I-5 and I-205 Toll Projects

MEMORANDUM

Date	10/6/20
To	Equity and Mobility Advisory Committee
From	Toll Projects Team
Subject	Purpose of use of StreetLight data on Toll Projects

StreetLight is a proprietary tool that uses smart devices (e.g., smart phones) to help the Toll Program understand existing driving patterns and user demographics within the corridor study area. For the I-205 Toll Project, this tool is envisioned to be used to provide a "snapshot" of driver behavior when faced with congestion. The information was collected by the I-205 Toll Project team for the pre-COVID months of February and March of 2019, and August and September of 2019. The trends identified in this information are used to conduct analyses and consists of the items below:

- Trip counts.
- Routing information.
- "Likely home location" of smart devices. Likely home location is determined by a device pining overnight within a residential area, the actual location is not known, only the likely Census block or block group.
- Demographic information (taken form the home location) of the smart device is identified through the distribution of race, income and other demographic information provided from the Census and American Community Surveys of the likely home block or block group. It is important to note that the demographics data is linked to smart devices that generate location-based data and not directly linked to individual trips.

It is important to recognize that while StreetLight's data is based on actual, historic information collected from travelers, it a) only provides a sample of the total trips being made, and b) also requires normalization and expansion of location-based data via a variety of algorithms. The identity of individual drivers or smart devices are not known by the project team. As such, its accuracy may be questioned. However, research to-date has shown that the accuracy of StreetLight's data for such data as trip origins and destinations and other travel pattern information increases as the sample size increases.

Additionally, since StreetLight's data is dependent upon smart phone tracking, it is possible that some inherent biases in the sample base may occur because a higher proportion of members of certain demographic groups may not be able to afford or desire smart phones; hence, these groups may be under-represented in StreetLight data.

Despite the limitations noted above, the project team believes this information source provides useful indicators for understanding travel patterns and user characteristics in the Project Corridor.

STREETLIGHT InSight

Our Methodology and Data Sources

Updated July 2019

StreetLight InSight® Metrics: Our Methodology and Data Sources

This white paper describes the data sources and methodology employed by StreetLight Data to develop travel pattern metrics. This document is relevant for all StreetLight InSight® metrics, whether they are available via the *StreetLight InSight* platform, via data API, or via custom delivery.

Table of Contents

Locational Data Sources and Probe Technologies 1

Data Processing Methodology 4

Measuring Sample Size 8

Locational Data Sources and Probe Technologies

StreetLight Data's metrics are currently derived from two types of locational "Big Data:" navigation-GPS data and Location-Based Services (LBS) data. StreetLight has incorporated and evaluated several other types of mobile data supply in the past, including cellular tower and ad-network derived data.

As the mobile data supply landscape has evolved and matured over time, we have determined that a combination of navigation-GPS data and LBS data is best suited to meet the needs of transportation planners. Our team phased out the use of cellular tower data because its low spatial precision and infrequent pinging frequency did not meet our standards for use in corridor studies, routing analyses, and many other Metrics. LBS data is suitable for these studies and offers a comparable sample size to cellular tower data.

As of July 2019, StreetLight's data repositories process analytics for about 79M devices, or ~28% of the adult U.S. and Canadian population, and about 13% of commercial truck trips. As detailed later in this report, sample size varies regionally, historically and by type of analysis conducted.

Our data supply grows each month as updated data sets are provided by suppliers. We currently use one major navigation-GPS data supplier, INRIX, and one LBS data supplier, Cuebiq. See Table 1, below, for more details on the different locational data sources StreetLight Data has recently evaluated.

Type	Pros	Cons	Notes
Cellular Tower: Derived from cellular tower "triangulation" and/or "multi-lateration" (100-2000m spatial precision)	Large sample size - Most telecom providers have over 30M devices Ability to infer home and work locations	Very poor spatial precision (average of several hundred meters) Infrequent pings for some suppliers High cost Consumers typically opt-out of data collection (vs. opt-in) No differentiation of personal and commercial trips Poor coverage in rural areas No capture of short trips No ability to reliably infer active modes of transportation	We haven't seen the U.S. cellular industry making investments to improve these weaknesses.
In-Vehicle Navigation-GPS: From connected cars and trucks (3-5m spatial precision)	Excellent spatial precision Very frequent pings Separates personal and commercial trips Opt-in for consumers	Usually lower sample size Difficulties inferring home/work (depending on supplier practices) No non-vehicular modes	This data has been traditionally used for speed products.
Location-Based Services: Mix of navigation-GPS, aGPS, and sensor proximity data from apps that "foreground" and "background" with locational data collection (5-25m spatial precision)	Very good spatial precision Frequent ping rate Superior ability to infer trip purpose and trip chains Ability to infer modes (walk/bike/transit/Gig Driving) accurately Large and growing sample size Opt-in for consumers	Less mature suppliers Variation in sample size and characteristics across suppliers requires more sophisticated data processing	Several players are emerging in this new market with very large sample sizes, opening up the possibility of a healthy, competitive supply base.
Ad-Network Derived Data: When user sees an ad on their phone, their location is recorded by the ad-network	Large sample size of individuals	Few pings per month mean inference of travel patterns is not feasible	This source should not be used until significant changes are made.

Table 1 – Overview of Big Data supply options for transportation analytics. StreetLight recommends and uses a mix-and-match approach currently focused on navigation-GPS and LBS data types.

Our Navigation-GPS and LBS Data Sources

In this section, we will explain why access to two different Big Data sources is uniquely beneficial for transportation professionals. First, it is important to note that *StreetLight InSight* is:

- The first and only on-demand platform for planners to process Big Data into customized transportation analytics to their unique specifications, including the type of Big Data they would like to use.
- The first and only online platform that automatically provides comprehensive sample size information for analyses. (See more information on sample size on page 8 of this report.)

We selected navigation-GPS and LBS data because they are complementary resources that provide unique and valuable travel pattern information for transportation planning. See Figure 1 below for a visualization of these data sources.

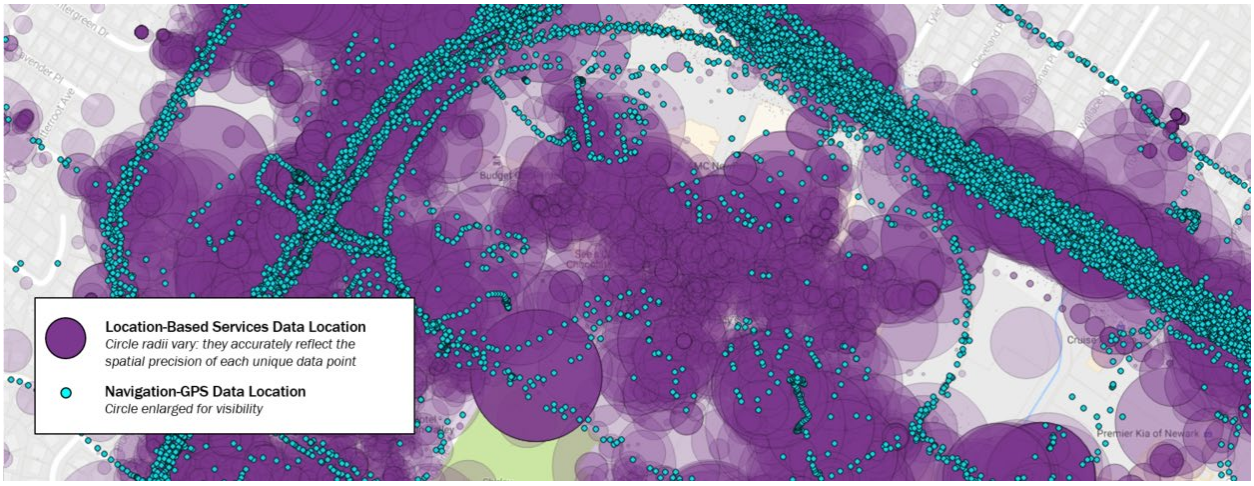

Figure 1 – Filtered visualization of a subset of unprocessed navigation-GPS and LBS data near a mall in Fremont, California.

