3-2020

Guidelines for Using StreetLight Data for Planning Tasks

Hong Yang
Old Dominion University, hyang@odu.edu

Mecit Cetin
Old Dominion University, mcetin@odu.edu

Qingyu Ma
Old Dominion University

Original Publication Citation

Yang, H., Cetin, M., & Ma, Q. (2020). *Guidelines for using StreetLight data for planning tasks* (Publication No. FHWA/VTRC 20-R23). Virginia Transportation Research Council (VTRC). https://www.virginiadot.org/vtrc/main/online_reports/pdf/20-r23.pdf

VTRC
Virginia Transportation Research Council

We Bring Innovation to Transportation

Guidelines for Using StreetLight Data for Planning Tasks

http://www.virginiadot.org/vtrc/main/online_reports/pdf/20-r23.pdf

HONG YANG, Ph.D.
Assistant Professor
Department of Computational Modeling & Simulation Engineering
Old Dominion University

MECIT CETIN, Ph.D.
Professor
Department of Civil and Environmental Engineering
Old Dominion University

QINGYU MA
Graduate Research Assistant
Department of Computational Modeling & Simulation Engineering
Old Dominion University

Final Report VTRC 20-R23

VIRGINIA TRANSPORTATION RESEARCH COUNCIL
530 Edgemont Road, Charlottesville, VA 22903-2454
vtrc.virginiadot.org

Standard Title Page - Report on Federally Funded Project

1. Report No.: FHWA/VTRC 20-R23	2. Government Accession No.:	3. Recipient's Catalog No.:
4. Title and Subtitle: Guidelines for Using StreetLight Data for Planning Tasks		5. Report Date: March 2020
		6. Performing Organization Code:
7. Author(s): Hong Yang, Ph.D., Mecit Cetin, Ph.D., and Qingyu Ma		8. Performing Organization Report No.: VTRC 20-R23
9. Performing Organization and Address: Virginia Transportation Research Council 530 Edgemont Road Charlottesville, VA 22903		10. Work Unit No. (TRAIS):
		11. Contract or Grant No.: 112110
12. Sponsoring Agencies' Name and Address: Virginia Department of Transportation 1401 E. Broad Street Richmond, VA 23219	Federal Highway Administration 400 North 8th Street, Room 750 Richmond, VA 23219-4825	13. Type of Report and Period Covered: Final Contract
		14. Sponsoring Agency Code:

15. Supplementary Notes:
This is an SPR-B report.

16. Abstract:

The Virginia Department of Transportation (VDOT) has purchased a subscription to the StreetLight (SL) Data products that mainly offer origin-destination (OD) related metrics through crowdsourcing data. Users can manipulate a data source like this to quickly estimate origin-destination trip tables. Nonetheless, the SL metrics heavily rely on the data points sampled from smartphone applications and global positioning services (GPS) devices, which may be subject to potential bias and coverage issues. In particular, the quality of the SL metrics in relation to meeting the needs of various VDOT work tasks is not clear. Guidelines on the use of the SL metrics are of interest to VDOT.

This study aimed to help VDOT understand the performance of the SL metrics in different application contexts. Specifically, existing studies that examined the potential of SL metrics have been reviewed and summarized. In addition, the experiences, comments, and concerns of existing users and potential users have been collected through online surveys. The developed surveys were primarily distributed to VDOT engineers and planners as well as other professionals in planning organizations and consultants in Virginia. Their typical applications of the SL metrics have been identified and feedback has been used to guide and inform the design of the guidelines.

To support the development of a set of guidelines, the quality of the SL metrics has been independently evaluated with six testing scenarios covering annual average daily traffic (AADT), origin-destination trips, traffic flow on road links, turning movements at intersections, and truck traffic. The research team has sought ground-truth data from different sources such as continuous count stations, toll transaction data, VDOT's internal traffic estimations, etc. Several methods were used to perform the comparison between the benchmark data and the corresponding SL metrics. The evaluation results were mixed. The latest SL AADT estimates showed relatively small absolute percentage errors, whereas using the SL metrics to estimate OD trips, traffic counts on roadway segments and at intersections, and truck traffic did not show a relatively low and stable error rate. Large percentage errors were often found to be associated with lower volume levels estimated based on the SL metrics. In addition, using the SL metrics from individual periods as the input for estimating these traffic measures resulted in larger errors. Instead, the aggregation of data from multi-periods helped reduce the errors, especially for low volume conditions. Depending on project purposes, the aggregation can be based on metrics of multiple days, weeks, or months.

The results from the literature review, surveys, and independent evaluations were synthesized to help develop the guidelines for using SL data products. The guidelines focused on five main aspects: (1) a summary for using SL data for typical planning work tasks; (2) general guidance for data extraction and preparation; (3) using the SL metrics in typical application scenarios; (4) quality issues and calibration of the SL metrics; and (5) techniques and tools for working with the SL metrics. The developed guidelines were accompanied with illustrative examples to allow users to go through the given use cases.

Based on the results, the study recommends that VDOT's Transportation and Mobility Planning Division (TMPD) should encourage and support the use of the guidelines in projects involving SL data, and that TMPD should adopt a checklist (table) for reporting performance, calibration efforts, and benchmark data involved in projects that use the SL metrics.

17 Key Words: StreetLight Data, Data Quality, Origin and Destination, Traffic Volume, StreetLight Index, Traffic Planning		18. Distribution Statement: No restrictions. This document is available to the public through NTIS, Springfield, VA 22161.	
19. Security Classif. (of this report): Unclassified	20. Security Classif. (of this page): Unclassified	21. No. of Pages: 120	22. Price:

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

FINAL REPORT

GUIDELINES FOR USING STREETLIGHT DATA FOR PLANNING TASKS

Hong Yang, Ph.D.
Assistant Professor
Department of Computational Modeling & Simulation Engineering
Old Dominion University

Mecit Cetin, Ph.D.
Professor
Department of Civil and Environmental Engineering
Old Dominion University

Qingyu Ma
Graduate Research Assistant
Department of Computational Modeling & Simulation Engineering
Old Dominion University

VTRC Project Manager
John S. Miller, Ph.D., P.E., Virginia Transportation Research Council

In Cooperation with the U.S. Department of Transportation
Federal Highway Administration

Virginia Transportation Research Council
(A partnership of the Virginia Department of Transportation
and the University of Virginia since 1948)

Charlottesville, Virginia

March 2020
VTRC 20-R23

DISCLAIMER

The project that is the subject of this report was done under contract for the Virginia Department of Transportation, Virginia Transportation Research Council. The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Virginia Department of Transportation, the Commonwealth Transportation Board, or the Federal Highway Administration. This report does not constitute a standard, specification, or regulation. Any inclusion of manufacturer names, trade names, or trademarks is for identification purposes only and is not to be considered an endorsement.

Each contract report is peer reviewed and accepted for publication by staff of the Virginia Transportation Research Council with expertise in related technical areas. Final editing and proofreading of the report are performed by the contractor.

ABSTRACT

The Virginia Department of Transportation (VDOT) has purchased a subscription to the StreetLight (SL) Data products that mainly offer origin-destination (OD) related metrics through crowdsourcing data. Users can manipulate a data source like this to quickly estimate origin-destination trip tables. Nonetheless, the SL metrics heavily rely on the data points sampled from smartphone applications and global positioning services (GPS) devices, which may be subject to potential bias and coverage issues. In particular, the quality of the SL metrics in relation to meeting the needs of various VDOT work tasks is not clear. Guidelines on the use of the SL metrics are of interest to VDOT. This study aimed to help VDOT understand the performance of the SL metrics in different application contexts. Specifically, existing studies that examined the potential of SL metrics have been reviewed and summarized. In addition, the experiences, comments, and concerns from existing users and potential users have been collected through online surveys. The developed surveys were primarily distributed to VDOT engineers and planners as well as other professionals in planning organizations and consultants in Virginia. Their typical applications of the SL metrics have been identified and feedback has been used to guide and inform the design of the guidelines.

To support the development of a set of guidelines, the quality of the SL metrics has been independently evaluated with six testing scenarios covering annual average daily traffic (AADT), origin-destination trips, traffic flow on road links, turning movements at intersections, and truck traffic. The research team has sought ground-truth data from different sources such as continuous count stations, toll transaction data, VDOT's internal traffic estimations, etc. Several methods were used to perform the comparison between the benchmark data and the corresponding SL metrics. The evaluation results were mixed. The latest SL AADT estimates showed relatively small absolute percentage errors, whereas using the SL metrics to estimate OD trips, traffic counts on roadway segments and at intersections, and truck traffic did not show a relatively low and stable error rate. Large percentage errors were often found to be associated with lower volume levels estimated based on the SL metrics. In addition, using the SL metrics from individual periods as the input for estimating these traffic measures resulted in larger errors. Instead, the aggregation of data from multi-periods helped reduce the errors, especially for low volume conditions. Depending on project purposes, the aggregation can be based on metrics of multiple days, weeks, or months.

The results from the literature review, surveys, and independent evaluations were synthesized to help develop the guidelines for using SL data products. The guidelines focused on five main aspects: (1) summary of using SL data for typical planning work tasks; (2) general guidance of data extraction and preparation; (3) using the SL metrics in typical application scenarios; (4) quality issues and calibration of the SL metrics; and (5) techniques and tools for working with the SL metrics. The developed guidelines, shown in this report in Appendix A, were accompanied with illustrative examples to allow users to go through the given use cases.

Based on the results, the study recommends that VDOT should encourage and support the use of the guidelines in projects involving SL data. It is also recommended that VDOT should adopt a checklist (table) for reporting performance, calibration efforts, and benchmark data involved in projects that use the SL metrics. A sample checklist is shown in Appendix C.

FINAL REPORT

GUIDELINES FOR USING STREETLIGHT DATA FOR PLANNING TASKS

Hong Yang, Ph.D.
Assistant Professor
Department of Computational Modeling & Simulation Engineering
Old Dominion University

Mecit Cetin, Ph.D.
Professor
Department of Civil and Environmental Engineering
Old Dominion University

Qingyu Ma
Graduate Research Assistant
Department of Computational Modeling & Simulation Engineering
Old Dominion University

INTRODUCTION

Traditionally, travel surveys and diaries have been one of the main sources of data for understanding and quantifying travel behavior. Major travel surveys, conducted typically once a decade, include rich behavioral data such as trip purpose, household size, origin-destination of trips, travel mode, trip time, socio-demographic information, etc. Recently, alternative technologies have emerged for collecting travel data, such as license plate matching and Bluetooth media access control (MAC) address matching. Data from these alternative technologies, albeit not as rich as travel surveys, enable collecting a much larger sample cost effectively. In addition to matching unique identifiers (e.g., MAC address) between two observation points, tracking individual trip makers through their mobile devices has become another common data source. In particular, mobile devices (e.g., smartphones) equipped with global positioning services (GPS) and other tracking capabilities enable collecting high-resolution location data to infer trip time, trip length, route, and possibly the travel mode. When overlaid with land-use and activity-location data, it becomes possible to estimate the trip purpose as well from such high-resolution data. To take advantage of these emerging data sources, the Virginia Department of Transportation (VDOT) has purchased a subscription to StreetLight—an application that provides estimates of origin-destination vehicle trips. StreetLight capitalizes on the massive volume of geospatial information created by mobile phones to generate estimates of ODs, trip purpose, and travel times for personal and commercial trips. The StreetLight (SL) platform enables the users to design, run, and visualize customized queries like origin-destinations and link flows that may be disaggregated by time of day and trip purpose.

While there are promising aspects of using such data, it is imperative to fully understand any potential biases and inaccuracies before SL data are used in planning and other VDOT activities. For example, since the main data sources of SL data are from smartphone applications and commercial vehicles' GPS devices, the data can be biased due to the affordability of smartphones and cellular data plans within low-income groups and uneven market penetrations among the trips being analyzed. Thus, a set of questions are warranted. For example, what types of applications can take advantage of the SL data? Are the SL data accurate enough in each application? If not, what are the gaps, how accurate are the estimates, etc.? How can a transportation planner extract and analyze the data? To answer such questions, this research project conducted a comprehensive evaluation of the SL data to support VDOT in making the best use of the data in its various work tasks.

PURPOSE AND SCOPE

The primary goals of this project are to (i) determine which existing VDOT work tasks and associated methodologies might be enhanced by SL data; (ii) assess whether the accuracy provided by SL is sufficient for use in the relevant studies; and (iii) develop basic guidelines for using the SL data under various applications.

To accomplish these goals, a survey of existing practices in using SL data within and beyond VDOT was conducted to determine its applicable areas, advantages, and limitations. The accuracy of using SL data in each type of planning and other tasks was determined considering the spatiotemporal coverage, resolution, and specific metrics needed. More specifically, the research team interviewed VDOT travel demand modeling groups and related users (e.g., VDOT Traffic Engineering Division (TED), Metropolitan Planning Organizations (MPOs), and Planning District Commissions (PDCs)), to understand their current SL data usage practices. Their feedback regarding the SL data applicability, specific data elements, data coverage, resolution, quality, data processing practices as well as concerns and limitations were summarized to create an "application list" for mapping each type of tasks with the necessary SL data elements. Considering the variety of work tasks, the SL data used in projects of different scales were compared with other well-known data sources (e.g., loop detectors). This involved evaluation of the SL data used in: (i) regional/zonal-level, (ii) corridor-level (e.g., highways and arterials), and (iii) site/station-level (e.g., ports, intersections, bridges / tunnels, and road segments) studies.

Built upon a comprehensive literature review, survey, and comparative evaluations, guidelines for using the SL data with functionality as of July 2019 were developed and related recommendations were made available to VDOT. It should be noted that the scope of this work is a snapshot in time since the SL Platform is evolving. Therefore, a decision had to be made to evaluate the SL Platform given its capabilities at the time period work was conducted. The researchers used the version of SL Platform between September 2018 to August 2019.

METHODS

To achieve the project goals, the following tasks were conducted:

1. Perform a literature review on the use of SL data.
2. Develop and distribute surveys to existing and potential users of SL data in Virginia.
3. Perform comparative studies to evaluate the use of SL data.

4. Develop guidelines for using SL data.

Literature Review

The literature review identified studies on the use of SL data. Major areas of the literature review included the examination of applications and practices in using SL for addressing transportation issues and evaluating the accuracy of SL data. The reviewed references were identified through research databases and search engines including Google Scholar, Transportation Research Board's Transport Research International Documentation (TRID), Web of Science, and Scopus. Research articles, publicly available presentation files, and reported information on webpages that are related to SL data were explored and synthesized.

Users and Non-User Surveys

Following the results of the literature review, the research team developed an online survey for transportation professionals to elicit input about their experience in (considering) using SL data in their typical projects. The survey questions covered application scopes, concerns, suggestions, etc. regarding the use of SL data. The research team received responses from professionals working in different groups of VDOT, MPOs/PDCs, and consultants primarily in Virginia. Two survey instruments were designed. A longer survey was designed for the current users and a shorter one for non-users. The survey sample was obtained by distributing the surveys to SL users in Virginia, attendees of recent SL Data User conferences, consultants, and contacts provided by the TRP such as MPO/PDC staff and VDOT district staff. These surveys can be found in Appendix C.

Evaluation of SL Data

The research team at Old Dominion University (ODU) compared the SL Data with a set of data from other resources, including VDOT annual average daily traffic (AADT) data, State and local traffic detector data, toll-transactions data on I-66, and bike-sharing data. The comparisons were performed at site level, sensor station level, and corridor/zone level. The goal was to evaluate whether the SL metrics can be a good proxy for actual traffic metrics such as traffic count, OD trips, and AADT. In order to evaluate the performance of the SL Index, performance measures were defined and the preparation of all the benchmark data was delineated.

Evaluation Scenarios

The research team focused on five main categories for the evaluation of the quality of the SL data: (a) SL annual average daily traffic (AADT) estimate; (b) SL Index between different zones for the purpose of estimating OD trips; (c) SL Index captured by the middle filter (pass-through zone/gateway) on specific road links; (d) SL Index representing the turning traffic at intersections; and (e) SL Index representing truck traffic approach intersections. Figure 1 provides illustrations for the five categories of analyses. The first category denotes the SL AADT estimation in a year for selected road links. The second category tests the capability of the SL

Index as a source for estimating travel demand between zones. The third category assesses whether the SL Index can be a good source for quantifying the variation of link flow. The fourth one evaluates the potential of the SL Index for describing turning movement at intersections. The last category examines whether the SL Index based on navigation-GPS data is good for depicting truck traffic at intersections approaches. Due to data availability, the evaluation of truck metrics was only limited to the selected intersections. This could be extended to freeway sites if vehicle classification data were available.

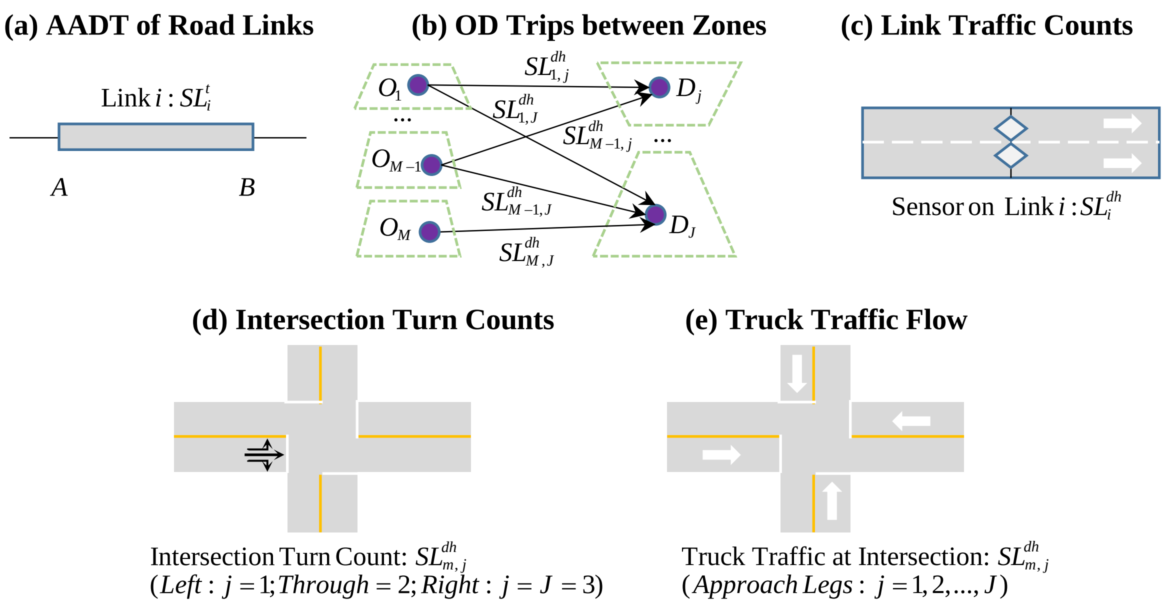


Figure 1. Design of Evaluation Scenarios.

Defining Performance Measures

Other than the SL AADT estimates, the other SL Index represents a relative indicator of the actual traffic measurements of interest (e.g., OD trips, traffic counts, etc.). When referring to an actual (or true) measurement we will use Y, and $\hat{Y}$ when referring to an estimate of Y. By actual measurements in this report, we are referring to agency-reported data or observed measurements from vehicle detector sensors in the field that are assumed to provide reliable data. The SL Index SL has no unit and is supposed to only reflect the relative level of the actual measurement Y. In practice, the SL Index usually needs to be converted to an estimate of a specific traffic measurement (e.g., volume and trips). If we have the actual traffic measurement Y and its corresponding estimate $\hat{Y}$, we can compute a set of performance measures, for example, their difference. The value of $\hat{Y}$ is obtained through the following procedures for the three evaluation scenarios:

- **Estimate of AADT:** Since SL publishes its own AADT estimates ($\hat{Y}$) through the SL Platform, these published estimates $\hat{Y}_i^{year}$ for all links $i=1,2,...,I$ in a given year will be used without performing any calibration. The difference between the estimated AADT and the actual AADT is indicated as ε_i^{year}, where $\varepsilon_i^{year} = Y_i^{year} - \hat{Y}_i^{year}$.

- **Estimate of OD trips:** The SL Index in the OD analysis reflects the relative level of trips between origins and destinations. Thus, SL Index needs to be scaled so that it represents the number of trips from an origin to a destination. As shown in Figure 1(b), assuming the SL Index between an origin zone m and a destination zone j is $SL_{m,j}^{dh}$ at time h on day d and the corresponding actual demand is $Y_{m,j}^{dh}$. $SL_{m,j}^{dh}$ needs to be scaled to produce $\hat{Y}_{m,j}^{dh}$ which can then be compared with $Y_{m,j}^{dh}$. Since we know the SL Indexes from an origin zone m to all the other destination zones, we can use these indexes to calculate the proportion or percentage of the demand from zone m to any destination $j\,(j=1,2,...,J)$ as follows:

$$p_{m,j}^{dh} = \frac{SL_{m,j}^{dh}}{\sum_{j=1}^{J} SL_{m,j}^{dh}} \times 100 \qquad (1)$$

Given the total trips $\sum_{j=1}^{J} Y_{m,j}^{dh}$ leaving from the origin zone m, the result of Eq. (1) will be used in the following equation for the estimation of $\hat{Y}_{m,j}^{dh}$.

$$\hat{Y}_{m,j}^{dh} = p_{m,j}^{dh} \times \sum_{j=1}^{J} Y_{m,j}^{dh} \qquad (2)$$

Alternatively, one can also calculate the proportion of trips from any origin zone $m\,(m=1,2,...,M)$ to a specific destination j as follows:

$$p_{m,j}^{dh} = \frac{SL_{m,j}^{dh}}{\sum_{m=1}^{M} SL_{m,j}^{dh}} \times 100 \qquad (3)$$

Given the total trips $\sum_{m=1}^{M} Y_{m,j}^{dh}$ arriving the destination zone j, the result of Eq. (3) will be used in the subsequent estimation for obtaining $\hat{Y}_{m,j}^{dh}$.

$$\hat{Y}_{m,j}^{dh} = p_{m,j}^{dh} \times \sum_{m=1}^{M} Y_{m,j}^{dh} \qquad (4)$$

Eqs. (2) and (4) only represent whether the analysis is based on the available information of trip production and attractions, respectively. In this project, we used Eqs. (1) and (2) for the estimation of $\hat{Y}_{m,j}^{dh}$. Then, the difference between the estimated trips and the actual trips is calculated as $\varepsilon_{m,j}^{dh}$, where $\varepsilon_{m,j}^{dh} = Y_{m,j}^{dh} - \hat{Y}_{m,j}^{dh}$.

- **Estimate of traffic counts:** The scenario shown in Figure 1(c) represents the analysis of the SL Index associated with the specific road link. The index SL_i^{dh} at time h on day d can be easily obtained from the SL Platform through the placement of a pass-through zone on the target link i. SL_i^{dh} does not represent the approximate of traffic volume. Instead, it is just an indicator that reflects the relative level of the traffic volume. To make the index useful, it needs to be converted to an estimated volume through a calibration process. Since SL_i^{dh} and its corresponding actual traffic volume Y_i^{dh} are both time series data, a high SL_i^{dh} should reflect a high volume condition Y_i^{dh}. For example, if $SL_i^{dh} = 100$ and $SL_i^{dh+1} = 200$, then the corresponding actual traffic measurement Y_i^{dh+1} is expected to be twice of Y_i^{dh}. The estimate $\hat{Y}_t$ can be obtained through a conversion model: $\hat{Y}_i^{dh} = f(SL_i^{dh})$, where $f()$ is a transformation function. Since the vendor does not provide specific guidance for converting the SL Index to volume or trips, one rational way to do so is through a linear model:

$$\hat{Y}_i^{dh} = f(SL_i^{dh}) = \alpha + \beta \times SL_i^{dh} \tag{5}$$

where, α and β are constant parameters. This conversion model implies that the traffic measurement will be linearly represented by the SL Index. Based on Eq. (5), we can obtain the estimate $\hat{Y}_i^{dh}$. In reality, due to various issues such as unstable sampling rate at different times, the $\hat{Y}_i^{dh}$ based on SL Index SL_i^{dh} will not always match with the actual measurement Y_i^{dh}. Thus, we can represent the ground-truth Y_i^{dh} as follows:

$$Y_i^{dh} = f(SL_i^{dh}) + \varepsilon_i^{dh} = \hat{Y}_i^{dh} + \varepsilon_i^{dh} = \alpha + \beta \times SL_i^{dh} + \varepsilon_i^{dh} \tag{6}$$

where, $\varepsilon_i^{dh} = Y_i^{dh} - \hat{Y}_i^{dh}$ reflects the difference between the actual measurement and the estimate based on SL Index.

- **Estimate of turn counts and truck volumes at intersections:** The same approach presented in the estimation of traffic volume for road links is adopted for estimating turn counts and truck volume at intersections. Each SL Index for a turning movement direction is treated as the SL Index for a link flow in equations (5) and (6). Similarly, the SL Index for truck traffic at an approach leg is also treated as the link flow in using the in equations (5) and (6).

The following two key performance measures based on ε^t and the corresponding actual observation Y^t at time t are proposed for quantifying the performance of SL Index:

- **Percentage Error** (PE): This describes the relative difference between the estimated traffic measurement based on SL Index for a specific analysis unit (e.g., link, OD pair, etc.) and the actual traffic measurements at time t:

$$PE^t = \frac{-\varepsilon^t}{Y^t} \times 100 \tag{7}$$

- **Absolute Percentage Error** (APE): This represents the absolute relative difference between the estimated traffic measurement based on SL Index and the actual traffic measurements at time t :

$$APE^t = \frac{|\varepsilon^t|}{Y^t} \times 100 = |PE^t| \tag{8}$$

The obtained performance measures PE and APE can be further analyzed based on different objectives. For example, each performance measure calculated for different time periods and units are grouped by (i) actual volumes (trips); (ii) the corresponding $\hat{Y}^t$ based on SL Index; (iii) facility types; (iv) time periods, etc. Other indicators such as the mean, median, and 95% confidence interval (CI) of each performance measure in each group of analysis are also reported for comparison.

Benchmark Data Preparation

The research team collected different types of benchmark data to evaluate the performance of the SL Index in approximating traffic flow, OD trips, and AADT. With the exception that "observed truck traffic at intersections" was compared to the SL navigation GPS index, the benchmarks used in the other analysis were compared to the SL personal LBS index. This is because the personal GPS data source is sampled from all types of vehicles, which is consistent with benchmark data.

VDOT AADT Data

Since SL published 2017 and 2018 AADT estimates, the research team also obtained the AADT data published by VDOT for the same two years (VDOT, 2019). The SL AADT data cover the Interstate, arterials, and primary routes in Virginia. The VDOT AADT data include 8,857 and 8,868 unique road links in 2017 and 2018, respectively. One attribute in the data table indicates the quality (QA) of AADT for each link. The glossary of terms in the downloaded PDF of the VDOT traffic data indicates that the AADT of the links with QA=A were estimated based on the average of complete continuous count data measured by traffic sensors. In total, 955 and 921 links with QA=A were included in the data of 2017 and 2018, respectively. The sensor information including coordinates, sensor types, and sensor stationIDs, was obtained from VDOT. The unique sensor stationIDs were used to match with links through the comparison with the LinkIDs. Among the links with QA=A, we extracted the ones whose traffic sensors are in-road sensors (e.g., loop detectors). Links with other types of sensors such as radar were not used. These efforts aimed to obtain the most reliable data as the benchmark. Finally, 204 and 193 links were selected for 2017 and 2018, respectively. Most of the links were present in both years' data. In total, 225 unique links were included in the two sets of links. Their VDOT AADT values were used as the benchmark for evaluating the AADT estimates by the SL Platform. The corresponding AADT estimates by SL were obtained through gateways placed at the sensor locations on the SL Platform. Figure 2 shows the geographic distribution of the gateways on the selected links across Virginia. Table 1 summarizes the selected links by road systems.

Table 1. Selected Links for AADT Comparison

	Samples by Year	
Road Systems	**2017**	**2018**
Interstate (IS)	83	81
State Route (SR)	47	45
US Route (US)	67	62
Others	7	5
Total	**204**	**193**

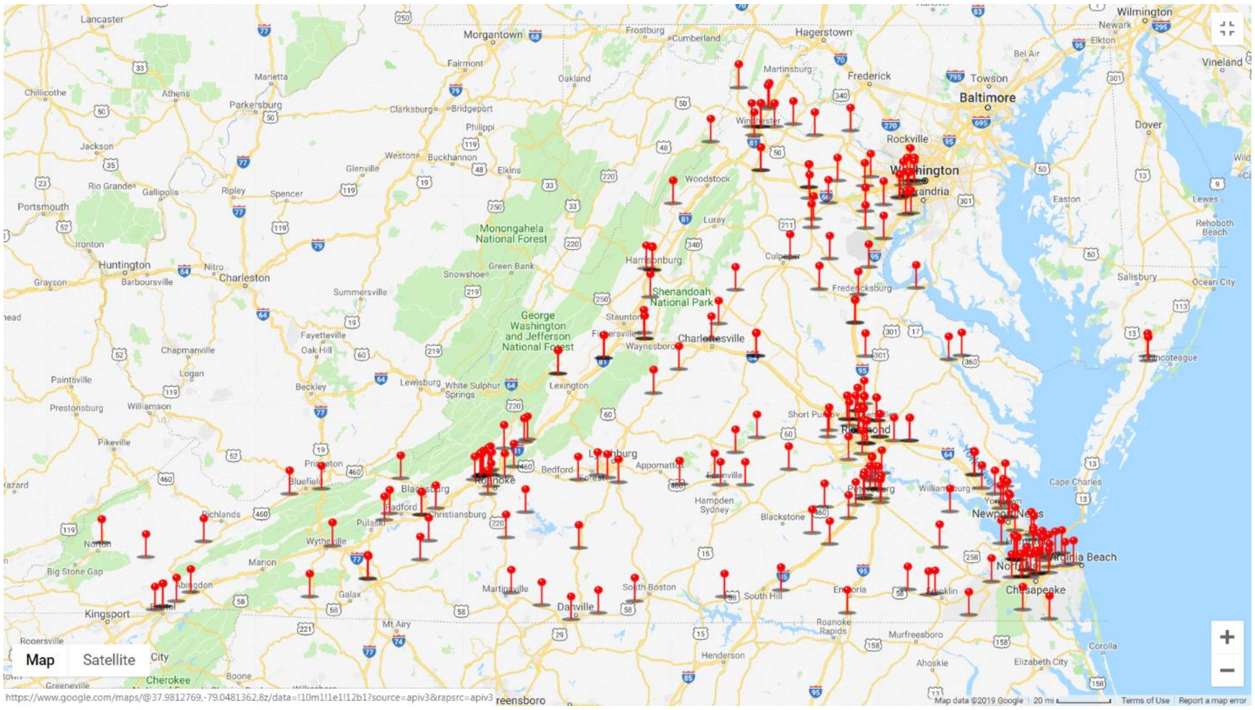
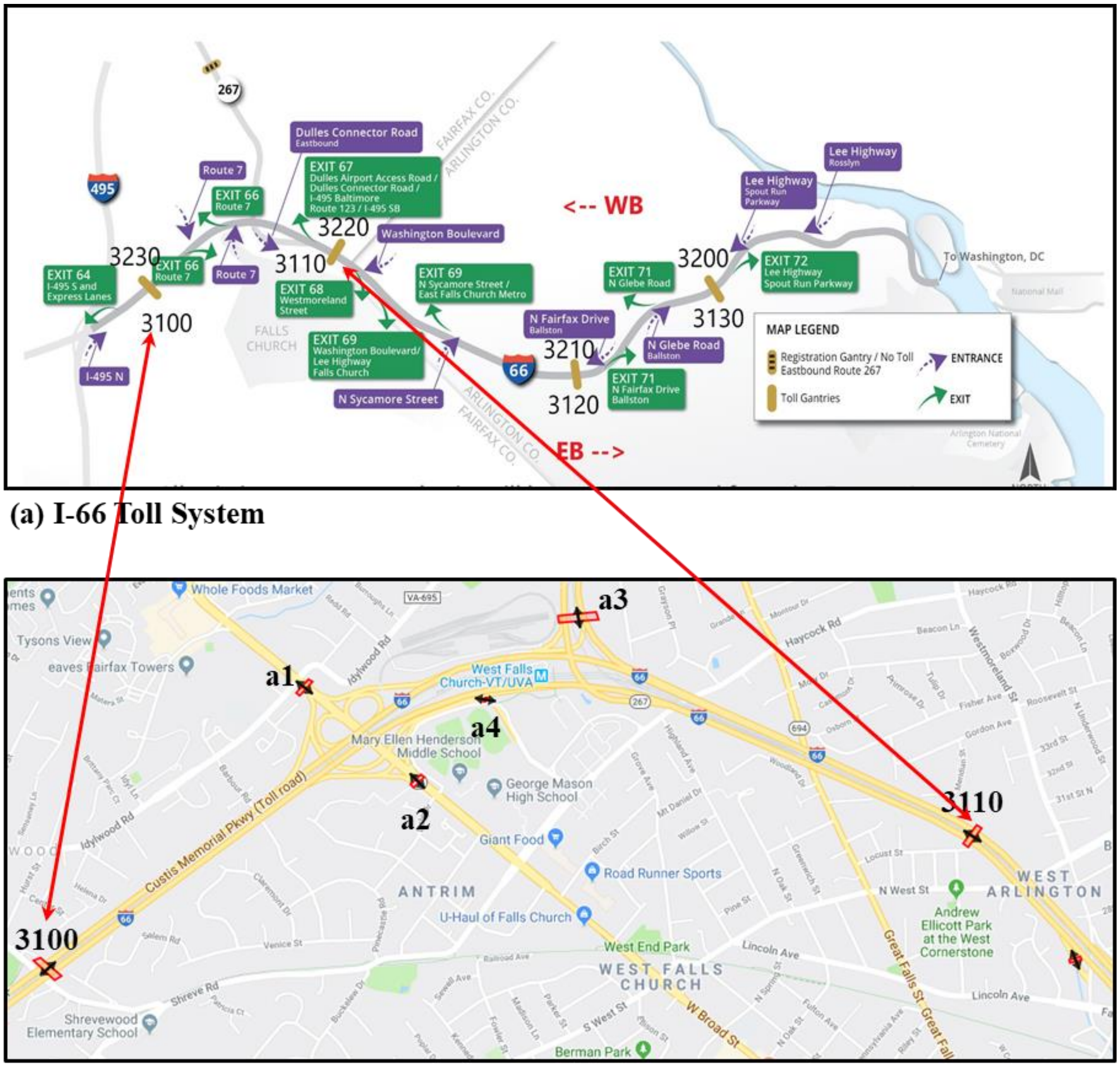


Figure 2. Locations of Selected Links for AADT Evaluation [Note: The selected links presented in at least one of the two years – 2017 and 2018]. © 2019 Google Maps.

OD Trips Based on Toll Transaction Data

The research team obtained the toll transaction data archived by the Electronic Toll System of the I-66 Express Lanes inside the Beltway in Northern Virginia (Virginia Department of Transportation 2018). The toll system has four toll gantries distributed between exit 64 and exit 72 along I-66. The hours of operations are 5:30am-9:30am in Eastbound (EB) lanes and 3pm-7pm in Westbound (WB) lanes during the weekdays. Figure 3(a) shows the layout of the toll gantries. From the timestamps of each vehicle passing the toll gantries, it is possible to derive an actual OD matrix for the traffic during each toll operation period. Since the minimum aggregation interval of the SL Index is by hour, we extracted the toll transaction data for 3 hours (6am-9am) in the morning and 4 hours (3pm-7pm) in the afternoon for all weekdays in May 2018, with the exception of Memorial Day (Monday, May 28, 2018). In total, hourly OD matrices for 22 weekdays were prepared based on the obtained toll transaction data. The corresponding OD matrices for the SL Index were extracted from the SL Platform. It should be noted that we deploy pass-through zones between two consecutive gantries to cover all possible trips that entered or left the toll system. Figure 3(b) provides an example showing that pass-through zones a1, a2, a3, and a4 were deployed for collecting the SL Index for traffic passing gantry 3100 but leaving I-66 before reaching gantry 3110.



(a) I-66 Toll System

(b) Example of OD Zone Configurations on SL Platform

Figure 3. Collecting OD Data between Zones along a Highway Corridor. [Original Figure 3(a) from (Virginia Department of Transportation 2018), annotation added by the authors.] Accessed May 16, 2019. Reprinted With Permission.

Traffic Count Data from Virginia Beach

In order to assess the quality of the SL Index in estimating traffic flow on local roads, the research team identified a set of locations in the City of Virginia Beach to compare the traffic volume measured by traffic detectors deployed on the road segments with the corresponding SL Index obtained at the same sites. We primarily used the archived traffic count database system (TCDS) of the City of Virginia Beach (City of Virginia Beach 2019). We extracted the hourly data from the TCDS for 21 sites with different levels of annual average daily traffic (AADT). Among the studied sites, 19 sites have directional traffic counts and 2 sites only have one-way measurements. In total, this offers us $19 \times 2 + 2 = 40$ segments for comparisons. The map in Figure 4 shows the locations of these studied sites. All selected sites had continuous traffic count in April 2018. The pass-through zones were deployed at the sensor locations on SL Platform to retrieve the corresponding hourly SL Index for each site in April 2018. It should be noted that the SL Index may not be available for some sites at some hours due to low sample sizes during these periods. In such cases, no comparisons were made. The plot in Figure 4 provides an example of the raw measurements for one site.

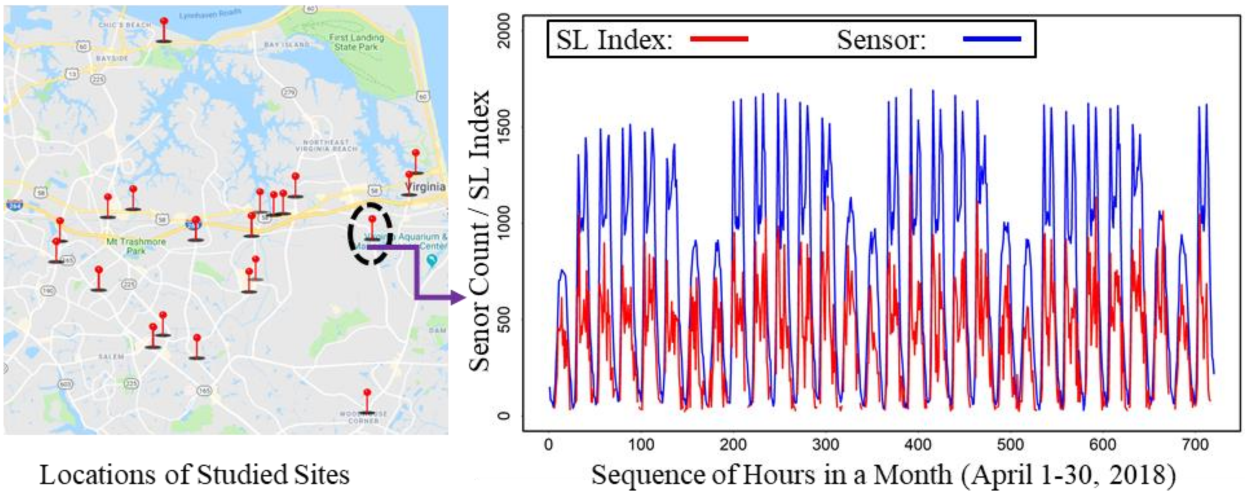


Figure 4. Distribution of Tested Sites and an Example of SL Index *vs.* Observed Sensor Counts. Left map: Accessed July 8, 2019. Reprinted With Permission.

VDOT Traffic Count Data for Links with Hourly Measurements

The research team received the hourly traffic count data from VDOT for the same links selected in the AADT analysis. As described above, these links cover different types of roads. In total, 193 links with hourly traffic count data collected in April 2018 were used for evaluation. Like AADT analysis, we extracted the corresponding SL Index based on the gateways placed at the sensor site on each link.

