STREETLIGHT InSight

StreetLight AADT 2020 Methodology and Validation White Paper

United States

Version 1.0

May 2021

Table of Contents

Introduction: Why We Developed an AADT Metric

Annual average daily traffic (AADT) volumes are the core input to many important parts of the transportation practice. With federal government 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. In addition, by 2026, all state DOTs are required to collect Model Inventory of Roadway Elements (MIRE) Fundamental Data Elements (FDEs), including AADT, on all public roads to collect Highway Safety Improvement Program (HSIP) funds. This in turn has increased the need for a quick, easy, and cost-effective AADT measure and its extension, vehicle miles traveled (VMT). Modeling AADT counts from regional travel demand models leads to inaccuracies, as does temporary count expansion. Additionally, temporary-counts are cumbersome and expensive. As an improved alternative, StreetLight Data, Inc. ("StreetLight") developed an AADT for urban or rural road segments across the U.S. that outperforms industry-standard accuracy targets and can be used by industry practitioners for reporting, traffic impact studies, and more.

What's New in AADT 2020

StreetLight's AADT 2020 Metric uses new and improved methods, building on lessons learned from AADT models of prior years, our customers, and our participation in the FHWA Pooled Fund (TAC Study Number: TPF-5(384)). This results in continued accuracy improvements. Some highlights and improvements of the new AADT Metric are:

- StreetLight has obtained access to training data captured from 6,692 unique permanent counter data points (3,032 unique permanent counter locations) across 25 states in the U.S.
- An Extreme Gradient Boosting model was selected as the primary method for estimating AADT values, a change from Random Forest models used in previous years' AADT models.
- A new type of "hybrid" machine-learning model was implemented, further improving estimations, especially on very low- and high-volume roads.
- Overall estimation accuracy has significantly improved from previous years' AADT, across a wide range of road sizes.

Defining Target Accuracy

Our first task 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 while being compatible with the StreetLight InSight® transportation analytics platform. This meant that the algorithm had to be computationally elegant and scalable.

To evaluate accuracy, we grouped permanent counter stations by AADT, and reported on results within those groupings (labeled road size bin). On some occasions, we used 10 road size bins for more detailed results, while in others we used three road size bins to more generally describe high-, medium-, and low-volume roads. Errors measured as a percentage of AADT naturally increase with smaller roads, and groupings by road AADT allow for more visibility into where the errors of the model lie. For summary metrics to describe the general error across an AADT grouping, we report on mean absolute percentage error (MAPE, see equation below), and normalized root mean square error (NRMSE, see equation below). These metrics are standard measures of error for predictive machine-learning models and are helpful to give an idea of what an expected error is for most stations. We detail both metrics, because they are both commonly used in literature, and highlight different aspects of model performance. MAPE describes errors well on small roads and treats errors of all sizes equally, and in contrast, NRMSE penalizes large errors more, making it more sensitive to the accuracy of AADT estimation on high-volume roads.

The equation for mean absolute percentage error (MAPE) is shown below:

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

The equation for normalized root mean square error (NRMSE) is shown below:

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

Although the summary of errors is useful for getting a sense of model performance, we find that describing the percentile spread of error across the test set is additionally illustrative of where error lies. The 68th percentile absolute error is reported, as it represents one standard deviation of the mean within a standard bell curve, and thus is a useful descriptor of the typical error across the road segments. As MAPE and NRMSE may be sensitive to outliers, the 68th percentile absolute error can provide more visibility into expected “typical” error. The 95th percentile absolute error is provided to measure the spread of errors across a broader array of sites, representing the upper limit of the expected absolute error. Finally, we present the median percentage error to give an indication of bias in our model, which also indicates how accurate an aggregate value (such as the sum of VMT on all roads in this bin) will be. Values close to 0 suggest that our model has low bias, while positive values would indicate overestimation, and negative values indicate underestimation.

For AADT model Metrics with sufficient sample size, we also computed the 95th percentile error, which is labeled “95% Traffic Count Error (TCE) Range (%).” For this metric, percentage error and the log of AADT was fit to a quantile regression. From this fit, the larger absolute value between the 2.5th and 97.5th error range is reported.

Evaluating model accuracy

When evaluating model accuracy, it is useful to consider how well AADT estimation compares to current estimation approaches. One widely used approach is to collect a two-day temporary count from a link, and expand that count to estimate AADT using calibration of nearby permanent counters ("short-term count expansion"). Research performed by Battelle for FHWA[1] included study of expected error from short-term count expansion. Table 1 shows the computed errors we compiled from short-term expansion in their study. It should be noted that Table 1 represents a conservative case of current AADT factorization methods, which assumes short-term counts are available on a link for the year AADT is being estimated, and that reference counts from permanent counters from the same roadway classification type and year are available for the region. This is not the case for most links across the United States.

Krile et al.[2] simulated model errors from same-year 48-hour temporary count (SY-TC) expansion. There are a few differences between StreetLight's approach to accuracy metrics and the Krile study. First, we examine 2020 AADT alone, while the Krile study looked at estimated AADT values from 2000 to 2012. In addition, we generate a single 2020 AADT estimate for each of the 2,469 bi-directional stations, while the Krile et al. (2015) study utilized 206 total stations, with multiple comparisons within each station (hundreds of pairs of 48-hour count to yearly AADT values for each station). The Krile study had only one counter with fewer than 500 AADT, which is not enough to draw a conclusion about the accuracy of temporary-count expansion for such small roads. Thus, there is no widely accepted target for accuracy in this very low range.

Minnesota DOT volunteered their assessment of how often the short-term same-year expansion method in Table 1 is used. For roads under 2,000 AADT, many of which are not in the federal aid system, over half of the roadways lacked any actual counts at all, and for roads under 55,000 AADT, less than half had a short-term count from the same year (it should be noted that the FHWA's Traffic Monitoring Guide recommends a three-year cycle for counting such roads).