Location-Based Services (LBS) Data

LBS data can be processed into personal travel patterns at a comprehensive scale. Its fairly high spatial precision and regular ping rate allow for capturing trips as well as activity patterns (i.e., home and work locations), trip purpose, and demographics. This makes it an ideal alternative to data derived from cellular towers, which also has a large sample size but unfortunately lacks spatial precision and pings infrequently.

Cuebiq, our LBS data supplier, provides pieces of software (called SDKs) to developers of mobile apps to facilitate LBS. These smartphone apps include couponing, dating, weather, tourism, productivity, locating nearby services (i.e., finding the closest restaurants, banks, or gas stations), and many more apps, all of which utilize their users' location in the physical world as part of their value. The apps collect anonymous user locations when they are operating in the foreground. In addition, these apps may collect anonymous user locations when operating in the background. This "background" data collection occurs when the device is moving. LBS software

collects data with WiFi proximity, a-GPS and several other technologies. In fact, locations may be collected when devices are without cell coverage or in airplane mode. Additionally, all the data that StreetLight uses has better than 20-meter spatial precision. (Similarly, our partner INRIX collects some LBS data from navigation-oriented smart phone apps).

Navigation-GPS Data

Navigation-GPS data has a smaller sample size than LBS data, but it does differentiate commercial truck trips from personal vehicle trips. This makes navigation-GPS data ideal for commercial travel pattern analyses. Navigation-GPS data is also suitable for very fine resolution personal vehicle travel analyses (e.g.: speed along a very short road segment) because of its extremely high spatial precision and very frequent ping rate.

INRIX, our navigation-GPS data supplier, provides data that comes from commercial fleet navigation systems, navigation-GPS devices in personal vehicles, and turn-by-turn navigation smartphone apps. (These apps produce data that are like the LBS data described above). Segmented analytics for medium-duty and heavy-duty commercial trucks are available. For commercial trucks, if the vehicle's on-board fleet management system is within INRIX's partner system, INRIX (and thus StreetLight) will collect a ping every one to three minutes whenever the vehicle is on, even if the driver is not actively using navigation.

For personal vehicles, if the vehicle is in INRIX's partner system and has a navigation console, INRIX (and thus StreetLight) will collect a "ping" every few seconds whenever the vehicle is on, even if the driver is not actively using the navigation system. This provides a very complete picture of vehicles' travel patterns and certainty that the trips are in vehicles.

Data Processing Methodology

The following section contains an overview of the fundamental methodology that StreetLight Data uses to develop all metrics. Each *StreetLight InSight* metric has specific methodological details which can be shared with clients as needed by request.

Step 1 – ETL (Extract Transform and Load)

First, we pull data in bulk batches from our suppliers' secure cloud environments. This can occur daily, weekly, or monthly, depending on the supplier. The data do not contain any personally identifying information. They have been de-identified by suppliers before they are obtained by StreetLight. StreetLight Data does not possess data that contains any personally identifying information.

The ETL process not only pulls the data from one environment securely to another, but also eliminates corrupted or spurious points, reorganizes data, and indexes it for faster retrieval and more efficient storage.

Step 2 – Data Cleaning and Quality Assurance

After the ETL process, we run several automated, rigorous quality assurance tests to establish key parameters of the data. To give a few examples, we conduct tests to:

- Verify that the volume of data has not changed unexpectedly,
- Ensure the data is properly geolocated,
- Confirm the data shares similar patterns to the previous batch of data from that particular supplier.

In addition, StreetLight staff visually and manually reviews key statistics about each data set. If anomalies or flaws are found, the data are reviewed by StreetLight in detail. Any concerns are escalated to our suppliers for further discussion.

Step 3 – Create Trips and Activities

For any type of data supply, the next step is to group the data into key patterns. For example, for navigation-GPS data, a series of data points whose first time-stamp is early in the morning, travels at reasonable speeds for a number of minutes, and then stands still for several minutes, could be grouped into a probable “trip.” For LBS data, we follow a similar approach. However, since LBS data continues to ping while the device is at the destination, we see clusters of pings in close proximity at the beginnings and ends of trips.

Step 4 – Contextualize

Next, StreetLight integrates other “contextual” data sets to add richness and improve accuracy of the mobile data. These include road networks and information like speed limits and directionality, land use data, parcel data, and census data, and more.

For example, a “trip” from a navigation-GPS or LBS device is a series of connected dots. If the traveler turns a corner but the device is only pinging every ten seconds, then that intersection might be “missed” when all the device’s pings are connected to form a complete trip. StreetLight utilizes road network information including speed limits and directionality, to “lock” the trip to the road network. This “locking” process ensures that the complete route of the vehicle is represented, even though discrepancies in ping frequency may occur. Figure 2, below, illustrates this process.

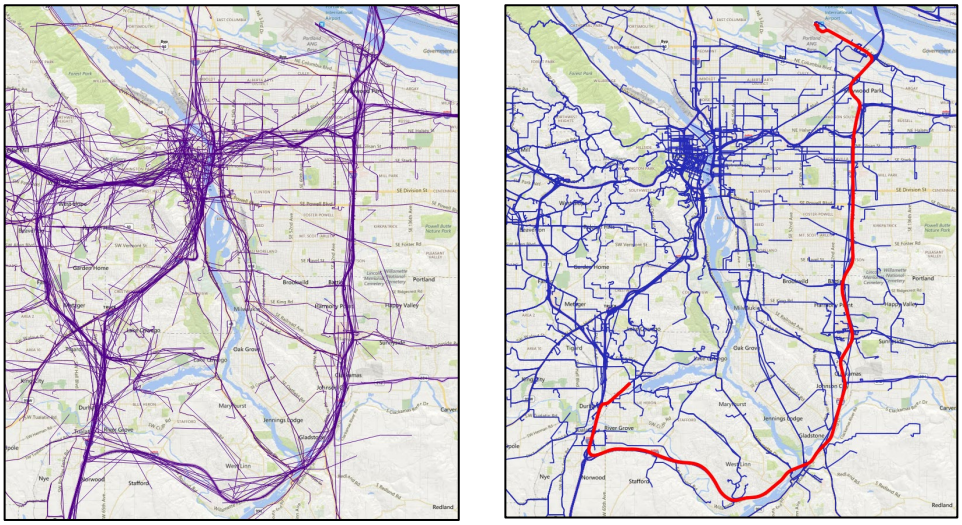

Figure 2: “Unlocked” Trips becoming locked trips.

