local	1	690	691	1,523	1,978	1,977	NA	2,709	2,707
	Total Urban	**15**	**3,842**	**3,857**	**35,026**	**9,449**	**9,549**	**44,180**	**11,013**	**11,421**
Grand Total		**35**	**4,608**	**4,643**	**21,960**	**8,460**	**8,562**	**32,447**	**10,446**	**10,834**

Note: "—" mean not available and "NA" mean not applicable.

Table 4. Descriptive Statistics of Traffic Volume Data for the El Paso TxDOT District.

Rural/ Urban	Functional Class	Number of Sites			Average AADT			St. Deviation AADT		
		Perm.	Short Term	All	Perm.	Short Term	All	Perm.	Short Term	All
Rural	1) Interstate	1	4	5	15,312	20,631	19,567	NA	5,895	5,632
	2) Principal Arterial—Freeways	—	—	—	—	—	—	—	—	—
	3) Principal Arterial—Other	1	32	33	2,460	3,895	3,851	NA	4,776	4,707
	4) Minor Arterial	1	34	35	1,121	5,626	5,497	NA	7,031	6,968
	5) Major Collector	—	47	47	—	1,598	1,598	—	1,375	1,375
	6) Minor Collector	—	11	11	—	1,325	1,325	—	494	494
	7) Local	—	7	7	—	2,215	2,215	—	2,079	2,079
	Total Rural	**3**	**135**	**138**	**6,298**	**3,730**	**3,786**	**7,835**	**5,543**	**5,576**
Urban	1) Interstate	3	14	17	91,999	73,781	76,996	76,842	62,788	63,186
	2) Principal Arterial—Freeways	2	12	14	36,544	27,195	28,531	35,132	31,983	31,177
	3) Principal Arterial—Other	—	477	477	—	20,175	20,175	—	10,341	10,341
	4) Minor Arterial	1	541	542	15,080	8,511	8,523	NA	6,049	6,050
	5) Major Collector	—	590	590	—	6,123	6,123	—	5,764	5,764
	6) Minor Collector	—	10	10	—	4,698	4,698	—	2,752	2,752
	7) Local	1	651	652	1,523	1,965	1,965	NA	2,703	2,701
	Total Urban	**7**	**2,295**	**2,302**	**52,241**	**8,944**	**9,076**	**60,871**	**11,760**	**12,378**
Grand Total		**10**	**2,430**	**2,440**	**38,458**	**8,654**	**8,777**	**54,556**	**11,564**	**12,157**

Note: "—" means not available and "NA" means not applicable.

Table 5. Descriptive Statistics of Traffic Volume Data for the Laredo TxDOT District.

Rural/ Urban	*Functional Class*	*Number of Sites*			*Average AADT*			*St. Deviation AADT*		
		Perm.	*Short Term*	*All*	*Perm.*	*Short Term*	*All*	*Perm.*	*Short Term*	*All*
Rural	1) Interstate	3	6	9	20,196	14,831	16,619	2,591	11,370	9,470
	2) Principal Arterial—Freeways	—	—	—	—	—	—	—	—	—
	3) Principal Arterial—Other	7	122	129	4,497	4,824	4,807	1,679	2,385	2,349
	4) Minor Arterial	1	55	56	5,501	2,711	2,761	NA	1,679	1,705
	5) Major Collector	—	116	116	—	1,920	1,920	—	1,889	1,889
	6) Minor Collector	—	21	21	—	1,046	1,046	—	669	669
	7) Local	—	—	—	—	—	—	—	—	—
	Total Rural	**11**	**320**	**331**	**8,870**	**3,348**	**3,531**	**7,486**	**3,233**	**3,575**
Urban	1) Interstate	—	6	6	—	51,875	51,875	—	25,835	25,835
	2) Principal Arterial—Freeways	—	1	1	—	13,612	13,612	—	NA	NA
	3) Principal Arterial—Other	3	148	151	25,142	15,916	16,099	11,169	11,622	11,649
	4) Minor Arterial	—	59	59	—	9,277	9,277	—	8,167	8,167
	5) Major Collector	—	90	90	—	6,324	6,324	—	5,314	5,314
	6) Minor Collector	—	6	6	—	1,983	1,983	—	1,546	1,546
	7) Local	—	5	5	—	5,564	5,564	—	6,196	6,196
	Total Urban	**3**	**315**	**318**	**25,142**	**12,180**	**12,302**	**11,169**	**12,062**	**12,103**
Grand Total		**14**	**635**	**649**	**12,357**	**7,729**	**7,829**	**10,503**	**9,841**	**9,870**

Note: "—" mean not available and "NA" mean not applicable.

Table 6. Descriptive Statistics of Traffic Volume Data for the Pharr TxDOT District.

