EVALUATING STREETLIGHT ESTIMATES OF ANNUAL AVERAGE DAILY TRAFFIC IN OREGON

Final Report

Oregon Department of Transportation

EVALUATING STREETLIGHT ESTIMATES OF ANNUAL AVERAGE DAILY TRAFFIC IN OREGON

Final Report

by

Josh Roll

for

Oregon Department of Transportation
Research Section & Traffic Data Section
555 13th Street NE
Salem OR 97301

and

Federal Highway Administration
1200 New Jersey Avenue SE
Washington, DC 20590

June 2019

Technical Report Documentation Page

1. Report No. OR-RD-19-11	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle EVALUATING STREETLIGHT ESTIMATES OF ANNUAL AVERAGE DAILY TRAFFIC IN OREGON		5. Report Date June 2019 6. Performing Organization Code
7. Author(s) Josh Roll		8. Performing Organization Report No.
9. Performing Organization Name and Address Oregon Department of Transportation Research Section 555 13th Street NE, Suite 1 Salem, OR 97301		10. Work Unit No. (TRAIS) 11. Contract or Grant No.
12. Sponsoring Agency Name and Address Oregon Dept. of Transportation Research Section 555 13th Street NE, Suite 1 Salem, OR 97301	Federal Highway Admin. 1200 New Jersey Avenue SE Washington, DC 20590	13. Type of Report and Period Covered Final Report 14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract: This report summarizes the evaluation of Streetlight Data's Annual Average Daily Traffic (AADT) product by comparing with Oregon Department of Transportation's automatic traffic recorder data and AADT estimates derived from short term counts in Bend, MPO. Using an evaluation methodology established by Turner et al. (2008), the report concludes that accuracy was measured to be 18% (N = 172) median (mean equals 25%) absolute percent error for the automatic traffic recorder comparison and 32% (N = 66) median (mean equals 59%) absolute percent error for the short term based AADT estimate comparison. Other measures of data quality were reasonable but the accessibility data quality measure would suffer for network wide (all segments in the network) data accessibility since the setting gates for Streetlight Data's process would be time consuming. Data quality is difficult to fully answer outside the context of a use of the data and its recommended that future evaluations apply the AADT product in a safety, vehicle miles traveled, or air quality analysis to determine the magnitude of difference between currently available data (from models or traditional traffic counts) and these third party data products. Future evaluations should also look at any potential costs savings of purchasing third party data compared with traditional traffic count data collection.

17. Key Words		18. Distribution Statement Copies available from NTIS, and online at www.oregon.gov/ODOT/TD/TP_RES/	
19. Security Classification (of this report) Unclassified	20. Security Classification (of this page) Unclassified	21. No. of Pages 35	22. Price

Technical Report Form DOT F 1700.7 (8-72) Reproduction of completed page authorized Printed on recycled paper

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS					APPROXIMATE CONVERSIONS FROM SI UNITS				
Symbol	When You Know	Multiply By	To Find	Symbol	Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH					**LENGTH**		
in	inches	25.4	millimeters	mm	mm	millimeters	0.039	inches	in
ft	feet	0.305	meters	m	m	meters	3.28	feet	ft
yd	yards	0.914	meters	m	m	meters	1.09	yards	yd
mi	miles	1.61	kilometers	km	km	kilometers	0.621	miles	mi
		AREA					**AREA**		
in^2	square inches	645.2	millimeters squared	mm^2	mm^2	millimeters squared	0.0016	square inches	in^2
ft^2	square feet	0.093	meters squared	m^2	m^2	meters squared	10.764	square feet	ft^2
yd^2	square yards	0.836	meters squared	m^2	m^2	meters squared	1.196	square yards	yd^2
ac	acres	0.405	hectares	ha	ha	hectares	2.47	acres	ac
mi^2	square miles	2.59	kilometers squared	km^2	km^2	kilometers squared	0.386	square miles	mi^2
		VOLUME					**VOLUME**		
fl oz	fluid ounces	29.57	milliliters	ml	ml	milliliters	0.034	fluid ounces	fl oz
gal	gallons	3.785	liters	L	L	liters	0.264	gallons	gal
ft^3	cubic feet	0.028	meters cubed	m^3	m^3	meters cubed	35.315	cubic feet	ft^3
yd^3	cubic yards	0.765	meters cubed	m^3	m^3	meters cubed	1.308	cubic yards	yd^3
~NOTE: Volumes greater than 1000 L shall be shown in m^3.									
		MASS					**MASS**		
oz	ounces	28.35	grams	g	g	grams	0.035	ounces	oz
lb	pounds	0.454	kilograms	kg	kg	kilograms	2.205	pounds	lb
T	short tons (2000 lb)	0.907	megagrams	Mg	Mg	megagrams	1.102	short tons (2000 lb)	T
		TEMPERATURE (exact)					**TEMPERATURE (exact)**		
°F	Fahrenheit	(F-32)/1.8	Celsius	°C	°C	Celsius	1.8C+32	Fahrenheit	°F

*SI is the symbol for the International System of Measurement

ACKNOWLEDGEMENTS

Thank you to the Technical Advisory Committee members listed below.

Name	Section	Position
Don Crownover	Traffic Data	Transportation Systems Monitoring Unit Team Leader
Jennifer Campbell	Transportation Data	HPMS Coordinator
Brian Dunn	Transportation Planning and Analysis	Section Manager
Alex Bettinardi	Transportation Planning and Analysis	Senior Integrates Analysis Engineer
Christina McDaniels-Wilson	Traffic Engineering Services	Highway Safety Engineer

DISCLAIMER

This document is disseminated under the sponsorship of the Oregon Department of Transportation and the United States Department of Transportation in the interest of information exchange. The State of Oregon and the United States Government assume no liability of its contents or use thereof.

The contents of this report reflect the view of the authors who are solely responsible for the facts and accuracy of the material presented. The contents do not necessarily reflect the official views of the Oregon Department of Transportation or the United States Department of Transportation.

The State of Oregon and the United States Government do not endorse products of manufacturers. Trademarks or manufacturers' names appear herein only because they are considered essential to the object of this document.

This report does not constitute a standard, specification, or regulation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

The ubiquity of smartphones has sprouted a new industry where firms now buy, process, package and sell user information that purports to offer information about aggregate travel activity. Oregon Department of Transportation (ODOT) is increasingly being propositioned by these firms but staff do not yet have a full understanding of the data products' capabilities or limitations. Additionally, the agency requires a standard method for evaluating these products to help make decisions on potential purchases where multiple vendor options exist.

The following report applies a methodology for evaluating a firm called Streetlight Data's Annual Average Daily Traffic (AADT) product by comparing that firm's estimates with data collected by ODOT's Traffic Monitoring Unit to determine the quality of the product. The comparison is presented using data quality metrics derived from established sources summarized in Turner et. al. (2008). This report offers some insight into the usability of a specific firm's AADT product by disclosing an objective assessment of the product based on these data quality metrics. As mentioned in Turner et al. (2008), the issue of data quality often times comes down to the end use of the data. Since this evaluation does not go further and employ these data in a use case such as a safety analysis or project planning effort, the evaluation is limited to the data quality metrics. Ideally these metrics offer agency staff, and others, insight into the usability of these data.

