Title

New Data and Methods for Estimating Regional Truck Movements

Authors

Dion, Francois, PhD
Yang, Mingyuan
Patire, Anthony, PhD

Publication Date

2023-09-01

DOI

10.7922/G2JM27ZP

RESEARCH REPORT

Institute of Transportation Studies | UNIVERSITY OF CALIFORNIA

New Data and Methods for Estimating Regional Truck Movements

Francois Dion, Ph.D., P.E., Senior Research and Development Engineer, California Partners for Advanced Transportation Technology, University of California, Berkeley

Mingyuan Yang, Ph.D. Candidate in Transportation Engineering, Department of Civil and Environmental Engineering, University of California, Berkeley

Anthony Patire, Ph.D., Senior Research and Development Engineer, California Partners for Advanced Transportation Technology, University of California, Berkeley

September 2023

Berkeley Institute of Transportation Studies

Report No.: UC-ITS-2022-25 | DOI: 10.7922/G2JM27ZP

Technical Report Documentation Page

1. Report No. UC-ITS-2022-25	**2. Government Accession No.** N/A	**3. Recipient's Catalog No.** N/A
4. Title and Subtitle New Data and Methods for Estimating Regional Truck Movements		**5. Report Date** September 2023
		6. Performing Organization Code ITS Berkeley
7. Author(s) Francois Dion, Ph.D., P.E. (Michigan), https://orcid.org/0000-0002-4555-8791; Mingyuan Yang; Anthony Patire, Ph.D., https://orcid.org/0000-0002-3109-4164		**8. Performing Organization Report No.** N/A
9. Performing Organization Name and Address Institute of Transportation Studies, Berkeley 109 McLaughlin Hall, MC1720 Berkeley, CA 94720-1720		**10. Work Unit No.** N/A
		11. Contract or Grant No. UC-ITS-2022-25
12. Sponsoring Agency Name and Address The University of California Institute of Transportation Studies www.ucits.org		**13. Type of Report and Period Covered** Final Report (August 2021 – August 2022)
		14. Sponsoring Agency Code UC ITS
15. Supplementary Notes DOI:10.7922/G2JM27ZP		

16. Abstract

This report describes how current methods of estimating truck traffic volumes from existing fixed roadway sensors could be improved by using tracking data collected from commercial truck fleets and other connected technology sources (e.g., onboard GPS-enabled navigation systems and smartphones supplied by third-party vendors). Using Caltrans District 1 in Northern California as an example, the study first reviews existing fixed-location data collection capabilities and highlights gaps in the ability to monitor truck movements. It then reviews emerging data sources and analyzes the analytical capabilities of StreetLight 2021, a commercial software package. The study then looks at the Sample Trip Count and uncalibrated Index values obtained from three weigh-in-motion (WIM) and twelve Traffic Census stations operated by Caltrans in District 1. The study suggests improvements to StreetLight's "single-factor" calibration process which limits its ability to convert raw truck count data into accurate traffic volume estimates across an area, and suggests how improved truck-related calibration data can be extracted from the truck classification counts obtained from Caltrans' WIM and Traffic Census stations. The report compares uncalibrated StreetLight Index values to observed truck counts to assess data quality and evaluates the impacts of considering alternate calibration data sets and analysis periods. Two test cases are presented to highlight issues with the single-factor calibration process. The report concludes that probe data analytical platforms such as StreetLight can be used to obtain rough estimates of truck volumes on roadway segments or to analyze routing patterns. The results further indicate that the accuracy of volume estimates depends heavily on the availability of sufficiently large samples of tracking data and stable and representative month-by-month calibration data across multiple reference locations.

17. Key Words Trucks, traffic volume, traffic counts, fleet management, weigh in motion, data collection, data quality		**18. Distribution Statement** No restrictions.	
19. Security Classification (of this report) Unclassified	**20. Security Classification (of this page)** Unclassified	**21. No. of Pages** 109	**22. Price** N/A

Form Dot F 1700.7 (8-72) Reproduction of completed page authorized

About the UC Institute of Transportation Studies

The University of California Institute of Transportation Studies (UC ITS) is a network of faculty, research, and administrative staff, and students dedicated to advancing the state of the art in transportation engineering, planning, and policy for the people of California. Established by the Legislature in 1947, ITS has branches at UC Berkeley, UC Davis, UC Irvine, and UCLA.

Acknowledgments

This study was made possible with funding received by the University of California Institute of Transportation Studies from the State of California through the Road Repair and Accountability Act of 2017 (Senate Bill 1). The authors would like to thank the State of California for its support of university-based research, and especially for the funding received for this project. The authors would also like to thank Dr. Hashim Rizvi, from the Office of System Management, Traffic Operations Division at Caltrans, who assisted the project team in accessing the Caltrans District 1 StreetLight multimode subscription, obtaining WIM data, and provided valuable insights into data use by Caltrans. The authors would also like to thank Lisa Hockaday, from Caltrans District 1, for providing the project team with Traffic Census counts.

Disclaimer

Institute of Transportation Studies | UNIVERSITY OF CALIFORNIA

New Data and Methods for Estimating Regional Truck Movements

Francois Dion, Ph.D., P.E., Senior Research and Development Engineer, California Partners for Advanced Transportation Technology, University of California, Berkeley

Mingyuan Yang, Ph.D. Candidate in Transportation Engineering, Department of Civil and Environmental Engineering, University of California, Berkeley

Anthony Patire, Ph.D., Senior Research and Development Engineer, California Partners for Advanced Transportation Technology, University of California, Berkeley

September 2023

Berkeley Institute of Transportation Studies

Report No.: UC-ITS-2022-25 | DOI: 10.7922/G2JM27

Table of Contents

Table of Contents

List of Tables

List of Figures

Executive Summary

Executive Summary

Determining where trucks are traveling is crucial for the planning and maintenance of transportation networks. In California, information regarding truck movements is primarily derived from a network of fixed traffic monitoring stations. While data collected from these stations can be used to classify passing trucks, determine their travel direction, and assess their proportion within the general traffic, they provide limited information about trip origins and destinations and the routes taken between stations. Estimating truck movements within a region thus largely depends on extrapolating data between known collection points. Although this can be done with relative ease in simple networks offering few alternate routes, it can be a difficult task in complex networks without a multitude of supporting observation points.

A potential solution to the above problem is using vehicle tracking data, also known as probe data, supplied by third-party vendors to fill in gaps in truck monitoring. This data is collected from individual onboard vehicle monitors or GPS-enabled navigation devices in the vehicle. It is typically used by fleet operators to manage their business, but it can also be used to provide accurate information about truck movements not available from roadside monitors. Several vendors compile and disseminate this information, with StreetLight being one such available data set.

StreetLight provides Sample Trip Counts based on the information it collects to produce an Index value representing a relative traffic intensity. This Index can then be calibrated based on empirical traffic counts from a set of "calibration points" to scale its value up or down to create more representative weekly, monthly and yearly traffic volume estimates. To test the accuracy and reliability of this process, this study evaluates the ability of StreetLight to estimate truck counts using actual truck count data collected between December 2019 and January 2022 from three weigh-in-motion (WIM) and 12 Traffic Census monitoring stations operated by Caltrans District 1 in Northern California. Based on the analysis of the results, ways to improve the procedure are then suggested.

To test the accuracy of both all-vehicles and truck-only volume estimates produced by StreetLight in relatively simple settings the authors developed two showcases. One showcase predicted traffic volume along an isolated section of a divided highway and the other at an intersection in a residential area. In each case, calibration data was collected from either two one-way roads or one two-way road and used together with data contained in StreetLight's underlying trip databases to predict volumes at a separate nearby location.

For the first showcase, since the calibration zones and the prediction zone were nearby the assumption was that traffic volume would be similar. The overall two-way traffic volume estimates for the test road closely matched expectations. In addition, the estimated truck proportions also closely matched the actual ratios of medium-duty and heavy-duty trucks. However, separate estimates made in each travel direction failed to correctly apportion traffic volumes. While a 40/60 split was expected, a 50/50 split was obtained. This can be attributed to the way the calibration process averages calibration values obtained from each of the two one-way calibration zones. Some improvement was obtained by running a fourth test to estimate traffic in each travel direction using only eastbound calibration data to predict eastbound volumes and using only westbound data to predict westbound volumes. This approach resulted in a much closer 38/62 split.

These results indicate that appropriate considerations must be given to selecting suitable sets of calibration zones. Improved predictions are likely to be obtained if relevant sets of calibration zones are used. This means selecting calibration zones that closely match the characteristics of the prediction zone (such as direction of travel) to ensure that data from dissimilar calibration zones does not unduly influence calculating the calibration factor.

For the second showcase, three approaches to an intersection were chosen to serve as calibration zones to predict traffic volumes on the fourth approach. Unlike the first showcase, however, there was no reason to assume that the traffic volumes in the first three approaches would match the flows in the fourth as traffic from each approach can either turn left, go straight, or turn right. Here again, results when two-way traffic was used to calculate the calibration factor to predict two-way travel on the test approach differed from when data was combined for traffic traveling on individual lanes in opposite directions due to the program's averaging approach.

Including data from more locations reduces the potential that data from one zone could play an oversized impact in the averaging process. StreetLight, for instance, recommends using between 10 and 20 calibration points. Selecting calibration zones sharing relatively similar index-to-volume ratios and having similar traffic characteristics to the prediction zone further ensures that the resulting calibration factor will not overly skew the estimates.

One problem encountered that could affect the usefulness of this technique relates to the fact that sudden reductions in sample trip counts can lead to inaccurate estimates. Such drop-offs could result from a data provider suddenly stopping to provide information or due to technical glitches. For District 1, this occurred in February and May 2021 following two successive drop-offs in fleet tracking data from suppliers to StreetLight. While the two drop-offs affected the program's Index values, the latter one had a more pronounced effect as it resulted in very low trip counts and sample sizes that likely did not accurately reflect average traffic characteristics at each location.

Another issue is that StreetLight data only distinguishes between medium-duty trucks, defined as commercial vehicles weighing between 14,000 and 26,000 lbs., and heavy-duty trucks, those above 26,000 lbs. Light-duty trucks, under 14,000 lbs. are not considered by the program due to a lack of tracking data. This is likely due to difficulty in distinguishing commercial light-duty trucks, which include pickup trucks, sport utility vehicles (SUVs), vans, and minivans, and delivery vehicles, from the general traffic. Since this categorization does not match the vehicle classification based on the number of axles, truck configuration (tandem or single), and number of trailers that is normally used by Caltrans, it is necessary to convert Caltrans classification data into a suitable StreetLight format, which can potentially introduce errors.

Finally, the authors recommend various means of improving the program's calibration capabilities. StreetLight users can provide local average daily traffic counts and truck percentages to produce estimates more closely representing observed traffic volumes. This can be done by first calculating a local adjustment factor for each calibration location and then averaging the factors to produce a single value that can be applied areawide to convert all Index values into Calibrated Index values. StreetLight derives a uniform adjustment factor meant to be applied to all locations by simply taking the average of all the local adjustment factors associated with each calibration location. This treats data equally from all calibration zones. The process does not consider that data from certain calibration zones might be more reliable or more representative of overall truck movements than others, or that geographical factors may influence the relationship between calibration and prediction zones.

One suggestion to address this issue is to introduce weights into the single-factor determination process to reflect the fact that not all calibration data are necessarily equal. Weights could be used to allow the single-factor determination process to consider the distance between a prediction zone and a set of calibration zones, on the assumption that data from calibration zones closest to the prediction zone should have a greater impact on the accuracy of the prediction than data from calibration zones located further away.

Overall, the study's authors conclude that StreetLight can be used to obtain rough estimates of truck volumes if adequate and appropriate calibration data are provided. StreetLight produced Calibrated Index values corresponding to between 80 and 120 percent of observed volumes. However, they also found that the accuracy of the estimates deteriorated to no better than between 50 and 400 percent of observed counts when questionable calibration data were used or when issues affected StreetLight's underlying tracking data.

Based on the evaluation, the researchers found that:

- Using calibration data from stable data sources is important. For each calibration location, the data quality was assessed by calculating the ratio between the uncalibrated StreetLight Index value and the calculated medium-duty and heavy-duty truck counts. Stations with stable ratios were deemed more desirable, as this is a sign of more consistent underlying data. High ratio variability, as well as sudden ratio changes, were used as criteria for excluding specific locations or specific periods from consideration.
- A robust method is required to convert the 15-class, axle-based counts reported from the District's WIM and Census stations into the simple medium-duty/heavy-duty truck categorization used by StreetLight. This can reasonably be done with WIM data, as measured weights are available. However, a method needs to be devised to apportion vehicles from other data sources.
- Checks should be made to ensure that sufficient samples of observed trips support the analyses being considered. Issues may arise when the underlying samples are too low, as low counts may not adequately capture the range of trips that may be observed at a given location.
- One issue associated with the single-factor calibration process is that it treats each reference location equally. This process ignores the fact that locations with higher sample trip counts may provide more reliable or representative information than those with limited data. It also ignores potential geographic connections between calibration and prediction zones.

To help increase the probability of obtaining reasonable analytic results, the authors recommend a procedure for setting up the analyses which includes steps for preparing calibration data, assessing calibration data quality, and determining adequate calibration data sets.

Contents

Introduction

The ability to determine where trucks are traveling plays a crucial role in the maintenance of transportation networks. Knowing where trucks are coming from and where they are going helps identify routes supporting freight activities, prioritize needed roadway improvements related to trucking, and allocate appropriate funds for road maintenance and development.

In California, information on truck movements is primarily derived from data collected from fixed traffic monitoring stations installed at key locations on the State Highway network. Examples include weigh-in-motion (WIM) stations operated by the California Department of Transportation (Caltrans) and fixed and rotating count locations supporting Caltrans' Traffic Census program. Data collected from these locations typically include information about each observed truck, such as its length, number of axles, and weight. In many cases, the travel direction is also provided or can be inferred based on the location of the sensor. The proportion of trucks within the traffic can also be determined so long as all vehicles are counted.

While valuable in characterizing truck traffic at specific locations, these counts provide limited information about truck origins and destinations and the routes taken in between. Not enough information is collected to directly determine truck movements within an area and assess their impact on the local road network and nearby communities. In this context, estimating truck movements within a region largely depends on extrapolating data between known collection points. While this can be done with relative ease in simple networks offering few alternate routes, it can be very difficult to correctly estimate the routes taken in more complex networks without a multitude of supporting observation points. This typically results in less accurate transportation studies.

A potential solution to this problem could be found by using vehicle tracking data, also know as probe vehicle data. Whereas sensor data is generally produced by some type of sensor installed in or next to a roadway that collects data on all vehicles passing that particular point, probe data is collected by monitoring the position of individual vehicles (called "probes") by detecting an onboard electronic identification tag or from a GPS device in the vehicle.

Given the increasing use of vehicle tracking technologies by fleet operators to monitor the movements of their vehicles, either through onboard GPS devices or cellular phone applications, various vendors have developed data aggregators to collect, aggregate, and analyze these vehicle tracking data. While the initial purpose was to provide fleet operators with detailed operational reports, some vendors, such as StreetLight and INRIX, have started offering transportation agencies reports summarizing vehicle movements within particular areas. These emerging data sources have opened the possibility of combining vehicle information obtained from fixed truck monitoring locations with routing data derived from tracked vehicles to build a more comprehensive picture of truck activities within an area, including truck origins and destinations.

Some of the specific benefits that Caltrans could derive from a better understanding of truck movements within a region include:

- Better estimates of current truck vehicle miles traveled on state highways
- More reliable truck volume projections for assessing the impacts of infrastructure projects, designing new roads, planning pavement rehabilitations, or conducting load factor characterization for structures
- Better apportioning of roadway maintenance funds to sections heavily traveled by trucks
- Improved calculation of truck-related accident rates
- Improved consideration of truck-related effects on road safety
- Reduced reliance on ad-hoc vehicle classification studies to cover areas without permanent monitoring stations
- Ability to use automated data collection approaches to characterize truck demand along given roadway segments

Beyond Caltrans, enhanced truck movement information may also:

- Allow the California Highway Patrol (CHP) to better plan where truck enforcement actions may be needed
- Allow regional planning organizations to better account for truck movements when developing transportation plans
- Allow the California Air Resource Board to estimate truck-related emissions more accurately

Study Objectives

While the information collected by these probe data aggregators can supplement current truck-related data collection efforts by Caltrans and facilitate the development of more accurate representations of truck movements within a region, some uncertainties remain. For example, one issue is whether the derived vehicle movements and estimated counts would be a true reflection of the actual truck traffic within an area or a biased representation due to the specific fleets or types of vehicles being monitored.

This project assesses how information generated from third-party probe data aggregators could be used in conjunction with data from weigh-in-motion, Traffic Census, and other fixed count locations to improve descriptions of truck movements within a region. The study:

- Assesses gaps associated with current data collection efforts
- Evaluates alternate data collection opportunities offered by third-party aggregators relying on vehicle tracking
- Determines how truck volumes derived from third-party probe data compare to existing classification counts
- Develop recommendations for obtaining reliable estimates of truck movements using both traditional and third-party data

Study Area

The study focused on Caltrans District 1 in the northwest corner of California, which covers Lake, Mendocino, Humboldt, and Del Norte counties (see Figure 1). This district was selected as it was the only one with an existing subscription to a probe data analytical platform, namely, StreetLight, when the study began.

Figure 1. Study Location: Caltrans District 1

Report Outline

The report is organized as follows:

- Identification of gaps in truck data collection within Caltrans District 1 based on the identified existing data sources
- Overview of emerging probe-based data collection opportunities
- General description of the analytical capabilities offered by the StreetLight platform
- Evaluation of the accuracy and reliability of StreetLight truck outputs for analyses in Caltrans District 1
- Exploration of potential improvements to the StreetLight calibration process

The report concludes with recommendations on using StreetLight and if so, on how probe-based data may be used in conjunction with traditional data sources to obtain reasonable estimates of truck volumes on key regional roadways.

The following two technical appendices are also provided:

- Review of existing truck data sources within Caltrans District 1
- Review of Key Mobile Data Vendors

Assessment of Current Data Gaps

This section presents an assessment of gaps in the collection of truck data within Caltrans District 1. from traditional fixed-location sources.

Summary of Truck Data Availability

Figure 2 presents a summary of the directional truck classification data that was collected with the help of Caltrans staff for the period from January 2019 to January 2022. Vehicle classification and weight data were both available continuously only from the three WIM stations operated within the district.

In addition to the WIM data, continuous or near-continuous vehicle classification counts were also available from nine locations where Caltrans maintains automated traffic counters in support of its Traffic Census program. Six of these are located along US-101, while the others are on CA-299 near the district's eastern boundary and on CA-20 and CA-1 near where the two highways meet south of Fort Bragg. These locations provided classification counts, but no weight information.

Three additional Traffic Census locations appear to have started providing continuous classification data in mid/late 2020. These include a station on US-101 near Orick, another on US-101 north of Ukiah, and one on CA-20 north of Ukiah. Occasional data is further available from two additional stations, one on CA-197 north of Crescent City, and one on CA-20 near Willis.

In addition to the data collection points shown in Figure 2, 24-hour, bi-directional counts were available from the locations that were mapped earlier in Figure 44. These are locations for which observed or estimated AADT truck volumes are reported by Caltrans on an annual basis, with data from each location typically updated once every three years. For locations at the intersection of two or more roads, AADT statistics were typically provided for each approaching direction.

Figure 3 graphically summarizes the degree to which truck data is available on the primary road network of District 1. Segments highlighted in green are those for which weight and classification data are available. These typically correspond to segments where WIM stations are located. Segments highlighted in yellow are those for which directional truck classification counts from the Traffic Census program are available, either on a continuous or sporadic basis. Segments highlighted in brown are those for which only AADT truck data are available. Finally, segments highlighted in red are those for which no truck data is available.

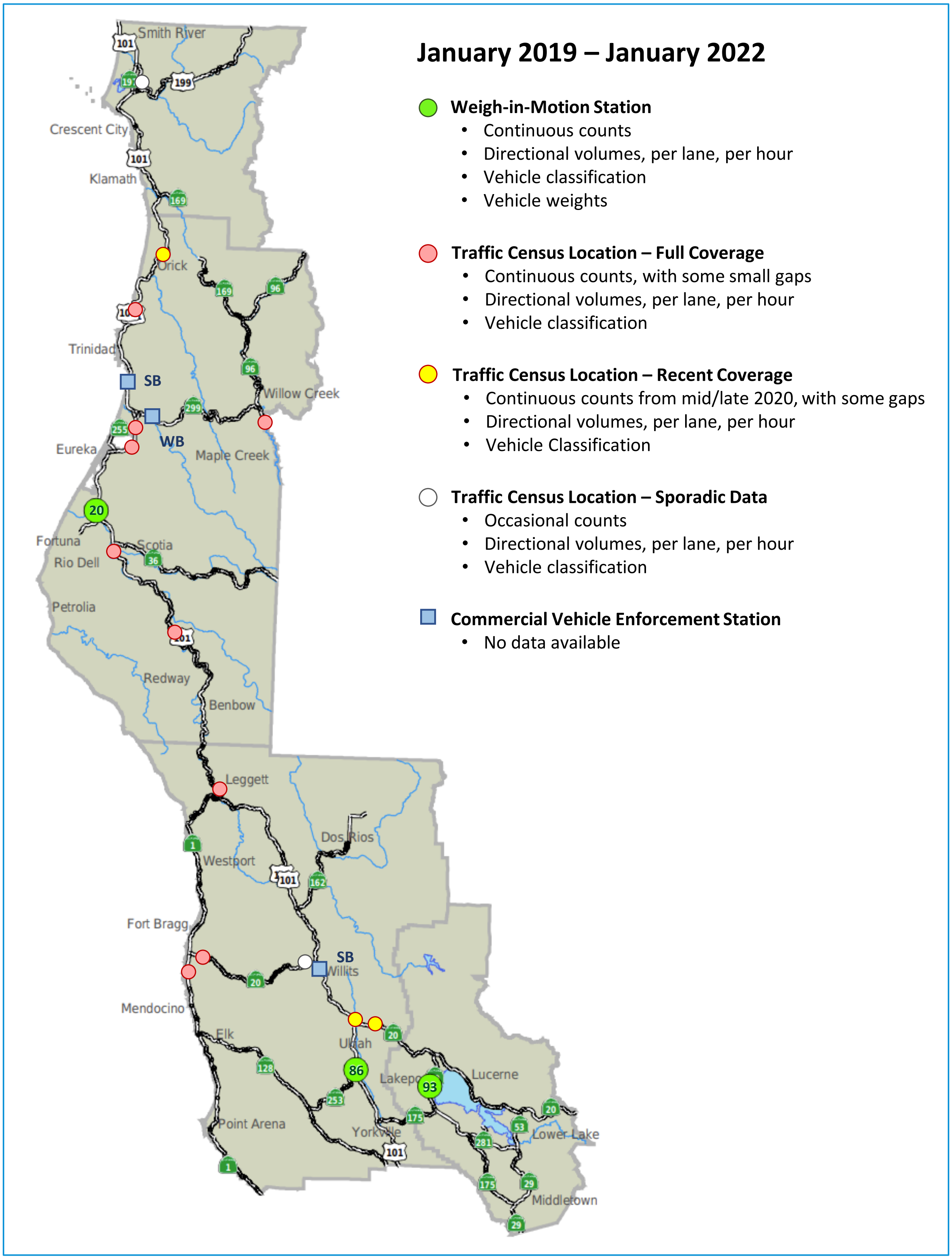


Figure 2. District 1 Directional Classification Truck Data Availability

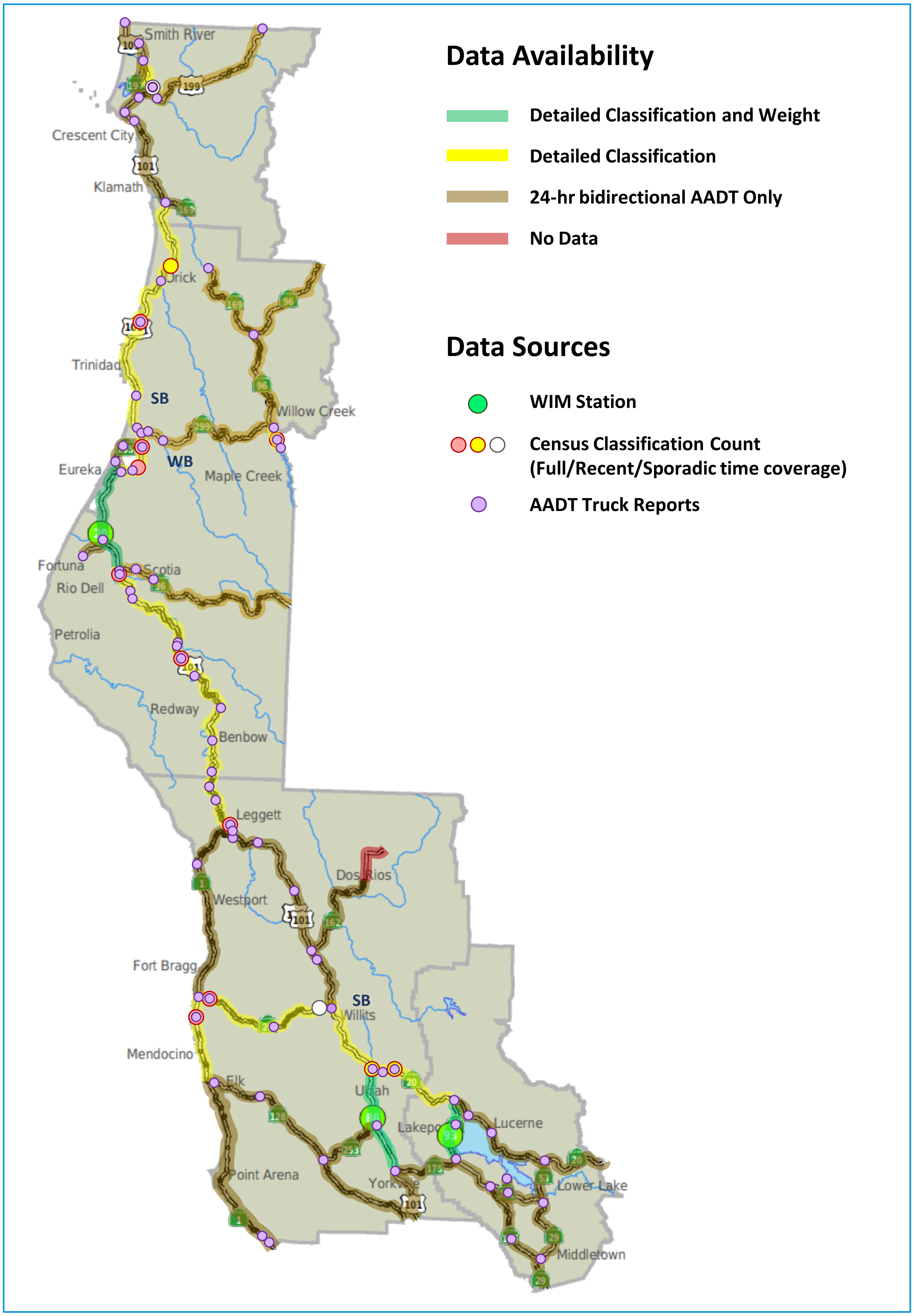


Figure 3. District 1 Truck Data Availability by Road Segment

Data Gaps

The following summary is an assessment of data availability and key data gaps associated with the current monitoring of truck movements within Caltrans District 1:

- Truck weight information is only available continuously from the three weigh-in-motion stations located within the district. While weight is also measured at truck scales operated by the CHP, these data are only retained for trucks with violations and are not typically shared with other agencies.
- Continuous truck classification data is available from 15 locations across the district. This includes the three WIM stations located in the district and 12 automated vehicle classification stations along US-101 between Ukiah and Orick, on CA-20 and CA-1 near Fort Bragg, CA-20 near Ukiah, and CA-299 at the district boundary.
- For most roadways outside of US-101 and sections of CA-20 between Ukiah and Fort Bragg, the only source of available traffic data is the AADT truck statistics produced by the Caltrans Traffic Census Program. Since these statistics typically only provide 24-hour, bi-directional counts, additional data collection efforts must be conducted to obtain truck volumes in a specific direction or for a specific period, or the distribution of truck types beyond a simple categorization based on the number of axles (2, 3, 4, or 5).
- Weight distributions are usually extrapolated for roadway segments for which no direct weight observations exist. This means that for most roadway segments within the district assessments of truck pavement impacts, and thus pavement needs assessments, are based primarily on estimates and may not always be realistic.
- None of the available data sources collecting information from fixed locations track vehicle movements, i.e., where a truck is coming from, or going to, or what route it is taking.
- Very little information is available from the areas around Clear Lake, Eureka, and Crescent City to help quantify truck movements.

Emerging Data Sources

This section presents a general review of existing traffic data collection capabilities, with a particular focus on those provided by mobile data sources. It describes the data collection methods associated with point detection methods, segment detection methods, and mobile data sources. The information presented is an update on information that was first compiled in 2019 for an early 2021 PATH report on hybrid data implementation (Khan et al. 2021).

Existing data collection methods can be categorized into three broad groups, illustrated in Figure 4:

- Point detection
- Segment detection
- Mobile data

Key data collection methods, technologies, advantages, and disadvantages associated with each category are described in more detail in the subsections that follow.

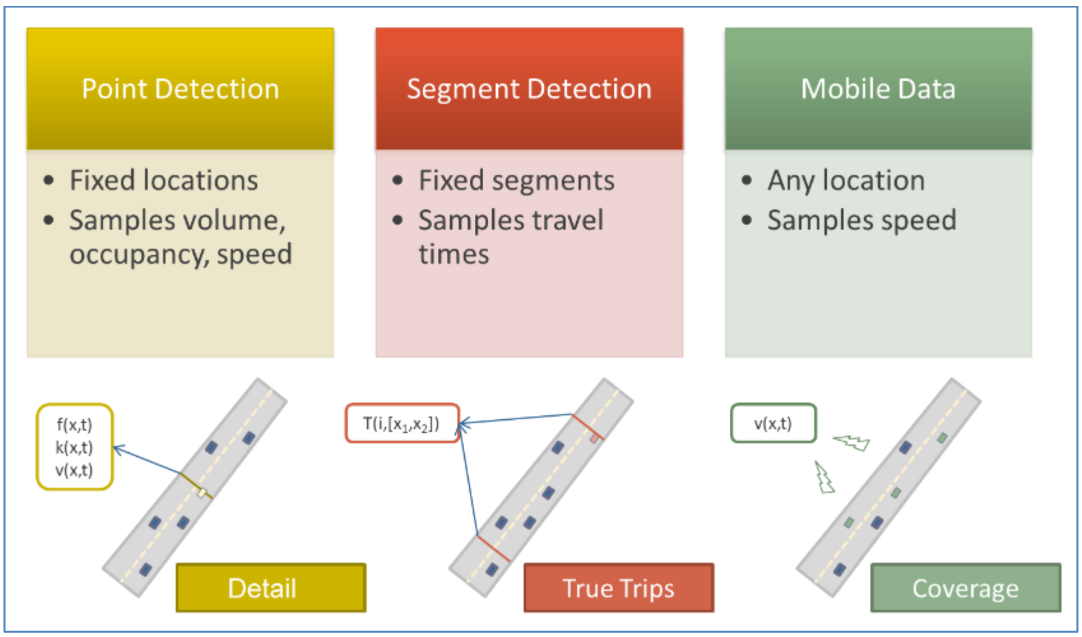


Figure 4. Comparison of traffic data collection methods (Bayen et al. 2013)

Point-based collection methods

Point-based data collection methods measure traffic flows and/or speeds at one dedicated location. Key examples include inductive loops, radar-based sensors, and weigh-in-motion devices. This type of data collection is currently the primary source of truck traffic information used by Caltrans in District 1. The main strength of these methods is that they capture the complete cross-section of all vehicles passing by a given location, and therefore provide reliable measures of flow and speed, within the capabilities of each technology. However, their main disadvantage is that they provide no direct information about what happens between those locations. For example, there is no way to detect a traffic incident between two point detectors until a disruption in traffic resulting from the incident propagates upstream or downstream to the detectors. Even then, the exact position of the incident between the detectors would remain unknown. Another disadvantage is that these detection methods require the installation and maintenance of dedicated infrastructure, which can be costly.