Turn Counts at Intersections

From the TCDS of the City of Virginia Beach, we extracted the traffic turn counts at 17 intersections. Figure 5 shows the location and a zone configuration example for turning traffic analysis. For example, we are interested in the traffic passing through Zone 1 and making a right turn to Zone 4, going through to Zone 3, and making a left turn to Zone 2 at this intersection. Four zones need to be drawn to capture these turn counts. Since only short-term counts were collected by the City, we extracted the data collected between 7:00am and 8:00am on 09/07/2017 and 09/09/2017. For each intersection, left turn count, through traffic, and right turn count at one of its approach legs were extracted from the TCDS. The corresponding hourly SL Indexes were extracted from the SL Platform. Although there were observed turn counts for each selected intersection, SL Indexes were not available for some of the turning directions at some intersections.

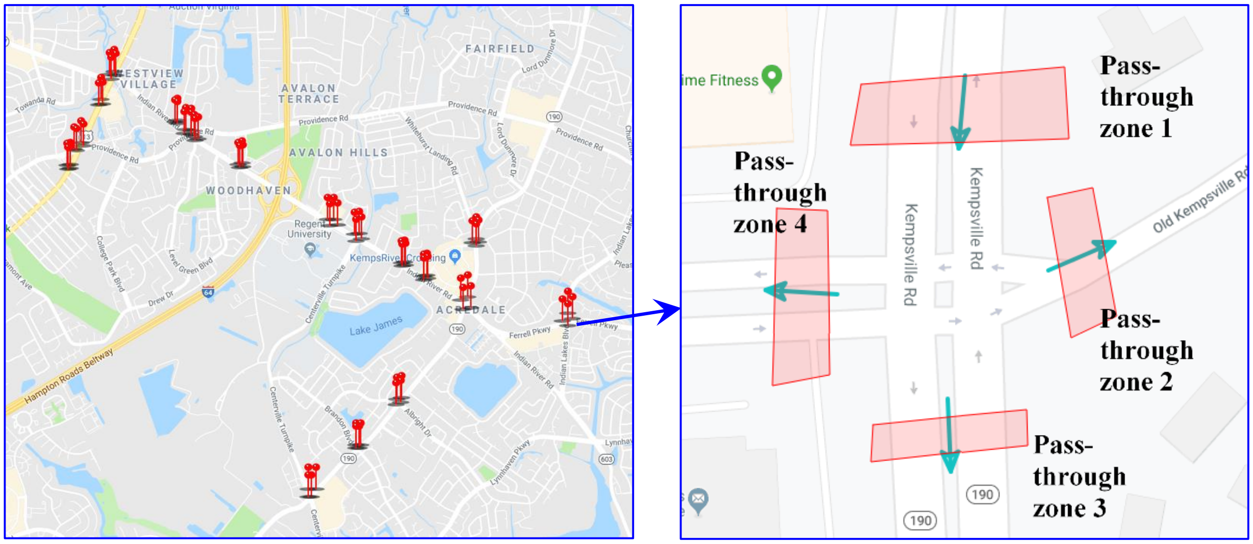


Figure 5. Locations and a Zone Configuration Example for Turning Movement Data Collection.

Truck Volume at Intersections

From the TCDS of the City of Virginia Beach, we extracted the hourly truck volume approaching the intersections as mentioned in the previous section (see Figure 6). The truck volumes represent the short-term counts of trucks passing each approach leg of every intersections. We extracted the data collected between 7:00am and 8:00am on 09/07/2017 and 09/09/2017. The observed truck volumes for every approach leg of the 17 intersections were prepared for comparison. The corresponding SL Indexes were extracted from the SL Platform. It should be noted that some SL Indexes were not available.

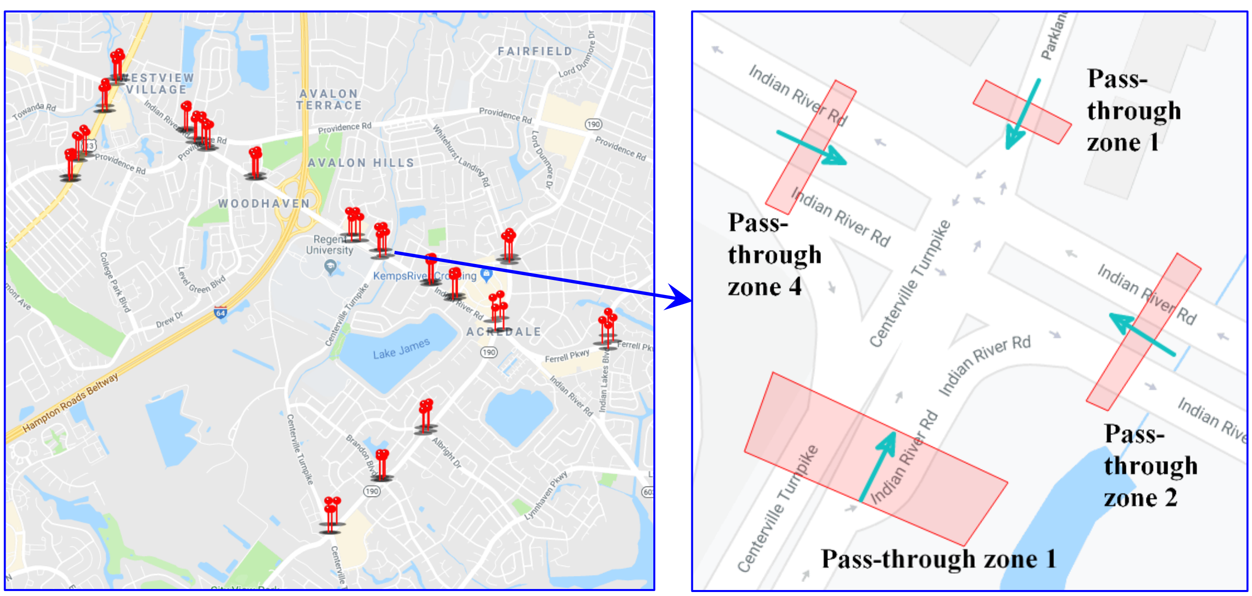


Figure 6. Locations and a Zone Configuration Example for Truck Traffic Data Collection. Accessed July 20, 2019. Reprinted With Permission.

Development of Guidelines for Using SL Data

Based on the findings from the literature review, user surveys, and independent evaluations with real-world data, the ODU research team developed a draft set of guidelines for using the data products from the SL Data. The guidelines are organized according to five main aspects: (1) summaries for related planning work tasks that may use SL data, (2) general guidance of data extraction from the SL Platform, (3) typical applications in different planning tasks, (4) data quality and calibration, and (5) possible tools and techniques that may support the use of the SL products.

RESULTS AND DISCUSSION

Literature Review

SL data provide a convenient way of extracting well-organized travel information, such as an OD trip index, estimation of AADT, zone-activity analysis and have drawn growing attention from transportation professionals. We reviewed the existing literature on research using SL data to explore the dynamic roles and potentials that SL might serve in transportation projects. Some of the typical applications are summarized in Table 2.

OD analysis serves as one of the major products that users can incorporate into their projects. It has been frequently utilized to assist users in several aspects such as analyzing travel demands and exploring travel behaviors. For example, Fehr & Peers (2016) used SL OD analysis to explore travel trends in Park City of Utah. Meanwhile, Georgia Department of Transportation (2016) utilized SL OD analysis to calculate OD matrix indexes in the Downtown Connector of Georgia. Later, CDM Smith (2017) examined an OD matrix in Lake/Orange Connector Expressway in West Orange County and East Lake County of Florida using SL OD analysis. In addition, SL OD analysis has also been used in smaller regions such as public transportation stations. For instance, SSTI utilized SL OD pattern analysis at the Zinfandel station for solving the last mile problem (State Smart Transportation Initiative, 2017; McCahill, 2017). They conclude that the performance of SL OD analysis is satisfactory except for that the quality of SL data required further validation. Meanwhile, the OD analysis (Shay, 2017) was adopted and found to be useful for visualizing freight patterns originating near the Rickenbacker International Airport of Ohio. These practices imply that SL OD analysis has been frequently used in regions such as public transportation stations, downtown areas, census tracts, cities, and counties.

Another important product of SL is the AADT estimates. AADT estimations in 2017 and 2018 are provided by the SL Platform, which has also been applied in areas without AADT estimations by DOTs. For example, Coates (2017) used the SL AADT estimates to analyze AADT on the campus of Miami University (in Florida). Since AADT estimation is quite useful and has been of interest to many agencies, there are multiple attempts to evaluate the quality of the SL AADT estimate. Granato (2017) validated Ohio DOT and SL estimation of daily trucks and reached an R-squared value of 0.7555. Similarly, Minnesota Department of Transportation (2017) found that 2016 SL estimated ADTs of I-70 and I-75 tend to be higher than the AADT (2010) by the Miami Valley Regional Planning Commission (MVPRC). All in all, it should be noted that the quality of AADT varies under different traffic volumes. Turner and Koeneman (2017) adopted the mean absolute percentage error (MAPE) to evaluate the data quality of SL AADT. According to their conclusion, the MAPE values associated with traffic volumes at [300-5,000], [5,000-10,000], [10,000-20,000], [20,000-50,000], and [50,000-Inf.] are 68%, 58%, 44%, 29%, and 34% respectively. Likewise, StreetLight Data (2019) also proclaimed that the root mean square error (RMSE) of AADT can range from 21% for high AADT to 47% for lower AADT. These results suggest that the data quality of the SL AADT estimates varies among different levels of traffic volumes.

Besides AADT, researchers have also tested the accuracy of SL from different aspects. It was found that the spatial precision for GPS and location-based services (LBS) data is about 5 and 20 meters, respectively (StreetLight Data, 2018), and thus data might be inaccurate on some local roads with major roads nearby. Also, the penetration ratio of mobile users is unknown and can be biased with income level, age, and time. As such, SL has scaled and calibrated the data to achieve high confidence in data accuracy. For example, the correlation between unscaled and uncalibrated sample data provided by SL and the observed AADT valued provided by MnDOT is 0.79, whereas it was improved to 0.85 after SL’s scaling and calibration process (Turner and Koeneman, 2017). However, the data processing process of SL is not clear to users. Many projects using SL have compared SL Indexes with other benchmark data. The benchmark data come from surveys, license plate system, on-road sensors, video records, etc. Venkatanarayana and Fontaine (2018) validated an OD matrix that was estimated through multiple data types sourced including AirSage, selected Bluetooth, and automated license plated readers. The mean censored average percentage difference for all OD matrices was significantly higher than the 5% rate and most OD pairs were found to be unacceptable. On the other hand, StreetLight Data (2018) also examined the license plate accuracy on one Friday on certain road segments in Napa County. The numbers in different travel modes such as personal LBS and navigation GPS have shown error rates of 9% (personal LBS) and 5% (commercial GPS), respectively. Therefore, the validation process is inevitable because the suggestions from SL are often difficult to apply to different scenarios.

Given that it appears that simply relying on SL data will not produce robust and accurate results, researchers have compared and integrated SL with other data sources such as surveys, camera systems, and vehicle plate systems to address the accuracy issue. For instance, Miller and O’Kelly (2016) found that SL data outperformed AirSage and the American Trucking Research Institute (ATRI) data when estimating trip length, trip purpose, internal/external flows, and external stations. Kuppam et al. (2017) combined SL data with other types of data such as surveys to develop the tour-based truck travel models. Moreover, Fehr & Peers (2016) integrated SL data with vehicle classification count data, license plate analysis data, and survey data to estimate the OD pattern on U.S. 101 in and between Ventura, Santa Barbara, and San Luis Obispo counties of the Central Coast region of California. Kimley-Horn (2018) compared the data from the Metropolitan Washington Council of Governments (MWCOG) and SL and found that trip patterns of the SL Index were more accurate on yearly OD travel patterns, but MWCOG data offered a more reasonable magnitude of vehicular trips relative to the Rosslyn study area. Later, Schiffer (2016) proposed the integration of SL data with AirSage data, ATRI data, and Skycomp data to conduct a better estimation of the OD matrix. Picado (2017) examined

Southeast Florida origin-destination datasets derived from the SL Index based on the calibrated scale factor and the AADT observed by FDOT in 2015 and 2016.

In addition to the literature mentioned above, there is other research addressing problems with other attribute information provided by SL, such as truck GPS navigation, speed, trip time, trip length, and other demographic features (e.g., trip purpose). For example, Hong (2017) incorporated SL metrics in April 2015 to classify medium and light trucks. Likewise, Federal Highway Administration (2018) analyzed truck travel mode in MAG (Maricopa Association of Governments)-PAG (Pima Association of Governments) mega-region based on SL data. Cambridge Systematics (2017) used SL data to analyze truck travel mode patterns in the San Joaquin Valley. Similarly, Harrison (2017) extracted SL mobile and contextual data and revealed the opportunities for their applications in public transit planning aspects such as the optimal place to expand transit in Virginia. There are some other studies on attribute information, for example, Aultman-Hall and Dowds (2017), DenBleyker et al. (2018), Komanduri et al. (2017), Avner (2018), SANDAG (2018), and Schiffer (2016). A major concern of these applications is that the precision issues can be problematic along roadways with heavy congestion, sharp curvatures, or dense streets.

In summary, SL provides indexes based on massive LBS and GPS data and these data have been used for various applications including truck trips/demographics, OD estimation, travel demand and behavior modeling and analysis, congestion analysis, and AADT estimations. OD Index and AADT can be validated and integrated with other sources of data. SL data accuracy tends to be a concern under low volume conditions, and thus needs additional effort for calibration. Validation of SL data still needs specific experiments under different scenarios.

Table 2. Summary of SL Data Usage

Reference	Purpose	Data Quality Evaluation	Data Fusion
(CDM Smith, 2017)	Access the viability of a new toll road project, Estimate OD matrix for travel demand model	Model prediction vs 2015 base-year conditions RMSE (root mean square error); volume/count (V/C) ratio	Start with SL Lake/Orange County Traffic Counts for refinement
(Wahlstedt, 2017)	Estimate OD for operational modeling and simulation in VISSIM	Descriptive Comparison of benefits and limitations (Visual observation, license plate survey, AirSage, SL)	SL
(Tillery and Pourabdollahi, 2016)	Tourism travel demand, Florida DOT, Simulation	Descriptive Comparison of benefits and limitations (Survey, AirSage, SL, Visit Florida data)	NA
(Napa Valley Transportation Authority, 2014)	Travel behavior study in Napa County	NA	Strat with SL vehicle classification count/ license plate matching/ Winery regression analysis/ surveys for refinement
(McAtee, 2017)	Validate trip distribution in South Michigan	Peak travel time distance (survey, Skims, SL) Personal/commercial vehicle trip patterns (SL, survey) Percentage error	NA
(Herman and Tong, 2017)	OD data collection & estimation I-95 Corridor OD estimation, traffic flow pattern at STARS Route 7	Weaving Pattern (SL, Bluetooth) Percentage error	SL
(Turner and Koeneman, 2017)	Traffic volume and AADT estimation	AADT/average annual hourly volume under different traffic volume level, (SL, MnDOT data) MAPE, MAD.MSD	SL
(Aultman-Hall and Dowds, 2017)	Travel demand model and travel behavior	Descriptive comparison of benefits and limitations (SL, AirSage, paper/telephone/online survey, mobile device app)	NA
(Cambridge Systematics, 2017)	Truck Trip distribution, OD matrix, goods movement study	Share of long-distance trips (SL, 2008 OD survey)	SL
(Miller and O'Kelly, 2016)	Travel data assessment Station and internal TAZ	Trip length, EE IE/EI analysis (EE) (SL, ATRI, ODOT data) Percentage error	SL
(StreetLight Data, 2019)	Validation of SL 2018 Data	AADT of SL in 2017 and 2018, AADT between SL and diverse DOT data MAPE, RMSE	NA

Reference	Purpose	Data Quality Evaluation	Data Fusion
(Roll, 2019)	Evaluation SL estimated of AADT in Oregon	Compared short term based AADT and automatic traffic recorder between SL and ODOT data Percentage error, APE	NA
(Avner, 2018)	Travel demand models	Interval vs external flow (MWCOG, SL) External trip distribution (AirSage, Regional Model, SL, NHTS 2009 Add On) Percentage error	SL
(Picado, 2017)	Travel trip OD travel demand	Scale factor calibration based on FDOT AADT in 2015 and 2016 Percentage error	SL
(SANDAG, 2018)	Truck flow modeling and visualization	NA	SL
(Kimley-Horn, 2018)	Travel demand model	Trip pattern +OD pattern in different time period (SL, MWCOG, traffic counts) Percentage error	NA

Note: MAPE: mean absolute percent error MAD: mean absolute difference MSD: mean signed difference APE: absolute percentage error.

Users and Non-User Surveys

The research team at ODU conducted a survey for both users and non-users of SL metrics in Spring 2019. The survey URLs were sent to 201 potential SL data users. A total of 32 existing SL users responded to the survey and 16 non-users of SL completed the survey, for a response rate of 23.9%. Below is a summary of the survey results for some key questions.

Survey Summary of Users

Background of the Surveyed Users

Respondents were asked to indicate in which state their work organization was located. Table 3 summarizes the results. The majority of respondents (84%) indicated that their work organization is located in Virginia. The 27 respondents whose work organization is located in Virginia were then asked which type of organization they work for. 30% indicated that they work for the VDOT District office, followed by 19% who work for a Metropolitan/Rural Planning Organization (MPO/RPO). Additionally, 15% of those respondents reported that they work for VTRC. Respondents who reported they worked outside of Virginia (n=5), indicated they work for a State DOT consultant (n=2), another organization outside of Virginia (n=2) or a MPO/RPO outside of Virginia (n=1).

Table 3. Affiliation of Surveyed Users

Respondents who reported they work in VA (N=27): **"*I work for…*"**	**% (n)**
VDOT District Office	29.6% (8)
A MPO/RPO (Metropolitan/Rural Planning Organization) in VA	18.5% (5)
VDOT Central Office TMPD	11.1% (3)
VTRC	14.8% (4)
A PDC (Planning District Commissions) in VA	7.4% (2)
A VDOT consultant	7.4% (2)
Other (including NVTA and University)	7.4% (2)
VDOT Central Office (other than TMPD)	3.7% (1)

Note: VDOT: Virginia Department of Transportation TMPD: Transportation and Mobility Planning Division NVTA: Northern Virginia Transportation Authority.

Focus Areas when Using SL Metrics

The focus areas when using SL metrics were asked and the results are shown in Figure 7. Traffic demand modeling was reported by 69% of respondents when asked to describe focus areas that are typically looked at when using SL metrics. 56% of respondents examine traffic congestion when using SL metrics and 25% of users focus on commercial traffic/freight/logistics SL data. Other focus areas include estimating the relative number of trips on certain roadways, parking demand traffic pattern analysis, analysis of traffic speed, volume, and trips.

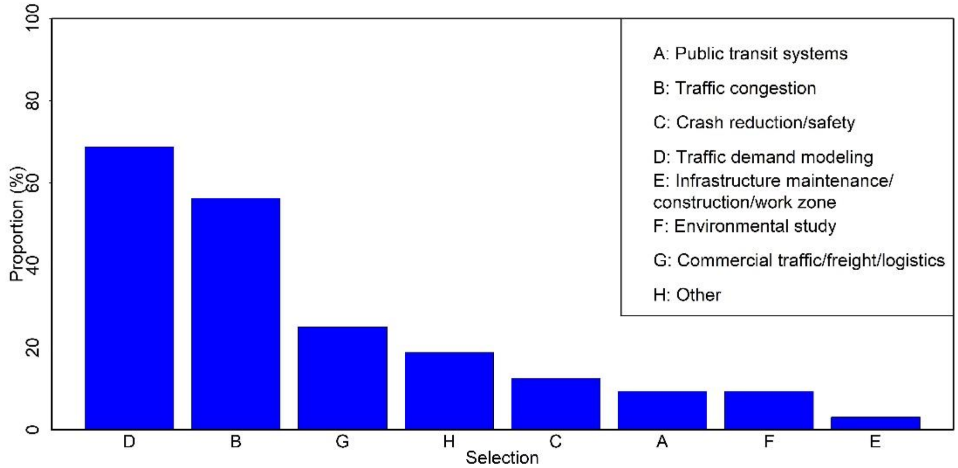


Figure 7. Selected Focus Areas When Using SL Metrics.

How SL Metrics Are Used in Typical Projects

Users were asked to think about the typical project(s) they perform and answer a series of questions regarding them. According to Table 4, 94% of users reported using SL metrics for OD analysis and 60% of users used SL metrics for traffic flow/volume analysis. Additionally, 44% used the SL metrics for route choice analysis. Respondents were also asked which major elements are included in typical projects that use SL metrics. Over three-quarters of users (78%) reported their typical projects include freeway and arterial data elements. An additional 66% include urban street elements and 40% examine TAZ/Census tract elements.

Table 4. Typical Application Scenarios

Considering your TYPICAL project(s), for which of the following specific tasks have you employed the StreetLight Data metrics?	**% (n)**
OD analysis	93.8% (30)
Traffic flow/volume analysis	59.4% (19)
Route choice analysis	43.8% (14)
Travel time analysis	34.4% (11)
Road speed analysis	28.1% (9)
Network analysis	28.1% (9)
Travel mode analysis	15.6% (5)
Attraction analysis	9.4% (3)
Mode choice analysis	9.4% (3)
Socioeconomic-factor/demographic analysis	9.4% (3)
Other (e.g., home location)	3.1% (1)

Used Functionalities of SL Platform

Users were asked what functionalities of the SL Platform were used by them and the responses are shown in Figure 8. 91% of respondents indicated that they have primarily used OD analysis when accessing the SL Platform. 34% have performed zone activity analysis as well as segment analysis and 25% use the functionality of visitor home and work analysis when using the SL Platform.

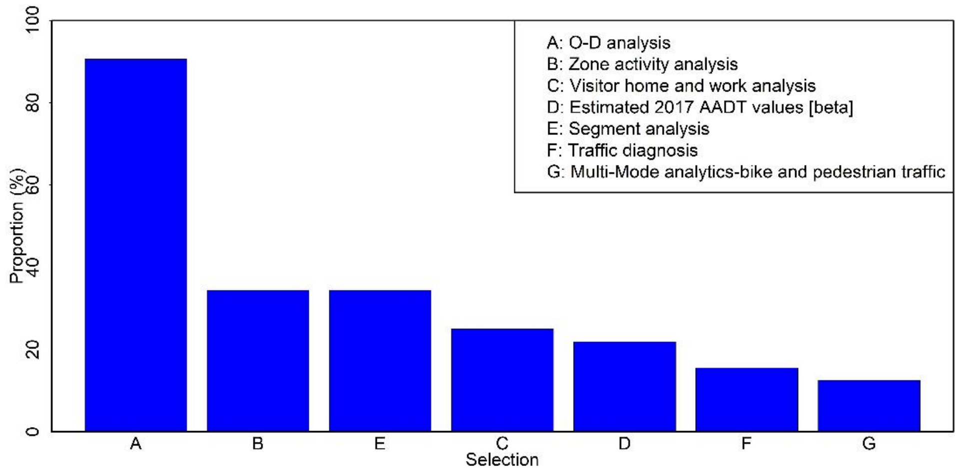


Figure 8. Functionalities of SL Platform Primarily Used in Typical Projects.

Geographic Areas and Spatial/Temporal Scales when Using SL Metrics?

We further explored the geographic locations of the projects that used the SL metrics and the spatiotemporal scales of the used metrics. The responses are presented in Table 5, Table 6 and Table 7, respectively. According to Table 5, the vast majority of the SL metrics used were located in an urban area (97%).

Table 5. Geographic Areas of the Typical Projects Using SL Metrics

The TYPICAL Project(s) that have used StreetLight Data metrics are primarily in a/an:	**% (n)**
Urban area	96.9% (31)
Rural area	43.8% (14)
Port(s)	9.4% (3)
Other (e.g., Restricted roads, toll roads, parkways and suburbs)	6.3% (2)
Airport(s)	3.1% (1)

Based on Table 6, at least 50% of respondents indicated that they required corridor and region/county level zone sets when using SL metrics (59% and 50% respectively).

Table 7 shows that 78% reported that daily data is the temporal scale they typically need for their projects and nearly 60% indicated they need hourly data for their typical projects. The majority of respondents (72%) indicated that they do not typically need SL metrics on a smaller level than an hour. 28% of users have had the need for data on an interval level smaller than an hour.

Table 6. Spatial Scale of the Used SL Metrics in the Typical Projects

What is the approximate spatial coverage level of your zone sets in your TYPICAL Project(s) using StreetLight Data metrics?	**% (n)**
Corridor level	59.4% (19)
Region/County level	50.0% (16)
Intersection level	46.9% (15)
Neighborhood level	31.3% (10)
City/Township level	31.3% (10)
State level	15.6% (5)
Other (e.g., Block groups and Tax level)	6.3% (2)

Table 7. Temporal Scale of the Used SL Metrics in the Typical Projects

What is the temporal scale of analysis primarily applied in your TYPICAL Project(s) that have used the StreetLight Data metrics?	**% (n)**
Daily	78.1% (25)
Hourly	59.4% (19)
Monthly	37.5% (12)
Weekly	34.4% (11)
Seasonal	28.1% (9)
Yearly	28.1% (9)

Users Satisfaction with the Sample Size of the SL Metrics in their Typical Projects

Respondents were then asked about their satisfaction with the sample size produced from SL metrics and Figure 9 shows that 65% of users were somewhat satisfied with sample size. Another 13% were somewhat dissatisfied and an additional 13% indicated that they do not check sample size. Only about 7% of users indicated that they were very satisfied with the sample size from SL metrics.

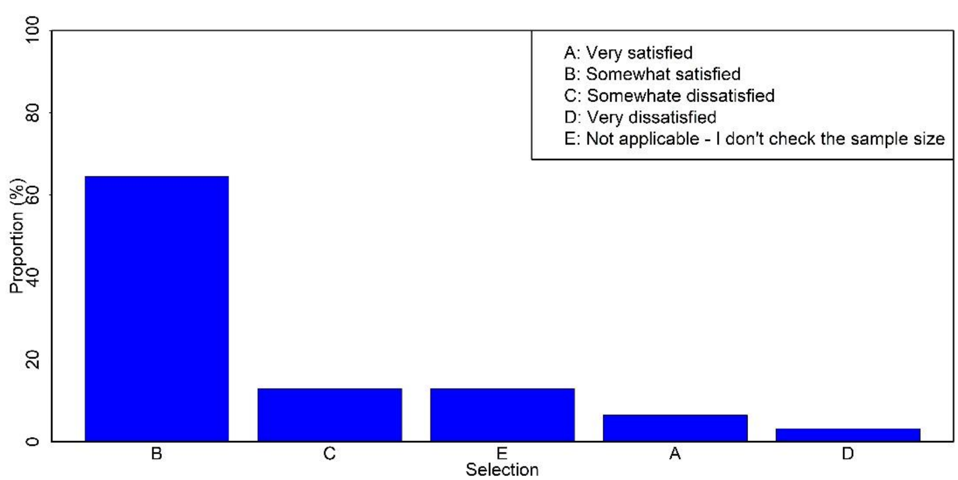


Figure 9. Satisfaction with SL Data Sample Size.

Frequency of Use of SL Metrics

In regard to frequency of use, Figure 10 shows that 50% of users utilize SL metrics less than once per month. 31% of users employ SL metrics 1-2 times per month. Only 3% of users use SL metrics 7-10 times per month.

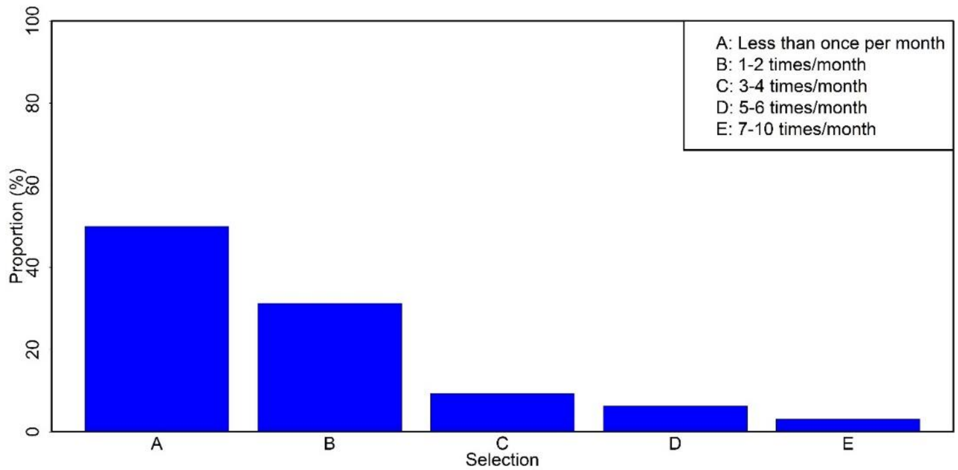


Figure 10. Monthly Frequency of Using SL Metrics.

Alternative Data Source Potential

The survey asked respondents to select applicable statements in regards to collecting data without SL products and results are summarized in Table 8 and Table 9. 44% of users answered that they would collect alternative data but it would be very time consuming and costly. Only 6% indicated they would do just fine without SL metrics for most of their projects because they can obtain alternate data sources relatively easily. Respondents were also asked if they agreed or disagreed that using SL metrics enables them to better perform their job responsibilities and 53% of users agreed, while only 6% of users disagreed that using SL metrics enables them to better perform their job responsibilities.

Table 8. Opinions on Data Sources Without SL Metrics

Without StreetLight Data products, which of the following statements would MOSTLY apply to you?	**% (n)**
I would collect alternative data but it would be very time consuming and costly	43.8% (14)
I would use other existing data but the results might not be as useful, reliable, or accurate	37.5% (12)
Other **(**i.e., would use other available data; educated guesses of experts and OD from travel demand model results, but not as reliable or accurate.**)**	12.5% (4)
I would do just fine WITHOUT StreetLight Data metrics for most of my projects (I can obtain alternative data sources relatively easily)	6.3% (2)

Table 9. Overall Opinion on the Use of the SL Metrics

Please indicate if you agree or disagree with the following statement: *"Using StreetLight Data metrics enables me to better perform my job responsibilities."*	**% (n)**
Agree	53.1% (17)
Strongly agree	31.3% (10)
Neither agree nor disagree	9.4% (3)
Disagree	6.3% (2)
Strongly disagree	0.0% (0)

Other Areas of Concerns and Challenges

Users were asked to share their concerns and challenges with SL metrics and the most frequently (9 out of 15 respondents) listed concerns were data inconsistencies and errors/quality of the data. Small sample sizes, additional training needed, not being user friendly, and needing additional data not provided by SL were also listed as concerns and challenges.

Survey Summary of Non-Users

Background of the Surveyed Non-users

Non-users were asked who they are primarily employed by in Virginia. Table 10 summarizes the results. 25% work for VDOT Central Office Transportation and Mobility Planning Division (TMPD) and 18% of respondents work for a VDOT District Office. The survey also asked respondents what their main job responsibility is at their respective agency. Half of the respondents (50%) are planners and 25% of respondents answered they have some other main job responsibility. Only 6% of non-users indicated they were a project manager. All non-users mentioned that they had heard of SL Platform; 75% of them have access to SL data; and 81% of them have considered possible use of SL metrics in projects.

Table 10. Background of the Surveyed Non-users

In Virginia I work for…	**% (n)**
VDOT Central Office TMPD	25.0% (4)
VDOT District Office	18.8% (3)
A MPO/RPO (Metropolitan/Rural Planning Organization) in VA	18.8% (3)
VTRC	12.5% (2)
A PDC (Planning District Commissions) in VA	12.5% (2)
Other (i.e., DRPT, University)	12.5% (2)

Note: VDOT: Virginia Department of Transportation TMPD: Transportation and Mobility Planning Division DRPT: Department of Rail & Public Transportation.

Reasons for Not Using SL metrics

The survey asked non-users about the main reasons they have not used SL metrics. Figure 11 shows that 38% of non-users indicated that they need more training before using the data and only 6% of non-users answered they find better alternative data sources. Over half of non-users (50%) reported they had not used SL metrics for other reasons (i.e., concerns about incomplete data, not very familiar, not needed, etc.).

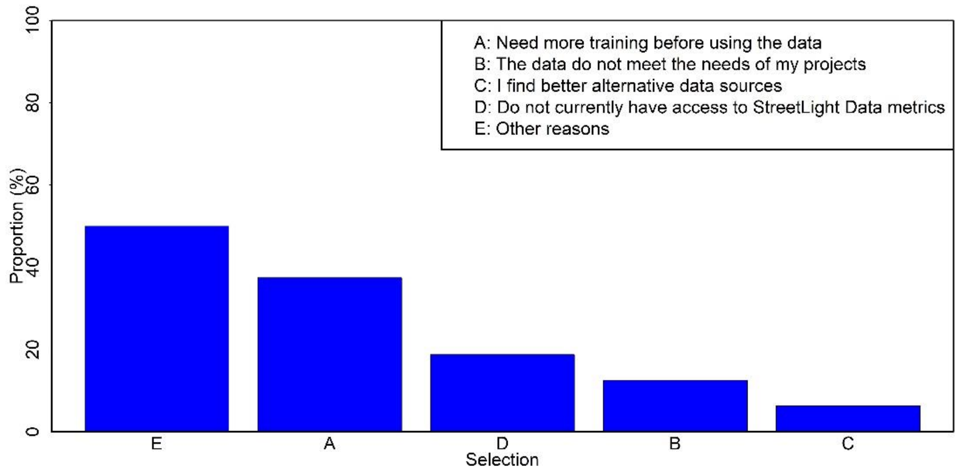
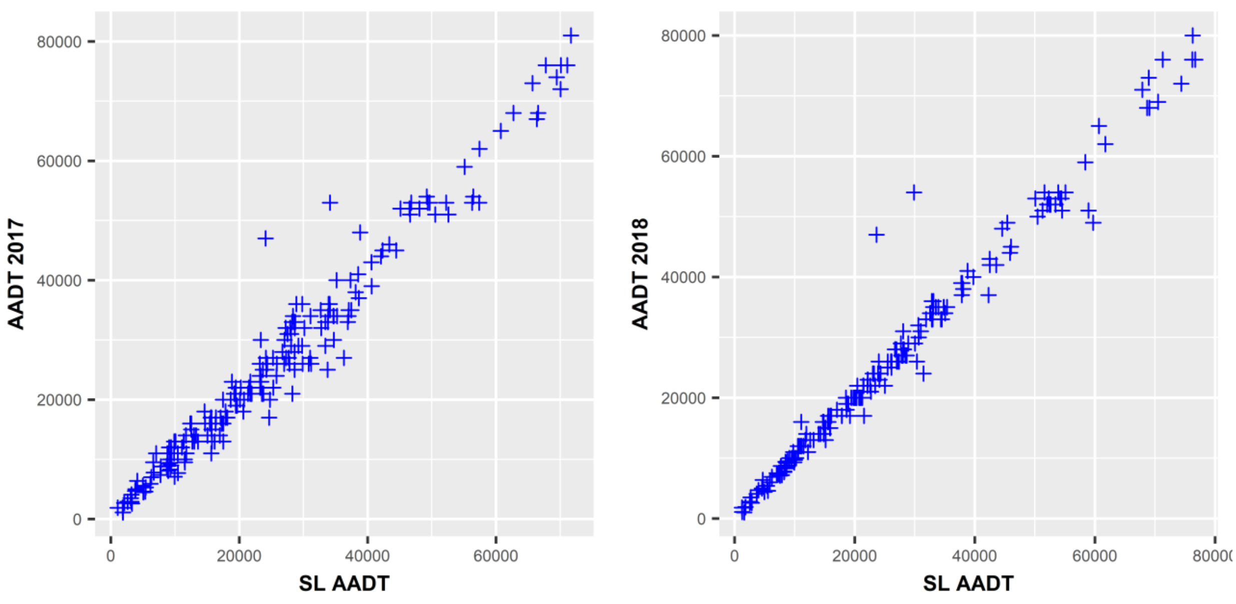


Figure 11. Main Reasons Non-Users Have Not Used SL Metrics.

Evaluation of SL Data

SL AADT Estimate vs. VDOT AADT

The VDOT AADT data were used as the benchmark to assess the quality of the SL AADT estimate. Figure 12 shows the relationship between the SL AADT estimate and the AADT published by VDOT for the calendar years 2017 and 2018. It can be seen that the SL AADT estimates in both years have a clear linear trend with the ones published by VDOT. In particular, the SL AADT estimate in 2018 shows a better linear relationship with the VDOT AADT as most of the points are well aligned along the diagonal of the scatter plot.



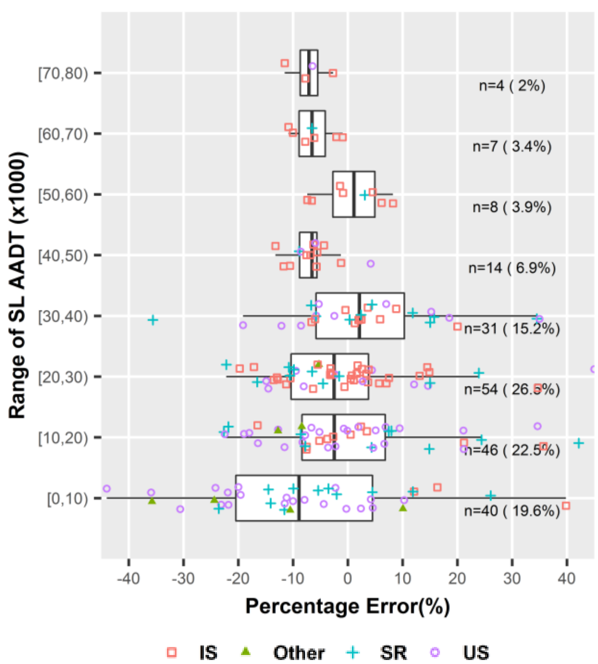
(a) 2017 SL AADT vs. VDOT AADT **(b) 2018 SL AADT vs. VDOT AADT**

Figure 12. Relationship between SL AADT Estimate and VDOT AADT.

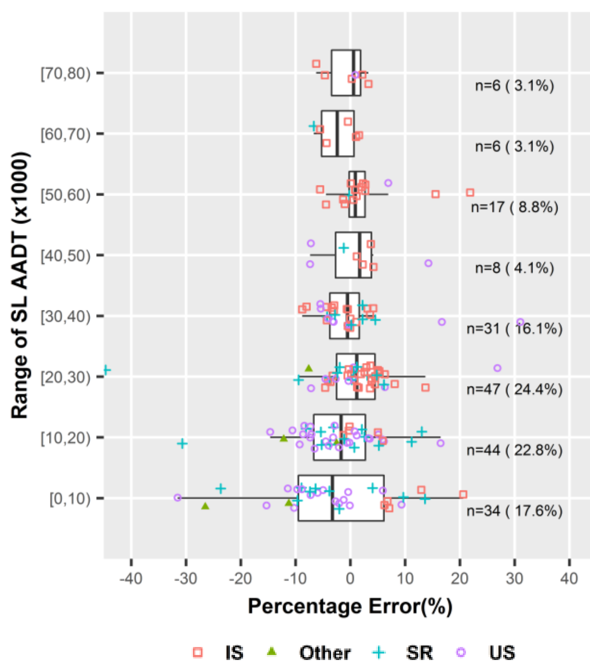
Note: AADT: Annual Average Daily Traffic.

The raw data shown in Figure 12 were used to calculate the performance measures percentage error (PE) and absolute percentage error (APE) for each year. The calculated results were organized in boxplots shown in Figure 13, which presents distributions of the calculated PE and APE according to the level of SL AADT estimate in each year. The four types of colored symbols in Figure 13 represent the type of each link (e.g., Interstate, State Route, etc.) described in the data preparation section. The texts to the right of each box represent the sample size and proportion of selected links in each level of SL AADT estimate.