1.1 EVALUATION METRICS

In order to evaluate the products that Streetlight Data™ provides data quality metrics discussed in Turner (2004) are employed which include five metrics described in Table 2.1 below. The table includes the data quality element, a description of each element, as well as an example. These metrics were developed after synthesizing various sources from multiple fields including computer science, defense, information systems as well as intelligent transportation systems. Each of these data quality elements will be reported on for Streetlight Data's AADT product.

Table 1.1: Data Quality Metrics

Data Quality	Characteristics Description	Example Metric
Accuracy	A quality of that which is free of error. A qualitative assessment of freedom from error, with a high assessment corresponding to a small error. (FIPS Pub 11-3)	Percent of values that are correct when compared to the actual value. For example, the estimate of annual traffic compared to the actual value.
Completeness	Completeness is the degree to which values are present in the attributes that require them. (Data Quality Foundation)	Percent of data fields having values entered into them.
Validity	The quality of data that is founded on an adequate system of classification and is rigorous enough to compel acceptance. (DOD 8320.1-M)	Percent of recorded traffic count values that fall within operational expectations.
Timeliness	As a synonym for currency, timeliness represents the degree to which specified data values are up to date. (Data Quality Management and Technology)	Percent of data available within a specified threshold time frame (e.g., days, hours, minutes).
Accessibility (also referred to as usability)	Includes simple technical accessibility as well as ability to manipulate based on user needs. (Strong et al. 1997)	Percent of freeway centerline miles with sensor coverage and average sensor spacing.

*Adapted from Turner et al. 2004

1.2 STREETLIGHT PROCESS AND SOURCES

Streetlight is one of many emerging transportation technology firms that have entered the bigger data marketplace. These firms process and sell the location and movement data captured on mobile phones and global positional system (GPS) devices. The types of mobile data that Streetlight Data purchases and uses include data from location-based service enabled smartphone applications where the application collects the user's location at various times while the app is running and in some cases even when the app is not currently in use. Other mobile data come from GPS devices that both passenger and commercial vehicles use for navigation. These data are considered samples that are then combined with other sources of information necessary like US Census and traffic counts, to estimate more comprehensive travel activity measures.

2.0 ANNUAL AVERAGE DAILY TRAFFIC FOR VEHICLES

One of the products of interest provided by Streetlight is the AADT estimates for vehicles. The below section describes the data and methods used to derive these estimates.

2.1 DATA SOURCES AND METHODS

The 2017 Streetlight AADT data product is created using multiple data sources including location-based service data points, navigation-GPS for personal and commercial travel and US Census data among others like weather data and traffic counts from permanent counters throughout the United States. Below are short descriptions gleaned from Streetlights documentation that describes their methods for estimating AADT (Streetlight 2018).

- Location-based Services (LBS)

  Locations-based service (LBS) data are generated by smartphone applications that rely on a device's geographic location in the physical world, like a weather app, shopping app or social media app. These LBS data are used to determine trip ends including the home location of the device as well as the destinations. The firm employs data from 365 days to fully depict seasonal variations in travel patterns.

- Navigation GPS Trips

  Navigation GPS data are created from connected personal and commercial vehicles that use GPS navigation devices including those on smart phones. Similar to the LBS data, Streetlight Data uses observations from 365 days to fully depict seasonal variations in travel patterns.

- US Census

  LBS data are adjusted using information from US Census in order to represent all the travel within a given geography. An example from Streetlight's documentation states that," if ten devices in our sample "live" on a block with 100 people, each of those devices is scaled up by a factor of 10. If ten devices "live" on a block with 50 people, each is scaled by a factor of 5. ". Population density is also described as being taken into account but no other details are provided.

- Weather Data

  The documentation describes the use of weather data to account for areas with extreme precipitation events that might affect traffic patterns.

- Permanent Traffic Count Data

> The use of permanent traffic count data from 2,605 automatic traffic recorders are used to train a machine learning algorithm that employs the above mentioned data to estimate AADT. Based on the documentation, the counts data comes from states through the US and include both urban and rural roads.

Using the above data sources the firm purports to have tried using two methods of machine learning including an Ordinary Least Squares (OLS) regression approach and a Random Forest approach. The OLS regression approach was not chosen due to stated inaccuracy compared with the Random Forest though no documentation of the testing is provided.

The Random Forest method is a standard machine learning technique that can produce accurate estimates when used in predictive modeling but is less intuitive than regression models where independent variables can be explained using a theoretical foundation. Instead, Random Forest techniques split dependent data into training and test data with features, or single and grouped independent variables many times and average multiple decision trees with the goal of reducing variance and improving prediction power

Validation tests were constructed to test the accuracy of the model. A random k-fold test was done using all of the data but the number of folds was not described though common practice is 10 folds. Another cross validation test was done where each state was used as a unique fold. Results of these validation tests were not presented in the Streetlight Data documentation.

2.2 STREETLIGHT PLATFORM INPUTS/OUTPUTS AND USER PLATFORM

The above data sources and methods are available for users through a web platform that allows users to upload and download data as well as perform some analyses and visualization. The data input process requires shape file data to be formatted into an expected structure with specified attributes required for the upload to proceed properly.

The gate_width attribute is important of what Streetlight Data calls zones. The gate_width decides what trips, presumably from GPS and estimated trips data, are counted on a given network link. The gate width can be specified in two ways. The first is by using default values based on the functional classification of the road where higher class roads like interstates and principle arterials have wider gates compared to collectors and local streets. The second is by specifying them in the input data. The latter was done for this work since it was found the default values were not always wide enough on higher functional classification roads to 'catch' all the traffic. Figure 2.1 below shows a properly specified gate where the entire facility is captured including both directions of travel and an improperly specified gate where one side of the interstate is missed. These gate width polygon features (shown below) are an output of the web platform and are helpful in ensuring proper gate width and placement.

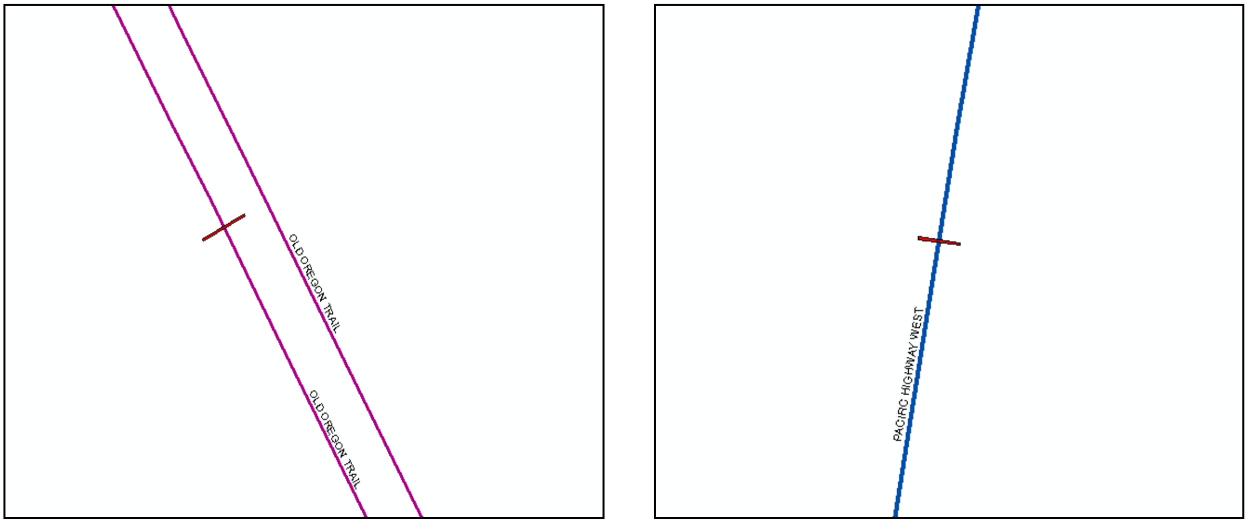

Figure 2.1: Example of proper (left panel) and improper (right panel) specification of gate width on Interstate 5 in Oregon

Since the purpose of this evaluation is to compare the 2017 AADT Streetlight Data product with traffic counts from permanent count locations that do collect traffic in both directions of travel, it's important to ensure that the Streetlight Data platform captures the entire facility where counts are collected.

