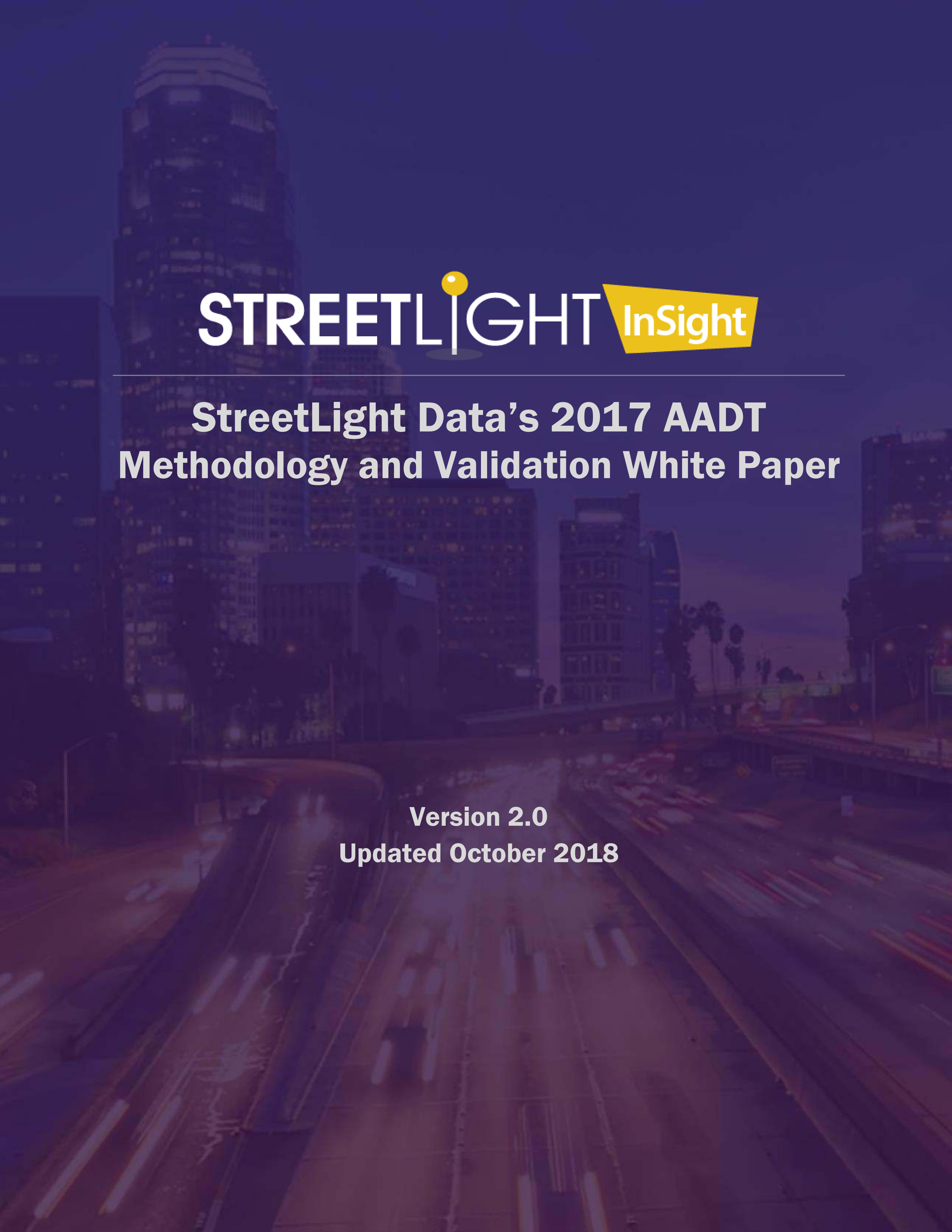


StreetLight Data's 2017 AADT Methodology and Validation White Paper

Version 2.0

Updated October 2018

StreetLight Data's 2017 AADT Methodology and Validation White Paper

This white paper provides technical detail about the methodology, algorithm development, validation, and data sources used in StreetLight Data's 2017 AADT Metrics. This white paper was first published in August 2018 and is updated periodically as new validation is performed.