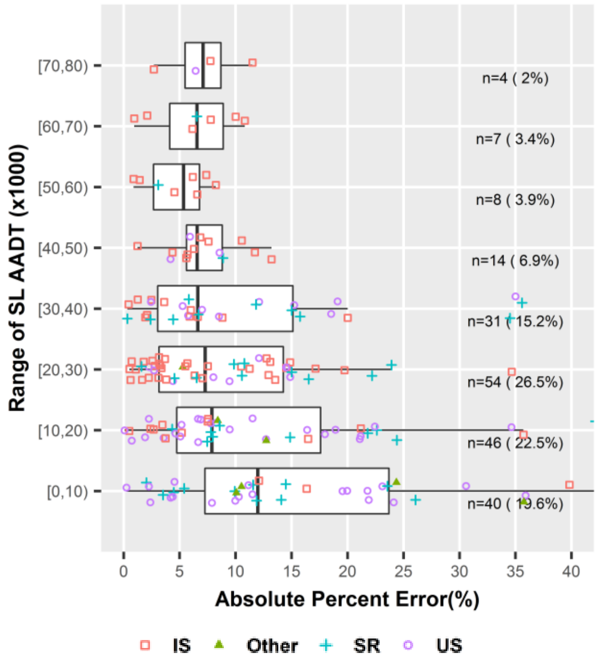
In Figure 13(a) and (b), the negative percentage errors mean that the SL AADT estimate is lower than the AADT published by VDOT and positive values mean that the SL AADT estimate is higher than the ones published by VDOT. Based on the two charts we can see that, overall, the percentage error decreases for links with higher SL AADT estimate. The percentage errors in both years have larger variance for the links with lower SL AADT estimate (e.g., 0-20,000 veh/day), and the majority of these links are the State Routes and/or US Routes. The variance of PE for the 2018 SL AADT estimate tends to be smaller than that of 2017. Most Interstate highways with SL AADT estimate over 20,000 veh/day are associated with PE between -10% and 10% in 2017 and between -5% and 5% in 2018. Figure 13(c) and (d) show the distributions of calculated APE for 2017 and 2018, respectively. Comparing these two graphs we can see that 2018 SL AADT estimate is better than 2017 SL AADT estimate because of its much lower APE. APEs for the majority of the links in 2018 are less than 5%. In contrast, there are still more links that have APEs greater than 10%, especially for those with SL AADT estimate below 30,000 veh/day.



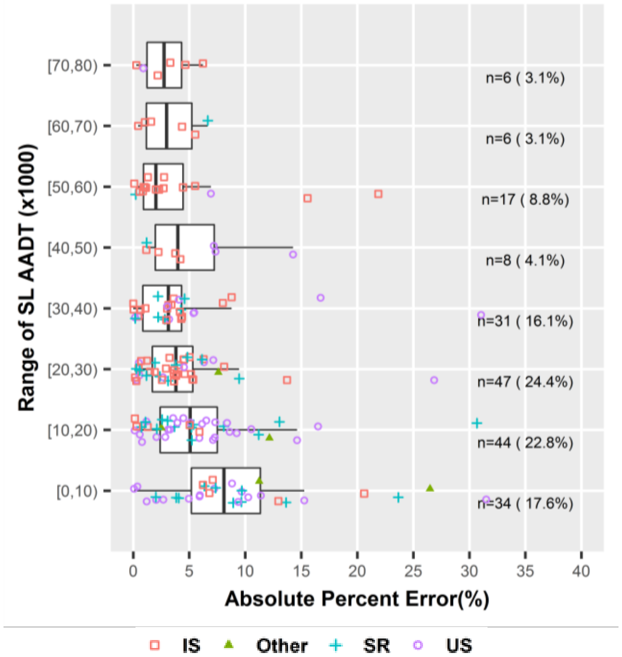
(a) PE Distribution for 2017 SL AADT



(b) PE Distribution for 2018 SL AADT



(c) APE Distribution for 2017 SL AADT



(d) APE Distribution for 2018 SL AADT

Figure 13. Performance of SL AADT Estimate in 2017 and 2018. Note: AADT: Annual Average Daily Traffic PE: Percentage Error APE: Absolute Percentage Error IS: Interstate type of roads SR: State Route type of roads US: U.S. type of roads Other: other type of roads.

Table 11 and Table 12 further summarize the median, mean, and 95% CI of APE for each level of SL AADT estimates in 2017 and 2018, respectively. The 0-10,000 range has the highest median APE with 12.0% in 2017 and 8.1% in 2018. For all the 204 links with 2017 AADT, the median and mean APE are 7.8% and 11.5%, respectively. Likewise, the median and mean APE for the 193 links with 2018 AADT are 4.06% and 6.18%, respectively. Samples with an estimated AADT over 40,000 veh/day have smaller ranges and a reduced upper limit of 95% CI, which imply a more stable APE for road links with higher levels of SL AADT estimates.

Table 11. Absolute Percent Error Summary by the Range of SL AADT (2017)

Range of SL AADT Estimate	Absolute Percent Error			Samples
	Median	Mean	95% CI	
70,000 - 80,000	7.1%	7.1%	[3.0%, 11.2%]	4
60,000 - 70,000	6.5%	6.3%	[1.1%, 10.7%]	7
50,000 - 60,000	5.4%	4.8%	[1.0%, 8.1%]	8
40,000 - 50,000	6.6%	7.2%	[2.2%, 12.7%]	14
30,000 - 40,000	6.6%	10.3%	[0.4%, 35.2%]	31
20,000 - 30,000	7.3%	10.2%	[0.7%, 41.6%]	54
10,000 - 20,000	7.9%	11.6%	[0.5%, 35.6%]	46
0 - 10,000	12.0%	18.2%	[2.0%, 69.8%]	40
All Sites	**7.8%**	**11.5%**	**[0.53%, 42.0%]**	**204**

Table 12. Absolute Percent Error Summary by the Range of SL AADT (2018)

Range of SL AADT Estimate	Absolute Percent Error			Samples
	Median	Mean	95% CI	
70,000 - 80,000	2.7%	2.9%	[0.4%, 6.0%]	6
60,000 - 70,000	3.0%	3.3%	[0.5%, 6.5%]	6
50,000 - 60,000	2.0%	4.1%	[0.1%, 19.4%]	17
40,000 - 50,000	4.0%	5.2%	[1.2%, 13.1%]	8
30,000 - 40,000	3.1%	4.3%	[0.1%, 20.3%]	31
20,000 - 30,000	3.8%	6.1%	[0.3%, 42.0%]	47
10,000 - 20,000	5.1%	5.9%	[0.2%, 16.4%]	44
0 - 10,000	8.1%	10.8%	[0.3%, 32.9%]	34
All Sites	**4.1%**	**6.2%**	**[0.2%, 30.8%]**	**193**

Table 13 further summarizes the APE without considering the road types. We can see that over 75% of the links have APEs below 15% in 2017, whereas over 92% of the links have APEs below 15% in 2018.

Table 13. Proportion of Road Links in Each Level of Absolute Percentage Error

Absolute Percentage Error	2017		2018	
	# Links	Proportion	# Links	Proportion
0~5%	62	30.4%	114	59.1%
5~10%	57	28.0%	52	26.9%
10~15%	36	17.6%	12	6.2%
15~20%	16	7.8%	4	2.1%
20%+	33	16.2%	11	5.7%
Total	**204**	**100.0%**	**193**	**100.0%**

In summary, the 2018 SL AADT estimate shows improved performance compared with the 2017 SL AADT estimate. These findings are consistent with the reported improvement efforts made by SL towards its AADT metrics (StreetLight Data, 2019) because the continuous sensors in VA were used for calibrating the SL AADT 2018.

For the comparisons that were performed above, it deserves mentioning two important facts when interpreting the comparative results. First, it should be noted that the published VDOT AADT data were rounded to the nearest 100 if the volume is under 10,000 veh/day and to the nearest 1,000 if over 10,000 veh/day. This may have some impact on the calculated error outcomes. For example, suppose both the SL AADT and the actual AADT are 10,400, due to the rounding issue, the actual AADT will be rounded to 10,000 and therefore the PE will be (10,400-10,000)/10,400≈3.8%. Likewise, if the actual AADT and SL AADT are 50,400, the PE will be (50,400-50,000)/50,400≈0.8%. The impact of rounded AADT on the error tends to be more noticeable in the lower AADTs, which may contribute in part to the higher error in the 10-30k AADT range. The more precise the comparison enables the unrounded AADT to exclude this impact.

Second, one would expect the percentage error will appear to be greater for lower AADTs when compared to higher AADTs. This is due to the nature of the equations used to calculate the performance measures: the actual AADT is used as the denominator and given a constant difference (e.g., difference =500) only its scale (e.g., 10,000 vs. 50,000) will affect the PE, for example, (10,500-10,000)/10,000=5% and (50,500-50,000)/50,000=1%. On the other hand, if both scenarios have the same PE (e.g., 5%), the actual volume deviation could be notably different. For example, (10,500-10,000)/10,000=5% and (52,500-50,000)/50,000=5%, indicating that the actual deviation (=2,500 veh/day) is 5 times higher for the high-volume condition.

StreetLight Index vs. Actual OD Trips Derived from Toll Transaction Data

The actual OD trips derived from toll transaction data and the corresponding SL Index were prepared for each hour. Let $Y_{m,j}^{dh}$ represent the observed trips from origin m to destination j for h^{th} hour on the data collection day d. For our studied scenario, there were 10 possible OD pairs in the Eastbound direction and 10 OD pairs in the Westbound direction. Two OD pairs were excluded from analysis: the trips leaving from zone 3130 Eastbound and the trips leaving from zone 3230 since these two only have one destination. This results in 18 OD pairs in total after combining both directions' data. Further, we only extracted the data collected on Tuesday, Wednesday, and Thursday for evaluation. This resulted in $d = 1, 2, ..., 15$ (Note: Only 15 days in May, 2018 are either Tuesday, Wednesday, or Thursday). In addition, $h = 1, 2, 3$ for ODs in Eastbound as there are 3 sampled toll operation hours on each weekday and $h = 1, 2, 3, 4$ for ODs in Westbound with 4 toll operation hours on each weekday. $Y_{m,j}^{dh}$ was obtained based on toll transaction data and $SL_{m,j}^{dh}$ was derived from the raw SL Index. These raw data were used to perform the following types of analysis:

- Type 1: Estimate $\hat{Y}_{m,j}^{dh}$ via Eqs. (1) and (2).; and compare the difference between $\hat{Y}_{m,j}^{dh}$ and $Y_{m,j}^{dh}$;
- Type 2: The observed hourly trips and raw SL Index were first aggregated to obtain the average of observed trips and SL Index for hour h over D days, respectively. Let $Y_{m,j}^{h}$ and $SL_{m,j}^{h}$ be the averaged results. Then we will estimate $\hat{Y}_{m,j}^{h}$ based on $Y_{m,j}^{h}$ and $SL_{m,j}^{h}$ and compare the difference between $\hat{Y}_{m,j}^{h}$ and $Y_{m,j}^{h}$. In this case study, both $D = 5$ and $D = 15$ were considered. Since we have 15 days for analysis, if $D = 5$, we will obtain three samples (sample 1: average of days 1 to 5; sample 2: average of days 6 to 10; and sample 3: average of days 11 to 15) for each hour, which are represented by $Y_{m,j}^{h_n}$ $(n = 1, 2, 3)$ through the following calculation:

$$Y_{m,j}^{h_n} = \frac{1}{D} \times \sum_{d=1+D\times(n-1)}^{d=D\times n} Y_{m,j}^{dh} \qquad (\text{with } D = 5 \text{ and } n = 1, 2, 3) \tag{9}$$

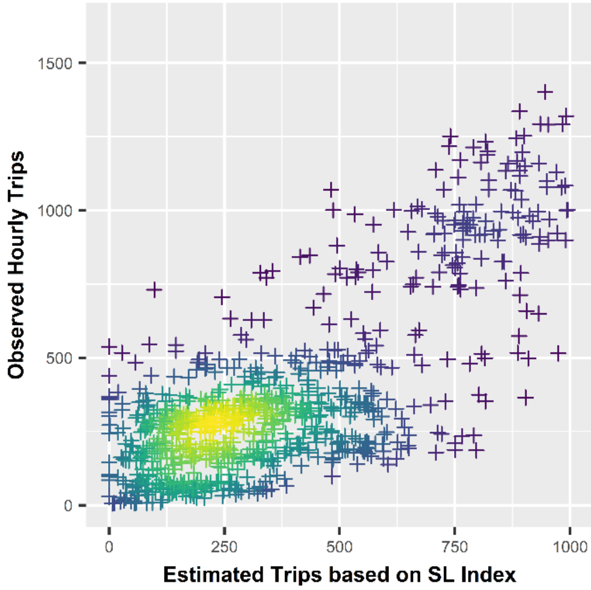
If $D = 15$, that means we will average the hourly observations $Y_{m,j}^{dh}$ for all the selected 15 days. Thus, we will obtain one sample for each hour $Y_{m,j}^{h_n}$ based on the following equation:

$$Y_{m,j}^{h_n} = \frac{1}{D} \times \sum_{d=1+D\times(n-1)}^{d=D\times n} Y_{m,j}^{dh} \qquad (\text{with } D = 15 \text{ and } n = 1) \tag{10}$$

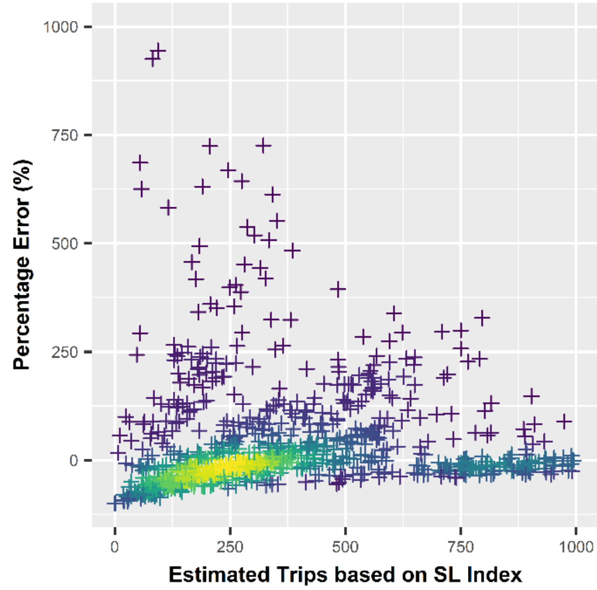
Likewise, $SL_{m,j}^{h_n}$ will be calculated using the equations similar to Eq. (9) or Eq. (10) accordingly; The calculated $Y_{m,j}^{h_n}$ and $SL_{m,j}^{h_n}$ will be used in Eqs. (1) and (2) to estimate the corresponding average hourly trips $\hat{Y}_{m,j}^{h_n}$. Finally, the difference between $\hat{Y}_{m,j}^{h_n}$ and $Y_{m,j}^{h_n}$ will be compared. Given $D = 5$ and $D = 15$, we denote the comparison as Type 2(A): $D = 5$ and Type 2(B): $D = 15$, respectively. Equivalently, Type 1 comparison is just a special case for $D = 1$ and $n = 1, 2, ..., 15$ (Type 1: $D = 1$).

The scatter plots of the observed trips among OD pairs and the corresponding estimates of trips are shown in Figure 14(a) , Figure 15(a), and Figure 16(a) based on the type of analysis. The light-colored area means that there were more observations clustered in that region. The estimated hourly trips based on SL Index were compared with the actual observations to calculate the performance measures (e.g., PE, APE). Likewise, Figure 14(b), Figure 15(b), and Figure 16(b) demonstrate the final calculated results of the estimated hourly trips based on the SL Index versus their associated percentage errors in each type of analysis. We can see that the percentage error has a very large variance if the estimated hourly trips were below 500.

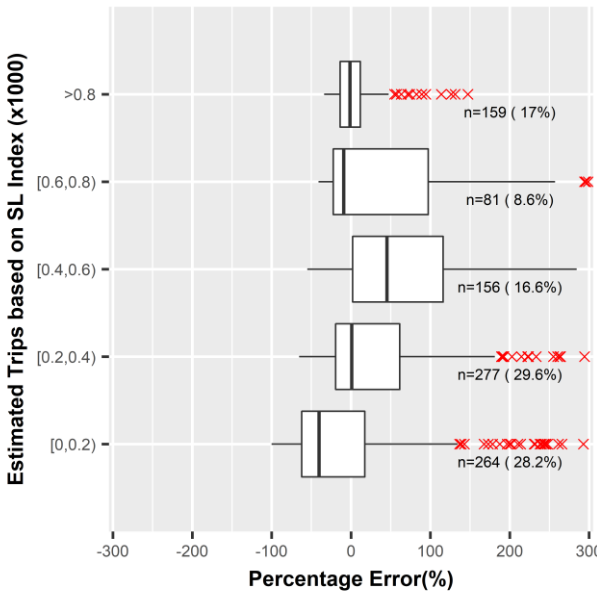
For each type of analysis, charts (c) and (d) in Figure 14, Figure 15, or Figure 16 show the distributions of PE and APE, respectively. These results show that if the OD trips were estimated based on the SL Index, the distributions of PE and APE both are not consistent with the changes of estimated hourly trips. The variation of the errors for those with estimated hourly trips above 800 is less and the overall magnitude of errors for this group is lower. More importantly, comparing the results of the three types of analysis, it is shown that the error based on the average of multiple days' SL Index and the corresponding average of actual observations are less variant and the errors tend to be smaller if the average over more days' data were used. Figure 16's result should be used with caution as its sample size is not large in some ranges of the estimated trips. Overall, it would be better to use the averaged index as the input for estimating OD trips. Estimation for the daily trips between the zones is not recommended.



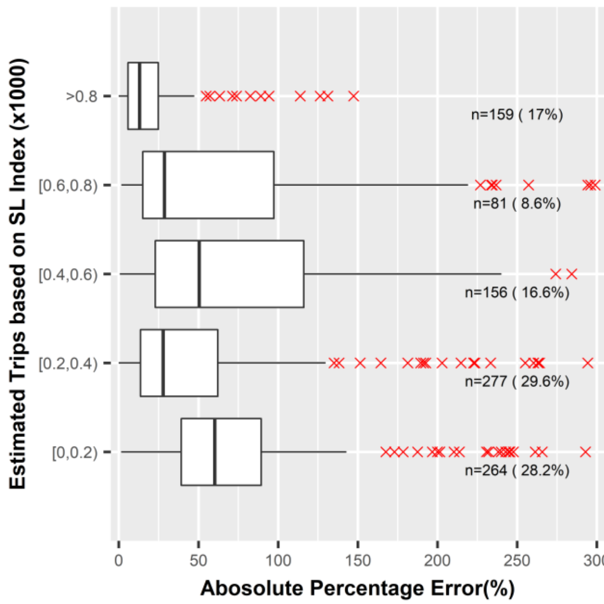
(a) Estimated Trips vs. Observed Trips



(b) Percentage Error vs. Estimated Trips



(c) Distribution of Percentage Error



(d) Distribution of Absolute Percentage Error

Figure 14. Type 1 Comparison: Estimated Trips based on the Original SL Index.

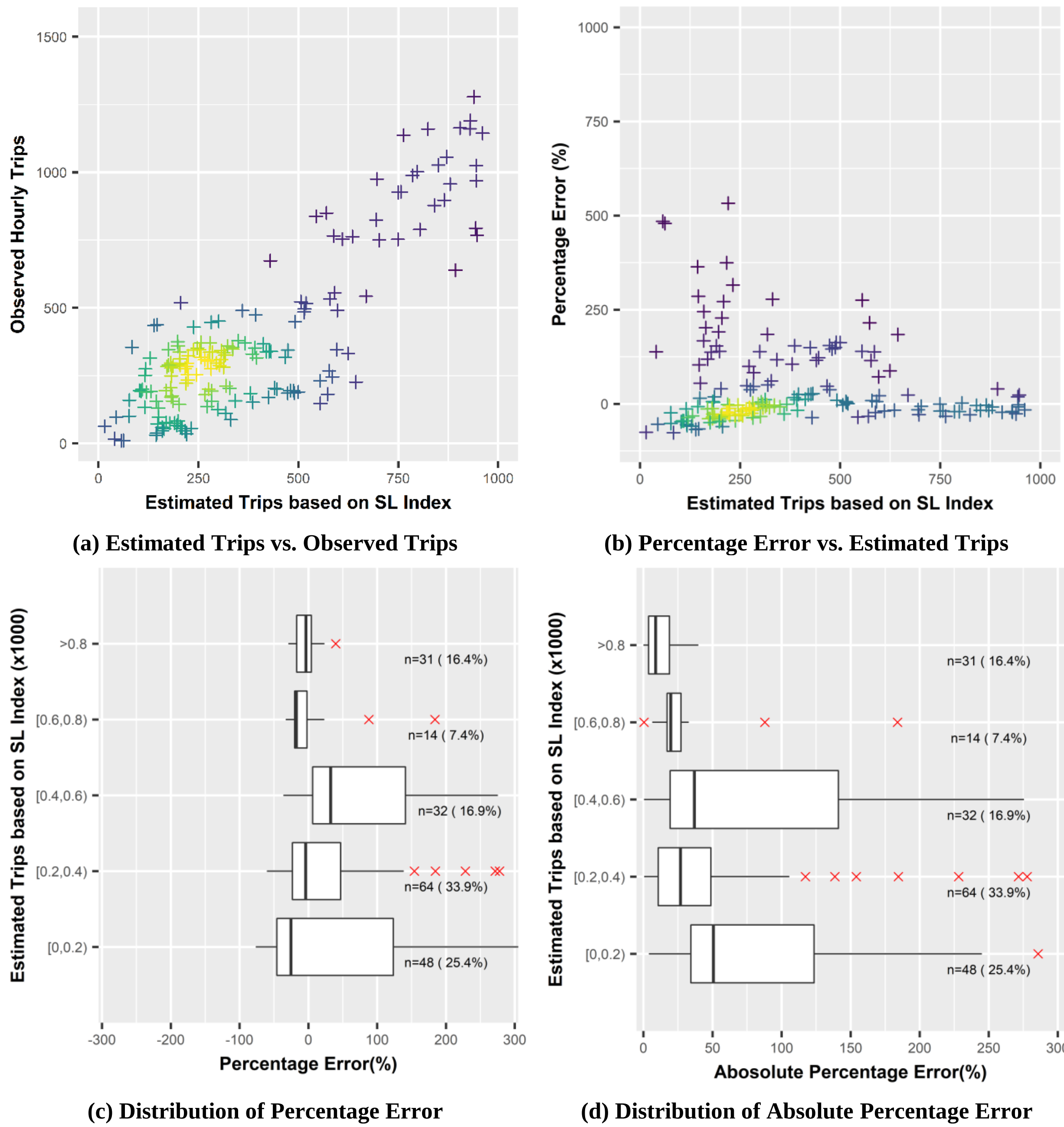
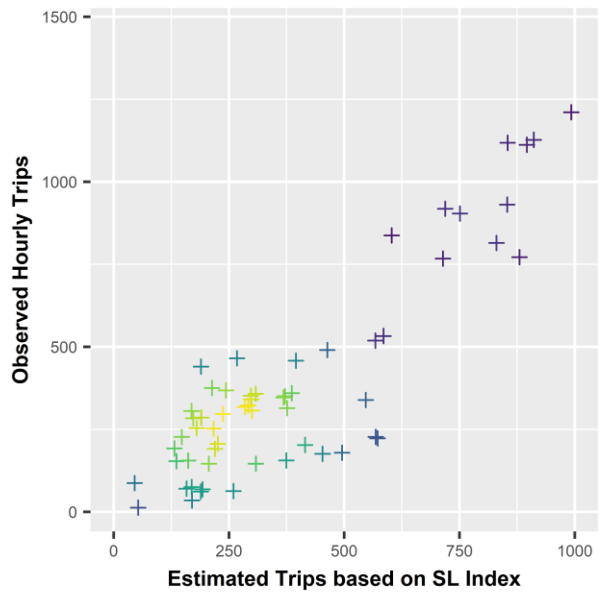
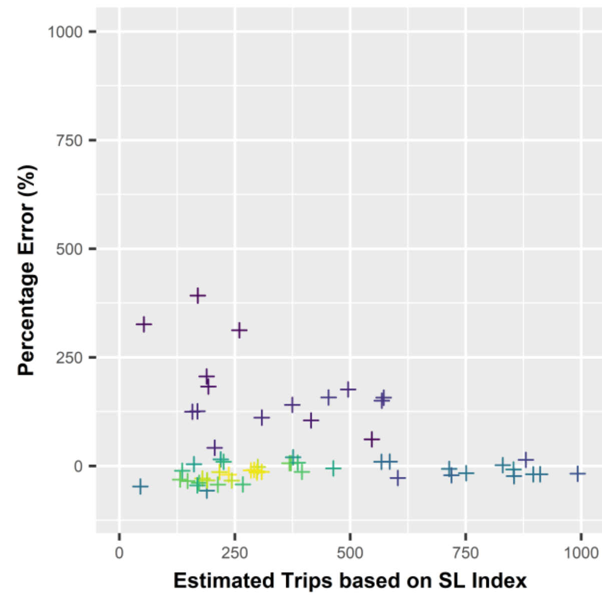


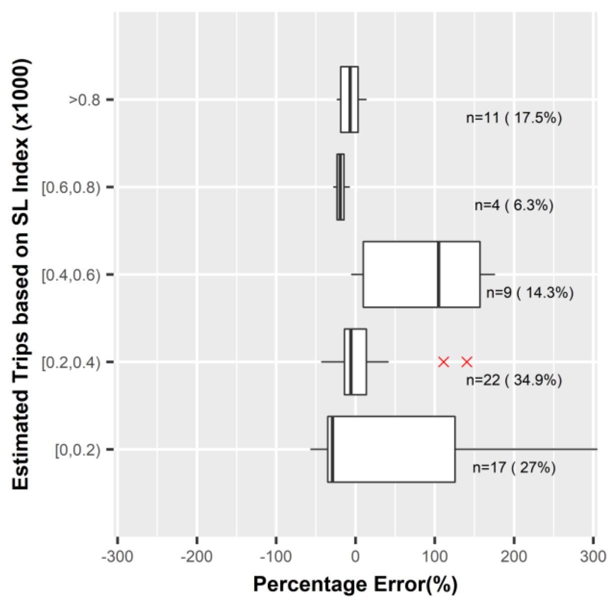
Figure 15. Type 2(A) Comparison: Estimated Trips based on 5-day Average SL Index.



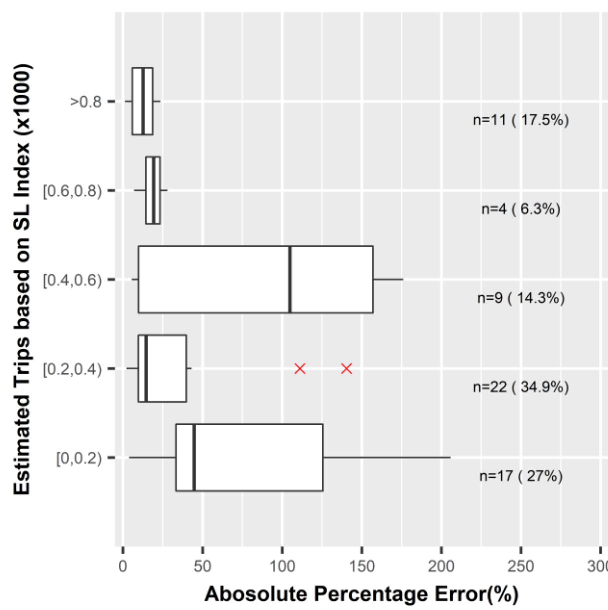
(a) Estimated Trips vs. Observed Trips



(b) Percentage Error vs. Estimated Trips



(c) Distribution of Percentage Error



(d) Distribution of Absolute Percentage Error

Figure 16. Type 2(B) Comparison: Estimated Trips based on 15-day Average SL Index.

Table 14 summarizes the APE based on the range of estimated trips based on SL Index. The majority of the levels for median, mean, and 95% CI were improved when multi-day averages were considered in the estimation. The median and mean APEs also decreased when the level of estimated trips increased. Table 15 further examines the APE without considering the levels of estimated OD trips. We can see that more than 45% of the samples have an APE over 40% and about 32% of the samples have an APE less than 20% in the Type 1 ($D=1$: No average over days) analysis. In contrast, about 39% and 49% of the samples have an APE less than 20% in the 5-day and 15-day average scenarios, respectively. The proportion of samples with an APE over 40% were also reduced below 36% in these two types of analysis. Since the 15-day average scenario has a smaller sample size, the results for some cases may not be very reliable. However, every level of estimated trips in the 5-day average scenario has 10 or more samples. Its results will be more reasonable compared to that of the 15-day average scenario. With similar aggregation, we can expect that the median APE for most will be likely to be below 20% if the estimated trips were above 600. The zones with a higher estimated number of trips (above 800 vph) tend to have smaller APEs.

Table 14. Absolute Percent Error Summary by the Range of Estimated Trips based on SL Index

Aggregation Over D Days	Range of Estimated Trips	Absolute Percent Error			Samples
		Median	Mean	95% CI	
D=1	>800	13.0%	20.3%	[0.3%, 95.2%]	159
	600 – 800	28.6%	72.4%	[3.3%, 298.9%]	81
	400 – 600	50.5%	75.8%	[1.4%, 232.6%]	156
	200 – 400	27.8%	78.8%	[1.4%, 520.3%]	277
	0 – 200	60.3%	94.3%	[7.0%, 531.3%]	264
	All Samples	**35.4%**	**72.2%**	**[1.4%, 401.5%]**	937
D=5	>800	8.8%	12.4%	[0.2%, 31.7%]	31
	600 – 800	19.7%	35.3%	[2.4%, 152.7%]	14
	400 – 600	36.9%	77.5%	[2.5%, 228.7%]	32
	200 – 400	26.8%	62.4%	[1.0%, 340.8%]	64
	0 – 200	50.6%	94.6%	[6.4%, 458.9%]	48
	All Samples	**27.2%**	**62.9%**	**[0.8%, 330.2%]**	**189**
D=15	>800	12.6%	11.8%	[1.5%, 22.5%]	11
	600 – 800	19.3%	18.4%	[7.7%, 27.5%]	4
	400 – 600	104.8%	92.4%	[6.3%, 172.3%]	9
	200 – 400	14.5%	40.9%	[4.3%, 222.1%]	22
	0 – 200	44.7%	101.3%	[6.9%, 365.7%]	17
	All Samples	**20.0%**	**58.08%**	**[2.1%, 318.5%]**	63

Table 15. Proportion of Sampled ODs in Each Level of APE

Aggregation Over D Days	APE	# OD Samples	Proportion
D=1	0~10%	160	17.1%
	10~20%	148	15.7%
	20~30%	114	12.2%
	30~40%	86	9.2%
	40~50%	59	6.3%
	50%+	370	39.5%
	Total	**937**	**100.0%**
D=5	0~10%	42	22.2%
	10~20%	32	17.0%
	20~30%	25	13.2%
	30~40%	23	12.2%
	40~50%	11	5.8%
	50%+	56	29.6%
	Total	**189**	**100.0%**
D=15	0~10%	16	25.4%
	10~20%	15	23.9%
	20~30%	5	7.9%
	30~40%	6	9.5%
	40%~50%	5	7.9%
	50%+	16	25.4%
	Total	**63**	**100.0%**

StreetLight Index vs. Traffic Counts from the City of Virginia Beach

The hourly traffic count data were obtained for 40 sites in Virginia Beach. Let Y_i^{dh} represent the actual volume collected at h^{th} hour on day d at the site i and SL_i^{dh} be the corresponding SL Index. Further, we only extracted the data collected on Tuesday, Wednesday, and Thursday for evaluation. This resulted in $d = 1, 2, ..., 12$ (Note: Only 12 days in April 2018 are either Tuesday, Wednesday, or Thursday). In addition, data were collected between 6am and 8pm on each day for each site. Therefore, there are 14 hours of data for a day and $h = 1, 2, ..., 14$. However, the data for some hours were not available for some sites. If the data were not available for a given period over 12 days, they were not included in the analysis. If the data were only available for a given period over several days, they were used in analysis and the max of h was set to the number of available days for that site. This resulted in 6,710 samples (based on all available SL_i^{dh}) in total. These raw data were used to perform the following types of analysis:

- Type 1: Estimate $\hat{Y}_i^{dh}$ through the Eqs. (5) and (6); and compare the difference between $\hat{Y}_i^{dh}$ and Y_i^{dh}; and
- Type 2: The observed hourly volume and raw SL Index were first aggregated to obtain the average of observed volume and SL Index for hour h over D days, respectively. Let Y_i^h and SL_i^h be the averaged results. Then we estimate $\hat{Y}_i^h$ through the Eqs. (5) and (6) and compare it with Y_i^h. More specifically, in this case study both $D = 5$ and $D = 12$ were considered. Since we have 12 selected weekdays for analysis, if $D = 5$, we will obtain two samples (sample 1: the average of days 1 to 5; and sample 2: the average of days 6 to 10) for each hour. These two samples are denoted as $Y_i^{h_n}$ $(n = 1, 2)$ and computed based on the following equation:

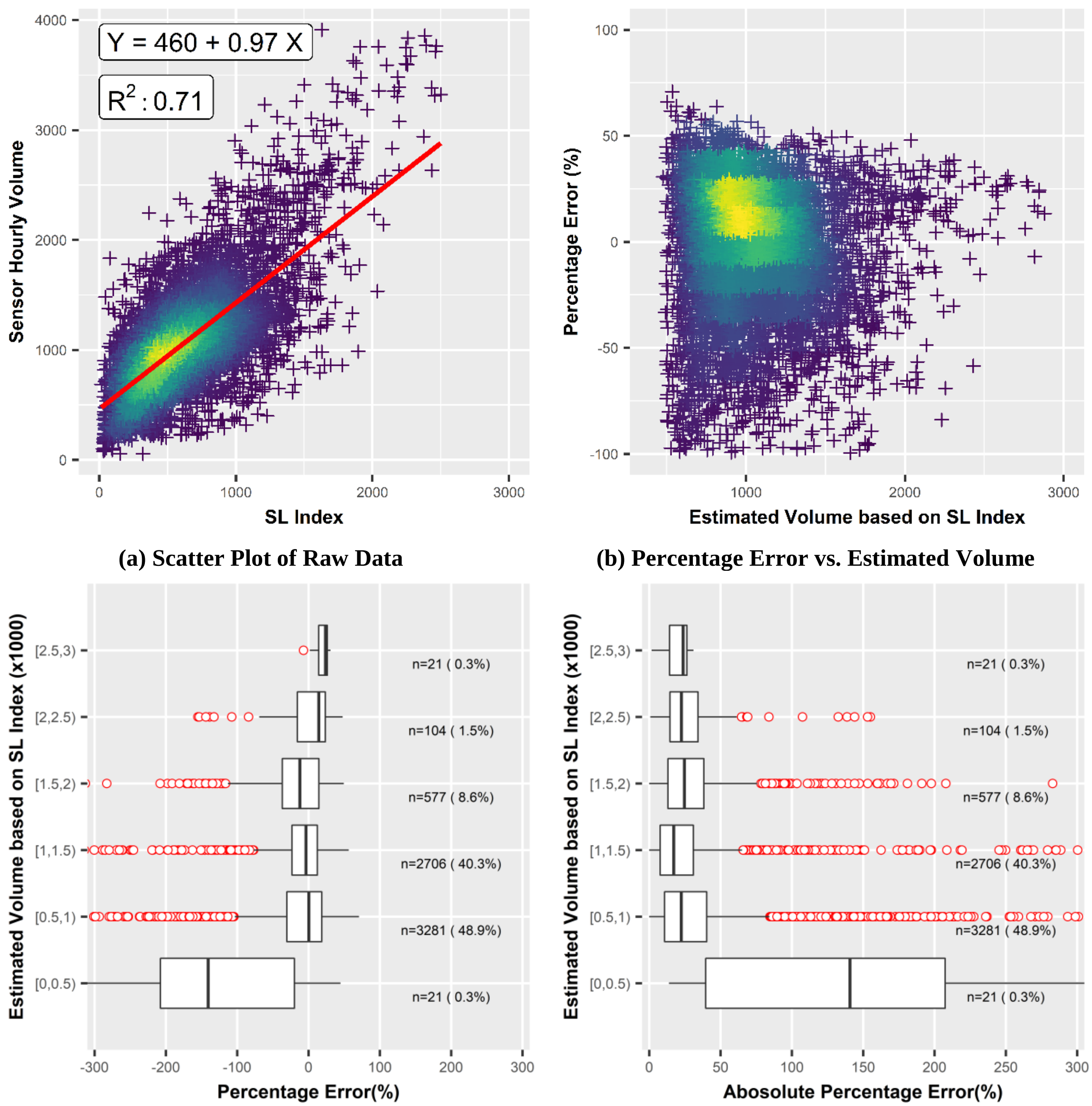
$$Y_i^{h_n} = \frac{1}{D} \times \sum_{d=1+D\times(n-1)}^{d=D\times n} Y_i^{dh} \quad (\text{with } D = 5 \text{ and } n = 1, 2) \tag{11}$$

If $D = 12$, that means we will average the hourly observations Y_i^{dh} for all the selected 12 days. Thus, we will obtain only one sample for each hour $Y_i^{h_n}$ $(n = 1)$ based on the following equation:

$$Y_i^{h_n} = \frac{1}{D} \times \sum_{d=1+D\times(n-1)}^{d=D\times n} Y_i^{dh} \quad (\text{with } D = 12 \text{ and } n = 1) \tag{12}$$

Likewise, $SL_i^{h_n}$ will be calculated using the equations similar to Eq. (9) or Eq. (10) accordingly; The calculated $Y_i^{h_n}$ and $SL_i^{h_n}$ will be used in Eqs. (5) and (6); to estimate the corresponding average hourly volume $\hat{Y}_i^{h_n}$. Finally, the difference between $\hat{Y}_i^{h_n}$ and $Y_i^{h_n}$ will be compared. Given $D = 5$ and $D = 12$, we denote the comparison as Type 2(A): $D = 5$ and Type 2(B): $D = 12$, respectively. Equivalently, Type 1 comparison is just a special case for $D = 1$ and $n = 1, 2, ..., 12$ (Type 1: $D = 1$).

The scatter plot of the SL Index and the observed hourly volume is shown in Figure 17(a). The light-colored area means that there were more observations clustered in that region. The two sets of data were used to establish the conversion model shown as the line inFigure 17(a) $Y = 460 + 0.97X$. The estimated hourly volumes based on SL Index were compared with the actual traffic count to quantify the performance measures. Figure 17(b) shows the final calculated results of estimated hourly volume based on the SL Index versus their associated percentage errors. We can see that the percentage error can vary for most of the estimated volumes. For example, when the estimated volume is less than 1,000 vph, the possible PE can range between -100% and 75%. Figure 17(c) and (d) show the distributions of PE and APE, respectively. These results show that if the traffic volumes were estimated based on the original SL Index, the distributions of their PE and APE both are centered for the cases with volume greater than 500 vph and the values tend to be smaller when the estimated volume is higher. For the cases with an estimated volume below 500 vph, the associated errors have a large variation.