As another example, if a device that creates LBS data regularly pings on a block with residential land use, and those pings often occur overnight, there is a high probability that the owner of the device owner lives on that block/block group. This allows us to associate “home-based” trips and a “likely home location” to that device. In addition, we can append distribution of income and other demographics for residents of that census block to that device. That device can then “carry” that distribution everywhere else it goes. (Our demographic data sources for the U.S. are the Census and American Community Surveys. In Canada, our source is Manifold Data.) This allows us to normalize the LBS sample to the population, and to add richness to analytics of travelers such as trip purpose and demographics.

Step 5 – More Quality Assurance

After patterns and context are established, additional automatic quality assurance tests are conducted to flag patterns that appear suspicious or unusual. For example, if a trip appears to start at 50 miles per hour in the middle of a four-lane highway, that start is flagged as “bad.” Flagged trips and activities are not deleted from databases altogether, but they are filtered out from *StreetLight InSight* queries and metrics.

Step 6 – Normalize

Next, the data is normalized along several different parameters to create the StreetLight Index. As all data suppliers change their sample size regularly (usually increasing it), monthly normalization occurs.

For LBS devices, we perform a population-level normalization for each month of data. For each census block, StreetLight measures the number of devices in that sample that appear to live there, and makes a ratio to the total population that are reported to live there. A device from a census block that has 1,000 residents and 200 StreetLight devices will be scaled differently everywhere in comparison to a device from a census block that has 1,000 residents and 500

StreetLight devices. Thus, the StreetLight Index for LBS data is normalized to adjust for any population sampling bias. It is not yet “expanded” to estimate the actual flow of travel.

For navigation-GPS trips, StreetLight uses a set of public loop counters at certain highway locations to measure the change in trip activity each month. Then it compares this ratio to the ratio of trips at the location, and normalizes appropriately. In addition, StreetLight systemically performs adjustments to best estimate total, normalized trips based on external calibration points. Such calibration points include public, high-quality vehicle count sensors (for example, those in PEMs systems, or the TMAS repository) as well as reports from surveys and other externally validated sources. Thus, the StreetLight Index for GPS data is normalized to adjust for change in our sample size. It is not normalized for population sampling bias (because we cannot infer home blocks for GPS data). This is one of the reasons we recommend LBS data for all personal travel analytics. The StreetLight Index for GPS data is not yet “expanded” to estimate the actual flow of travel.

Step 7 – Store Clean Data in Secure Data Repository

After being made into patterns, checked for quality assurance, normalized, and contextualized, the data is stored in a proprietary format. This enables extremely efficient responses to queries via the *StreetLight InSight* platform. By the time the data reaches this step, it takes up less than 5% of the initial space of the data before ETL. However, no information has been lost, and contextual richness has been added.

Step 8 – Aggregate in Response to Queries

Whenever a user runs a metric query via *StreetLight InSight*, our platform automatically pulls the relevant trips from the data repository and aggregates the results. For example, if a user wants to know the share of trips from origin zone A to destination zone B vs. destination zone C during September 2017, they specify these parameters in *StreetLight InSight*. Trips that originated in origin zone A and ended in either destination zone B or destination C during September 2017 will be pulled from the data repositories, aggregated appropriately, and organized into the desired metrics.

Results always describe aggregate behavior, never the behavior of individuals.

Step 9 – Final Metric Quality Assurance

Before delivering results to the user, final metric quality assurance steps are automatically performed. First, *StreetLight InSight* determines if the analysis zones are appropriate. If they are nonviable polygon shapes, outside of the coverage area (for example, in an ocean) or too small (for example, analyzing trips that end at a single household) the zone will be flagged for review. If a metric returns a result with too few trips or activities to be statistically valid or to protect privacy, the result will be flagged. When results are flagged, StreetLight’s support team personally reviews the results to determine if they are appropriate to deliver from a

statistical/privacy perspective. The support team then personally discusses the best next steps with the user.

In general, *StreetLight InSight* response time varies according to the size and complexity of the user's query. Some runs take two seconds. Some take two minutes. Some take several hours. Users receive email notifications when longer projects are complete, and they can also monitor progress within *StreetLight InSight*. Results can be viewed as interactive maps and charts within the platform, or downloaded as CSV and shapefiles to be used in other tools.

Measuring Sample Size

StreetLight's Big Data resources include about 79M devices in the U.S. and Canada, which covers approximately 28% of these countries' combined adult population. However, clients should not expect a 28% penetration rate for all *StreetLight InSight* analyses they run. Penetration rates for individual analyses can range from as small as 1% to as large as 35%.

As is the case with any Big Data provider, sample size and penetration rate for a given analysis depend on the specific parameters used in the study. The reason is that some data are useful for certain analyses, but are not useful for others. For example, a device may deliver high-quality, clean location data for one study, but messy, unusable location data – or no data at all – for another. Efficiently identifying the data that are "useful" for a particular analysis is a critical component of the data science value that differentiates StreetLight Data. Because penetration rates vary, sample sizes are automatically provided for almost all *StreetLight InSight* analyses[1]. This allows users to calculate penetration rates and to better evaluate the representativeness of the sample. Sample size values also are useful to clients who wish to normalize *StreetLight InSight* results through additional statistical analysis.

For LBS analyses, sample size is currently provided as the number of unique devices and/or number of trips for LBS analyses, depending on the type of analysis. These values should be thought of as most similar to "person trips." Including both the number of devices and trips for all LBS analyses is in our product roadmap. Sample size is provided as number of trips for navigation-GPS analyses. These should be thought of as "vehicle trips."

---

[1] Sample sizes are not automatically provided for AADT or Traffic Diagnostics Projects. They are available by request. These analyses use a very large volume of location data, so providing sample sizes automatically via *StreetLight InSight* would negatively impact data processing speeds.

In general, though not always, the trip sample size for commercial navigation-GPS data will be higher than the device (truck) sample size. Commercial trucks that are in active use typically take many trips per week that are often on set routes; thus, they are more likely to have up-to-date fleet management tools, and that means they are more likely to be included in StreetLight's navigation-GPS data set. Trucks that are more rarely used are less likely to be included in the data set.

In general, though not always, the trip sample size for LBS data will be lower than the device (person) sample size. The reason is that not all devices in StreetLight's database capture every single trip perfectly. To illustrate, consider this hypothetical example:

- 8:00AM: Device creates location data at expected home location
- 2:00PM: Device creates location data at sports arena

This device has created useful information for analyzing the home locations of visitors to the arena. However, since the device didn't create any location data on the trip to arena, perhaps because it was off, then the route taken and the travel time cannot be calculated with certainty. As result, it could not be used in an analysis of road activity on an arterial near the arena.

As another example, consider a device that generates regular pings for each trip taken over 10 days. However, the user deletes the smart phone app that created that data, and it stops pinging. That device then disappears for the last 20 days of the month. The device's data can still be used, but the trip penetration for the month is only 33% of this person's trips, not 100%.

Typical daily trip penetration rates are between 1 and 5% of all trips on any one specific day. StreetLight's pricing and data structure encourage looking at many days of data. The costs are the same for analyzing an average day across three months and analyzing a single day. Thus, we encourage clients to evaluate the total sample across the entire study period instead of focusing on per-day penetration rates.