Another important element related to the zone is the gate's location on the input network link of interest. By default the gate will be placed over the segments mid-point though it can also be specified in the input data. The gate location, if improperly located, can produce AADT estimates with high error. For example in Figure 2.2 below the default location falls over a complex interchange at the Interstate 5 and Interstate 84 junction and captures links other than those assigned to the permanent count location, in this case station 26-120. For this evaluation four such locations were found and were removed from the comparisons below. More work could have been done to more properly place the gate but that was not within the scope of this evaluation.

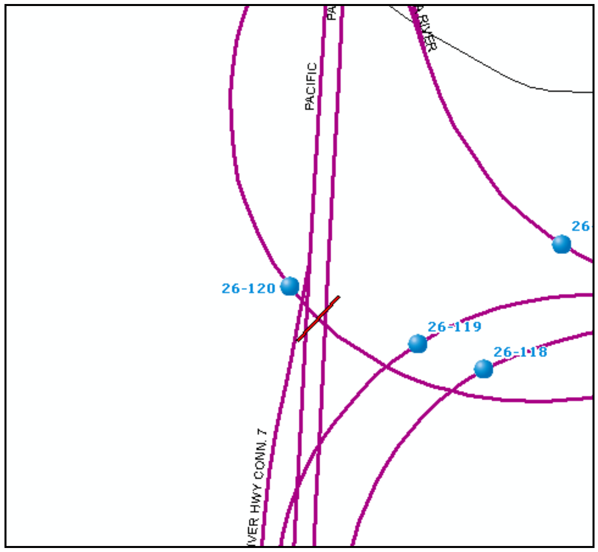

Figure 2.2: Example of poor location placement of gate at Interstate 5 and Interstate 84 interchange in Oregon

Once the user enters the network segments of interest and determines a few other toggles, the platform begins processing the request. For the 180 segments entered into the platform the process time was less than one hour.

2.3 ATR COMPARISON

The below section details the ODOT data used for comparison of the Streetlight AADT estimates followed by an evaluation of the data product using the data quality metrics described above.

2.3.1 ODOT ATR Data

Input data for the AADT evaluation included network links from the Table of Potential Samples (TOPS) which is ODOT's network line feature class data used in reporting to the Federal Highways Administration (FHWA) for purposes of accounting to the Highway Performance Monitoring Program (HPMS). In order to assess the accuracy of the Streetlight 2017 AADT estimates traffic count data from 180 automatic traffic recorders (ATR), or permanent count location are used. Table 2.1 below summarizes the count of ATR count and the associated TOPS segment length in miles.

Table 2.1: ODOT ATR Site Summary

Functional Classification	Number of ATRs	TOPS Length (mi.)
Interstate	42	83.3
Principal Arterial - Freeways and Expressways	10	12.6
Principal Arterial - Other	85	184.4
Minor Arterial	30	95.8
Major Collector	6	22.9

The figure below shows the relative locations of the ATRs throughout the state while also showing the functional classification system for the state and non-state (owned) systems.

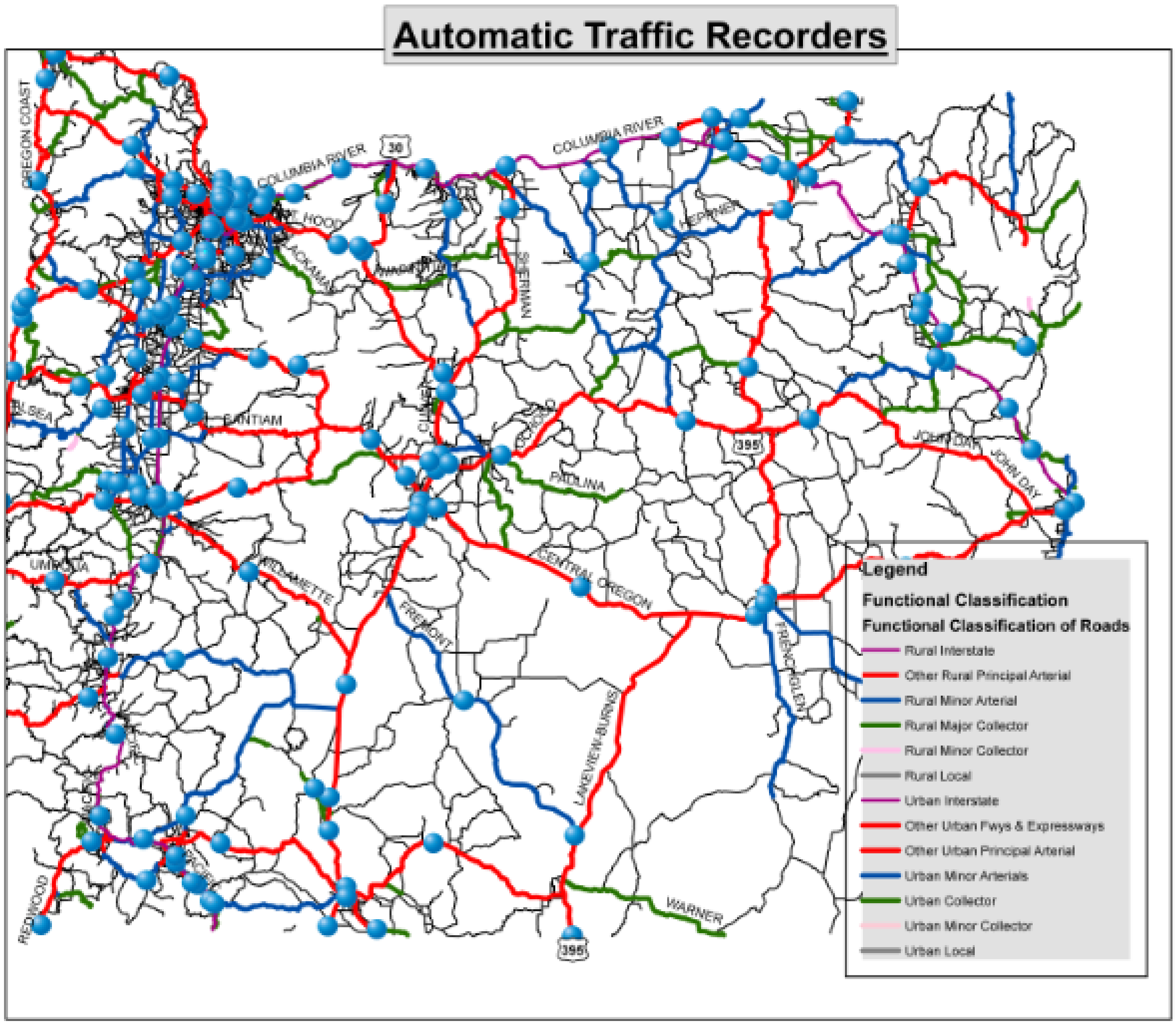

Figure 2.3: Automatic traffic recorder (permanent count location) in Oregon

The most direct way of comparing the estimates from the Streetlight Data platform is to compare these estimates with AADT estimates from the ATRs that ODOT manages across the state. For the comparison year 2017, ODOT has AADT estimates at 180 locations across each functional classification. The 2017 Streetlight AADT estimates will be evaluated based on the data quality criteria described in Section 3.1.