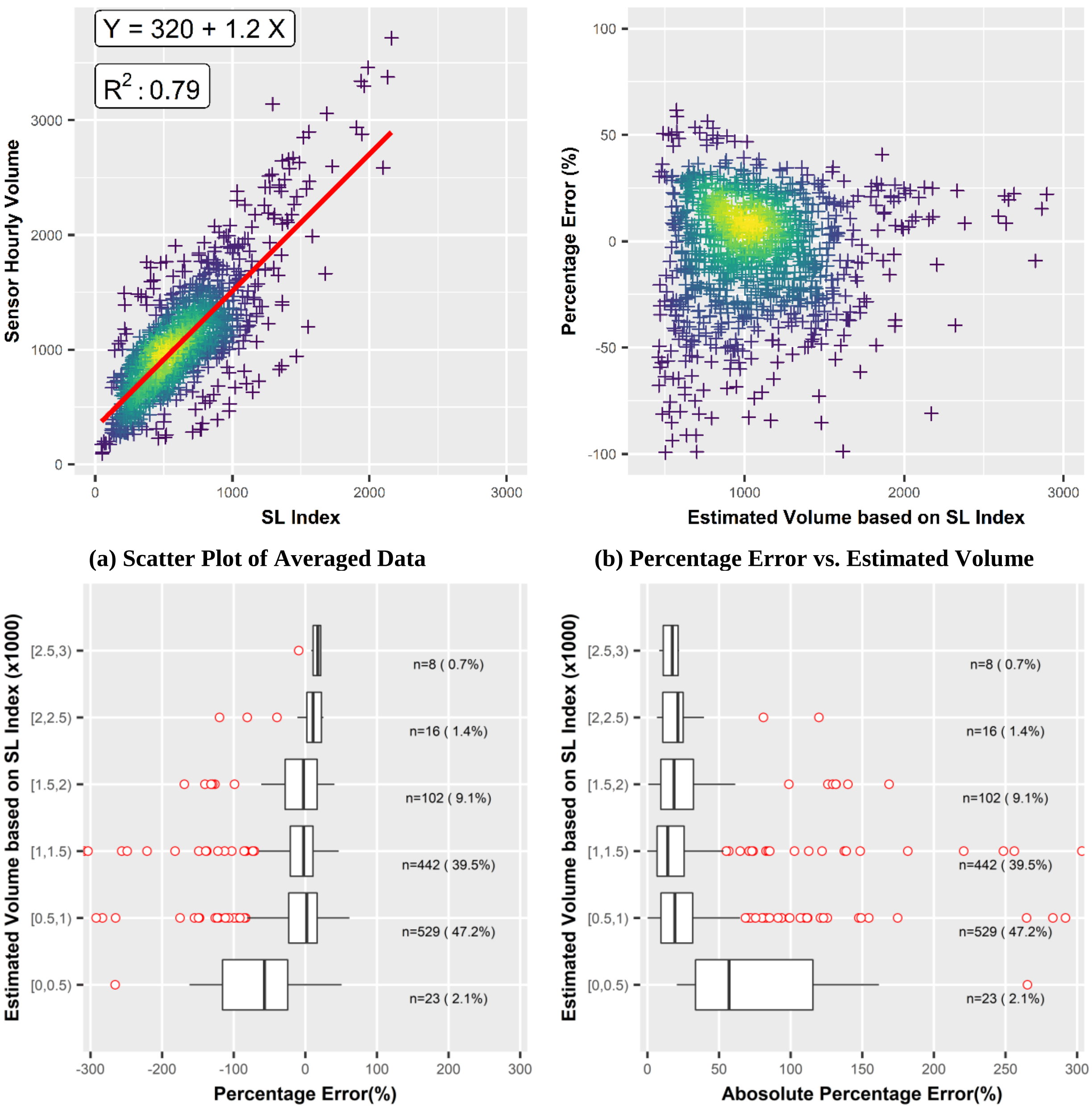


(a) Scatter Plot of Raw Data **(b) Percentage Error vs. Estimated Volume**

(c) Distribution of Percentage Error **(d) Distribution of APE**

Figure 17. Type 1 Comparison: Estimated Volume based on Original SL Index for Roads in Virginia Beach.

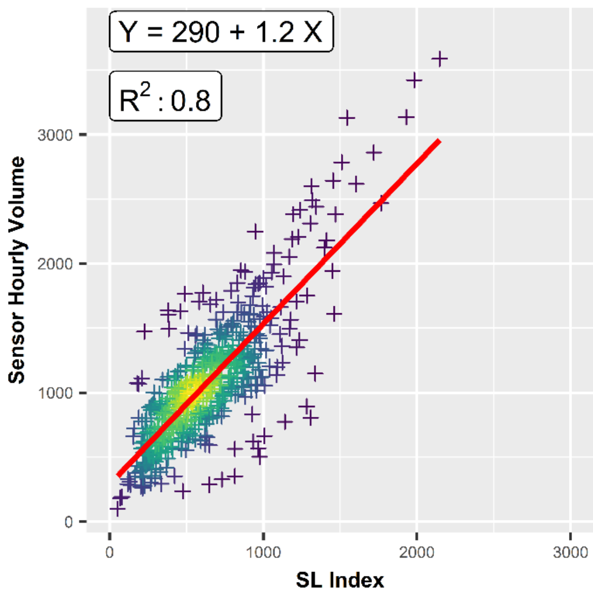
Instead of using the raw SL Index to estimate traffic volume, Figure 18 and Figure 19 show the comparative results based on the average SL Index over 5 days and 12 days, respectively. Compared to the results in Figure 17, it is clear that the PE and APE both have smaller variations with the use of the average SL Index. The values for APE tend to be lower. This suggests the benefits of using the average SL Index as the variable to estimate traffic volume. In real-world projects, the average data of a study site over multiple time periods are often preferred to the instance of a single value of individual periods. This is to make sure that the typical conditions of the site will be well represented.



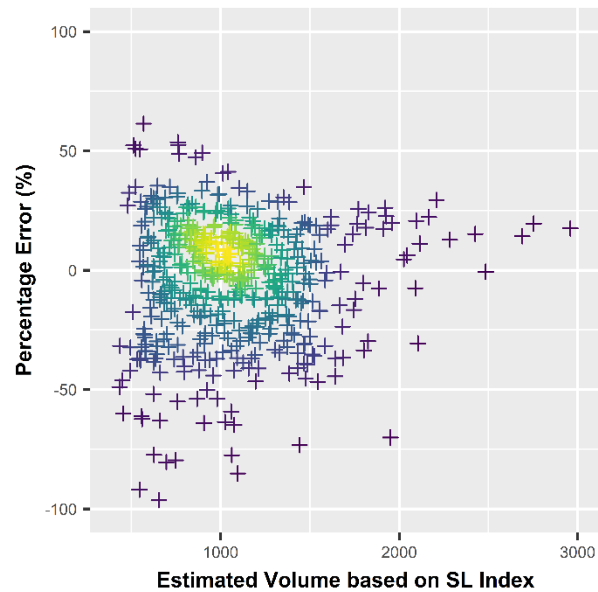
(a) Scatter Plot of Averaged Data **(b) Percentage Error vs. Estimated Volume**

(c) Distribution of Percentage Error **(d) Distribution of APE**

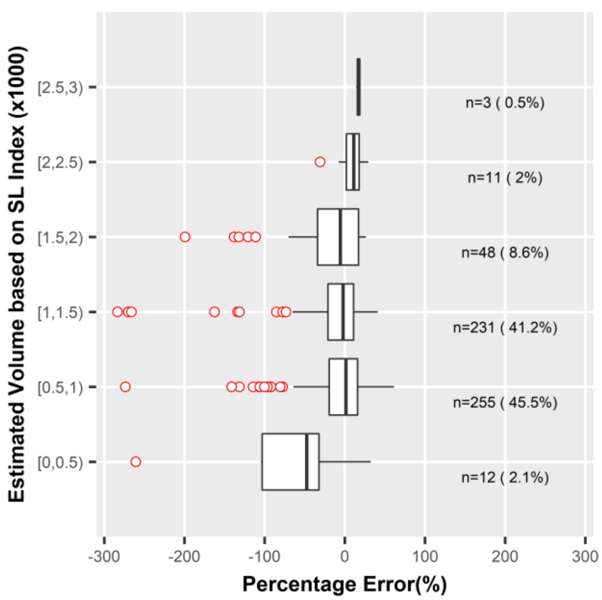
Figure 18. Type 2(A) Comparison: Estimated Volume based on 5-day Average SL Index for Roads in Virginia Beach.



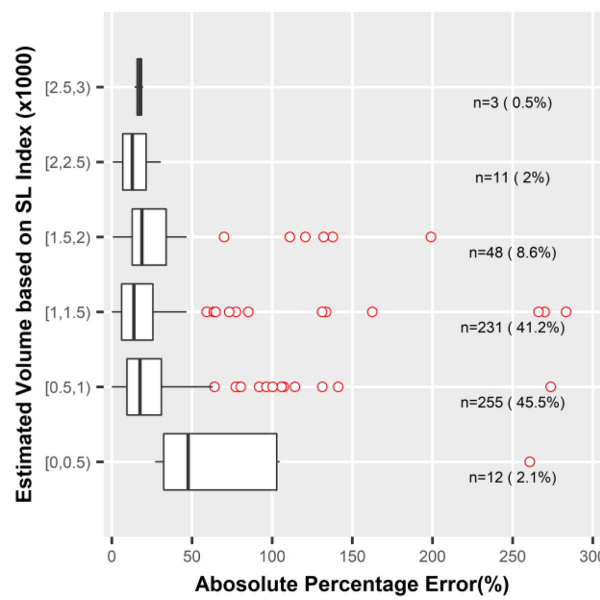
(a) Scatter Plot of Averaged Data



(b) Percentage Error vs. Estimated Volume



(c) Distribution of Percentage Error



(d) Distribution of APE

Figure 19. Type 2(B) Comparison: Estimated Volume based on 15-day Average SL Index for Roads in Virginia Beach.

Table 16 presents the median, mean, and 95% CI of the APE according to the range of estimated volume based on the SL Index. We can see that the median APEs for the estimated volumes of 500 vph or more are all lower than 25%. In particular, the use of average SL Index helped reduce the median APEs, most of which became less than 20%. For those below 500 vph, their median APE is relatively high but can be reduced to 48% if the average SL Index was used. The estimates based on the SL Index are less accurate at the low level of estimated volume and should be used with caution in projects. With an aggregation of 12-day data, most of 95% CIs were reduced in terms of the ranges and upper limits.

Table 17 further examines the APE without considering the levels of estimated volumes. Without using an average SL Index, we can see that less than 45% of the sampled cases with estimated trips having APE below 20%. About 18% of the sampled cases with estimated trips have an APE greater than 50%. In contrast, the analyses in cases when $D=5$ and $D=12$ used average SL Index to estimate volume and over 55% of the sampled cases now have an APE below 20%. Rather than using the original SL Index to estimate short-term traffic, these results further highlight the benefits of using SL Index to estimate average traffic volume.

Table 16. Absolute Percentage Error Summary based on Data from Virginia Beach

Aggregation Over D Days	**Range of Estimated Volume**	**Absolute Percent Error**			**Samples**
		Median	**Mean**	**95% CI**	
D=1	2,500 – 3,000	23.7%	20.3%	[4.0%, 30.4%]	21
	2,000 – 2,500	22.6%	32.1%	[3.2%, 140.9%]	104
	1,500 – 2,000	24.6%	34.1%	[1.6%, 146%]	577
	1,000 – 1,500	17.1%	25.9%	[0.7%, 109.8%]	2,706
	500 -1,000	22.5%	34.9%	[1.2%, 157.4%]	3,281
	0 - 500	140.7%	165.2%	[14.0%, 497.4%]	21
	All Samples	**20.4%**	**31.51%**	**[0.9%, 146.3%]**	**6,710**
D=5	2,500 – 3,000	17.4%	16.2%	[8.5%, 22.2%]	8
	2,000 – 2,500	21.3%	28.1%	[7.4%, 105.2%]	16
	1,500 – 2,000	18.6%	27.4%	[1.4%, 131.6%]	102
	1,000 – 1,500	14.2%	22.6%	[0.5%, 112.4%]	442
	500 -1,000	19.2%	26.8%	[1.4%, 109.3%]	529
	0 - 500	57.1%	86.1%	[24.9%, 288.3%]	23
	All Samples	**17.7%**	**26.3%**	**[1%, 122.9%]**	**1,120**
D=12	2,500 – 3,000	17.5%	17.1%	[14.4%, 19.5%]	3
	2,000 – 2,500	12.8%	14.7%	[1.6%, 30.4%]	11
	1,500 – 2,000	18.8%	32.7%	[1.4%, 136.9%]	48
	1,000 – 1,500	13.8%	21.9%	[1.1%, 96.6%]	231
	500 -1,000	17.6%	24.4%	[1.1%, 98.9%]	255
	0 - 500	47.6%	74.4%	[28.4%, 217.8%]	12
	All Samples	**17.1%**	**25.0%**	**[1.0%, 114.4%]**	**560**

Table 17. Proportion of Samples in Each Level of APE based on Data from Virginia Beach

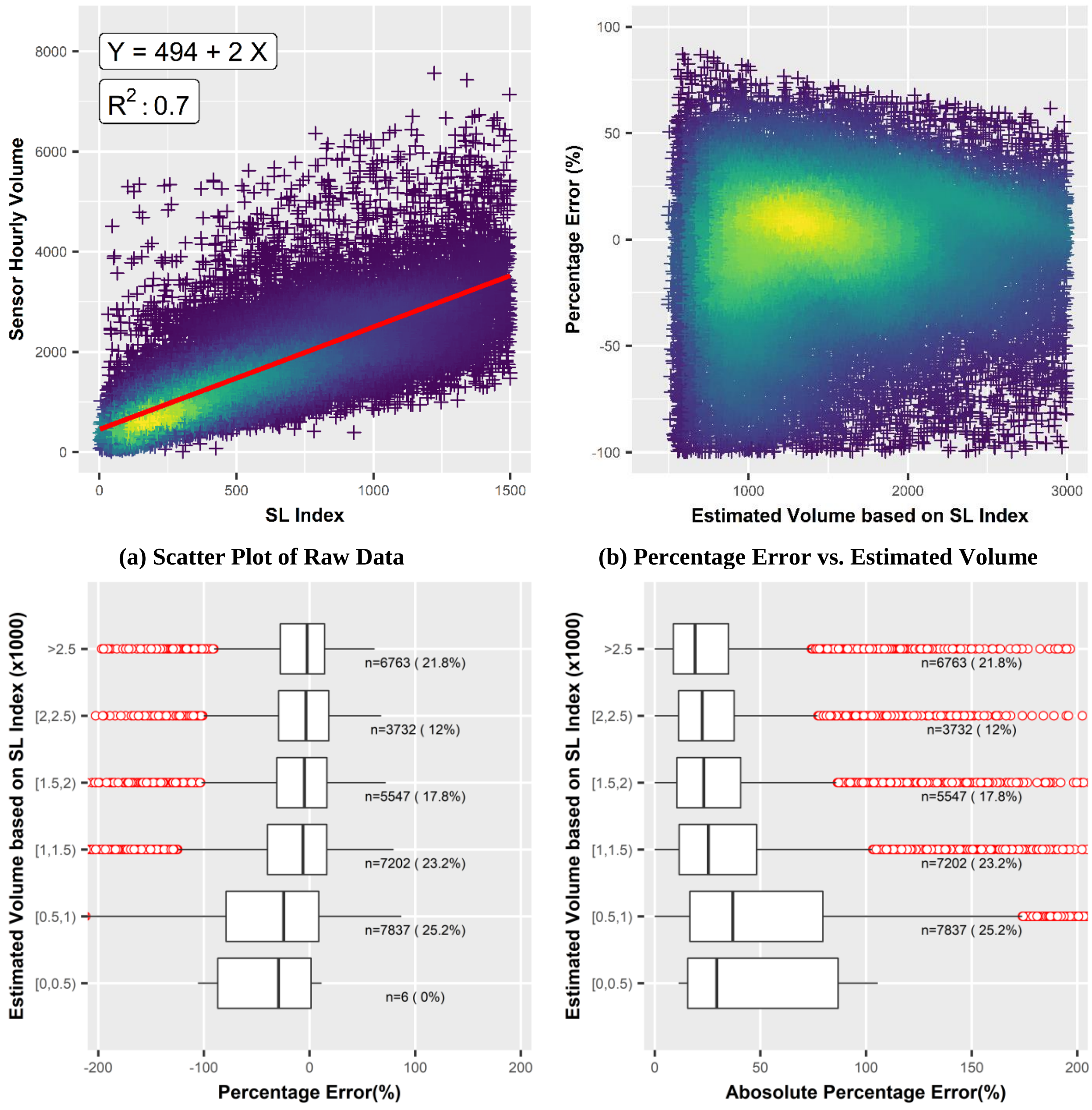
Type of Analysis	APE	# Samples	Proportion
D=1	0~10%	3,904	23.4%
	10~20%	3,635	21.8%
	20~30%	3,010	18.0%
	30~40%	2,015	12.1%
	40~50%	1,164	7.0%
	50%+	2,945	17.7%
	Total	**16,673**	**100.0%**
D=5	0~10%	333	29.7%
	10~20%	282	25.3%
	20~30%	228	20.4%
	30~40%	108	9.6%
	40~50%	59	5.3%
	50%+	109	9.7%
	Total	**1,120**	**100.0%**
D=12	0~10%	168	30.0%
	10~20%	163	29.1%
	20~30%	94	16.8%
	30~40%	63	11.3%
	40~50%	22	3.9%
	50%+	50	8.9%
	Total	**560**	**100%**

StreetLight Index vs. Traffic Counts of VDOT Road Links with In-road Sensors

The hourly traffic count data were obtained for the 193 VDOT road links used in the 2018 AADT evaluation. Let Y_i^{dh} represent the actual volume collected at h^{th} hour on day d at the site i, and SL_i^{dh} is the corresponding SL Index. Further, we only extracted the data collected on Tuesday, Wednesday, and Thursday for evaluation. This resulted in $d = 1, 2, ..., 12$ (Note: Only 12 days in April 2018 are either Tuesday, Wednesday, or Thursday). In addition, data collected between 6am and 8pm on each day for each site were collected. Therefore, there are 14 hours of data for a day and $h = 1, 2, ..., 14$. Some sites may not have SL Indexes available for some hours. Those hours' data were filtered out. Then, we performed the same analysis as the one presented for the Virginia Beach data.

The scatter plot of the two sets of raw data is shown in Figure 20(a). The light-colored area means that there were more observations clustered in that region. The two sets of data were used to establish the conversion model shown as the line in Figure 20(a): $Y = 450 + 2X$. The estimated hourly volumes based on the SL Index were then compared with the actual traffic count to quantify the performance measures (e.g., PE and APE). Figure 20(b) shows the final calculated results of estimated hourly volume based on SL Index versus their associated percentage errors. Like the previous evaluation scenario, we can see that the percentage error can vary for most of the estimated volumes. For example, when the estimated volume is less than 1,000 vph, the possible PE can also range between -100% and 75%. Figure 20(c) and (d) show the distributions of PE and APE, respectively. These results show that if the traffic volumes were estimated based on the raw SL Index, the distributions of their PE and APE both are centered and the values tend to be smaller when the estimated volume is higher.

Figure 21 and Figure 22 present similar results based on the use of the average SL Index. Like the analysis in the case of the Virginia Beach data, the use of the average SL Index produced an improved estimation of the average hourly traffic volume. The PE and APE both are less dispersed and the majority of the APEs are below 25%. Consistent with early findings, the errors of SL Index in the lower estimated volume conditions were higher. The cases when $D = 5$ and $D = 12$ are preferred if the estimated hourly volume is over 500 vph. The average of the SL Index is advised to be a better way to describe the typical traffic condition of a given site.



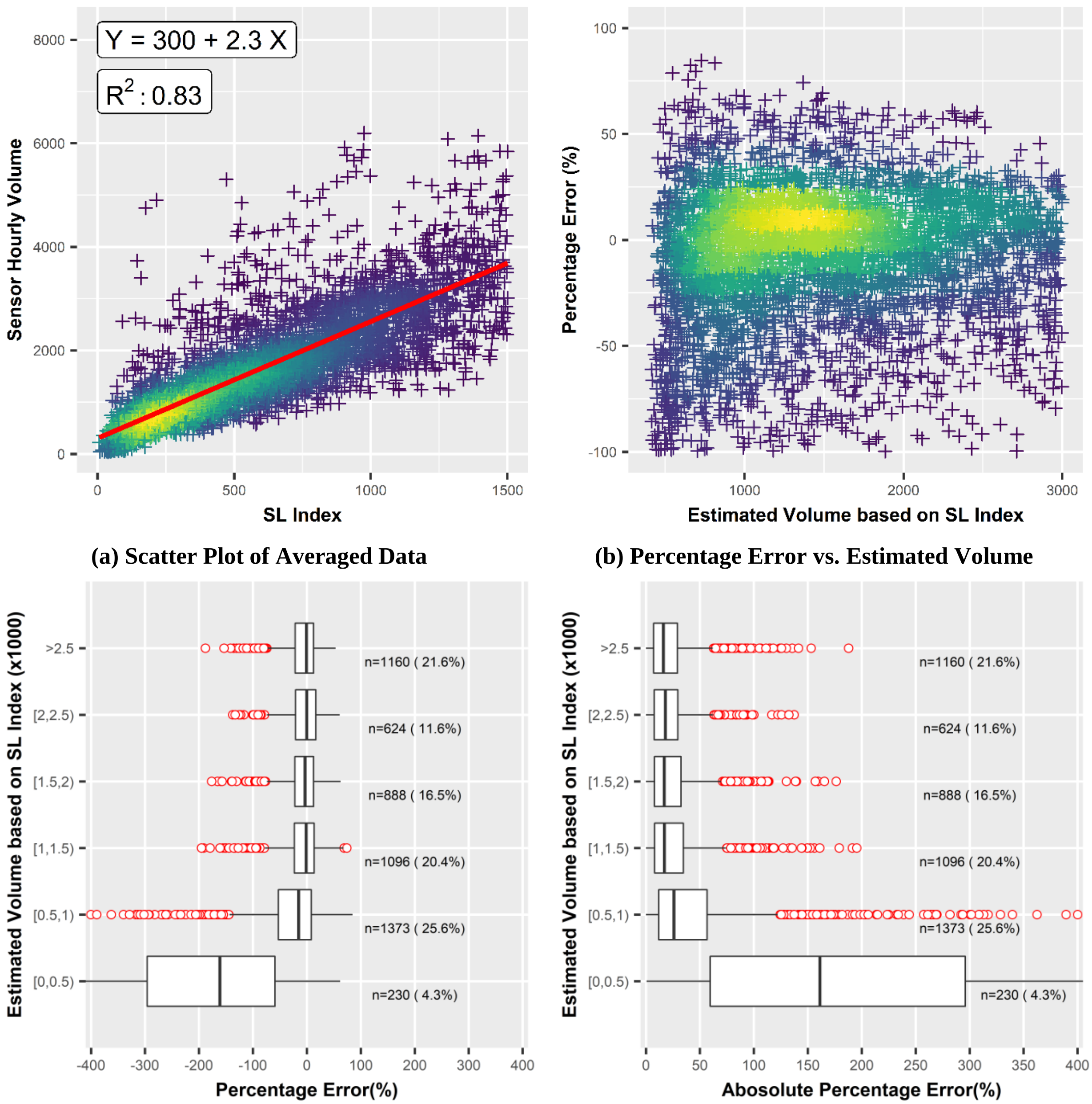
(a) Scatter Plot of Raw Data

(b) Percentage Error vs. Estimated Volume

(c) Distribution of Percentage Error

(d) Distribution of APE

Figure 20. Type 1 Comparison: Estimated Volume based on Original SL Index for VDOT Road Links.



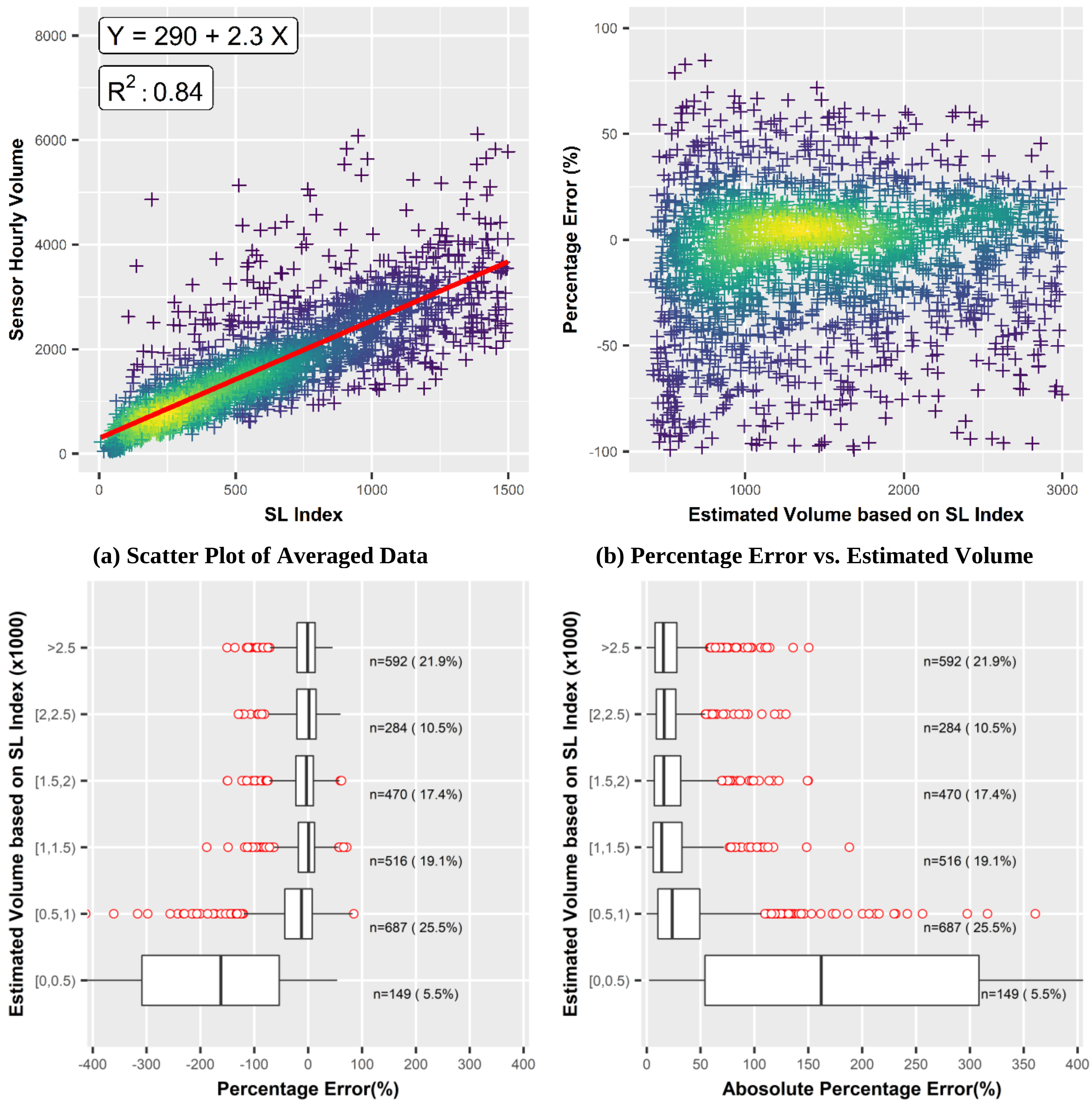
(a) Scatter Plot of Averaged Data

(b) Percentage Error vs. Estimated Volume

(c) Distribution of Percentage Error

(d) Distribution of APE

Figure 21. Type 2(A) Comparison: Estimated Volume based on 5-day Averaged SL Index for VDOT Road Links.



(a) Scatter Plot of Averaged Data **(b) Percentage Error vs. Estimated Volume**

(c) Distribution of Percentage Error **(d) Distribution of APE**

Figure 22. Type 2(B) Comparison: Estimated Volume based on 12-day Averaged SL Index for VDOT Road Links.

Table 18 presents the median, mean, and 95% CI of the APE according to the range of estimated volume based on SL Index. We can see that the median APEs for the estimated volumes above 1,500 vph are lower than 25%. For those below 1,500 vph, their APEs are between 25% and 37%. This suggests that the estimates based on the SL Index are less accurate. With the use of the average SL Index, the errors were reduced in analyses with $D=5$ and $D=12$. With the aggregation of 12-day data, most of the 95% CIs were reduced in terms of the ranges and upper limits. Table 19 further examines the APE without considering the levels of estimated volumes. If the raw SL Index was used, we can see that only about 40% of the sampled cases have APEs below 20%. About 30% of the sampled cases have APEs greater than 40%. In contrast, the proportions of sampled cases with APEs below 20% are over 50% if the average SL

Index was used. The cases with APEs above 40% were also reduced below 25%. Once again, these results suggest the importance of averaging the SL Index across multiple days to obtain better estimates for the hourly traffic volume of a typical day.

Table 18. APE Summary based on Link Volume Data from VDOT

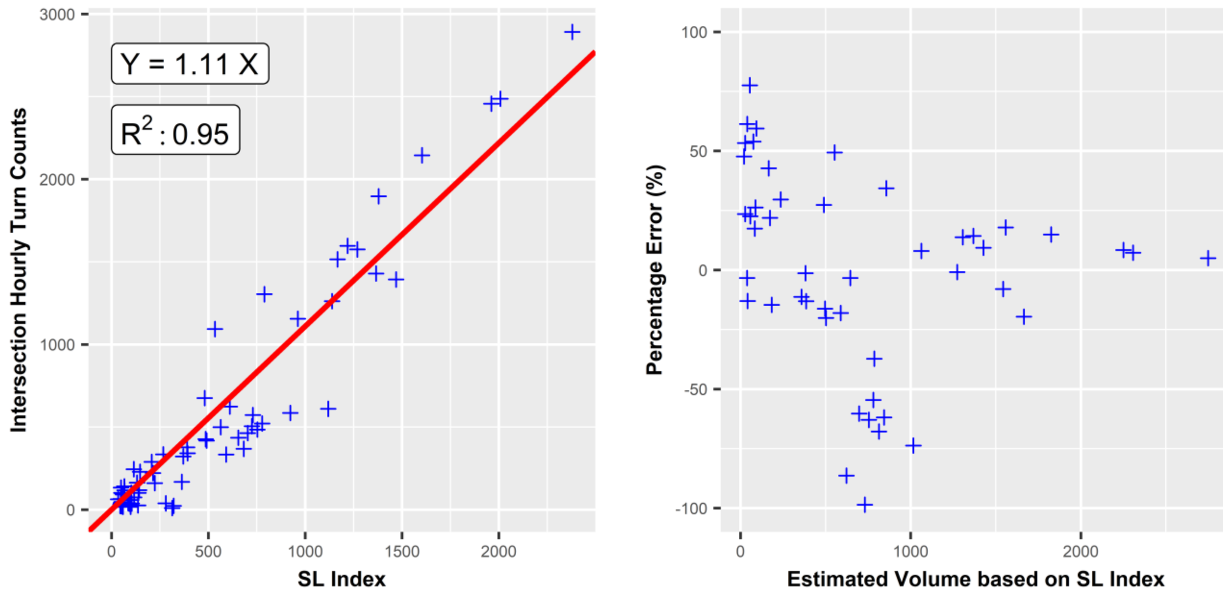
Aggregation Over D Days	**Range of Estimated Volume**	**Absolute Percent Error**			**Samples**
		Median	**Mean**	**95% CI**	
D=1	>2,500	19.1%	26.3%	[0.9%, 97.9%]	6,763
	2,000 – 2,500	22.5%	29.3%	[1.0%, 107.2%]	3,732
	1,500 – 2,000	23.2%	31.3%	[1.1%, 122.0%]	5,547
	1,000 – 1,500	25.4%	39.0%	[1.2%, 140.9%]	7,202
	500 -1,000	37.0%	80.3%	[1.7%, 445.7%]	7,836
	0 – 500	29.4%	136.5%	[11.3%, 565.9%]	6
	All Samples	**24.9%**	**44.1%**	**[1.2%, 203.1%]**	**31,086**
D=5	>2,500	16.2%	21.9%	[0.6%, 81.1%]	1160
	2,000 – 2,500	18.0%	23.0%	[0.8%, 81.3%]	624
	1,500 – 2,000	17.0%	24.4%	[0.7%, 88.4%]	888
	1,000 – 1,500	17.0%	26.0%	[1.0%, 97.2%]	1,096
	500 -1,000	25.9%	47.5%	[1.3%, 233.4%]	1,373
	0 – 500	161.2%	211.2%	[4.7%, 696.3%]	230
	All Samples	**19.4%**	**37.9%**	**[0.9%, 205.4%]**	**5,371**
D=12	>2,500	15.4%	21.5%	[1.0%, 77.3%]	592
	2,000 – 2,500	16.3%	22.2%	[1.1%, 80.5%]	284
	1,500 – 2,000	16.0%	23.5%	[0.5%, 81.2%]	470
	1,000 – 1,500	13.9%	23.2%	[0.6%, 95.9%]	516
	500 -1,000	23.7%	38.5%	[0.9%, 152.0%]	687
	0 - 500	162.0%	205%	[4.8%, 620.0%]	149
	All Samples	**18.3%**	**36.7%**	**[0.8%, 215.7%]**	**2,698**

Table 19. Proportion of Samples in Each Level of APE based on Link Volume Data from VDOT

Aggregation Over *D* Days	APE	# Samples	Proportion
D=1	0~10%	6,791	21.8%
	10~20%	6,130	19.7%
	20~30%	4,900	15.8%
	30~40%	3,654	11.8%
	40~50%	2,430	7.8%
	50%+	7,181	23.1%
	Total	**31,086**	**100.0%**
D=5	0~10%	1,514	28.2%
	10~20%	1,241	23.1%
	20~30%	801	14.9%
	30~40%	477	8.9%
	40~50%	319	5.9%
	50%+	1,019	19.0%
	Total	**5,371**	**100.0%**
D=12	0~10%	816	30.2%
	10~20%	611	22.6%
	20~30%	399	14.8%
	30~40%	229	8.5%
	40~50%	148	5.5%
	50%+	495	18.4%
	Total	**2,698**	**100.0%**

StreetLight Index vs. Observed Turn Counts at Intersections

The extracted SL Indexes were used to estimate the turn counts at the selected intersections. In total, 93 samples (either left turn, right turn, or through traffic) were obtained from the TCDS. The corresponding SL Indexes were also obtained but only 69 have data. Based on the matched data, their scatter plot is shown in Figure 23(a). The line in the figure represents the best-fit regression model. Note that the initial model has a negative intercept which was not reasonable when SL Index is 0. Thus, the intercept was forced to be zero to make the model more meaningful. With the estimated hourly turn counts, the percentage errors were calculated and shown in Figure 23(b). It can be seen that there were some large errors when the estimated volumes were below 1,000 vph. The errors were between -25% and 25% when the estimated volumes were greater than 1,000 vph. The APEs were also calculated and the mean, median, and 95% CI are 153.1%, 39.9%, and [7%, 941.4%], respectively. Please note that the large mean value was attributed to low-volume samples with large PEs. The median should be a better indicator to represent the overall performance in the studied scenario. Table 20 further summarizes the number of samples in each level of APEs. It was found that about 30% of the estimated turn counts have APEs less than 20%. More than 40% of the estimated volumes have errors over 50%.



(a) SL Index vs. Turn Counts **(b) Percentage Error vs. Estimated Volume**

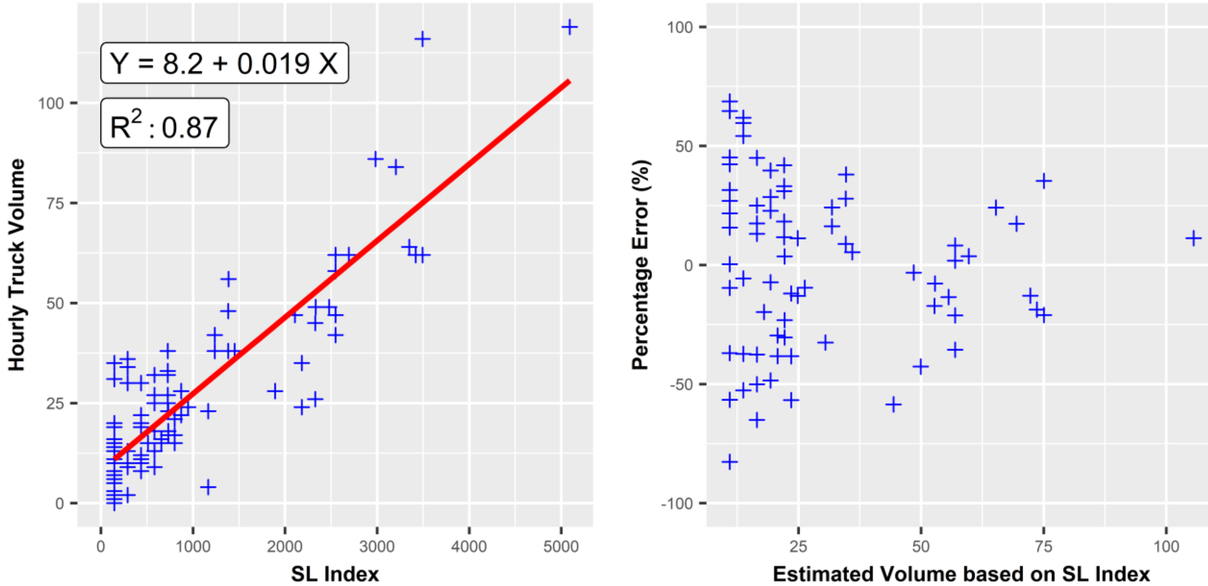
Figure 23. Evaluation of SL Index for Intersection Turn Counts Analysis.

Table 20. Proportion of Samples in Each Level of APE based on Intersection Turn Count Data

Absolute Percentage Error	# Samples	Proportion
0~10%	7	10.2%
10~20%	13	18.8%
20~30%	8	11.6%
30~40%	7	10.2%
40~50%	5	7.2%
50%+	29	42.0%
Total	**69**	**100.0%**

StreetLight Index vs. Observed Truck Traffic at Intersections

In total, 126 samples of truck traffic counts were obtained from the TCDS. Meanwhile, the SL Indexes for them were retrieved from the SL Platform. Only 95 have available SL Indexes. These indexes were used to fit a regression model for estimating truck traffic. The results were shown in Figure 24. Despite small numbers of trucks observed in each hour, the corresponding SL Indexes were quite large. Due to the low sample size, the SL Indexes built on navigation-GPS data for commercial vehicles are not comparable to the ones based on LBS data for personal travel. Also, it was found that the lower bound values of the SL Indexes were limited to a positive number (see Figure 24(a)). Figure 24(b) shows the calculated PEs of the estimated volume. As most of the observed were truck volumes were below 100 vph, the PEs were dispersed widely. The corresponding mean, median, and 95% CI of the APEs were found to be 73.8%, 31%, and [3.3%, 586.7%], respectively. Like the turn count analysis, the mean value should be used with caution. The median should be a better indicator to represent the overall performance in the studied scenario. Table 21 shows the samples at each level of APEs. It was found that about one-third of the estimated truck volumes have APEs less than 20%. About 27% of the estimated volumes have an error over 50%.



(a) SL Index vs. Truck Volume **(b) Percentage Error vs. Estimated Volume**

Figure 24. Evaluation of SL Index for Truck Traffic Analysis.

Table 21. Proportion of Samples in Each Level of Absolute Percentage Error based on Truck Traffic Data

Absolute Percentage Error	# Samples	Proportion
0~10%	13	13.7%
10~20%	19	20.0%
20~30%	13	13.7%
30~40%	18	18.9%
40~50%	6	6.3%
50%+	26	27.4%
Total	**95**	**100.0%**

CONCLUSIONS

The findings in this research are a starting point for VDOT to develop a set of guidelines for using the SL metrics in its planning projects. In conjunction with the guidelines developed in this project, the following major conclusions are drawn:

- *SL data accuracy tends to be problematic under low volume conditions (e.g., AADT under 20,000, volume under 500 vph).* Despite the practices in using SL Index in real projects, its quality was often questioned and only limited evaluation and calibration efforts were performed by existing studies.