[1] Krile, R., Todt, F., Schroeder, J., & Jessberger, S. (2016). Assessing roadway traffic count duration and frequency impacts on annual average daily traffic estimation: assessing accuracy issues related to annual factoring (No. FHWA-PL-16-012). United States. Federal Highway Administration.
[2] Krile, R., Todt, F., Schroeder, J. (2015). Assessing Roadway Traffic Count Duration and Frequency Impacts on Annual Average Daily Traffic Estimation (FHWA-PL-16-008). United States. Federal Highway Administration.

Thus, we illustrate how the SY-TC errors compare to a more typical estimation (Typ-NC) arising from an older or similar count in Table 2. No comprehensive or complete data source was found to describe the accuracy of typical techniques, such as different-year temporary counts or "similar" segments, akin to the Krile report for same-year two-day expansions. Therefore, a summary of reports was combined with expert input to create the reference comparison metrics.

Table 1 below illustrates the same-year temporary-count (SY-TC) targets across 10 categories of road size. Note that there are no SY-TC targets available for roads with AADT under 500 (Road Size Bin A).

Road Size Bin	**Median Error (Bias) (%)**	**95% TCE Error Range (%)**	**68th Percentile Absolute Error (%)**	**95th Percentile Absolute Error (%)**	**MAPE (%)**	**NRMSE (%)**
B: 500 – 1,999	0.0	34.2	11.7	26.9	10.0	12.9
C: 2000 – 4,999	2.2	30.8	10.7	33.5	10.4	17.2
D: 5,000 – 9,999	3.1	28.5	9.6	28.1	9.2	13.9
E: 10,000 – 19,999	1.2	26.7	9.2	27.8	8.9	12.9
F: 20,000 – 34,999	0.8	25.7	8.3	24.3	8.1	13.2
G: 35,000 – 54,999	0.4	24.8	8.3	19.3	7.2	9.7
H: 55,000 – 84,999	-0.2	24.1	6.0	14.4	5.2	7.2
I: 85,000 – 124,999	0.0	23.5	4.8	14.7	4.6	6.7
J: > 125,000+	3.0	23.3	7.0	17.6	6.2	9.9

Table 1: Bias and absolute percentage error from short-term counter expansion in estimating AADT across 10 categories of road size for a same-year 48-hour count. This assumes short-term counts are

available on a link for the year AADT is being estimated, and that reference counts from permanent counters from the same roadway classification type and year are available for the region. This is not the case for most links across the United States, especially for small roads (see introduction). Data from Krile et al. (2015).

Table 2 includes 95% TCE Error Ranges for three categories of roads with high, medium, and low volumes, as well as MAPE and NRMSE. In this table we include estimation targets for the same-year temporary-count targets, as well as the typical situations with no counts for the same calendar year. Note that Typ-NC targets do not have target 95% TCE Error Ranges available and there are no targets available, for roads with AADT under 500.

AADT Volume Range	Method	95% TCE Error Range (%)	Median Bias (%)	MAPE (%)	NRMSE (%)
500 – 4,999 (low)	Same-Year Temporary Counts	34	-0.1	10.2	18.0
	Typical "No Count" estimates	Not available	Not available	50	65
5,000 – 54,999 (medium)	Same-Year Temporary Counts	28	1.1	8.6	14.2
	Typical "No Count" estimates	Not available	2	18	27
55,000+ (high)	Same-Year Temporary Counts	24	1.4	5.3	9.5
	Typical "No Count" estimates	Not available	1.5	20	12

Table 2: 48-hour same-year temporary-count expansion as represented by Krile et al. (2015) and typical situations with no counts from the same calendar year for key statistical indicators.

Calculating the Accuracy of StreetLight's AADT 2020 Metric for the U.S.

We trained our machine-learning model on a set of 6,692 unique permanent AADT counters from 25 states across the U.S., with outliers removed. To evaluate our model performance, we used two different validation techniques. In order to report on error, the stratified k(10) fold cross-validation was used to generate test sets for metrics. The individual counter stations were randomly divided into 10 groups, stratified by AADT grouping. For each fold, a given model was trained on nine of the other folds, and the chosen fold was used to generate error metrics for the model. Cross-validation was used because it allows for every counter to be tested, so that the model performance can be evaluated across as many unique types of roads and regions throughout the United States as possible. This metric represents how well we expect our model to perform for locations well represented by our training data, including the states we incorporated into building our model.

We also used a technique called "leave-one-state-out (LOSO) cross-validation," training our model on a set of counters from 24 different states. The model was tested 25 times, each time blind to the permanent counters of a different test state. Because our model was not trained on all 48 contiguous states, the LOSO error allows us to understand the expected error for geographic regions not represented in our trained model. All cross-validated results were generated exclusively from bi-directional counters, which totaled 2,469 locations in our sample.

As shown in Figure 1, the actual and estimated AADT values through the cross-validation are highly correlated with an extremely high R^2 and the fitted line is very close to a 45-degree line, which indicates that the performance of our model is excellent, with minimal bias.

