EVALUATING THIRD-PARTY DATA FOR STATEWIDE TRAFFIC VOLUME ESTIMATION

Josh Roll, ODOT
Joseph Broach, Metro/PSU

EXECUTIVE SUMMARY

We evaluated StreetLight Data AADT, a third-party proprietary user data product, by comparing the firm's traffic volume estimates with those derived from continuous and short-duration traffic counts throughout Oregon. We also considered other aspects affecting such data's quality, including: completeness, validity, timeliness, and accessibility/usability.

- **Meets expectations of completeness and timeliness; reasonably accessible**
- **High accuracy (<10% MAPE) on highest volume facilities (75,000+ AADT), declining as volumes fall to lower levels (16-20% MAPE, 10,000-75,000 AADT; 22-28% MAPE, 1000-10,000 AADT)**
- **Tracked factored, short-duration counts fairly closely on higher volume facilities (10,000+ AADT) but diverged considerably on lower volume roads**
- **Validity challenging to assess when both sample data and factoring methods proprietary**

OBJECTIVES

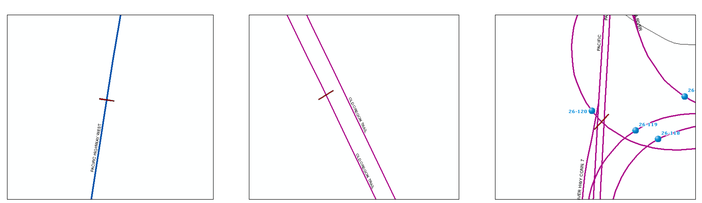
Despite recent advances, reliable traffic volume estimates cover only a small subset of any given transportation network. Recently, proprietary data products have emerged promising much broader network coverage, but quality has not been widely evaluated. We wanted to assess one of these products from a public agency's point of view. We also wanted to test the suitability of a more general evaluation framework in this arena. Are such data products, derived from user location data, a viable alternative to traditional traffic count programs?

DATA AND METHODS

We used three data sources to compare common locations across Oregon:

- **StreetLight Data**: AADT estimates derived from location-based service (in-app location) and navigation (in-vehicle GPS) data, factored using a proprietary algorithm combining census, weather, and permanent traffic counter calibration data (2017 AADT v2 product)
- **Automated traffic recorders (ATRs):** 175 locations (63% outside urban areas)
- **Factored, short-duration counts:** 66 sites in Bend MPO from HPMS reporting

StreetLight Data estimates were extracted via their web platform. Locations are specified as "gates" with specified width submitted as shapefiles. Setting gates properly is sometimes challenging in complex locations, and incorrect gating can have potentially large unwanted impacts on results (**Figure 1**).



(a) (b) (c)

FIGURE 1 Examples showing (a) correct gate, (b) incorrect gate missing one travel lane, and (c) gate at complex location that would "catch" volumes from both the off-ramp and travel lanes [blue dots show ATR locations]

RESULTS

Figure 2 compares the third-party estimates to automated traffic counts for the same 175 locations on a variety of roadways around the state. StreetLight volumes tend to underestimate recorded volumes on all roadway types except for the lowest volume facilities. Accuracy tends to fall with volume, and nearly half of sites reviewed deviate from automated count estimates by more than 20%.

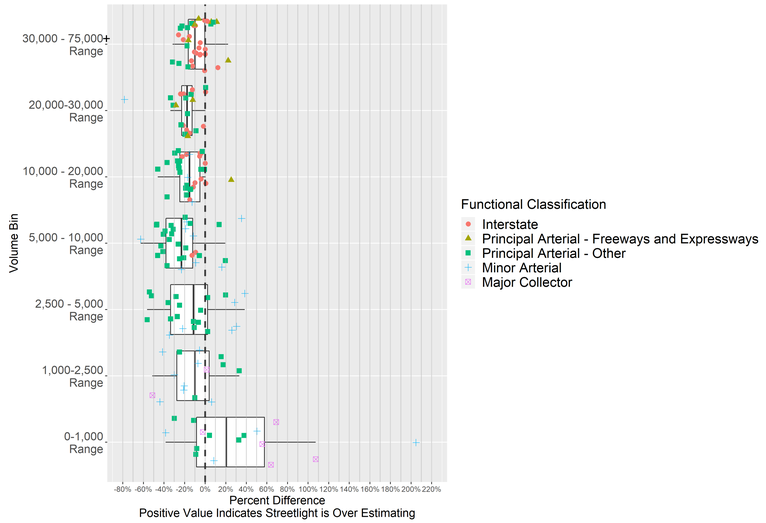


FIGURE 2 Percent error comparing ODOT automated counts and StreetLight AADT (Note: one site in 0-1000 bin not shown due to scale)