About StreetLight Data

StreetLight Data 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, StreetLight measures diverse 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 3,000 global projects every month.

STREETLIGHT DATA

STREETLIGHT InSight

StreetLight Volume Methodology & Validation White Paper

Updated August 2019

This white paper provides technical detail about the methodology, algorithm development, validation, and data sources used in StreetLight Data's Volume output. This white paper was first published in August 2019 and is updated periodically as new validation is performed.

Table of Contents

Introduction 3

Methodology 3

- Estimated Volume for Roads 3
  - Data Sources 3
  - Algorithm Details 4
- Estimated Volume for Areas 5
  - Data Sources 5
  - Algorithm Details 5
- Estimated Volume for Origin-Destination Analyses 6

Validation 8

- Zone Activity Volume for Roads 8
  - Data Sources and Methods 8
  - Validation Results: Zone Activity Volume 8
  - Validation Results: Seasonality 10
- Origin-Destination Volume for Roads 11
  - Data Sources and Methods 11
  - Validation Results: Origin-Destination Volume 12

About StreetLight Data 14

Introduction

StreetLight's underlying data sample varies month to month, and the resulting trip counts and normalized Index values, while valuable for cross zone or cross project comparison, do not represent estimated trip counts. The goal of the new StreetLight Volume output is to provide an estimate of average daily traffic, and to allow for time-series analysis, or comparison of actual traffic changes over time. This Volume output provides a quick, easy, and cost-effective way to measure traffic at the yearly, monthly, daily, and even hourly level. Volume estimates can be derived for any location, such as a road, park, TAZ or user-defined special area. It can also be used to estimate zone-to-zone traffic, providing accurate estimates for work like turning movement studies and travel demand models. StreetLight Volume is available for analyses in the U.S., and soon for Canada.

Methodology

Estimated Volume for Roads

DATA SOURCES

In order to create an estimate of the actual number of cars on the road at a variety of points in time, the analysis combined multiple models to create optimal results. At a high level: StreetLight's machine-learning models predict expected seasonal changes at a location over time, and use the Streetlight Data AADT (annual average daily traffic) to calibrate seasonal changes to an estimated volume.

Following is a brief overview of StreetLight AADT methodology and data sources. To get more detailed information, please refer to the StreetLight AADT white paper.

The StreetLight AADT blends together the following data sources to provide the best prediction of annual average daily traffic at a given location:

1. Location-Based Services trips.
2. Navigation GPS trips - personal and commercial.
3. U.S. Census and Manifold super demographics which are derived from Statistics Canada.
4. Open Street Maps data reflecting road classification, density of commercial activity, and more.
5. Weather data.
6. AADT counts, derived from permanent traffic recorders, including a mix of small and large, urban and rural locations. StreetLight uses 11,000+ counts across the U.S. and Canada to develop and validate AADT.

Using a combination of the features described above, the analysis applied a Random Forest model to estimate AADT at each location. It then performed several types of cross-validation to ensure the model worked well in different scenarios (across states, road types etc.) The validation work proved that the actual and estimated AADT values through the cross validation were correlated with a very high R^2 (.96) which indicates that the performance of the model is excellent without bias.

In order to estimate variation in traffic volume across time, analysis relied on permanent traffic recorders (PTR) deployed on roads across the U.S. which count the number of cars constantly. This constant counting allows StreetLight to evaluate monthly average daily traffic (MADT) metrics to assess monthly variation in trip volume at a particular location.

Creating a monthly traffic model demanded promptly published data on how many cars were historically present on a road each month. Quality data at the monthly level is not readily available from all states, thus StreetLight had to narrow PTR counters to those that met a high standard of frequency and quality. This left 474 counters across eight states: Colorado, Georgia, Indiana, Michigan, Massachusetts, Montana, Ohio, and Rhode Island. This is a subset of those used in the AADT calibration process.

Algorithm Details

With the MADT data derived from counter locations across the county, a distinct linear model was trained and generated for each month of the year. Using a series of spatial and temporal features, the linear model predicts a seasonal factor for the expected change in traffic for that month relative to the yearly average for 2018 (AADT). Seasonal factors are represented as a percent change from the yearly average, so a month with more traffic than the yearly average will have a positive seasonal factor (say +10%), while a month with less traffic than the yearly average will have a negative seasonal factor (say -15%).

The resulting model allows us to ingest monthly data samples that vary in size, and then predict monthly trip volumes that correspond with seasonal variation. In Figure 1, LBS trips at a single location are translated into “seasonal change” in LBS across months (left, green). Each dot represents data from a different month at a specified location. In running the model, the seasonal factor can be translated into MADT (right, blue).

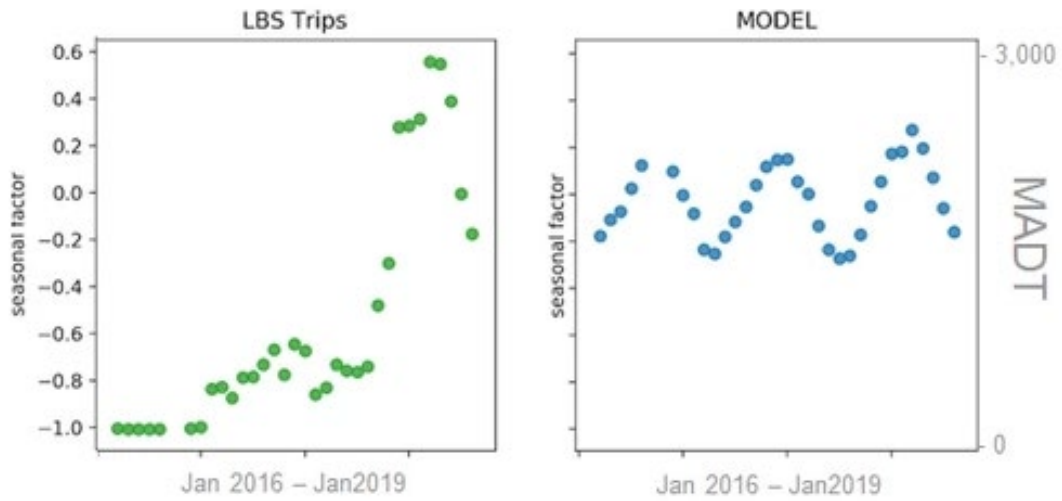

Figure 1: Unadjusted LBS sample trip counts (left) show sample growth over time vs. MADT model output, which corrects and normalizes the input data.

Estimated Volume for Areas

DATA SOURCES

Calibrating LBS data to the volume of large areas is less straightforward than calibrating to expected road volume without reliable "truth" data representing the real-world number of trips that start or end in large areas. The most consistent and reliable validation and training data available is for roads. Thus, StreetLight used its well-validated method of estimating traffic on roads to infer expected volume to areas.

In order to estimate trips to or from an area, the process followed this high-level method:

1. Sample nearby roads with trips in the zone area.
2. Obtain an estimate of MADT for the sampled road, as described previously.
3. Use the estimated MADTs from the roads near the area to calibrate and generate an estimate of volume in the area.