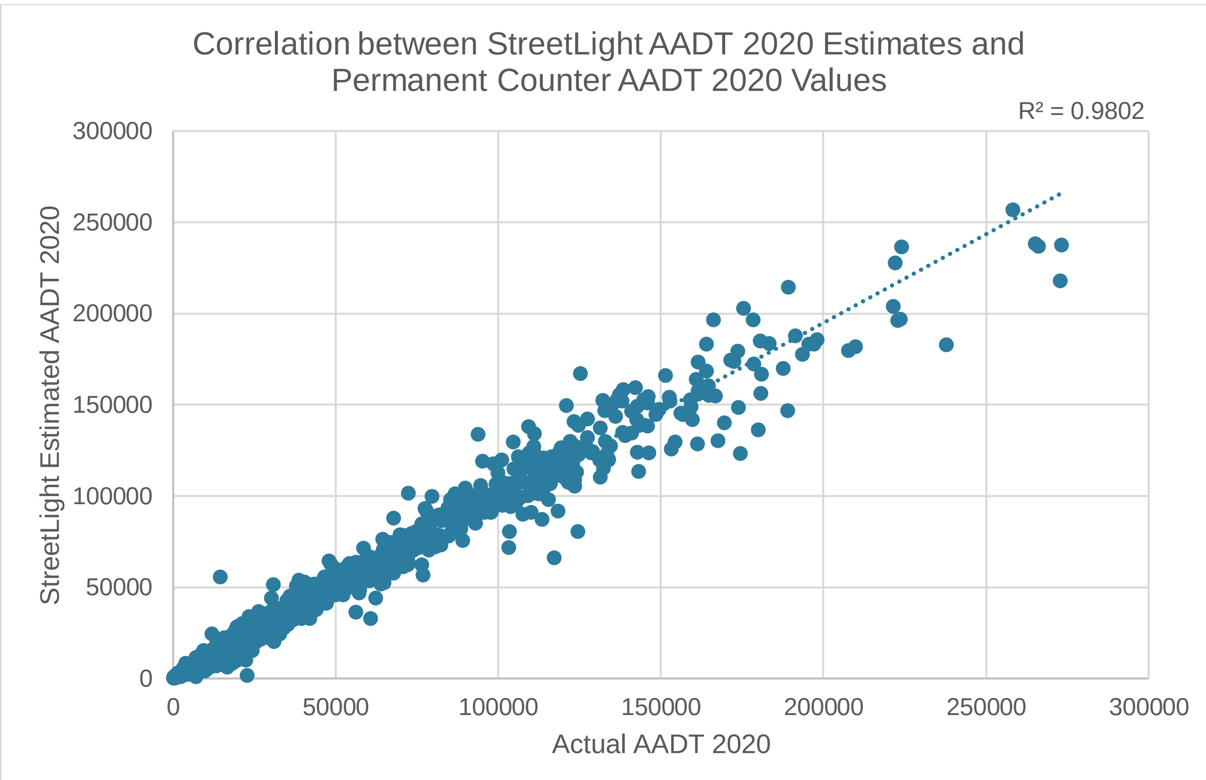

Figure 1: StreetLight AADT 2020 for test data compared to permanent counter 2020 AADT (U.S.). No outliers were removed.

Tables 3a and 3b show the estimation accuracy based on our two methods of cross-validation. In the following tables we utilize the 10 categories of road size (A-J) in order to gain further insight into the model's performance. The K-fold results represent random testing and are most representative of model error for states included in the training set.

Road Size	Count	50th Percentile	68th Abs Percentile	95th Abs Percentile	MAPE (%)	NRMSE (%)
A: 0 – 499	21	66.6	98.0	130.1	74.8	82.3
B: 500 – 1,999	248	12.0	30.4	69.5	25.4	30.8
C: 2,000 – 4,999	393	0.2	18.4	39.9	14.9	20.8

D: 5,000 – 9,999	386	-0.3	14.1	34.0	12.5	17.9
E: 10,000 – 19,999	451	0.2	12.0	32.4	11.7	20.5
F: 20,000 – 34,999	339	-1.5	9.3	23.9	8.8	12.9
G: 35,000 – 54,999	204	0.1	7.5	19.1	7.3	9.6
H: 55,000 – 84,999	158	-1.0	8.4	20.7	7.5	10.6
I: 85,000 – 124,999	170	0.6	7.5	22.7	7.0	10.5
J: 125,000 +	99	-3.4	11.7	22.1	8.9	11.8

Table 3a: Cross validation results for K-fold method for StreetLight's 2020 AADT Metric

The leave-one-state-out method results are most relevant for states whose permanent counts for 2020 AADT values were made available for training the model. As we would expect, the leave-one-state-out method produces slightly more variation relative to the standard K-fold method.

Road Size	Count	50th Percentile	68th Abs Percentile	95th Abs Percentile	MAPE (%)	NRMSE (%)
A: 0 – 499	21	86.7	115.8	223.8	104.5	116.3
B: 500 – 1,999	248	19.3	36.9	87.8	32.1	38.9
C: 2,000 – 4,999	393	4.3	21.8	52.3	19.1	25.1
D: 5,000 – 9,999	386	-0.9	16.9	37.9	14.5	20.0
E: 10,000 – 19,999	451	-0.2	14.1	35.1	13.1	22.6
F: 20,000 – 34,999	339	-1.2	9.8	28.8	9.5	14.0
G: 35,000 – 54,999	204	0.3	8.8	24.3	8.0	10.9
H: 55,000 – 84,999	158	-1.7	10.4	24.1	9.0	12.6
I: 85,000 – 124,999	170	-0.1	11.5	38.1	11.5	17.0
J: 125,000 +	99	-3.6	13.6	35.2	12.8	16.7

Table 3b: Cross-validation results for leave-one-state-out method for StreetLight's 2020 AADT Metric. Count refers to the number of unique bi-directional sites evaluated within each AADT range category.

Overall, we're pleased with these results, especially the strength of the estimates on higher-volume roads and the minimal bias, as shown in the 50th percentile column. Later sections will illustrate how these Metrics are an improvement upon prior years' AADT models.

Next, we use coarser road size bins in order to compare model error between StreetLight's AADT and expansion from short-term counts. We calculated errors from the Krile et al. (2015)

study[3] discussed earlier and have presented them in the table below to allow comparison within the three presented groupings by road volume AADT (low, medium, high).

Table 4 below shows StreetLight's cross-validated results following K-fold method compared to the targets.