Table of Contents

Introduction: Why We Developed a New AADT Metric

Annualized Average Daily Traffic (AADT) volumes are the core input to many important parts of the transportation practice. With new federal requirements put in place by the Moving Ahead for Progress in the 21st Century Act (MAP-21), local and state government agencies must increase reporting of performance metrics for planning and funding allocation. This in turn has increased need for quick, easy, cost-effective AADT measures (and its extension, vehicle miles traveled, or "VMT"). Modeling AADT counts leads to inaccuracies, as does temporary count expansion. Additionally, temporary counts are cumbersome and expensive.

StreetLight's 2017 AADT Metrics ("2017 AADT") are an upgrade and improvement over our previously released 2016 AADT Metrics ("2016 AADT"). We decided to upgrade and launch this new set of AADT Metrics for several reasons:

1. A year has passed, so we wanted to offer 2017 counts.
2. We wanted to incorporate client feedback on the accuracy of our 2016 AADT estimates for different geographies and road types.
3. There has been progress in the field on this topic overall, which gave us the opportunity to incorporate third-party findings.

Defining Target Accuracy

Our first question was to define what was “accurate enough” for our AADT. We had to balance the need for accuracy with the need to build an algorithm that could calculate an AADT very quickly and that is also compatible with the StreetLight InSight® transportation analytics platform. This meant that the algorithm had to be computationally elegant and scalable.

Our goal was to develop AADT estimates that are more accurate than a temporary count expansion. A paper by researchers by the University of Texas found that the absolute error for temporary count expansion ranges from 4.9 – 83%. The mean absolute percentage errors range from 11.5-18.5%, depending on rural/urban divide and functional class (see Table 2 in: Gadda, S., A. Mangoon, and K. Kockelman. “Estimates of AADT: Quantifying the Uncertainty.” 11th World Conference on Transport Research. Location: Berkeley CA, United States. Date: 2007-6-24 to 2007-6-28). Ranges were not available by AADT class, so we set our overall target stretching from 20% for lower AADT bins to 12% for higher AADT bins. The equation for mean absolute percentage error (MAPE) is shown below.

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

We also wanted our Metrics to be as accurate as an AADT estimated from a model. There are more sources on acceptable error in this body of work, as summarized by the Travel Modeling Improvement Program (see Figure 1). All the sources evaluate error in terms of Root Mean Square Error (RMSE) as a percent of AADT. The equation is:

$$RMSE\ as\ \% \ = 100 * \frac{\sqrt{\left(\frac{1}{n}\right) * \ (Traditional\ AADT - Big\ Data\ AADT)^2}}{(\frac{1}{n} * \sum_{i=1}^{n}(Traditional\ AADT)}$$

As you can see in Figure 1, below, many DOTs generally agree on the acceptable RMSE. We took the average of acceptable RMSEs for each AADT range as the target for our results (we merged the 0-1,000 and 1,000-2,500 ranges due to low availability of permanent counters). The targets for RMSE as % of AADT range from 20% for high AADT to 68% for lower AADT bins. Note -just because the RMSE values are higher than the MAPE does not mean the RMSE-conveyed results are “worse” – they are simply a different calculation.