ALGORITHM DETAILS

In order to estimate volume for a specified area, the algorithm selects a subset of roads with trips that start, end, or pass through the zone area. See Figure 2 for an example, where a specified area (shaded) is accompanied by a subset of randomly sampled roads (orange gates) in the surrounding area. The number of sample roads will depend on the size and location of the area zone.

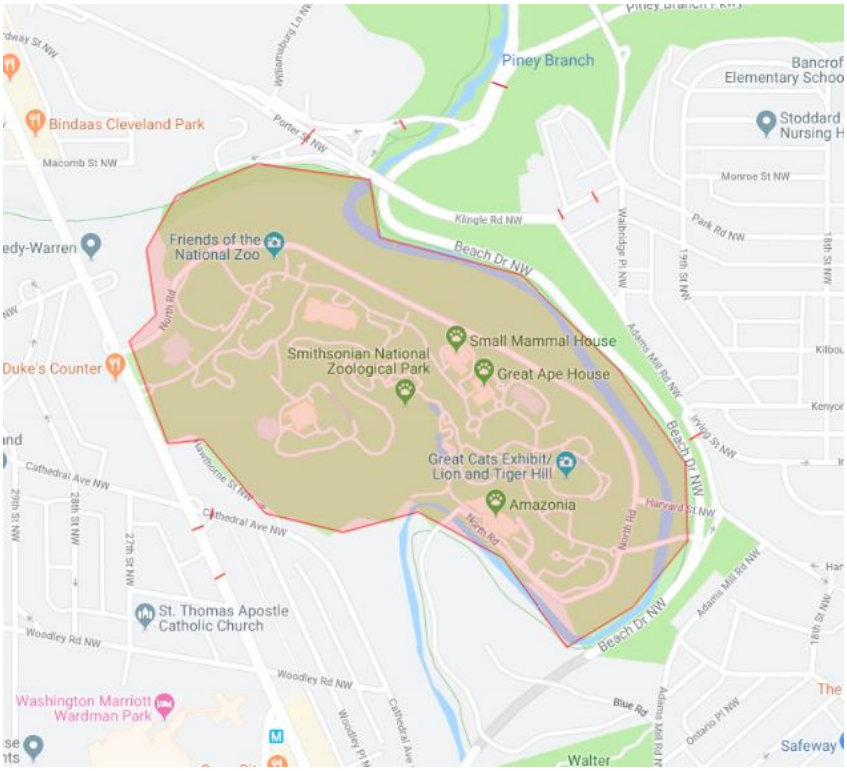

Figure 2: Example area zone with selected gates (red lines) used to calculate MADT for trip starts and stops to the area.

For each sampled road, the system will do the following:

1. Run a pass-through Zone Activity analysis for an estimate of MADT from each sampled road.
2. Use the ratio of LBS through the road, and LBS trips through the zone area to estimate zone area volume. This is based on the assumption that: LBS road/LBS area = actual road/actual area.
3. Calculate the weighted average volume estimate from all the sub-sampled roads to choose a final estimated StreetLight Volume.

Based on seasonal factors associated with the months included in the analysis, this results in an estimated volume for the defined area based on trip starts and ends.

Estimated Volume for Origin-Destination Analyses

Once Volume outputs were estimated for individual zones (both pass-through and area zones) these were applied to origin-destination analyses, which allowed for evaluating how many trips span between locations. The goal is to generate an O-D Volume that allows for comparisons across time, and provides a number that represents a reasonable estimate of the real-world number of trips.

This was accomplished via the following approach:

1. Calculate the total Zone Activity Volume for each O-D zone (described in the previous sections).
2. Return the LBS trip counts between each O-D zone.
3. Use Iterative Proportional Fitting (IPF) to scale the LBS O-D counts to Volume based on the estimated volume at each O-D.

Iterative Proportional Fitting is a technique used to adjust the counts in a table so that they add up to specified totals (or "marginal totals") for both columns and rows. In this case, the adjusted data (called "seed" cells) is the LBS trip counts between each O-D pair. Using an adjusted Zone Activity Volume for each O-D as the marginal totals, then scaling the LBS trip counts with IPF adds up to the expected Zone Activity Volumes. This approach follows well-established practices in the transportation industry.[1]

In addition to a two-dimensional matrix used in an O-D project, the IPF technique can also be applied to a three-dimensional matrix to derive volume estimates for an Origin-Destination with Middle Filter (ODMF) zone configuration.

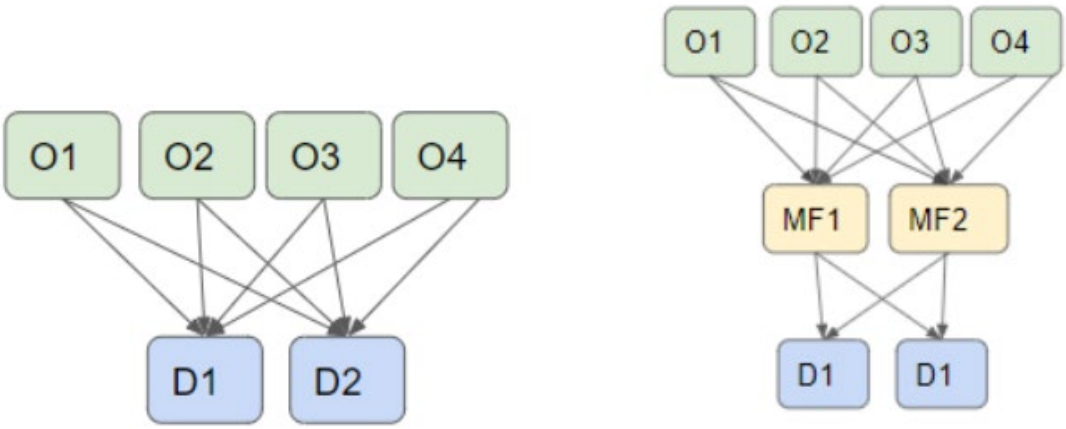

Figure 3: Example Origin-Destination analysis configuration (left) and Origin-Destination analysis with Middle Filter configuration (right) used in IPF calculations.

[1] CDM Smith, A. Horowitz, T. Creasey, R. Pendyalam, and M. Chen. NCHRP Report 765: *Highway Traffic Data for Urbanized Area Project Planning and Design*. TRB, National Research Council, Washington, D.C., 2014. Pg 161

Validation

Zone Activity Volume for Roads

Data Sources and Methods

In order to validate monthly Volume output, we created a zone set that contained 495 permanent counter locations across the continental U.S. These locations were not used to train the original model, but had sufficient MADT data reported across time so they could be used as a direct point of comparison. These locations, obtained from state DOTs, included counters dispersed across 15 U.S. states, including urban, suburban, and rural locations, as well as a variety of road sizes and classifications. Figure 4 shows those zone locations.