Segment-based collection methods

Segment-based collection methods provide trip times for preset road segments the data collection device can identify the same vehicle at two different locations and match the data collected at each observation point. Examples of data sources in this category include:

- Toll-tag readers
- License plate readers
- Magnetometers
- Bluetooth MAC address readers
- Readers of WIFI MAC addresses

Each of these methods can measure travel times between the two locations. The number of vehicles that are matched can vary with the specifics of each technology and its deployment. With Bluetooth and WIFI, this rate also depends on how many of the vehicles are equipped with these consumer devices. In practice, the sample size is generally large enough to provide useful median travel times. As with the point-based collection methods, segment-based methods require dedicated field infrastructure.

None of these data collection methods are currently used in Caltrans District 1.

Mobile data sources

Mobile data collection methods rely on the proliferation of GPS-enabled mobile devices and data networks to identify the position of individual vehicles over time. This offers two key advantages:

- No additional field equipment is necessary beyond the existing cellular network
- Data can be obtained from virtually any location where cellular coverage exists

Depending on the underlying technology used, mobile data sources can further be divided into the following four categories:

- **Smartphone applications.** These use location-based services running on GPS-enabled smartphones to periodically obtain information about the location of each device. Depending on the application, the rate at which the vehicle's location is updated may vary from once every few seconds to several minutes apart. This data collection method is one of the main data streams offered by INRIX.
- **In-vehicle GPS navigation devices.** GPS-enabled navigation devices are embedded in a vehicle's dashboard to obtain information about the vehicle's location, as well as predictive information about future locations when the dynamic navigation feature is activated. The services offered by the devices are very similar, if not identical, to smartphone applications. However, a major distinction with smartphone applications is that in-vehicle navigational devices are marketed, sold, and installed by vehicle manufacturers. Since they are attached to a specific vehicle, it may also be possible to associate information about the vehicle type with the data being collected if that information is available.
- **Fleet telematics.** These use vehicle tracking data generated by onboard GPS devices installed in commercial vehicles to manage their movements. These devices are often installed on commercial truck fleets, rental cars, taxis, transit buses, etc. Many of these fleets agree to allow traffic information aggregators to use their collected data to estimate current traffic conditions and to archive it for historical reference. A particular benefit of fleet telematics data is that it can often be linked to a specific type of vehicle which allows for aggregate data analysis. This is the primary source of truck data offered by INRIX and StreetLight.
- **Connected Vehicles.** Some newer vehicles have onboard telematics modules that collect data about the vehicle and its internal diagnostics. Selected data may then be periodically transmitted to the driver of the vehicle, the vehicle manufacturer, and other vehicles using various communication technologies. The transmitted data can include the vehicle's GPS position, its speed and heading, acceleration and braking data, its vehicle identification number (VIN), and vehicle make and model, along with information about its current operating conditions.

Analysis of Streetlight Capabilities

This section provides an overview of data collection and analysis capabilities offered by the StreetLight platform based on information retrieved in March 2022 from the online StreetLight Help Center and white papers published by StreetLight Data.

The analysis in this report focused on the capabilities offered by StreetLight Insight to the team had full access to this analytical platform for data covering Caltrans District 1 through a Multimode Regional subscription from Caltrans at the time of the study. Such access was not available for any other Caltrans districts.

Key elements reviewed in this section include:

- Data sources
- Data processing methodology
- Modes of travel
- Truck categorization
- Analysis types
- Output metrics
- Local calibration capabilities

Data Sources

StreetLight uses the following two types of data to track the movements of vehicles, as illustrated in Figure 5 (StreetLight 2020):

- **Location-based service data.** Trip data obtained from smartphone applications that use opt-in location-based services. This is comprised of data collected by Cuebiq through software provided to mobile application developers to facilitate location services. Data collected through these sources provide a location ping on average every 200 seconds.
- **Navigation GPS data.** Data from GPS-based navigation systems within vehicles. This is comprised of data supplied by INRIX from commercial fleet navigation systems, navigation devices in personal vehicles, and turn-by-turn navigation applications on smartphones. This is the only source of data for commercial vehicles. For commercial trucks, location data is collected every one to three minutes whenever the vehicle is in operation, even if the driver is not actively using navigation. For passenger cars, if the vehicle is within the INRIX partner system and has a navigation console, location data is connected every few seconds whenever the vehicle is running.

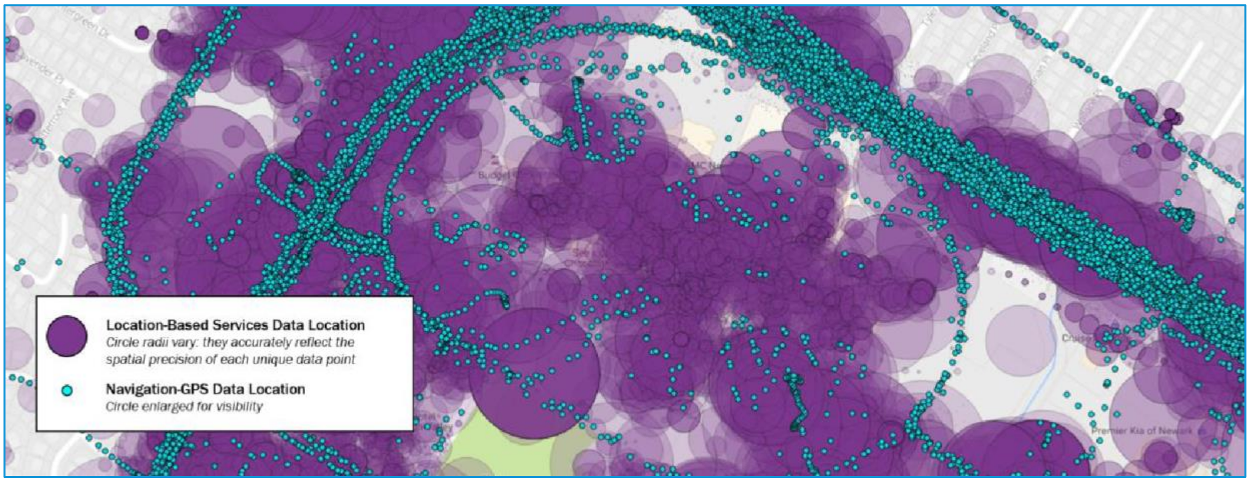


Figure 5. Location-Based and Navigation GPS Data (StreetLight, 2020)

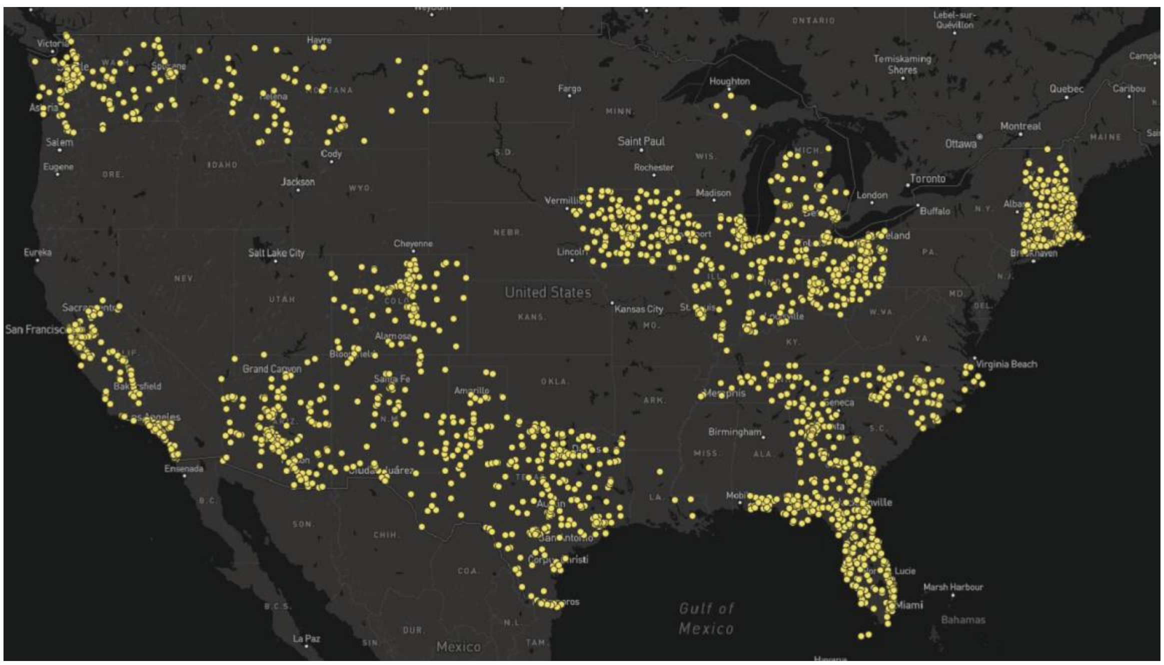


Figure 6. Permanent Traffic Counters used for Training Monthly ADT (StreetLight, 2021)

Additional underlying data used to assist with the generation of trip metrics include:

- Demographic statistics derived from U.S. Census data
- OpenStreetMap data reflecting road classification and density of commercial activity
- Weather data
- Reference traffic counts from over 10,000 permanent traffic count stations located in various environments across the United States, as shown in Figure 6

StreetLight documents indicate that location-based data is periodically obtained from over 110 million devices in the United States and Canada. Sample sizes for location-based devices will generally be lower than the number of devices tracked in a given area as all devices are not always in use all the time and not all trips made are captured perfectly. Depending on the location and type of analysis conducted, this translates into effective penetration rates that can range from one to 35 percent.

Penetration rates for commercial trucks are harder to pinpoint. Due to the nature of commercial freight activities, trip sample sizes for navigation-GPS data will be generally higher than the number of trucks tracked. This is because actively used commercial trucks typically take many trips per week, often on set routes. Trucks being tracked are also more likely to come from companies managing large fleets of vehicles, as these are more likely to have up-to-date fleet management tools. Trucks from smaller companies or that are more rarely used are less likely to be included in the data set. This results in a potential bias in the truck data being collected.

Data Processing Methodology

StreetLight collected tracking data is typically processed according to the following seven steps that are incorporated into the analytical platform:

- **Step 1 - Data Extraction, Transformation, and Loading.** Retrieval of anonymous data from suppliers, on a daily, weekly, monthly, or custom basis. This step also includes processes to eliminate corrupted data.
- **Step 2 - Data Cleaning and Quality Assurance.** Checks to verify that the volume of data has not changed unexpectedly, is properly geolocated, and shares similar patterns to batches previously received from each supplier.
- **Step 3 – Creation of Trips and Activities.** Groups collected data into key patterns, such as data points collected at a particular time or representing movement at a particular speed, that can be used to associate the data with a given trip.
- **Step 4 - Data Contextualization.** Integration of additional data from other sources to add richness and improve accuracy, such as speed limits, road directionality, land use data, parcel data, etc. This is where sets of data points associated with a given trip are locked to the road network and used to determine whether a trip originated from a home or office, or made by bus, walking, or biking.
- **Step 5 - Additional Quality Assurance.** Executes additional tests to flag patterns that appear suspicious or unusual
- **Step 6 - Data Normalization.** Adjusts data along several parameters to create the StreetLight Index. This is performed monthly to account for changes in sample sizes or observed traffic volumes at reference locations.
- **Step 7 – Data Storage.** Storage of processed data in a database to support queries

Modes of Travel

Depending on the analysis type, and subject to local data availability, trip analyses can be performed for the following modes of travel:

- **All Vehicles.** Trips based on data obtained from smartphone applications using opt-in location-based services. This includes all trips captured by mobile devices carried by individuals in passenger cars, trucks, buses, or while walking or riding bikes or scooters if the captured movements are behaving like a motorized vehicle. This is the default mode for most analysis types.
- **Truck.** Trips from commercial vehicles obtained from GPS navigation sources.
- **Bicycle.** Bicycle trips
- **Pedestrian.** Pedestrian trips
- **Bus.** Trips on transit buses
- **Rail.** Trips on commuter rail systems

Trips by bus, rail, bicycle, or walking are identified using a machine learning algorithm that attempts to find specific characteristics in the collected location-based data points. This process may further be helped by using data from mode-tagged location-based services, such as applications designed to track daily walking or biking trips, as well as validated bicycle and pedestrian counts. For bus and rail trips, trip identification is only possible where transit routes are defined in Open Street Map, as this permits a determination of whether trips go through known bus stops or transit stations.

The machine learning process does not identify a specific mode associated with a given trip but can estimate the probability that a trip may be made using each mode considered. The mode assigned to a given trip in the analyses is usually the one with the higher probability.

Truck Categorization

StreetLight uses the following binary categorization for trucks:

- **Medium-duty trucks.** Commercial vehicles weighing between 14,000 and 26,000 lbs.
- **Heavy-duty trucks.** Commercial vehicles weighing over 26,000 lbs.

Due to a lack of tracking data, StreetLight does not currently provide metrics on light-duty commercial vehicles.

The above classification is not determined by StreetLight, but rather by the data providers that obtain GPS navigation data from commercial fleet management systems. At the time of the study, the primary provider of such data to StreetLight was INRIX.

Analysis Types

The StreetLight platform currently offers the following types of analysis:

- **Zone Activity.** Analysis of traffic starting in, stopping in, or passing through a group of locations
- **Origin-Destination.** Analysis of traffic traveling from one group of locations to another
- **Origin-Destination through Middle Filters.** Analysis looking at trips going from one group of locations to another, but through specific in-between locations
- **Top Routes for Zones.** Analysis to see how traffic flows to and from an area or a road segment
- **Top Routes between Origins and Destinations.** Most popular routes for trips between the selected origins and destinations
- **Trip to/from Pre-set Geography.** Trips from one group of locations to another group, but through a specific filter of locations
- **Segment Analysis.** Trip information for a specific road segment, from one pass-through gate to another.
- **AADT.** Measurement of Average Annual Daily Traffic for any road segment in the contiguous United States.
- **Turning Movement Counts.** Analysis of vehicle movements at an intersection to determine whether the traffic is turning left, right, or continuing straight.

Not all analysis types are available for every mode of travel. Table 1 provides a summary of what is possible with each travel mode.

Table 1. Summary of StreetLight Analysis Capabilities

Analysis Type	Travel Mode					
	All Vehicles	Trucks	Bus	Rail	Bicycles	Pedestrians
Zone Activity	X	X	X	X	X	X
Origin-Destination	X	X	X	X	X	X
Origin-Destination with Middle Filter	X	X	X	X	X	
Trips to/from Pre-Set Geography	X	X			X	X
Top Routes Between Origins and Destinations	X	X				
Top Routes for Zones	X	X				
Segment Analysis	X	X				
AADT	X					
Turning Movement Counts	X					

Output Metrics

Depending on the analyses considered, key metrics that can be produced by StreetLight include one or more of the following:

- **StreetLight Sample Trip Count.** Value representing the total number of observed trips in an underlying StreetLight data sample. This is a total across all days within the analysis period. These values are not adjusted for seasonal variations, penetration rates, or other factors.
- **StreetLight Index.** Normalized value derived from the sample trip counts. This metric does not represent an actual volume, but a value derived from Sample Trip Counts accounting for variation in sample sizes across space and time. It is produced to allow the comparison of metrics across zones and analyses. This is the default output metric for analyses focusing on trucks, pedestrians, and bicycles.
- **Single-Factor Calibrated Index.** Estimated trip count based on a comparison to a reference set of values. This reference set can either be StreetLight AADT estimates, or actual count data provided by the user. This index is obtained through a process that creates a single normalization factor from the provided data that is then applied to the analysis results to scale the output metrics up or down.
- **StreetLight Volume.** Estimated number of vehicle trips within or between zones. This is comparable to real-world count data and is based on an algorithm trained with real-world data and seasonal factors. This is the default output metric for analyses based on location data from smartphones. It is not yet available for truck-related analyses.

Local Calibration Capabilities

As indicated earlier, the StreetLight Index is the default output metric for analysis involving trucks. This index does not represent an estimated volume, but a normalized value expressing trip intensity derived from sample trip counts and aimed at facilitating comparisons across space and time. It is however possible to scale the resulting indices towards more representative volume estimates using the Single Factor Calibrated Index output metric.

As illustrated in Figure 7, which shows the calibration data input window, StreetLight typically requests the following data to be specified for the calibration process:

- **Calibration Data Type.** Annual Daily Traffic (ADT), Average Annual Daily Traffic (AADT), Average Weekly Daily Traffic (AWDT), or Annual Average Weekly Traffic (AAWDT)
- **Calibration Value.** Daily traffic volume at the calibration location, corresponding to the calibration data type specified above.
- **Personal Traffic Ratio.** Percentage of passenger cars within the calibration volume. A default value of 0.96 (96%) is assumed.

- **Commercial Medium Duty Traffic Ratio.** Ratio of medium-duty trucks within the calibration volume. A default value of 0.02 is assumed, with suggested values between 0.005 and 0.01 for local arterials, and between 0.01 and 0.05 for highways.
- **Commercial Heavy Duty Traffic Ratio.** Ratio of heavy-duty trucks within the calibration volume. A default value of 0.02 is assumed, with suggested values between 0.0 and 0.005 for local arterials, and between 0.01 and 0.05 for highways.
- **Collection Method.** Optional field indicating how the calibration data was collected

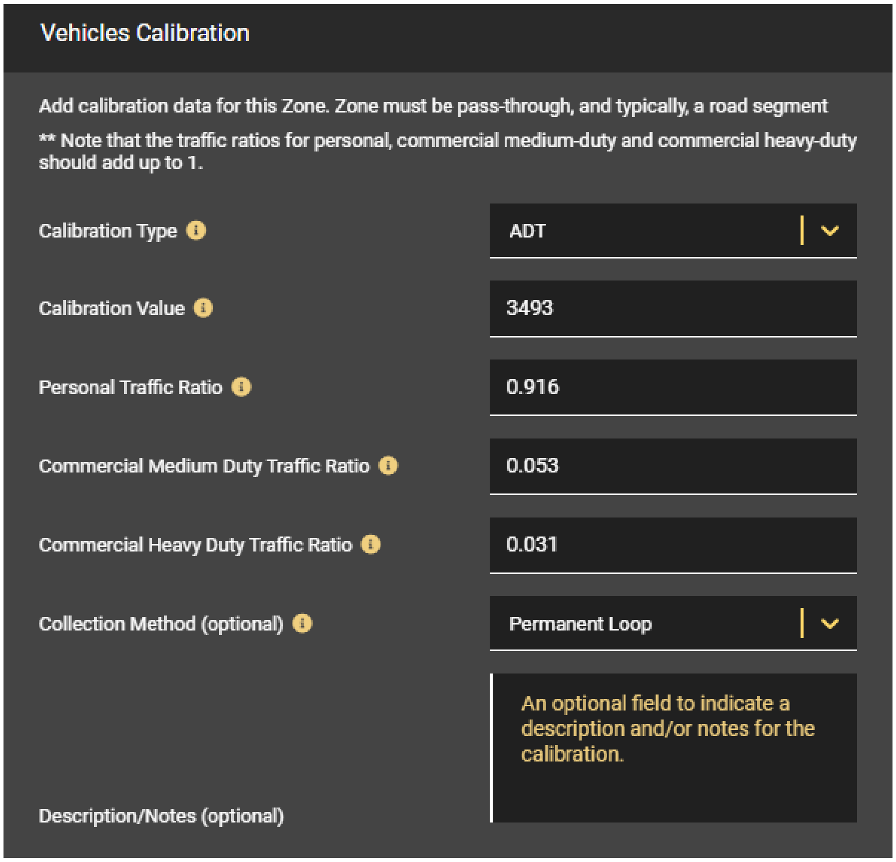


Figure 7. User-Defined Calibration Count Data Parameters

Output Examples

This section presents a few examples of truck-related analysis results from the StreetLight platform. These include:

- **Figure 8.** Top routes traveled by medium- and heavy-duty trucks from the Loleta WIM station on weekdays between January and June 2021
- **Figure 9.** Top routes traveled by medium- and heavy-duty trucks from all three District 1 WIM stations on weekdays between January and June 2021

- **Figure 10.** Observed medium-duty and heavy-duty truck trips at 20 different locations on weekdays between January and June 2021

- **Figure 11.** Characteristics of trips observed at a particular WIM station on weekdays between 6 AM and 7 PM.

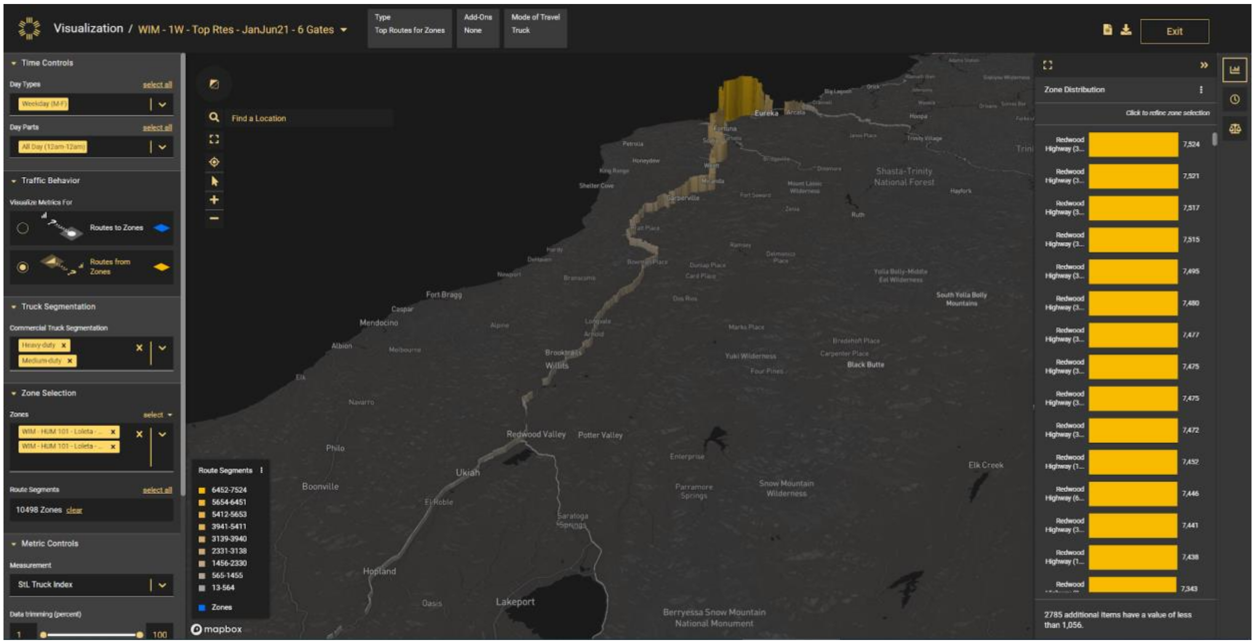


Figure 8. Analysis Example 1 – Top Routes from Loleta WIM Station

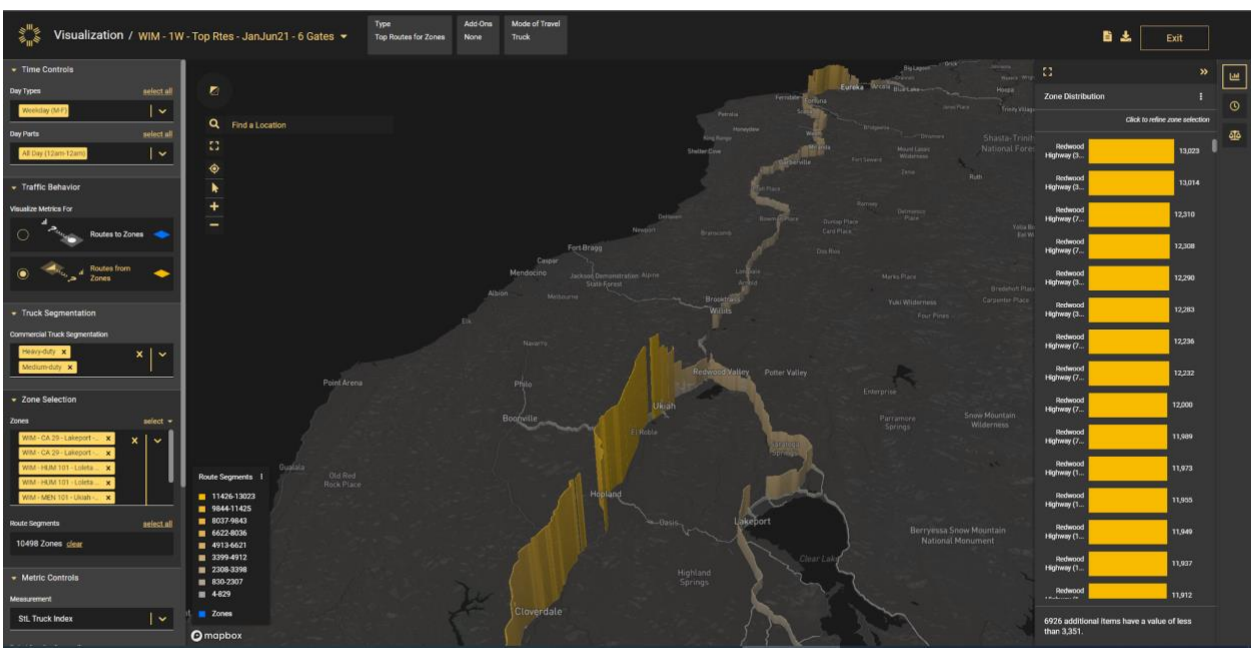


Figure 9. Analysis Example 2 – Top Routes from All WIM Stations

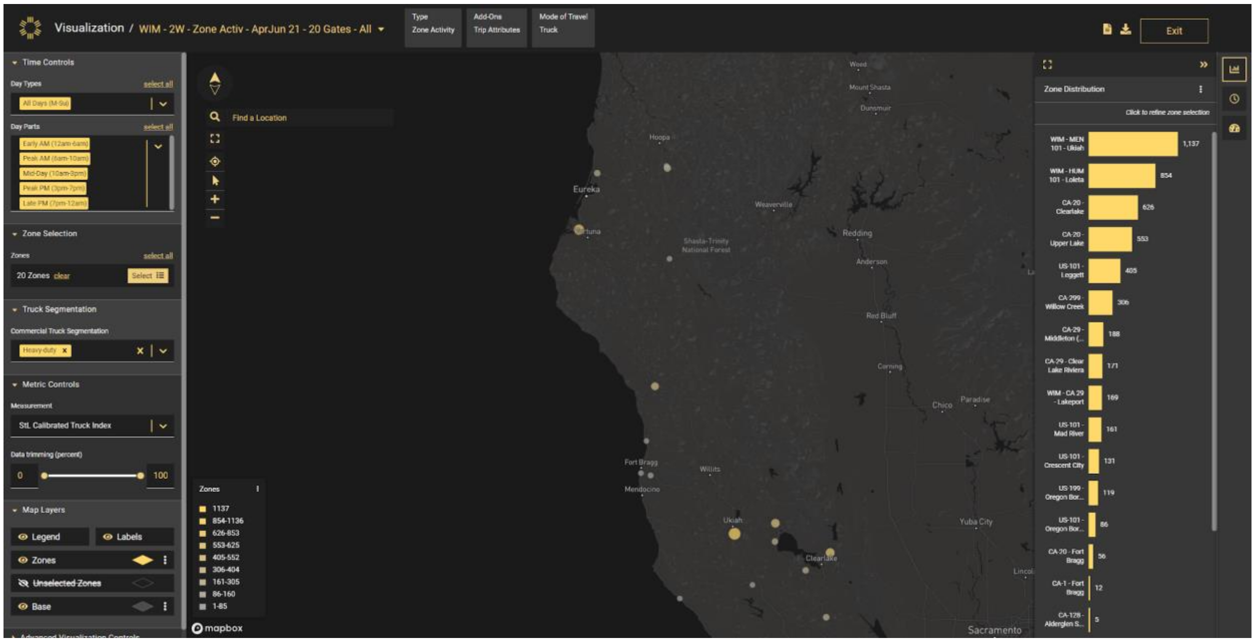


Figure 10. Analysis Example 3 – Zone Activity Analysis

Figure 11. Analysis Example 4 – Trip Characteristics Analysis

Analysis of StreetLight Truck Outputs

This section analyzes the ability of the StreetLight platform to potentially fill in the gaps in truck monitoring that were observed for Caltrans District 1. Specific elements covered in the section include:

- Review of StreetLight default output metrics for truck analyses
- Review of the single-factor calibration process used to convert StreetLight Index values into Calibrated Index values meant to represent more closely actual volumes
- Review of the Calibrated Index values produced for the road segments covering the three District 1 WIM stations under various sets of calibration data and for alternate analysis periods
- Summary observations

Streetlight Default Output Metrics for Truck Analyses

The two following metrics are the default outputs produced by StreetLight for truck-related analyses:

- StreetLight Sample Trip Counts (Truck Trips)
- StreetLight Index (Truck Trips)

These metrics are not adjusted; they represent a simple compilation of data contained in StreetLight's underlying datasets. Understanding what they represent is important as this helps assess the quality of the outputs produced and determine whether additional or different data might be required.

Streetlight Sample Trip Counts (Truck Trips)

The Streetlight Sample Trip Count represents the number of observed trips that have been captured by vehicle tracking technologies. This will typically be a fraction of all trips that may have occurred. As an example, StreetLight reported an all-vehicle Sample Trip Count of 13,018 vehicles for December 2020, and medium-duty and heavy-duty truck counts of 3,347 and 304 vehicles. For the same month, the WIM station counted 513,989 vehicles, with 98,651 of these being trucks. This means that the Sample Trip Counts represented just 3.1 percent of all vehicles and 3.7 percent of trucks.

For this project, sample classification counts were collected for the three WIM and 12 Traffic Census stations with full or recent coverage shown earlier in Figure 2. Data were collected for each month between January 2019 and December 2021 for both medium-duty and heavy-duty trucks. For each month, counts were further retrieved for all weekdays (Monday to Friday), all weekend days (Saturday-Sunday), and all days within a month. The resulting data are shown in Figure 12 through Figure 14.