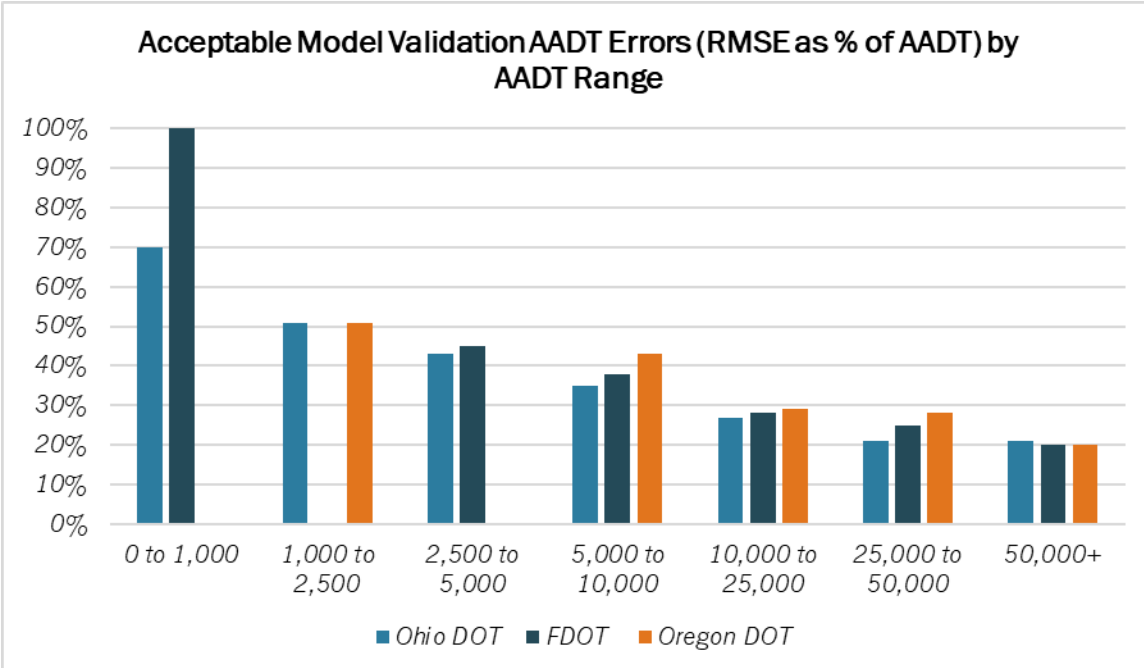


Figure 1: Source – Figure 9.8 in Travel Model Improvement Program, "Travel Model Validation and Reasonableness Checking Manual Second Edition." September 24, 2010.

Calculating the Accuracy of StreetLight's 2017 AADT Metrics

We trained our algorithm on a set of 2,441 permanent AADT counters from across the country. Then we tested the algorithm on a set of 164 permanent AADT counters. This test was "blind," which means the results below do not include data used for the calibration or training of our algorithm. The results you see in Figure 2 are actually the averages of many training and testing "folds." More detail on this process is in the following sections.