2.3.2 ATR Comparison – Accuracy

In order to measure accuracy of the Streetlight AADT estimates, 2017 AADT values from ATRs managed by ODOT are used for comparison. These could be fixed with some effort but was not done for this evaluation. Accuracy will be measures using percent error and absolute percent error (APE). These calculation methods are presented in the two equations below.

$$Percent\_Error = \frac{(AADT_{Streetlight} - AADT_{ODOT})}{AADT_{ODOT}}$$

(2-1)

$$Absolute\_Percent\_Error = \left|\frac{(AADT_{Streetlight} - AADT_{ODOT})}{AADT_{ODOT}}\right|$$

(2-2)

Where:

$AADT_{Streetlight}$ is the annual average daily traffic estimate from Streetlight

$AADT_{ODOT}$ is the annual average daily traffic estimate from ODOT ATR

Of the 180 ATRs in Oregon with quality data, only 173 are used in the comparison. Two of the sites submitted through the Streetlight platform did not get an AADT estimate. A review of the ODOT data at these sites revealed both had AADT 300 or less and the Streetlight Data platform will not return a value if it estimates 400 or less. Five other locations were removed because the gate was located over multiple segments making the resulting Streetlight Data AADT estimate not directly comparable to the ATR data.

Table 3.2 below shows the APE by urban area with median, mean and max values presented as well as the number of sites in each urban area. The result of all sites is presented in the final row of the table and shows that the median and mean APE is 18 percent and 26 percent respectively. The Portland urban area had the lowest measures of error, likely due to many of the sites in that urban area being on the interstate system where volumes are larger and prone to less error (see more on error by volume bin below). Rural sites exhibit higher error compared to all the sites and show the highest maximum error.

Table 2.2: ATR Comparison Absolute Percent Error Summary by Urban Area

	Absolute Percent Error			
Urban Area	**Median**	**Mean**	**Max**	**# Sites**
Albany	43%	43%	55%	2
Bend	20%	19%	24%	4
Eugene	14%	17%	44%	6
Grants Pass	32%	32%	33%	2
Medford	14%	16%	37%	6
Milton-Freewater	16%	16%	16%	1
Portland	4%	15%	82%	23
Rural	23%	29%	197%	111
Salem	11%	30%	90%	4
Small Urban	14%	24%	98%	21
All Observations	18%	26%	197%	173

It’s important to note that rural sites are defined as those residing outside the federal urban area boundaries but may still serve urban area residents making interurban trips. In order to show how error varies geographically, Figure 2.4 is presented below which describes the median absolute error by county. Rural counties generally reveal higher median error though this not always the case which raises some questions about the underlying method that might lead to these variations in results. The number of ATRs in each county is shown as a numeric value next to the county name in the figure below. No data is shown in Wallowa and Wheeler counties since not ATR devices were present in these areas.

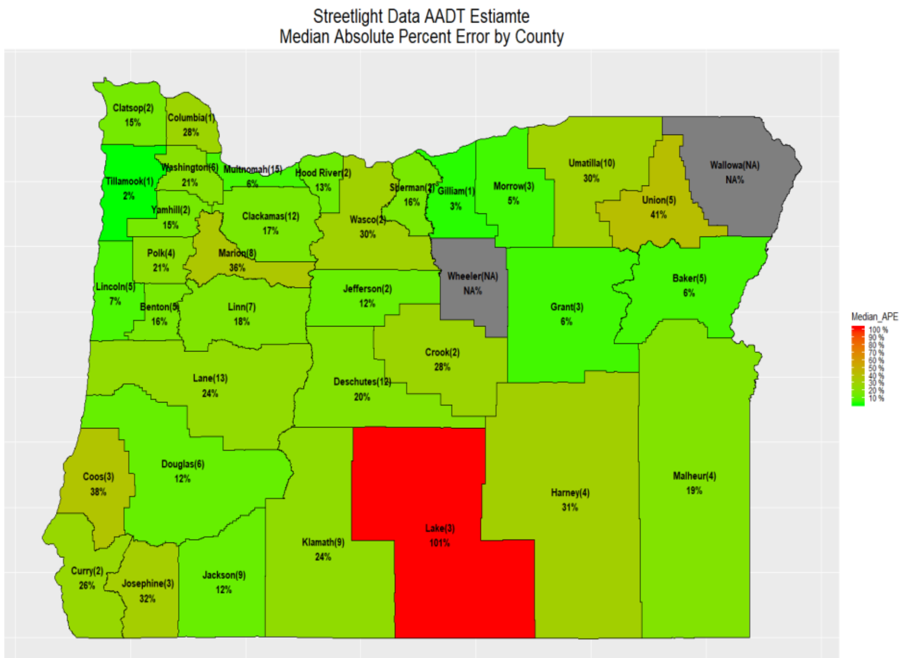

Figure 2.4: Median absolute percent error by county

The figure below shows the percent difference between the two AADT estimates based on volume bins. This figure describes the error using both a box and whisker chart and as well as the percent error values using points. The functional classification is also shown using different colors and symbols. For reference the 0% error line is shown as a dotted line. Any values on the right side of the line indicate ATR sites where the Streetlight estimate was higher than the ODOT data. Generally, as the overall volume at the site increases the percent difference decrease.