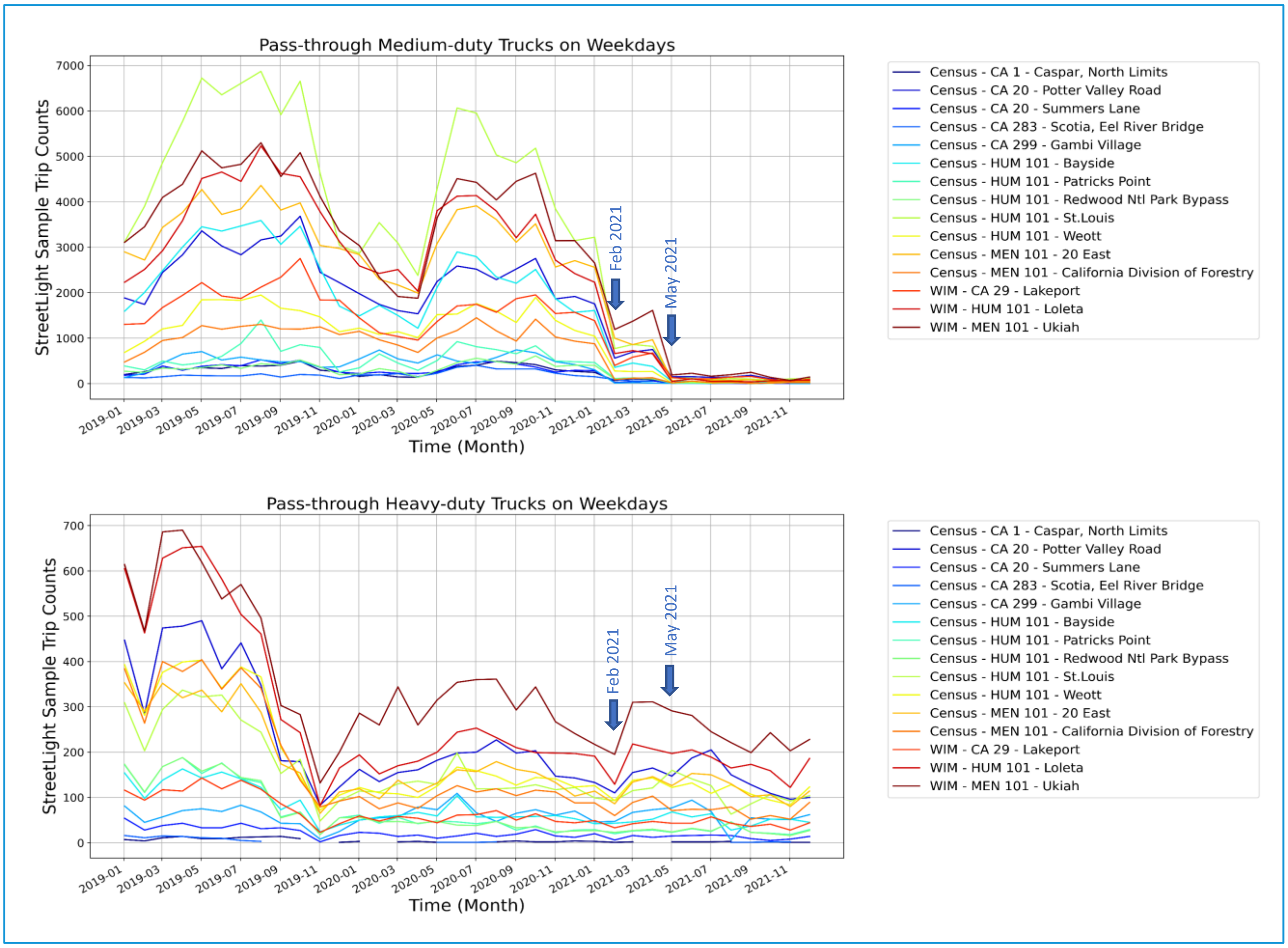


Figure 12. Monthly StreetLight Sample Trip Counts for Analysis Locations, Weekdays

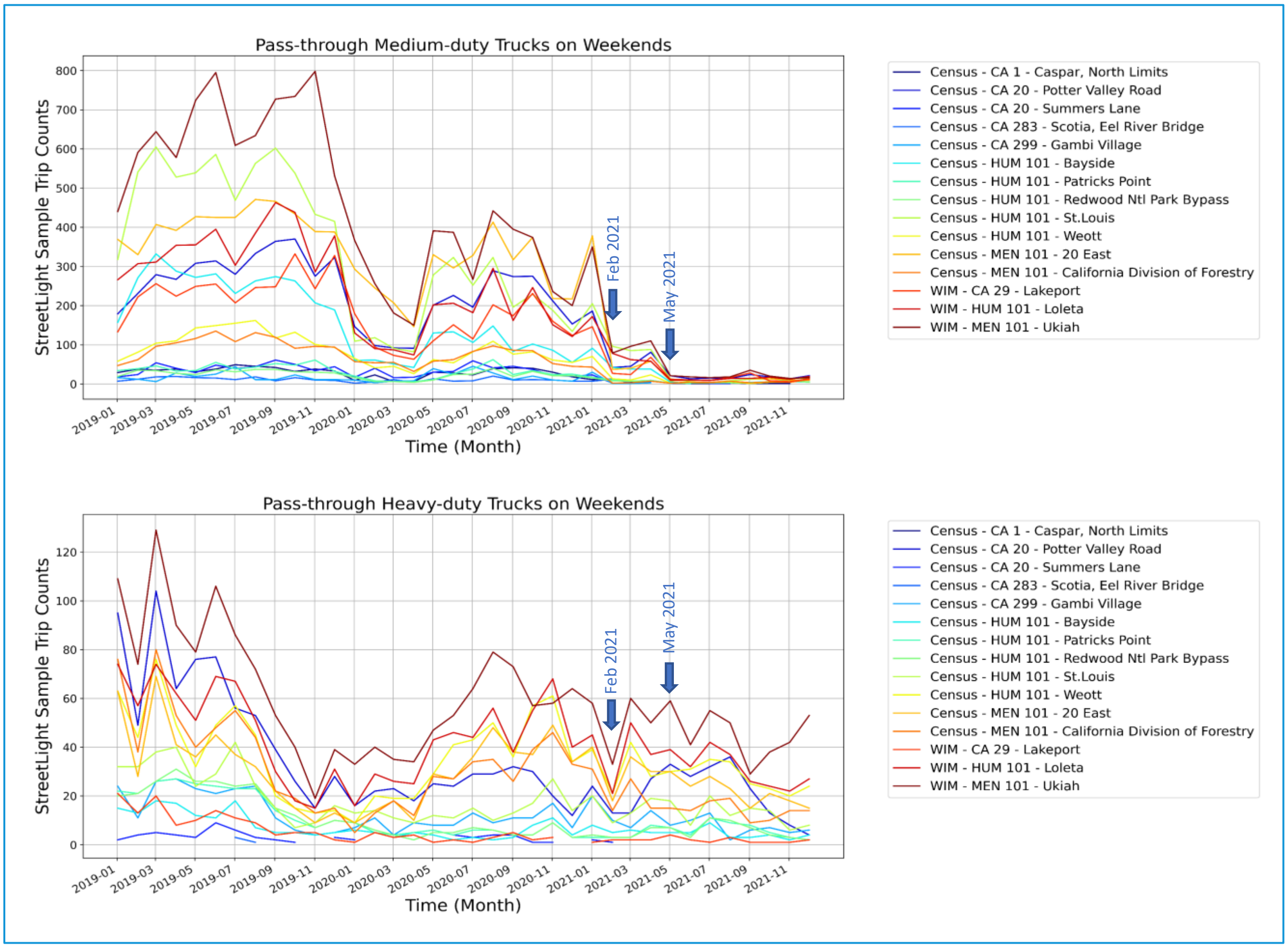


Figure 13. Monthly StreetLight Sample Trip Counts for Analysis Locations, Weekends

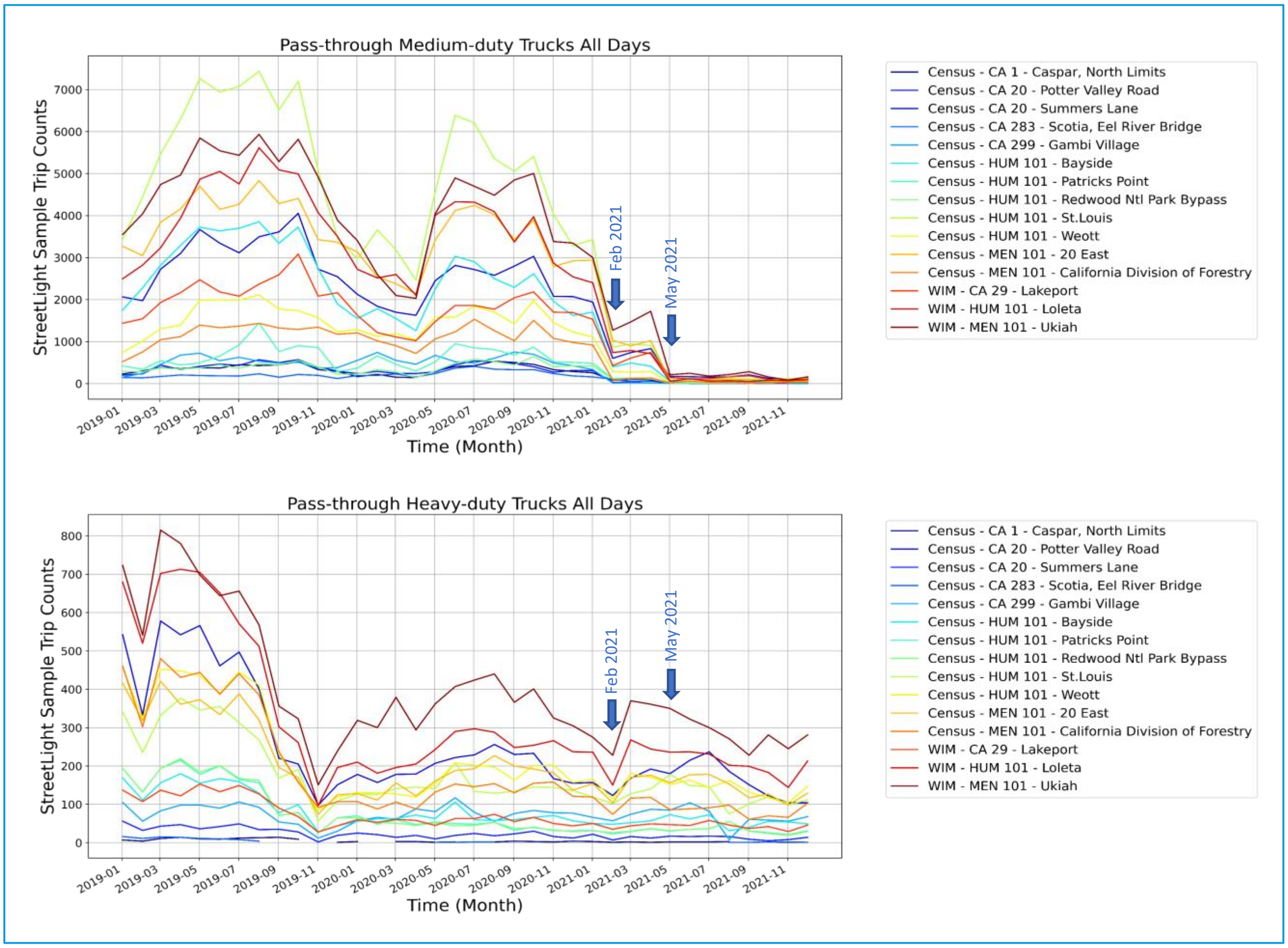


Figure 14. Monthly StreetLight Sample Trip Counts for Analysis Locations, All Days

Unlike other Streetlight outputs, Sample Trip Counts are not converted to an average day or adjusted for seasonal variations, penetration rate, or other factors. They represent the total number of trips within the StreetLight database related to the requested analysis. As can be observed, weekday monthly Sample Trip Counts for medium-duty trucks range from a few hundred to nearly 7,000, while heavy-duty truck counts remain below 700. Weekend counts are expectedly much lower, never exceeding 850 for medium-duty trucks and 130 for heavy-duty trucks.

There were significant drop-offs in the Sample Trip Counts for medium-duty trucks in 2021, in both the weekday and weekend data. The first drop-off occurred in February 2021 and the second in May 2021. Similar drop-offs were not observed for heavy-duty trucks. Based on discussions with StreetLight staff, these appear to be the result of two fleet data providers that successively stopped supplying tracking data to the StreetLight platform. The lack of impacts on the heavy-duty truck samples suggests that in this case the providers mainly supplied data about medium-duty trucks. These drop-offs are important, as they may affect the robustness of the Streetlight Index calculations, in addition to potentially affecting the single-factor calibration process.

Streetlight Index (Truck Trips)

The Streetlight Index for Truck Trips is the default metric for truck-related analyses. This is a normalized value derived from the Sample Trip Count associated with the analysis being conducted. The Index does not represent actual volumes, but a relative trip intensity accounting for variations in sample sizes across space and time. It is produced to allow relative comparisons of statistics across zones and periods. As will be detailed later, it is also a fundamental element of the single-factor calibration method.

StreetLight Index values were retrieved for each month between January 2019 and January 2022 for the three WIM and 12 Traffic Census stations with full or recent coverage shown in Figure 2. The resulting data are shown in Figure 15 to Figure 17, which respectively show the retrieved indices for an average weekday (Monday to Friday), average weekend day (Saturday-Sunday), and an average day of the month. In each figure, the top diagram illustrates the returned Index values for medium-duty trucks while the bottom graph shows the values for heavy-duty trucks.

Comparing the Index values to the Sample Trip Counts shown in Figure 12 to Figure 14 highlights the different nature of each metric. While monthly Sample Trip Counts remain below 6,000 on weekdays and below 130 on weekend days, Index values reach values up to 40,000 on weekdays and up to 4,500 on weekends. As indicated earlier, while the Sample Trip Counts report the number of trips captured by tracking technologies over each month, the Index values represent a relative average daily trip intensity for each month based on the sample trips.

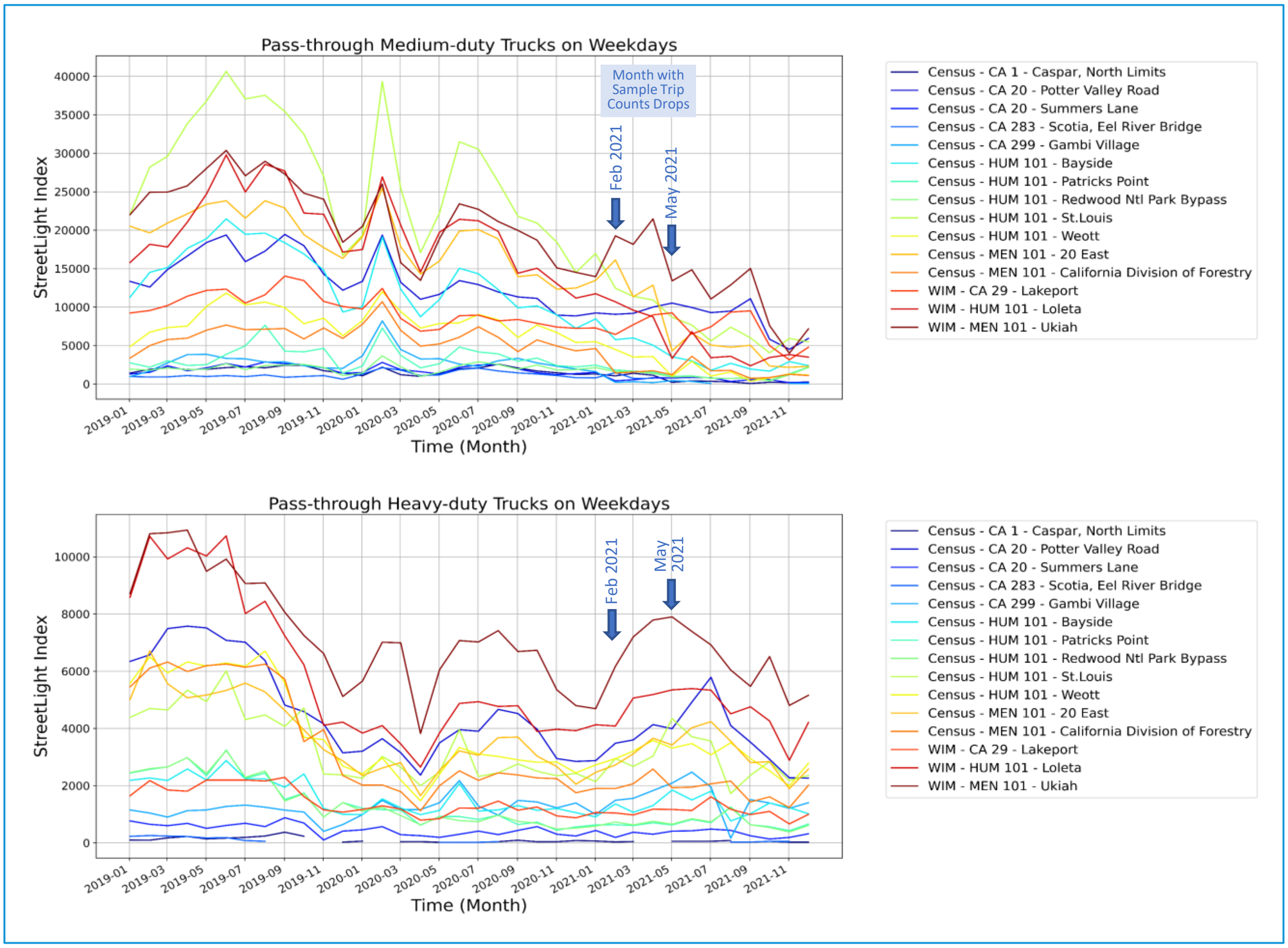


Figure 15. Monthly StreetLight Truck Index for Analysis Locations, Weekdays

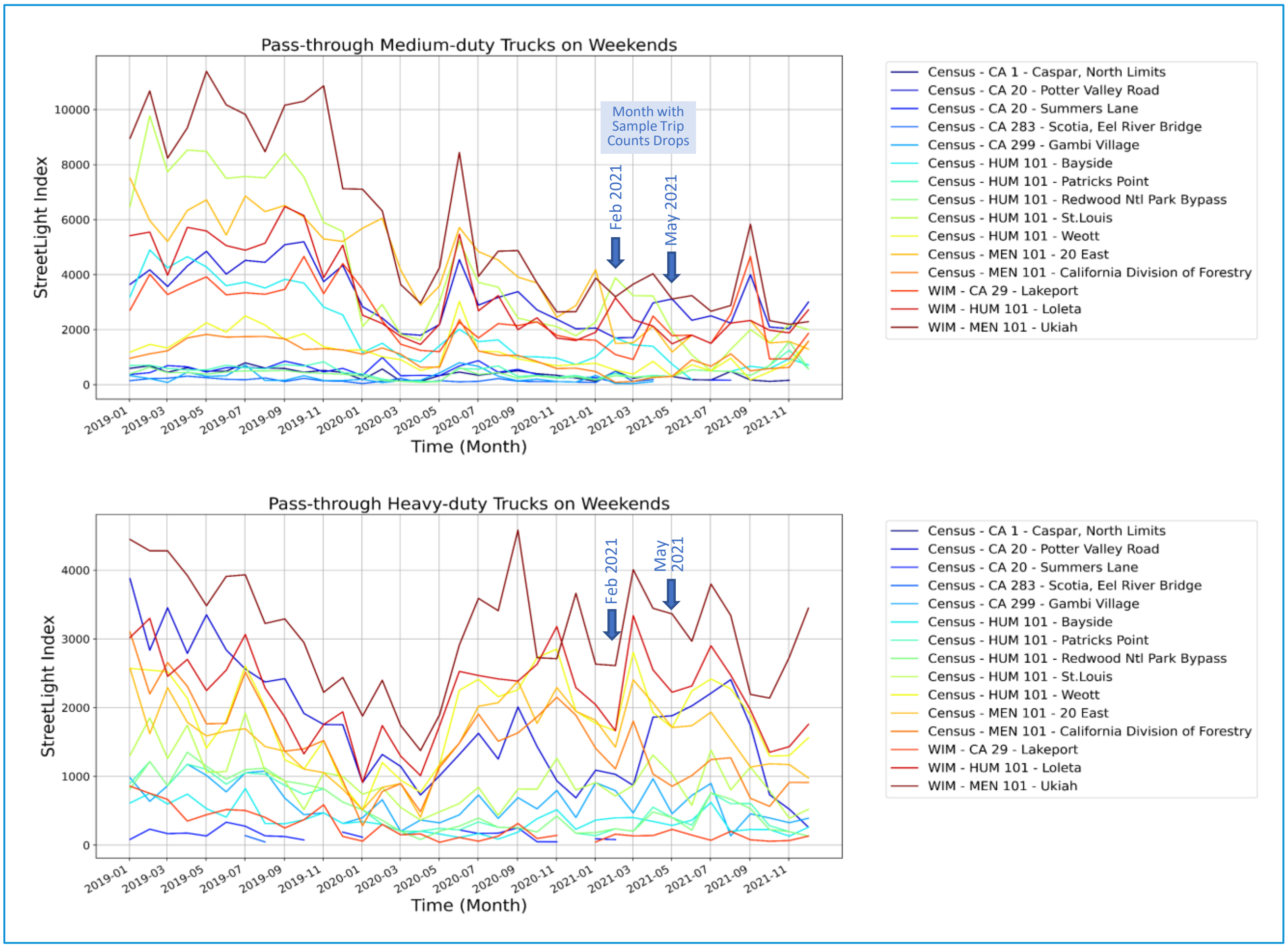


Figure 16. Monthly StreetLight Truck Index for Analysis Locations, Weekends

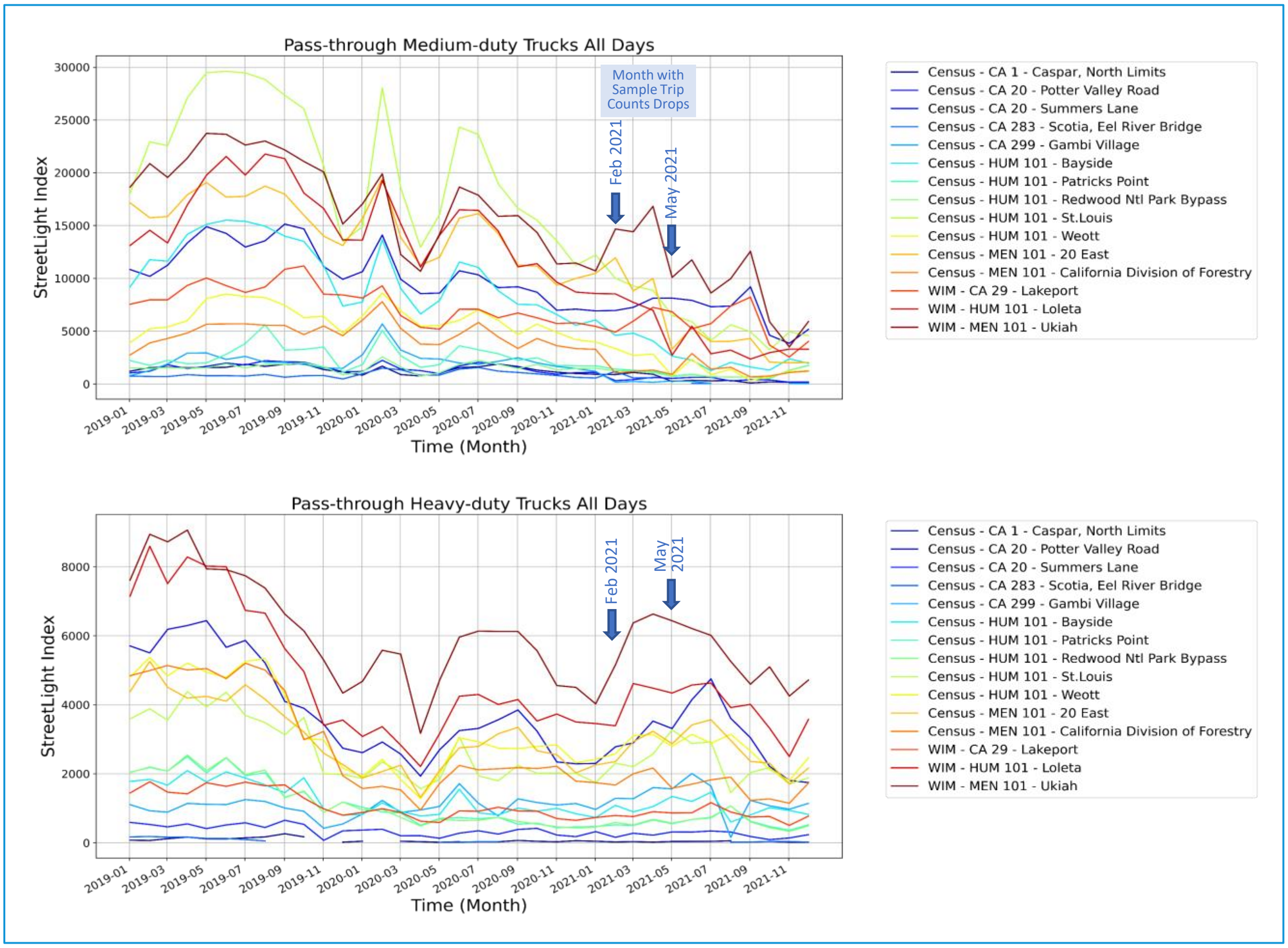


Figure 17. Monthly StreetLight Truck Index for Analysis Locations, All Days

More significant is how the Sample Trip Count and Index values can sometimes behave similarly and sometimes differently. While relatively similar month-by-month trends can be observed throughout 2019 and early 2020, particularly in terms of seasonal patterns, significant divergences appear in the medium-duty truck data when moving into 2021. These divergences can be partly linked to the February 2021 and May 2021 drop-offs in Sample Trip Counts described earlier as these drop-offs would have forced StreetLight to rely on much smaller samples of observed trips, and thus potentially less representative data, to produce Index values for 2021.

The above observations indicate that it is crucial to understand how StreetLight Index values are produced as this can help explain variations in the metrics that are generated and assist in better evaluating their reliability. While fluctuations in Sample Trip Counts can explain some of the observed variations, other underlying factors are also likely at play. This might partly explain why the May 2021 decline in Sample Trip Counts does not appear to reduce Index values by a corresponding magnitude. StreetLight staff indicated that truck Index calculations partly rely on data collected at a set of fixed traffic counting stations. For truck analyses, data collected around Sacramento has a major influence, as stations in this area are used to determine seasonal and annual trends. Some of the observed differences between Sample Trip Counts and the StreetLight Index could therefore partly reflect changes in trucking activities measured around Sacramento.

Optional Calibration Data Inputs

Within the StreetLight platform, optional calibration data can be supplied to steer the StreetLight Index towards values that may more realistically reflect observed volumes. Potential calibration data typically consist of the following elements:

- Average daily traffic volume on the roadway segment through the zone being considered
- Percentage of passenger cars, medium-duty trucks, and heavy-duty trucks within the specified average daily traffic volume

Specific recommendations from StreetLight on the selection of a set of suitable calibration data include the following:

- Data from 6 to 10 locations should be used at a minimum, and ideally 10-20 locations
- Use of data from permanent counters is preferred
- Data from the same types of road segments, days, periods, and settings should be used
- Use count data that represent typical conditions
- Avoid using data from highly congested locations

For District 1, the two best sources of calibration data are vehicle classification counts produced by WIM and Traffic Census stations spread across District 1. As shown earlier in Figure 2, these include:

- Two WIM stations along US-101 and one along CA-29
- 7 Traffic Census stations along US-101 and 7 others spread across the district

The following sections present a summary of activities that were conducted to extract suitable calibration data from the above-mentioned sources. Specific elements covered include:

- Review of calibration
- Example of single-factor calculation
- Available WIM all-vehicle counts
- Available all-vehicle counts from Traffic Census stations
- Categorization of WIM and Traffic Census counts into medium-duty and heavy-duty trucks
- Consideration of buses
- Resulting truck calibration data

Calibration approach

Within the StreetLight platform, calibration refers to the process of adjusting the default calculated Index values to values representing more closely, but not necessarily exactly, observed real-world counts. The process relies on user-provided calibration data and typically consists of the following two steps:

- For each calibration location, a local adjustment factor is calculated from the average daily traffic volume and truck percentages supplied
- The single adjustment factors obtained for each calibration location are averaged to obtain a single overall adjustment factor
- The resulting overall adjustment factor is then applied to the Streetlight Index values associated with each prediction location to create Calibrated Index values

The resulting Calibrated Index values represent an estimate of the number of trips that would be observed on a particular roadway segment or pass-through zone based on the supplied calibration data. However, as will be demonstrated later, while the process can generally bring the StreetLight Index values closer to observed volumes, it cannot guarantee high-accuracy results if large differences exist across space and time between the flow patterns captured by the calibration data and the Index values produced by StreetLight.

Single Factor Calculation Example

The following example, replicated from information provided in the Streetlight Help Center on the topic of calibration at *https://support.streetlightdata.com/hc/en-us/articles/360024521892*, illustrates how calibration data is used to adjust analysis outputs.

Table 2. Calculation of Calibration Single-Factor Adjustment Parameter

Calibration Zone	StreetLight Index	Calibration Volume	Calibration Factor
Zone 1	1000	10,000	**10.00**
Zone 2	180	1,000	**5.56**
Zone 3	40	500	**12.50**
		Average of Calibration Factors	**9.35**

In Table 2, the second column lists the StreetLight Index that would be associated with each calibration zone, while the third column lists the user-provided calibration volume for each of the zones. In the last column, a calibration factor is determined by simply dividing the user-supplied calibration volume, such as counts from WIM or Traffic Census stations, by the associated StreetLight Index. This represents the adjustment that would be needed to bring the Index to a value matching the observed count. As can be observed, the resulting factors vary between 5.56 and 12.50. At the end of the process, an overall calibration factor of 9.35 is determined by simply averaging the local factors.

An element to keep in mind here is that the above calculations treat each data location equally. Since the overall factor is determined by simply averaging all the local factors, the calculation ignores that data from sources with higher Sample Trip Counts might be more reliable than locations with limited data or that potential geographic connections might exist between calibration and prediction locations. Consider for instance two calibration zones named A and B, and a prediction zone named C. If zone A is much closer to zone C than zone B, small changes in the data provided by zone A could be expected to have a greater impact on the estimations from zone C than if the changes were to occur in zone B.

Table 3. Example of Single-Factor Calibration Adjustment

Origin	Destination	StreetLight Index	Calibration Factor	Calibrated Index (Estimated Trips)
Zone 1	Zone 2	100	**9.35**	**935**
Zone 1	Zone 3	250	**9.35**	**2,338**
Zone 1	Zone 4	350	**9.35**	**3,273**
Zone 2	Zone 1	200	**9.35**	**1,870**

Table 3 further indicates how the resulting calibration factor is used to convert Streetlight Index values into calibrated indices estimating trip counts. For each pair of origin-destination zones, the adjustment simply consists of multiplying the Index normally returned by StreetLight by the calibration factor calculated over all the calibration zones. As an example, this process converts the Index value of 100 associated with trips between Zones 1 and 2 into an estimated 935 trips.

Available Weigh-in-Motion (WIM) Counts

Figure 18 and Figure 19 show the average all-vehicle daily weekday and weekend traffic that were retrieved from the three District 1 WIM stations for each month between January 2019 and October 2021, except for May 2019 at the Loleta station and between January 2019 and March 2020 at the Ukiah station, when technical issues resulted in no data being returned.

As can be observed, all-vehicle counts from the three stations share similar patterns across the months. All the counts first exhibit annual cyclical patterns with peak traffic in the summer and low traffic in December/January. They also all further show a significant drop in observed traffic in April 2020, at the start of the period with stay-at-home orders from the Covid-19 pandemic, followed by partial recovery over the following four months.