AADT Volume Range	Method (n)	95% TCE Error Range (%)	Median Bias (%)	MAPE (%)	NRMSE (%)
0 – 499 (very low)	Same-Year Temporary Counts (77)	Unknown	Unknown	Unknown	Unknown
	Typical "No Count" estimates	Unknown	Unknown	Unknown	Unknown
	AADT 2020 K-fold (21)	58.5	49.4	66.41	76.11
500 – 4,999 (low)	Same-Year Temporary Counts (77)	34	-0.1	10.2	18.0
	Typical "No Count" estimates	Not available	Not available	50	65
	AADT 2020 K-fold (641)	40.50	4.3	19.0	24.0
5,000 – 54,999 (medium)	Same-Year Temporary Counts (103)	28	1.1	8.6	14.2
	Typical "No Count" estimates	Not available	2	18	27
	AADT 2020 K-fold (1380)	27.88	-0.7	10.6	15.4
55,000+ (high)	Same-Year Temporary Counts (25)	24	1.4	5.3	9.5
	Typical "No Count" estimates	Not available	1.5	20	12
	AADT 2020 K-fold (427)	15.65	-0.9	8.1	12.7

Table 4: StreetLight's AADT 2020 cross-validated results compared to 48-hour same year temporary count expansion as represented by Krile et al (2015) and typical situations with no counts for key statistical indicators. N refers to the number of unique bi-directional sites evaluated within each AADT volume range category.

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

For medium and large roads, the AADT 2020 model performs competitively with 48-hour same-year temporary counts across most accuracy metrics and consistently better than the typical situations with no counts. We find that our model may perform better for some states, which may be due to the enriched availability of permanent counts within the state used to calibrate our model, both in terms of geographic spread and diversity of road types. Below we highlight a number of states from the 2020 AADT model that performed particularly well relative to same-year temporary-count TCE targets. State-based 95% TCE Metrics in Table 5 below were derived from the K-fold cross-validation method.

AADT Volume Range	**Target 95% TCE Error Range for Same-Year Temporary Counts**	**State(n)**			
		Georgia (163)	**Massachusetts (135)**	**Ohio (159)**	**Washington (146)**
500 – 4,999 (low)	34	38.79	38.93	38.73	39.94
5,000 – 54,999 (medium)	28	27.36	24.82	24.56	27.89
55,000+ (high)	24	13.69	16.25	16.35	15.78

Table 5: StreetLight's AADT 2020 cross-validated results (K-fold) relative to 95% TCE targets based on a 48-hour same-year temporary-count expansion for specific states. N refers to the number of unique bidirectional sites evaluated within each state.

Table 6 uses the 10 categories of road size and compares the target MAPE results across same-year temporary-counts and typical "no count" estimates (such as years where a count is extrapolated) to the cross-validated results from StreetLight's AADT 2020 model. Where the AADT 2020 model performs better than typical "no count" estimates (the case for all categories), MAPE values are bold. Where the AADT 2020 model is within 1% of the same-year temporary-count targets, values are highlighted in blue, while places where the model is within 5% of the same-year temporary-count targets are highlighted in yellow. We note that the targets set for same-year temporary count targets are very conservative – they assume that all data collection processing, factoring, and collector functioned perfectly. The final column indicates the individual states represented in the AADT 2020 model with MAPE values that individually exceed the same-year temporary-count targets. We have also included the median bias targets and achievement in all categories. A low median bias means that the estimates can be used accurately for aggregate AADT functions (such as the sum of all VMT in a state for a certain road class).

AADT Volume Range	Target Median Error (%) Bias for Same-Year Temporary Counts	AADT 2020 Median Error (%) Bias	Target MAPE (%) for Same-Year Temporary Counts	Target MAPE (%) for Typical "No Count" Estimates	AADT 2020 MAPE (%)	States(n) that exceed Same-Year Temporary-Count MAPE target with AADT 2020 model
A: <= 499	Unknown	66.6	Unknown	Unknown	**74.8**	--
B: 500 – 1,999	0.01	12	10	50	**25.4**	Louisiana(2)
C: 2000 – 4,999	-0.3	0.2	10.4	50	**14.9**	Connecticut(1), Florida(30), Illinois(13), Indiana(9), Louisiana(3), Minnesota(1), Ohio(16)
D: 5,000 – 9,999	0.4	-0.3	9.2	18	**12.5**	Connecticut(4), Georgia(26), Indiana(10), Iowa(27), Louisiana(6), Ohio(23), West Virginia(5)
E: 10,000 – 19,999	0.9	0.2	8.9	18	**11.7**	Indiana(11), Iowa(26), Michigan(4), Minnesota(3), North Carolina(19), Ohio(25), Virginia(51)
F: 20,000 – 34,999	1.3	-1.5	8.1	18	**8.8**	Connecticut(4), Indiana(8), Iowa(10), Michigan(9), Montana(4), North Carolina(19), Ohio(37), Texas(31), West Virginia(1)
G: 35,000 – 54,999	3.2	0.1	7.2	18	**7.3**	Arizona(7), Georgia(15), Indiana(8), Iowa(1), Massachusetts(18), New Mexico(1), Ohio(27), Texas(21), Virginia(8)

H: 55,000 – 84,999	2.2	-1	5.3	20	**7.5**	Connecticut(8), Louisiana(1), Massachusetts(32)
I: 85,000 – 124,999	0	0.6	4.6	20	**7.0**	Indiana(4), Michigan(8), New Hampshire(1), Washington(15)
J: > 125,000+	2.4	-3.6	6.2	20	**8.9**	Connecticut(1), Georgia(19)

Table 6: StreetLight's AADT 2020 cross-validated results (K-fold) relative to MAPE and Median Bias targets based on a 48-hour same-year temporary-count expansion and typical "no count" estimates. N refers to the number of unique bidirectional sites evaluated within each state.