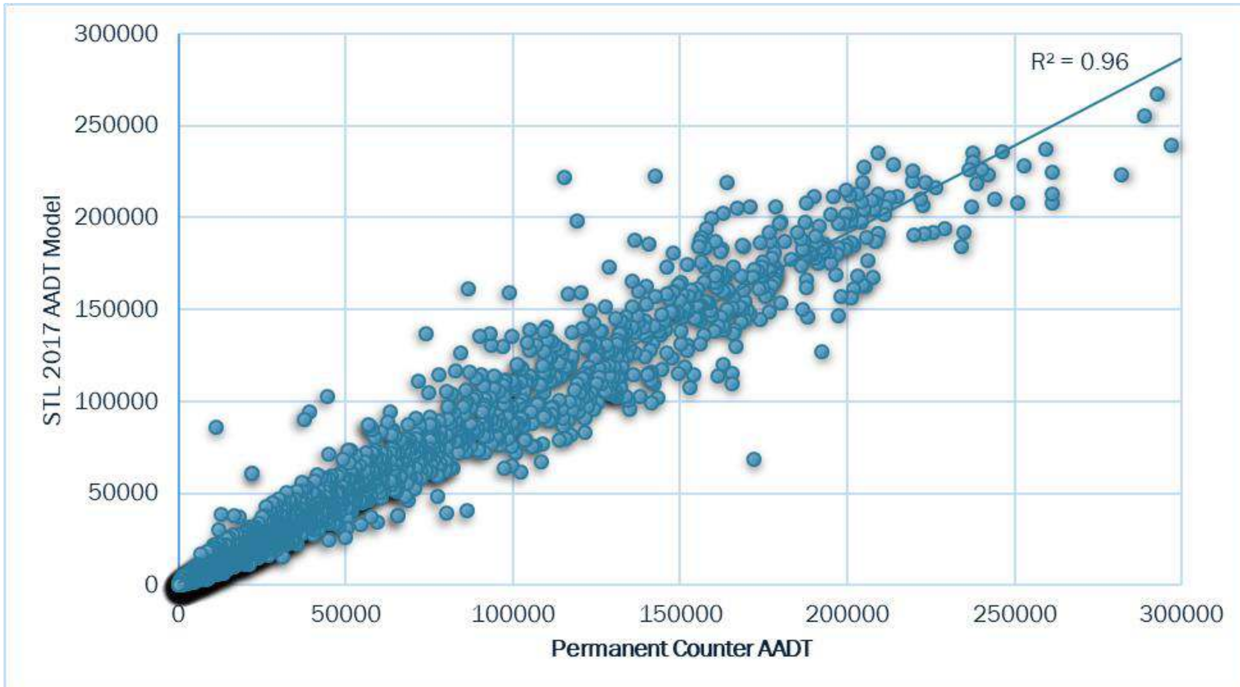


Figure 2 – StreetLight 2017 AADT for Test Data compared to Permanent Counter AADT. R2 is 0.96. No outliers were removed.

Tables 1a and 1b show the absolute error and RMSE of StreetLight's 2017 AADT compared to target errors.

AADT Range	# of Segments	Target Abs. Error	StreetLight Algorithm Mean Abs. Error	Delta to Target (positive means "better than target")
50,000+	795	12%	12.7%	-1%
25,000- 49,999	386	16%	15.7%	0%
10,000 – 24,999	509	20%	20.6%	-1%
5,000- 9,999	350	20%	23.5%	-4%
0 – 4,999	564	Not available - too few in comparison paper. See Table 1b.	43.3%	NA

Table 1a: Key Results from StreetLight's 2017 National Validation Test –Absolute Error

AADT Range	# of segments	Target RMSE as % of Average AADT	StreetLight Algorithm's RMSE as % Average AADT	Delta to Target (positive means "better than target")
50,000+	795	20%	15.80%	4%
25,000- 49,999	386	25%	20.80%	4%
10,000 – 24,999	509	28%	31.40%	-3%
5,000- 9,999	350	39%	31.50%	7%
2500 – 4,999	270	44%	36.10%	8%
0 – 2,499	294	68%	58.80%	9%

Table 1b: Key Results from StreetLight's 2017 National Validation Test – RMSE

We are happy with these results. In addition to exceeding or being very close to our desired targets for all classes of roads, our 2017 AADT Metrics perform well on both small and large roads. As detailed in the following section, the results also compare very favorably to other AADT counts derived from Big Data. We welcome opportunities to conduct additional validation studies for major clients who are interested in the results for their specific locales.

Comparing Our Results with Other Big Data Derived AADT Counts

Louisiana State University 2018 Big Data AADT Comparison for Louisiana DOT

A recent publication by the Louisiana Department of Transportation and written by researchers at Louisiana State University (LSU) showed results of Big Data-derived AADT estimates that utilized a combination of cellular-tower derived data and modeling software for expansion obtained from Streetlytics. (https://www.ltrc.lsu.edu/pdf/2018/FR_591.pdf).

There are a few differences between our approach and the approaches described in this paper: LSU reported their results as "difference" (not the same as RMSE or MAPE), as an attempt to mitigate for the higher share of temporary counts used in their test set. The "difference" equation is:

$$Difference = \left|\frac{(Traditional\ AADT - Big\ Data\ AADT)}{(Traditional\ AADT + Big\ Data\ AADT)/2}\right| * 100\%.$$

Therefore, we recut our own AADT Metrics to follow the "difference" calculation to allow readers to compare the results (see Table 2).