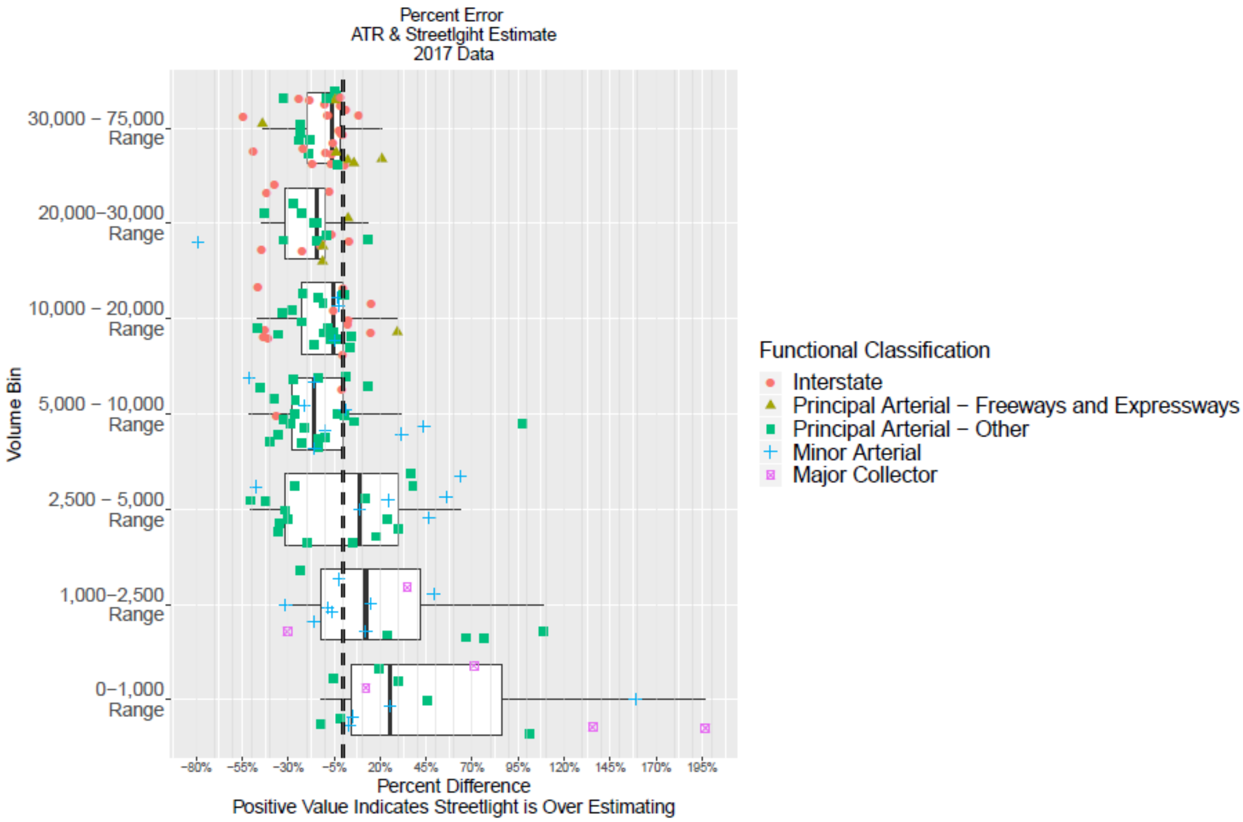

Figure 2.5: Percent error comparing ODOT ATR AADT estimates and streetlight AADT

Figure 2.6 below shows the absolute percent error using similar charting elements including box and whiskers and points representing individual location comparisons. This figure also shows the total number of observations and median error in text to the right of each box and whisker plot. This chart shows a similar general outcome where higher volume sites have less error with the 75,000 + volume bin having a median absolute error of just 3 percent. The 2,500 – 5,000 volume bin had the highest median absolute percent error with 32 percent. This graphic is presented in Table 2.3 below.

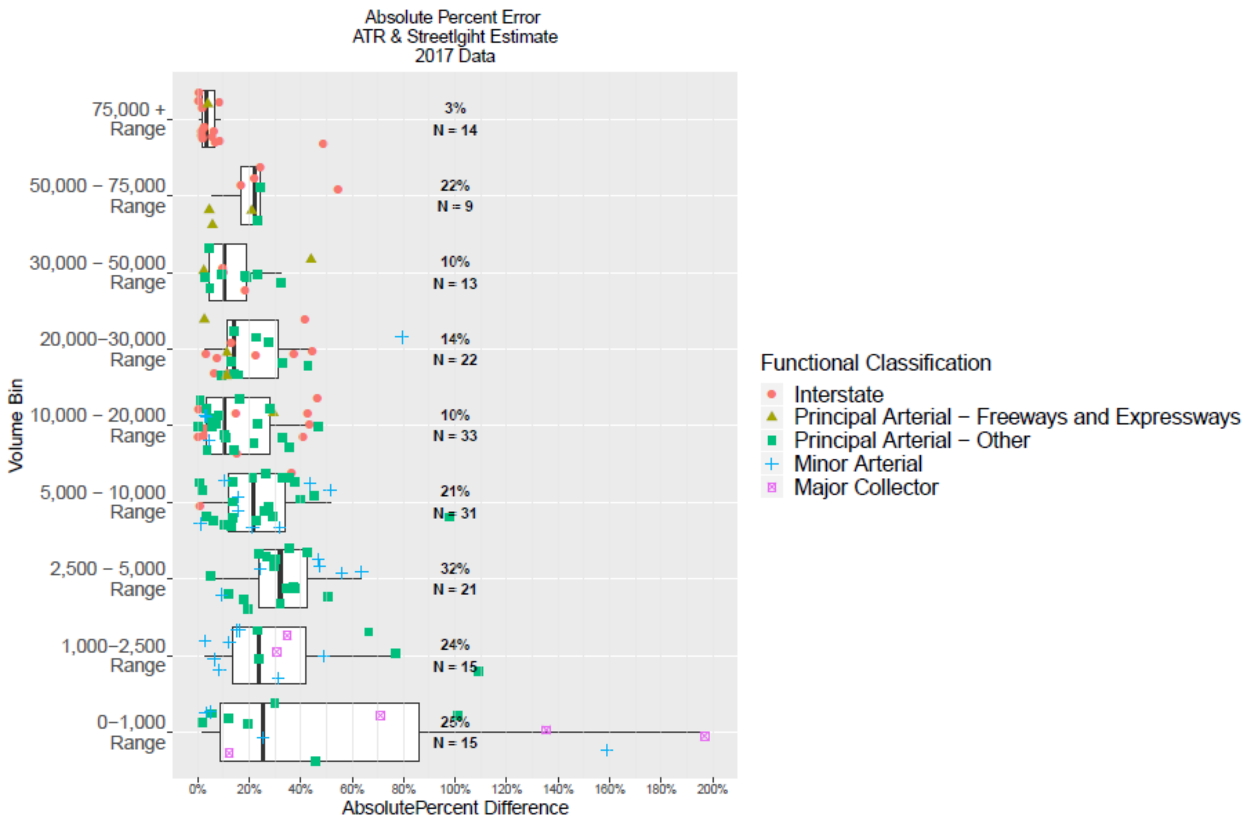

Figure 2.6: Absolute percent error comparing ODOT ATR AADT estimates and streetlight AADT

Table 2.3: Absolute Percent Error Summary by Volume Bin

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

The table below summarizes the absolute present error irrespective of volume showing that 57 of the sites exhibited absolute percent error of 10 percent or less with another 36 sites showed absolute error of between 10 and 20 percent absolute error, or 54% of the total number of sites.

Table 2.4: Absolute Percent Error Summary for ATRs

Error	# Sites	Proportion of Total
0-10%	57	33%
10-20%	36	21%
20-30%	28	16%
30-40%	20	12%
40-50%	17	10%
50%+	15	9%
Total	173	100%

The accuracy results presented above are the product of three iterations of setting gate widths. Default widths were used in the first iteration of submitting locations to Streetlight Data's platform followed by a review of the gates for locations where error looked particular high. Many of these sites had gate widths that did not adequately cover the facility and were adjusted. This was done twice and mostly focused on higher volume roads. Additional work could be done to better set the location of the gate and define its width to possibly improve these results.

2.3.3 ATR Comparison – Completeness

Of the 180 ATR sites submitted to the Streetlight platform 173, or 96% returned estimated AADT values for use. Of the seven locations, two were a related to an issue on Streetlight's side of the processing while the other five related to the placement of gates and could be fixed so the completeness could be either 96% or 99% depending on which value is selected.