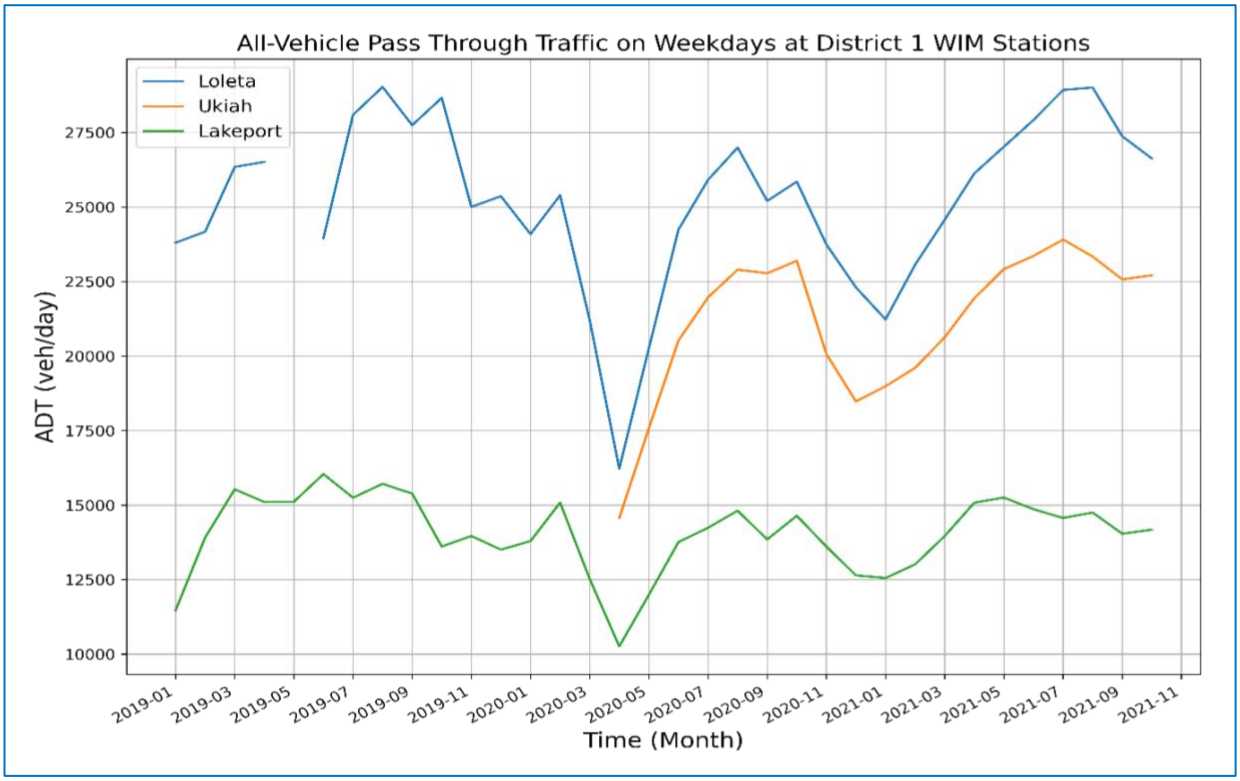


Figure 18. Monthly ADTs from WIM Stations, All Vehicles, Weekdays

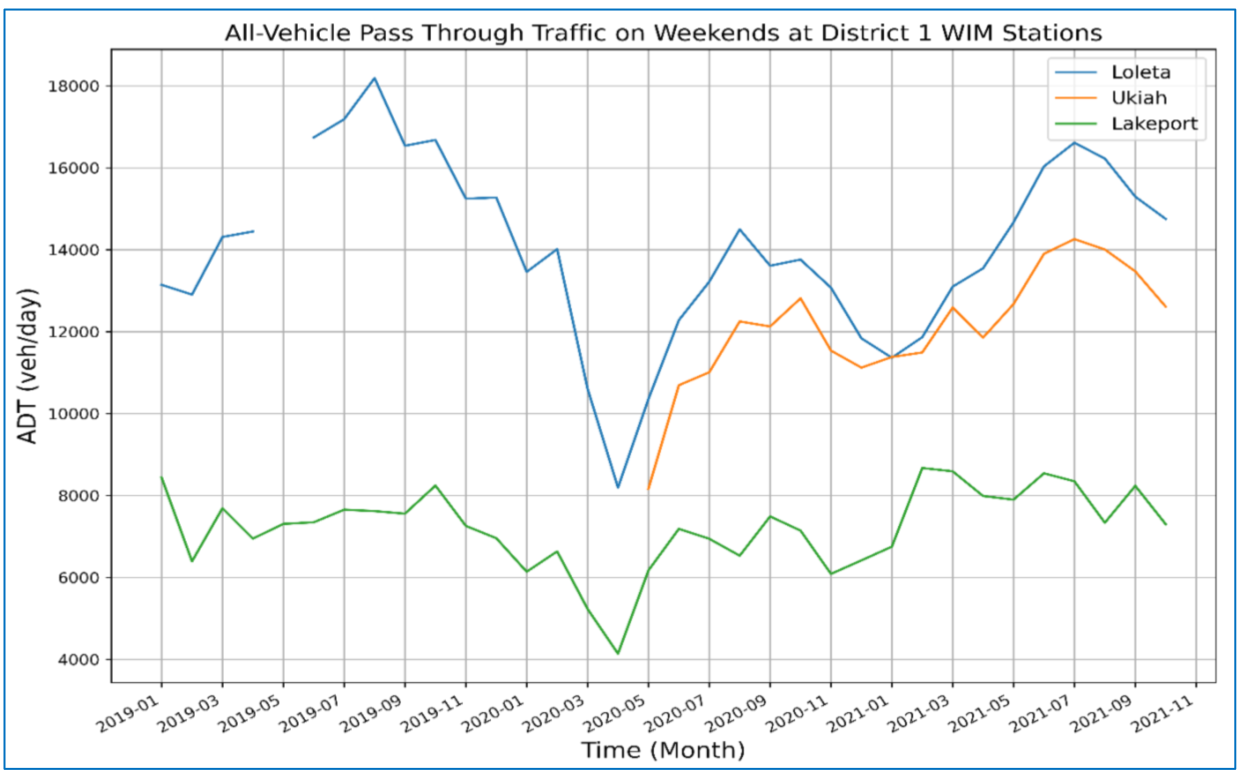


Figure 19. Monthly ADTs from WIM Stations, All Vehicles, Weekends

In terms of volumes, the Loleta station on US-101 typically has the highest recorded traffic, while the Ukiah station further north on US-101 ranks second. However, while Loleta has the highest overall traffic, Ukiah generally has the highest truck volumes. The Lakeport station on CA-29 usually has much lower traffic than the two other stations. Weekday and weekend traffic also share similar patterns. The main difference is in the magnitude of counts, with weekend traffic being expectedly much lower.

However, some inconsistencies can be found when comparing the WIM counts to the Streetlight Index values shown in Figure 15 and Figure 16. In May 2021, the weekday StreetLight Index values for medium-duty trucks dropped significantly while the WIM counts increased at all stations. This is likely an issue internal to StreetLight caused by the May 2021 Sample Trip Count drop-off described earlier.

The above divergences are important to note, as they indicate that changes in the underlying data supporting the StreetLight analyses could occasionally affect truck volume estimates in a way that does not reflect actual traffic trends.

Traffic Census All-Vehicle Counts

Figure 20 and Figure 21 show the average daily weekday and weekend traffic recorded at the 14 Traffic Census stations identified earlier. This is for all vehicles. This includes data from seven stations along US-101 and seven other stations spread across the district. Data from the US-101 stations are shown in a separate figure as these are located on the main north-south road going across District 1 and tend to carry more traffic, particularly in the southern portion of the district.

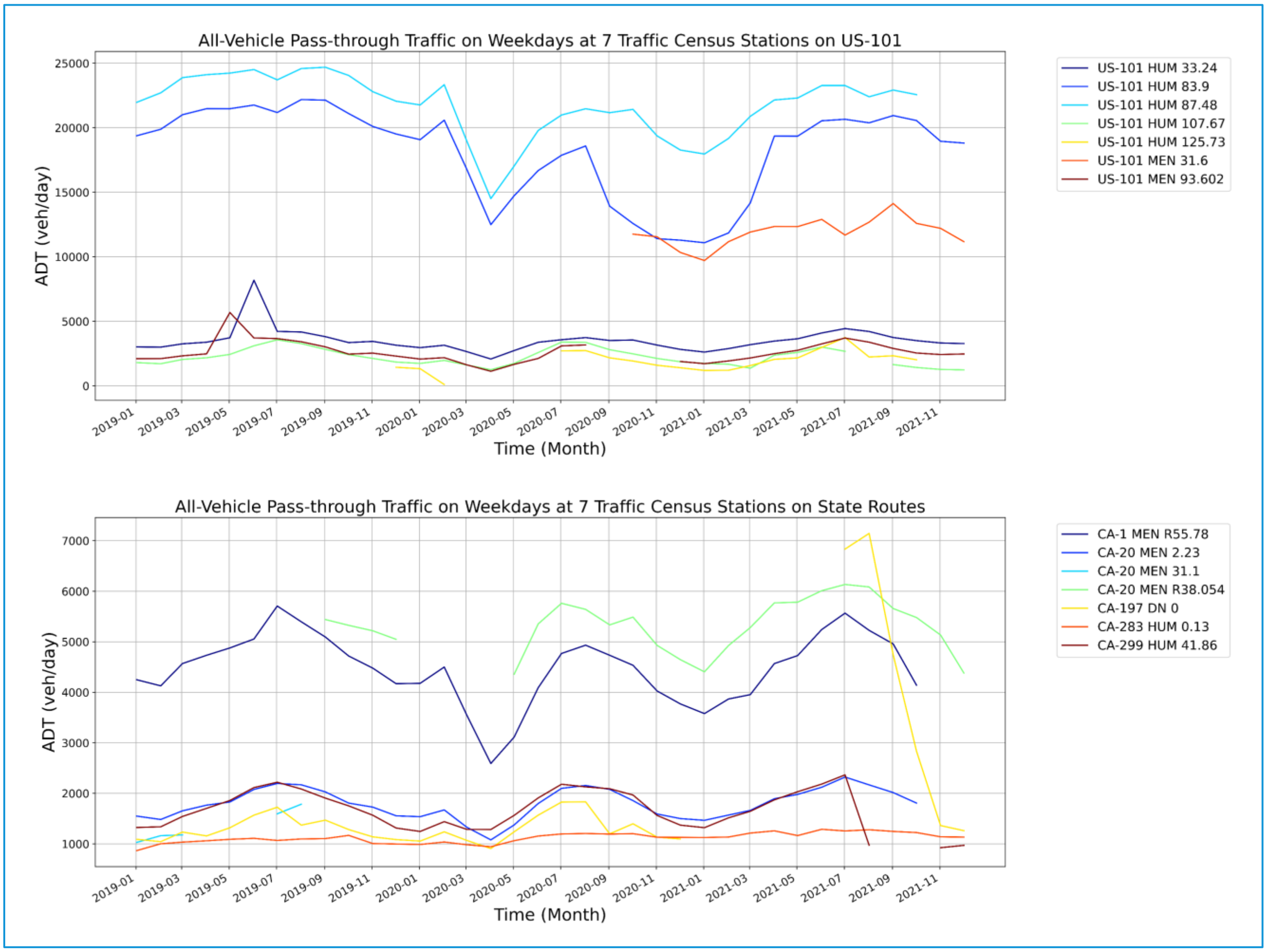


Figure 20. Monthly ADTs from Traffic Census Stations, All Vehicles, Weekdays

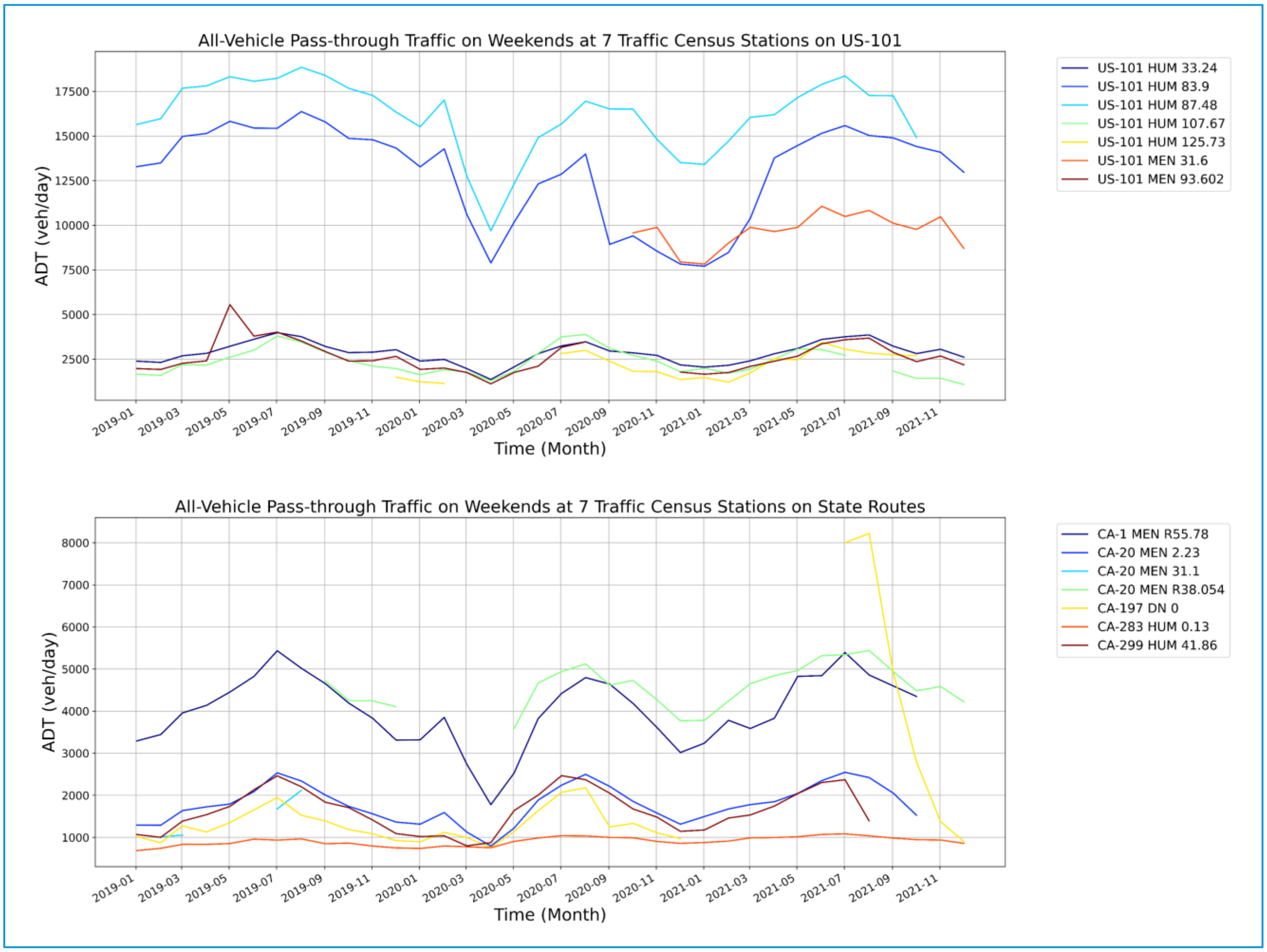


Figure 21. Monthly ADTs from Traffic Census Stations, All-Vehicles, Weekends

Some broken lines can be noticed in each graph. These are the results of missing data. For each month, the mapped ADT is the average over all available weekday or weekend days. The two most problematic stations were the Broaddus Creek (CA 20 MEN 31.1) and Junction 199 (CA 197 DN 0) stations. The Broaddus Creek station produced data for only a handful of months in 2019 while the Junction 199 station returned unrealistically large counts of Class 5 trucks throughout the entire analysis period and high numbers of passenger cars between July 2021 and October 2021. The station also provided no counts from January to June 2021. This led to their exclusion from further analyses, leaving only 12 stations from which to extract calibration data.

Similar to the WIM data, most of the Traffic Census stations share similar traffic patterns over time. This includes a significant decline in traffic in April 2020 related to the Covid-19 pandemic and cyclical summer peaks and winter lows. Weekday and weekend traffic show similar patterns, except in the magnitude of the captured volumes. The same discrepancies between the captured values and produced Index values that were noted for the WIM data can also be observed here.

Medium-Duty and Heavy-Duty Truck Categorization

Based on the fleet tracking data obtained by StreetLight, which typically comes from INRIX, the following two categories of trucks are defined within the StreetLight platform:

- **Medium-duty trucks**, representing trucks with gross vehicle weight between 14,000 lbs. and 26,000 lbs.
- **Heavy-duty trucks**, representing trucks heavier than 26,000 lbs.

There is no categorization for light-duty trucks due to a lack of tracking data. StreetLight assumes these vehicles are part of the passenger car category.

As indicated earlier, the ratios of medium-duty and heavy-duty trucks are input into the calibration process. If no value is provided, StreetLight will assume that two percent of the traffic is made up of medium-duty trucks and two percent of heavy-duty trucks. This is not reflective of District 1 traffic. Across the three WIM stations, medium-duty trucks comprise between three percent and ten percent of traffic, while heavy-duty trucks comprise between one percent and six percent. Since using the default values would not be representative, we used district-specific data for the analyses.

We used classification data from both WIM and Traffic Census stations to estimate the proportions of medium-duty and heavy-duty trucks. A challenge in doing this was that both sources do not use the same categories as StreetLight to classify trucks:

- WIM stations classify vehicles into 15 axle-based classes defined by Caltrans as shown earlier in Figure 38 or by their measured weight, using bins with a 10,000 lbs. increment
- Traffic Census stations only classify trucks in terms of the 15 classes defined by Caltrans

For the WIM data, the measured weights allow dividing vehicles between those weighing less than 14,000 lbs., 14,000 to 26,000 lbs., and more than 26,000 lbs. to match the passenger vehicle, medium-duty truck, and heavy-duty truck categorization used by StreetLight. However, this was not expected to produce perfect matches as measured weights are not gross vehicle weights. Measured weight depends on whether a truck is loaded and the specific weight of what is being carried. Gross vehicle weight is a static number representing the potential weight of a truck when fully loaded. Since the measured weight can be lower than the gross vehicle weight, a slight bias towards mischaracterizing some heavy-duty trucks as medium-duty trucks and medium-duty trucks as light-duty trucks, or passenger vehicles, may exist.

For the Traffic Census data, vehicles were classified into one of four categories based on their assigned classes:

- Passenger cars included classes 1 through 4 (motorcycles, passenger cars, pickup trucks, and buses)
- Medium-duty trucks included classes 5, 6, and 7 (single-unit trucks)
- Heavy-duty trucks included classes 8 to 14 (single- and multi-trailers)
- Vehicles in class 15 were not considered, as this class represents unclassified vehicles and errors

A more refined classification could potentially be achieved by first using WIM data to determine the typical proportions of trucks weighing between 14,000 and 26,000 lbs., and over 26,000 lbs. within each of the Caltrans vehicle classes and then using this average characterization to apportion the Traffic Census data into medium-duty and heavy-duty trucks.

Consideration of Buses

Bus weight can vary between 24,000 lbs. and 40,000 lbs. Based on the vehicle classification used by StreetLight, bus weights fall within the range of weights used to identify medium-duty and heavy-duty trucks. Technically, buses could therefore be considered to belong to one or both of the truck categories. However, buses were not considered in this study as StreetLight analyses bus trips through a separate module. This results in calculated truck metrics that do not typically include buses.

Within StreetLight, buses are indirectly considered within the All-Vehicle metrics. Mobile devices held by passengers typically result in the creation of an individual trip for each person holding a device being tracked. The selected output metric then determines how trips that might have been made on a bus would be considered and reported:

- **Sample Trip Count.** All tracked devices are reported as individual rides or trips
- **StreetLight Index.** All tracked devices are considered as individual trips, resulting in an Index that estimates the intensity of trips, rather than a traffic volume
- **StreetLight Volume.** A bus is reported as a single vehicle if StreetLight can associate all the individual recorded trips made while riding the bus to a single vehicle. Since this is not always the case, a single bus trip may thus be represented as involving more than one vehicle. As indicated earlier, this metric is not available yet for truck-related analyses.

Resulting Truck Calibration Data

Figure 22 to Figure 24 present the estimated average daily traffic counts for medium-duty and heavy-duty trucks that were derived from the available WIM and Traffic Census counts for each month between January 2019 and December 2021. As the StreetLight Sample Trip Counts and StreetLight Index data presented earlier, the figures respectively illustrate count estimates that were produced for an average weekday, weekend day, and day of the month.

As shown earlier in Figure 7, StreetLight's calibration process relies on users providing an all-vehicle average daily traffic volume and the percentages of personal traffic (passenger cars and light-duty trucks), commercial medium-duty trucks, and commercial heavy-duty trucks for each calibration location. In this context, all-vehicle average daily traffic volumes for weekday and weekend analyses were taken directly from the counts illustrated in Figure 20 or Figure 21. For analyses considering both weekday and weekend activities, the traffic volumes were further determined by summing the data from both figures. After obtaining the average all-vehicle volumes, the percentages of vehicle types were calculated for each analysis by comparing the truck counts shown in Figure 22 to Figure 24 to the corresponding average all-vehicle counts over the given analysis period.

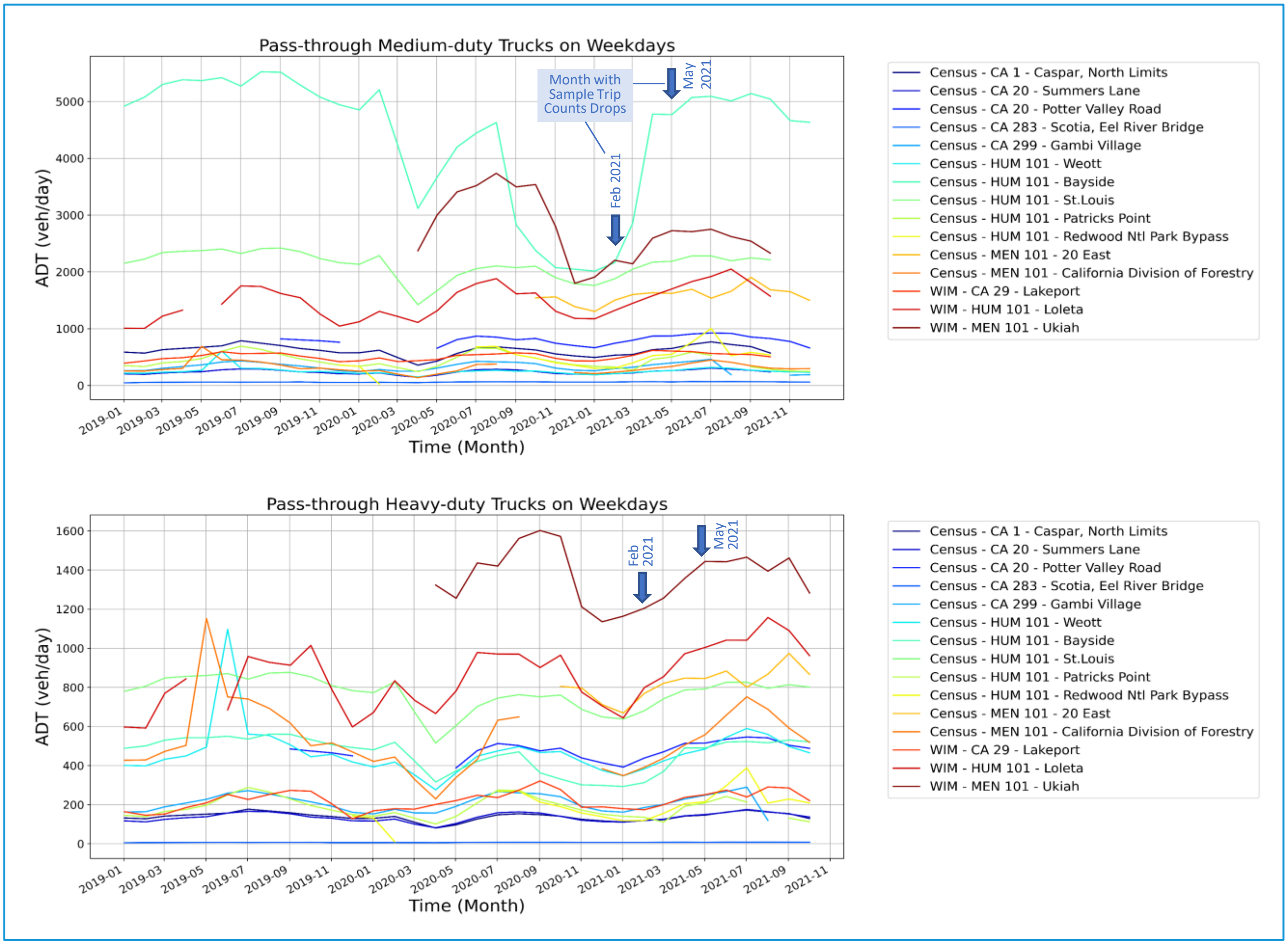


Figure 22. Estimated Monthly Truck ADTs from WIM and Census Stations, Weekdays.

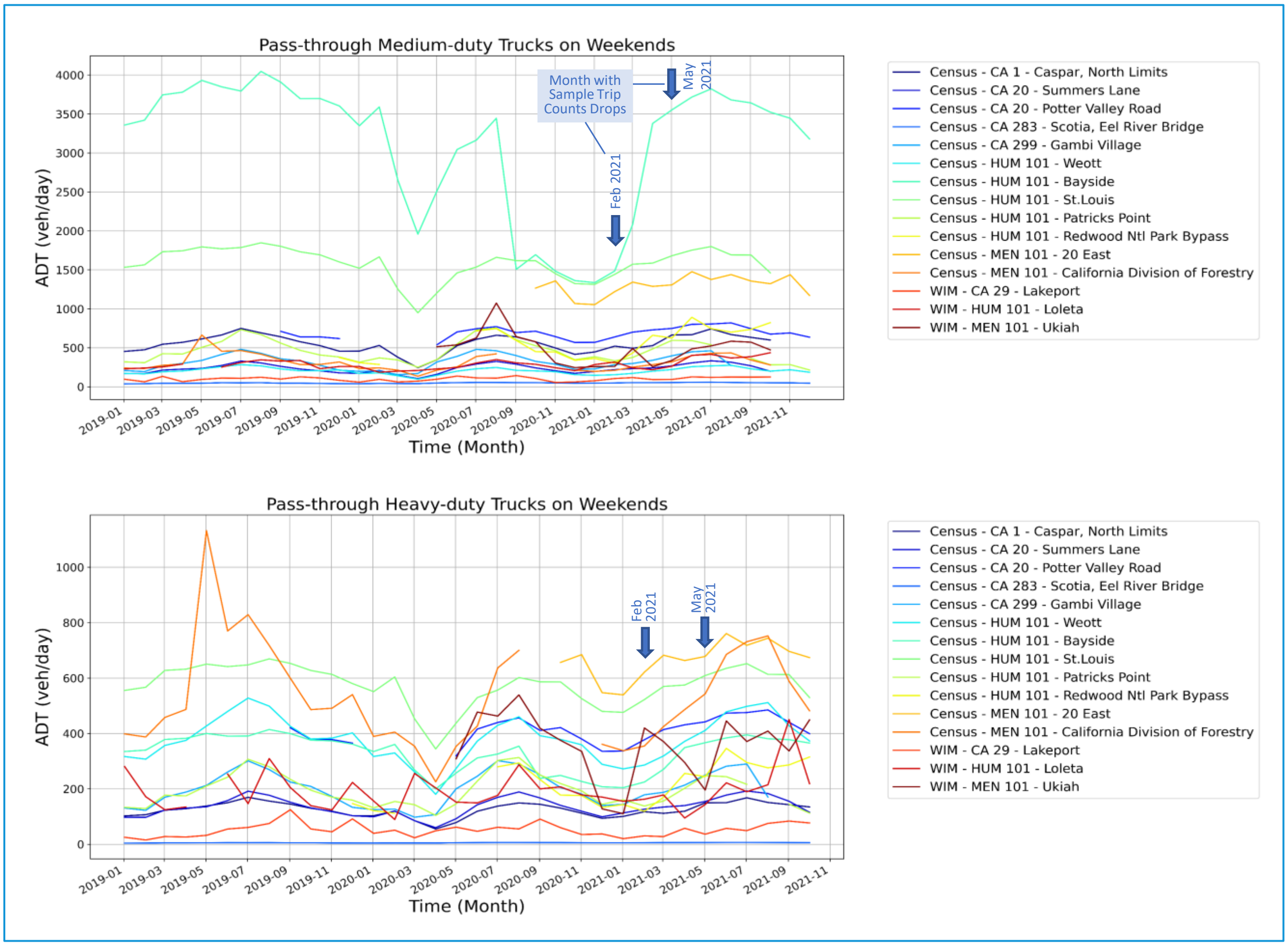


Figure 23. Estimated Monthly Truck ADTs from WIM and Census Stations, Weekends

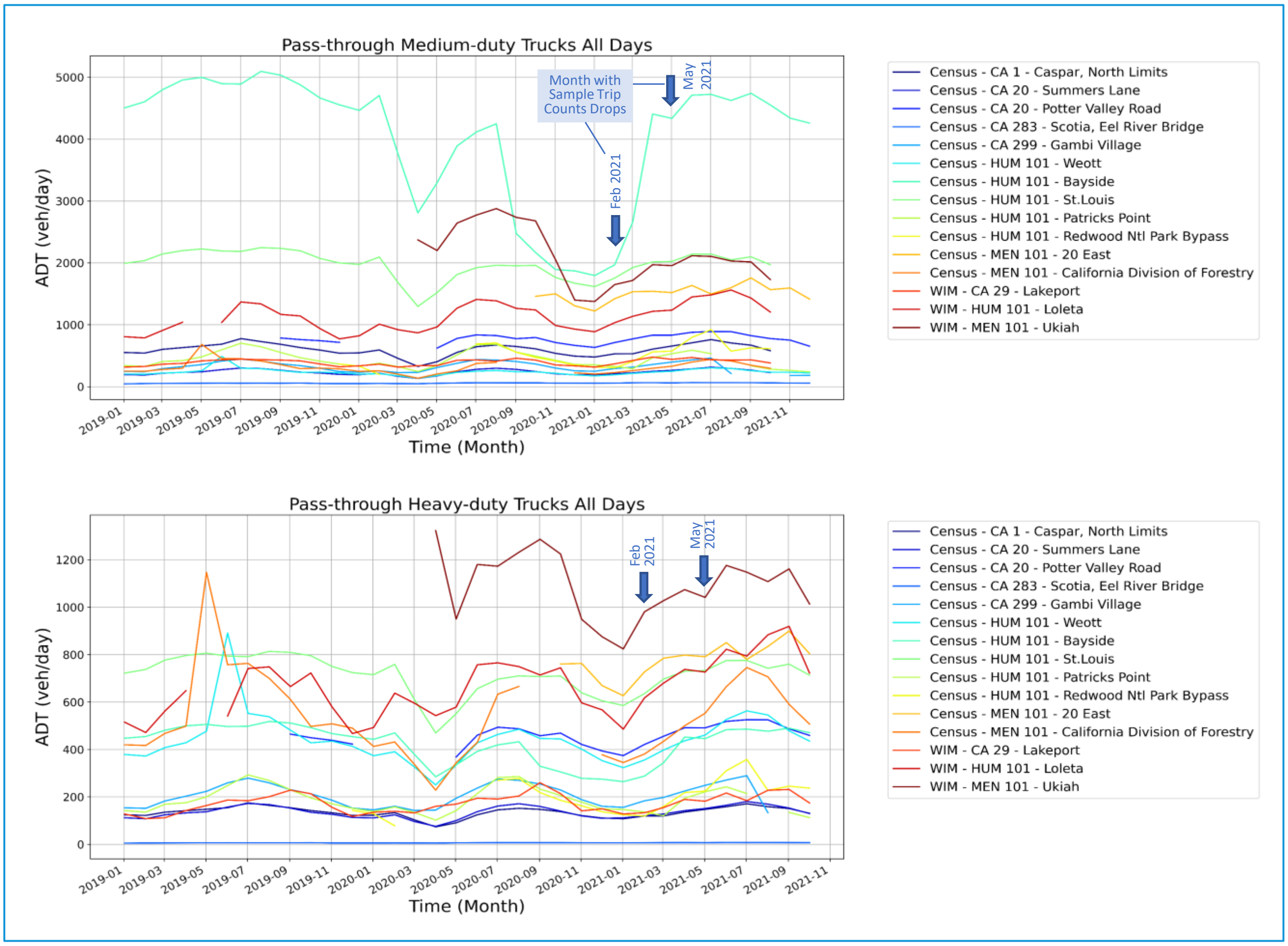


Figure 24. Estimated Monthly Truck ADTs from WIM and Census Stations, All Days

For comparative purposes, Figure 25 presents the average ADTs and truck percentages that were estimated over an analysis period extending from January 2019 to October 2021 for all stations considered as potential calibration locations for all days of the week. Based on the illustrated data, the following observations can be made:

- All-vehicle traffic volumes along US-101 vary significantly depending on the section considered. Peak traffic areas appear to be around Eureka, with ADTs around 40,000 vehicles/day, and in the southern end of the district around Ukiah, with ADTs around 20,000 vehicles/day. Sections between Ukiah and Eureka and north of Eureka have ADTs between 4,000 and 6,000 vehicles/day.
- The estimated percentage of medium-duty truck traffic along US-101 varies between 3.6 percent and 12.3 percent, for an average of 7.9 percent across the various stations
- The estimated percentage of medium-duty truck traffic on roads outside the US-101 varies between 2.7 percent and 9.5 percent, for an average of 6.1 percent across the various stations
- The estimated percentage of heavy-duty trucks along US-101 varies between 1.3 percent and 10.2 percent, for an average of 4.6 percent across the various stations

- The estimated percentage of heavy-duty trucks on roads outside the US-101 varies between 0.3 percent and 6.1 percent, for an average of 2.9 percent across the various stations

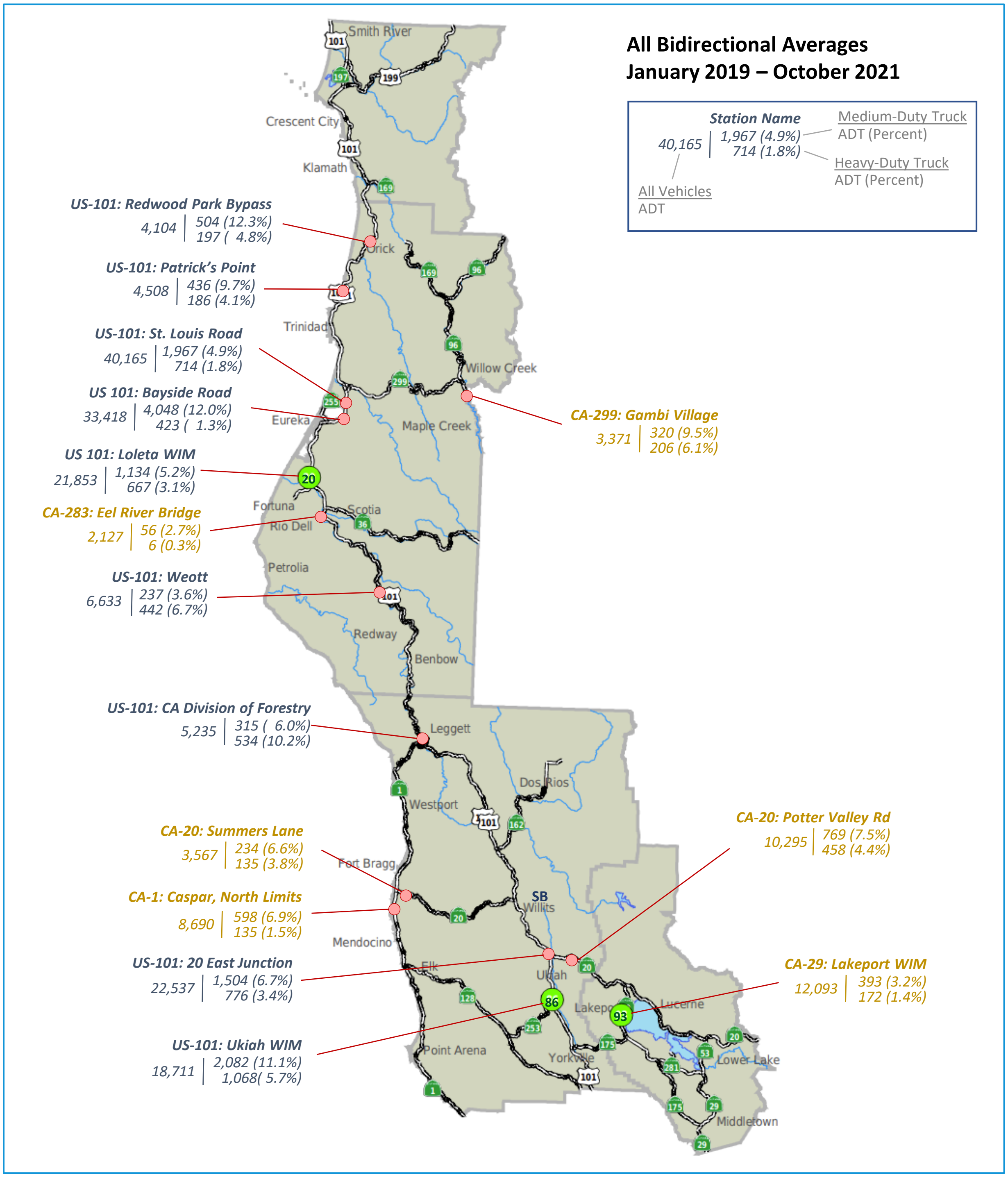


Figure 25. Estimated Average Truck ADTs for Calibration Locations over the Analysis Period

Analysis of StreetLight Estimated Truck Outputs for District 1

This section presents an analysis of various truck-related outputs produced by StreetLight for Caltrans District 1. Specific elements discussed herein include:

- Stability of StreetLight Index relative to actual counts
- Analysis of outputs using different sets of calibration data
- Analysis of outputs covering different analysis periods
- Showcases highlighting potential issues with the single-factor calibration process

StreetLight Index Stability Relative to Observed Counts

To assess the stability of the StreetLight Index to observed counts, the ratios between the StreetLight Index and observed truck counts were calculated for each month between January 2019 and December 2021 for the three WIM and 12 Traffic Census stations from which truck ADTs were estimated. This index-to-volume ratio represents the adjustment needed to bring the StreetLight index to a value exactly matching an observed count. In this context, a value of 1.0 would mean that the Index matches exactly the observed count, while a value greater than 1.0 would indicate that the index underestimates the actual count, and a value lower than 1.0 that an overestimation has occurred.

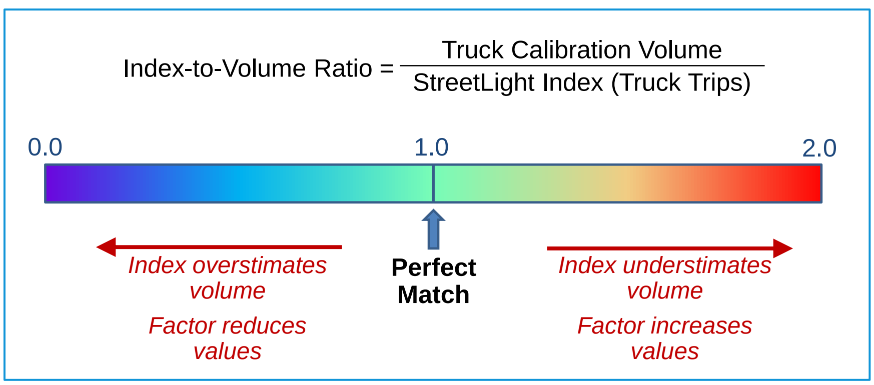


Figure 26. Color-Code Value of Index-to-Volume Ratios

To help with the assessment, the ratios are presented in the color-coded format of Figure 26. A perfect match would show as green, an overestimation as a shade of blue, and an underestimation as a shade of red. Darker shades of red or blue would indicate that a greater adjustment is needed. The colors are not an indication of worsening or better data quality, but simply an indication of the magnitude of the adjustment needed to bring the StreetLight Index value closer to the observed counts.

Figure 27 to Figure 29 present the results of the ratio calculations that were performed for each month between January 2019 and December 2021 for data reflecting an average weekday, an average weekend day, and an average day of the month. In each figure, the top diagram represents the ratios associated with medium-duty truck data, while the bottom figure provides ratios related to heavy-duty trucks. Months shown in white correspond to periods for which no WIM or Traffic Census counts are available.

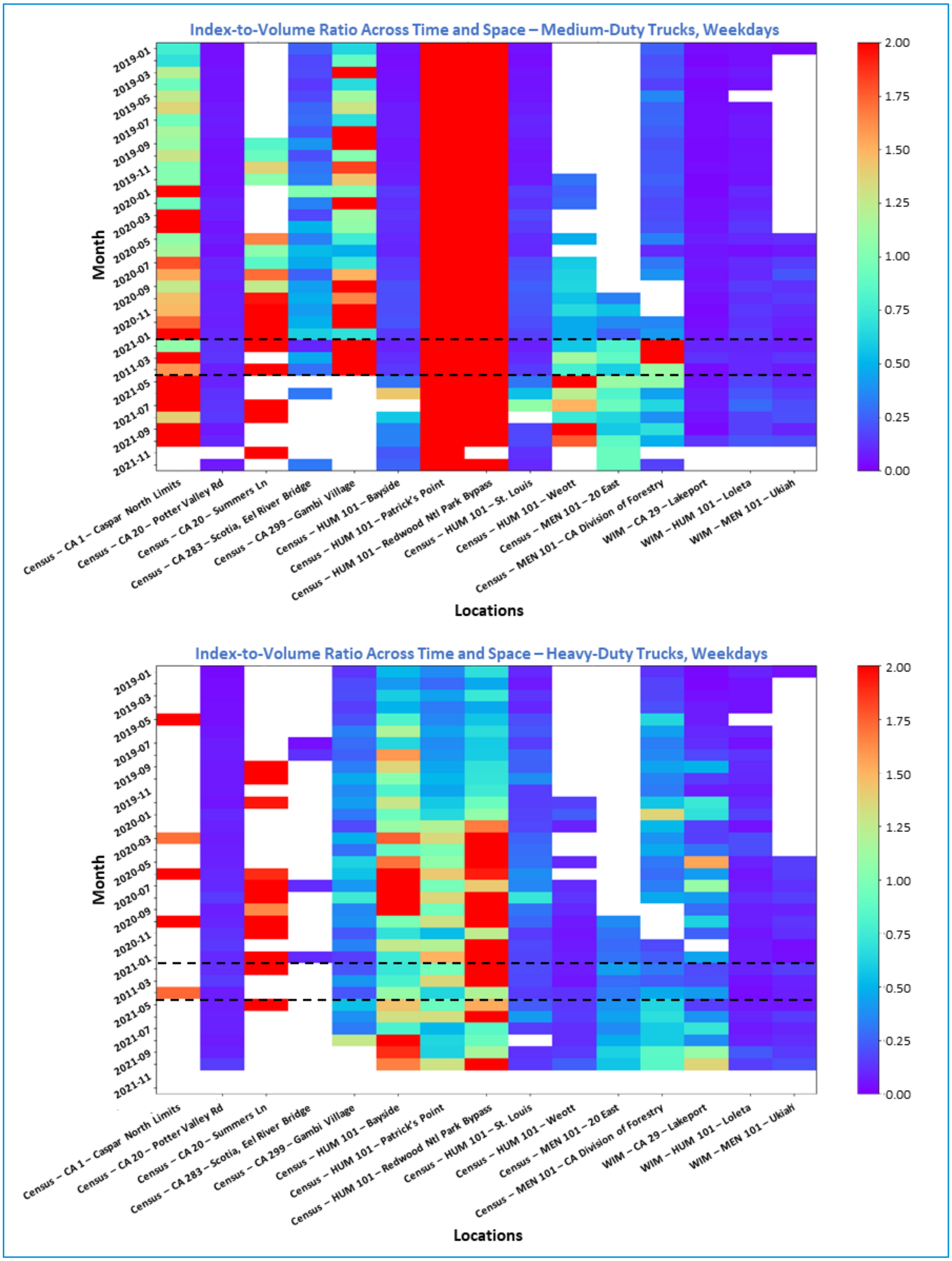


Figure 27. Monthly Index-to-Volume Ratios for WIM and Traffic Census Locations, Weekdays

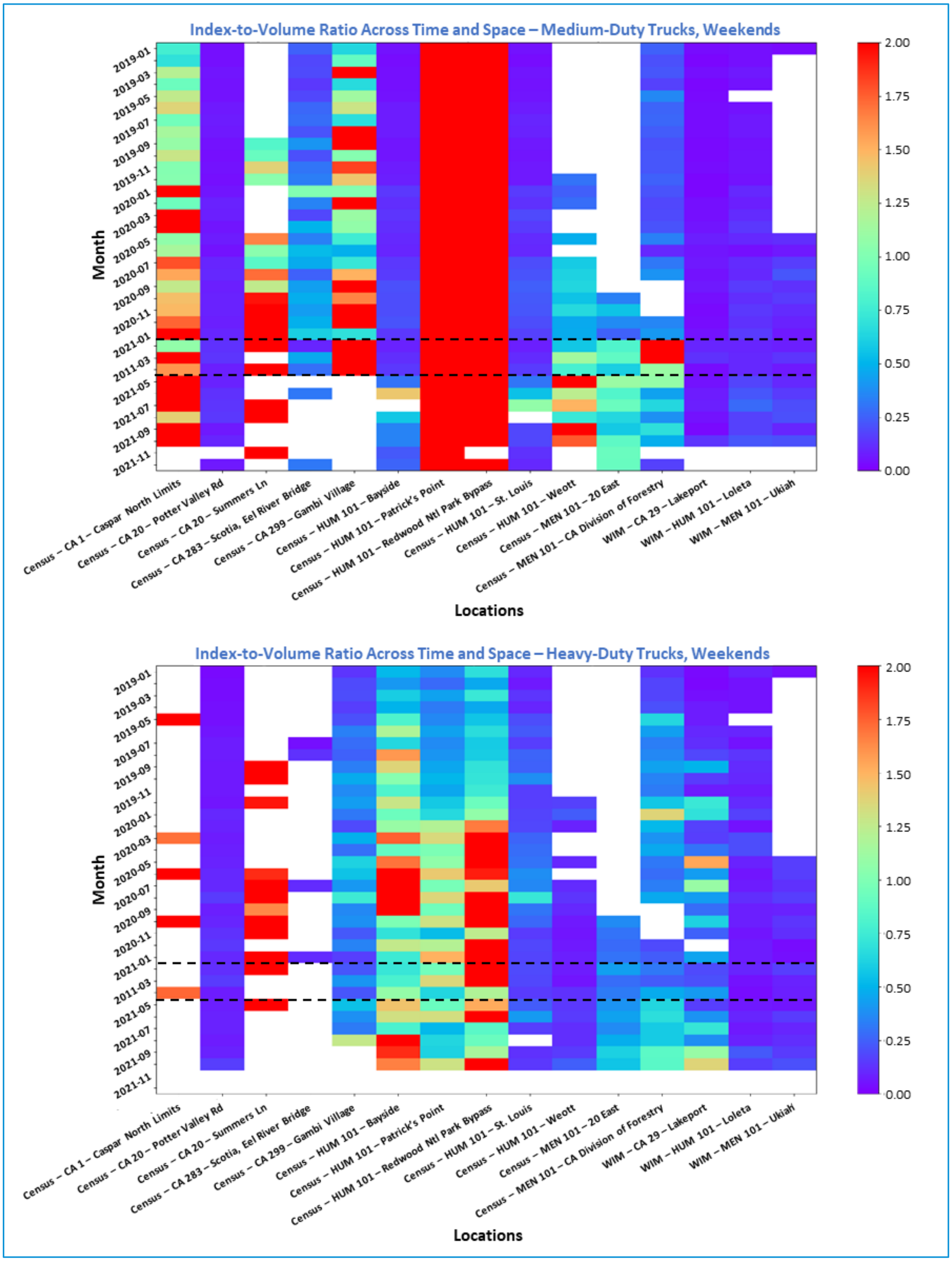


Figure 28. Monthly Index-to-Volume Ratios for WIM and Traffic Census Locations, Weekends

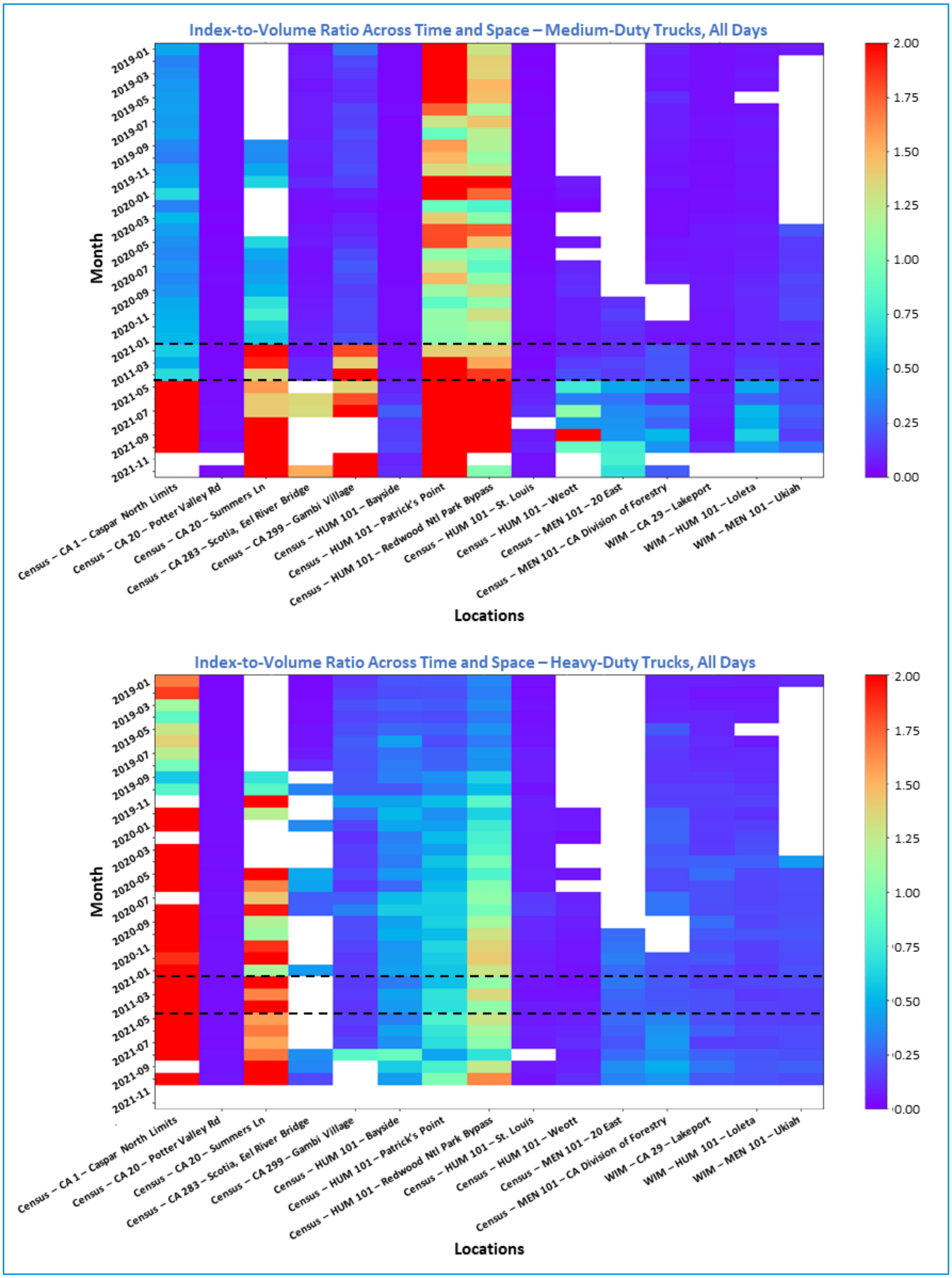


Figure 29. Monthly Index-to-Volume Ratios for WIM and Traffic Census Locations, All Days

The key aspect to look at in the color-coded diagrams is whether the calculated index-to-volume ratios for a particular station are consistent over time, i.e., whether similar colors are assigned to each station. A range of colors would indicate a variable relationship between the StreetLight Index and the corresponding monthly count. This could be indicative of potential issues with either the field counts or the underlying tracking data used by StreetLight to produce index values.

Given this, we conclude that for weekday ratios for medium-duty trucks:

- Significant variability is observed across the entire analysis period for the Patrick's Point (US-101) and Redwood National Park Bypass (US-101) stations
- Increased variability is observed from February 2021 onward at the Summers Lane (CA-20) and Gambi Village (CA-299) stations
- Increased variability is observed from May 2021 onward at the Caspar (CA-1), Eel River Bridge (CA-283), Weott (US-101), 20 East (US-101), California Division of Forestry (US-101), and Loleta (US-101) stations

For weekday ratios for heavy-duty trucks:

- Only ratios from Caspar (CA-1), Summers Lane (CA-20), and Redwood National Park Bypass (US-101) exhibit significant variability across the entire analysis period

A slight change in the average index-to-volume ratio was observed at the Patrick's Point (US-101) station after May 2021. There is no noticeable change in variability during February 2021 or May 2021, when drop-offs in Sample Trip Counts occurred. For weekend ratios, for medium-duty trucks:

- Significant variability is observed across the entire analysis period for a different set of stations than on weekdays. This set includes the Caspar (CA-1), Summers Lane (CA-20), and Gambi Village (CA-299) stations.
- The ratios from the California Division of Forestry (US-101) station exhibited increased variability from February 2021 onward
- The Bayside (US-101), St. Louis (US-101), and Weott (US-101) stations all exhibited increased variability starting in May 2021

For Weekend ratios for heavy-duty trucks:

- In addition to the Caspar (CA-1), Summers Lane (CA-20), and Redwood National Park Bypass (US-101) stations, the Bayside (US-101), Patrick's Point (US-101), and Lakeport (CA-29) stations show significant variability across the entire analysis period
- Ratios from the Gambi Village (CA-299), St. Louis (US-101), and Division of Forestry (US-101) stations show moderate variability across the analysis period
- There was no noticeable change in variability between February 2021 and May 2021, when drop-offs in Sample Trip Counts occurred

Based on the above observations, we conclude:

- Increased ratio variability appears to start in February or May 2021 at several stations. Referring to Figure 12 to Figure 14, this corresponds to months with significant drop-offs in Sample Trip Counts. This correspondence between the drop-offs in Sample Trip Count and data variability highlights the importance of having sufficiently large data samples to conduct reliable analyses.
- The lack of impact around February and May 2021 for heavy-duty trucks suggests that the drop-offs in the number of underlying sample trips mainly affected data-capturing trips made by medium-duty trucks. This highlights the importance of verifying data quality for each vehicle type considered.
- Stations exhibiting ratio variability across the entire analysis period may have continuous underlying issues, such as unreliable counting. Data collected from these stations should therefore be reviewed for accuracy and be excluded from use in calibration until the cause, or causes, of the variability is ascertained.

Overall, the analyses indicate that data provided by some stations are more stable than others and that changes in the underlying data collection may affect the reliability of the data. This point to a need to carefully select representative locations to the used as calibration reference points.

Impacts of Alternate Sets of Calibration Data – WIM Data Only

This section analyses the effect of using different sets of calibration data to produce Calibrated Index values for the three WIM and 12 Traffic Census stations. We undertook two sets of experiments: one exclusively using WIM data for the calibration, and a second set adding Traffic Census data into the mix. All analyses used January 2021 to June 2021 calibration data to estimate average truck volumes at the reference locations over the same period. This period was selected since it is the one for which data was available at a maximum number of locations.

For the experiment exclusively using WIM data for the calibration, three sets of sub-experiments were conducted:

- Using data from all three WIM data sets for predicting traffic volumes at all three stations
- Using data from two stations for predicting traffic volume at the third station
- Using data from one station for predicting traffic volumes at the other two stations

Results from these experiments are shown in Table 4 through Table 6 and illustrated graphically in Figure 30. The accuracy of the truck volume estimates is presented as the ratio of the resulting Calibrated StreetLight Index to the actual volume observed at the location. In this context, ratios closer to 1.0 reflect better predictions, while ratios above 1.0 represent overestimates while ratios below 1.0 represent underestimates.

Table 4. Predicted Volume Ratios, All-Days, All WIM Stations as Calibration Set

Vehicle Type	Loleta	Ukiah	Lakeport
Medium-Duty	0.81	0.86	1.66
Heavy-Duty	1.12	1.04	0.90
All Trucks	*0.92*	*0.92*	*1.45*

Notes: Ratios: Calibrated StreetLight Volume / Observed WIM Count
Calibration/prediction period: January-June 2021

Table 5. Predicted Volume Ratios, All-Days, Two WIM Stations as Calibration Set

Vehicle Type	Loleta	Ukiah	Lakeport
Calibration Data: Lolita and Lakeport			
Medium-Duty	-	0.79	-
Heavy-Duty	-	1.06	-
All Trucks	-	*0.88*	-
Calibration Data: Ukiah and Lakeport			
Medium-Duty	0.72	-	-
Heavy-Duty	1.19	-	-
All Trucks	*0.89*	-	-
Calibration Data: Loleta and Ukiah			
Medium-Duty	-	-	1.99
Heavy-Duty	-	-	0.83
All Trucks	-	-	*1.66*

Notes: Ratios: Calibrated StreetLight Volume / Observed WIM Count
Calibration/prediction period: January-June 2021

Table 6. Predicted Volume Ratios, All-Days, One WIM Station as Calibration Set

Vehicle Type	Loleta	Ukiah	Lakeport
Calibration Data: Loleta			
Medium-Duty	-	1.05	2.04
Heavy-Duty	-	0.92	0.79
All Trucks	-	*1.00*	*1.68*
Calibration Data: Lakeport			
Medium-Duty	0.49	0.52	-
Heavy-Duty	1.29	1.20	-
All Trucks	*0.79*	*0.77*	-
Calibration Data: Ukiah			
Medium-Duty	0.94	-	1.94
Heavy-Duty	1.09	-	0.87
All Trucks	*1.00*	-	*1.63*

Notes: Ratios: Calibrated StreetLight Volume / Observed WIM Count
Calibration/prediction period: January-June 2021

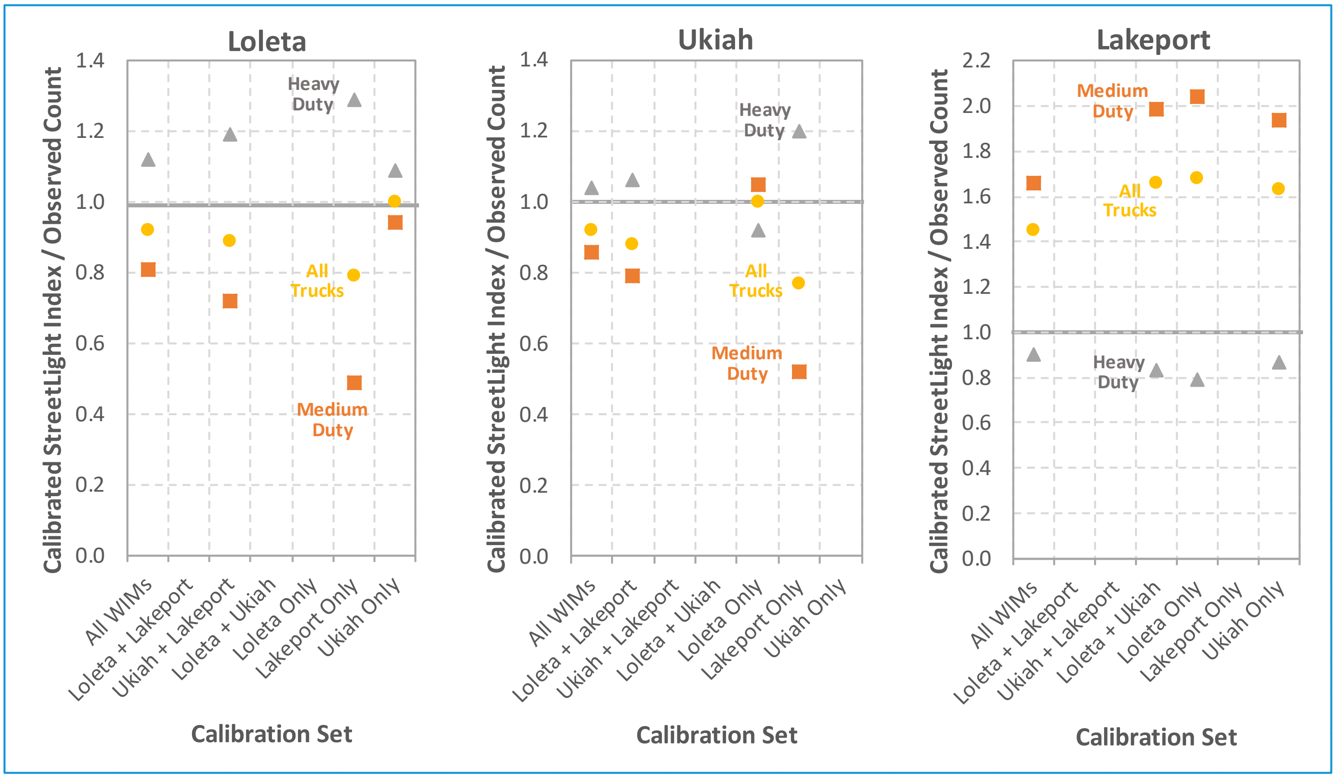


Figure 30. Predicted Volume Ratios, All-Days, Alternate WIM Station Calibration Sets

For all stations, the experiment results indicate that using data from all three WIM stations appears to produce the best ratios for both medium-duty and heavy-duty trucks across all stations, i.e., estimates that align reasonably well with observed counts at all stations. Removing one or two datasets generally results in higher or lower ratios, i.e., in more diverging estimates. However, exceptions can be observed for the Loleta station when using only calibration data from Ukiah, and the Ukiah station when using only data from Loleta. In both cases, the predicted and calibration stations are on US-101. This suggests that a potential geographical connection might exist between the two stations and that considering only data along US-101 may provide better predictions for US-101 stations as data from Lakeport on CA-29 may capture traffic with different characteristics.

Across all scenarios, results from the Lakeport station are generally worse than those from Ukiah and Loleta. This can be attributed to the much lower Sample Trip Counts coming from this station, as indicated in Figure 12 to Figure 14. These smaller sample sizes reduce the quality of the reference data upon which the StreetLight Index is calculated and thus make the estimates less reliable.

The above results were somewhat expected as the calibration data come from the same locations and period over which the estimates are made. As shown in Figure 27 through Figure 29, the index-to-volume ratios associated with the three WIM stations are relatively similar between January and June 2021. As will be demonstrated later, this temporal and spatial consistency plays an important role in StreetLight's ability to produce Calibrated Index values that reasonably represent observed traffic at each location.

Impacts of Alternate Sets of Calibration Data – WIM and Traffic Census Data

To further assess whether improvements could be achieved by adding more calibration points, additional experiments were conducted in which data from Traffic Census stations were added to the WIM data. This resulted in four additional experiments:

- Calibration using data from all WIM and Traffic Census stations
- Calibration using data from the WIM stations and three Traffic Census stations with the highest ADTs. These are the Bayside, St. Louis, and 20 East stations along US-101, which all have ADTs similar to or exceeding those from the WIM stations.
- Calibration using data from WIM stations and nine Traffic Census stations with the lowest ADTs
- Calibration using data from WIM stations and three Traffic Census stations having similar Index-to-volume ratios. These are the Weott, 20 East, and California Division of Forestry stations.