In addition, the paper suggests that the outputs tuned the model for a particular state. Therefore, we show both our nationwide results and our results if our algorithm were tuned for performance in one state. (Note: This state-specific tuning is available for state and regional-level for clients who invest in our 2017 AADT Metrics at scale.). We chose Washington State because many of the counters in that sample were located on low AADT roads.

Finally, the LSU paper removed 4.7% of results as outliers before presenting final outcomes. We do not remove outliers when reporting our results.

The weighted average "difference" for the LSU data, as reported in this document, was 44% (see table 3, page 27 of the report) improving to 38% for roads above 300 AADT. Our average difference on a national scale is 4.5% for all roads and 4.4% for roads above 300 AADT.

Table 2: Performance of AADT Counts Measured in "Difference" from LSU/Streetlytics Research and StreetLight Data's 2017 AADT.

AADT Range	*LSU/Streetlytics vs. Traditional Data (% Difference)*	*StreetLight National vs. Traditional Data (% Difference)*	*StreetLight Single State Tuned vs. Traditional Data (% Difference)*
0-300	41%	28%	27%
300-2500	35%	10%	9%
2500-5000	28%	7%	6%
5000-10000	23%	6%	5%
10000+	29%	4%	3%

Fundamentally, we think our better outcomes are due to three key reasons:

1) Our use of modern Machine Learning techniques, as opposed to the Universal Kriging approach used by the approach evaluated by LSU (See: Selby, B. and Kockelman, K. *Spatial prediction of AADT in unmeasured locations by universal kriging*. in *Transportation Research Board 90th annual meeting*. 2011.)
2) Our combination of data sources, with strong spatial precision and thus superior route locking potential, compared to the cellular data, means that we're building from a superior core of data.
3) As a disclosure, our test set had more very large (20,000+ AADT) roads than the LSU validation, and more permanent counters to compare to. This makes the top bin (10,000+) look extremely good for StreetLight. However, that should not affect our smaller bin comparisons.

If you know of other published research about Big Data-derived AADT, please send it to StreetLight. We could like to expand this comparison section of our white paper.

Data Sources Used in Our 2017 AADT Metrics

In developing our 2016 AADT we used three types of processed Big Data: location-based services data points, navigation-GPS personal trips, and commercial navigation-GPS commercial trips data. Our 2017 AADT uses more data sources and takes advantage of our improved location-based services data processing techniques. As with 2016 AADT, the results of combining these data sources are vastly superior to one or two of them alone. These were the inputs:

Input 1: Location-Based Services Trips Data

Location-based services (LBS) data are created by smartphone applications that provide a service that depends upon on a device's geographic location in the physical world; for example, shopping apps, weather apps, or dating apps. We used these data points for our 2016 AADT. Since then, we developed new algorithmic processing techniques to link those data points into trips. We put our new and improved LBS trips to good use in our 2017 AADT. A full 12 months of LBS trips data (365 days) went into our 2017 AADT, which means that they capture seasonal variation.

Input 2: Navigation GPS Trips – Personal and Commercial

The navigation-GPS data we use are created by connected personal and commercial vehicles. Our data set is tagged by vehicle type: personal vehicle, heavy-duty commercial vehicle, or medium-duty commercial vehicle. We included counts of both personal and commercial vehicles in both our 2016 and 2017 AADT. Since roads vary heavily in the share of commercial trucks (and in the share of medium vs. heavy duty trucks) having both is critical. We use a full 12 months of data (365 days), which means our AADT Metrics capture seasonal variation.

For both LBS and all navigation-GPS trips, we also analyzed the ratio of trips between different day parts and day types.