Rural/ Urban	Functional Class	Number of Sites			Average AADT			St. Deviation AADT		
		Perm.	Short Term	All	Perm.	Short Term	All	Perm.	Short Term	All
Rural	1) Interstate	—	—	—	—	—	—	—	—	—
	2) Principal Arterial—Freeways	2	—	2	4,873	—	4,873	1,042	NA	1,042
	3) Principal Arterial—Other	3	62	65	35,122	9,737	10,909	26,392	6,889	9,789
	4) Minor Arterial	1	68	69	11,625	3,363	3,483	NA	2,659	2,821
	5) Major Collector	—	169	169	—	1,558	1,558	—	1,335	1,335
	6) Minor Collector	—	11	11	—	761	761	—	224	224
	7) Local	—	1	1	—	654	654	—	NA	NA
	Total Rural	**6**	**311**	**317**	**21,123**	**3,552**	**3,885**	**22,806**	**4,687**	**5,961**
Urban	1) Interstate	—	9	9	—	13,806	13,806	—	11,213	11,213
	2) Principal Arterial—Freeways	—	4	4	—	57,109	57,109	—	49,308	49,308
	3) Principal Arterial—Other	5	419	424	16,855	16,462	16,467	14,378	8,385	8,452
	4) Minor Arterial	—	198	198	—	8,966	8,966	—	5,366	5,366
	5) Major Collector	—	539	539	—	5,149	5,149	—	4,951	4,951
	6) Minor Collector	—	29	29	—	2,838	2,838	—	2,243	2,243
	7) Local	—	34	34	—	1,689	1,689	—	1,623	1,623
	Total Urban	**5**	**1,232**	**1,237**	**16,855**	**9,692**	**9,721**	**14,378**	**9,008**	**9,039**
Grand Total		**11**	**1,543**	**1,554**	**19,183**	**8,454**	**8,530**	**18,647**	**8,676**	**8,820**

Note: "—" mean not available and "NA" mean not applicable.

2.2.2 Passively Collected Data

Streetlight Data provided TTI with unscaled, uncalibrated trip counts as well as AADT estimates for the two study areas. Note that the collection of probe data and estimation of AADT for roads at border crossings has some difficulties compared to (inland) roads that are far from border regions. For example, when cell phones switch data providers between the two countries, unusual probe data points may be created. Further, long vehicle queues at border crossings may result in false short trips. For the first study area (i.e., POEs), StreetLight Data provided the following data:

- **GPS COM Trip Count**: Raw uncalibrated count of GPS-based commercial vehicle trips in 2017 (*5*). Every raw count included the sum of all GPS-equipped commercial vehicles that traveled along a roadway segment throughout the year. Data were separately provided for the two directions of travel, where applicable, at select POEs. Navigation-GPS data originate from connected cars, smartphones using GPS navigation, and connected commercial trucks. Navigation-GPS data are precise, vehicle specific, and the only data source that offers specific metrics for commercial trucks.
- **GPS POV Trip Count**: Raw uncalibrated count of GPS-based private vehicle trips in year 2017. Every raw count included the sum of all GPS-equipped private vehicles that traveled along a roadway segment throughout the year. Data were separately provided for the two directions of travel, where applicable, at select POEs.
- **LBS POV Trip Count**: Raw uncalibrated count of LBS-based private vehicle trips for the period 1/1/2017–4/30/2017 (*6*). Every raw count included the sum of all private vehicles (with LBS devices) that traveled along a roadway segment during this four-month period. Data were separately provided for the two directions of travel, where applicable, at select POEs. LBS data originate from smartphone applications that use opt-in location-based services. The LBS data include all personal travel modes, such as cars, trucks, bikes, and pedestrians.
- **AADT**: AADT estimates for 17 POEs in year 2017. Each AADT estimate captured the total traffic volume in both directions of travel (i.e., sum of northbound and southbound traffic volumes). StreetLight Data developed nearly all the analytics for estimating 2017 AADT values at no cost to this CIITR research project. TTI did provide StreetLight Data with sample POE traffic volume data for model calibration purposes. According to StreetLight Data, the sample POE traffic volume data were not used to calibrate its AADT estimation models.

For the second study area, which included 4,643 count locations within the three border TxDOT districts, TTI downloaded the following data for each location using StreetLight's InSight tool:

- **GPS COM Trip Count**: Raw uncalibrated count of GPS-based commercial vehicle trips in both directions of travel in 2017. Every raw count included the sum of all GPS devices that traveled along a roadway segment throughout the year.
- **GPS POV Trip Count**: Raw uncalibrated count of GPS-based private vehicle trips in both directions of travel in 2017. Every raw count included the sum of all GPS devices that traveled along a roadway segment throughout the year.

- **LBS POV Trip Count**: Raw uncalibrated count of LBS-based private vehicle trips in both directions of travel in 2017. Every raw count included the sum of all LBS devices that traveled along a roadway segment throughout the year.
- **AADT**: AADT estimate that captures the total traffic volume in both directions of travel, if available.