- *The survey results show that many users are highly interested in using SL data in their typical projects but there are also concerns and challenges in using the data. Surveys for both existing users and non-users were successfully distributed to transportation professionals within and beyond VDOT.* It was found that many of them expected to use SL metrics for traffic demand modeling and traffic congestion analysis. In particular, the top three application scenarios include OD analysis, traffic flow/volume analysis, and route choice analysis with the SL metrics. The most frequently listed concern was data inconsistencies and errors/quality of the data. About 16% of surveyed users are somewhat or very dissatisfied with the sample size of SL data. The need for additional training, the interface not being user-friendly, and the need for additional data types (beyond what is provided by SL) were also listed as concerns and challenges.

- *The 2018 AADT estimates by SL have a mean APE of 6.2%, which has shown higher accuracy than 2017 AADT whose mean APE is 11.5% based on data from all sites. The benchmarks are the published AADT by VDOT in both years.* The estimates for road links with higher AADT tend to be more accurate.

- *The OD estimates based on the SL Index are often associated with errors. When compared to the actual OD demand, especially for OD pairs with estimated hourly trips below 600, the median APEs can be as high as 60% for the analyzed ODs.* Averaging the SL Indexes across multiple hours of different days (e.g., 5) will help reduce the estimation error, especially for low volume conditions. Depending on project purposes, the aggregation can be based on metrics of multiple days, weeks, or months.

- *The SL Index can be used for traffic volume estimation, but the estimates are subject to notable errors for some roads with different levels of traffic volumes.* The assessment results from the case studies imply that most of the SL Index estimates are associated with median absolute percentage errors as high as 25%. The errors will be greater for road links with volumes less than 500 vph.

- *The SL Index can be used for estimating intersection turn counts, but the estimates will be associated with widely dispersed error rates for low-volume turning traffic.* Most error rates associated with estimated volumes greater than 1,000 vph were found to be within ±25%.

- *The seasonal and weekday/weekend influence on SL data quality is essentially reflected by the actual traffic volume levels and corresponding samples of the SL data*.

- *Using the SL Index for estimating truck traffic approaching intersections is less reliable in low truck volume conditions.* The case study in Virginia Beach showed that the estimated truck volumes at only one-third of the sampled sites have absolute percentage errors less than 20%.

- *The guidelines to use SL metrics were developed based on the literature review, surveys, and comparative evaluations.*

RECOMMENDATIONS

1. *VDOT TMPD should make the guidelines available to users inside VDOT via a designated web site.* The guidelines shown in Appendix A provide instructions complementing the existing SL tutorials and include examples regarding the use of SL metrics for typical planning tasks and key issues and challenges on SL data extraction, processing, calibration, and analysis.

2. *VDOT TMPD should adopt a checklist or a table to track how SL metrics are utilized and calibrated in different applications.* As indicated by previous studies and the results from the comparative evaluations in this study, the quality of the SL metrics is not always consistent in different applications. Although the latest SL AADT estimates showed relatively good performance, the quality of the SL Index when used for estimating OD trips, truck volumes, and traffic counts at intersections or on highways exhibits large variation. Thus, it is necessary for VDOT project managers to know the achieved performance, calibration efforts, and involved benchmark data in projects that use the SL metrics. Appendix C provides a sample checklist. SL data users are recommended to fill in the checklist based on their project information and submit the checklist to VDOT project managers. By providing the checklist, VDOT project managers will have an improved understanding of the use of the SL products.

FUTURE WORK

As SL keeps updating its data products, it would be valuable to understand the data quality and potentials of the new products. In particular, VDOT may want to evaluate the newly published SL Volume product. The SL Volume became available on SL Platform in summer 2019. The SL Volume has the potential to replace the SL Index which would simplify working with SL data since there would not be a need to calibrate SL Index to estimate volumes. Currently, there is no assessment regarding the quality of SL Volume metrics. It is unknown whether the SL Volume estimates can accurately represent the actual traffic flow conditions.

IMPLEMENTATION AND BENEFITS

Implementation

The project team has sought feedback from VDOT staff within and beyond TMPD regarding the developed guidelines. According to the Executive Review (September 25, 2019), the guidelines (Recommendation 1) and template (Recommendation 2) may be posted to the TPRAC SharePoint site or the TMPD site (e.g., Pathways for Planning Data). Then, if the guidelines or templates are updated, the updated version may be provided at those locations. A tutorial session for the guidelines was conducted on October 30, 2019, during the Transportation Planning Research Advisory Committee (TPRAC) Meeting at VDOT’s Bristol District Office.

The guidelines, once available, could be shared with a number of user groups that include (1) the survey respondents [e.g., on-call consultants, district planners, TMPD, and so forth), (2) the users of SL, and (3) the PDC representative staff. This will include sharing the web link to the developed visualization tool to allow users to explore example comparisons between different benchmark data and SL data products. The tool will be maintained by the research team at ODU for at least two years after the publication of this report.

Benefits

Implementing the study recommendations is expected to lead to more accurate and efficient use of the SL data products for VDOT projects.

The primary benefits of implementing Recommendation 1 are the facilitation of the proper and efficient use of the SL products in VDOT projects. It will allow both existing and potential users (e.g., planners and engineers) in VDOT and its partners (e.g., MPOs, consulting companies, etc.) to have an informative reference with various examples to determine whether the SL data products are appropriate for their projects. Users will be able to learn the key steps for working with the data efficiently and be aware of the related issues and challenges in using the data.

The primary benefit of implementing Recommendation 2 is improved information to understanding the outcomes of projects using SL data products. There were a number of VDOT projects on the SL Platform that have used the SL data metrics but no formal checklist or table was available to VDOT users, collaborators, and consultants. Clearly defining the information that need to be reported will allow VDOT to have additional transparent facts in assessing the quality of each project. Meanwhile, the users of SL data products will be able to follow the checklist (table) to summarize key facts about their work related to the use of SL data metrics.

ACKNOWLEDGMENTS

The authors are grateful for the great guidance and support provided by the study's technical review panel, including the project champions Brad Reed (Staunton District Planner) and Chuck Proctor (Culpeper District Planner), Mike Fitch, Amy O'Leary, Cathy McGhee, John Miller, Ram Venkatanarayana, and Kevin Wright at VTRC, Jungwook Jun at VDOT TMPD, Sulabh Aryal at the Richmond TPO, Stephen Haynes at VDOT Fredericksburg District Division, and John Simpkins and Richard Duran at the FHWA Virginia Division Office. In particular, their input for survey designs and guideline development was deeply appreciated. The authors also thank the other VDOT staff that have helped share their experience and comments at different stages of the project. In addition, the authors appreciate Wenling Chen, Dan Dunnavant, Brooke Kordas, Michael Sledge, and Hamlin Williams at VDOT TED for their help in collecting multiple important traffic datasets used in the project. Last but not least, the authors appreciate the TPRAC committee to provide us opportunity to conduct the tutorial session during its meeting in fall 2019.

REFERENCES

Aultman-Hall, L., and Dowds, J. Designing the All-in-One Vermont Transportation Survey. *Report No. 17-004. University of Vermont Transportation Research Center*, 2017.

Avner, J. Using Big Data in Small and Medium Sized Regions. 2018. http://onlinepubs.trb.org/onlinepubs/Conferences/2018/Tools/JAvner.pdf. Accessed November 8, 2018.

Cambridge Systematics. San Joaquin Valley I-5/SR 99 Goods Movement Study Strategic Programs and Their Feasibility Assessment. 2017. http://sjvcogs.org/wp-content/uploads/2016/08/FresnoCOG_SJV_I5_SR99_GMS_DR2_20160707_DRAFT_FINAL_v3_20160718_small.pdf. Accessed October 11, 2018.

CDM Smith. Lake/Orange Connector Preliminary Traffic & Revenue Study Report 2017. https://www.cfxway.com/wp-content/uploads/2017/09/Lake_Orange-Connector-TR-Letter-Report_Final.pdf.

City of Virginia Beach. Transportation Data Management System. 2019. https://vbgov.ms2soft.com. Accessed May 12, 2019.

Coates, A.M. Application of Streetlight Data for Miami University Circulation Study and Wisconsin DOT. WRA, 2017. https://www.otdmug.org/wordpress/wp-content/uploads/2017/09/Baker-OTDMUG-Presentation-09-08-17.pdf. Accessed November 8, 2018.

DenBleyker, A., DeFries, T., Palacios, C., Koupal, J., Manzo, C., Bowman, N., Vaz, N., and Schewel, L. Using Telematics Data to Improve the National Emissions Inventory. *Transportation Research Board 97th Annual Meeting*, 2018.

Federal Highway Administration. Mega-Regional Multi-Modal Agent-Based Behavioral Freight Model. 2018. http://www.azmag.gov/Portals/0/Documents/MagContent/TRANS_2017-02-13_SHRP2-TRANS_2017-06-06-C20-MAG-Next-Generation-Freight-Demand-Model-Update.pdf. Accessed October 11, 2018.

Fehr & Peers. Park City Transportation Demand Management Final Report. 2016. http://52.26.130.11/home/showdocument?id=32722. Accessed October 23, 2018.

Fehr & Peers. Central Coast Origin-Destination Survey. Santa Barbara Council of Governments (SBCAG), 2016. http://www.sbcag.org/uploads/2/4/5/4/24540302/central_coast_o-d_survey_final_report_7-8-2016.pdf. Accessed December 10, 2018.

Georgia Department of Transportation. Existing Volume Development and Origin-Destination Data. Downtown Connector Study, 2016. http://www.dot.ga.gov/BuildSmart/Studies/Documents/DowntownConnector/DataReport.pdf. Accessed November 8, 2018.

Granato, S. Various uses for INRIX/Streetlight data: Ohio plus border area. *Ohio Department of Transportation (ODOT)*, 2017.

Harrison, K. Improving Last-Mile Connections to Transit: An Exploration of Data and Analysis Tools. *16th TRB National Transportation Planning Applications Conference*, 2017.

Herman, M., and Tong, J. Use Big Data and Modeling Tools To Decipher Traffic Patterns: Case Studies in Virginia. *2017 Virginia Section Insitute of Transportation Engineers (VASITE) Annual Meeting*, Virginia Beach, VA, 2017. June 28-30, 2017.

Hong, S. Expanding the Uses of Truck GPS Data in Transportation Planning and Analysis. *50th TRB Applications Conference*, Raleigh, NC., 2017.

Kimley-Horn. Core of Rosslyn Transportation Study Existing Conditions Report. *Kimley Horn*, Reston, VA, 2018.

Komanduri, A., Schewel, L., Beagan, D., and Wong, D. Using Big Data to Develop Comparative Commercial Vehicle Metrics for Port Traffic at Major Ports in the US. *WRA*, 2017.

Kuppam, A., Lemp, J., Selby, B., Livshits, V., Vallabhaneni, L., and Hong, S. Tour-Based Truck Travel Models using Truck GPS Data. *16th TRB National Transportation Planning Applications Conference*, 2017.

McAtee, S. Validating Trip Distribution in Southeast Michigan Using GPS Data. *16th TRB National Transportation Planning Applications Conference*, Raleigh, NC, United States, 2017. May 14-18, 2017.

McCahill, C. Improving Last-Mile Connections to Transit: An Exploration of Data and Analysis Tools. *CNU*, 2017.

Miller, H.J., and O'Kelly, M.E. Progress report: Estimating External Travel Using Purchased Third-Party Data. 2016. https://www.otdmug.org/wordpress/wp-content/uploads/2016/12/Dec2016_OSU_Archived_Data_v3.pdf. Accessed October 15, 2018.

Minnesota Department of Transportation. AADT Comparison - Streetlight Data. 2017. https://www.otdmug.org/wordpress/wp-content/uploads/2017/09/Parikh_OTDMUG_Streetlight.pdf. Accessed April 13, 2019.

Napa Valley Transportation Authority. Napa County Travel Behavior Study Napa County Joint Board of Supervisors and Planning Committee Meeting, 2014. https://www.nvta.ca.gov/sites/default/files/Napa%20County%20Travel%20Behavior%20Study.pdf. Accessed October 11, 2018.

Picado, R. Southeast Florida Origin-Destination Travel Survey. *Final Report, Miami-Dade Transportation Planning Organization*, Miami, Florida, 2017.

Roll, J. Evaluating Streetlight Estimates of Annual Average Daily Traffic in Oregon. Salem, OR, United States, 2019.

San Diego Association of Governments (SANDAG). Visualizing Truck Flows Based Upon Industry Data and Using Truck Visualization as A Planning Tool. 2018. https://www.sandag.org/uploads/publicationid/publicationid_4542_24796.pdf. Accessed November 19, 2018.

Schiffer, R.G. Alternate Methodologies for Origin-Destination Data Collection. *Tools of the Trade Conference on Transportation Planning for Samll and Medium-sized Communities*, 2016.

Shay, N. Using Streetlight InSight Data for a Small Area Study: Rickenbacker Area. Mid-Ohio Regional Planning Commission, 2017. https://www.otdmug.org/wordpress/wp-content/uploads/2017/09/20170908_MORPC_Shay_OTDMUG.pdf. Accessed November 18, 2018.

State Smart Transportation Initiative. Understanding Trip-Making with Big Data - A Connecting Sacramento Summary Brief. 2017. https://www.ssti.us/wp/wp-content/uploads/2017/07/SSTI_Connecting_Sacramento_Tripmaking.pdf. Accessed February 12, 2019.

StreetLight Data. StreetLight Data's AADT 2018 Methodology and Validation White Paper. https://www.streetlightdata.com, 2019. Accessed July 15, 2019.

StreetLight Data. https://www.streetlightdata.com/resources#case-study. 2018. Accessed March 15, 2019.

Tillery, R., and Pourabdollahi, Z. Florida Statewide Tourism Travel Demand Model: Development of a Behavior-Based Framework. *AICP*, *RS&H Inc*, 2016.

Turner, S., and Koeneman, P. Using Mobile Device Samples to Estimate Traffic Volumes. *Minnesota. Dept. of Transportation. Research Services & Library*, 2017.

Venkatanarayana, R., and Fontaine, M.D. Assessing the Quality of Private Sector Origin-Destination Data. *Transportation Reserach Board 97th Annual Meeting*, Washington, DC, 2018.

Virginia Department of Transportation. 66 Express Lanes: Inside the Beltway, Richmond. 2018. http://66expresslanes.org/. Accessed August 5, 2019.

Wahlstedt, M.R. Using GPS Based Origin-Destination Data to Improve Traffic Studies. 2017. https://www.dot.state.oh.us/engineering/OTEC/2017Presentations/52/Wahlstedt_52_Version1.pdf. Accessed May 2, 2019.

APPENDIX A

DEVELOPMENT OF GUIDELINES FOR USING SL METRICS

Based on the findings from the literature review, user surveys, and independent evaluations with real-world data, the ODU research team developed a draft set of guidelines for using the data products from the SL Data. Then, the guidelines were modified based on feedback from members of the VDOT Technical Review Panel (TRP) who reviewed them in August 2019. For example, one suggestion was to include one-page summaries for each planning work task; accordingly, ten such summaries were developed to complement the more detailed instructions on the use of SL metrics.

In summary, the guidelines include five major components covering (1) summaries for related planning work tasks that may use SL data, (2) general guidance of data extraction from the SL Platform, (3) typical applications in different planning tasks, (4) data quality and calibration, and (5) possible tools and techniques that may support the use of the SL products.

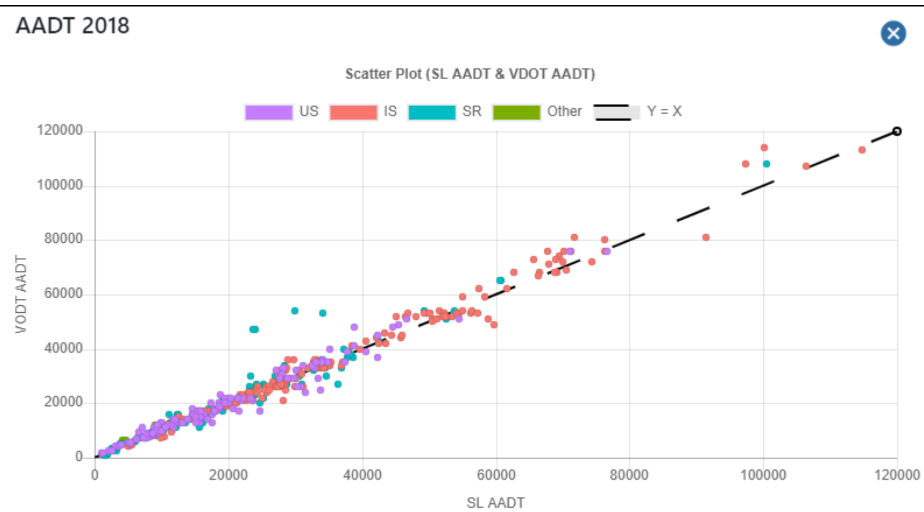
Summary of Using SL Data in Planning Work Tasks

[G1] One-page summary tables for identified work tasks that may use SL data

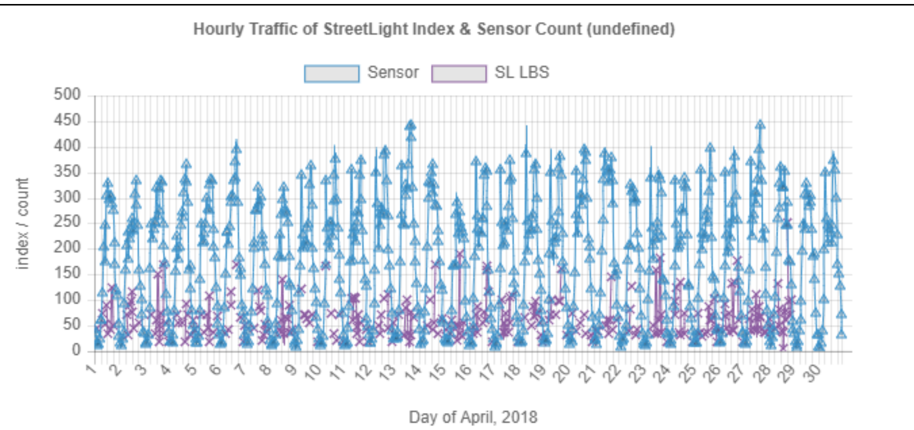
There are ten types of planning projects that VDOT staff frequently need to perform with SL data. A one-page summary for each of these ten types follows.

- Work Task 1: Projects Requiring Estimated AADT
- Work Task 2: Projects Requiring Estimated Traffic Volumes on Roadway Segments
- Work Task 3: Projects Requiring Estimated Intersection Turning Movement Traffic
- Work Task 4: Projects Requiring Truck Traffic Throughout Analysis at Intersections
- Work Task 5: Projects Requiring Estimated Travel Demand between Different OD Zones
- Work Task 6: Projects Requiring Route Choice Analysis between OD Zones
- Work Task 7: Projects Requiring External Traffic Analysis
- Work Task 8: Projects Requiring Traffic Congestion Analysis
- Work Task 9: Projects Requiring Bicycle Traffic Analysis
- Work Task 10: Projects Requiring Traffic Analysis for Low-Volume Roads

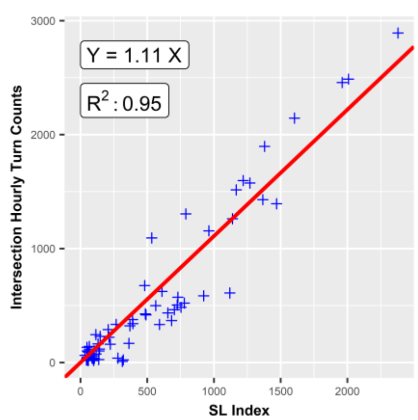
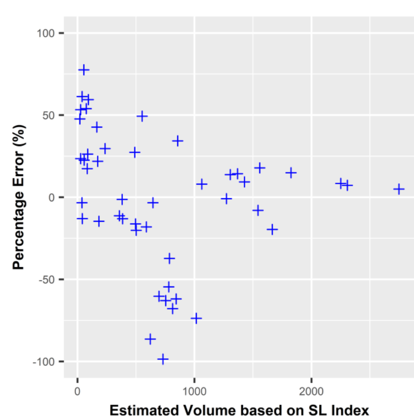
Work Task 1: Projects Requiring Estimated AADT

Guideline	Description
SL Data Source	Blended GPS and LBS data
SL Metrics Availability	Year 2017 and Year 2018.
Supported Temporal Analysis	2017 and 2018 analyses are enabled.
SL Analysis Type	Estimated AADT Values
SL Data Extraction Key Components	Deploy a (bi-)directional pass-through zone on each roadway segment.
Analysis Techniques and Tools Needed	Excel will be sufficient for quantitative analysis, e.g., filtering and finding AADT at each roadway segment, etc.
Expected Error	Most road segments (about 60% for tested links in this project) have an absolute percentage error (APE) less than 10% in 2017 and APE less than 5% in 2018. The errors of the segments with estimated AADT below 10,000 tend to be greater. Rounding to the hundredths or thousandths will also account for some difference between VDOT's AADT and the SL AADT.
SL Data Gaps	Some locations do not have SL AADT available due to low sample size.
Calibration Suggested	Not very necessary for the links with SL AADT over 10,000 in 2018 and links with SL AADT over 40,000 in 2017. Otherwise see **[G21]**.
Possible Benchmark Data	AADT reported by State and/or local transportation agencies
Calibration Procedure	Simple linear regression model **(See [G21])**
Other Comments	It should be noted that the SL AADT 2017 and SL AADT 2018 were estimated with different algorithms. The AADT 2018 has improved performance. The year-to-year SL AADT comparison is not suggested because the estimation algorithms changed in 2018.
Visual Demo: *The graph shows comparison between the 2018 SL AADT estimate and the 2018 AADT estimate published by VDOT. Most of the points are well aligned with the reference line Y=X in the graph, indicating the relative good performance of 2018 SL AADT estimate.*	

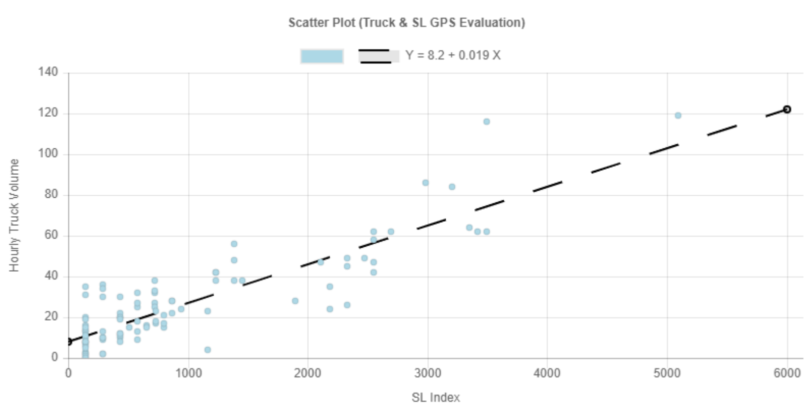
Work Task 2: Projects Requiring Estimated Traffic Volumes on Roadway Segments

Guideline	Description
SL Data Source	Personal LBS / Commercial GPS
SL Metrics Availability	The SL Index is available: 01/2016 to 04/2019 (Personal); 01/2014 to 02/2019 (Commercial)
Supported Temporal Analysis	Both hourly and daily analyses are enabled.
SL Analysis Type	Zone Activity Analysis
SL Data Extraction Key Components	Deploy a (bi-)directional pass-through zone on each roadway segment. **(See [G4]** and **[G5])**
Analysis Techniques and Tools Needed	Excel will be sufficient for quantitative analysis, e.g., filtering and finding the SL Index for each road segment; visualizing the time series SL Index of each link, etc. Other tools such as ArcMap will be helpful for preparing a large number of zones. **(See [G8]** and **[G23])**
Expected Error	Depending on the range of traffic volume. Most of the absolute percentage errors can be less than 25% if the average of SL Indexes over five days were used to estimated traffic volumes. (The average of 5 days or more was preferred in the case study.) Error tends to be greater and unstable at low volume conditions (e.g., 500 vph).
SL Data Gaps	Some hours do not have SL Index. **(See [G6])** The change of SL Index is not always consistent with that of actual volume.
Calibration Suggested	Strongly suggested
Possible Benchmark Data	Loop detector data or other traffic sensor volume data are preferred Prefer to have benchmark data from the similar functional class of subject roads
Calibration Procedure	Check the scatter plot of SL Index vs. benchmark data. If no linear trend is observed, SL Index should not be used. If a linear trend is clear, building a liner regression model to relate the SL Index to the benchmark data. The established model will be used to covert new SL indexes into estimated link volumes. **(See [G21])** Calibration can be done in Excel by creating a scatter plot and obtaining the linear trend line equation.
Other Comments	Not recommended for real-time analysis; and averaged SL Indexes over multiple periods (such as 5-day) are suggested considering the dispersion of daily (hourly) indexes. Experiments show that the hourly average of raw SL Indexes from five or more days helped reduce the error in estimating the average traffic flow.
Visual Demo: *The graph shows the time series of SL Index and the corresponding volume measured by a traffic detector on a road segment.*	

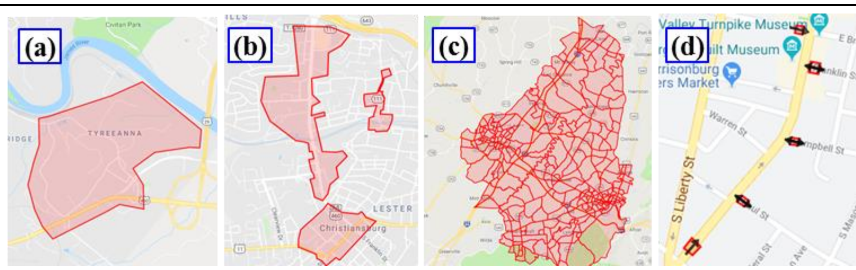
Work Task 3: Projects Requiring Estimated Intersection Turning Movement Traffic

Guideline	Description
SL Data Source	Personal LBS
SL Metrics Availability	SL Index is available: 01/2016 to 04/2019
Supported Temporal Analysis	Both hourly and daily analyses are enabled.
SL Analysis Type	OD Analysis
SL Data Extraction Key Components	Deploy a bi-directional pass-through zone on each link connecting to each intersection. **(See [G15])**
Analysis Techniques and Tools Needed	Excel will be sufficient for quantitative analysis, e.g., filtering and finding SL Index for each turning direction by pairing origin zone and destination zone; calculating the percentage of each turning direction (left, straight, right), etc.
Expected Error	Depending on the range of turning movement counts. Absolute percentage errors can be as high as 25% for sites with hourly turn counts less than 1,000 vph. Possibly obtaining greater errors in cases with less hourly turn counts.
SL Data Gaps	Some hours do not have SL Index at turns with lower traffic; and SL Index below 500 is less reliable due to small sample size.
Calibration Suggested	Strongly suggested
Possible Benchmark Data	Manually observed turning movement counts; or Turning movement counts from traffic sensors/cameras (e.g., automatic traffic recorder)
Calibration Procedure	Check the scatter plot of SL Index vs. benchmark data. If no linear trend is observed, SL Index should not be used. If a linear trend is clear, building a liner regression model to relate the SL Index to the benchmark data. The established model will be used to covert new SL index into estimated turning movement volume. **(See [G21])** Calibration can be done in Excel by creating a scatter plot and obtaining the linear trend line equation.
Other Comments	Not recommended for real-time analysis. Recommended for hourly analysis for scenarios with estimated turn counts above 1,000 vph. Recommended for analysis with daily average SL Index.
Visual Demo: *The left graph shows the scatter plot and the calibrated equation for estimating hourly turn counts with SL Index. The right graph illustrates the errors when converting SL Index to estimated hourly turn counts.at a set of intersection approach legs.*	 

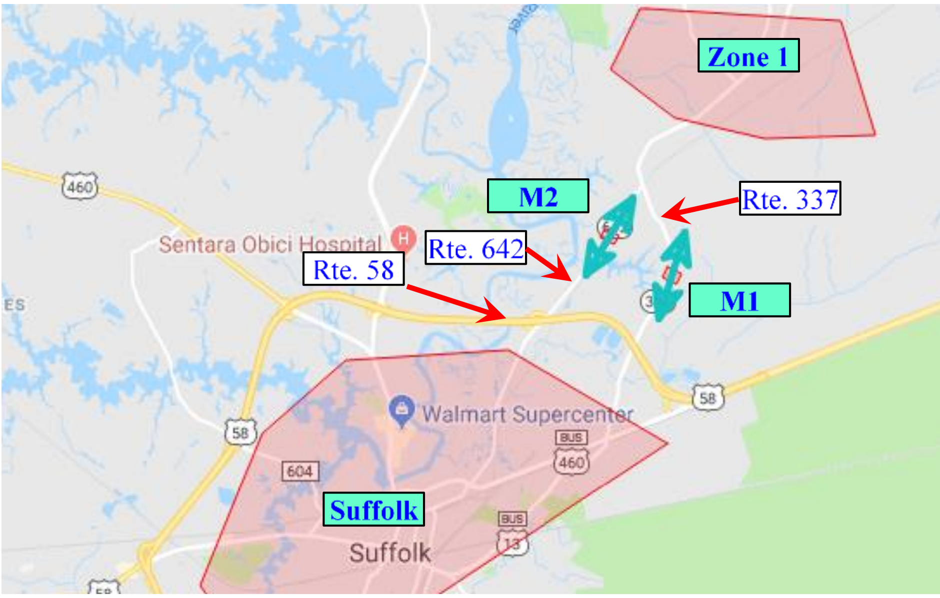
Work Task 4: Projects Requiring Truck Traffic Throughout Analysis at Intersections

Guideline	Description
SL Data Source	Commercial GPS
SL Metrics Availability	SL Index is available: 01/2014 to 02/2019
Temporal Analysis	Both hourly and daily analyses are enabled.
SL Analysis Type	Zone activity analysis
SL Data Extraction Key Components	Deploy a (bi-)directional pass-through zone on each target link connecting to the intersection. **(See [G4]** and **[G5])**
Analysis Techniques and Tools Needed	Excel will be sufficient for most of the quantitative analyses, e.g., filtering and finding SL Index for each target link; visualizing the time series SL Index of each link; and comparing with other reference data. Other tools such as ArcMap will be helpful for creating a large number of zones. **(See [G8]** and **[G23])**
Expected Error	Depending on the range of truck traffic volume. Sites with low truck volume (e.g., 25 vph) tend to have greater errors (e.g., absolute percentage errors can be over 50%).
SL Data Gaps	Some hours do not have SL Index. **(See [G6])** SL Index below 200 is not reliable due to small sample size.
Calibration Suggested	Strongly suggested
Possible Benchmark Data	Observed truck traffic count or vehicle classification data (e.g., weight-in-motion data) collected on the similar type of roadways.
Calibration Procedure	Check the scatter plot of SL Index vs. benchmark data. If no linear trend is observed, SL Index should not be used. If a linear trend is clear, building a linear regression model to relate the SL Index to the benchmark data. The established model will be used to covert new SL index into estimated truck volume. **(See [G21])** Calibration can be done in Excel by creating a scatter plot and obtaining the linear trend line equation.
Other Comments	Not recommended for real-time analysis. Recommended for averaged hourly/daily analysis for approach legs of intersections with estimated volume above 25 vph. Similar analysis can be done for truck traffic on road segments. Depending on the volume of truck traffic, it is anticipated that SL Index may have different errors.
Visual Demo: *The graph illustrates the relationship between the raw SL Index and the hourly truck volume observed at a set of intersection approach legs.*	

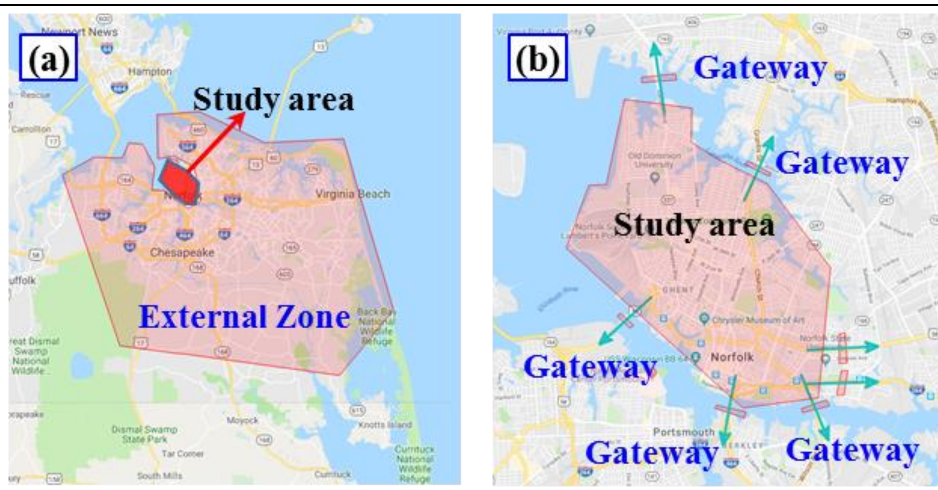
Work Task 5: Projects Involved Estimated Travel Demand between Different OD Zones

Guideline	Description
SL Data Source	Personal LBS / Commercial GPS
SL Metrics Availability	SL Index is available: 01/2016 to 04/2019 (Personal); 01/2014 to 02/2019 (Commercial)
Temporal Analysis	Both hourly and daily analyses are enabled.
SL Analysis Type	OD analysis
SL Data Extraction Key Components	Draw or import a zone set (TAZs, census tracts, etc.). **(See [G8]** and **[G23])** Pass-through zone should be placed on road and larger zones represent areas.
Analysis Techniques and Tools Needed	Excel will be sufficient for quantitative analysis, e.g., filtering and finding SL Index between different OD zones; visualizing the time series SL Index of each OD pair or each zone, etc. To perform spatial analysis, tools such as ArcMap can be used.
Expected Error	Depend on the demand between each OD pair. Higher demand (e.g., >600 trips/hour) tends to have lower absolute percentage errors (APE) (e.g., APE<25%). Lower demand is likely to have greater (e.g., APE>50%) and unstable APE due to the low sampled trips between the zones.
SL Data Gaps	Some OD pairs may not have observed SL samples and therefore the estimated SL Index is subject to error. **(See [G6])** Higher SL Index does not guarantee higher demand. **(See [G18])**
Calibration Suggested	Suggested
Possible Benchmark Data	Travel survey data (in large spatial areas) Vehicle tracking data through license plate readers, GPS, etc. (in corridor level OD analysis)
Calibration Procedure	Obtain the actual/estimated traffic production/attraction for each zone based on the benchmark data. **(See [G12])** Distribute the production/attraction based on the calculated proportions of SL Index associated with each OD pair. This will obtain an estimated OD matrix. **(See [G22])** Establish the linear regression model between the estimated OD matrix and OD matrix derived from the benchmark data. **(See [G21])**
Other Comments	Recommended for using the proportion of SL Index. Additional attributes can be retrieved such as trip purpose (home/work) but difficult to verify their accuracy. Suggested use multiple-period (e.g., the same hour of 5 days) average SL Index in analysis instead of disaggregated (e.g., individual hours) SL Index. OD pairs with low estimated demand (e.g., <600 trips/hour) based on SL Index should be used with caution.
Visual Demo: *The four graphs show (a) a single zone, (b) multiple zones, (c) a large zone set, and (d) zones along roadways created by SL users.*	
Visual Demo: Accessed August 20, 2019. Reprinted With Permission.	

Work Task 6: Projects Involved Route Choice Analysis between OD Zones

Guideline	Description
SL Data Source	Personal LBS/Commercial GPS
SL Metrics Availability	SL Index is available: 01/2016 to 04/2019 (Personal); 01/2014 to 02/2019 (Commercial)
Supported Temporal Analysis	Both hourly and daily analyses are enabled.
SL Analysis Type	OD Analysis with Middle Filter
SL Data Extraction Key Components	Deploy an origin zone, a destination, and a middle filter zones on each route. Middle filters should be placed at critical sites to capture more trips using the target routes. **(See [G14])**
Analysis Techniques and Tools Needed	Excel will be sufficient for quantitative analysis, e.g., comparing road choices with different middle filter zones, etc.
Expected Error	When sample size is low, the SL index will be less reliable. **(See [G18])** The detected trips may not exactly follow candidate routes because they may use other alternative routes to reach the middle filter zones.
SL Data Gaps	Only one middle filter can be selected at a time and therefore the full path of each trip will not be fully traceable. The route choice analysis cannot be conducted for short-term studies (e.g., 15 minutes, 30 minutes, etc.).
Calibration Suggested	Suggested
Possible Benchmark Data	Trajectory data; and others that can track vehicles at criterial locations along target routes (e.g., toll transaction data, videos, weigh-in-motion data, etc.)
Calibration Procedure	Establish a linear regression model between the SL Index and the observed trips from benchmark data of the target routes. **(See [G21])**
Other Comments	If there were more alternative routes that can lead traffic to the middle filer zones, the route analysis based on the SL Index will be less reliable.
Visual Demo: *The graph shows an example of route choice analysis from Zone 1, to Suffolk. Rte. 337 and Rte. 642 are two routes of interest. Two middle filter zones M1 and M2 are needed.* *However, it is impossible to obtain all trips that fully used a specific route (i.e., Rte. 337 vs. Rte. 642). For example, some vehicles may use Zone 1 → Rte.337 →M1 → Rte.58 → Rte.642 → Suffolk.*	
Visual Demo: Accessed August 20, 2019. Reprinted With Permission.	