Input 3: US Census

We normalized our LBS trips using the US Census. Normalizing is an important step to adjust a sample that is not perfectly distributed. To better understand this method, please visit the StreetLight Data website. In short, 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. This adjusts for variation in geographic distribution, which is correlated with demographic factors like income. We also looked at the population density near the road in question, which is a proxy for identifying a road as rural or urban

Input 4: Open Street Maps Data

We included features commonly extractable from Open Street Maps (OSM) such as road geography, speed limits, number of lanes, availability of parking, road classification, and other factors. We know all OSM features are not always available for every road. Our algorithm is factored to adjust to a different set of coefficients if no OSM feature data is available. We also use the OSM to "lock" a trip to a route by connecting pings along the most viable network path a vehicle can take.

Input 5: Weather Data

We included data on precipitation and temperature to account for areas that have extreme precipitation events (like snow storms) on a regular basis and might experience different travel patterns as a result.

Input 6: Training and Testing AADTs using Permanent Loop Counters

We researched extensively to find well-cleaned permanent counter data. We wanted our data to be spread across the US, between small and large roads, urban and rural. The biggest challenge was finding permanent counter data for small rural roads. The map and charts below show the locations of the 2,605 counter data points we used to develop our algorithm.

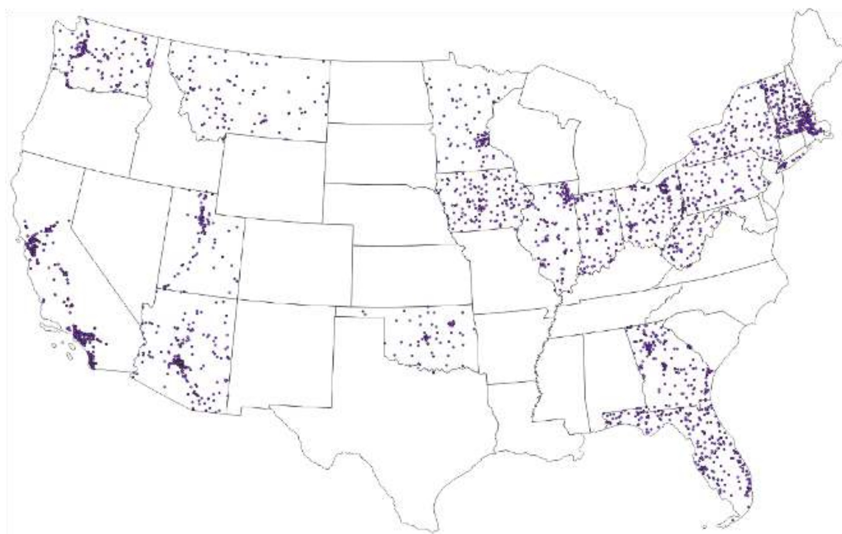


Figure 3: Map of all Permanent Counters Used for Training and Testing the 2017 AADT Metric.

State	# Permanent Counters	State	# Permanent Counters
AZ	232	NY	144
FL	243	NH	65
GA	181	OH	146
ID	116	OK	68
IN	90	CA	272
IA	147	PA	90
MA	193	UT	108
MN	84	VT	82
MT	97	WA	175
		WV	72

AADT Range	# Permanent Counters
50,000+	795
25,000- 49,999	386
10,000 – 24,999	509
5,000 - 9,999	350
2500 – 4,999	270
0 – 2,499	294

Table 3: State and AADT Range Spread for all Permanent Counters Used for Training and Testing the 2017 AADT Metric.

Selecting and Testing the Algorithm

We weighed dozens of different algorithmic approaches to develop our 2017 AADT Metrics. In this section, we will provide an overview of our major decisions.

First, we checked to see how our normalized LBS trips, which are our strongest, largest data set, were correlated with AADT. The results are shown below. As you can see, the correlation is fairly strong. As such, the remainder of the task was to use machine learning to reduce error and improve correlation as compared to scaling to counts from LBS trips alone.