The results of the above experiments are reported in Table 7 and represented graphically in Figure 31.

Table 7. Predicted Volume Ratios, All-Days, Alternate Calibration Sets

Vehicle Type	Loleta	Ukiah	Lakeport
Calibration Data: 3 WIM Stations			
Medium-Duty	0.81	0.86	1.66
Heavy-Duty	1.12	1.04	0.90
All Trucks	*0.92*	*0.92*	*1.45*
Calibration Data: 3 WIM + 12 Traffic Census Stations			
Medium-Duty	1.13	1.21	2.33
Heavy-Duty	1.49	1.39	1.18
All Trucks	*1.26*	*1.27*	*2.00*
Calibration Data: 3 WIM + 3 Traffic Census Stations with Highest ADT			
Medium-Duty	0.96	1.02	1.98
Heavy-Duty	1.63	1.52	1.29
All Trucks	*1.21*	*1.21*	*1.78*
Calibration Data: 3 WIM + 9 Traffic Census Stations with Lowest ADT			
Medium	1.13	1.21	2.34
Heavy	1.02	0.96	0.81
All Trucks	*1.09*	*1.12*	*1.90*
Calibration Data: 3 WIM + 3 Traffic Census Stations with Similar index-to-volume ratios as WIM stations			
Medium	0.80	0.85	1.65
Heavy	1.07	1.00	0.85
All Trucks	*0.90*	*0.91*	*1.42*

Notes: Ratios: Calibrated StreetLight Volume / Observed WIM Count Calibration/prediction period: January-June 2021

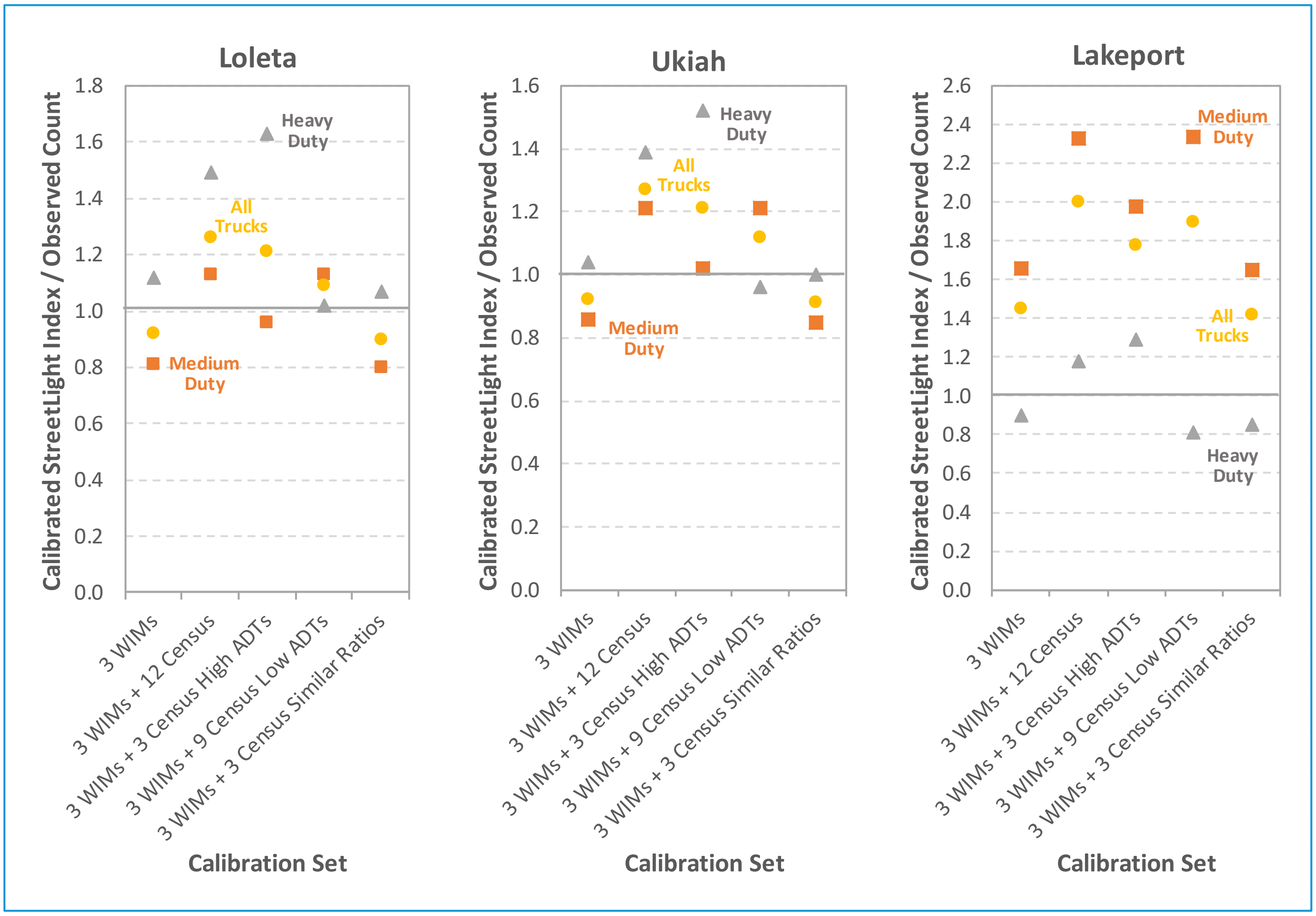


Figure 31. Predicted Volume Ratios, All-Days, Alternate WIM/Traffic Census Calibration Sets

Based on the illustrated data, we conclude:

- Adding data from the 12 Traffic Census stations generally leads to worse ratios when compared to using only WIM data
- Adding data from the Traffic Census stations with the highest ADTs produces the best results for medium-duty trucks at the Loleta and Ukiah stations. The Lakeport ratios are also better but still worse than using only WIM data. However, including this data also generally worsens the heavy-duty truck ratios.
- Adding data from the nine Traffic Census stations with the lowest ADTs produces ratios for medium-duty trucks that are similar to those obtained when using data from all 12 stations and still worse than when using only WIM data. However, the best ratios for heavy-duty trucks are obtained here for the Loleta and Ukiah stations.
- Using data collected between January 2021 and June 2021 from the three US-101 Traffic Census Stations having index-to-volume ratios similar to the WIM stations produces ratios for medium-duty and heavy-duty trucks that are close to those obtained when considering only WIM data

The above experimental results strongly suggest that a correlation exists between the accuracy of the StreetLight predictions and the quality of available calibration data, as expressed by the calculated Index-to-volume shown in Figure 27 through Figure 29:

- In the second experiment, the deteriorating ratios for both medium-duty and heavy-duty trucks are likely the result of adding in the calibration data from several stations having highly variable index-to-volume ratios
- In the third experiment, the Bayside, St. Louis, and 20 East stations have index-to-volume ratios for medium-duty trucks between January and June 2021 that are similar to those from the three WIM stations. Adding data from these stations is thus less likely to negatively affect the estimation process and could even strengthen it, explaining the observed improvements for the Lolita and Ukiah stations over using WIM data only.
- While results from the third experiment show improving ratios for medium-duty trucks, they also show deteriorating ratios for heavy-duty trucks. In this case, the data from the Bayside station, and to some extent the 20 East station, appear to deviate a bit more from the WIM station data than with the medium-duty truck. This lowers the average consistency of calibration across the six calibration stations and could explain the deteriorating estimates. The relatively low Sample Trip Counts associated with heavy trucks could have had an influence.
- In the fourth experiment, the medium-duty truck ratios are affected by the same factors as in the second experiment as all the stations with highly variable index-to-volume ratios are in the group of nine stations considered. However, the improving ratios for heavy-duty trucks may be the result of considering fewer inconsistent data, as only the Caspar, Summers Lane, and Redwood National Park Bypass stations have index-to-volume ratios diverging significantly from the WIM stations.
- Results from the last experiment, which only considers WIM and Traffic Census stations having similar index-to-volume ratios, confirm that consistency of calibration data plays an important role in obtaining reasonable volume estimates as this scenario appears to provide the best overall results.

The best overall results appear to be obtained when using only data from the WIM stations or a set of stations exhibiting similar index-to-volume ratios. Deteriorating estimations are obtained when using a more eclectic mix of stations. This is likely due to the calibration process averaging data from all available stations irrespective of the quality of the associated data. In this context, using data from locations exhibiting similar index-to-volume ratios likely produces a single calibration factor that reasonably matches the adjustment ratios estimated at individual locations. In turn, this leads to relatively reasonable volume estimates at each location. However, if stations exhibiting a wide range of ratios are used, there is then an increasing likelihood that the calculated single calibration factor will differ from the local adjustment ratios. This results in adjustments that will in some cases overshoot what should be done and in other cases undershoot it.

The above results suggest that the best approach for developing adequate calibration sets is to select calibration locations having comparable index-to-volume ratios. Referring to Figure 27 through Figure 29, this means considering stations with similar colors and leads to the following considerations regarding District 1 data:

- Data from Caspar (CA 1), Summers Lane (CA 20), Patrick's Point (US 101), and Redwood National Park Bypass (US 101) should be excluded from any analysis as these stations constantly exhibit calibration factors that are significantly different than from other stations
- Data from El River Bridge (CA 283), and Gambi Village (Ca 299) may also need to be excluded depending on the period considered
- Particular attention should be paid to selecting calibration data from stations with high observed traffic volumes and high StreetLight Sample Trip Counts, as these locations are more likely to contain quality data

Impacts of Analysis Period

This section reports on experiments that were conducted to evaluate the impacts of considering different analysis periods. In each case, the prediction period is set to correspond to the calibration period. The objective here was to see if prediction outcomes would be significantly affected by variations in underlying Sample Trip Counts, i.e., underlying tracking data.

At the heart of this experiment is the fact that truck Sample Trip Counts significantly dropped off twice in 2021, the first time in February and the second time in May. By May 2021, Sample Trip Counts were only a small fraction of what they were in late 2020 and January 2021. Thus, we made volume predictions for the three WIM stations using calibration data from all three WIM and 12 Traffic Census stations for the following three-month periods around the two months when data drop-offs occurred:

- October 2020 to December 2020 (before the drop-offs)
- January 2021 to March 2021 (crossing the February drop-offs)
- April 2021 to June 2021 (crossing the May drop-offs)
- July 2021 to September 2021 (after both drop-offs)

Table 8 provides the volume estimations obtained in the form of the ratio between the Calibrated StreetLight index and the observed counts, with a ratio of 1.0 indicating a perfect match between the two values. Figure 32 graphs the results.

Table 8. Predicted Volume Ratios, All-Days, Different Analysis Periods

Vehicle Type	Loleta	Ukiah	Lakeport
October – December 2020			
Medium-Duty	1.22	0.78	2.01
Heavy-Duty	1.64	1.41	1.31
All Trucks	*1.38*	*0.99*	*1.79*
January – March 2021			
Medium-Duty	1.19	1.13	2.14
Heavy-Duty	1.85	1.55	1.54
All Trucks	*1.43*	*1.29*	*1.98*
April – June 2021			
Medium-Duty	1.10	1.77	3.50
Heavy-Duty	1.55	1.59	1.19
All Trucks	*1.27*	*1.70*	*2.81*
July – September 2021			
Medium-Duty	1.08	1.96	4.17
Heavy-Duty	1.39	1.57	1.21
All Trucks	*1.20*	*1.82*	*3.11*

Notes: Ratios: Calibrated StreetLight Volume / Observed WIM Count Calibration data: All WIM and Traffic Census stations; Prediction period: Same as calibration period

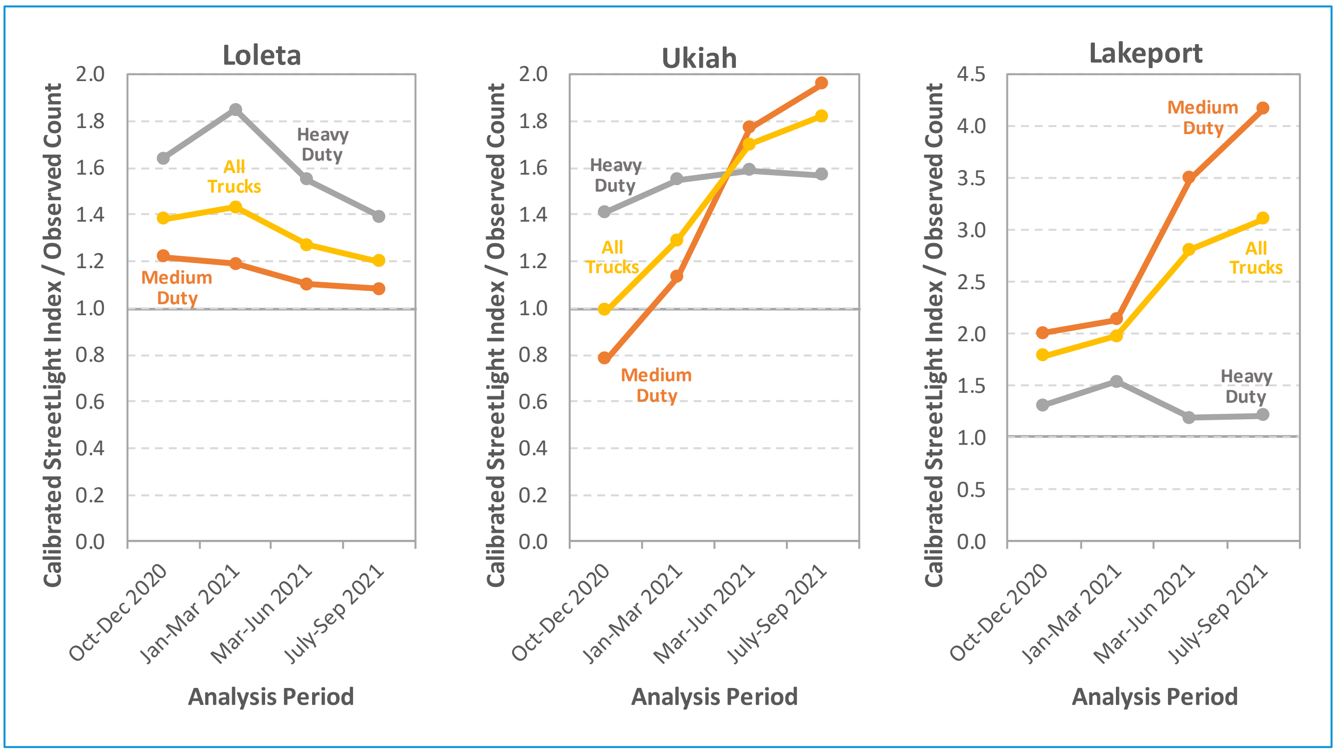


Figure 32. Predicted Volume Ratios, All-Days, Different Analysis Periods

Based on the analysis data, we conclude:

- Estimation ratios for the medium-duty trucks at the Ukiah and Lakeport stations generally worsen over time. This is likely a direct consequence of the reduced Sample Trip Counts producing reference data progressively less representative of the medium-duty truck traffic.
- Medium-duty truck predictions for the Loleta station improve slightly over time. This may be because the drop-offs in Sample Trip Counts in February 2021 and May 2021 resulted in a negligible loss of tracking data from vehicles typically passing by the station.
- Ratios for heavy-duty trucks generally worsen from the October-December 2020 period to the January-March 2021 period, and either stabilize or improve over the subsequent periods. This suggests that the February 2021 data drop-off likely affected the predictions but that the May 2021 drop-off only had a marginal impact, if any. Comparing the Sample Trip Counts from medium-duty and heavy-duty trucks in Figure 12 though Figure 15, this might be because the May 2021 drop-off only resulted in a small loss of tracking data from heavy-duty trucks.

Based on the above analysis, we further recommend the following regarding the selection of analysis periods for District 1 based on the available calibration data:

- While the prediction ratios worsened from the October-December 2020 period to the January-March 2021 period, the differences are somewhat minimal. This is an indication that despite the drop-off in Sample Trip Counts sufficient tracking data may remain to make reasonable predictions. This leads to the conclusion that data from February, March, and April 2021 remain usable as calibration data.
- Due to the significant change that occurs in May 2021 in the calculated index-to-volume ratios caused by the drop off in Sample Trip Counts occurring in that month, as shown in Figure 27 through Figure 29, data from before May 2021 should not be used concurrently with data from or after May 2021 in the analyses. Using data from both periods runs the risk of considering calibration data representing different groups of vehicles or trips with different characteristics.
- Data from May 2021 onward should be used with caution due to the very small Trip Sample Counts associated with this period. These low counts may cause the calculated adjustment factors for each calibration point to fluctuate widely on a month-by-month basis. In turn, these fluctuations might lead to the determination of a single calibration factor that may vary as widely, thus resulting in less reliable estimates.

While the above recommendations are specific to the analysis of District 1 data, the analyses that were conducted here could be replicate with data sets in other districts to assess the use of alternate reference periods may affect the calibration process.

Test Showcases

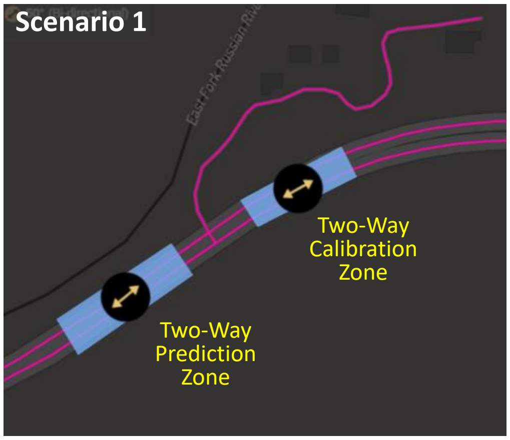
This section presents two showcases that were developed to test the accuracy of the all-vehicles and truck volume estimates produced by StreetLight in relatively simple settings. The first showcase estimates traffic volume along an isolated divided highway segment and the second at an intersection in a residential area. Both locations used for the showcases are actual locations within District 1 and use fictious volumes rounded to the thousand as calibration data to better illustrate the test outcomes.

Both showcases include four analysis scenarios:

- Prediction for a bidirectional (2W) zone using a bidirectional (2W) calibration zone
- Prediction for a bidirectional (2W) zone using two one-way (1W) calibration zones
- Prediction for two one-way (1W) zones using two one-way (1W) calibration zones
- Separate directional predictions for each one-way (1W) zone using the corresponding one-way (1W) calibration zone

Highway Showcase

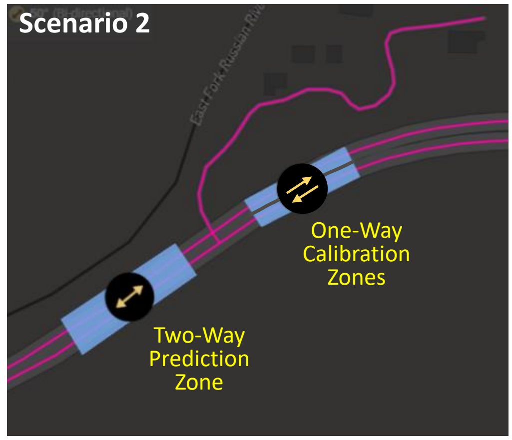
Figure 33 illustrates the four setups for the divided highway showcase. In each scenario, data from the single or dual calibration zones are used in conjunction with data contained in StreetLight's underlying trip databases to predict volumes at a nearby location. In all cases, analyses are conducted using an analysis period extending from January 1 to December 31, 2019.



Truck Percentages
- Medium-duty: 7.5%
- Heavy-duty: 2.5%

Calibration Zone ADT (two-way)
- All vehicles: 10,000

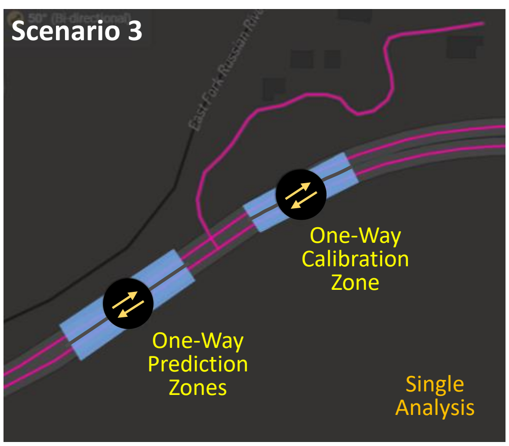
Prediction Zone Calibrated Index (two-way)
- All vehicles: 9,991
- Medium-duty: 751
- Heavy-duty: 250



Truck Percentages
- Medium-duty: 7.5%
- Heavy-duty: 2.5%

Calibration Zone ADT (one-ways)
- All vehicles: 4,000 / 6,000 → 10,000

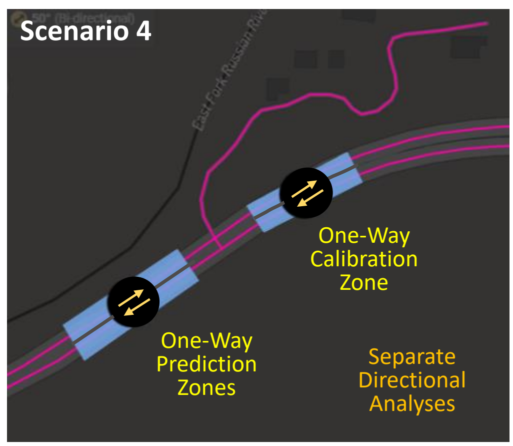
Prediction Zone Calibrated Index (two-way)
- All vehicles: 10,063
- Medium-duty: 749
- Heavy-duty: 249



Truck Percentages
- Medium-duty: 7.5%
- Heavy-duty: 2.5%

Calibration Zone ADTs (one-ways)
- All vehicles: 4,000 / 6,000 → 10,000

Prediction Zone Calibrated Index (one-ways)
- All vehicles: 4,995 / 5,057 → 10,052
- Medium-duty: 361 / 389 → 750
- Heavy-duty: 118 / 130 → 248



Truck Percentages
- Medium-duty: 7.5%
- Heavy-duty: 2.5%

Calibration Zone ADTs (one-ways)
- All vehicles: 4,000 / 6,000 → 10,000

Prediction Zone Calibrated Index (one-ways)
- All vehicles: 3,844 / 6,259 → 10,103
- Medium-duty: 291 / 463 → 754
- Heavy-duty: 97 / 154 → 251

Figure 33. Showcase 1 - Isolated Highway Segment Scenarios

Because of the proximity between the calibration and prediction zones, estimated truck volumes (Calibrated Truck Index) at the prediction zones were expected to be similar to the volumes assigned to the calibration zones. While a local road exists between the calibration and prediction zones, it plays a very minor role in the calculation as this road typically carries very little traffic.

Results from the first two scenarios show bi-directional volume estimates that closely match the expectations. Bidirectional flow estimates of 9,991 and 10,063 vehicles were obtained while a value of 10,000 was anticipated. In both cases, the resulting truck proportions further closely match the ratios of medium-duty and heavy-duty trucks assigned to the calibration zone.

Results from Scenario 3, however, fail to correctly apportion volumes per direction. While a 40/60 split was expected, a 50/50 split was obtained. This is attributed to the averaging process built into the single-factor calibration process, which adjusts values based on the average correction factor obtained from the two one-way calibration zones.

To improve the predictions, a fourth scenario executed separate analyses for each travel direction. The first analysis predicted eastbound volume using only eastbound calibration data, while the second one used westbound data to predict westbound volumes. This divided approach resulted in a much closer 38/62 split.

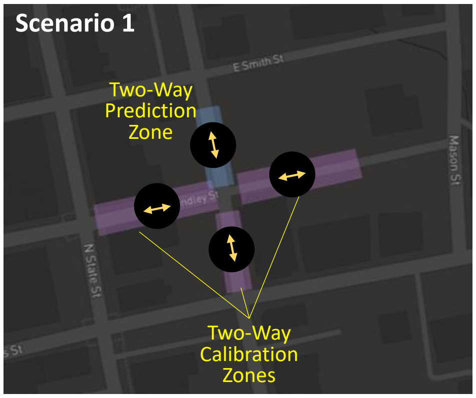
These results indicate that appropriate considerations must be given to selecting suitable sets of calibration zones. In particular, improved predictions are likely to be obtained if relevant sets of calibration zones are used. This means selecting calibration zones that closely match the characteristics of the prediction zone (such as direction of travel) to ensure that data from dissimilar calibration zones does not unduly influence calculating the calibration factor.

Intersection Showcase

In this showcase, three approaches to an intersection are defined as calibration zones while the fourth approach is the prediction zone, as shown in Figure 34. The goal is to predict bidirectional or unidirectional volumes on the fourth approach using the user-provided volumes from the three calibration zones and underlying traffic movement information provided by StreetLight over an analysis period extending from January 1 to December 31, 2019. In this case, volumes in the prediction zone were not expected to match the sum of flows in the calibration zones as traffic from each approach can either turn left, go straight, or turn right.

In Scenario 1, we obtained a bi-directional volume of 7,786 for the prediction zone using two-way calibration zones. In Scenario 2, the estimated bi-directional volume on the prediction zone fell to 7,087 when consider the same analysis setup but replacing two-way calibration zones with one-way calibration zones. As with the previous showcase, these results indicate that switching from two-way to one-way calibration zones has potential significant effects on the prediction results due to the averaging process built within StreetLight's single-factor calibration process.

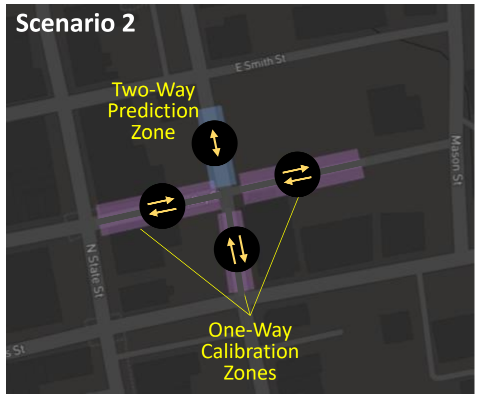
In Scenario 3, predicting volumes on two one-way zones produced single directional volumes of 4,910 and 2,177, for an overall bidirectional volume of 7,087, matching the result of Scenario 2.



Truck Percentages
- Medium-duty: 4.0%
- Heavy-duty: 1.0%

Calibration Zone ADT (two-way)
- All vehicles: 2,000 / 6,000 / 2,000

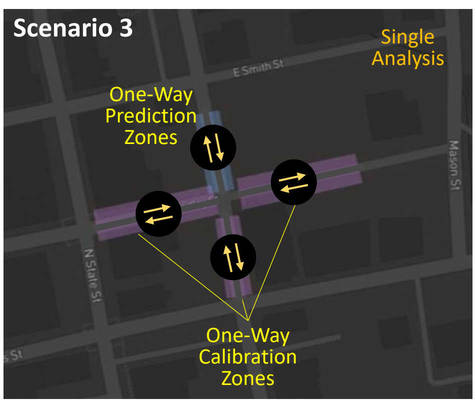
Prediction Zone Calibrated Index (two-way
- All vehicles: 7,786
- Medium-duty: 373
- Heavy-duty: 0



Truck Percentages
- Medium-duty: 4.0%
- Heavy-duty: 1.0%

Calibration Zone ADT (one-ways)
- All vehicles: Out: 1,500 / 4,500 / 1500
  In: 500 / 1,500 / 500

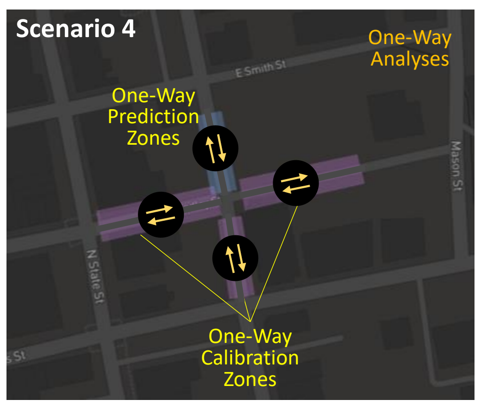
Prediction Zone Calibrated Index (two-way)
- All vehicles: 7,087
- Medium-duty: 362
- Heavy-duty: 0



Truck Percentages
- Medium-duty: 4.0%
- Heavy-duty: 1.0%

Calibration Zone ADTs (one-ways)
- All vehicles: Out: 1,500 / 4,500 / 1500
  In: 500 / 1,500 / 500

Prediction Zone Calibrated Index (one-ways)
- All vehicles: 4,910 / 2,177 → 7,087
- Medium-duty: 288 / 75 → 363
- Heavy-duty: 0 / 0 → 0



Truck Percentages
- Medium-duty: 4.0%
- Heavy-duty: 1.0%

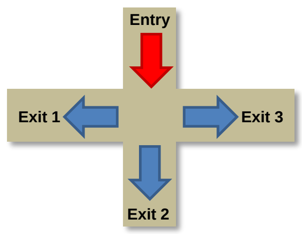
Calibration Zone ADTs (one-ways)
- All vehicles: Out: 1,500 / 4,500 / 1500
  In: 500 / 1,500 / 500

Prediction Zone Calibrated Index (one-ways)
- All vehicles: 114,004 / 1,403 → 115,507
- Medium-duty: 456 / 46 → 502
- Heavy-duty: 0 / 0 → 0

Figure 34. Showcase 2 - Intersection in Residential Area

In Scenario 4, however, conducting separate analyses for each direction of travel resulted in a very large volume estimate to be produced for one direction. This can again be attributed to the averaging principle used in the single-factor calibration process. This is best explained by an example. Consider Figure 35, which replicates the setup of Scenario 4 for predicting the volume entering the intersection. The diagram on the top indicates the calibration volume provided for the three calibration zones considered, in this case, the intersection exits. The Sample Trip Count, StreetLight Index, and Calibrated Index returned by StreetLight for the four legs are also provided as reference data. Two sets of calibration calculations are shown at the bottom. On the left are calculations using data from the three calibration zones, as was done in Scenario 4, while calculations on the right show what happens after having removed from consideration data from Exit 1, the exit with the smallest Sample Trip Count and smallest Index.

	Sample Trip Count	StreetLight Index	Calibrated Index
All-Vehicles	13,715	1,244	114,004
Medium-Duty Trucks	834	292	456
Heavy-Duty Trucks	9	10	--



Exit 1:

Calibration Volume	Sample Trip Count	StreetLight Index	Calibrated Index
1,500	63	6	526
20	--	--	--
15	--	--	--

Exit 3:

Calibration Volume	Sample Trip Count	StreetLight Index	Calibrated Index
1,500	2,017	174	15,933
60	94	33	51
15	2	2	--

Exit 2:

Calibration Volume	Sample Trip Count	StreetLight Index	Calibrated Index
4,500	10,090	906	82,997
480	396	139	216
45	11	13	--

Calibration Using All Exit Segments

Calibration Zone	StreetLight Index	User-Supplied Calibration Volume	Local Calibration Factor
Exit 1	6	1,500	250.00
Exit 2	906	4,500	4.97
Exit 3	174	1,500	8.62
Average of Calibration Factors			87.86

Section	StreetLight Index	Calibration Factor	Calculated Calibrated Index	StreetLight Returned Calibrated Index
Exit 1	6	87.86	527	526
Exit 2	906	87.86	79,603	82,977
Exit 3	174	87.86	15,288	15,933
Entry	1244	87.86	109,300	114,004

Calibration Using Two Exit Segments with Highest Index

Calibration Zone	StreetLight Index	User-Supplied Calibration Volume	Local Calibration Factor
Exit 1	*Not Used*	*Not Used*	*Not Used*
Exit 2	906	4,500	4.97
Exit 3	174	1,500	8.62
Average of Calibration Factors			6.79

Section	StreetLight Index	Calibration Factor	Calculated Calibrated Index	StreetLight Returned Calibrated Index
Exit 1	6	6.79	41	39
Exit 2	906	6.79	6,155	6,157
Exit 3	174	6.79	1,182	1,182
Entry	1244	6.79	8,415	8,457

Figure 35. Showcase 2 – Scenario 4

Calculations using all three calibration zones explain how a very high Calibrated Index can be obtained in some cases. In this case, Exit 1 has a local calibration factor of 250.00, while Exit 2 and Exit 3 have local factors of 4.97 and 8.62 respectively. Averaging the three factors then results in a single calibration factor of 87.86. Applying this factor to the Index values associated with each exit results in a slight underestimate of volumes for Exit 1 but in large overestimates of volumes for Exit 2 and Exit 3. This is the product of the oversized influence that the data from Exit 1 have on the determination of the overall calibration factor in this case. While both Exit 1 and Exit 3 have the same calibration volumes, Exit 1 exerts a much larger influence on the calibration due to its significantly lower associated Index.