StreetLight Data follows three main steps to develop AADT estimates (*7*):

- **Step 1**: Process and combine GPS and LBS data that StreetLight obtains from various data providers.
- **Step 2**: Normalize GPS and LBS trip counts (derived from Step 1) using non-traffic data, such as U.S. Census socioeconomic and demographic data.
- **Step 3**: Calibrate the estimates developed in Step 2 using machine learning algorithms. StreetLight Data uses actual traffic volume data that public agencies collect primarily from continuous count stations that are permanently installed at select locations of the network.

All passively collected data and the details of the traffic volume estimation models are the intellectual property of StreetLight Data and considered confidential.

CHAPTER 3: EVALUATION OF PASSIVELY COLLECTED DATA

3.1 INTRODUCTION

This chapter describes the analysis performed to determine (a) the penetration rate of raw uncalibrated trip samples, and (b) the accuracy of AADT estimates developed by StreetLight using probe and other types of data. For clarity, the analysis was separately conducted for the two study areas. To address the research questions listed in Section 1.1 and identify potential hidden trends in the data, TTI performed the analysis and aggregated the results, where applicable, by mobile device type (GPS versus LBS), traffic type (POV versus COM), TxDOT district, roadway functional class, and rural/urban designation. Sections 3.2 and 3.3 present the results for the two study areas, respectively.

3.2 FIRST STUDY AREA: PORTS OF ENTRY

The penetration rate analysis and the evaluation of StreetLight Data AADT estimates are presented in Subsections 3.2.1 and 3.2.2, respectively.

3.2.1 Penetration Rate

TTI calculated the penetration rate of the three different types of trip samples (GPS COM, GPS POV, and LBS POV trip counts) as well as the total PR for each direction of travel at each POE as follows:

$$PR_{i,GPS\ COM} = \frac{GPS\ COM\ Trip\ Count_i}{AADT_{i,Trucks} \times 365\ days} \tag{2}$$

$$PR_{i,GPS\ POV} = \frac{GPS\ POV\ Trip\ Count_i}{AADT_{i,POV} \times 365\ days} \tag{3}$$

$$PR_{i,LBS\ POV} = \frac{LBS\ POV\ Trip\ Count_i}{AADT_{i,POV} \times 120\ days} \tag{4}$$

$$PR_{i,T} = \frac{GPS\ COM\ Trip\ Count_i + GPS\ POV\ Trip\ Count_i + LBS\ POV\ Trip\ Count_i \times \frac{365}{120}}{(AADT_{i,Trucks} + AADT_{i,POV}) \times 365\ days} \tag{5}$$

Where:

$PR_{i,GPS\ COM}$	=	penetration rate of GPS COM trip count at POE *i*.
$PR_{i,GPS\ POV}$	=	penetration rate of GPS POV trip count at POE *i*.
$PR_{i,LBS\ POV}$	=	penetration rate of LBS POV trip count at POE *i*.
$PR_{i,T}$	=	total penetration rate at POE *i*.
$GPS\ COM\ Trip\ Count_i$	=	raw uncalibrated number of GPS COM trips at POE *i* in 2017 (365 days).

GPS POV Trip Count$_i$	=	raw uncalibrated number of GPS POV trips at POE *i* in 2017 (365 days).
LBS POV Trip Count$_i$	=	raw uncalibrated number of LBS POV trips at POE *i* between 1/1/2017 and 4/30/2017 (120 days).
$AADT_{i,Trucks}$	=	truck AADT provided by CBP and local agencies (see Section 2.1) for POE *i* in 2017.
$AADT_{i,POV}$	=	POV AADT provided by CBP and local agencies (see Section .2.1) for POE *i* in 2017.

Penetration rates were calculated when both traffic volume data (i.e., AADT) and probe-based trip counts were available for each direction of travel at each POE. In other words, if the AADT or a trip count was not available for a given direction of a POE, the penetration rate could not be calculated. After determining directional penetration rates at each POE, where applicable, the researchers calculated average penetration rates separately for the northbound traffic and the southbound traffic across all POEs. To ensure data confidentiality, the results of the analysis are aggregated and presented by direction of travel in Table 7.

Table 7. Average Penetration Rate of GPS and LBS Mobile Device Samples at POEs.

Direction of Travel	***Average Penetration Rate (Sample Size*)***			
	GPS COM	***GPS POV***	***LBS POV***	***Total (GPS COM + GPS POV + LBS POV)***
Southbound	20.56% (6)	0.06% (8)	0.84% (8)	1.23% (9)
Northbound	3.22% (13)	0.02% (25)	0.85% (23)	0.99% (26)
Grand Average	**8.70% (19)**	**0.03% (33)**	**0.85% (31)**	**1.06% (35)**

* The red numbers in parentheses denote the total number of data points (i.e., sample size) that were included in the calculation of each average penetration rate. Each data point corresponds to the northbound or the southbound traffic of each POE.