2.3.4 ATR Comparison – Timeliness

The Streetlight AADT data represent 2017 which is the latest year available for comparison with ODOT data and so the timeliness is as current as any other available data set. Streetlight Data first released their 2017 AADT estimates as a product in August of 2018.

2.3.5 ATR Comparison – Accessibility

Access to the Streetlight 2017 AADT data required some spatial merging of ATR points with a network data set followed by some additional data elements added to conform to the template required by Streetlight Data platform. The web platform is then relatively simple though some minor trouble shooting required Streetlight customer support staff time. Once the platform processed AADT estimates, they were easily downloaded and analyzed. The input data manipulation, preparation and output analyses were performed in R, an open source statistical and data analysis software. It would be cumbersome to do any of these tasks in any platform other than one like R where scripts can be written to repeat tasks and handle summarizing data.

2.3.6 Data Quality Assessment Summary

Using the data quality metrics described above, the table below summarizes the results of the ATR Comparison evaluation. Accuracy of the Streetlight AADT product is very accurate based on measures of APE on higher volume roads but decrease as volume decreases. Only 53% of the comparison sites had error of 20% or less. This error could likely be reduced with additional work defining the gate locations and width. The data is very complete with 99% of the required locations being processed properly through the Streetlight platform. Validity was not assessed. The timeliness of the product was reasonable and is as up to date as agency data. Lastly, the vendor's platform was reasonably easy to work with though required some input and output processing that would require a platform with more functionality than some practioners may have in order to get the most out of the AADT product. The table below summarizes these findings.

Table 2.5: Data Quality Summary – ATR Comparison

Data Quality	ATR Comparison
Accuracy	Median APE 18% ; 53% of the comparison sites exhibited 20% error or less
Completeness	Vendor AADT estimates available at 99% of required sites
Validity	Not Assessed
Timeliness	Reasonable currency, as up to date as agency data
Accessibility (also referred to as usability)	Relatively accessible platform, with some moderate data processing tasks for input and output data which benefits from R

2.4 AADT ESTIMATE COMPARISON IN BEND MPO

In order to add additional insight on the Streetlight Data AADT product the following section compares the Streetlight Data AADT estimates to AADT estimates derived from short term traffic counts collected by ODOT as part of their reporting to FHWA for the HPMS requirements. The data quality rubric applied above will be used again below.

2.4.1 ODOT Data

AADT estimates from short terms counts are known to have less accuracy when compared to empirical counts due to error resulting from factoring methods. A study by Kockelman et al (2007) found that using data from Florida, 48 hour count data factored up to represent an annual figure had nearly 12% error on average. Another study by Sharma et al (1996) that evaluating AADT estimated from ATR data in Minnesota concluded that estimated AADT counts have between 11% and 95% error with inaccuracy falling marginally when 72-hour counts are used instead of 24 hour short term counts. No documentation of likely error in ODOT data is currently available.

This section initially used AADT estimates from 229 sites in the Bend MPO area but after discussion with the technical advisory committee about data quality issues at many of these sites, it was decided that only sites where a count had been taken within the last three years would be used. This reduced the number of sites from 229 to 66 sites. These remaining sites are summarized below in Table 2.6. Network data used to apply count site location to for input into the Streetlight Data platform is an all streets network used for multimodal analysis and includes all links in the system. Figure 2.7 below shows the location of ODOT AADT estimate sites in the Bend urban area.

Table 2.6: ODOT AADT Estimate Site Summary

Functional Classification	Number of Sites	Network Segment Length (mi.)
Principal Arterial - Other	21	1.5
Minor Arterial	19	1.6
Collector	26	2.1

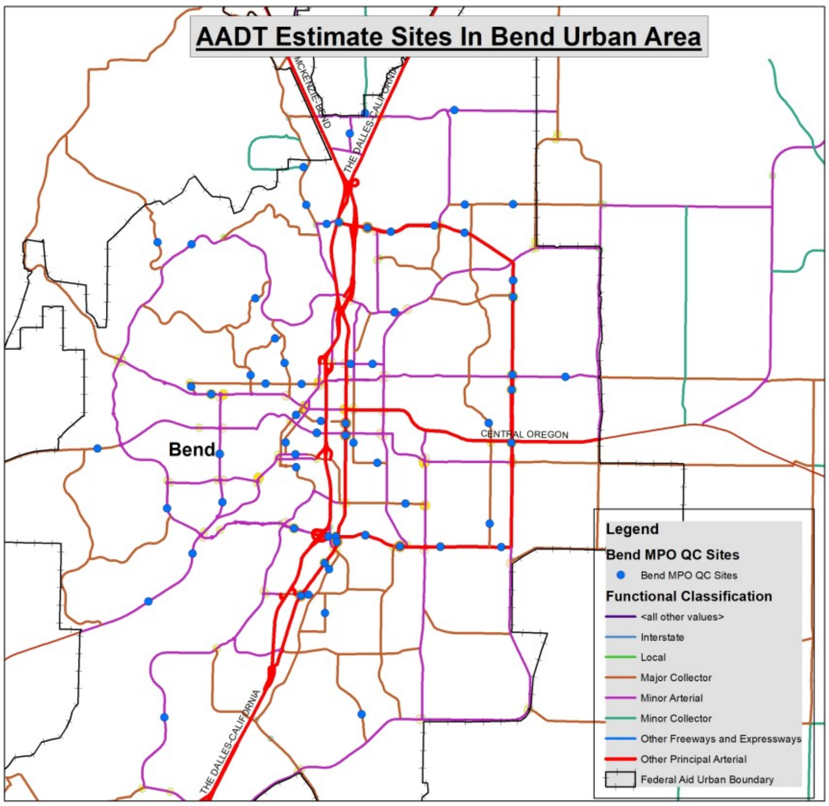

Figure 2.7: ODOT AADT estimate sites in Bend urban area

2.4.2 Bend AADT Estimate Comparison – Accuracy

The table below summarizes the absolute percent error by functional classification as well as a summary for all sites. For all sites the median absolute percent error is 32 percent with principle arterials showing the lowest error of 13 percent while collector facilities were highest with 83 percent error.

Table 2.7: ODOT AADT Comparison Absolute Percent Error Summary by Urban Area

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

Similar to the findings found in the ATR comparison above, volume of the site determined the accuracy of the Streetlight Data estimates with higher volume roads being exhibiting much less error. The figure below shows the percent error by volume bin. In 56 of the 66 comparisons, Streetlight Data over estimated volume compared to the AADT estimates.