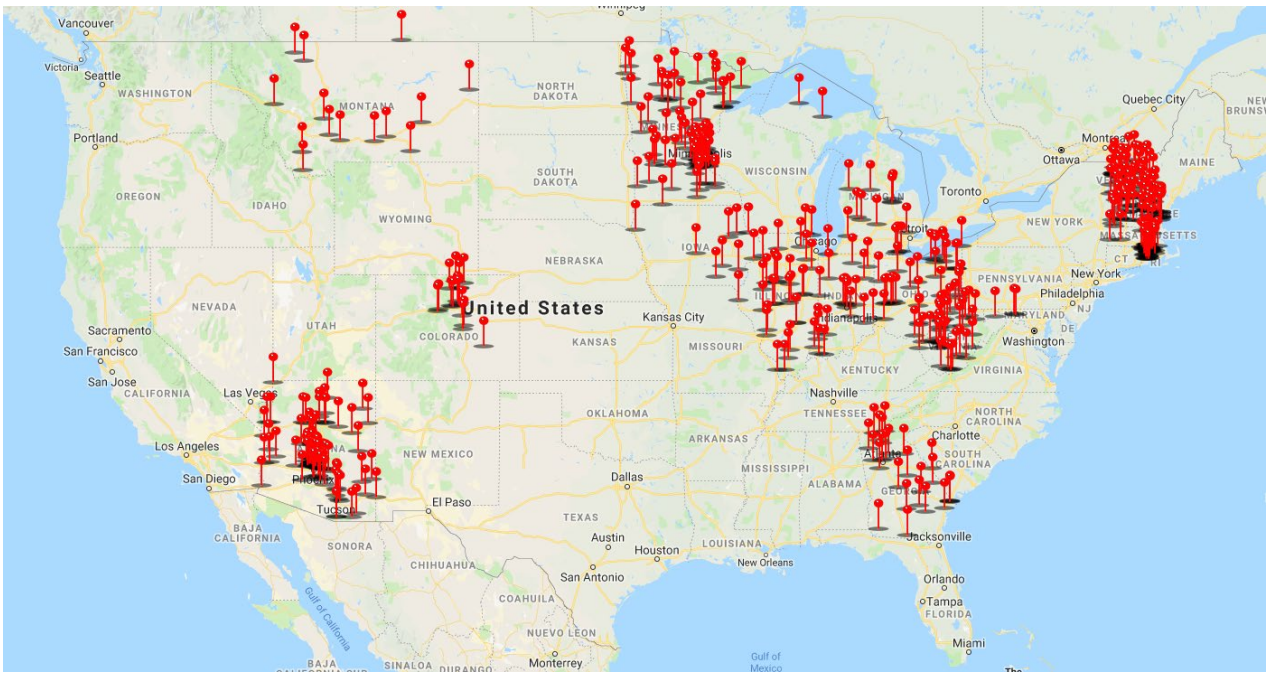

Figure 4: Counter locations across the U.S. used for MADT validation.

Validation was performed using these 495 counter zones in a series of Zone Activity Volume analyses within StreetLight InSight® for each calendar month in 2018. StreetLight Volume results were directly compared to the MADT values for accuracy. In total there were 5074 data points for comparison (each counter included data for a subset of months with 2018, but not necessarily all months within the calendar year).

Validation Results: Zone Activity Volume

Directly comparing the StreetLight Volume results to the reported MADT, there is a very high correlation. With no outlier deletion, the R^2 value is 0.979, indicating a strong relationship between StreetLight Volume estimates and real-world counts.

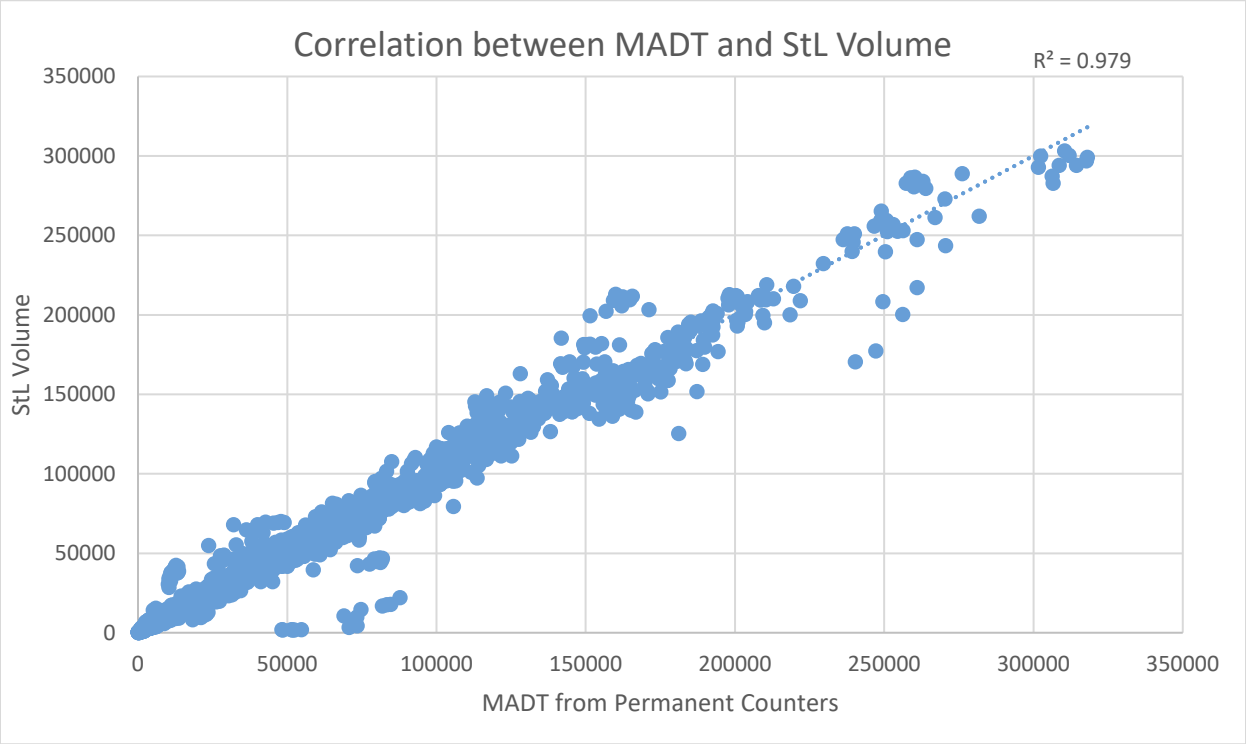

Figure 5: StreetLight Volume compared to published MADT values.

In addition to correlation, we also evaluated the mean absolute percentage error (MAPE) and root means square error as percent of average MADT (RMSE as %) by road size, expecting to have more accurate estimations on larger roads with higher MADT values. Table 1 compares the MAPE and RMSE to published target errors. The results fall within the target error range across all road sizes.

Road Size	Count	Target MAPE	MAPE	Target RMSE/Average MADT	RMSE/Average MADT
<2.5K	594	Not available	31%	47%	37%
2.5K-5K	586	Not available	12%	36%	17%
5K-10K	1011	20%	15%	29%	20%
10K-25K	1336	20%	13%	25%	25%
25K-50K	647	16%	10%	22%	17%
50K+	900	12%	8%	21%	13%

[2] See Table 2 in: Gadda, S., A. Mangoon, and K. Kockelman. Estimates of AADT: Quantifying the Uncertainty. 11th World Conference on Transport Research, Berkeley CA, 6-24-2007 to 6-28-2007.

VALIDATION RESULTS: SEASONALITY

In addition to evaluating the direct comparison between StreetLight Volume output and MADT across all locations, the analysis also examined some specific locations to validate the model's ability to accurately capture seasonal trends. Counter locations were randomly selected that had 11 or 12 monthly counts in 2018. In comparing results, trend lines reflected a similar seasonal pattern, while also being closely aligned in volume.