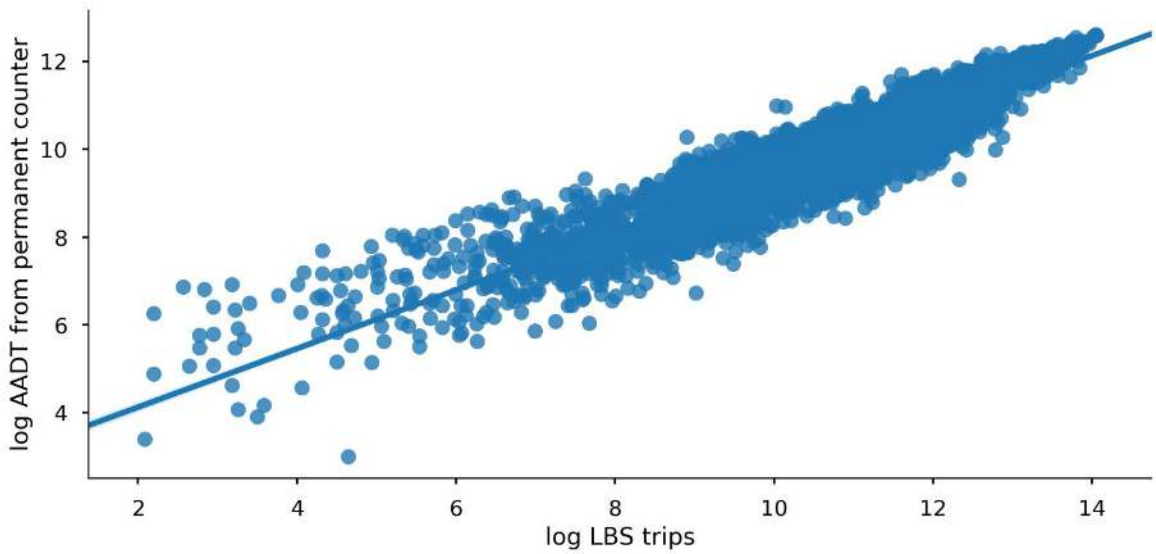


Figure 4: Correlation of population-normalized LBS to permanent loop counter data – the single best predictor value. The rest of the machine-learning work aimed at improving over these results.

The choice came down to two options for machine learning techniques: Ordinary Least Squares (OLS) and Random Forest.

First, we tried OLS, a multivariate equation framework with machine learning. The benefit of a multivariate regression technique is that it is easier to explain – you are more-or-less building a classic y = mx + b style equation. The disadvantage, as we found, is that the results were not as accurate as we wanted and the model was prone to throwing outliers.

We also tried a “random forest” model. For more on random forests, see Wikipedia (https://en.wikipedia.org/wiki/Random_forest#Algorithm) and our summary in Box 1.

Box 1: The Random Forest Model

This approach is similar to a decision tree, but it uses several decision trees. An example may help to illustrate this. Let's say you want to predict whether a patient entering an emergency room is at high risk or not. A decision tree may look like this: If age is over 50, blood pressure is over 150, and temperature is above 100F, then the patient is high risk. That's a decision tree. It is very interpretative, but does not have much predictive power alone. Random Forest uses a lot of decision trees (say, an ensemble), where each tree is a little bit different from the others. When we get a new patient, we take the majority vote of the decision tree ensemble to get a final result.

The different trees use random samples of observations and subset of features to train. For example, instead of considering age, blood pressure, and temperature, we may train one tree with age and blood pressure, another with blood pressure and temperature, another with age and temperature, and so on if we had more features. The key is that the trees become a bit different (less correlated), so when we average the results, we get a “diverse” answer. The idea is if you have a bunch of poor decision makers and put them in a room together to form a committee, they'll start making better decisions. And if each decision maker comes with a different perspective, you'll get better results.

The benefits of a random forest model for AADT estimation are that it is more accurate: In particular, it did a far better job of handling unusual roads (such as small ones or ones with extremely high commercial traffic near ports/warehouses). The detriment is that the algorithm is a little less intuitive to explain to non-data scientists. In the end, we think that the accuracy and algorithmic robustness for unusual roads and outliers was more important.