The most important results from the POE penetration rate analysis are summarized below:

- The grand average penetration rate for all three mobile device types combined is 1.06 percent.
- The total average penetration rate in the southbound direction (1.23 percent) is higher than that in the northbound direction (0.99 percent). Noteworthy is that GPS COM data have a significantly higher penetration rate in the southbound direction (20.56 percent) than in the northbound direction (3.22 percent).
- Among all three mobile device types, GPS COM data have the highest grand average penetration rate (8.7 percent), followed by LBS POV data (0.85 percent). In the absence of detailed information about the amount and quality of data that StreetLight obtains from various external data providers, it may be difficult to draw safe conclusions about this finding. A potential explanation is that drivers of commercial vehicles rely more on GPS navigation devices that tend to be more precise than LBS devices, particularly in rural areas with limited or no internet coverage.
- The grand average penetration rate of GPS POV data is negligible (0.03 percent). A potential explanation behind this finding may be attributed to the extensive and increasing use of smartphone devices over the last few years by owners of private

vehicles for navigation purposes, thus essentially limiting the use of traditional GPS navigation devices.

- The average penetration rate of LBS POV data is nearly the same (≈0.85 percent) in both directions of travel.

3.2.2 AADT Accuracy

StreetLight Data provided TTI with bi-directional AADT estimates for 17 POEs. Each AADT estimate captured both commercial and privately owned vehicle traffic in both directions of travel. Of all 17 AADT estimates, TTI evaluated the accuracy of 10 estimates for which northbound and southbound traffic volumes were provided by CBP and other local agencies (see Section 2.2.1). To determine the accuracy of StreetLight AADT estimates, TTI calculated the following metrics:

$$MSD\ (vehicles) = \frac{1}{n}\sum_{i=1}^{n}\left(AADT_{Estimated,i} - AADT_{Observed,i}\right) \tag{6}$$

$$MAD\ (vehicles) = \frac{1}{n}\sum_{i=1}^{n}\left(\left|AADT_{Estimated,i} - AADT_{Observed,i}\right|\right) \tag{7}$$

$$MAPE\ (\%) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\left|AADT_{Estimated,i} - AADT_{Observed,i}\right|}{AADT_{Observed,i}}\right) \times 100 \tag{8}$$

$$ACV\ (\%) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Standard\ Deviation(AADT_{Estimated,i}, AADT_{Observed,i})}{(AADT_{Estimated,i} + AADT_{Observed,i})/2}\right) \times 100 \tag{9}$$

Where:

MSD	=	mean signed difference.
MAD	=	mean absolute difference.
MAPE	=	mean absolute percent error.
ACV	=	average coefficient of variation.
$AADT_{Estimated,\,i}$	=	StreetLight AADT estimate for the i^{th} site.
$AADT_{Observed,\,i}$	=	observed AADT for the i^{th} site.
n	=	total number of sites (10 POEs) included in the evaluation.

The AADT accuracy metrics are aggregated and presented in Table 8 for three AADT groups, as well as for all 10 POEs together (last row).

Table 8. Accuracy Measures of StreetLight Data AADT Estimates.

AADT Range	*Number of POEs*	*MSD*	*MAD*	*MAPE*	*ACV*
0–5,000	4	−323	1,017	33.6%	26.7%
5,001–10,000	4	−1,885	1,885	26.6%	23.8%
>10,000	2	−8,224	8,224	44.7%	40.7%
Total	**10**	**−2,528**	**2,806**	**33.0%**	**28.3%**

The MAPE and ACV for all 10 POEs are 33.0 percent and 28.3 percent, respectively. Overall, the MAPEs (33.6 percent and 26.6 percent) for the first two AADT groups are smaller than the corresponding errors (68 percent and 58 percent) reported in a similar study conducted in 2017 using data from Minnesota (*7*). Another finding is that StreetLight Data AADT estimates tend to be lower (in magnitude) than the observed AADT values. Of all 10 AADT estimates included in the analysis, nine are lower than the observed AADTs. One potential explanation behind this finding is the small penetration rate of mobile devices at POEs.

TTI developed a scatterplot (Figure 2) between observed AADT values and StreetLight Data AADT estimates and fitted two trendlines to (a) better understand the relationship between the observed and the estimated AADT, and (b) explore if AADT estimates can be further improved. The observed AADT was the dependent variable (y axis), and the StreetLight Data AADT estimate was the independent variable (x axis). The chart shows 10 data points that correspond to the 10 POEs of the study, a black dichotomous line (y=x), and two dotted trendlines that have been fitted to the data:

- Linear (blue) trendline without intercept: $y=1.4917x$, $R^2=0.8207$.
- Nonlinear second-degree polynomial (red) trendline without intercept: $y=6*10^{-5}x^2+0.9329x$, $R^2=0.8572$.