Comparing Cross-Validation Results Among AADT 2018, 2019, and 2020

When comparing our previous AADT Metrics, we see that our AADT 2020 machine-learning model performs better than both the 2019 model, and the 2018 model. In the table below, we compare the cross-validation results of the MAPE values among AADT 2018, 2019, and 2020. Through the comparison, it can be seen that AADT 2020 improves results for small roads of AADT over 500 through large-traffic-volume roads. This improvement in the 2020 AADT model is due to updated modeling methodology.

AADT Volume Range	StreetLight 2018 AADT		StreetLight 2019 AADT		StreetLight 2020 AADT	
	n	MAPE (%)	n	MAPE (%)	n	MAPE (%)
A: <= 499	20	840.4	48	42.5	21	74.8
B: 500 – 1,999	150	27.2	350	27.7	248	25.4
C: 2000 – 4,999	226	21.1	636	19.3	393	14.9
D: 5,000 – 9,999	249	17.4	632	14.7	386	12.5
E: 10,000 – 19,999	254	16.2	710	12.7	451	11.7
F: 20,000 – 34,999	185	13.7	593	10.5	339	8.8
G: 35,000 – 54,999	130	13.9	371	9.3	204	7.3
H: 55,000 – 84,999	102	11.5	258	8.3	158	7.5
I: 85,000 – 124,999	86	11.5	187	8.6	170	7.0
J: > 125,000+	93	11.1	152	8.4	99	8.9

Table 7: Comparison of cross-validation results of mean percentage error (MAPE) metric among StreetLight's AADT models for 2018, 2019, and 2020 (U.S.).

Data Sources Used in Our AADT 2020 Metric

Our AADT 2020 Metric blends the following data sources to create our best estimation of Annual Average Daily Traffic:

Input 1: Location-Based Services Trips Data

Location-Based Services (LBS) data is created by smartphone applications providing a service that depends upon on a device's geographic location in the physical world – for example, shopping apps, weather apps, or direction-finding apps. We use algorithmic-processing techniques to link these data points into trips. We sampled trips throughout 2020 in order to create the best model possible for the 2020 calendar year.

Input 2: Navigation GPS Trips – Commercial

The navigation GPS data we use is created by connected commercial vehicles. Our data set is tagged by vehicle type: heavy-duty commercial vehicle, or medium-duty commercial vehicle. Because roads vary heavily in the share of commercial trucks (and in the share of medium- vs. heavy-duty trucks), having a combination of data sources from commercial vehicles is critical.

Input 3: U.S. Census Data

We normalized our LBS trips using the U.S. Census. Normalizing is an important step to adjust a sample that is not perfectly distributed. In short, if 10 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), as well as employment and income data pertinent to the surrounding area. To better understand this method, please read the analysis on our website, "Larger and More Representative Sample Size."

Input 4: OpenStreetMap Data

We included features commonly extractable from OpenStreetMap (OSM) such as road geography, speed limits, number of lanes, availability of parking, road classification, and other factors. We know not all OSM features are 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 OSM to "lock" (map match) 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 weather events (like snowstorms) on a regular basis and might experience different travel patterns as a result.

Input 6: Training and Testing AADT Using Permanent Loop Counters

We researched extensively to find well-cleaned permanent counter data. We wanted our data to be spread across the U.S., between small and large roads, urban and rural. The biggest challenge was finding permanent counter data for small rural roads. The following maps and charts show the locations of the 6,692 unique counter data points in the U.S. that we used to develop our algorithm.

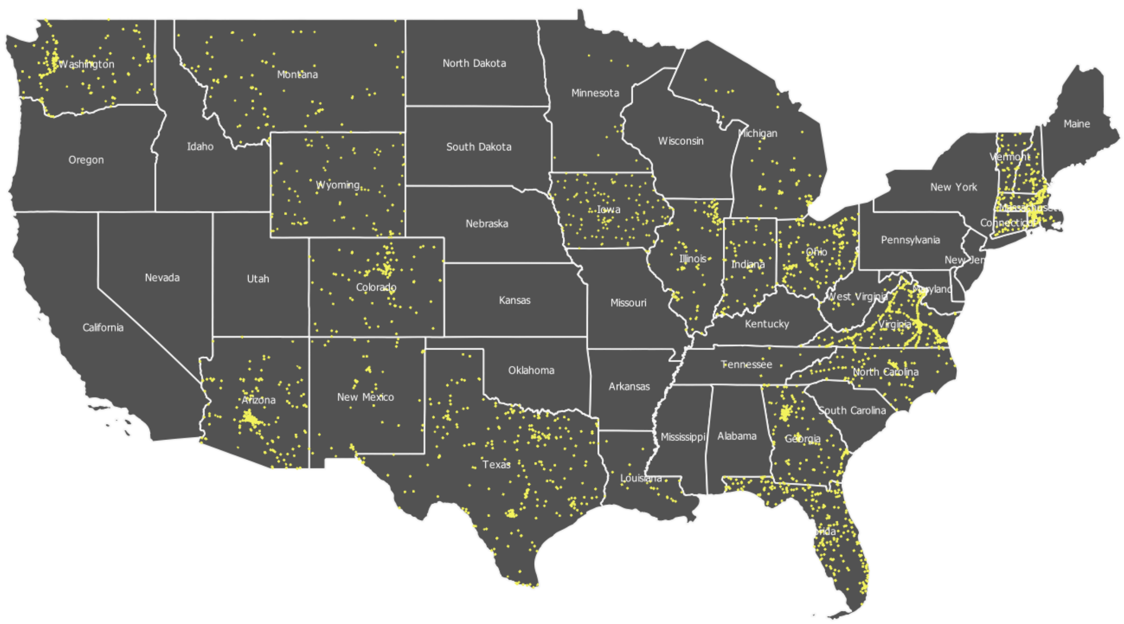

Figure 2: Map of all permanent counters in the U.S. used for training and testing the AADT 2020 Metric.