Table 1 presents the absolute percent error between StreetLight volumes and ATR counts by AADT bin. Accuracy drops off below 75,000 and again below 10,000 AADT, with estimates on links below 1000 AADT becoming highly erratic. StreetLight Data also supplies site-specific confidence intervals (CIs) around their AADT estimates. Overall, we found the precision of estimates to be somewhat worse than suggested by the provided CIs, and this was true across most volume bins.

TABLE 1 Absolute percent error comparing ODOT automated counts and StreetLight AADT

	Absolute Percent Error					Sites where Error within Streetlight-supplied 90% Confidence Interval
AADT Range	Median	Mean	Min	Max	n	%
75,000 +	5%	6%	0.02%	12%	14	100%
50,000 - 75,000	15%	16%	5.9%	26%	9	78%
30,000 - 50,000	16%	16%	5.6%	32%	13	100%
20,000-30,000	18%	20%	0.4%	78%	22	73%
10,000 - 20,000	17%	17%	0.1%	46%	33	82%
5,000 - 10,000	25%	28%	5.7%	63%	31	65%
2,500 - 5,000	27%	26%	2.3%	56%	21	76%
1,000-2,500	20%	22%	1.4%	51%	15	93%
0-1,000	38%	88%	2.2%	758%	17	88%
All Sites	18%	27%	0.02%	758%	175	81%

Figure 3 shows the correspondence between AADT estimates based on short-duration counts in Bend (medium-sized MPO) and StreetLight Data's estimates at the same locations. Streetlight estimates tracked the factored counts fairly closely for roadways over 10,000 AADT, with some downward bias. For lower-volume roads, estimates diverged considerably, with Streetlight showing a strong upward bias.

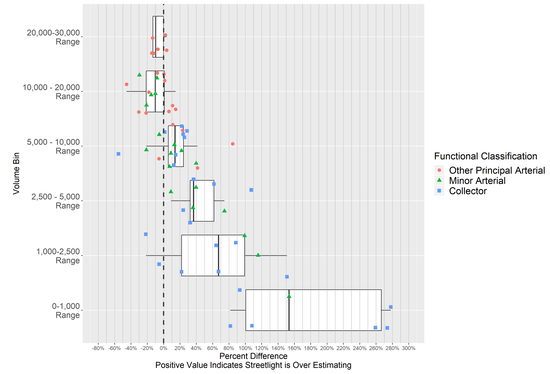


FIGURE 3 Percent difference between ODOT short-term AADT estimates and StreetLight AADT

DISCUSSION

Streetlight Data's AADT tool was able to provide usable volume estimates for all available traffic monitoring sites. The user-selected "gates" provide excellent flexibility to define locations but also introduce an additional potential source of error. Accessing the data was not overly burdensome, requiring an estimated 60 agency staff hours total for the 241 sites reviewed.

Accuracy was fairly good on higher volume facilities, approaching the expected performance of expanded 48-hour counts on average. At the same time, nearly half of reviewed locations had error rates in excess of 20%, and there was some evidence of bias (mostly undercounting) in the estimates as well as a fair number of outliers. StreetLight's own confidence intervals appear somewhat too narrow. Whether the current accuracy would meet agency needs will depend on the roadway and application context.

Validity of third party data products is difficult to assess when both the base data and methodology are largely proprietary. Furthermore, the nature of this data, ultimately derived from users' personal location information may raise additional issues for agencies around public trust.

Future work will seek to better distinguish and quantify context-dependent sources of error, evaluate the AADT estimates in an application setting, and potentially expand the comparison to other user data products.

ACKNOWLEDGMENTS

The paper authors would like to acknowledge the staff from StreetLight Data, who assisted in processing some of the inputs, and the Portland State University Faculty Association (PSUFA), which provided funding in support of the project.