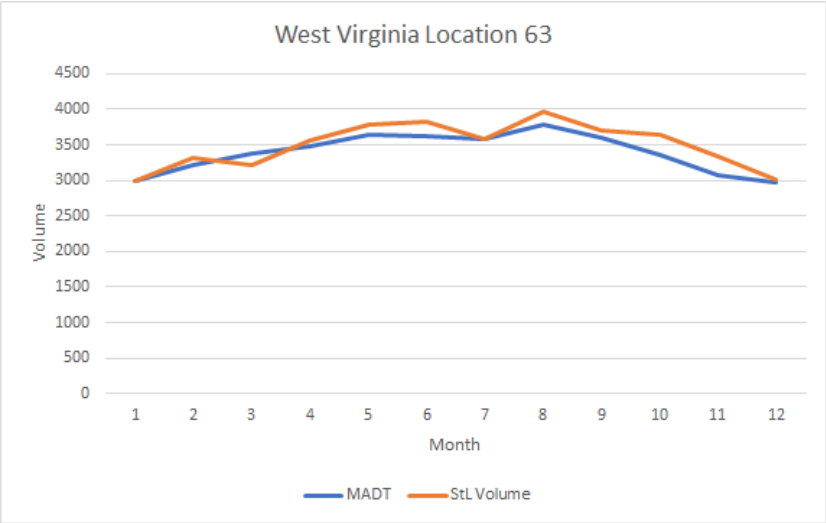

Figure 6: Monthly variation in StreetLight Volume and MADT across 2018 – sample mid-volume West Virginia location.

Testing both high- and low-volume roads confirmed the ability to report seasonal trends across all types of locations. Figure 7 below shows a higher volume road (~20K MADT). In this case, the StreetLight Volume estimate aligns very closely with the MADT values.

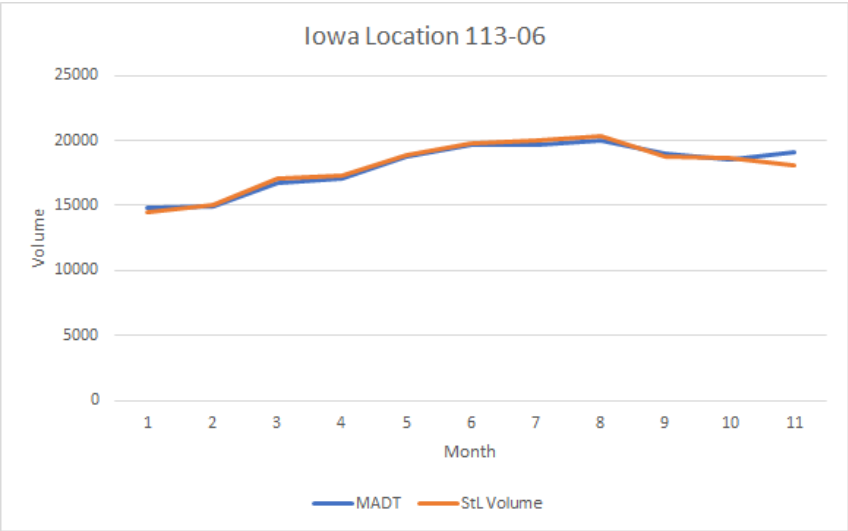

Figure 7: Monthly variation in StreetLight Volume and MADT across 11 months in 2018 – sample high-volume Iowa location.

Figure 8 depicts a very low-volume rural road in Montana with an MADT range between 200 and 1000 across the year. In this case, while slightly less extreme than the reported MADT numbers, the StreetLight Volume is still able to capture the seasonal peaks very accurately, with lows in the winter months and clear peak in July. These results give confidence in the model's ability to accurately predict seasonal trends, even when locations experience low-traffic volumes.

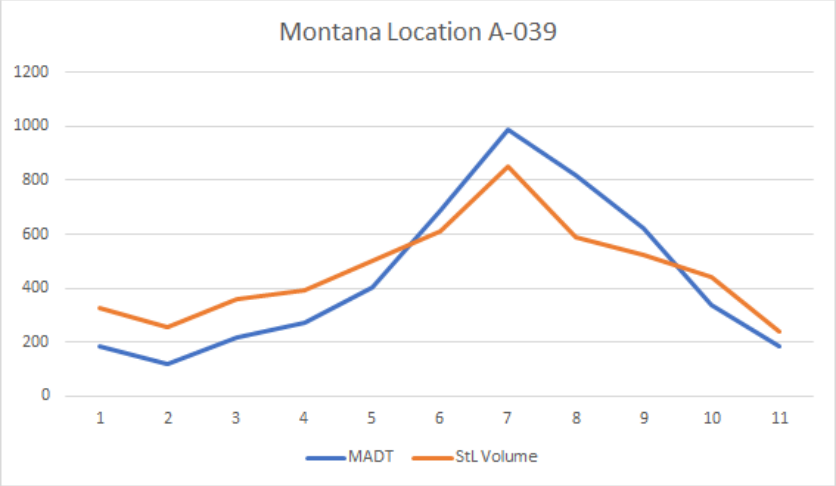

Figure 8: Monthly variation in StreetLight Volume and MADT across 11 months in 2018 – sample low-volume rural Montana location.

Origin-Destination Volume for Roads

DATA SOURCES AND METHODS

For validating Volume performance in an O-D analysis, StreetLight Volume results were compared to turning-movement counts published by Hennepin County in Minnesota[3]. A turning movement is an O-D study where each inbound road is the origin and each outbound road is the destination. Turning movements were chosen because validation data for turning movement studies is far more readily available than other types of O-D data.

The validation used data from five locations throughout the county, all of which were gathered on different dates in 2017. For each location, trips were manually counted between 6:00 a.m. and 6:00 p.m.

In order to perform a direct comparison between the Hennepin County locations and the StreetLight Volume output, we created zones in the *StreetLight InSight* platform that mirrored these five intersections. Then the platform ran an O-D analysis for the calendar year, structuring the query to match the specific weekday and hourly period from which the data were collected.

3 http://hennepin.maps.arcgis.com/apps/webappviewer/index.html?id=14c650982d904132a4854f399c71e1f2

For example, if site A used a Tue-Thu 8:00 a.m. to 10:00 a.m. definition of peak, the validation also used this definition of peak. The analysis closely mirrored the original study for direct comparison of turning movement counts and ratios.

Validation Results: Origin-Destination Volume

At each of the five locations, data was evaluated for eastbound, westbound, northbound, and southbound traffic as the origin, with left, right, and thru traffic as the destination. In total, this created 60 data points for comparison.

Without deleting any outliers, there was a high correlation between StreetLight Volume and the Hennepin turning movement counts, with an R^2 value of 0.947.