AADT Range	# Permanent Counters U.S.
Arizona	586
Colorado	292
Connecticut	59
Florida	577
Georgia	469
Illinois	226
Indiana	148
Iowa	135
Louisiana	51
Massachusetts	356
Michigan	153
Minnesota	12
Montana	214
New Hampshire	128
New Mexico	125
North Carolina	248
Ohio	434

Rhode Island	208
Tennessee	17
Texas	776
Vermont	87
Virginia	831
Washington	391
West Virginia	39
Wyoming	130

Table 8: Permanent counters by state used for training and testing the AADT 2020 Metric.

Selecting and Testing the Algorithm

We considered and evaluated dozens of different algorithmic approaches when developing our AADT 2020 Metric. In this section, we will provide an overview of our major decisions.

First, we checked to see how our normalized LBS trips, which comprise our strongest and 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 (compared to scaling to counts from LBS trips alone).

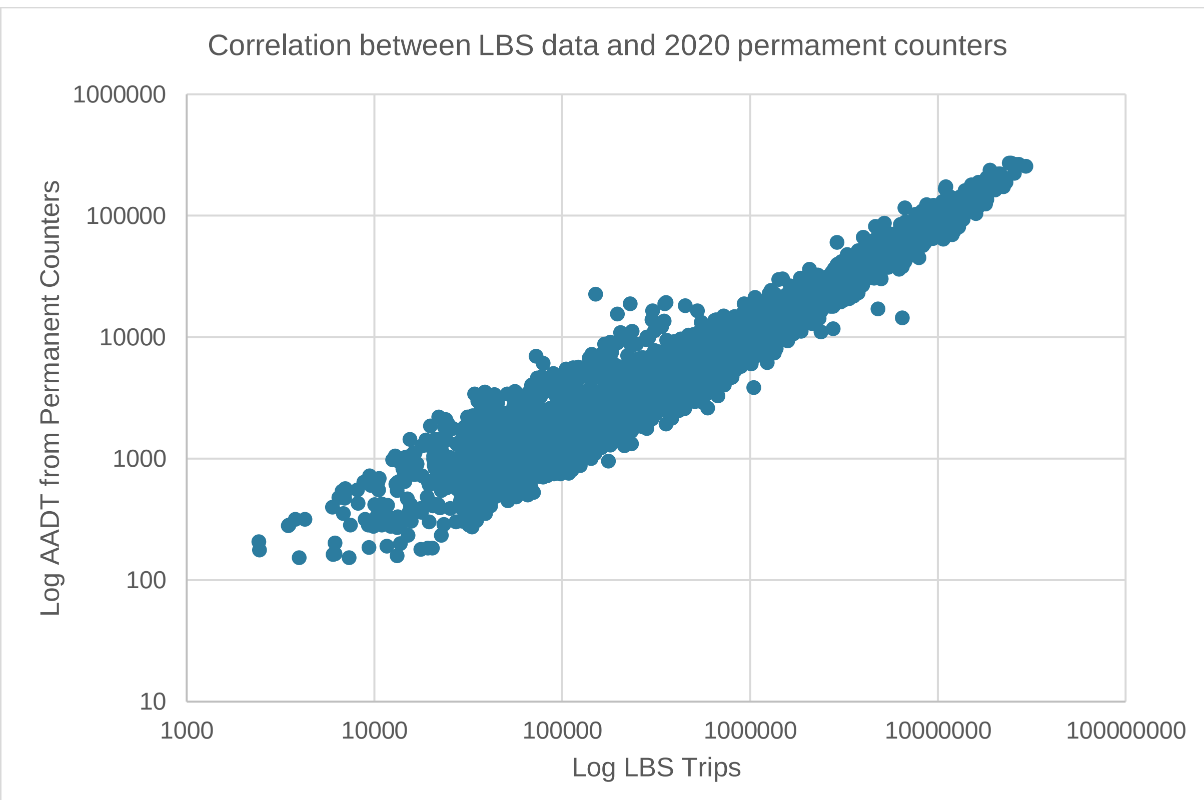

Figure 3: Correlation of population-normalized LBS to permanent loop counter data – the single best predictor value. The rest of the machine-learning work aimed at improving these results.

The choice came down to three options for machine-learning techniques: ordinary least squares (OLS), random forest, and gradient boosting.

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, as it is more or less building a classic y = mx + b style equation. The disadvantage, as we've 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, which we've relied on in our prior years' versions of our AADT models. The benefit of a random forest model for AADT estimation is 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). In the end, we thought that the accuracy and algorithmic robustness for unusual roads and outliers were more important.

Finally, we explored gradient boosting and extreme gradient boosting models, which are also tree-based models, like random forest. We found these options to have less spread of errors, and lower NRMSE as compared to the random forest algorithm. The gradient boosting algorithm adds additional complexity to the random forest algorithm by fitting errors as the model is built, which can further boost model performance. Extreme gradient boosting has faster model run-time than traditional gradient boosting algorithms, and because the model error was similar, we selected the extreme gradient boosting model as the final model for 2020 AADT estimation. For more information on extreme gradient boosting, see our summary in Box 1 below.

Box 1: The Extreme Gradient Boosting Model

An Extreme Gradient Boosting model is a specific implementation of the gradient boosting method. This approach is similar to a decision tree, but it uses several decision trees. For example, let's predict whether a patient entering an emergency room is at high risk. 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. Gradient boosting uses a lot of decision trees (say, an ensemble), where each tree is a little bit different from the others. When a new patient arrives, we take the majority vote of the decision tree ensemble to get a final result. Gradient boosting models build each tree one at a time, while alternatives like random forest models build each tree independently.

The different trees use random samples of observations and subsets 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 the results are combined, we get a "diverse" answer. The idea behind this model is that a bunch of poor decision-makers put together in a room to form a committee will start making better decisions. If each decision-maker comes with a different perspective, that creates better results.