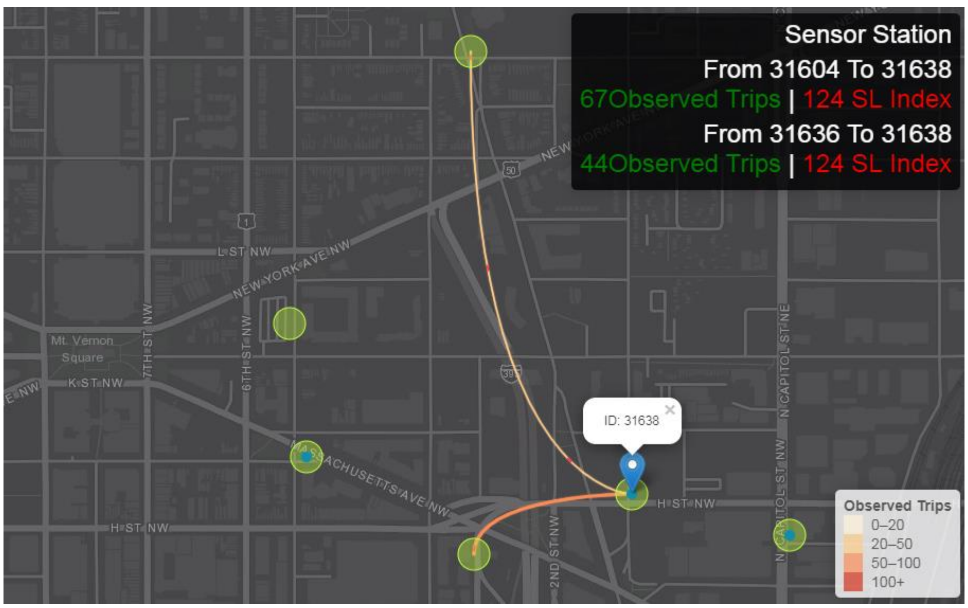
Work Task 7: Projects Involved External Traffic Analysis

Guideline	Description
SL Data Source	Personal LBS/Commercial GPS
SL Metrics Availability	SL Index is available: 01/2016 to 04/2019 (Personal); 01/2014 to 02/2019 (Commercial)
Supported Temporal Analysis	Both hourly and daily analyses are enabled.
SL Analysis Type	OD Analysis/Zone Activity Analysis
SL Data Extraction Key Components	If there are too many gateways connecting the study area to the external areas, drawing a big non-pass-through zone containing all outside areas will be more feasible and then use OD Analysis. Otherwise, draw pass-through zones (gateways) on each road segment connecting to the study area and apply Zone Activity Analysis. **(See [G13])**
Analysis Techniques and Tools Needed	Excel will be sufficient for quantitative analysis, e.g., filtering and finding SL Index between different zones; visualizing the time series SL Index of each zone, etc.
Expected Error	Two major sources of errors should be noted: (1) the non-pass-through external zone is not large enough to represent all the external areas; likewise, if gateways were used, not all roads linking to the target zone are considered; and (2) SL index may be inaccurate due to small sample size.
SL Data Gaps	The analysis will be more reliable for areas with limited access points (e.g., ports, airports, areas with natural constraints, etc.). It will be difficult to gather external traffic for an open area with numerous access points (e.g., downtown Richmond).
Calibration Suggested	Suggested
Possible Benchmark Data	Traffic count data collected at the access points (gateways). Trajectory (OD) of vehicles traveling between the study area and external areas.
Calibration Procedure	Establish a linear regression model between SL Index and observed external trips. **(See [G21])**
Other Comments	Percentage information can be calculated and compared between the study area and different external areas. However, this is subject to error, especially for the pairs with low SL indexes. Make sure external zones (gateways) are deployed appropriately. **(See [G13])**
Visual Demo: *The graphs show the example of using (a) non-pass-through zone as the external zone and (b) multiple pass-through zones (gateways) to capture traffic going to external areas.*	
Visual Demo: Accessed August 20, 2019. Reprinted With Permission	

Work Task 8: Projects Involved Traffic Congestion Analysis

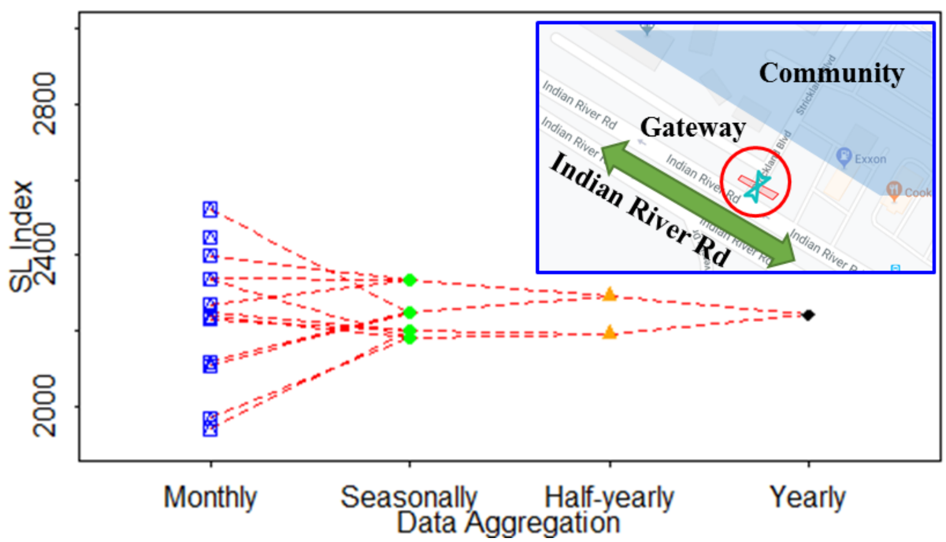
Guideline	Description
SL Data Source	Personal LBS
SL Metrics Availability	SL Index is available: 01/2016 to 04/2019
Supported Temporal Analysis	Both hourly and daily analyses are enabled.
SL Analysis Type	Segment analysis
SL Data Extraction Key Components	Deploy line segments (drawing or uploading a polyline layer) **(See [G9])**
Analysis Techniques and Tools Needed	Excel will be sufficient for quantitative analysis, e.g., filtering and finding SL estimated speed and travel time on each road segment; show the time series plot of speed/travel time of the selected road segment, etc.
Expected Error	Since the middle gateway is generated automatically at the center of each road segment, it may not be at the user's target site. Adjustment is needed. **(See [G9])** The speed and travel time estimates based on SL data may not often represent the true speed and travel time of vehicles fully used the segment.
SL Data Gaps	The minimum time interval of the speed and travel time update is by hours. They are not sensitive enough to capture short-term (e.g., 15 min.) variations.
Calibration Suggested	Strongly suggested
Possible Benchmark Data	Sensors (e.g., radar or loop detectors) capturing speed of road segments. Other data sources that can extract travel time of the passing vehicles (e.g., Bluetooth data, toll transaction data, etc.).
Calibration Procedure	Pairing the SL estimated speed/travel time with the benchmark data and establish a linear regression model. **(See [G21])**
Other Comments	Avoid conducting segment analysis for the segments with many parallel routes that pass the same middle filter zone.
Visual Demo: *The example on the right side shows the road segments drawn on SL Platform: (1) A to D; and (2) B to C. We can see that the middle zones of the road segments are at the same position. However, the analysis results can be different due to their set-ups of the "start-zone" and "end-zone".* *The key information provided by SL are average speed and travel time of the segment.*	
Visual Demo: Accessed August 20, 2019. Reprinted With Permission	

Work Task 9: Projects Involved Bicycle Traffic Analysis

Guideline	Description
SL Data Source	Personal LBS (Bicycle Type)
SL Metrics Availability	Averaged SL Index is available: Average of 4/18, 5/18, 6/18, 9/18, 10/18 and 11/18. (Note: As of July 2019, it was the average of 5/17, 6/17, 5/18, and 6/18)
Supported Temporal Analysis	Both hourly and daily analyses of the average days are enabled.
SL Analysis Type	Both zone activity analysis and OD analysis can be performed.
SL Data Extraction Key Components	Deploy a (bi-)directional pass-through zone on each bike roads (lanes). Select the Bicycle layer in the OpenStreetMap (OSM) Layer list.
Analysis Techniques and Tools Needed	Excel will be sufficient for quantitative analysis, e.g., filtering and finding SL index associated with each zone.
Expected Error	High errors are expected because of the limited bike trip data.
SL Data Gaps	The bike trip data are biased (depending on the vendors' raw data sources); for example, the bike-sharing data were not well captured by the SL Index. **(See [G17])**
Calibration Suggested	Strongly suggested (but may not be feasible due to unavailable benchmark data)
Possible Benchmark Data	Bike counts Bike-sharing system transaction records
Calibration Procedure	Unless continuous bicycles traffic data were available, it would difficult to compared and establish the calibration model as does in vehicle OD analysis.
Other Comments	Currently, this data source is very limited and unreliable due to the low sampling rate and biased raw data; and It is not recommended to use the bicycle SL Index before further verification has been performed.
Visual Demo: *The example shows extracted OD bike trips from the Capital Bike-Sharing System in D.C. The corresponding SL Index did not change according to the observed bike trips between stations.*	

Bike trip data source: Capital Bikeshare. SL Index data source. StreetLight Data

Work Task 10: Projects Focused on Low-Volume Road Traffic Analysis

Guideline	Description
SL Data Source	Personal LBS
SL Metrics Availability	SL Index is available: 01/2016 to 04/2019
Supported Temporal Analysis	Only daily analyses are enabled.
SL Analysis Type	Both zone activity analysis and OD analysis can be performed.
SL Data Extraction Key Components	Deploy (bi-)directional gateways on each road segment. **(See [G4])**
Analysis Techniques and Tools Needed	Excel will be sufficient for quantitative analysis, e.g., filtering and finding SL index associated with each zone.
Expected Error	Errors may be generated since SL index varies month to month.
SL Data Gaps	Daily index can be retrieved from time periods with different length. (e.g., month, season, half year, year)
Calibration Suggested	Strongly suggested
Possible Benchmark Data	Sensor counts (e.g., loop detectors) Video records
Calibration Procedure	Establish a linear regression model between the SL Index and the observed counts from benchmark data of the target roads. **(See [G21])**
Other Comments	With more months involved, the larger sample size will alleviate the error caused by the low sample size. On neighborhood roadways, seasonal and half-year aggregation are suggested.
Visual Demo: *The graph shows the extracted SL Index from a road connecting a neighborhood to the Indian River Rd in Virginia Beach. The daily SL index were retrieved in different temporal aggregations: monthly, seasonally, every half year, and yearly.* *In this case, the monthly aggregation is not recommended due to its large variation across the months.*	

General Guidance for Data Extraction and Preparation

[G2] Data Availability and Accessibility

The SL Platform provides 2017 and 2018 AADT estimates and SL Index between January 2014 and February 2019 (as of July 2019) that mainly allows for OD analysis, OD analysis with middle filter, OD to preset geography (e.g., census tracts), zone activity analysis, and segment analysis. Users can focus on personal or commercial travel analysis. In addition, as of July 2019, only limited travel analysis for bicycles and pedestrians are feasible based on the fused data of four months (May 2017, June 2017, May 2018, and June 2018). In August 2019, the average index based on the fuse of 6-month data (4/18, 5/18, 6/18, 9/18, 10/18 and 11/18) was provided. The temporal resolution for the SL Index can be as small as one hour.

[G3] Setting Zone Types

Users can draw zones or upload zones in the format of a shapefile or an Excel file to the SL Platform. If a target zone covers a large area (e.g., a traffic analysis zone (TAZs), a census tract, etc.), the zone is recommended to be a non-pass-through zone. A pass-through zone is suggested to be placed on a road segment.

Example*:* Figure 25 demonstrates the two types of zone configurations. If the goal is to create a zone for the Norfolk area, it should be set as a non-pass-through zone. If a large zone is set as a pass-through zone, this would likely trigger internal review by the SL Platform or have invalid metrics.

(a) Non-Pass-through Zone (Recommended) **(b) Pass-through Zone (Not Recommended)**

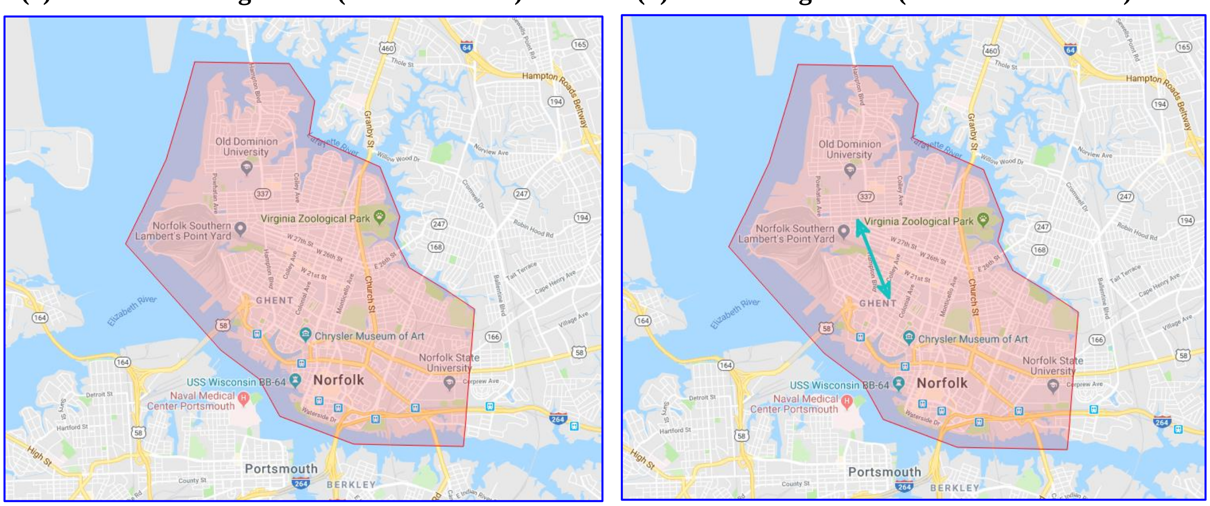


Figure 25. Setup Zones Covering a Large Area. Accessed July 25, 2019. Reprinted With Permission

[G4] Positioning of Zones on the Map

Users should manually verify the placement of each zone to confirm that it covers the target (e.g., a specific road segment or an area boundary of interest) shown on the OpenStreetMap layer on the SL Platform. Users should confirm whether the zone is a pass-through zone. If yes, users need to further determine whether the zone is bi-directional or not.

Example: Figure 26 provides an illustration of the creation of pass-through zones on a section of I-64. Zone 1 in the figure is not a correct configuration as it covers other adjacent roads (i.e., Crew Rd). The extracted SL Index will reflect the combination of traffic information of both I-64 and the adjacent roads. Zone 2's configuration is an example of a properly sized and placed zone for collecting bi-directional traffic data. Zones 3, 4, 5, and 6 are examples of properly sized and placed zones for gathering traffic data for a selected direction (i.e., Westbound of I-64). The arrow of each zone should be parallel to the road of interest.

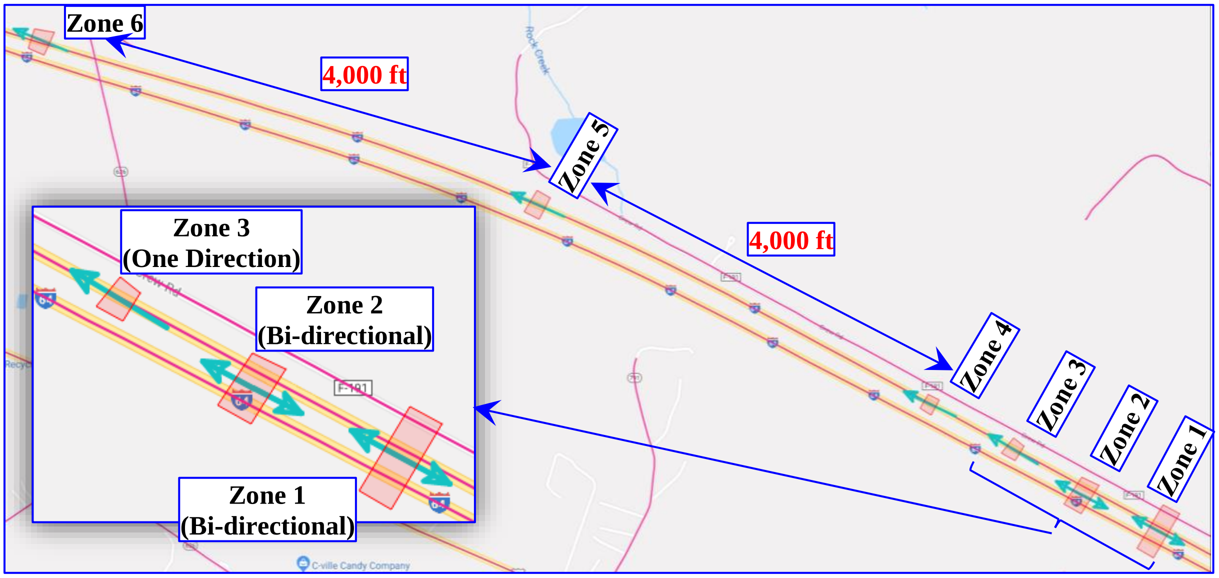


Figure 26. Example of Defining the Pass-through Zones. Accessed July 25, 2019. Reprinted With Permission.

[G5] Sensitivity of the SL Index for Adjacent Zones

Users need to be aware of the sensitivity of the SL Index to the positions of zones on road segments. Even if there is no exit or entrance along a road segment, the SL Index will not necessarily be identical if the pass-through zone is placed at different locations along the segment. Users should place each zone to its target site as precise as possible.

Example: SL Index data for a selected date were extracted for pass-through Zones 3, 4, 5, and 6 shown in Figure 26. There is no exit or entrance between the zones. These zones were placed at a distance of about 4,000ft between Zones 4, 5, and 6, and 1,000 ft between Zones 3 and 4. Despite their proximities, we can see that their indexes presented in the following Table 22 are not always identical for some hours when trip counts (samples) are available, e.g., SL Indexes for 7am-8am and 11am-12pm across the zones.

Table 22. Examples of Extracted SL for Adjacent Zones with Different Distances

Zone	Zone 3		Zone 4		Zone 5		Zone 6	
Time of Day (09/11/2018)	Trip Counts	SL Index	Trip Counts	SL Index	Trip Counts	SL Index	Trip Counts	SL Index
01: (12am-1am)	N/A	65	N/A	65	N/A	65	N/A	65
02: (1am-2am)	N/A	133	N/A	133	N/A	133	N/A	133
03: (2am-3am)	N/A	99	N/A	99	N/A	99	N/A	99

04: (3am-4am)	N/A	54	N/A	54	N/A	54	N/A	54
05: (4am-5am)	N/A	86	N/A	86	N/A	86	N/A	86
06: (5am-6am)	7	248	7	248	7	248	7	248
07: (6am-7am)	14	441	14	441	13	412	13	412
08: (7am-8am)	21	596	22	621	21	581	22	605
09: (8am-9am)	14	437	14	437	15	482	15	482
10: (9am-10am)	13	376	13	376	13	376	13	376
11: (10am-11am)	21	666	20	628	20	628	18	567
12: (11am-12noon)	14	432	15	468	13	416	16	502

[G6] SL Index with Low-sample Sizes or Missing Data

Users need to pay attention to the SL Index whose corresponding number of sampled trips is not available (trip counts=N/A) or low. For example, each hourly SL Index did not change for the first 6 hours across the four zones listed in Table 22. The missing or small samples require imputation of the corresponding SL Indexes. The imputation process will need certain assumptions (e.g., one may assume the N/A can be replaced with the average of indexes from historical hours) and is prone to introduce noise that can make the estimated SL Index less reliable.

[G7] Creating Multiple Projects for the Same Period

If users desire hourly SL Index data for the specific days of the week over multiple weeks, it is advised to create a separate project for each week. Otherwise, the extracted indexes will only represent the averaged values during selected time period.

***Example**:* In a project, two-week (02/10/2019-02/23/2019) SL Index data were extracted for Zones 3 shown in Figure 26. The extracted data for each hour can only provide the average results of the two Tuesdays (02/12/2019 and 02/19/2019) shown in Table 23. For instance, "SL Index=504" at time of day=12 represents the averaged value of the same hour of the two selected Tuesdays: (547+461)/2=504. If users are interested in extracting data for 02/12/2019 and 02/19/2019 separately, two separate projects need to be created on the SL Platform, with each project only covering one of the target dates.

Table 23. Examples of Extracted SL Index Averaged Over Different Days

Zone 3	**Avg. Tue: 02/12/19 & 02/19/19**		**Tue: 02/12/19**		**Tue: 02/19/19**	
Time of Day	**Trip Counts**	**SL Index**	**Trip Counts**	**SL Index**	**Trip Counts**	**SL Index**
01: (12am-1am)	3	23	N/A	18	N/A	27
02: (1am-2am)	3	36	N/A	50	N/A	25
03: (2am-3am)	4	46	N/A	32	N/A	61
04: (3am-4am)	4	38	N/A	23	N/A	54
05: (4am-5am)	5	51	N/A	45	N/A	59
06: (5am-6am)	24	248	16	362	8	133
07: (6am-7am)	55	566	20	412	35	720
08: (7am-8am)	78	762	34	707	44	819
09: (8am-9am)	69	740	31	693	38	785

10: (9am-10am)	73	766	35	677	38	857
11: (10am-11am)	53	592	22	484	31	700
12: (11am-12noon)	49	504	27	547	22	461

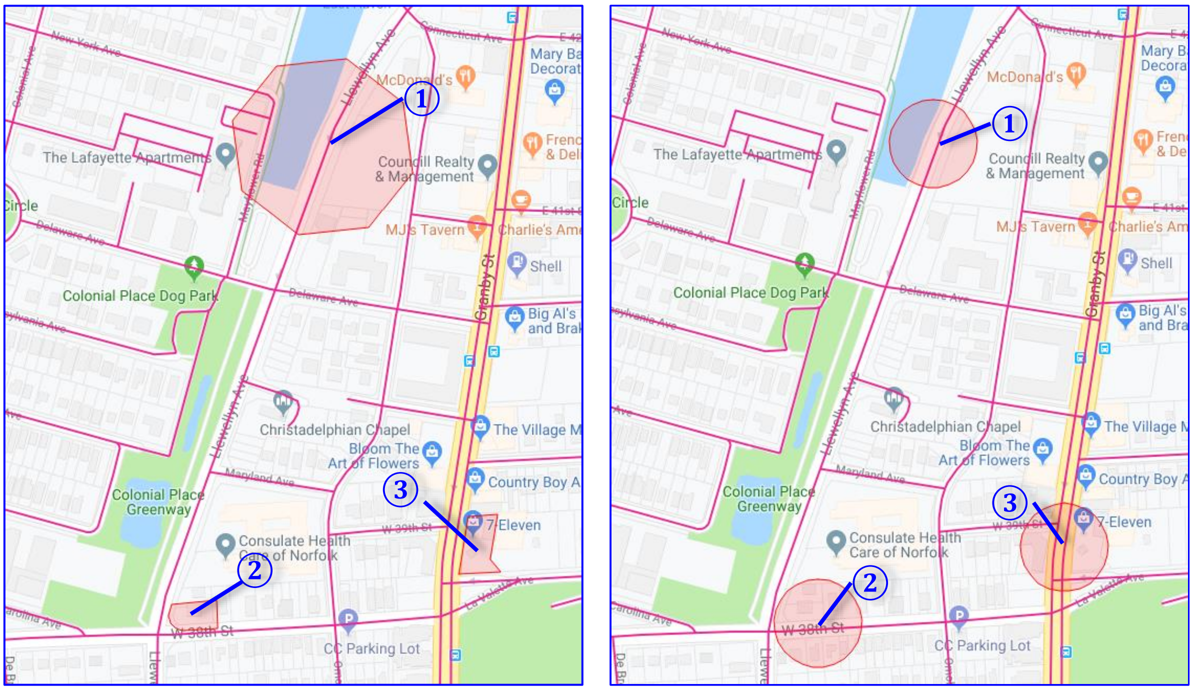
[G8] Creating a Zone Set with Numerous Zones

If SL Indexes for a large number of sites (e.g., 150 zones/links) are needed and the zone set is not available, users should prepare the basic zone set in advance because of the time-consuming zone drawing process on the SL Platform. Instead of using the "Upload Excel" option provided by the SL Platform, tools such as ArcMap that can create buffers for points are recommended to generate a layer of temporary zones and save it as a shapefile. After importing the shapefile of the temporary zones into the SL Platform, users should review and then edit each zone to make sure it covers the target site and indicate the direction and pass-through settings as needed.

***Example**:* Suppose it is needed to extract data from the SL Platform for a set of sites. The site information can be organized with the Excel template provided by the SL Platform (i.e., Figure 27(a)). If this file is directly imported into the SL Platform, the automatically generated zones will be like Figure 27(b), which may not provide good references to the site coordinates. One can see that two created polygons (zones 2 and 3) are off the road and it is difficult for users to determine which road the zone should be affiliated with. This is especially challenging for projects in urban areas with dense road networks. Instead, Figure 27(c) illustrates an alternative way to support the zone creation. First, a buffer layer was created in ArcMap such that each site's coordinates will be used as the location of a centroid and a circle with a user-defined radius will be created for the centroid. Then this buffer layer will be imported into the SL Platform and the outcome is illustrated in Figure 27(c). Finally, users can quickly go through each of these correctly referred "circle zone" in Figure 27(c), edit them as a rectangle to cover the road appropriately, and configure the direction and pass-through settings. This is found to be efficient as manually locating each zone on the embedded map of the SL Platform is very time-consuming.

Zone ID	Zone Name	Address	Latitude	Longitude
1	Granby		36.881221	-76.282
2	38 St		36.8806	-76.28452
3	Llewellyn Ave		36.884529	-76.28334
…	…	…	…	…

(a) Zone Information Needed for Creating Zone Set with the SL Excel Template



(b) Create Zones Using SL Excel Template **(c) Create Zones Using Buffered Area**

Figure 27. Preparation of a Zone Set with Many Zones. Accessed July 25, 2019. Reprinted With Permission.

[G9] Setting Links and Zones in Segment Analysis

When segment analysis is needed, users need to strictly select the segment of interest. The pass-through zone will be automatically placed in the middle of the drawn or imported links. If the pass-through needs to be placed at a different location (e.g., close to the upstream or downstream of the link), further zone editing effort is needed to redraw the pass-through zone at its target location. Users should be aware that the collected SL Index only reflects the sampled traffic that passed the starting point, the pass-through zone, and the ending point of the link. It does not represent the traffic using the full path of the link.

Example: Figure 28 shows an example of segment analyses for two links with different lengths. A→P1→D represents the longer link and B→P2→C denotes the shorter one. The pass-through zones P1 and P2 are at the same location. However, the overlap of P1 and P2 does not guarantee that the SL Index for each segment analysis will be identical as they represent traffic entering from different starting points and reaching different ending points.

Figure 28. Conducting Segment Analysis for Segments with Different Lengths. Accessed July 25, 2019. Reprinted With Permission.

Using SL Index in Typical Applications

[G10] Traffic Analyses and Applications Expecting Time-sensitive Data

The minimum temporal aggregation unit of the SL Index is an hour. Any traffic analysis or application that requires metrics aggregated in shorter time intervals (e.g., 5 minutes, 15 minutes, etc.) is NOT feasible with the SL Index. For example, the SL Index cannot be used to support projects involving real-time signal timing, dynamic speed control, etc.

[G11] Applications Using the SL AADT Estimate

The latest SL AADT estimate can be a useful source to obtain estimated AADT for roads with traffic volume above 20,000 veh/day. The associated absolute percentage errors of the SL AADT estimate are less than 10% for most of the representative Interstate highways, US routes, and State routes in Virginia. The propagated errors associated with the use of the SL AADT estimate in other predictive analysis should be evaluated.

[G12] Applications with OD Analysis

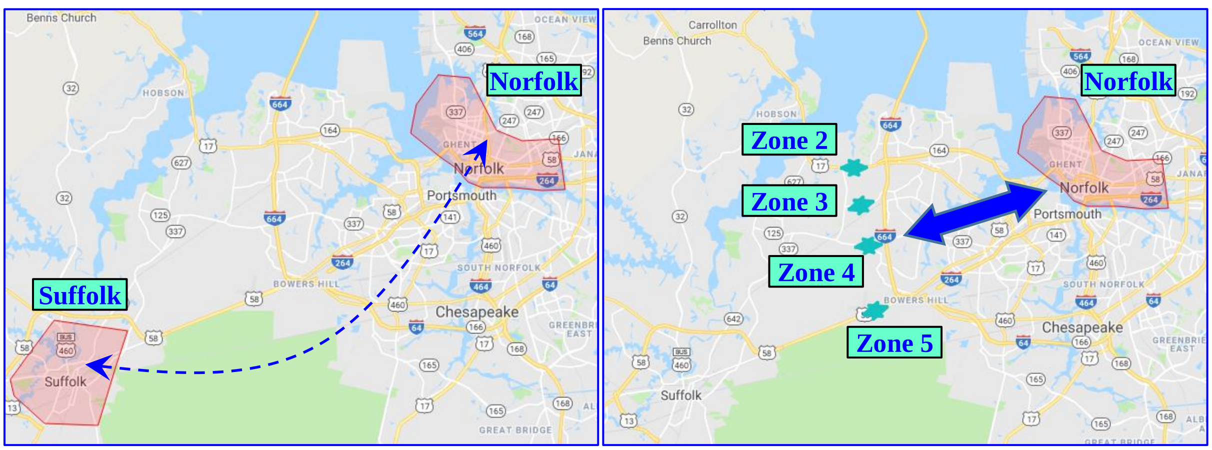
In OD analysis, users can obtain the SL Index between each OD pair. The obtained index only represents the relative number of trips traveling from an origin to a destination. Users cannot directly obtain the absolute number of trips between the zones. The SL Index can be used to compute the ratio of each OD pair's SL Index against the total SL Index of that zone.

Multiplying these ratios by the corresponding actual trip production or attraction of a zone will yield estimated number of trips between ODs. (See [G22])

[G13] Applications with External Traffic Analysis

If the OD analysis in a planning project involves traffic from external zones that are not bounded in a known area, users need to specify the critical roads that link the external area with the target zones and place pass-through zones on these critical roads to capture the external traffic as much as possible.

***Example**:* Figure 29 shows examples of OD analysis with and without external traffic. In Figure 29(a), users can perform OD analysis for trips traveling between the Norfolk zone and the Suffolk zone. However, if the traffic from the west side of I-664 to the Norfolk zone is of interest, the zone configuration in Figure 29(a) will not be able to capture all the trips. Instead, we can place Zones 2-5 on the major roads shown in Figure 29(b). These zones will serve as gateways to help capture most of the external traffic to the Norfolk zone. More precisely, pass-through zones are advised to be placed on all possible routes that link the external traffic to the target zone.



(a) OD Analysis without External Traffic **(b) OD Analysis with External Traffic**

Figure 29. OD Analyses with and without External Traffic. Accessed July 25, 2019. Reprinted With Permission.

[G14] Applications with Route Choice Analysis

The SL Index is not capable of providing precise route choice information in a project that involves many routes linking ODs of interest. In other words, the SL Index is not based on the vehicle traces that fully match each selected route. Instead, it uses middle-filter (pass-through) zones to check whether the trips between two zones passed the locations designated by the middle-filter zones. The trips that passed the middle-filter zones but only used a portion of the selected routes will also be counted in the SL Index.

***Example**:* Suppose we want to estimate how many motorists travel from Zone 1 to the Suffolk zone through Rte. 642 instead of Rte.337 shown in Figure 30. A typical zone setting will be similar to the one shown in Figure 30. This involves two middle-filter zones (M1 and M2) on the two target routes. We can only get the SL Index that represents trips crossed M1 or M2.

This does not guarantee that these trips travel the entire length of the corridors (Rte. 337 and Rte. 642) connecting these zones. For example, some trips may pass M1 and then follow the Westbound direction of Rte.58, and later turn to the Southbound direction of Rte. 642 to get to Suffolk.

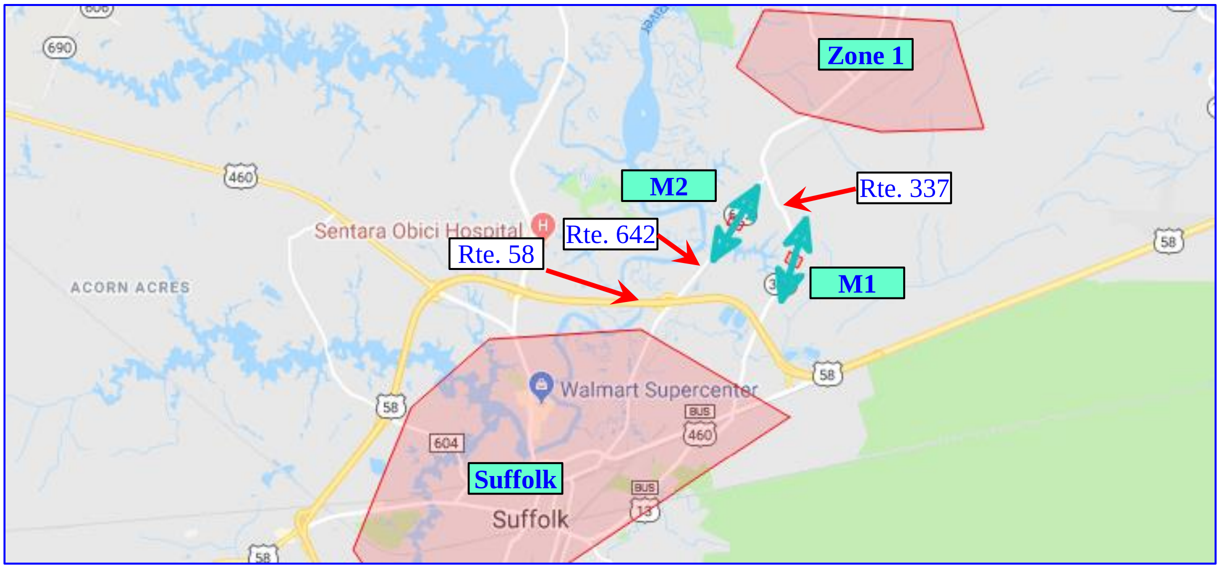


Figure 30. Route Choice Analysis. Accessed July 25, 2019. Reprinted With Permission.

[G15] Applications with Turning Movement Analysis

The SL Platform allows defining pass-through zones to extract the SL index for individual movements at an intersection or interchange. For a corridor/arterial with multiple interchanges or intersections, users need to be aware that the SL Index may not provide precise OD matrix information suitable to support applications such as calibrating a simulation model for a corridor/arterial. This is mainly limited by the incomplete information of route choices between zones separated far away from each other.

Example*:* Figure 31 shows an example of turning movement analysis for four intersections along the one-way 22nd St near the oceanfront of Virginia Beach. For each intersection A, B, C, or D, four pass-through zones were created. At each intersection, users can extract the SL Index representing the relative turning volume for each pair of the pass-through zones around the intersection (e.g. SL Index for Zone 1 to Zone 4 represents the right turning traffic). For the whole section, however, the SL Index from Zone 13 to Zone 4 represents not only the traffic continuously using the Westbound of 22nd St to Zone 4, but also traffic that used other alternative routes from Zone 13 to Zone 4 (e.g., Zone 13→D→22 ½ St→B→A→Zone 4).

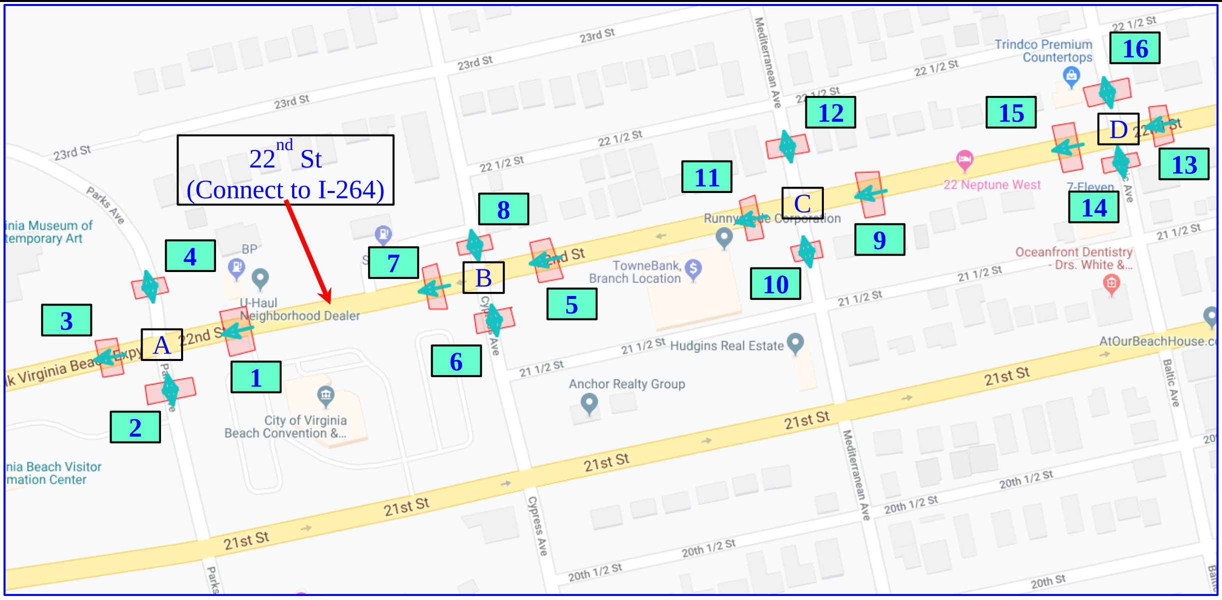


Figure 31. Turning Movement Analysis for an Arterial. Accessed July 25, 2019. Reprinted With Permission.

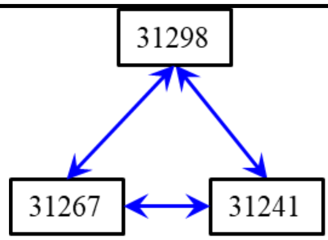
[G16] Applications with Truck Traffic Analysis

The SL Platform enables commercial truck studies by using the SL Index derived based on navigation GPS devices. Other than the difference in data sources, all the analysis process is the same as personal travel studies. Users should be aware that the analysis cannot distinguish trucks by their classes (e.g., two-axle, six-tire, single-unit trucks vs. five-axle single-trailer trucks). Also, whether the sampled navigation GPS devices can well represent the truck traffic in a place is not known because the sampled trucks may not be the major fleet there and the sampling process will be biased.

[G17] Applications with Bicycle Traffic Analysis

The SL Platform allows for extracting the SL Index for pedestrian/bicycle traffic analysis. Users can conduct similar analyses as the vehicular traffic. However, as of July 2019, the bicycle source data are limited to only four months (May 2017, June 2017, May 2018, and June 2018) and users cannot select the SL Index for specific days. The OD analysis for bicycle traffic is still challenging due to the very limited samples of cyclists.

***Example**:* Figure 32 shows an example of OD analysis using data from the Capital Bike-Sharing System. The average daily trips between bike stations with the IDs of 31298, 31267, and 31241 are shown in Figure 32(b) and the corresponding SL Index is shown in Figure 32(c). Notably, all the available SL Indexes are "1", whereas the actual trips varied among the station pairs. In addition, 2/3 of the ODs do not have SL Indexes. Thus, users should be aware of these issues when applying the SL Index for similar bicycle traffic analysis.



(a) Bike Stations

Zone	31298	31267	31241
31298	131	15	20
31267	33	124	126
31241	349	138	18

(b) Observed ODs

Zone	31298	31267	31241
31298	--	1	1
31267	--	--	1
31241	--	--	--

(c) SL Index

Figure 32. Comparison between Observed Bike ODs and SL Index.

Quality and Calibration of the SL Index

[G18] A Critical Premise on the Validation of the SL Index

The SL Index is built on sampled trips. SL has its own (proprietary) algorithms to normalize these sampled trips as the SL Index. The validation of the published SL Index is built on a strong premise that the higher the SL Index is, the larger the actual traffic volume (trip) count is. Users should bear in mind that this premise may not be always valid because the sampling rate in each period is always unknown and changes over time.

Example*:* In a project, the hourly SL Index data for two days (02/10/2019 and 02/23/2019) were extracted for Zone 3 in Figure 26 from the SL Platform. The data were summarized in Table 24. We can see that within the same day, higher sampled trips do not always mean higher SL Index (e.g., Trip Counts=31→ SL Index=693 on 02/12/2019 *vs.* Trip Counts=35→ SL Index=677 on 02/12/2019) and the same sampled trips do not mean the same SL Index (e.g., Trip Counts=35→ SL Index=677 on 02/12/2019 *vs.* Trip Counts=35→ SL Index=720 on 02/19/2019). There is no guarantee that "SL Index=720" means more actual traffic than that of "SL Index=677" because "35" trips could be sampled in either high or low volume conditions, depending on SL's source data.

Table 24. Extracted SL Index Associated with Different Sampled Trip Counts

Zone 3	**Tue: 02/12/19**		**Tue: 02/19/19**	
Time of Day	**Trip Counts**	**SL Index**	**Trip Counts**	**SL Index**
01: (12am-1am)	N/A	18	N/A	27
02: (1am-2am)	N/A	50	N/A	25
03: (2am-3am)	N/A	32	N/A	61
04: (3am-4am)	N/A	23	N/A	54
05: (4am-5am)	N/A	45	N/A	59
06: (5am-6am)	16	362	8	133
07: (6am-7am)	20	412	**35**	**720**
08: (7am-8am)	34	707	44	819
09: (8am-9am)	**31**	**693**	38	785
10: (9am-10am)	**35**	**677**	38	857
11: (10am-11am)	**22**	**484**	**31**	**700**
12: (11am-12noon)	27	547	**22**	**461**

[G19] Using Averaged Metrics

Although hourly SL metrics are available, their error rates often change periodically and are unknown. There is no clear information about the consistency of the metrics over time. Thus, to reduce the variances, users are advised to use the average of multi-period SL Indexes to represent the typical traffic condition of a selected analysis period in a studied scenario (e.g., analysis of traffic flow on a road link, analysis of trips between OD pairs). For example, if the traffic condition of 7 am-8am on weekdays is of interest, collecting the SL Indexes of 7am-8am for multiple representative weekdays (e.g., Tuesday, Wednesday, and Thursday) over a longer period (e.g., several weeks) and using their average in applications should be considered. Depending on project scopes, the length of the period will be different.