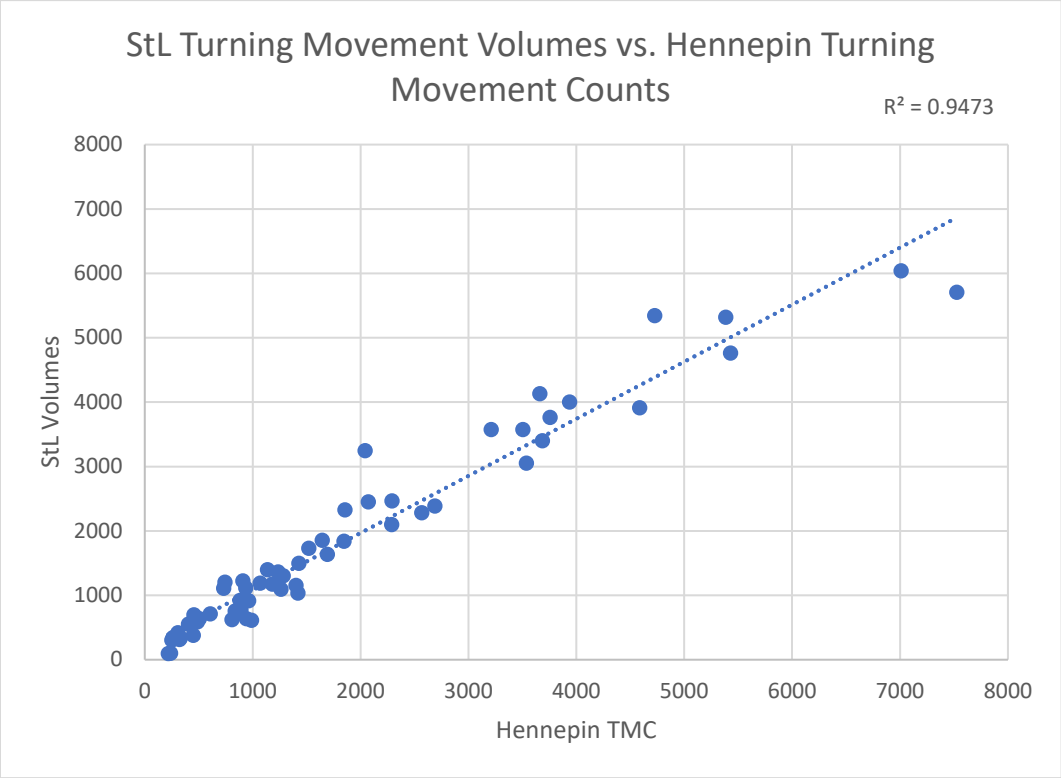

Figure 9: Correlation between Hennepin turning movement counts and StreetLight Volume.

In addition to the turning movement counts, the analysis also directly compared the turning movement ratios, represented as percentages of total origin zone traffic that traveled left, right, or directly through the intersection. The correlation for turning movement ratios was even higher, with an R^2 value of 0.976.

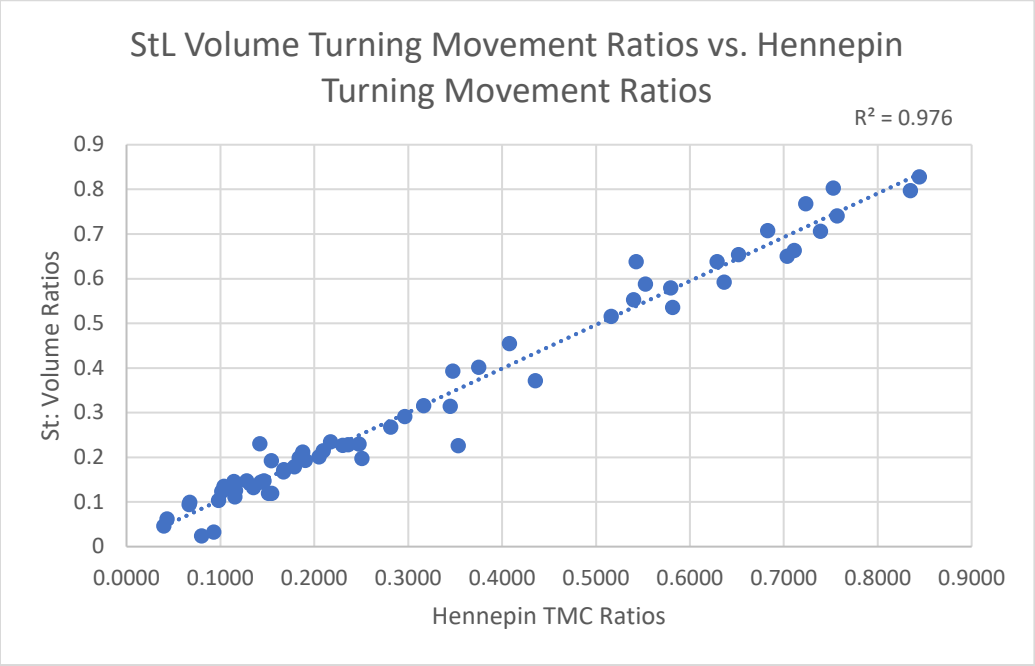

Figure 10: Correlation between Hennepin turning movement counts and StreetLight Volume.

The image below illustrates an individual intersection and the comparison between StreetLight Volume and turning movement counts along with turning movement ratios. Turning movement counts are very close, while turning movement ratios are nearly identical.

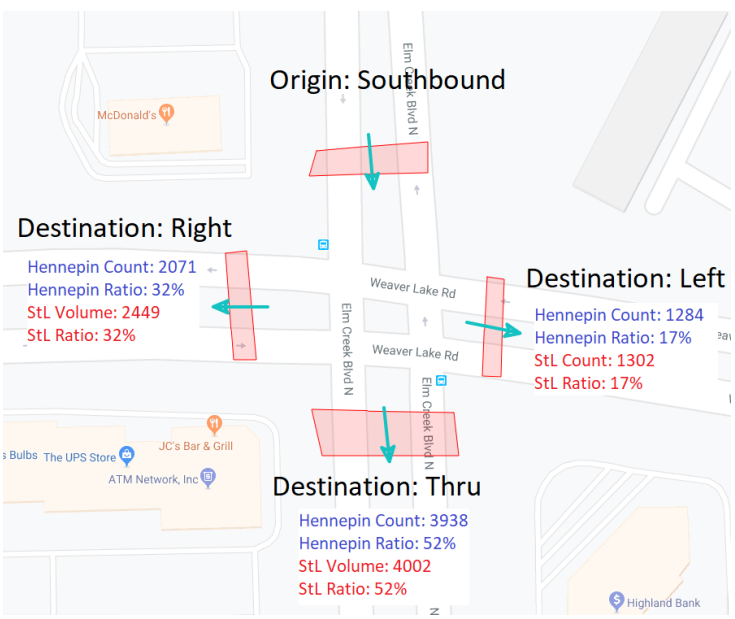

Figure 11: Southbound turning movement counts and ratios at location 4538.

Overall, these results are very promising and suggest that StreetLight Volume reliably captures seasonal trends, as well as O-D patterns.

In future iterations of the validation study, StreetLight will incorporate Zone Activity and O-D results for area zones, looking at validating trip counts that start or stop in the area (not pass-through). StreetLight welcomes any partner who has empirically measured counts for area zones that would like to share them for the purposes of validation.

About StreetLight Data

StreetLight Data 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, StreetLight measures diverse 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 3,000 global projects every month.

STREETLIGHT DATA