Ignoring the data from Exit 1 due to its very low associated Index results in a local calibration factor that is significantly higher than those from the two other exits, causing the single calibration factor to drop from 87.87 to 6.79, resulting in volume estimates of 6,157 on Exit 2 and 1,182 on Exit 3. Compared to the calibration volumes of 4,500 and 1,500 assigned to each exit, these predicted volumes are still significantly different but much closer. Differences still exist as the averaging process used for determining the calibration factor still results in overestimating the volume for Exit 2 and underestimating the volume for Exit 3 due to differences between the local factors and the resulting single calibration factor.

In conclusion, careful consideration must be made in the selection of an adequate set of calibration data. Including data from more locations reduces the potential that data from one zone could play an oversized impact in the averaging process. This is in line with StreetLight's recommendation for using between 10 and 20 calibration points. However, selecting zones sharing relatively similar index-to-volume ratios and having similar traffic characteristics to the prediction zones further ensures that the resulting single calibration factor will not differ too much from the individual ratios and would thus not overly skew the estimates.

Another issue illustrated in the example is that volumes cannot be predicted without a corresponding Streetlight Index. This is what occurs for Exit 1 in the example in Figure 35. In this case, the absence of underlying trip samples results in an inability for StreetLight to produce Index values for medium-duty and heavy-duty trucks. Because there are no Index values, no Calibrated Index can then be returned. This means it is important to always check whether there is underlying data within StreetLight, i.e., a Sample Trip Count or StreetLight Index greater than 0, to perform specific analyses.

However, a Calibrated Index may also not be produced when attempting to adjust very low Index values. This is what occurs for the heavy-duty truck adjustments on Exit 3. Unfortunately, the reason for this is not known. it might be due to the use of filters preventing the display of results with too low Index values.

Summary observations

The various analyses presented in this section indicate that StreetLight can be used to obtain rough estimates of truck movements within Caltrans District 1 or other districts if adequate calibration data are provided. In the best cases, StreetLight produced truck-related Calibrated Index values for District 1 WIM stations corresponding to between 80 percent and 120 percent of observed volumes. However, in the worst cases, the use of inadequate calibration data led to Calibrated Index values underestimating actual volumes by 50 percent or overestimating them by up to 400 percent.

The best predictions were obtained when using calibration data from the three WIM stations only or the three WIM and three Traffic Census stations sharing similar index-to-volume ratios. Adding data from other Traffic Census stations did not guarantee better estimates. This is potentially due to one or more of the following factors.

- **Inconsistent data quality across calibration stations.** Using calibration data from stable data sources is important. In this study, the quality of data obtained from WIM and Traffic Census stations was assessed by comparing for each station the ratio between the uncalibrated Index produced by StreetLight for medium-duty and heavy-duty trucks and the vehicle-specific counts provided by it. Locations with stable ratios were deemed more desirable, as this was viewed as an indication that the sample trips contained within the StreetLight database reflected observed trips consistently. High ratio variability, as well as sudden changes in average value, were used as criteria for excluding specific locations or periods from consideration.
- **Inconsistent truck data definitions.** The need to convert classification counts from the 15 axle-based classes normally used by Caltrans into three simple classes (all vehicles, medium-duty trucks, and heavy-duty trucks) introduces some errors. While weight data supplied by WIM stations easily allows dividing trucks into vehicles weighing below 14,000 lbs., between 14,000 and 26,000 lbs., and above 26,0000 lbs., Traffic Census counts do not provide information on vehicle weights so a method must thus be devised to apportion the trucks into the two types considered byStreetLight.
- **Limited Sample Trip Counts.** Less reliable predictions were generally obtained when conducting analyses for zone sets or periods with very low Sample Trip Counts. In such cases, the small sample sizes may not be sufficient to adequately capture the average characteristics of the trips that are normally observed at a given location. For analyses covering District 1, this means that caution should be exercised when conducting analyses for periods after January 2021, and particularly after May 2021, as the low Sample Trip Counts over this period produce less representative data and may thus lead to less reliable analyses. Caution should be exercised until Sample Trip Counts go back up.
- **Single-factor calibration process.** The calibration process treats data from each location equally. This is due to the simple averaging of adjustment factors associated with each calibration location. This process thus ignores that calibration data from sources with higher underlying Sample Trip Counts could be more reliable, that geographic factors may result in differences between calibration and prediction zones, or that it may be desirable to weight calibration data based on their known potential influence.

To reduce negative effects associated with averaging data, StreetLight suggests conducting separate calibrations for each direction of travel along a given roadway or across a zone. For each direction, separate sets of calibration zones would then be provided to best inform predictions along each direction of travel.

Despite the production of a Calibrated Index that can significantly differ from observed traffic counts, there are still advantages to using the StreetLight platform to conduct truck movement analyses. While counts obtained from fixed locations can be used to assess truck volumes, they do not provide information about where trucks are originating from or where they are going, nor about the route taken between the few existing observation points. In this context, estimating truck movements largely depends on general knowledge about their movements within the area of interest. While StreetLight may only rely on relatively small samples of tracking data, such samples reflect actual trucks and can be used to inform truck movement predictions. While the resulting predictions may not fully match actual truck movements, the ability to rely on a sample of observed trips may be sufficient to nudge an analysis towards more reasonable estimates in a way that may be hard for a human analyst to replicate.

Potential Calibration Process Improvements

This section explores two possible improvements to the calibration process used by StreetLight to produce Calibrated StreetLight Index values that more closely match observed trips:

- Weighted single-factor calibration
- Calibration outside the StreetLight platform

Weighted Single-Factor Calibration

As mentioned in the section detailing StreetLight's calibration process, StreetLight derives a uniform adjustment factor meant to be applied to all locations by simply taking the average of all the local adjustment factors associated with each calibration location. This approach treats data equally from all calibration zones. This process does not consider that data from certain calibration zones might be more reliable or more representative of overall truck movements than others, or that geographical factors may influence the relationship between calibration and prediction zones.

One suggestion to address this issue is to introduce weights into the single-factor determination process to reflect the fact that not all calibration data are necessarily equal. The example below illustrates how such weights could be used to allow the single-factor determination process to consider the distance between a prediction zone and a set of calibration zones. The basic principle highlighted here is that data from calibration zones closest to the prediction zone should have a greater impact on the accuracy of the prediction than data from calibration zones located further away.

Table 9. Geographic-Weighted Calibration Factor Calculation Example

Calibration Zone	StreetLight Index	Calibration Volume	Calibration Factor	Euclidean Distance with the Prediction Zone	Weighted Calibration Factor
Calibration Zone 1	1000	10,000	**10.00**	**2000**	**1.67**
Calibration Zone 2	180	1,000	**5.56**	**1000**	**3.71**
Calibration Zone 3	40	500	**12.50**	**2000**	**2.08**
				Sum of Calibration Factors	**7.46**

Replicating the example of Table 2, Table 9 indicates what would happen if the distance between the calibration and prediction zones is taken into consideration. For simplicity, Euclidean distance is used. However, a more accurate calculation method, such as using the distance along actual roads, could be applied if desired. In this case, the weighted calibration factor for each calibration zone is calculated based on the squared distance between it and the prediction zone, as shown below for calibration zone 1:

$$Weighted\ Factor = Factor_1 * \frac{\frac{1}{Dist_1{}^2}}{\frac{1}{Dist_1{}^2} + \frac{1}{Dist_2{}^2} + \frac{1}{Dist_3{}^2}}$$

$$= 10.00 * \frac{\frac{1}{2000^2}}{\frac{1}{2000^2} + \frac{1}{1000^2} + \frac{1}{2000^2}}$$

$$= 1.67$$

The overall adjustment factor is then taken as the sum of the individual factors instead of their average.

As can be observed, this calculation method would have produced an overall adjustment factor of 7.46 instead of 9.35. In the analytical context, this value reflects the greater importance given to the 5.56 factor associated with zone 2, the closest to the prediction zone, than to the 10.00 and 12.50 factors associated with zones 1 and 3. Weighting the data in this way pulls down the value assigned to the single calibration factor from 9.35 to 7.46.

While the above example uses distance as a weighting factor, alternate parameters could be used to reflect the relative importance of various calibration zones. Potential examples include weighting data according to:

- The Sample Trip Count associated with each calibration zone
- The source of data, such as whether data comes from a WIM or Traffic Census station
- The method used to collect data, such as whether the data was collected from a continuous count or a limited count covering only a few days
- User-defined values
- A combination of parameters

Calibration Outside the StreetLight Platform

While the previous section dealt with adding weights to the calibration process, it also implied that instead of letting StreetLight directly calculate the Calibrated Index values the StreetLight Index values would first have to be retrieved from the platform and the subsequent factor calculations and data adjustments would have to be performed outside the StreetLight platform.

Several attempts were made to implement the weighted calibration approach based on Euclidean distances, however, these attempts were met with various difficulties in obtaining reliably predicted volume ratios, i.e., Calibrated Index Values that would be not too far from observed volumes. Results from the outside calibration calculations suggest that additional research is still needed to determine a suitable calibration approach, such as exploring what would be appropriate weights factoring distance, Sample Trip Counts, or other desired calibration parameters.

Recommendations

Based on the various analyses conducted, this section provides general recommendations regarding the use of StreetLight to analyze truck movements for the three following aspects:

- Whether probe data analytical platforms can be used to conduct reasonable truck-related analyses within a region
- Methodology for setting up the StreetLight platform to enhance the reliability of outputted metrics
- Recommendations for future development of the StreetLight platform

Use of Probe Data Analytical Platforms

Probe data analytical platforms offered by third-party data providers can be used to obtain reasonable rough estimates of truck movements along various roadway segments or zones. When adequate calibration data was provided, StreetLight was for instance able to produce Calibrated Index values corresponding to between 80 percent and 120 percent of observed volumes, i.e., to estimate volumes with errors of 20 percent or less.

Probe data analytics platforms can help analyze truck movements:

- By accurately predicting the number of truck trips based on sampled truck counts, such platforms can be a valuable tool to help understand the routes normally followed by trucks from one point to another
- By providing a database of continuously collected sample trips, such platforms help to understand how truckers may react to incidents affecting roadway capacity, such as road closures for maintenance or emergencies such as a forest fire
- By obtaining estimates of truck volumes on road segments located between fixed sensors or in areas without coverage. While these estimates may carry some errors, their ability to provide estimates from a sample of observed trips may provide a more reliable, and quicker, way of obtaining such information than attempting to do so manually.

The ability of probe data analytical platforms to perform reliable analyses of truck movements further depends on the:

- Availability of sufficient underlying tracking data. This determines the ability to reliably represent average vehicle movements within an area of interest. Within StreetLight, such availability can be verified by ensuring that the Sample Trip Counts associated with a given analysis cover several hundred trips at a minimum, and ideally several thousand trips.
- Ability to obtain stable calibration counts from a sufficient number of fixed stations across the area of interest over the desired analysis period. This ensures that the adjustments made to convert Index values into Calibrated Index values adequately reflect observed traffic characteristics within the analysis area and period.

Potential limitations that may be associated with existing probe data analytical platforms include:

- Analyses focusing on light-duty trucks may not be possible as this type of vehicle is not typically tracked
- Trip analyses can only be produced for zones or roadway segments for which sample trips exist within the underlying tracking data
- The current simple categorization of trucks into medium-duty and heavy-duty vehicles based on their registered gross vehicle weight does not correspond to the axle-based classification scheme used by Caltrans
- Simply categorizing trucks as either medium-duty or heavy-duty does not provide enough detail to determine average axle loads with reasonable accuracy or to conduct reliable local factor analyses for roadway segments
- Underlying truck tracking data are likely to be biased toward characterizing truck movements belonging to firms operating large fleets as these are more likely to be tracked by a fleet management system. The data may significantly underrepresent movements by trucks operated by an owner-operator or by firms owning a single or a handful of vehicles.

StreetLight Calibration Setup Methodology

To maximize the likelihood of obtaining reasonable truck volume estimates, we recommend the following approach for using the StreetLight platform

- **Calibration data preparation:**
    - Identify non-congested locations with available classification counts
        - Weigh-in-motion data (ideal reference locations)
        - Traffic census stations
        - Others
    - For each identified reference location
        - Collect available classification counts.
        - Obtain an estimate of the average daily traffic volume for all vehicles for each direction of travel
        - Identify a method for apportioning truck classification counts into medium-duty (14,000-26,000 lbs. vehicles) and heavy-duty trucks (vehicles over 26,000 lbs.)
            - WIM stations: Use recorded weights
            - Traffic census counts: Use vehicle classification or another suitable method
        - Estimate the proportion of passenger cars, medium-duty, and heavy-duty trucks within each traffic direction

- **Define general calibration zone set**
    - For each location with available calibration data
        - Code in StreetLight a separate calibration zone for each direction, checking in each case the Calibration Data option
        - For each coded zone, enter the corresponding estimated average daily traffic volume and vehicle type percentages within the calibration data section
        - Save the calibration zone set
- **Refinement of calibration zone based on data quality**
    - For each coded calibration zone:
        - For each month within the analysis period:
            - Retrieve from StreetLight the monthly *Sample Trip Count* for the all-vehicles, medium-duty truck, and heavy-duty truck vehicle categories
            - For each vehicle type, divide the *observed average daily traffic volume* by the *Sample Trip Count* to obtain a local *Index-to-volume ratio*
        - For each vehicle type, tabulate the calculated monthly *Index-to-volume ratios* to obtain a time series across the analysis period
        - Remove from the set of potential calibration zones any zone for which the *Index-to-volume ratio* shows significant variability across the analysis period
    - Save the updated calibration zone set
- **Selection of calibration set for the desired analysis**
    - For the desired analysis, select between 6 and 20 (ideally between 10 and 20) representative calibration zones based on the following criteria:
        - Relatively consistent *Index-to-volume ratio* across all zones for all vehicle types considered
        - Zones covering similar types of road segments to those considered in the analysis
        - Zones without highly congested traffic
    - Save the resulting calibration zone set as a new set
- **Perform analysis**
    - In the *Basic Info* tab, define the analysis using the following element:
        - Select *Truck* as the mode of travel
        - Select *Single Factor Calibrated Index using user Counts (Truck Trips)* as the output metric

        - Select the *Add Calibration Zone Set* option, then select the zone set obtained at the end of the previous step as the calibration set.
    - In the *Time Periods* tab, select the desired period to analyze.
    - In the *Zones* tab, select the desired set of zones or segments for which truck trip estimates are to be obtained
    - Run analysis

We further recommend that separate calibration and analyses be conducted for each direction of travel. For instance, to obtain specific Calibrated Index values for the northbound and southbound traffic along a given roadway segment, separate calibrations should be set up for each direction. This means first preparing a set of calibration data to reflect northbound traffic and then another set to reflect southbound traffic. Each set would then be used in separate directional analyses.

While the above-recommended approach was developed for the StreetLight platform, similar considerations could be made for other analytical platforms.

Potential Future Developments

Future improvements for the StreetLight platform, and possibly other platforms, would be to weigh calibration data based on their quality or other evaluation metrics. Within the version of StreetLight that was used, all user-supplied calibration data were weighted equally. This requires users to carefully select the data provided. An alternate approach could be to incorporate the ability for users to attach weights to the data provided to reflect their relative importance. This would open the possibility for users to provide larger weights to data associated with zones or segments with higher underlying sample trip counts, zones closest to the analysis area, or data known to be more reliable. A particular advantage would be the potential to include more data points in areas with limited sources of calibration data.

References

Bayen, A. M., P. E. Sharafsaleh and A. D. Patire. 2013. *Hybrid Traffic Data Collection Roadmap: Objectives and Methods*. PATH Research Report UCB-ITS-PRR-2013-2, University of California, Berkeley, October 2013.

Caltrans. 2002a. Traffic Census Program. https://dot.ca.gov/ programs/traffic-operations/census. Accessed May 5, 2022.

Caltrans. Truck Volume AADT. 2002b. California Open Data Portal. https://data.ca.gov/dataset/truck-volumes-aadt. Accessed August 5, 2022.

Khan, S. M., N. Fournier, M. Mauch, A. Patire and A. Skbaradonis. 2021. *Hybrid Data Implementation*. PATH Research Report, University of California, Berkeley, February 2021.

StreetLight. 2020. StreetLight InSight Metrics: Our Methodology and Data Sources, StreetLight Data, July 2020.

StreetLight. 2021. StreetLight Volume Methodology and Validation White Paper, United States + Canada, StreetLight Data, July 2021.

Appendix A - Existing Truck Data Sources

The traditional truck data sources used to describe truck traffic within Caltrans District 1 include:

- Data from weigh-in-motion (WIM) stations operated by Caltrans
- Data from commercial vehicle enforcement (CVEF) facilities operated by the CHP
- Truck data captured by the Performance Measurement System (PeMS)
- Classification counts from Traffic Census stations
- Annual Average Daily Traffic (AADT) truck statistics from the Caltrans Traffic Census Program
- National Performance Management Research Data Set (NPMRDS)
- Data from the California Statewide Freight Forecasting Model (CSFFM)

Data from Weigh-in-Motion (WIM) Stations

Caltrans maintains three WIM sites in District 1. The specific locations are shown in Figure 36 and include two stations along US-101, one near Lolita south of Eureka and the other south of Ukiah, and an additional station on CA-29 near Lakeport. These three stations are only used for data collection, not inspection or enforcement.

District 1 WIM Stations

ID	County	Route	Postmile	Location
20	Humboldt	101	65.6	Loleta
86	Mendocino	101	21.9	Ukiah
93	Lake	29	44.4	Lakeport

Figure 36. District 1 Weigh-in-Motion Stations

Figure 37 presents a photo of the Lolita station on US-101. All mainline WIM sensors used by Caltrans are bending plates on frames embedded in concrete. As a vehicle travels over the plate, the weight associated with each axle is determined based on the degree the plate is bent. This can be done while the vehicle is traveling at normal traffic speed. Inductive loops are also installed before and after the WIM sensor array to measure vehicle speed and overall vehicle length.

Figure 37. Weigh-in-Motion Station at Loleta on US-101

WIM stations typically gather and store data continuously in the roadside cabinet. Information captured at the WIM stations is then automatically sent to a data management system hosted on a Caltrans server that allows data to be queried based on location or date. Information typically collected from each passing truck by WIM stations includes:

- Axle spacing
- Axle weights
- Gross vehicle weight
- Caltrans vehicle classification
- Vehicle speed
- Vehicle overall length
- Weight violation flag
- Day/time of observation
- Direction of travel
- Lane of travel

Figure 38 illustrates the vehicle classification system used by Caltrans to characterize traffic. These classes are similar to those used by the Federal Highway Administration (FHWA), except for Classes 14 and 15. Classes 1, 2, and 3 include motorcycles, passenger cars, and pickup trucks, while Class 4 includes buses. Trucks are included in Classes 5 to 14. Single-unit trucks with two axles are in Class 5, three axles in Class 6, and four or more axles in Class 7. Classes 8, 9, and 10 respectively represent single trailers with 3 or 4 axles, 5 axles, and 6 or more axles. Classes 11 to 14 similarly represent various categories of multi-trailer trucks based on the number of axles. Finally, Class 15 covers vehicles that could not be associated with any of the previous classes.

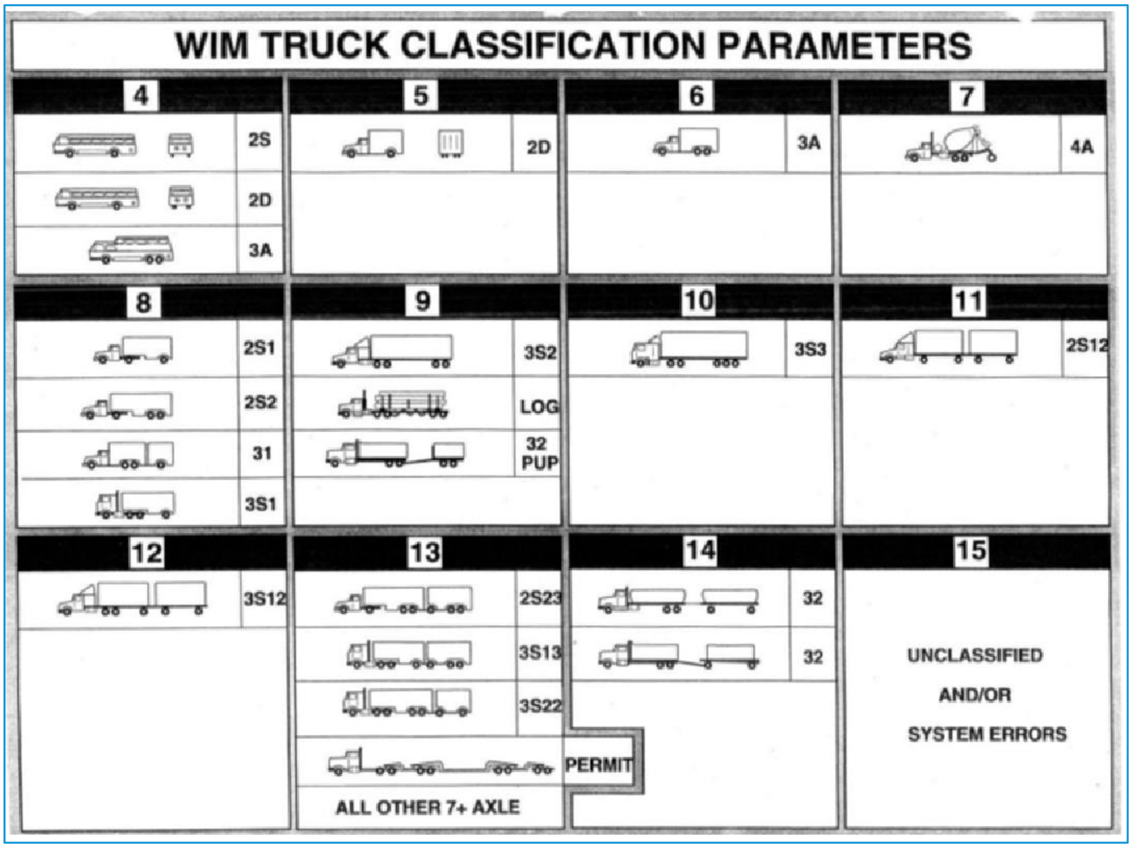


Figure 38. Caltrans Axle-Based Vehicle Classes

Data from Commercial Vehicle Enforcement Facilities (CVEF)

Figure 39 maps the location of the Commercial Vehicle Enforcement Facilities (CVEF) operated by the CHP within District 1 and on roads connected to District 1. These are all Class D facilities, i.e., facilities located at strategic points on major and secondary highway routes for the primary purpose of weighing vehicles. While they all have scales, these facilities may only have a limited open area for the inspection of vehicle equipment and may not be operating all the time. Their operational hours are based on need and are typically determined based on average daily truck traffic, peak truck traffic hours, and seasonal needs.

Based on our discussions with CVEF operators, information about passing trucks is typically only recorded for vehicles flagged with a violation. No data characterizing all passing trucks are therefore available.

District 1 Commercial Vehicle Enforcement Facilities

ID	County	Route	Postmile	Location	Type	Class
36	Humboldt	101 SB	97.2	Little River	Platform	D
37	Humboldt	299 WB	7.4	Buckhorn	Platform	D
39	Mendocino	101 SB	41.2	Ridgewood (Old Willis)	Platform	D

Nearby Commercial Vehicle Enforcement Facilities

ID	County	Route	Postmile	Location	Type	Class
38	Shasta	299 EB	12.7	Wiskeytown	Platform	D
45	Marin	101 SB	15.2	St. Vincents	Platform	D
46	Marin	101 NB	14.1	Terra Linda	Platform	D

District 1 Weight-in-Motion Stations

ID	County	Route	Postmile	Location
20	Humboldt	101	65.6	Loleta
86	Mendocino	101	21.9	Ukiah
93	Lake	29	44.4	Lakeport

Figure 39. CHP Commercial Vehicle Enforcement Facilities in District 1

Truck Data Captured by PeMS

Figure 40 maps the Traffic Census stations within District 1 employing the California Performance Measuring System (PeMS). For most locations, the only data available are directional general traffic volume counts up to December 2018.[1]

[1] We are not certain why there is no data past December 2018 but it is probably due to the processing algorithms used by PeMS to assess data quality which are currently designed to reject all data for any days for which a potential problem is flagged. Data since 2019 may have been rejected simply due to a change in how the data is provided to PeMS or other technical reasons.

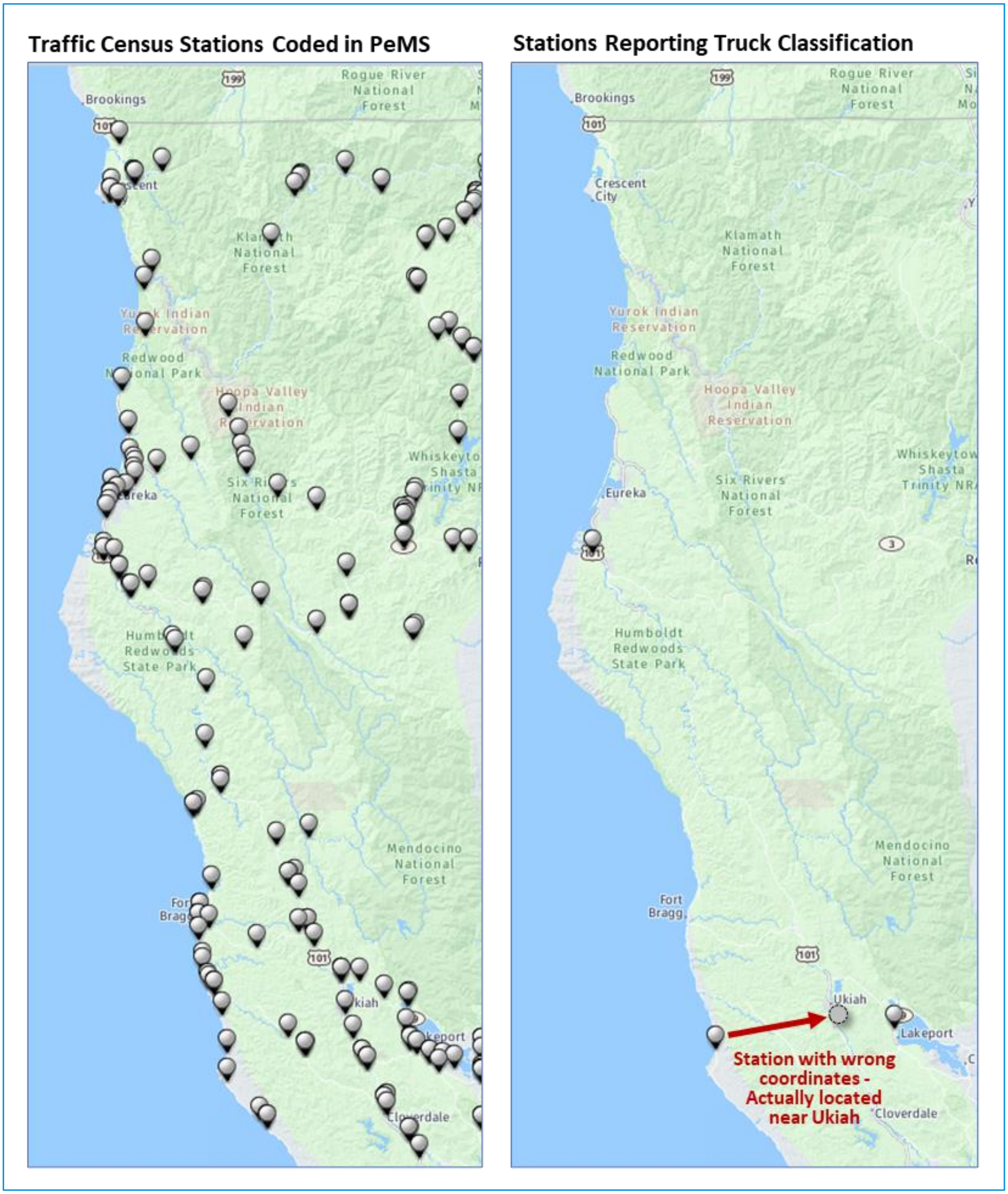
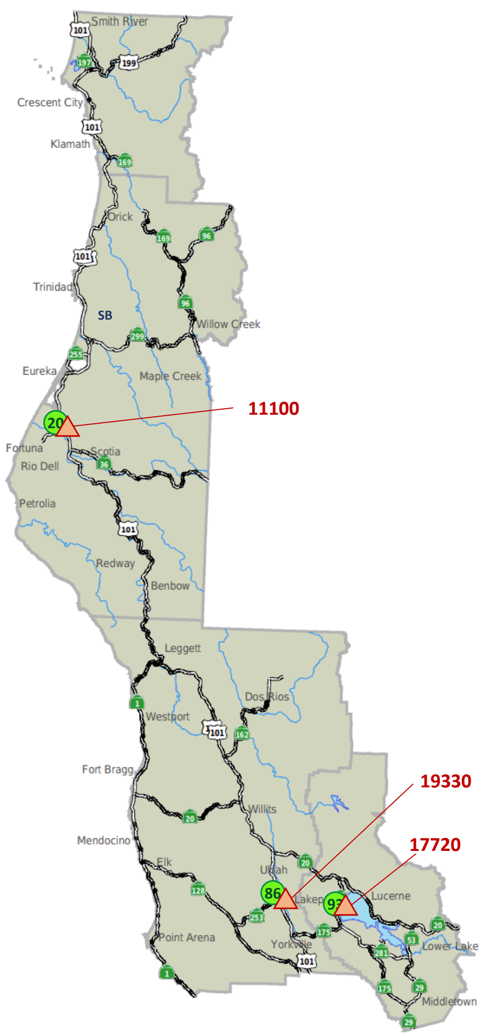


Figure 40. Traffic Census Stations Registered in PeMS

Where data is available, all-vehicle volume counts are typically provided for the following periods:

- Weekday AM peak (6 AM - 10 AM)
- Weekday PM peak (3 PM - 7 PM),
- Weekday midday off-peak (10 AM – 3 PM)
- Weekday evening/night off-peak (7 PM – 6 AM),
- Saturdays
- Sundays

Depending on the location, data may be available every month or only for specific months. This is due to how the Traffic Census program operates. Due to limited resources, data is sampled continuously by moving traffic counting instruments from one location to another to revisit each important location every three years. Only stations with strategic importance have permanent counting instruments.