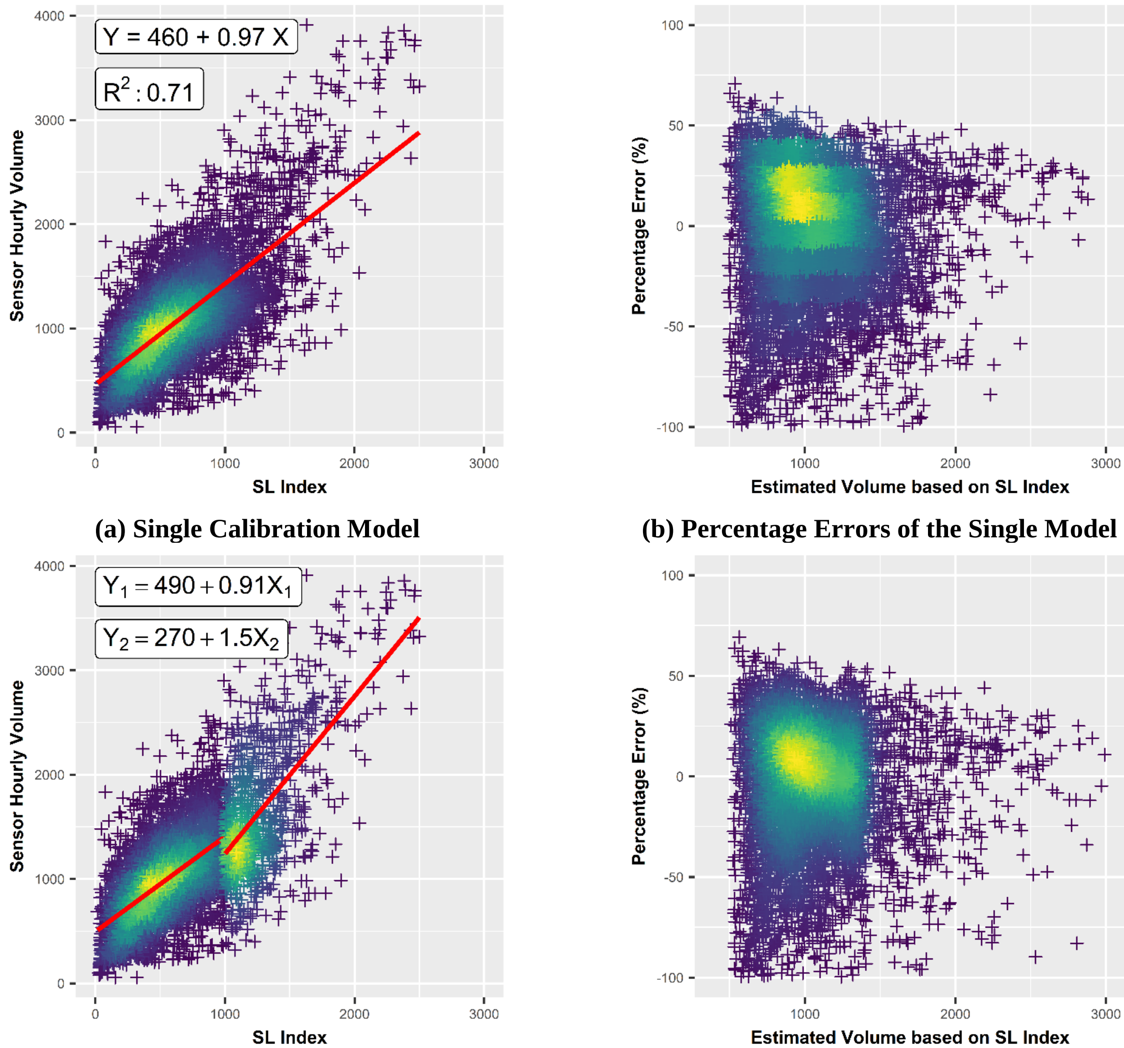
[G20] Preparation of Benchmark Data

Other than the SL AADT estimate, the SL Index cannot be directly used as the traffic volume or trips. In case traffic volume or trips are of interest in a project, users first need to collect some benchmark data for calibrating the SL Index as the estimates of traffic volume or trips. Depending on the application scenarios, typical benchmark data include the traffic counts from traffic sensors, human observations, survey data, turning counts at intersections, vehicle trajectory data, vehicle classification data, etc.

[G21] Calibration of the SL Index

Users should calibrate each type of SL Index before using it in an actual project. As a rule of thumb, a set of benchmark data (e.g., sensor counts, observed trips, etc.) is needed for building a calibration (i.e., regression) model to relate the SL Index or its derivatives to the benchmark data: $Y = f(X)$, where Y represents the benchmark data and X denotes the SL Index or its derivatives. Sometimes tiered calibration models based on the levels of SL Index, facility types, etc. will be preferred to reduce the estimation errors.

Example*:* Figure 33 illustrates the calibration of the SL Index for estimating traffic volume of roads. Traffic sensor data of roads with different volumes were used as the benchmark and the corresponding SL Indexes were extracted from the SL Platform. Figure 33(a) shows the scatter plot of these two types of data and a regression model $Y = 460 + 0.97X$ can be built to relate the SL Index (X) to the sensor count (Y). For other similar sites without traffic count data, this model can be applied to convert their SL Indexes into estimated traffic volumes. Certainly, these estimates are subject to errors like the one shown in Figure 33(b). For another improved approach, we built a regression model for sites with SL Index over 1,000 and another model for others with SL Index below 1,000. Figure 33(c) and (d) show the two estimated models and the corresponding errors. The equation $Y_1 = 490 + 0.91X_1$ represents the estimated model for SL Index less than 1,000. The other one represents the estimated model for SL Index over 1,000. With these tiered models, the errors have been reduced. Specifically, the mean absolute percentage error (MAPE) has been decreased from 31.51% with the single calibration model to 23.42% with the tiered models.



(a) Single Calibration Model **(b) Percentage Errors of the Single Model**

(c) Tiered Calibration Models **(d) Percentage Errors of the Tiered Models**

Figure 33. Calibrating SL Index for Estimating Traffic Volume.

[G22] OD Matrix Estimation

In order to build a travel demand OD matrix based on the SL Index, users need to have either the number of trip productions or the number of attractions of each zone. Assuming the trip productions are available, the zone-to-zone SL Index matrix will be first extracted and the ratio of each SL Index against the total SL Index leaving an origin zone will be calculated. This ratio will be used as the proportion of trips from this origin zone to one of its destination zones. Multiplying the ratio by the actual known trip productions of the origin zone, the corresponding estimate of the trips will be obtained for each OD pair. The calibration process described in the previous section may be needed for further adjusting the estimated OD trips.

Example: In a project, the SL Index between each OD pair was extracted from the SL Platform. Suppose the results are shown in Figure 34(a). Users can calculate the proportion of the SL Index with respect to the total SL Index of each origin zone and the results are shown in Figure 34(b). Given the actual zone-level productions (i.e., Figure 34(c)), the estimate of OD trips for each OD pair is calculated as the product of the corresponding proportion of the trip productions of the origin zone. For example, the estimated number of trips from zones 2 to 1 is $2{,}500 \times 33.3\% = 833$. This results in the initial estimate of OD trips shown in Figure 34(d). Similarly, the attractions can be calculated by columns (in one column, all ODs share the same destination zone).

If the actual OD table is known, we can further build a linear regression model to relate the initial estimates in Figure 34(d) with the actual number of trips for each OD pair. The built regression model can be considered in a similar project without an actual OD table for further adjusting its initial estimate of OD trips.

Zone	1	2	3	Sum
1	50	100	250	400
2	100	80	120	300
3	50	200	150	400
Sum	200	330	570	1100

(a) Original SL Index Matrix

Zone	1	2	3	Sum
1	12.5%	25.0%	62.5%	100%
2	33.3%	26.7%	40.0%	100%
3	12.5%	50.0%	37.5%	100%
Sum	--	--	--	--

(b) Proportions based on SL Index

Total Productions	Trips
Zone 1	**1,000**
Zone 2	**2,500**
Zone 3	**1,650**
Sum	5,150

(c) Actual Zone-level Productions

Zone	1	2	3	Sum
1	125	250	625	**1,000**
2	833	667	1,000	**2,500**
3	206	825	619	**1,650**
Sum	1,165	1,742	2,244	5,150

(d) Initial Estimate of OD Trips

Figure 34. Estimating OD Trips based on the SL Index Matrix.

Techniques and Tools for Working with the SL Products

[G23] Techniques and Tools for Preparing the Input Data

For efficiency, users are expected to be familiar with GIS tools such as ArcMap to develop a basic zone set for a large number of zones in some projects.

Example: The steps described below illustrate the major process to create a set of zones with the coordinates of the target zones by using the ESRI Desktop ArcMap. The data sources of the coordinates could be any target sites with the latitude and longitude information (e.g., sensor locations in the VDOT SmarterRoads data portal). The first 6 steps generally convert the points to a shape file layer with the basic zones (with a shape of the user defined buffer). The developed buffer layer needs to have a projected coordinate system selected by users (e.g., WGS 1984 World Mercator). Other tools such as desktop Arc GIS Pro that can perform these tasks can also be considered.

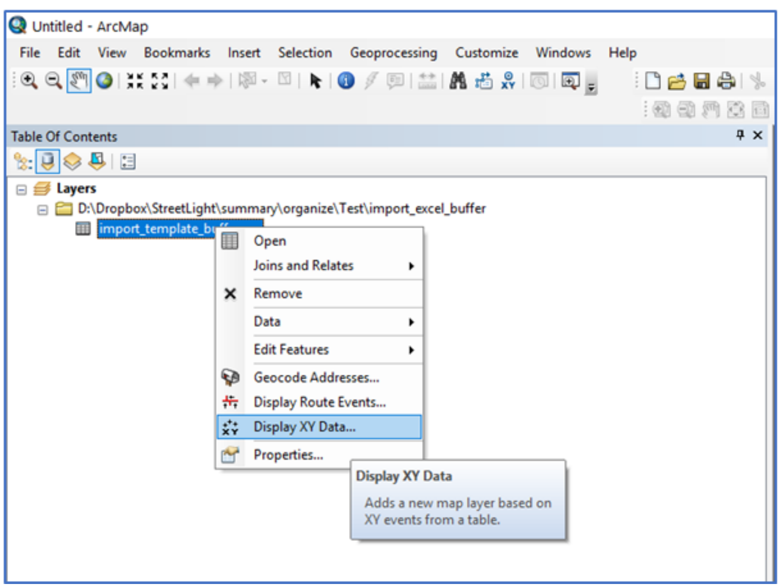
Given the projection, the buffer layer can be exported and zipped as the zone set input of the SL Platform. The imported zone further needs to be manually edited on the SL Platform to customize the travel direction and zone shape.

If users already have the polygon shapefiles with a projected coordinate system, it can be directly imported to SL Platform without using additional GIS tools. These polygon shapefiles could be the zone set from MPOs, census tract, etc.

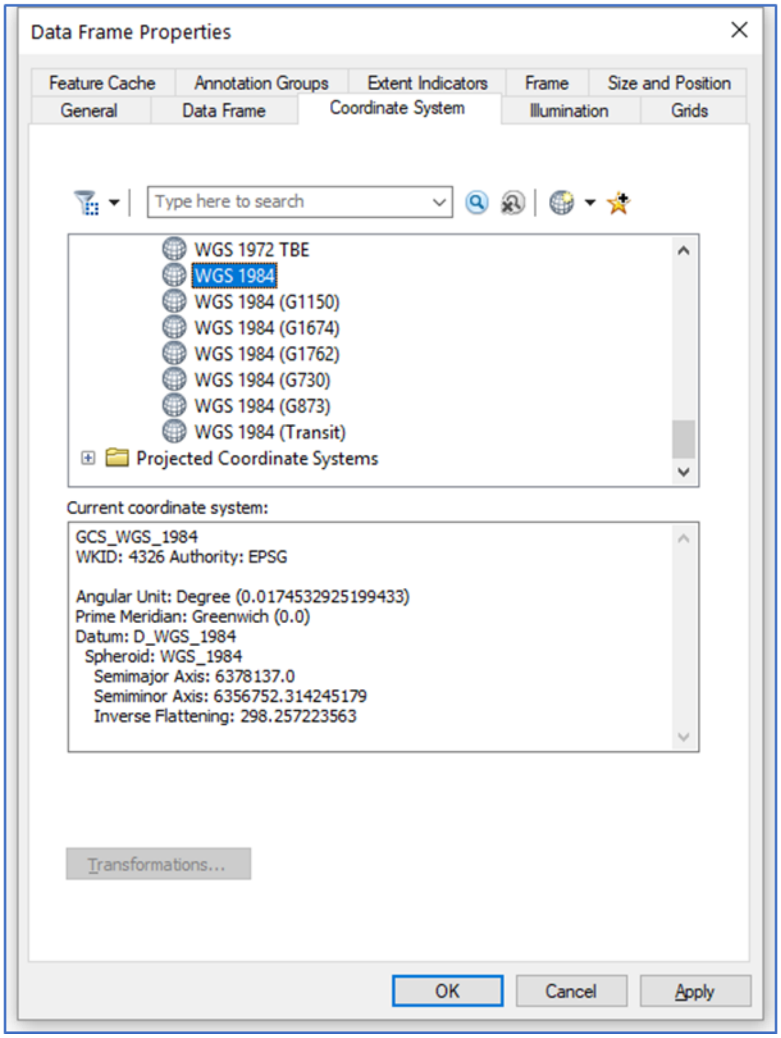
Step 1. Start from a .CSV format table with Latitude and Longitude.

Zone ID	Zone Name	Latitude	Longitude
1	Granby	36.88122	-76.282
2	38 St	36.8806	-76.2845
…	…	…	…
39	Llewellyn Ave	36.88453	-76.2833
40	W 35st	36.87843	-76.284

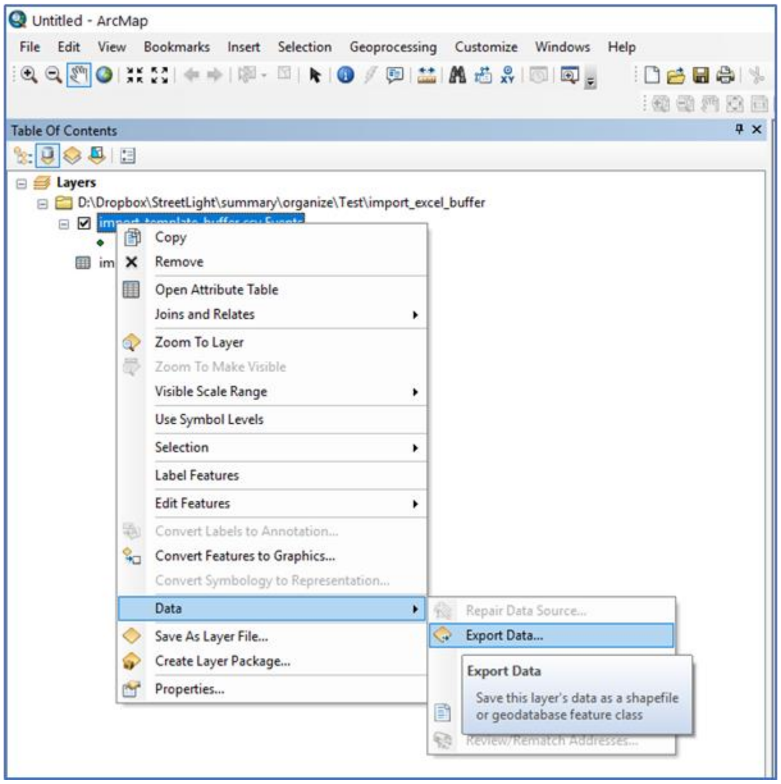
Step 2. Import the table to ArcMap and display XY Data.



Step 3. Set up the Coordinate System of the Data Frame: select Geographical Coordinate System (GCS) "WGS 1984"

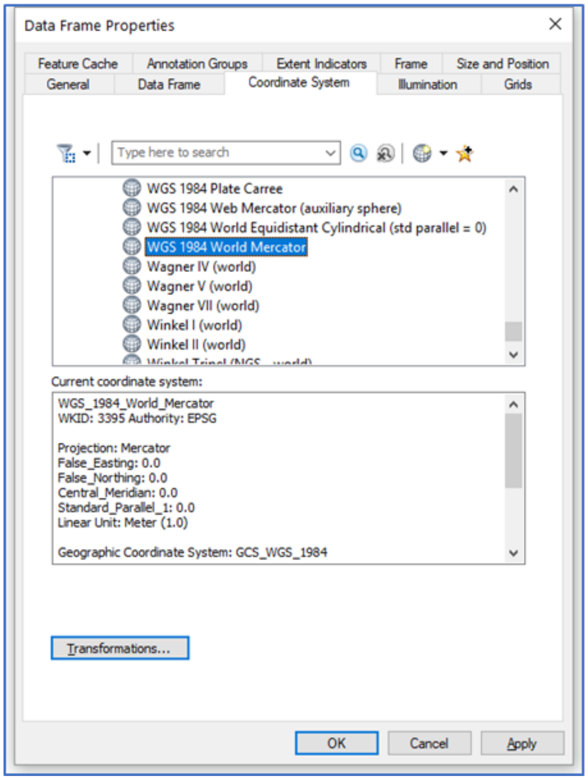


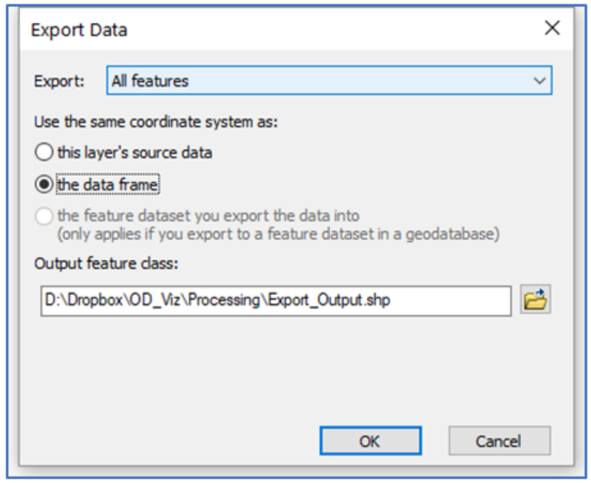
Step 4. Export the temporary displayed layer as a feature layer.

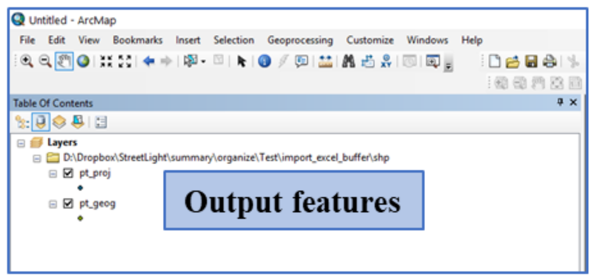


Step 5. Transfer the coordinate system from geographical coordinate system (GCS) to projected coordinate system (PCS)

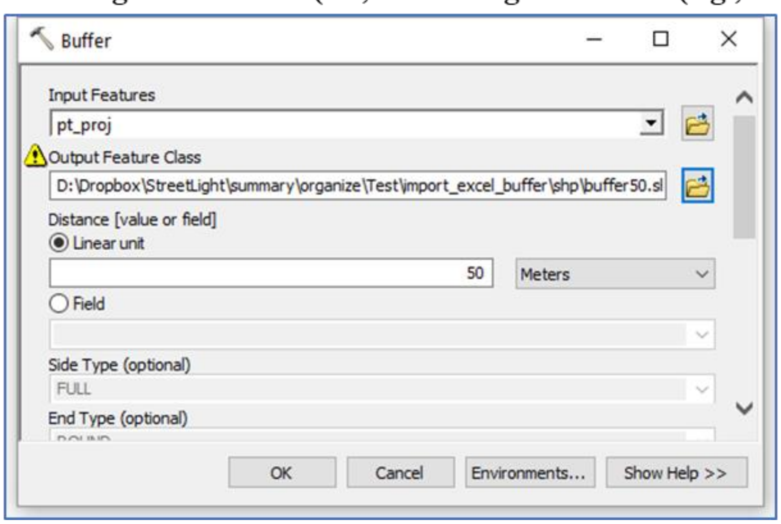
Firstly, change the Data Frame coordinate system to "PCS WGS 1984 World Mercator." Then export the feature layer using the same coordinate system as: "the data frame".



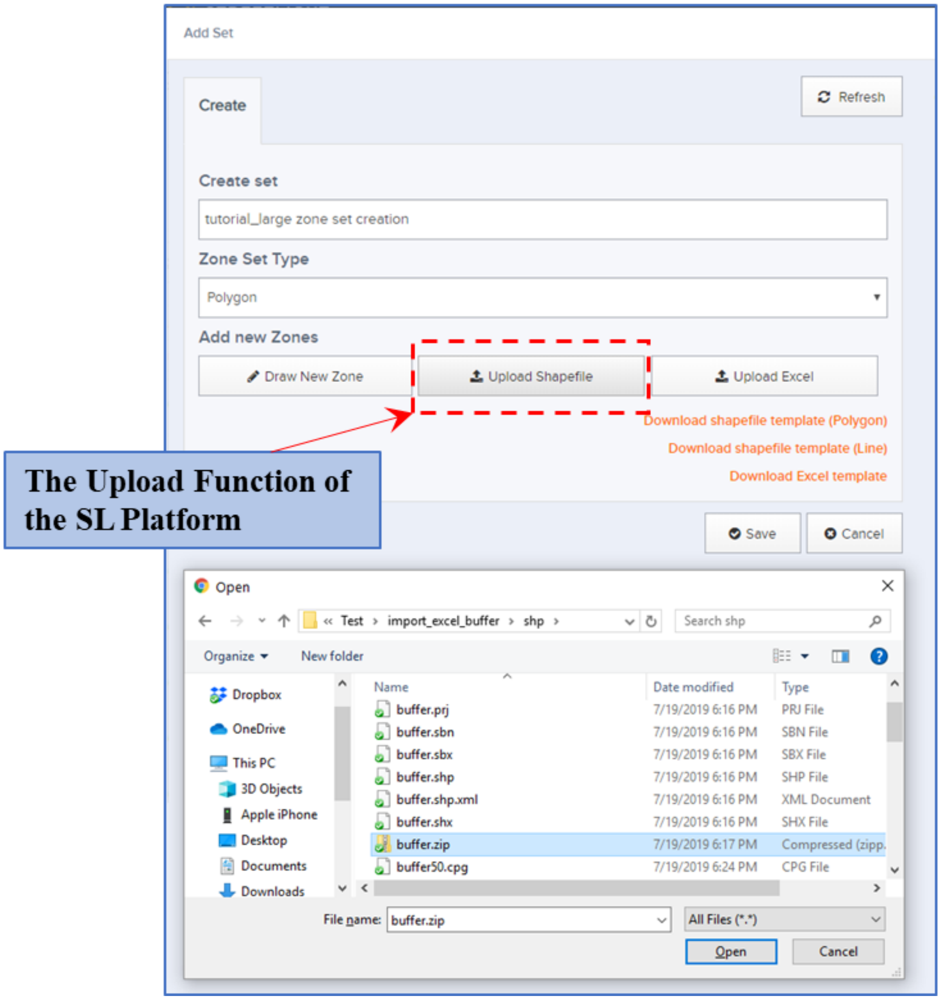




Step 6. Take the layer "pt_proj" from Step 5 as input and using the Buffer tool to create buffer with a given distance (i.e., Indicating linear unit (e.g., 50 meters)).

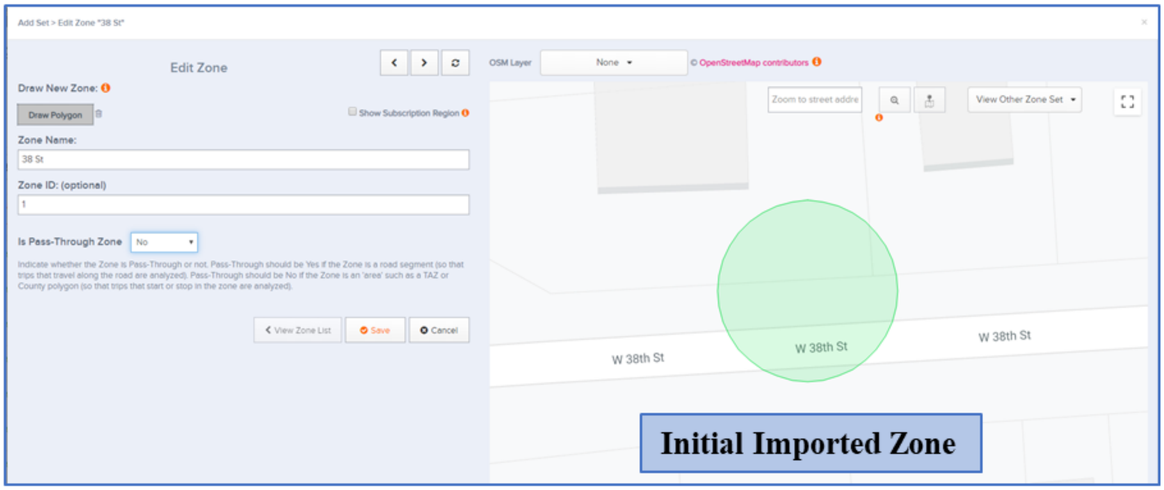


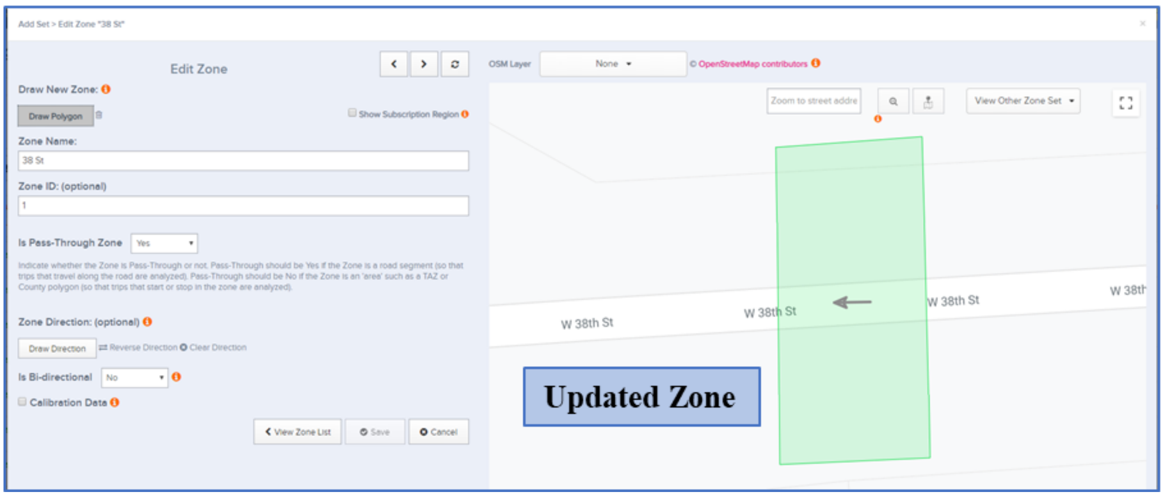
Step 7. Zip the buffer shapefiles and upload it to the SL Platform.



Visual Demo in Steps 7 and 8: Accessed August 20, 2019. Reprinted With Permission.

Step 8. Go through each zone on the SL Platform and manually edit the shape and attributes.





Visual Demo in Steps 7 and 8: Accessed August 20, 2019. Reprinted With Permission.

[G24] Techniques and Tools for Basic Analysis of the SL Index Data

Users may use any familiar tool (e.g., Excel) that can work with a .csv file to analyze the extracted SL Index data. SL Platform organizes the output file in tables that clearly describe the attributes of each column, for example, Zone ID, Zone Name, Day Type, SL Index, etc. Users can also download each project in a .zip file that can be imported into a spatial analysis tool such as ArcMap for further analysis and visualization.

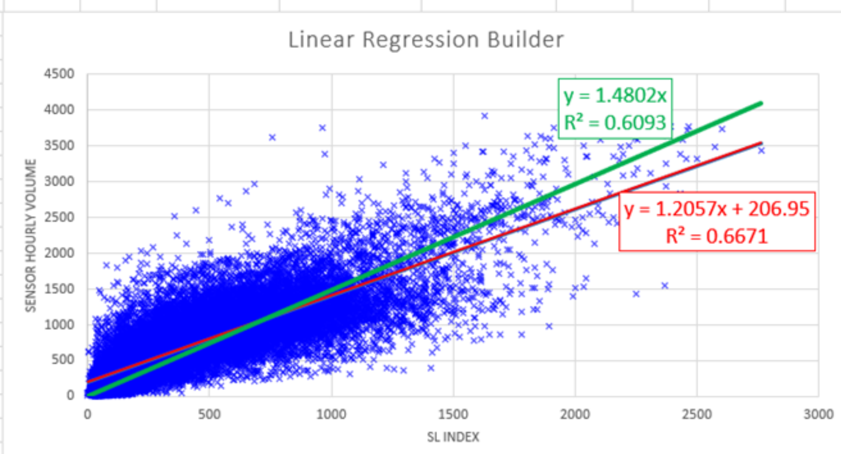
Example: An Excel template is provided for generating Linear Regression model between the SL Index and the benchmark sensor count. The steps described below illustrate the major process to use this template with customized user inputs.

Step 1. Replace the SL Index and Sensor Count with your own data.

ID	SL Index	Sensor Count
1	43	185
2	83	116
...	...	...
12345	207	236
12346	29	326

Step 2. The linear regression equations and plots will be automatically updated.

SL Index	Sensor Count
43	185
83	116
101	68
207	236
29	326
506	499
315	679
495	830
455	947
1051	1175
857	1187
653	1089
736	1120
479	1048
1368	1066
133	801
403	772
335	514
216	337



[G25] Techniques and Tools for Advanced Analysis of the SL Index Data

Users may need to be familiar with other statistical analysis tools such as R or SPSS to conduct advanced statistical modeling analysis and visualize the results through various types of charts. Users should be able to develop the calibration models with these tools and evaluate the model performance. For most of the descriptive analyses and visualization, Excel will be a convenient tool to meet the need.

Currently, SL provides a visualization tool, which has issues such as overlapping, hard to understand, limited functions. We developed an alternative web-based tool that enables users to directly filter and select on a map. http://senselane.com/od/bridge This is a demo showing retrieved SL Index results in a VDOT project. (Project type: OD with middlefilter).

APPENDIX B

ACRONYMS

Acronym	Full Name
AADT	Annual Average Daily Traffic
APE	Absolute Percentage Error
ATRI	American Trucking Research Institute
CI	Confidence Interval
GPS	Global Positioning System
IS	Interstate
LBS	Location-based Services
MAC	Media Access Control
MAG	Maricopa Association of Governments
MAPE	Mean Absolute Percentage Error
MPO	Metropolitan Planning Organization
OD	Origin-Destination
ODU	Old Dominion University
OSM	Open Street Map
PDC	Planning District Commissions
PE	Percentage Error
RMSE	Root Mean Square Error
RPO	Rural Planning Organization
SL	StreetLight
SR	State Route
TAZ	Traffic Analysis Zone
TCDS	Traffic Count Database System
TED	Traffic Engineering Division
TMPD	Transportation and Mobility Planning Division
TPO	Transportation Planning Organization
TRID	Transportation Research International Documentation
VA	Virginia
VASITE	Virginia Section Institute of Transportation Engineers
VDOT	Virginia Department of Transportation
VTRC	Virginia Transportation Research Council

APPENDIX C

SAMPLE TEMPLATE FOR COLLECTING DATA ON HOW SL DATA ARE USED IN SUPPORTING THE TASKS OF A VDOT PROJECT

Item	Description
Project Name	The name of the project.
Creation Date	The date the SL project created on the SL Platform.
Data Source	The SL data source used. (e.g., personal LBS or navigation GPS)
Analysis Period	The months or time period included in the analysis. (e.g., May and June 2018)
Specific Days	If specified, please list the day type/part selected. (e.g., Monday to Friday)
SL Analysis Type	Please specify which SL analysis is used. (e.g., OD Analysis)
Study Area	Please provide the study area of the project. (e.g., Hampton Road, VA)
SL Metrics Used	Please list which SL metric was used in the project. (e.g., AADT, SL index, trip duration, trip length, speed, etc.)
Additional Data Needed for Using SL Metrics	What were the other data that were collected for the project? (e.g., loop detector counts, travel demand of zones, etc.)
Data Quality	Has the quality of SL data been assessed for this project? If yes, what were the performance indicators (e.g., mean absolute percentage error) used? What was the range of accuracy?
Benchmark Data	What benchmark data were used if SL data quality assessment / calibration was performed?
Calibration Performed	Whether SL data calibration was performed in the project? (Yes or No)
Calibration Method	What was the method/procedure used for calibrating SL data? (e.g., linear regression model)
Limitations	What are the noted limitations of the SL data used in the project?
Additional Comments	Any additional comments on the use of SL data.

APPENDIX D

DEFINITION TABLE OF TERMINOLOGY

Category	Glossary	Description
Transportation Terms	AAHT	Annual Average Hourly Traffic (AAHT) is the average vehicle count per hour at a given location throughout the course of a year
	MADT	Monthly Average Daily Traffic (MADT) is the average vehicle count per day at a given location throughout the course of one month
	Transportation Model	A transportation model, specifically, typically uses a transportation system, plus location variables, to understand traffic flows
	AADT	Average Annualized Daily Traffic. Throughout the course of one year, the average vehicle count per day at a given location
	AADPT	AADPT stands for Annual Average Daily Pedestrian Traffic. AADPT represents the average pedestrian count per day at a given location throughout the course of one year and is estimated via loop counters, blue tooth sensors, temporary counters, and more
	AADBT	AADBT stands for Annual Average Daily Bicycle Traffic. AADBT represents the average bicycle count per day at a given location throughout the course of one year and is estimated via loop counters, Bluetooth sensors, temporary counters, and more
General Info	Marketing Consultant Subscription	A type of subscription available in *StreetLight InSight*. Marketing consultant subscriptions provide unlimited access to transportation data to consulting firms to use in their proposals
	Project Subscription	A type of subscription available in *StreetLight InSight*, also referred to as a "Pay Per Use" subscription
	Regional Subscription	This subscription type is available to some public agencies and their designated consultants and is only limited by a designated subscription region
	Evaluation	In certain cases, StreetLight offers access to our platform through a guided evaluation
Data Sources	Location-Based Services Data	Location-Based Services is one of our data sources, originating from smartphone apps that use opt-in location-based services
	Navigation-GPS Data	Navigation-GPS is one of our data sources that originates from connected cars, smartphones using GPS navigation, and connected commercial trucks
Downloads and Visualizations	Confidence Range	The StreetLight InSight® confidence range--commonly referred to as a prediction interval in statistics--is an estimate, within a given probability, of an interval in which the future observations will fall

APPENDIX E

SURVEY QUESTIONNAIRES

Two different hyperlinks (one for non-users of StreetLight and one for users of StreetLight) were emailed to prospective survey respondents so that they could complete the survey electronically.

Streetlight Data Questionnaire (for Non-Users)

This survey aims to collect opinions of transportation professionals who might be interested in using the StreetLight Data metrics in their projects in the future. StreetLight Data metrics provide users Origin-Destination Matrices, AADT estimates, etc. This survey is part of an on-going project led by a research team at Old Dominion University (ODU) with support from the Virginia Department of Transportation (VDOT) and the Virginia Transportation Research Council (VTRC). The Principle Investigators of the project are Drs. Hong Yang and Mecit Cetin at ODU. The collected information will help VDOT evaluate the usefulness of the StreetLight Data metrics and develop guidance for VDOT to make the best use of the data. The survey will take about 10 minutes to complete. If you have any questions about the survey, please contact Hong Yang (hyang@odu.edu) or Tancy Vandecar-Burdin (tvandeca@odu.edu). We greatly appreciate it if you could also share the survey with other colleagues or friends of interest.

Disclaimer: Your participation is voluntary and your responses are anonymous. The survey responses will only be analyzed and reported in an aggregated way.

1. In which STATE is your work organization located? (If VA, please answer Question 2; If other, please answer Question 3.)

- Enter STATE______ (pull down list of States)

2. [VIRGINA RESPONDENTS] I work for (Please select one):

a. VDOT Central Office TMPD ○
b. VDOT Central Office (other than TMPD) ○
c. VDOT District Office ○
d. VTRC ○
e. An MPO/RPO (Metropolitan/Rural Planning Organization) in VA ○
f. A PDC (Planning District Commissions) in VA ○
g. A VDOT consultant ○
h. A City or Municipality in VA ○
i. Others__________________

3. [RESPONDENTS FROM STATES OTHER THAN VIRGINIA] I work for (Please select one):

a. USDOT ○
b. State DOT ○
c. An MPO/RPO (Metropolitan/Rural Planning Organization) ○
d. A PDC (Planning District Commissions) ○
e. A State DOT consultant ○
f. A City or Municipality ○
g. Other__________________

4. What is your main job responsibility at your agency (organization)?
a. Project Manager ○
b. Designer ○
c. Planner ○
d. Data Analyst ○
e. Other__________________

5. Which of the following MAJOR data source(s) do you commonly use in your projects? (select all that apply)
a. AirSage data ☐
b. Bluetooth reader data ☐
c. License-plate matching ☐
d. Loop or other on-road sensor data ☐
e. Manual surveys ☐
f. 3rd part speed or travel time data (e.g., INRIX, HERE, TomTom, etc.) ☐
g. Bikeshare data ☐
h. Taxi O-D trips from GPS devices ☐
i. None of the above ☐
j. Other_____________________

6. In general, field data capturing actual O-D trips would be valuable for my transportation projects:
a. Strongly agree ○
b. Agree ○
c. Neither Agree or Disagree ○
d. Disagree ○
e. I don’t know ○

7. Have you heard about Streetlight Data? (If YES, please answer Questions 8-11. If NO, exit survey)
a. Yes ○
b. No ○

8. Do you know anyone in your work organization who has used Streetlight Data for a project?

a. Yes ○

b. No ○

9. Do you believe Streetlight Data metrics (e.g., O-D index) might be useful for some of your projects?

a. Yes ○

b. No ○

c. Don't know ○

10. Do you have access to Streetlight Data metrics?

a. Yes ○

b. No ○

11. Have you considered possible use of Streetlight Data metrics in projects? (If YES, please answer Questions 12 and 13. If NO, exit survey)

a. Yes ○

b. No ○

12. What are the main reasons you have not used Streetlight Data metrics? (select all that apply)

a. Need more training before using the data ☐

b. The data do not meet the needs of my projects ☐

c. I find better alternative data sources ☐

d. Do not currently have access to Streetlight Data metrics ☐

e. Other reasons___________________

13. Do you have any of the following suggestions for improving a Streetlight Data user's experience in the future? (select all that apply)

f. Increase sample sizes and coverage area ☐

g. Improve the interface of the Streetlight InSight Platform ☐

h. Further simplify data acquisition process ☐

i. Other suggestions___________________

StreetLight Data Questionnaire (for Existing Users)

This survey aims to collect opinions of transportation professionals who have used the StreetLight data in their projects. StreetLight Data metrics provide users Origin-Destination Matrices, AADT estimates, etc. This survey is part of an on-going project led by a research team at Old Dominion University (ODU) with support from the Virginia Department of Transportation (VDOT) and the Virginia Transportation Research Council (VTRC). The Principle Investigators of the project are Drs. Hong Yang and Mecit Cetin at ODU. The collected information will help VDOT evaluate the usefulness of StreetLight Data metrics and develop guidance for VDOT to make the best use of the data. The survey will take about 20 minutes to complete. If you have any questions about the survey, please contact Hong Yang (hyang@odu.edu) or Tancy Vandecar-Burdin (tvandeca@odu.edu). We greatly appreciate it if you could also share the survey with other colleagues or friends of interest.