The chart and the R^2 values suggest that the relationship between observed versus estimated AADT values is nonlinear at the 10 POEs. The two equations can be used to further adjust the StreetLight Data AADT estimates and improve their accuracy. All parameter coefficients are statistically significant at the 95 percent confidence level. The slope of the blue line (1.4917) confirms that StreetLight Data AADT estimates underestimate the observed AADT, as previously explained. Overall, as the use of mobile devices continues to increase and StreetLight Data obtains data from more vendors and enhances its traffic volume prediction methods, the accuracy of AADT estimates is expected to improve.

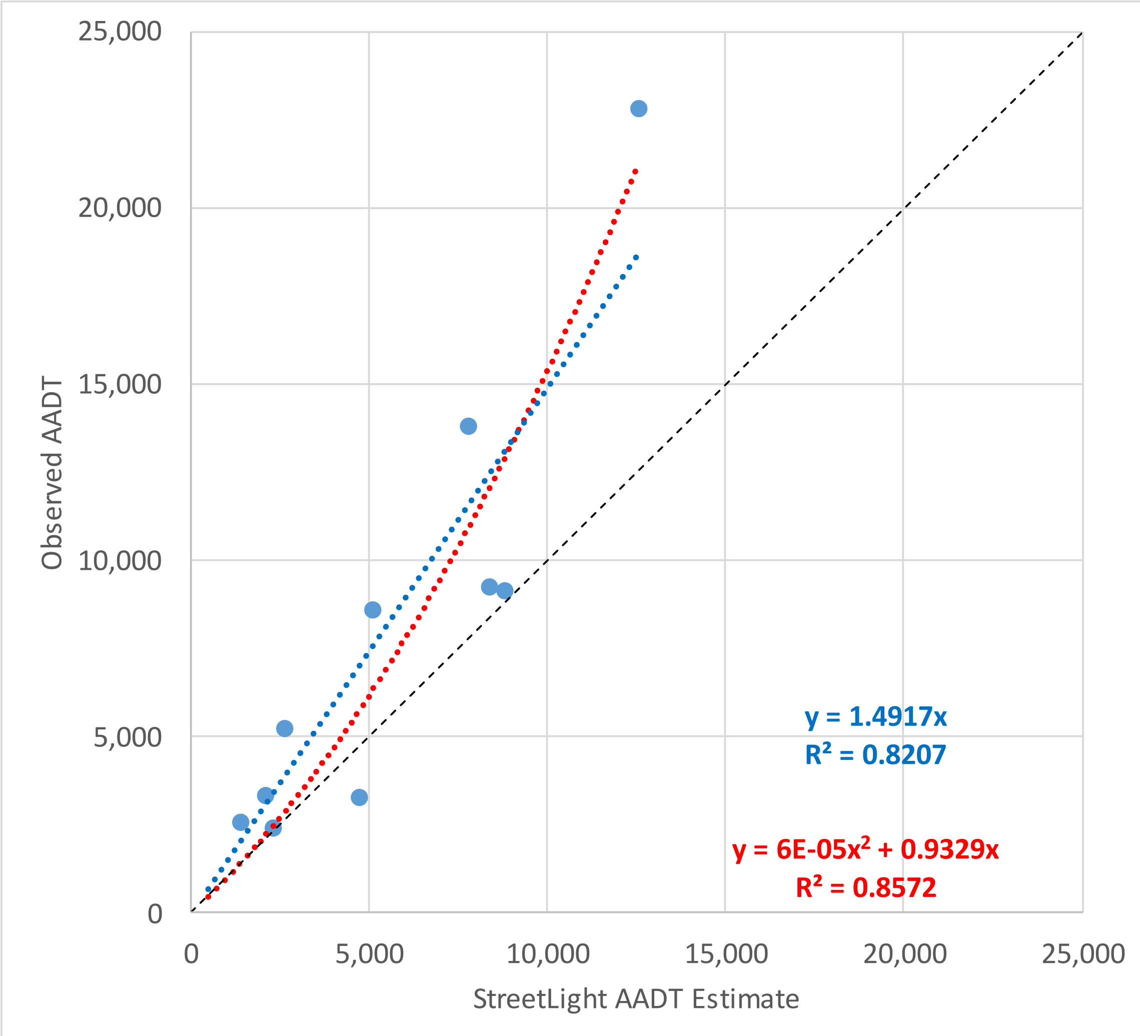

Figure 2. Observed versus Estimated AADT at POEs.