District 1 Weight-in-Motion Stations

ID	County	Route	Postmile	Location
20	Humboldt	101	65.6	Loleta
86	Mendocino	101	21.9	Ukiah
93	Lake	29	44.4	Lakeport

Matching Traffic Census Stations Registered in PeMS

WIM Server ID	County	Route	Postmile	Absolute Postmile	Location
11100	Humboldt	101	64.29	691.7	Loleta
19330	Mendocino	1	23.45	576.4	Ukiah
17720	Lake	29	44.46	98.9	Lakeport

Figure 41. Correspondence between WIM and PeMS Traffic Census Stations

The only stations currently logging truck-related data into PeMS are illustrated on the right side of Figure 40. As shown in Figure 41, these correspond to the three WIM stations described earlier. Please note that the WIM station on US-101 near Ukiah has been incorrectly mapped into the PeMS server as being located along CA-1. For the Lakeport and Ukiah stations, truck weight, vehicle classification, and general traffic volumes have been available for nearly every month over the past 10 years. On the other hand, the Ukiah station only has recorded data from November 2017, with large monthly gaps.

For the three District 1 WIM stations, data from the following time series were retrieved from PeMS for each class of vehicle:

- Daily truck volumes
- Average vehicle weight
- Average vehicle speed
- Average vehicle length
- Average wheelbase

In addition, retrieved data contained the following information for each vehicle:

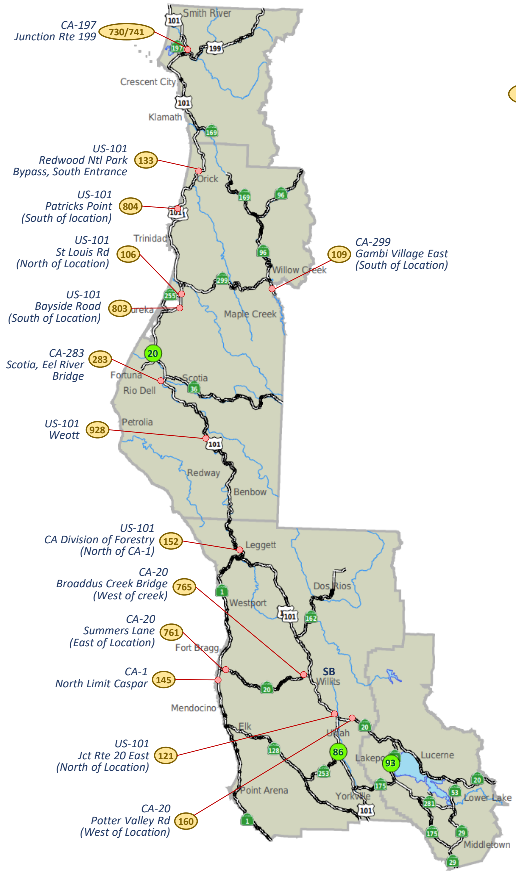
- Direction of travel
- Lane of travel
- Time of day
- Caltrans vehicle class

Classification Counts from Traffic Census Stations

While only WIM stations are currently sending truck data to PeMS, we requested and received data from Caltrans staff on vehicle classification counts conducted as part of the Traffic Census program from the 14 permanent locations mapped in Figure 42.

Figure 43 indicates the months for which classification data was available for each location between December 2018 and January 2022. For five of the 14 stations—stations 106, 109, and 928 in Humboldt County, and stations 145 and 761 in Mendocino County—data were available for every month within the 37-month search period. Stations 283, 803, and 804 in Humboldt County and station 152 in Mendocino County had missing data for only one, two, or three months. Data from the remaining five stations had varying coverage. Station 121 provided continuous data since October 2020, while stations 133 and 160 have generally provided continuous data with small gaps since December 2019 and September 2019, respectively. Stations 765 and 730/740 only provided sporadic data, potentially due to a lack of local data collection efforts following the normal Traffic Census three-year data collection program.

Caltrans staff can typically retrieve fairly comprehensive data from the WIM database. This includes 15-class truck counts by hour, week, day of the month, weight range (10,000 lbs. bins), and direction and lane of travel.



District Ad-Hoc Census Stations

ID	County	Route	Postmile	Leg	Location
106	Humbolt	101	87.48	A	St Louis
109	Humbolt	299	41.86	A	Gambi Village
121	Mendocino	101	31.60	A	Jct Rte 20 East
133	Humbolt	101	125.73	O	Redwood Park, South Entrance
145	Mendocino	1	55.78	O	Caspar, North Limits
152	Mendocino	101	93.60	O	California Division of Forestry
160	Mendocino	20	38.05	B	Potter Valley Road
283	Humbolt	283	0.13	O	Scotia, Eel River Bridge
730/741	Del Norte	197	0	A	Jct 199
761	Mendocino	20	2.23	A	Summers Lane
765	Mendocino	20	31.1	B	Broaddus Creek
803	Humbolt	101	83.9	B	Bayside
804	Humbolt	101	107.67	B	Patricks Point
928	Humbolt	101	33.24	A	Weott

Leg: A = North or east of location, B = South or west of location, O = On location

District 1 Weigh-in-Motion Stations

ID	County	Route	Postmile	Location
20	Humboldt	101	65.6	Loleta
86	Mendocino	101	21.9	Ukiah
93	Lake	29	44.4	Lakeport

Figure 42. District 1 Traffic Census Locations with Vehicle Classification Counts

Location ID	Route	County	Milepost	Month																																						Time Coverage
				Dec-18	Jan-19	Feb-19	Mar-19	Apr-19	May-19	Jun-19	Jul-19	Aug-19	Sep-19	Oct-19	Nov-19	Dec-19	Jan-20	Feb-20	Mar-20	Apr-20	May-20	Jun-20	Jul-20	Aug-20	Sep-20	Oct-20	Nov-20	Dec-20	Jan-21	Feb-21	Mar-21	Apr-21	May-21	Jun-21	Jul-21	Aug-21	Sep-21	Oct-21	Nov-21	Dec-21	Jan-22	
106	US-101	HUM	87.48	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	100%
109	CA-299	HUM	41.86	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	100%
121	US-101	MEN	31.60																							X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	42%
133	US-101	HUM	125.73													X	X	X			X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X				53%
145	CA-1	MEN	55.78	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	100%
152	US-101	MEN	93.60	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X				X	X	X	X	X	X	X	X	X	X	X	X	X	X	92%
160	CA-20	MEN	38.05										X	X	X	X					X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		63%
283	CA-283	HUM	0.13		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	97%
761	CA-20	MEN	2.23	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	100%
765	CA-20	MEN	31.10		X	X	X				X	X																														13%
803	US-101	HUM	83.90		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	97%
804	US-101	HUM	107.67		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		95%
928	US-101	HUM	33.24	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	100%
730/741	CA-197	DN	0.00																								X	X							X	X	X	X	X	X		21%

Figure 43. Traffic Census Program District 1 Classification Counts Time Coverage

Annual Average Daily Traffic (AADT) Truck Statistics from Caltrans Traffic Census Program

Figure 44 maps the District 1 locations for which average annual daily traffic (AADT) truck volumes are produced by the Caltrans Traffic Census Program. These typically reflect the total traffic and truck volumes observed in both directions every 24 hours. This data can be retrieved in Excel form from the Caltrans Traffic Census Program website (Caltrans 2022a) or manually copied from an ArcGIS website maintained by Caltrans (Caltrans 2022b).

An example of data provided for a given site is provided in Figure 45. This includes:

- Position of count relative to the identified location. This data indicates whether a count has been performed before the location of the intersection ("B"), after the intersection ("A"), or directly at the location ("O").
- The total number of vehicles observed in both directions over an average 24-hour period from the Traffic Volume on California State Highways booklet published annually by Caltrans
- The total number of trucks observed in both directions over an average 24-hour period
- Percentage of trucks within the overall traffic
- Number and percentage of trucks with 2, 3, 4, and 5 axles. Two-axle vehicles are 1½-ton trucks with dual rear tires but exclude pickups and vans with only four tires.
- Equivalent axle loading (EAL) of the observed truck traffic
- The latest year for which the truck percentages were either verified (code "V") or estimated (code "E"). Verified data are truck percentages based on counts performed at the site, while estimated data are truck percentages based on counts performed at other locations.

An important consideration regarding the AADT Truck Data is that the volumes presented for each location are not necessarily observed volumes. While data updates are produced every year, data for each location is typically collected on a three-year cycle, with each site normally visited once every three years. A further exception is for sites where volumes are considered static, where no new counts may be made until there is a notable change in traffic on the route. Reported volumes may therefore be actual volumes or estimates based on data from surrounding locations.

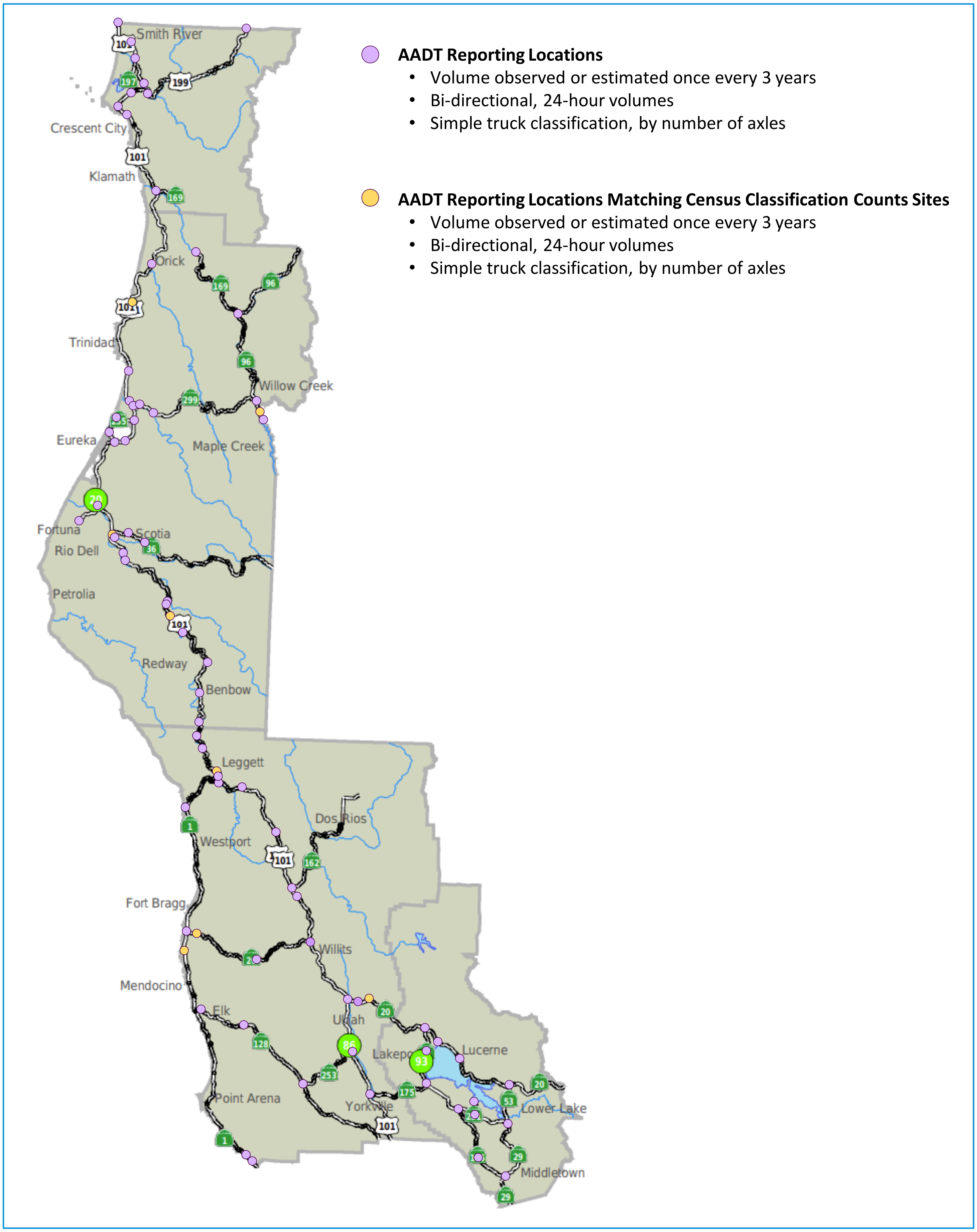


Figure 44. District 1 AADT Truck Reporting Locations

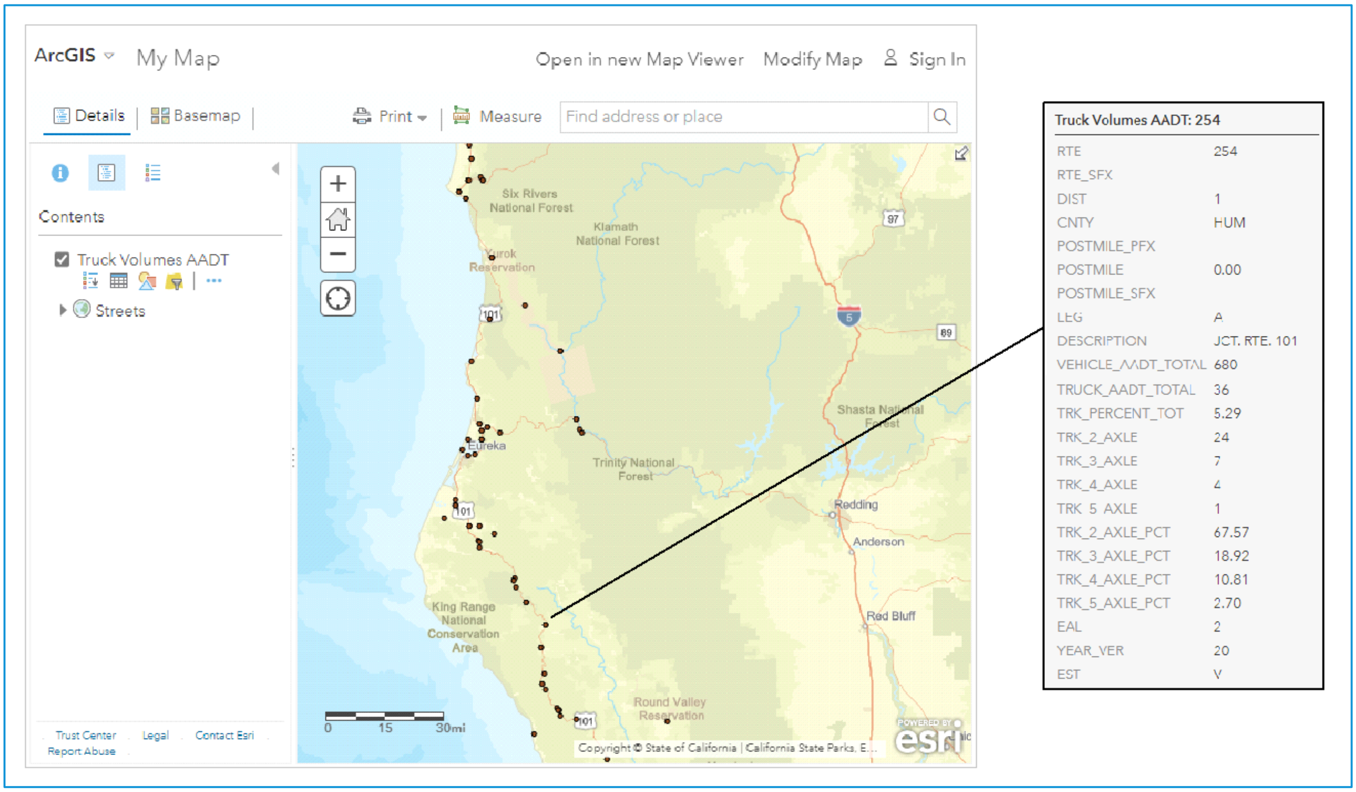


Figure 45. Detailed AADT Truck Data for A Given Location

National Performance Management Research Data Set (NPMRDS)

The National Performance Management Research Data Set (NPMRDS) is an archive of average travel times on the National Highway System collected from probe vehicles maintained by the FHWA. Average travel times are reported for passenger cars, freight trucks, and all vehicles, for each day at 5-minute intervals, without any smoothing, outlier adjustment, or imputation, for segments on the National Highway System and key border crossings with Canada and Mexico. Data are available for each year beginning in 2013. From 2013 to 2016, passenger car data were collected by HERE North America and truck data by the American Transportation Research Institute (ATRI). Since 2017, data have been collected by INRIX.

Access to the data is typically only provided to state Departments of Transportation and Metropolitan Planning Organizations, primarily to help them conduct their federal MAP-21 performance management activities.

An important limitation of this dataset is that it contains no volume data. Only travel time data are provided, as it is primarily intended to help conduct performance reviews and management activities. Since the primary focus of our study is to investigate issues related to estimating truck volumes along the California truck route network, no further efforts were made to investigate the potential uses of this dataset.

Data from the California Statewide Freight Forecasting Model

The California Statewide Freight Forecasting Model (CSFFM) is a travel forecasting model that has been developed on the Cube platform from Citilabs to help Caltrans and partner agencies better understand freight movements within California and their impacts on highway infrastructure, transportation networks, highway safety, energy consumption, and vehicle emissions. The development of this model originated from the need to develop a freight modeling system capable of evaluating the impacts of freight infrastructure enhancements and strategies for lessening traffic congestion, improving mobility and air quality, and reducing fuel consumption and vehicle emissions. As such, it was developed to help investigate impacts on freight movements associated with changes in socio-economic conditions, freight-related land-use policies, environmental policies, and multimodal infrastructure investments.

A key element of this model is its ability to forecast the flow of commodities by mode as a function of employment, establishment, and land-use variables on a model of California's primary road network and its key gateways. Using 2015 as a base year and a network of 97 analysis zones representing county or sub-county areas, regional flow forecasts can be made for 14 commodity groups. The resulting flows are then distributed between rail and truck networks, with the truck flows finally loaded onto a model of the primary road network using traffic assignment processes. One key output is average truck flows on roadway links by truck classes and time of day.

For this study, outputs from the model can be used to fill in gaps in available data to forecast what truck flows may be on roadway segments for which no data is available. However, these outputs must be used with some caution, as they are forecasts based on modeling surrounding conditions for a given base year. It is, therefore, possible that forecasted flows on specific roadway segments may differ from reality to some degree. Some of the limitations typically associated with planning models include:

- Travel surveys used to develop some elements of the model are expensive and time-consuming, and respondents may suffer from recall biases
- Models are computationally expensive as well as technically difficult to manage
- Models are geographically constrained and less accurate nearer to regional boundaries
- Most models are aggregated to a specific geography setting, such as a Traffic Analysis Zone (TAZ), making them unable to capture the nuance of local travel behavior
- Models are updated infrequently, typically once every few years, and often rely on travel surveys that may already be dated
- Models may be based on data from relatively small samples and limited geographic areas, especially in non-urban communities

Appendix B - Review of Key Mobile Data Vendors

This appendix provides a summary of key mobile data sources and mobile traffic data vendors. These include:

- FHWA National Performance Measure Research Data Set (NPMRDS)
- HERE Technologies
- TomTom
- INRIX
- StreetLight
- StreetLytics from Bentley Systems
- Replica

All the commercial vendors listed below depend heavily on mobile data sources for providing empirically based traffic speed and travel time information for roadway segments along with other commercially available roadway performance measures. Many also depend on secondary data sources, such as traffic volume and speed data from roadway sensors maintained by state DOTs, weather data, and/or incident data, to enhance services provided and/or for validation purposes. Even though these "big-data" vendors are all vying for market share, it is common to see data-sharing and collaboration or data-sharing agreements between individual vendors.

The last section of the appendix provides a summary table comparing their the offerings data delivery capabilities from each vendor.

FHWA National Performance Measure Research Data Set (NPMRDS)

As indicated in the Existing Truck Data Sources section, the National Performance Measure Research Data Set (NPMRDS) is a dataset acquired by FHWA for use in measuring transportation system performance. It contains empirically based speed and travel time data, averaged over 5-minute intervals, for passenger cars and commercial freight vehicles for a set of predetermined roadway segments that are part of the U.S. National Highway System and for 25+ key Canadian and Mexican border crossings. It is the default dataset for calculating the new Federal "PM3" system and freight performance measures.

The first NPMRDS dataset, covering 2013 to 2016, collected passenger car data from HERE North America LLC (formerly known as NAVTEQ and later Nokia) and freight truck data from the American Transportation Research Institute (ATRI). Since 2017, data have been provided by INRIX. The NPMRDS is made available free of charge to state Departments of Transportation and Metropolitan Planning Organizations for performance management.

HERE Technologies

HERE Technologies (https://www.here.com/), provides mapping, location data, and related services to individuals and companies. Since 2015, it has been majority-owned by a consortium of German automotive companies and Intel. Key products offered include:

- Automotive Products – Automobile software development kit (SDK) for connected embedded navigation solutions, real-time navigational data and services, anticipatory data and sensor support for advanced driver assistance systems (ADAS) and autonomous driving applications, weather data, locational EV charging station data, locational fuel price data, locational parking availability data, hazard warnings, intelligent sensor data for autonomous driving solutions, and real-time traffic data.
- Location Services Products – Fleet telematics, geocoding (mapping of geo-coordinates and addresses), interactive geo-visualization services, mobile SDK, interactive mapping, places and routing data, services, and products.
- Map Content and Positioning Products – Map data with visual places footprints, driver maneuver assistance (in-vehicle guidance for upcoming exits and lane splits), and smart positioning for mobile devices.
- Traffic Products – Real-time and historical traffic data, traffic analytics, and dashboards.

TomTom

TomTom (https://www.tomtom.com/en_us/) is a multinational developer and creator of location technologies and consumer electronics. The company has been collecting anonymous consumer-driven GPS-based measurements worldwide since 2008 and has used this data to build a historical traffic database.

TomTom's products include applications and products to aid drivers (navigation devices and trip apps and devices), the automotive industry (autonomous driving apps and support, HD maps, and map data for autonomous and traditional vehicles), and fleet management business solutions and products (enabling fleet management, vehicle tracking, fleet optimization, workforce management, green and safe driving, and business integration).

INRIX

INRIX (https://inrix.com) was founded in 2005 and currently collects anonymized data on traffic congestion, traffic incidents, parking, and weather-related road conditions from millions of data points daily in over 80 countries. These data are further combined and aggregated from in-vehicle devices and mobile devices, departments of transportation traffic data, cameras and sensors on roadways, and major events expected to impact traffic.

INRIX provides a variety of products, apps, and solutions for drivers, the automotive and trucking industries, and government agencies and their business partners, including:

- INRIX Drive Time – Real-time assessments of potential commute and travel times
- Roadway Analytics – Data as a service platform and tools to optimize roadway planning, performance monitoring, and the decision-making process
- Trip Trends – Cloud-based platform with access to trip counts, length, and duration data to facilitate the understanding of transportation trends
- Trip Analytics – Origin-destination data
- Corridor Analytics – In-depth analysis of traffic counts, miles traveled, and vehicle types by corridor, time of day, and day of the week
- Road Rules – A complete traffic signal database tool for cities to digitize, manage, and communicate the rules of their roadways, curbs, and sidewalks
- Signal Analytics – Platform for analyzing traffic signal operations based on anonymous probe-vehicle data
- Performance Measures Analytics – Transportation data and intelligence to help public agencies optimize roadway planning and decision-making
- Population Analytics – Combination of GPS and mobile data intelligence to help analyze and understand the movement of people
- AI Traffic – Uses artificial intelligence to provide instantaneous updates to real-time traffic conditions, pinpoint traffic speeds in different lanes, and deliver accurate arrival time estimates
- Volume-Traffic Count Data – Traffic volume dataset with nationwide coverage across 2.65 million miles of road that includes vehicle volume by street direction, time of day, and day of the week (in 15-minute bins by road segment)
- Speed Data – Historic speed and travel time data

StreetLight Data

StreetLight Data (https://www.streetlightdata.com/) was founded in 2011 to deliver empirically based data products on the movement of vehicles, bicycles, and pedestrians. Every month, the company takes in, indexes, and processes over 100 billion anonymized location records from smartphone applications using software provided by Cubeiq and truck data from in-vehicle navigation devices provided by INRIX. Data from various other sources, such as parcel and digital road network data and speed and count data from permanent traffic counters maintained by state departments of transportation, are also processed to help validate and enrich datasets.

StreetLight Data's traffic-related data products offered to private and public agency clients include trip duration and length, trip purpose, origin-destination metrics, and AADT estimates. Depending on data availability, travel modes considered include personal vehicles, medium and heavy freight trucks, public transit buses, rail services, bicycles, and pedestrians.

Bentley Systems' Streetlytics

Bentley Systems (https://www.bentley.com/en) is an infrastructure software development company. Part of its activity portfolio includes developing software for creating and managing roadways, bridges, and airports. In 2022, the company acquired Citilabs, the developer of the popular Cube travel demand forecasting model.

While Citilabs has long been associated with travel demand modeling software, services, and solutions, the company has in recent years expanded to include big-data transportation data analytics for private and government agency clients through their Streetlytics platform. This platform can provide historical traffic volumes and speeds on nearly all public roadways in California. The platform pulls data and information from billions of data points from GPS devices, cellular phones, connected car devices, Bluetooth devices, and ticketing systems, as well as from demographics and other empirically proven or "ground truth" data, to produce traffic-related utilization and performance metrics on public roadways. To accomplish this, Streetlytics employs a proprietary optimization process, which combines data of multiple types from multiple sources:

- Sampled location data from the movements of smartphones and vehicles
- Data on population movements based on travel behavior models applied to current household and employment data
- Ground truth measurement from a database of current traffic counts

The Streetlytics platform's key features and services include:

- Directional speed and volume data for roadways (including minor arterials and collector streets); hour-by-hour data by weekday type and month of the year
- Trip purpose and mode of travel data
- Route or itinerary data—routes used to travel between origins and destinations
- Home location and demographic characteristics of travelers

Replica

Replica (https://replicahq.com/) started in 2018 as a project within Alphabet's Sidewalk Labs seeking to understand how people move within cities. Replica then evolved into an independent company that developed a large-scale activity-based travel demand model that grew to cover the entire United States in 2021. Trip activities within California are included in a model that covers California and Nevada. This model is used in conjunction with various data feeds to perform travel analyses and generate synthetic trip data based on an underlying layer or observed data.

Key third-party data sources leveraged by the activity model include:

- Mobile location data from smartphone applications, cellular phone carriers, in-dash GPS-based car navigation devices, and commercial fleet management software
- Consumer and residential data
- Land use and real estate data
- Credit transaction data
- Ground truth data, such as traffic counts from PeMS, bicycle and pedestrian counts, transit boarding data, and household survey data

As of March 2022, Replica data for California is available only for September-November 2019.

Replica's primary source of freight data is INRIX. To adjust for varying regional coverage, a scaling factor based on observed estimates of total truck volumes to customer-sourced or internally generated ground truth data along freight routes is typically applied to each observed INRIX trip to produce flow estimates.

Comparison of Products

This section presents a matrix comparing key products and data delivery capabilities of each of the commercial vendors outlined in the previous section.

All the commercial vendors of "big-data" roadway traffic utilization and performance data depend heavily on smartphone applications, in-vehicle OEM navigational devices, and data from connected vehicles. To obtain these data, the vendors have business agreements with multiple cell phone manufacturers, carriers, and/or smartphone app providers. For example, INRIX has a free downloadable application, named "INRIX Traffic," that provides maps, navigational or route guidance information, and driver alerts. Similarly, through its ownership by a consortium of German automotive companies, HERE Technologies has unique data-sharing opportunities with these auto manufacturers.

The matrix of Table 10 summarizes data sources and relevant data products associated with each vendor, along with relevant supplemental information. In the table, "CELL/GPS/CV" indicates data sourced from a suite of smartphone applications, in-vehicle OEM navigation devices, and connected vehicles.

Aggregated traffic speed and travel time data were among the first types of data to become commercially available. The ability to provide trip origin-destination estimates came several years later. Providing traffic volume estimates is a relatively new feature as this is only made possible if sufficient ground truth data is collected. As the popularity of cell phones and vehicle route guidance apps grew, the amount of data available to these vendors increased, as did the reliability and accuracy of their traffic speed and travel time estimates (and the listing of products offered expanded).

INRIX currently appears to be the primary provider of commercial freight vehicle data, supplying data to the FHWA National Performance Management Research Data Set (NPMRDS), StreetLight, and Replica. Because of this, all these vendors tend to provide truck data analyses based on a similar truck characterization that only distinguishes between medium-duty and heavy-duty trucks. No data is typically provided for light-duty trucks due to the lack of a sufficient data source.

Table 10. Comparative Summary of Traffic Data Provider

Item	FHWA NPMRDS	HERE Technologies	TOMTOM	INRIX	StreetLight Data	Bentley Systems	Replica
Key Data Sources	HERE Technologies (2013-2016); INRIX (2017-Present)	CELL/GPS/CV Multiple sources	CELL/GPS/CV Multiple sources	CELL/GPS/CV Multiple sources	CELL/GPS/CV Multiple sources	CELL/GPS/CV (AirSage) Agency Traffic Counts	CELL/GPS/CV Multiple sources
Data Collection Method(s)	HERE Technologies or INRIX Methods	Smartphone App, Vehicle OEM device, and multiple others	Smartphone App, Vehicle OEM device, and multiple others	Smartphone App, Vehicle OEM device, and multiple others	Smartphone App, Vehicle OEM device, and multiple others	AirSage and Citilabs proprietary optimization process	Smartphone App, Vehicle OEM device, and multiple others
Main Data Products	Travel Times, Speeds	Travel Times, Speeds	Travel Times, Speeds	Travel Times, Speeds, O-D Trips, Estimated Volumes	Travel Times, Speeds, O-D Trips, Estimated Volumes	Travel Times, Speeds, O-D Trips, Estimated Volumes	Estimated Travel Times, O-D Trips, Estimated Volumes,
Separate Truck Data	YES (from INRIX)			YES	YES (from INRIX)		Yes (from INRIX)
Truck Categories	Medium Heavy 2			Medium Heavy 2	Medium Heavy 2		Medium Heavy 2

Item	FHWA NPMRDS	HERE Technologies	TOMTOM	INRIX	StreetLight Data	Bentley Systems	Replica
Additional Data Products Information	Historical speeds and travel times	Historical and real-time speeds and travel times; HOV lane speeds; split lane speeds [1]; incident feed, traffic safety warnings	Historical speeds and travel times, routes, O-D, incidents, bottlenecks	Historical and real-time speeds and travel times, bottlenecks, O-D, volumes, parking, population warehousing	Historical speeds and travel times; Trip duration, length, and purpose; vehicle AADT; O-D, Volume estimates (cars, peds, and bikes)	"Streetlytics" historical speed and travel times; O-D by block-group; AADT & hourly traffic volumes (trip Purpose and mode estimates)	Travel activity, Economic activity
Real-time Delivery Capability	NO	YES	YES	YES	NO	NO	No
Historical Delivery Capability	YES (monthly)	YES (daily)	YES	YES (daily)	YES (daily)	YES (as per client agreement)	YES (weekly)
Predictive Capacity	NO	YES	YES	YES	NO	NO	YES
Data validation reports?	YES	YES	?	YES	YES	YES	YES
Mapping Capability	Uses HERE mapping	Have own map products	Have own map products	Previously used Open Street Map (free) and TomTom (premium), now using HERE	Uses Open Street Maps	Uses HERE mapping	Uses Open Street Maps

Notes: [1] Speed estimates for two dissimilar lane groups, where the speed on one lane group differs from the speed on the second group. Typical locations are freeway diverges and freeway merges where one set of lanes or lane group is congested and the adjacent lane group flows freely.

[2] Truck categorization based on weight: Medium truck: 14,000-26,000 lbs.; Heavy truck: > 26,000 lbs.