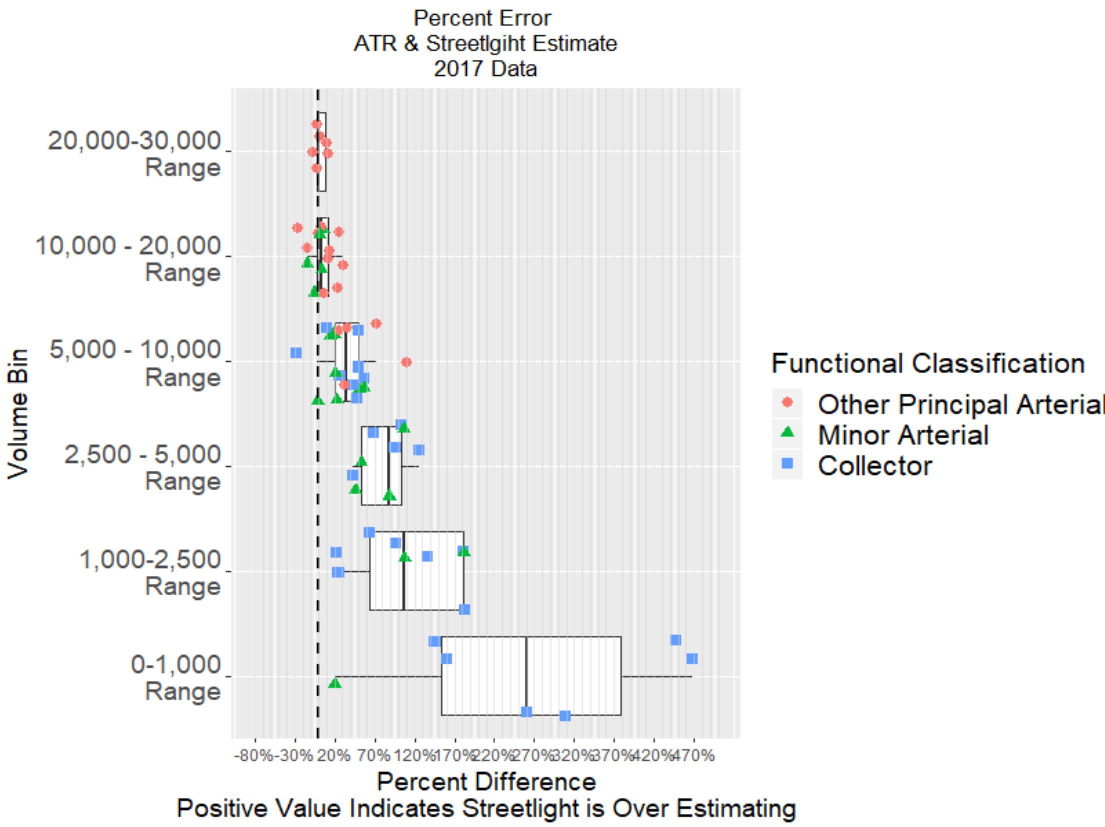

Figure 2.8: Percent error comparing ODOT short term AADT estimates and streetlight AADT

The absolute percent error distribution by volume bin is plotted below in Figure 2.9 and shows the 0- 1,000 AADT volume bin to have the highest median error of 260% while the lowest median absolute percent error is found in the 20,000 to 30,000 volume bin. Figure 3.8 does not show all volume bins since not all bins were represented in the short term AADT data.

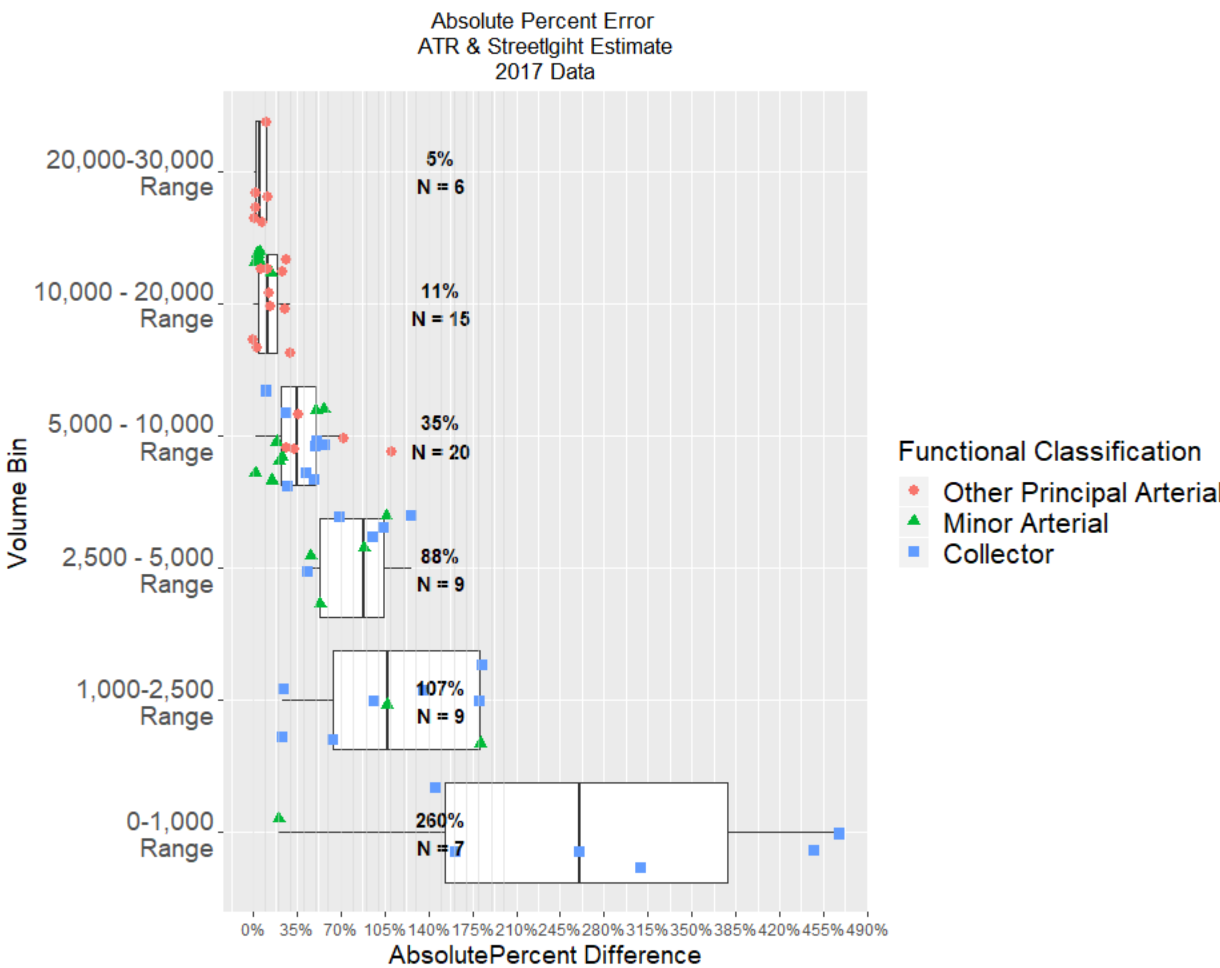

Figure 2.9: Percent error comparing ODOT short term AADT estimates and streetlight AADT

The table below summarizes the absolute present error irrespective of volume showing that 12 of the sites exhibited absolute percent error of 10 percent or less with another 9 sites showed absolute error of between 10 and 20 percent absolute error, or 14% of the total number of sites.