With our 2019 AADT Metric, we explored the benefits of using a hybrid-model approach, where the bulk of the results are predicted with a Random Forest model, but the tail ends of the model are predicted using linear regression. With our AADT 2020 Metric, we decided to take a similar hybrid approach, this time applying the Extreme Gradient Boosting model in combination with other approaches. After outlier removal, all 6,692 unique counter data points were trained for the main gradient boosting model, 3,286 of those locations were used for the lower gradient boosting model, and 136 of those locations were used for the upper linear regression.

This hybrid model provides a few benefits. First, it allows us to predict data beyond the range of permanent counters in our training data set. This means that if the highest AADT value in our permanent counter data set was 200,000, we'd still be able to predict values above that threshold. Second, we found that the extreme gradient boosting model alone struggled with performance on lower-volume roads. Having a specialized extreme gradient boosting model specifically for these lower-volume roads improved our prediction capabilities. Figure 4 below illustrates this new framework, where the low AADT Extreme Gradient Boosting model was trained on sites with AADT under 11,000, and the upper linear regression model was trained on sites with AADT above 120,000.

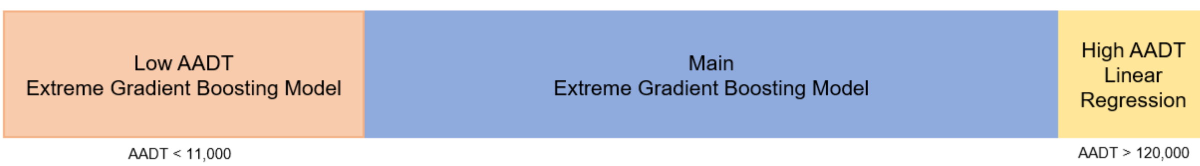

Figure 4: Structure of the hybrid model approach where an extreme gradient boosting model is combined with a specialized low AADT model and an upper linear regression tail. The model can predict values between 0 and infinity.

The next step was to decide which features (input variables) to include in our models. We tested hundreds of combinations. We wanted the most accurate results, but we wanted our algorithm to be scalable to anywhere in the U.S., and to be computationally efficient. In the end, we built three models that relied on a combination of 26 features captured from different data sources. We strive to avoid over-fitting by 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.

Figure 5 below illustrates a high-level flow diagram of how Big Data, specifically Location-Based Services (LBS) and navigation GPS data, and machine-learning models can be used to estimate AADT. First, the machine-learning model is trained to learn the relationship between the AADT derived from permanent counts and Big Data, along with contextual features influencing the AADT. Next the hyperparameters of the model are tuned through cross-validation to enhance the model performance and avoid overfitting. Finally, the model is applied to new stations with the input Big Data and contextual features and produces the estimated AADT.

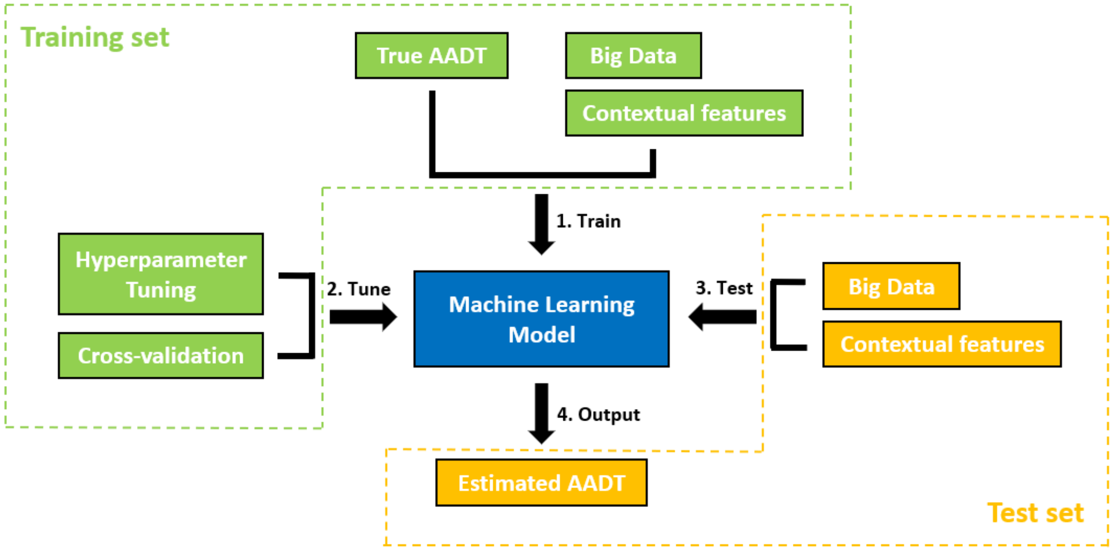

Figure 5: Flow diagram of AADT estimation using Big Data and machine learning.

Our AADT 2020 Metric in the StreetLight InSight® platform also includes a 90% confidence range for each AADT estimate to help locate the true AADT value. Our confidence range (also known as prediction interval) is an estimate of the interval within which the true AADT is expected to lie 90% of the time. To estimate the confidence range, the percent error and log of the predicted AADT values from the cross-validated data set were fit to a quantile regression. For each location in the cross-validated dataset, the predicted AADT value was mapped to the quantile regression, to determine the upper and lower confidence range limit for a 90% percentile confidence range. Figure 6 is a visual representation of the confidence ranges as they apply to high-, medium- and low-volume roads.