3.3 SECOND STUDY AREA: COUNT LOCATIONS IN BORDER REGIONS

In the absence of POV- and COM-specific MS2 traffic volume data, TTI calculated the total penetration rate of all mobile devices at each count location as follows:

$$PR_{i,Total} = \frac{GPS\ COM\ Trip\ Count_i + GPS\ POV\ Trip\ Count_i + GPS\ POV\ Trip\ Count_i}{AADT_{i,\ TxDOT} \times 365\ days} \quad (10)$$

Where:

$PR_{i,Total}$	= total penetration rate at site *i*.
$GPS\ COM\ Trip\ Count_i$	= raw uncalibrated number of GPS COM trips at site *i* in 2017 (365 days).
$GPS\ POV\ Trip\ Count_i$	= raw uncalibrated number of GPS POV trips at site *i* in 2017 (365 days).

LBS POV Trip Count$_i$ = raw uncalibrated number of LBS POV trips at site *i* in 2017 (365 days).

$AADT_{i,\ TxDOT}$ = 2017 AADT extracted from TxDOT's MS2 website for site *i*.

In addition, TTI compared StreetLight Data AADT estimates against permanent and short-duration traffic volume data extracted from TxDOT's MS2 platform. The AADT values calculated using permanent station data are typically considered to be representative of the actual traffic volumes at the permanent sites, assuming the dataset is complete and does not contain erroneous values. Therefore, under certain circumstances, the AADT calculated from permanent sites can be used for validation purposes. However, the AADT derived from a short-duration count is a value that has been estimated (not calculated) by applying one or more seasonal adjustment factors to the ADT of the count. As a result, the short-duration AADT values have an inherent estimation error, which does not make them appropriate for validation purposes. In this study, they were merely used as a comparison device.

Each StreetLight Data AADT estimate captured the total traffic in both directions of travel. TTI calculated the metrics presented in the previous section (Equations 6–9) to compare TxDOT traffic-based AADT values against StreetLight Data probe-based AADT estimates. Table 9 shows these metrics aggregated by five traffic volume ranges that TTI developed based on TxDOT MS2 data. The sixth column (MAPE) shows that the AADT accuracy gradually improves from low- to high-volume roads. The lowest MAPEs (13 percent) are observed within the last AADT range (>50,000 vpd). The results show that the grand average penetration rate is 0.86 percent. For completeness, Table 10 shows the same metrics disaggregated by rural/urban designation within each AADT range. To shed light on whether and how the penetration rates and the accuracy of AADT estimates are affected by various geographic and other roadway-related factors, Table 11 through Table 15 present the results aggregated by rural/urban designation (Table 11); roadway functional class (Table 12); TxDOT district (Table 13); TxDOT district and functional class (Table 14); and functional class and rural/urban designation (Table 15).

Table 9. Penetration Rate and Accuracy of StreetLight Data AADT Estimates by AADT Range.

AADT Range (vpd)	Number of Counts	PR	MSD	MAD	MAPE	ACV
401-5,000	2,400	1.04%	897	1,082	72%	31%
5001-10,000	857	0.76%	753	2,173	30%	18%
10,001-20,000	861	0.61%	-548	3,434	24%	18%
20,001-50,000	501	0.62%	-4,679	6,273	23%	19%
>50,000	24	0.75%	-12,491	13,782	13%	11%
Grand Total	**4,643**	**0.86%**	**-68**	**2,345**	**50%**	**25%**

Table 10. Penetration Rate and Accuracy of StreetLight Data AADT Estimates by AADT Range and Rural/Urban Designation.

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

The penetration rate on rural roads (1.72 percent) is significantly higher than that on urban roads (0.68 percent). This trend is consistent within all AADT ranges (Table 10). Though the penetration rate varies by TxDOT district, from 0.69 percent to 1.88 percent, it is consistently higher on rural roads in all three TxDOT districts. This trend is also observed within the first six roadway functional classes. Urban local roads have a higher penetration rate of mobile devices compared to the eight rural local roadway locations, which not surprisingly have the second lowest penetration rate among all functional classes (Table 15).

Table 11. Penetration Rate and Accuracy of StreetLight Data AADT Estimates by Rural/Urban Designation.

Rural/ Urban	Number of Counts	PR	MSD	MAD	MAPE	ACV
Rural	786	1.72%	214	1,349	63%	28.63%
Urban	3,857	0.68%	-126	2,548	47%	24.04%
Grand Total	**4,643**	**0.86%**	**-68**	**2,345**	**50%**	**25%**

Table 12. Penetration Rate and Accuracy of StreetLight Data AADT Estimates by Roadway Functional Class.