Table 2.8: Absolute Percent Error Summary for Short Term AADT Estimates

Error	# Sites	Proportion of Total
0-10%	12	18%
10-20%	9	14%
20-30%	11	17%
30-40%	3	5%
40-50%	5	8%
50%+	26	39%
Total	66	100%

The accuracy results presented above are the result of two iterations of setting gate widths. Default widths were used in the first iteration of submitting locations to Streetlight followed by a review of the gates for locations where error looked particular high. Many of these sites had gate widths that did not adequately cover the facility and were adjusted. This was done once and mostly focused on higher volume roads. Additional work could be done to better set the location of the gate and define its width to possibly improve these results. However, many comparison sites had gates that covered the facility properly but still revealed important differences in estimates between the ODOT and Streetlight estimates. Since these AADT estimates are derived from short term counts, which are known to have error, these comparisons may be less meaningful than those in the ATR section above.

2.4.3 Bend AADT Estimate Comparison – Completeness

Of the 229 ATR sites submitted to the Streetlight platform all were processed and returned with a value from the Streetlight platform resulting in 100 percent completeness.

2.4.4 Bend AADT Estimate Comparison – Timeliness

The Streetlight AADT data represent 2017 which is the latest year available for comparison with ODOT data and so the timeliness is as current as any other available data set. Streetlight first released their 2017 AADT estimates as a product in August of 2018.

2.4.5 Bend AADT Estimate Comparison – Accessibility

Access to the Streetlight Data's 2017 AADT product for comparison of data in this section was identical to the experience of the ATR data with some slight differences. More work is needed to make sure the gates are properly set and assigned an adequate width due to more site in these AADT data. The network data used in this comparison featured number of lanes which added a useful dimension to use when setting the gate widths and would be ideal to have at all the ATR locations.

2.4.6 Data Quality Assessment Summary

The table below summarizes the results of the Bend AADT estimate comparison evaluation. Accuracy of the Streetlight AADT product varies with more accurate Streetlight estimates on higher volume roads but decrease as volume decreases though no volume bin exhibited accuracy measures similar to the ATR comparison. Only 32% of the comparison sites had error of 20% or less though the use of AADT estimates derived from short term counts could help explain why these comparisons yielded higher error. This error could likely be reduced with additional work defining the gate locations and width. The data is complete with 100% of the required locations being processed properly through the Streetlight platform. Validity was not assessed. The timeliness of the product was reasonable and is as up to date as agency data. Lastly, since the product is identical to the data used in the ATR comparison the conclusion on use of the platform is similar. The table below summarizes these findings.

Table 2.9: Data Quality Summary – Short Term Count AADT Comparison

Data Quality	ATR Comparison
Accuracy	Median APE 32% ; 32% of the comparison sites exhibited 20% error or less
Completeness	Vendor AADT estimates available at 100% of required sites
Validity	Not Assessed
Timeliness	Reasonable currency, as up to date as agency data
Accessibility (also referred to as usability)	Relatively accessible platform, with some moderate data processing tasks for input and output data which benefits from R

3.0 REVIEW OF PAST EVALUATIONS

Little independent work has been published that assess the data quality of the Streetlight Data AADT product. One publically available evaluation was found that evaluated the AADT product in Minnesota. This work was published at the end of 2017 and is reported to have instructed further refinement of the process that Streetlight Data uses to estimate AADT.

In this evaluation for Minnesota Department of Transportation (MnDOT) worked with Streetlight to more fully develop the AADT algorithm and the public agency gave Streetlight Data access to data for 69 permanent count sites to use in training the algorithm (MnDOT 2017). Once the Streetlight Data algorithm was developed it was applied at 7,837 sites where AADT estimates were available from short term counts. The mean absolute percent error was 61% for all sites with less error at higher volume sites compared to lower volume sites. For example at sites with 50,000 AADT the MAPE was 34% (median 23%) while at sites with AADT between 300 and 5,000 the MAPE was 68 percent (median 31%). Streetlight estimates were mostly higher than the agency AADT estimates.

It should be noted that 1) it's unclear how the platform under develop in the MnDOT evaluation may have changed from the one evaluated in this report 2) even with MNDOT ATR data feeding the algorithm significant error resulted when comparing with other count sites in that state. Another evaluation is currently underway as of the writing of this report but no results are yet available.

Streetlytics is another third party data processing company that Louisiana Department of Transportation (LDOT) recently evaluated (LADOT 2018). Streetlytics is a product of a partnership between AirSage and Citilabs and uses mobile phone and GPS data to project traffic patterns through the Cube modeling platform. Like the Streetlight Data platform, the exact details of the analytic methods are not disclosed but appear to be similar. In LADOT's evaluation of traffic counts, it was found that absolute average error was 44.5 percent with error varying by volume bin.

As of the writing of this report there is a FHWA Pooled Fund underway titled *Exploring Non-Traditional Methods to Obtain Vehicle Volume and Class Data TPF-5(384)*.

4.0 CONCLUSION

This evaluation used available data from agency automatic traffic recorders and AADT estimates where confidence was high to assess the data quality of Streetlight Data's 2017 AADT product. Results showed that accuracy depends on the total volume of the site and can vary considerably between low and high volume bins. Other measures of data quality included in the rubric were relatively good. In order to get AADT estimates from the Streetlight Data platform for the whole state, considerable effort would be needed to properly set gates accurately across the network though compared to collecting data through traditional methods using on the ground sensors third party data likely represents a cost savings.

Accuracy may improve for ODOT sites if Oregon data were used in the Streetlight Data training algorithm though this would cause issues with doing a truly independent evaluation unless some agreement were reached with the firm where some sites were held out of the model training in order to ensure independence in any comparison.

Data accuracy is also a relative concept and depends on the ultimate use of the data. It's recommended that future evaluation work employ the Streetlight Data AADT product in a safety, vehicle miles traveled, or air quality analysis to determine the relative magnitude of difference in final results. Another recommendation is for Streetlight to quantify the uncertainty of their estimates in order to give users better information about the relative accuracy of the AADT product. Lastly, future evaluations should assess the cost of these data products with the cost of traditional counting techniques to understand if there are cost saving potential for the agency or at least an increase in overall information availability.

5.0 REFERENCES

Codjoe, J., Ashley, G., Saunders, W., & Louisiana Transportation Research Center. (2018). *Evaluating Cell Phone Data for AADT Estimation* (Final Report 591, Publication). Baton Rouge, LA: Louisiana Department of Transportation and Development. Retrieved from https://www.ltrc.lsu.edu/pdf/2018/ts_591.pdf.

Gadda, S., Kockelman, K. M., & Magoon, A. (2007). Estimates of AADT: Quantifying the uncertainty. In *World Conference on Transportation Research, Berkeley, Calif*.

Sharma, S. C., Gulati, B. M., & Rizak, S. N. (January 01, 1996). Statewide traffic volume studies and precision of AADT estimates. *Journal of Transportation Engineering, 122,* 6.

StreetLight Data, Inc. (2018) *Streetlight data's 2017 AADT methodology and validation white paper*. San Francisco, CA: StreetLight Data, Inc. Retrieved January 2019 from: https://www.streetlightdata.com/streetlight-volume-2017-aadt-methodology-and-validation-white-paper/

Turner, S. (2004). Defining and Measuring Traffic Data Quality. *Transportation Research Record: Journal of the Transportation Research Board, 1870*, 62-69. doi: org/10.3141/1870-08

Turner, S., Koeneman, P., Minnesota., & Texas A & M Transportation Institute. (2017). *Using mobile device samples to estimate traffic volumes*. St. Paul, Minn: Minnesota Department of Transportation, Research Services & Library. Retrieved from: www.dot.state.mn.us/research/reports/2017/201749.pdf