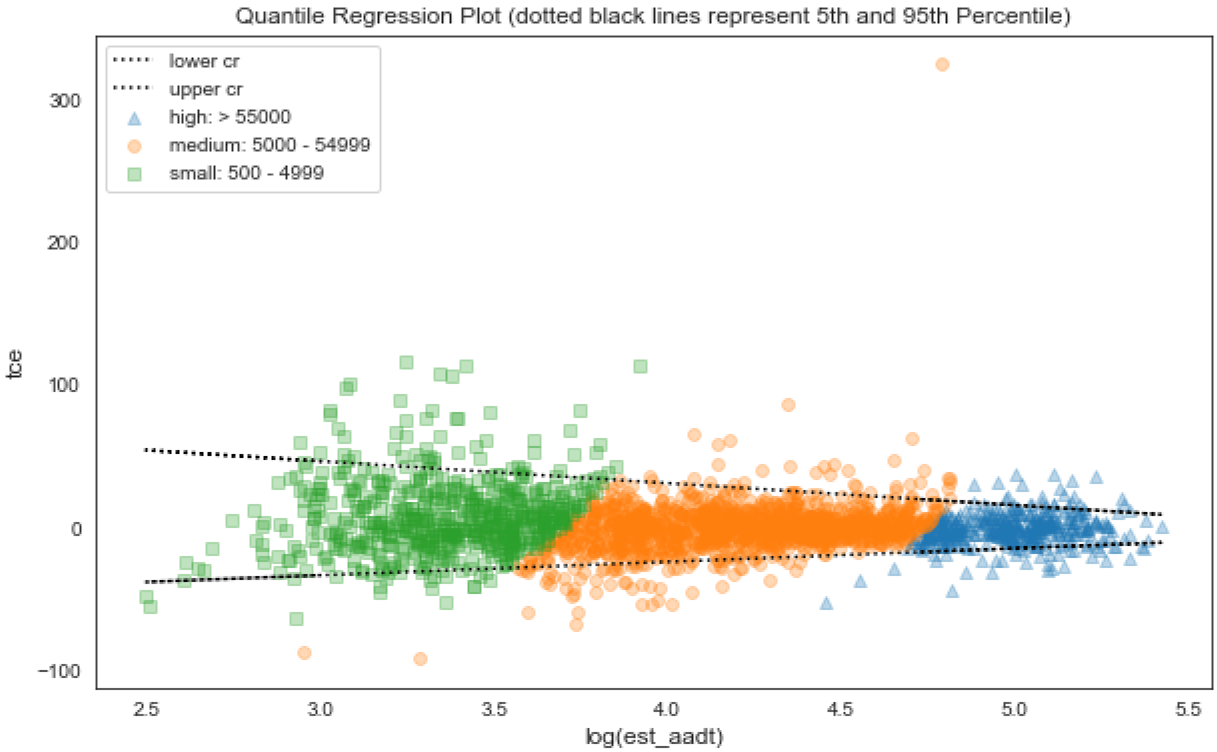

Figure 6: Quantile regression plot of the AADT 2020 model error for bi-directional AADT counts across three groupings of road volume.

Running AADT 2020 in StreetLight InSight®

If you have a StreetLight InSight® account with AADT enabled, choose to create a new analysis. Click the "Create Analysis" button under the AADT analysis option. Name your analysis, select the zone sets covering the roads of interest, and choose the AADT year "2020." Then click "Confirm Analysis" to begin processing.

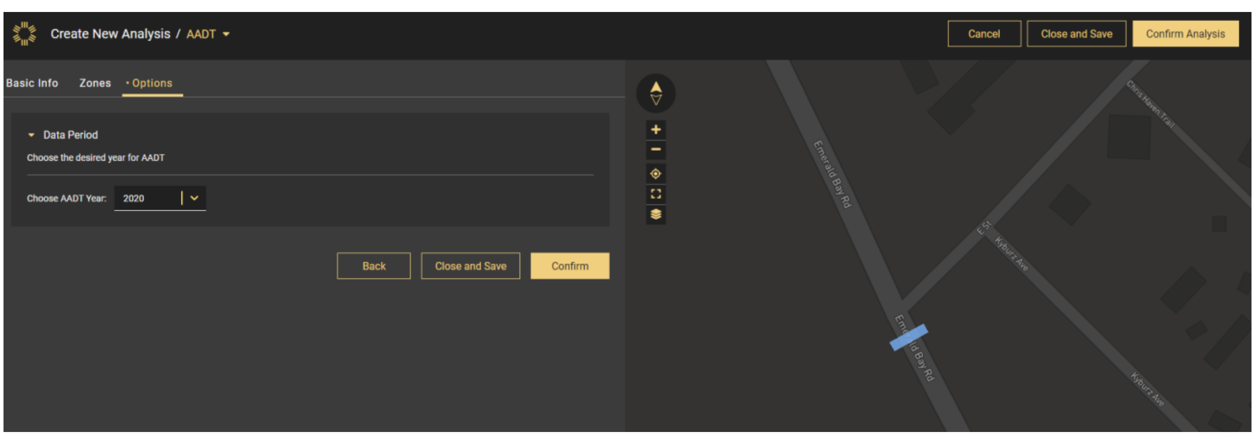

Figure 7: AADT 2020 as represented in the StreetLight InSight® software platform.

Our Next Steps

In future StreetLight InSight® releases, as new data and methods become available, StreetLight may choose to update the AADT model. We expect to at least publish an updated version in mid-to-late summer when the rest of the states publish their data. In addition, we will continue conducting validation and improvement studies on our AADT Metrics.

About StreetLight Data

StreetLight Data, Inc. ("StreetLight") pioneered the use of Big Data analytics to help transportation professionals solve their biggest problems. Applying proprietary machine-learning algorithms to over four trillion spatial data points over time, StreetLight measures multimodal travel patterns and makes them available on-demand via the world's first SaaS platform for mobility, StreetLight InSight®. From identifying sources of congestion to optimizing new infrastructure to planning for autonomous vehicles, StreetLight powers more than 6,000 global projects every month.

STREETLIGHT DATA