The next decision was to decide which features (input variables) to include in our random forest model. We tested hundreds of combinations. We wanted the most accurate results, but we wanted our algorithm to be scalable to anywhere in the US, and be computationally efficient.

In the end, we built a model with 12 features, which is an appropriate feature count for a training set of ~2,600. We see a lot of over-fitting in the transportation and big data industry, which means throwing far too many features into a machine learning model. This may make initial results look very good, but it also prevents the approach from scaling well outside of the research setting.

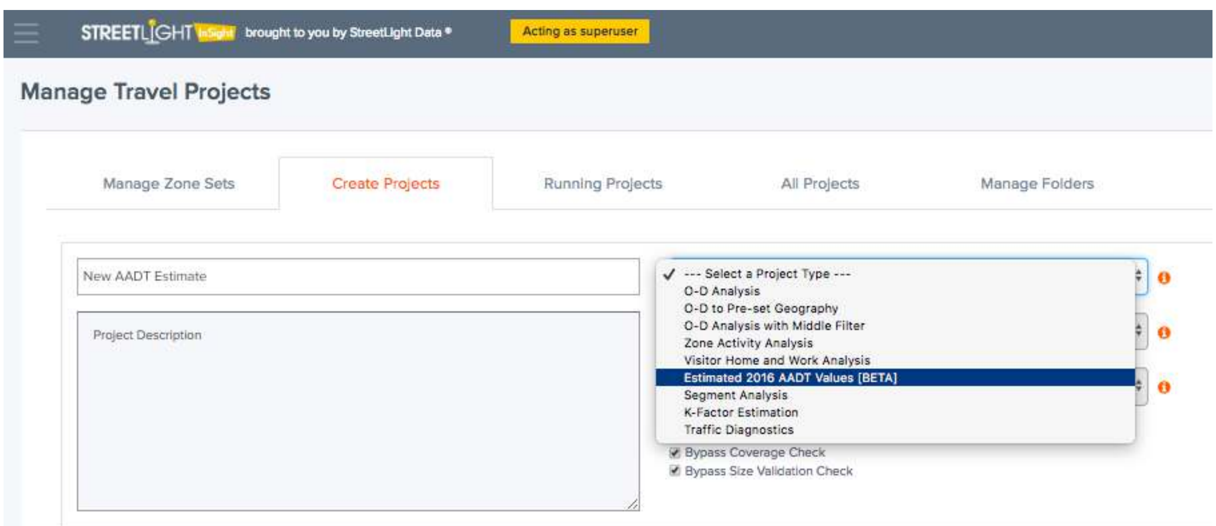
After we made the algorithm and feature choices for our model, we performed several types of cross validation. First, we ran the model several times, each time randomly designating different zones as "training set" and "test set." If the model output is very similar for each of these runs (or "folds"), you know you have a good model.

Next, we did another set of folds where we pulled out each state in turn as the "test" set. This was meant to mimic a new clients' experience because we trained our model on a set of data from a group of states. We wanted to know whether our model would perform well if someone ran AADTs in a new location. The answer again is yes, we showed strong match of our results with permanent counter data.

We did find that the state-based folds had more variation than the random folds. This means that we may be able to further improve results by tweaking the algorithm for an individual state or region. We are look forward to collaborating on this approach with clients who want to use our AADT Metric at large scale.

Running 2017 AADT Metrics in StreetLight InSight

If you have a *StreetLight InSight* account with AADT 2017 enabled, go to the "Create Projects" tab. Under Project Type, select "2017 AADT." Next, chose which Zones you want AADT for. Zones must be bidirectional pass-through road segments (either gate or line geometries).



Our Next Steps

In future *StreetLight InSight* releases, we will add unidirectional AADT as well as hourly and seasonal AADTs. We also will continue conducting validation studies on our 2017 AADT Metrics and will update this white paper as the results are made public.