Functional Class	Number of Counts	PR	MSD	MAD	MAPE	ACV
(1) Interstate	46	1.98%	-6186	8,159	29%	19.23%
(2) Principal Arterial—Other Freeways	21	1.04%	-4855	6,240	25%	18.48%
(3) Principal Arterial—Other	1,279	0.78%	-1809	3,598	29%	18.90%
(4) Minor Arterial	959	0.78%	-254	1,937	39%	21.32%
(5) Major Collector	1,551	1.02%	1319	2,001	59%	26.68%
(6) Minor Collector	88	0.93%	490	835	56%	27.49%
(7) Local	699	0.67%	769	1,067	83%	36.51%
Grand Total	**4,643**	**0.86%**	**-68**	**2,345**	**50%**	**25%**

Table 13. Penetration Rate and Accuracy of StreetLight Data AADT Estimates by TxDOT District.

TxDOT District	Number of Counts	PR	MSD	MAD	MAPE	ACV
El Paso	2,440	0.69%	106	2,501	53%	26%
Laredo	649	1.88%	469	2,343	61%	28%
Pharr	1,554	0.69%	-566	2,102	40%	21%
Grand Total	**4,643**	**0.86%**	**-68**	**2,345**	**50%**	**25%**

Table 14. Penetration Rate and Accuracy of StreetLight Data AADT Estimates by TxDOT District and Roadway Functional Class.

TxDOT District	Rural/ Urban	Number of Counts	PR	MSD	MAD	MAPE	ACV
El Paso	Rural	138	0.74%	-303	1,356	59%	29%
	Urban	2,302	0.69%	130	2,569	52%	26%
Laredo	Rural	331	2.66%	862	1,317	74%	31%
	Urban	318	1.06%	60	3,411	48%	24%
Pharr	Rural	317	1.17%	-236	1,380	54%	25%
	Urban	1,237	0.57%	-651	2,287	36%	20%
Grand Total		**4,643**	**0.86%**	**-68**	**2,345**	**50%**	**25%**

Table 15. Penetration Rate and Accuracy of StreetLight Data AADT Estimates by Roadway Functional Class and Rural/Urban Designation.

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

The main findings related to the accuracy of AADT estimates are summarized below:

- The grand average MAPE is 50 percent, which is lower than the corresponding MAPE (61 percent) reported in 2017 by Turner and Koeneman for Minnesota's transportation network (*7*). In general, the AADT accuracy consistently improves from lower to higher traffic volume roads.
- StreetLight Data AADT estimates tend to be higher than TxDOT AADT values (i.e., positive mean signed difference) for the first two AADT ranges, but this trend is reversed for the last three ranges (Table 9). Similarly, the mean signed difference is negative in the first four functional classes (Table 12) and becomes positive in the case of functional classes 5 through 7.
- The MAD gradually increases from low-volume roads to higher AADT roads (Table 9). Not surprisingly, this finding is also observed when the MAD is aggregated by roadway functional class (Table 12).
- StreetLight Data AADT estimates for urban roads tend to be more accurate than those for rural roads. This finding is consistent within functional classes 2 through 7, but the opposite trend is observed in the case of interstates (functional class 1); however, the sample size of functional class 1 is relatively small (46 count locations).

CHAPTER 4: FINDINGS AND CONCLUSIONS

This study examined the accuracy of probe-based AADT estimates at Texas-Mexico border crossings and counted roadways that are in proximity to the Mexican borders. In this project, StreetLight Data provided TTI with unscaled and uncalibrated GPS and LBS trip count data for commercial and privately owned vehicles, as well as probe-based AADT estimates for several locations in the two study areas. For each study area, TTI determined the penetration rate of mobile devices and compared StreetLight Data AADT estimates against traffic volume data collected by various state and local agencies in Texas. The main findings and conclusions from this study are summarized in this chapter.

The main research questions and the corresponding findings from this study are summarized below:

- **What is the penetration rate of mobile devices in the two study areas?**
    - The average penetration rate in the first study area, the POEs, was 1.06 percent. The penetration rate was higher (1.23 percent) in the southbound direction than in the northbound direction (0.99 percent) of the POEs.
    - The average penetration rate in the second study area (permanent and short-duration count locations within the three TxDOT districts) was 0.86 percent. The penetration rate on rural roads was significantly higher than that on urban roads. This trend was consistent within all AADT ranges. Though the penetration rate varied by TxDOT district, it was consistently higher on rural roads in all three TxDOT districts. This trend was also observed within the first six roadway functional classes. Urban local roads had a higher penetration rate of mobile devices compared to rural local roads, which not surprisingly had the lowest penetration rate among all roadway functional classes.
- **What is the penetration rate of mobile devices used in commercial vehicles versus privately owned vehicles?**
    - In the first study area, the penetration rates of GPS COM, LBS POV, and GPS POV mobile devices were 8.7 percent, 0.85 percent, and 0.03 percent, respectively. A potential explanation behind the relatively high capture rate of GPS COM devices is that it is easier for StreetLight Data to procure commercial truck data than LBS data. Noteworthy is that GPS COM devices had a significantly higher penetration rate in the southbound direction (20.56 percent) than in the northbound direction (3.22 percent) of the POEs. The small capture rate of GPS POV devices may be attributed to the extensive and increasing use of smartphone devices over the last few years by owners of private vehicles for navigation purposes, thus essentially limiting the use of more traditional GPS navigation devices.
- **What is the anticipated accuracy of probe-based AADT estimates in the two study areas?**
    - In the first study area, the MAPE that was estimated using data from 10 POEs was 33.0 percent. The StreetLight Data AADT estimates were lower than the observed AADT values at nine POEs. One potential explanation behind this finding is the fact that cell phone providers switch as vehicles cross the borders, creating (probe) data

anomalies. Another reason is that long vehicle queues (i.e., long wait times) at border crossings may be considered vehicle stops instead of single vehicle trips.
    - In the second study area, the grand average MAPE was 50 percent, which is lower than the corresponding MAPE (61 percent) reported in 2017 by Turner and Koeneman for Minnesota's transportation network (*7*).
- **Where do probe-based AADT estimates tend to be more accurate and why?**
    - In general, the AADT accuracy improved from low to high traffic volume roads. The AADT estimates were higher than TxDOT AADT values within the lower AADT ranges (401–50,000 vpd and 5,001–10,000 vpd), but this trend was reversed in the case of higher-volume roads (10,001–20,000, 20,001–50,000, >50,000 vpd). Similarly, the mean signed difference was negative in the first four functional classes and positive in the case of functional classes 5 through 7.
    - Overall, the AADT estimates for urban roads were more accurate (MAPE=47 percent) than those for rural roads (MAPE=63 percent). This finding was consistent within functional classes 2 through 7, but the opposite trend was observed in the case of interstates (functional class 1); however, the sample size within functional class 1 was relatively small (46 count locations).

As the use of mobile devices continues to increase and data providers continue to enhance their traffic volume prediction methods, the accuracy of AADT estimates is expected to improve. For example, the 2017 AADT estimates used in this project resulted in lower errors than those reported in a 2017 report that evaluated 2015 AADT estimates. Future evaluations of probe-based AADT estimates are needed using data from different states and regions that potentially have diverse traffic, geometric, demographic, socioeconomic, and weather characteristics. The ongoing FHWA pooled fund study "Independent Evaluation of Non-Traditional Methods to Obtain Annual Average Daily Traffic" (*8*) is expected to shed light on this topic and unravel other aspects and areas where improvements are needed.

The findings of this research may apply to other border regions across the United States. This presents an opportunity for technology transfer. In addition, obtaining AADT estimates from probe data could provide a common measure to assess the performance of traffic operations on both sides of border crossings. It could also yield time and cost savings for transportation agencies that either do not collect traffic volume data (e.g. Mexican agencies) or deploy expensive traffic equipment. It could also reduce safety risks to employees and contractors who typically go out in the field to install sensor devices in and on roadways. Further, it can assist agencies in meeting new federal requirements according to which States must have access to a series of Model Inventory Roadway Elements - Fundamental Data Elements, including AADT, for all public paved roads by 2026.

REFERENCES

1. *AASHTO Guidelines for Traffic Programs, 2nd Edition*. American Association of State Highway and Transportation Officials, Washington, DC, 2009.
2. *Traffic Monitoring Guide*. Federal Highway Administration, Washington, DC, October 2016.
3. MS2. Texas Department of Transportation, 2019. Accessed on August 31, 2019, at https://txdot.ms2soft.com/tcds/tsearch.asp?loc=Txdot&mod.
4. Statewide Traffic Analysis and Reporting System (STARS II). Texas Department of Transportation, 2019. Accessed on August 31, 2019, at https://www.txdot.gov/inside-txdot/division/transportation-planning/stars.html.
5. K. Gische. "Navigation-GPS Data." StreetLight Data, September 2019. Accessed on August 31, 2019, at https://support.streetlightdata.com/hc/en-us/articles/360018833651-Navigation-GPS-Data.
6. K. Gische. "Location-Based Services Data." StreetLight Data, September 2019. Accessed on August 31, 2019, at https://support.streetlightdata.com/hc/en-us/articles/360018833631-Location-Based-Services-Data.
7. S. Turner and P. Koeneman. *Using Mobile Device Samples to Estimate Traffic Volumes*. Final Report 2017-49, Prepared by the Texas A&M Transportation Institute for the Minnesota Department of Transportation, December 2017. Accessed on August 31, 2019, at http://mndot.gov/research/reports/2017/201749.pdf.
8. Federal Highway Administration. "Independent Evaluation of Non-Traditional Methods to Obtain Annual Average Daily Traffic." FHWA Pooled Fund Study, Performed by StreetLight Data Inc., the National Renewable Energy Lab, and Cambridge Systematics with the Texas A&M Transportation Institute, Washington, DC (ongoing project).