Disclaimer: Your participation is voluntary and your responses are anonymous. The survey responses will only be analyzed and reported in an aggregated way.

Part A

1. In which STATE is your work organization located? (If VA, please answer Question 2; If other, please answer Question 3.)

- Enter STATE______ (pull down list of States)

2. [VIRGINIA RESPONDENTS] I work for (Please select one):

a. VDOT Central Office TMPD ○
b. VDOT Central Office (other than TMPD) ○
c. VDOT District Office ○
d. VTRC ○
e. An MPO/RPO (Metropolitan/Rural Planning Organization) in VA ○
f. A PDC (Planning District Commissions) in VA ○
g. A VDOT consultant ○
h. A City or Municipality in VA ○
i. Other__________________

3. [RESPONDENTS FROM OTHER STATES] I work for (please select one):

a. USDOT ○
b. State DOT ○
c. An MPO/RPO (Metropolitan/Rural Planning Organization) ○
d. A PDC (Planning District Commissions) ○
e. A State DOT consultant ○
f. A City or Municipality ○
g. Other__________________

4. What is your main job responsibility at your agency (organization)?

a. Project Manager ○
b. Designer ○
c. Planner ○
d. Data Analyst ○
e. Other__________________

5. When was the last time you used the StreetLight Data metrics?

a. Within the past 3 months ○
b. 4-6 months ago ○
c. 7-12 months ago ○
d. More than 1 year ago ○

6. Please provide a title or a brief description of one or more of your TYPICAL project(s) that have used the StreetLight Data metrics. (Several questions that follow will focus on this/these TYPICAL project(s))

- Project Title/Description: ______________________

7. Which of the following focus areas describe your typical use of the StreetLight Data metrics? (select all that apply)

a. Public transit systems ☐
b. Traffic congestion ☐
c. Crash reduction/safety ☐
d. Traffic demand modeling ☐
e. Infrastructure maintenance/ construction/ work zone ☐
f. Environmental study ☐
g. Commercial traffic/freight/logistics ☐
h. Other______________________

8. Considering your TYPICAL Project(s), for which of the following specific tasks have you employed the StreetLight Data metrics? (select all that apply)

a. OD analysis ☐
b. Road speed analysis ☐
c. Traffic flow/volume analysis ☐
d. Attraction analysis ☐
e. Travel time analysis ☐
f. Route choice analysis ☐
g. Network analysis ☐
h. Mode choice analysis ☐
i. Travel purpose analysis ☐
j. Socioeconomic-factor/demographic analysis ☐

k. Other______________________

9. Which of the following major elements are included in your TYPICAL project(s) that have used StreetLight Data metrics? (select all that apply)

a. Freeway ☐
b. Arterial ☐
c. Urban street ☐
d. Port/airport ☐
e. TAZs / Census tract ☐
f. Toll road ☐
g. Ramp / Shoulder ☐
h. Traffic signal ☐
i. Work zone ☐
j. Parking facility ☐
k. Other______________________

10. Other than StreetLight Data metrics, which of the following MAJOR data source(s) have you used in your TYPICAL project(s)? (select all that apply)

k. AirSage data ☐
l. Bluetooth reader data ☐
m. License-plate matching ☐
n. Loop or other on-road sensor data ☐
o. Manual surveys ☐
p. Third-party speed or travel time data (e.g., INRIX, HERE, TomTom, etc.) ☐
q. Bikeshare data ☐
r. Taxi O-D trips from GPS devices ☐
s. None of the above ☐
t. Other______________________

11. Which of the following features of the StreetLight Data metrics have you used in your TYPICAL Project(s)? (select all that apply)

a. Non-passthrough zone ☐
b. Bi-directional passthrough zone ☐
c. Passthrough zone with one direction ☐
d. Road segments ☐
e. Other______________________

12. How did you create the zone sets in your TYPICAL Project(s) using StreetLight Data metrics? (select all that apply)

a. Draw on StreetLight InSight platform ☐
b. Upload customized GIS shapefiles from your local PC ☐
c. Use the existing zones created by others on StreetLight InSight platform ☐

d. Other______________________

13. Which of the functionalities available on the StreetLight InSight platform have you primarily used in your TYPICAL Project(s)? (select all that apply)

a. O-D analysis ☐
b. Zone activity analysis ☐
c. Visitor home and work analysis ☐
d. Estimated 2017 AADT values [beta] ☐
e. Segment analysis ☐
f. Traffic diagnostics ☐
g. Multi-Mode analytics - bike and pedestrian traffic ☐
h. Other______________________

14. Which types of StreetLight Data sources have you mainly used in your TYPICAL project(s)? (select all that apply)

a. Personal location-based services (LBS) ☐
b. Personal navigation-GPS ☐
c. Commercial navigation-GPS ☐

15. Which of the following StreetLight Data metrics have you primarily used as input in your TYPICAL project(s)? (select all that apply)

a. StreetLight trip index ☐
b. Raw trip count ☐
c. Other_______________________

16. Have you used any of the following metrics from StreetLight Data in your TYPICAL project(s)? (select all that apply)

a. Trip attribute (length/duration/speed/trip circuity) ☐
b. Trip purpose ☐
c. Census demographic ☐
d. None ☐

17. The TYPICAL Project(s) that have used StreetLight Data metrics are primarily in: (select all that apply)

a. Urban areas ☐
b. Rural areas ☐
c. Port(s) ☐
d. Airport(s) ☐
e. Other______________________

18. What is the approximate spatial coverage level of your zone sets in your TYPICAL project(s) using StreetLight Data metrics? (select all that apply)

a. Intersection level ☐
b. Corridor level ☐

c. Neighborhood level ☐
d. City / Township level ☐
e. Region / County level ☐
f. State level ☐
g. Other____________________

19. What is the temporal scale of analysis primarily applied in your TYPICAL project(s) that have used StreetLight Data metrics? (select all that apply)

a. Hourly ☐
b. Daily ☐
c. Weekly ☐
d. Monthly ☐
e. Seasonal ☐
f. Yearly ☐

20. The current minimum aggregation interval on the StreetLight InSight platform is by hour. Have you ever had the need for a smaller level of data aggregation (e.g., 15-min or 30-min instead of 1-hour aggregation) in your TYPICAL project(s)?

a. Yes ○
b. No ○

21. How satisfied are you with the sample size (e.g., actual number of trips) you got from StreetLight Data for your TYPICAL project(s)?

a. Very satisfied ○
b. Somewhat Satisfied ○
c. Somewhat dissatisfied ○
d. Very dissatisfied ○
e. Not applicable--I don't check the sample size ○

22. For the purpose of checking data quality in your TYPICAL project(s), please check all the following that apply:

a. I compare StreetLight Data metrics with another field data source ☐
b. I look for internal consistency (or biases) within different time periods or spatial locations within StreetLight Data metrics ☐
c. I calculate summary statistics and check variance/deviations/abnormal values within StreetLight Data metrics ☐
d. I compare StreetLight Data metrics to our/field expectations ☐
e. None of the above ☐
f. Other__________

23. For StreetLight Data metrics use, what is the MAXIMUM percentage of error that would be typically acceptable on your PLANNING level projects?

a. 0%-5% ○

b. 6%-10% ○
c. 11%-20% ○
d. 21%-30% ○
e. 31%-40% ○
f. > 40% ○
g. I don't deal with planning projects ○
h. No opinion ○

24. For StreetLight Data metrics use, what is the MAXIMUM percentage of error that would be typically acceptable on your DESIGN level projects?

a. 0%-5% ○
b. 6%-10% ○
c. 11%-20% ○
d. 21%-30% ○
e. 31%-40% ○
f. > 40% ○
g. I don't deal with design projects ○
h. No opinion ○

25. For StreetLight Data metrics use, what is the MAXIMUM percentage of error that would be typically acceptable on your OPERATIONS level projects?

a. 0%-5% ○
b. 6%-10% ○
c. 11%-20% ○
d. 21%-30% ○
e. 31%-40% ○
f. > 40% ○
g. I don't deal with operations projects ○
h. No opinion ○

26. Calibration is a way to improve the usefulness of the retrieved StreetLight Data metrics, especially for the indexes or percentage-formatted data. For example, StreetLight indexes can be calibrated to trip counts by multiplying a scaling factor to better describe pass-through trips. After downloading the metrics from the StreetLight InSight platform, have you calibrated the followings items in your TYPICAL Project(s)? (select all that apply)

a. AADT ☐
b. Direction flow ☐
c. OD table ☐
d. Others____________

Part B

(Please answer the questions below based on your overall experiences with StreetLight Data metrics.)

27. On average, how frequently do you use StreetLight Data metrics?

a. Less than once per month ○
b. 1-2 times / month ○
c. 3-4 times / month ○
d. 5-6 times / month ○
e. 7-10 times / month ○
f. > 10 times / month ○

28. Have you ever needed day-by-day StreetLight Data metrics for more than one week (e.g., needed the index for each day between 01/01/2019 to 01/14/2019; NOT the average index of these 14 days)?

a. Yes ○
b. No ○

29. Would you be interested in more advice or instructions on the use of the StreetLight Data metrics? (If YES, please answer Question 30. If NO, go to Question 31)

a. Yes ○
b. No ○

30. Please rank the following assistance/instruction needs from HIGH (Rank 1) to LOW (Rank 7) priorities:

Rank 1:____ 2:____ 3:____ 4:____ 5:____ 6:____ 7:____

a. How to upload customized zones
b. How to draw zones/lines
c. How to conduct O-D analysis
d. How to conduct road segments analysis
e. How to calibrate data
f. How to retrieve the sample size
g. How to identify the data error

31. Without StreetLight Data products, which of the following statements would MOSTLY apply to you?

a. I would do just fine WITHOUT StreetLight Data metrics for most of my projects (I can obtain alternative data sources relatively easily) ○
b. I would collect alternative data but it would be very time consuming or costly ○
c. I would use other existing data but the results might not be as useful, reliable, or accurate ○
d. Other_________ ○

32. How would you score your overall experience of using StreetLight Data metrics? (0: Very Poor -------- 10: Excellent)
Your Overall Evaluation Score: ________________________

33. Please provide your experiences / thoughts about challenges and difficulties when using StreetLight Data metrics. (if any)
Your Concerns: ________________________

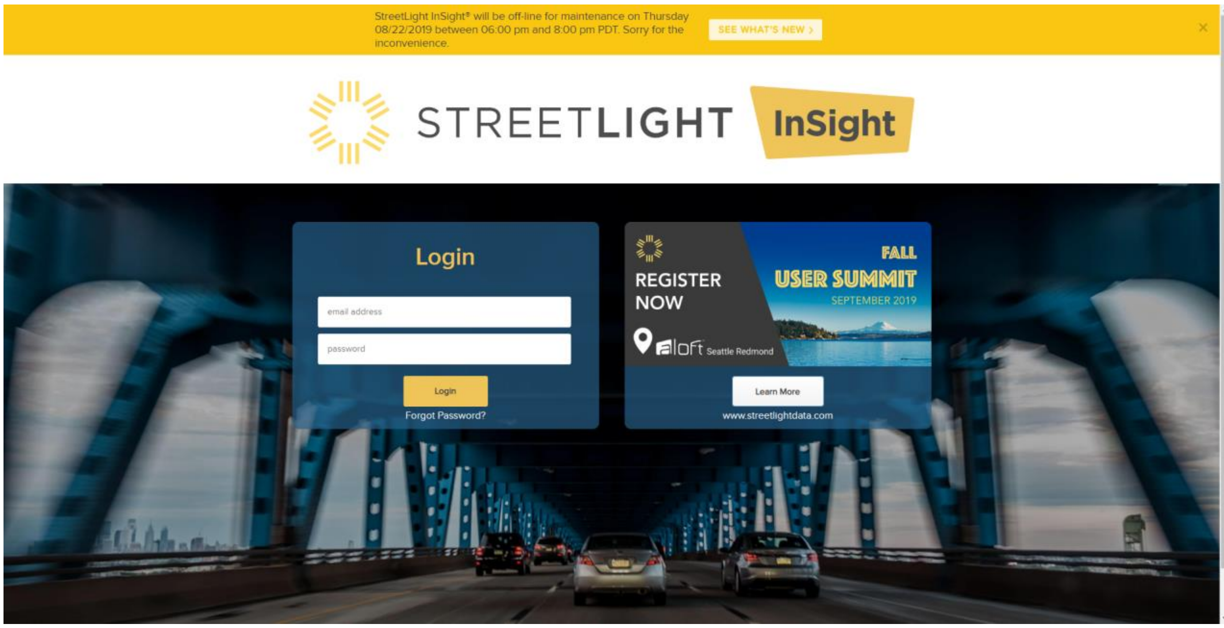
34. Do you have any other comments? (if any)
Comments: ________________________

APPENDIX F

STREETLIGHT DATA BASIC TUTORIALS

After logging into the StreetLight Data platform, there are three basic analysis: OD Analysis, Zone Activity Analysis, Segment Analysis.

Weblink: https://www.streetlightdata.com/

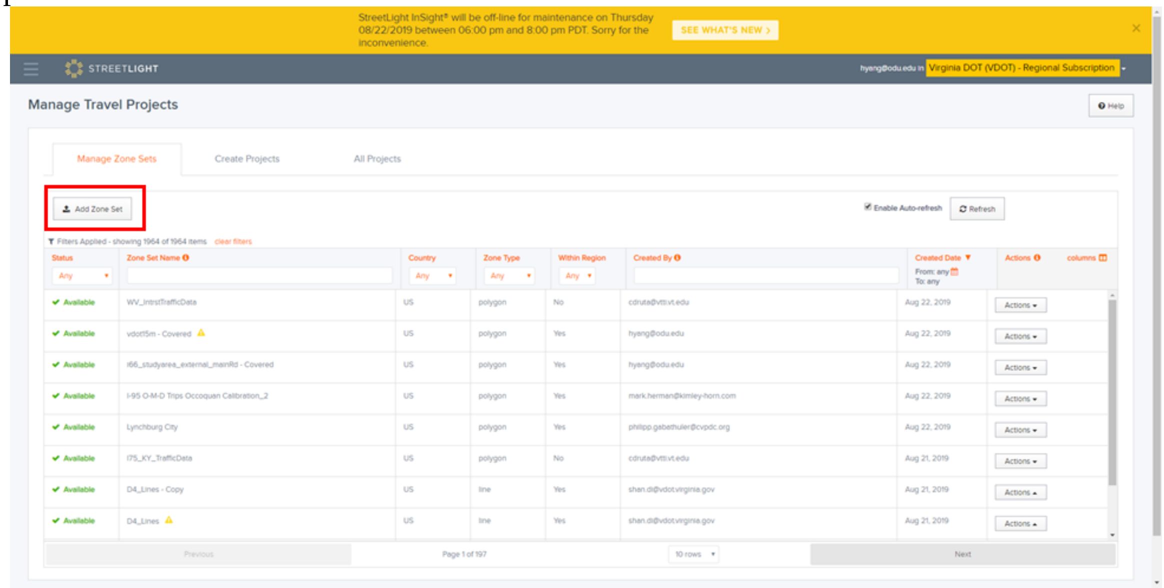


Accessed August 15, 2019. Reprinted With Permission.

Tutorial 1. OD Analysis

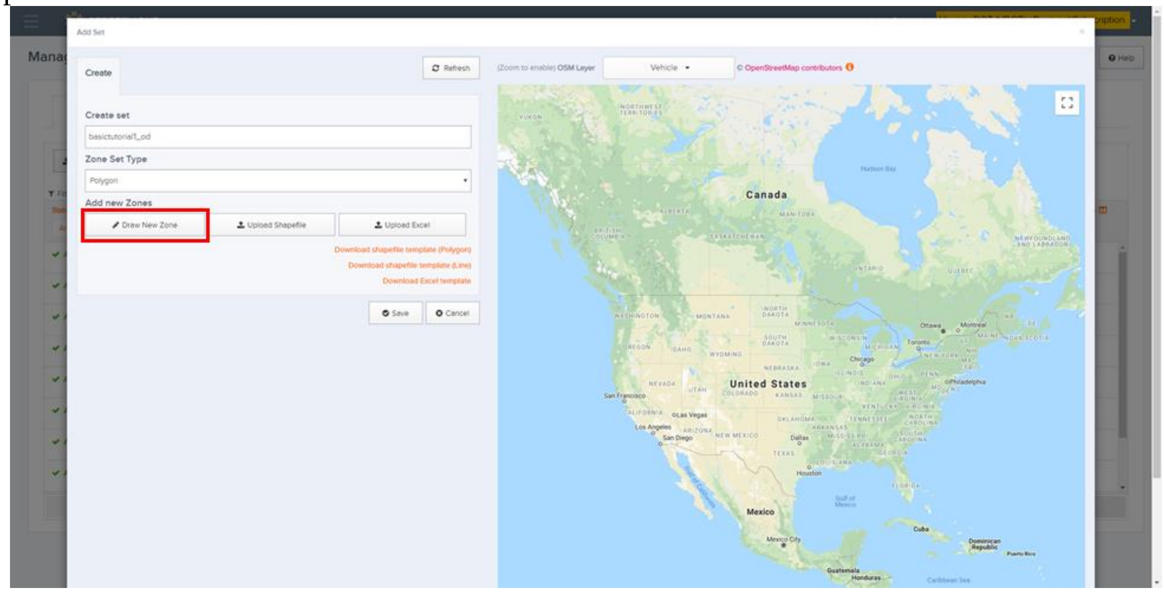
Goal: Explore the communication patterns between ***Norfolk*** *and* ***Hampton****.*

Step 1. Click “Add Zone Sets”.



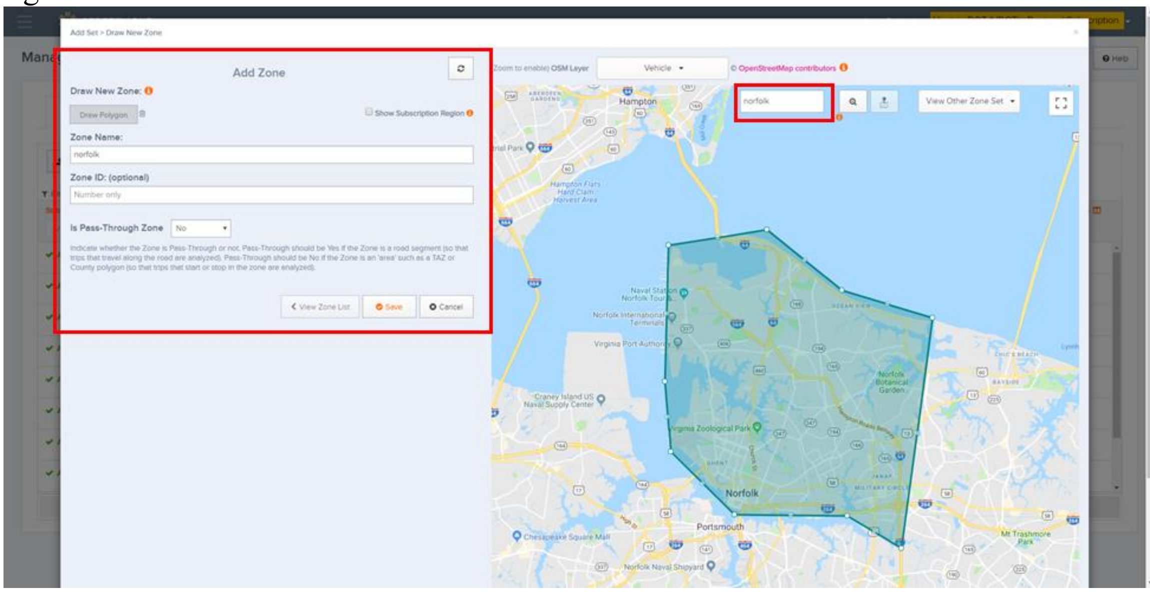
Accessed August 15, 2019. Reprinted With Permission.

Step 2. Click “Draw New Zone”.



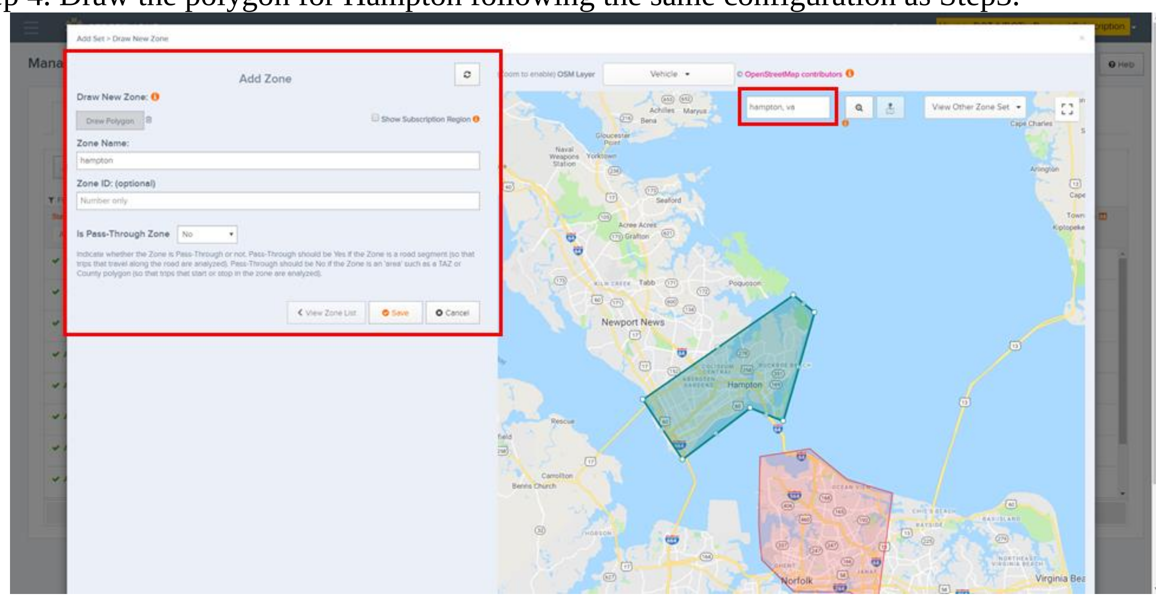
Accessed August 15, 2019. Reprinted With Permission.

Step 3. Type in “Norfolk” in the search box. Then, draw a polygon and set it as a “No” pass-through zone. Click “Save” to save the result.



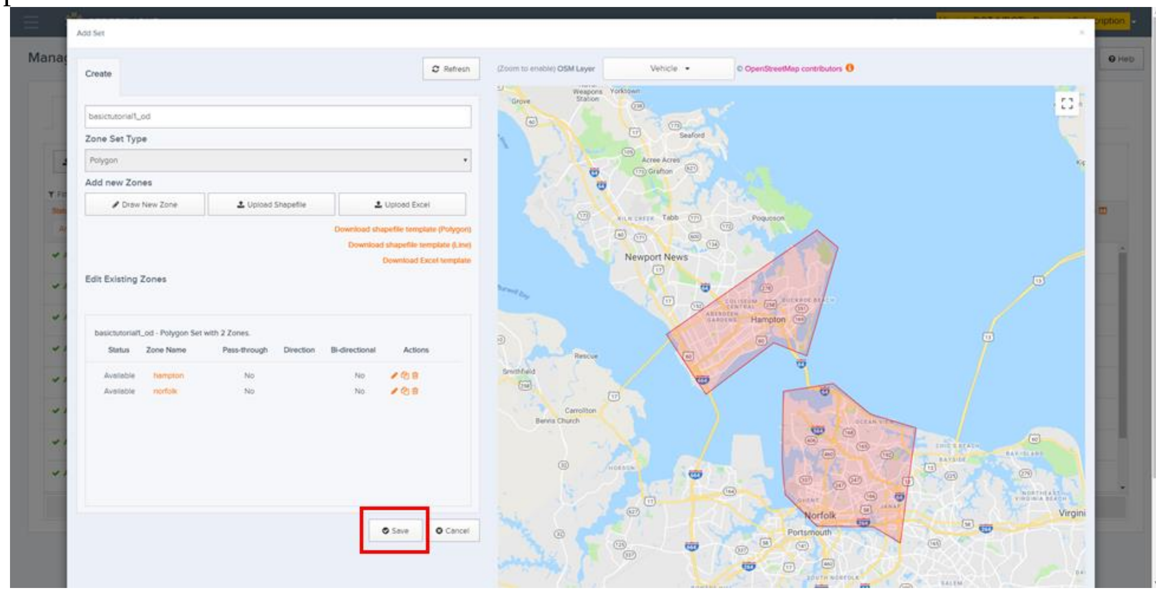
Accessed August 15, 2019. Reprinted With Permission.

Step 4. Draw the polygon for Hampton following the same configuration as Step3.



Accessed August 15, 2019. Reprinted With Permission.

Step 5. Click “Save” button to save the created zone set.



Accessed August 15, 2019. Reprinted With Permission.

Step 6. Click “Create Projects”, following the settings as below.

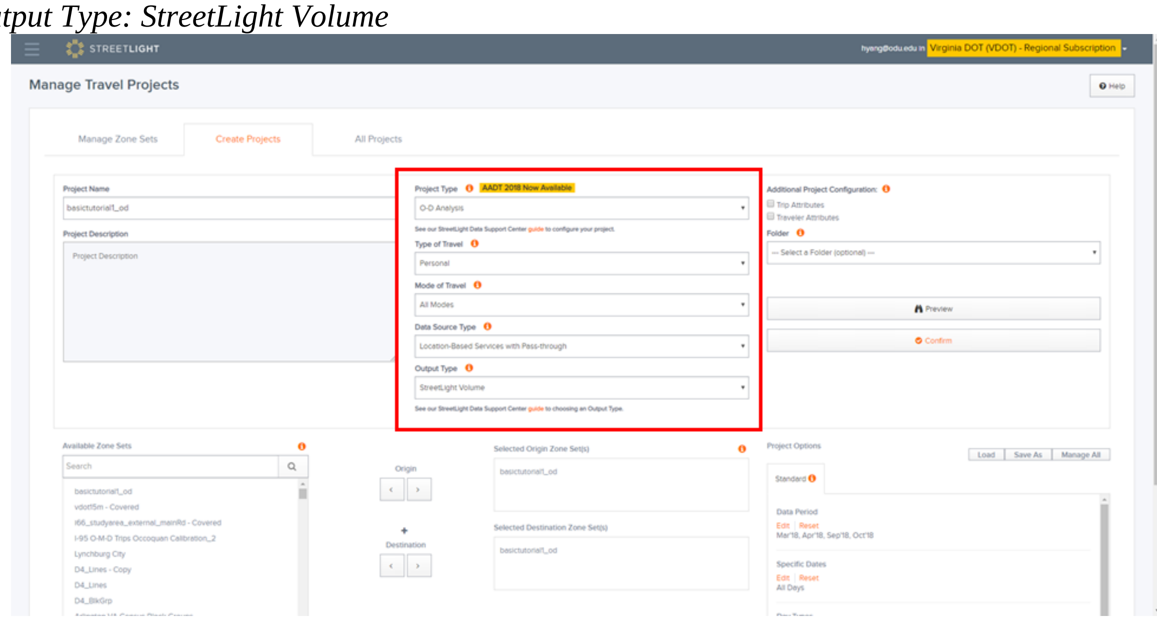
Project Type: O-D Analysis

Type of Travel: Personal

Mode of Travel: All Modes

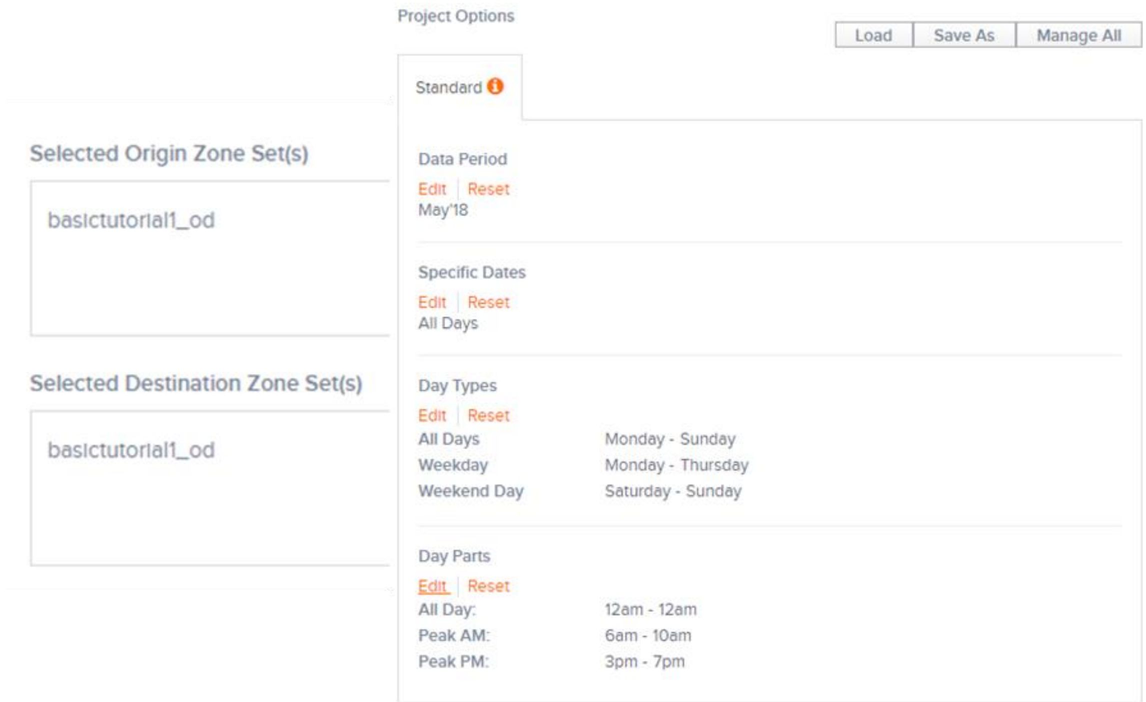
Data Source Type: Location-Based Services with Pass-through

Output Type: StreetLight Volume



Accessed August 15, 2019. Reprinted With Permission.

Step 7. Configure the Project Options. Make sure the created zone is set as both the Origin and Destination Zone Sets. Then click "Confirm" to build the project.



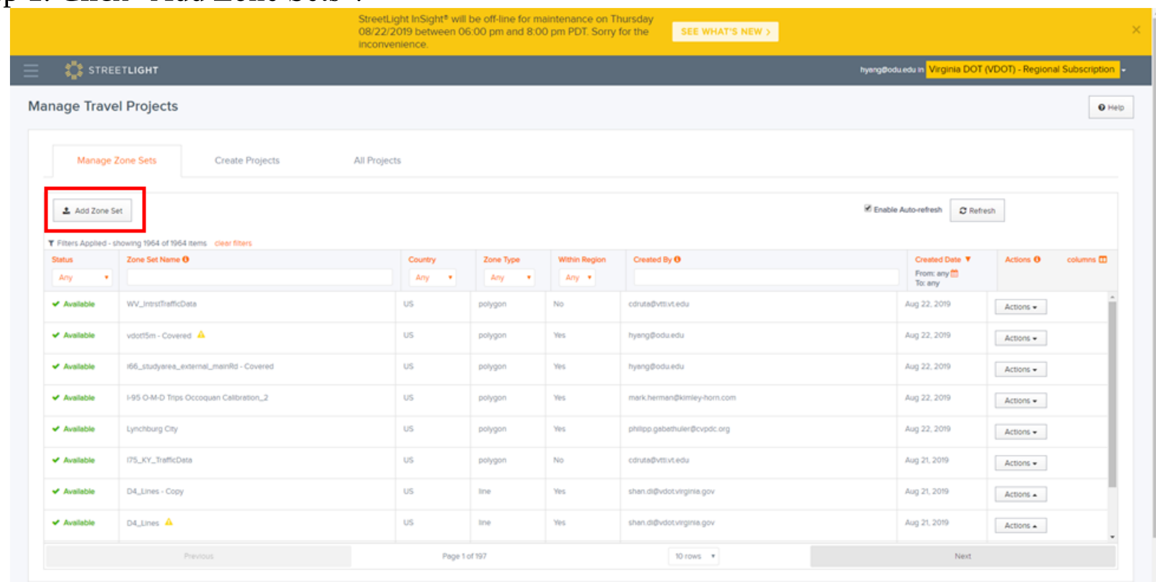
Accessed August 15, 2019. Reprinted With Permission.

Step 8. Find your project in the "All Projects" panel and download the analysis results. All results will be stored in a .zip file.
Step 9. Analyze the detailed results. See examples are shown in Tutorial 4.

Tutorial 2. Zone Activity Analysis

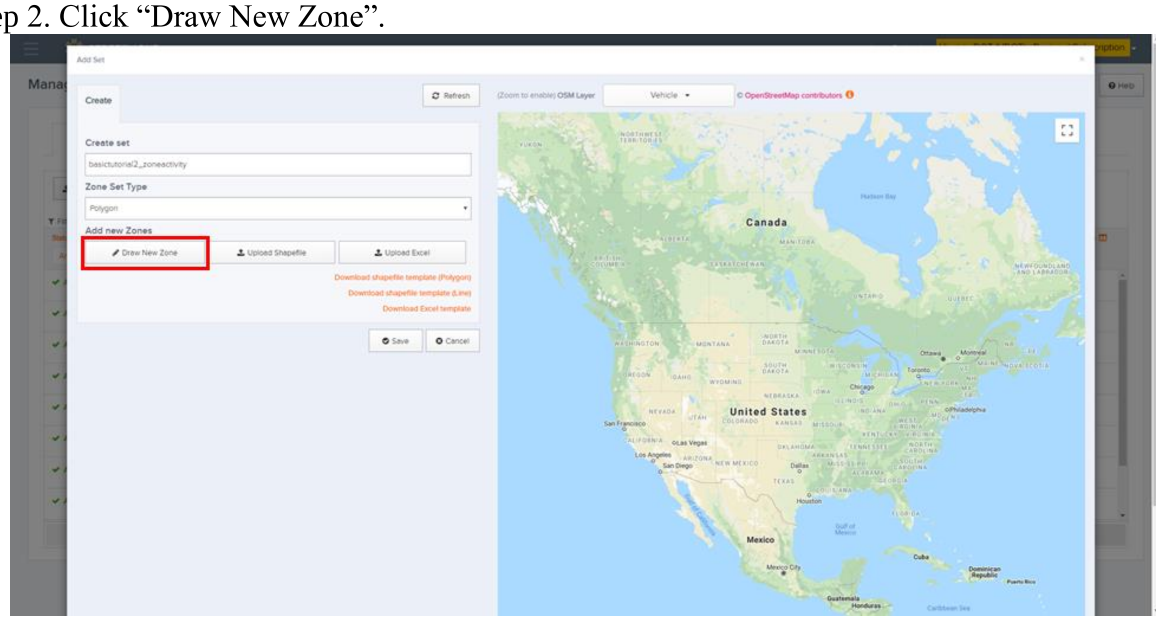
Goal: Explore the communication patterns on I-64 bridge.

Step 1. Click "Add Zone Sets".



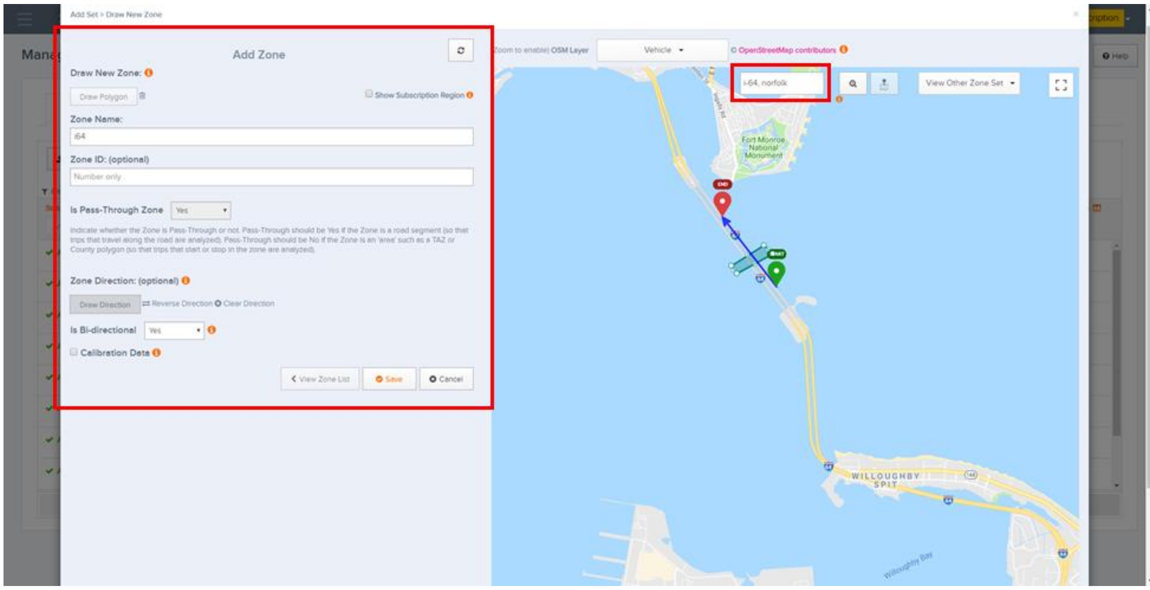
Accessed August 15, 2019. Reprinted With Permission.

Step 2. Click "Draw New Zone".



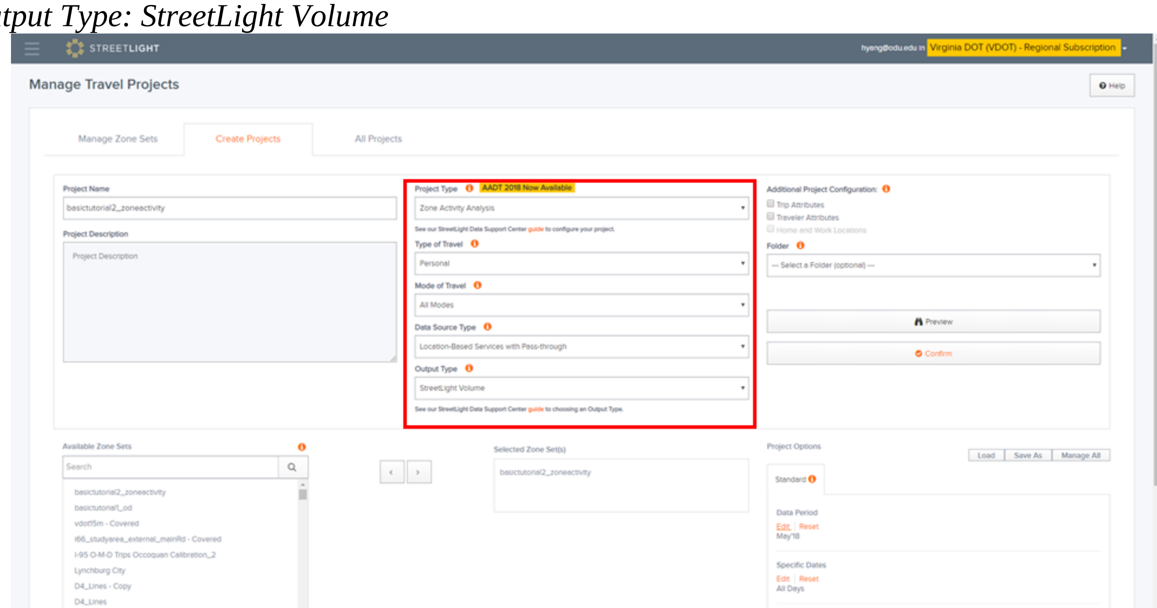
Accessed August 15, 2019. Reprinted With Permission.

Step 3. Firstly, type in “I-64, Norfolk” in the search box. Secondly, draw a polygon and set it as a “Yes” pass-through zone. Then, draw a direction and set it as “Bi-directional”. Click “Save” to save the result.



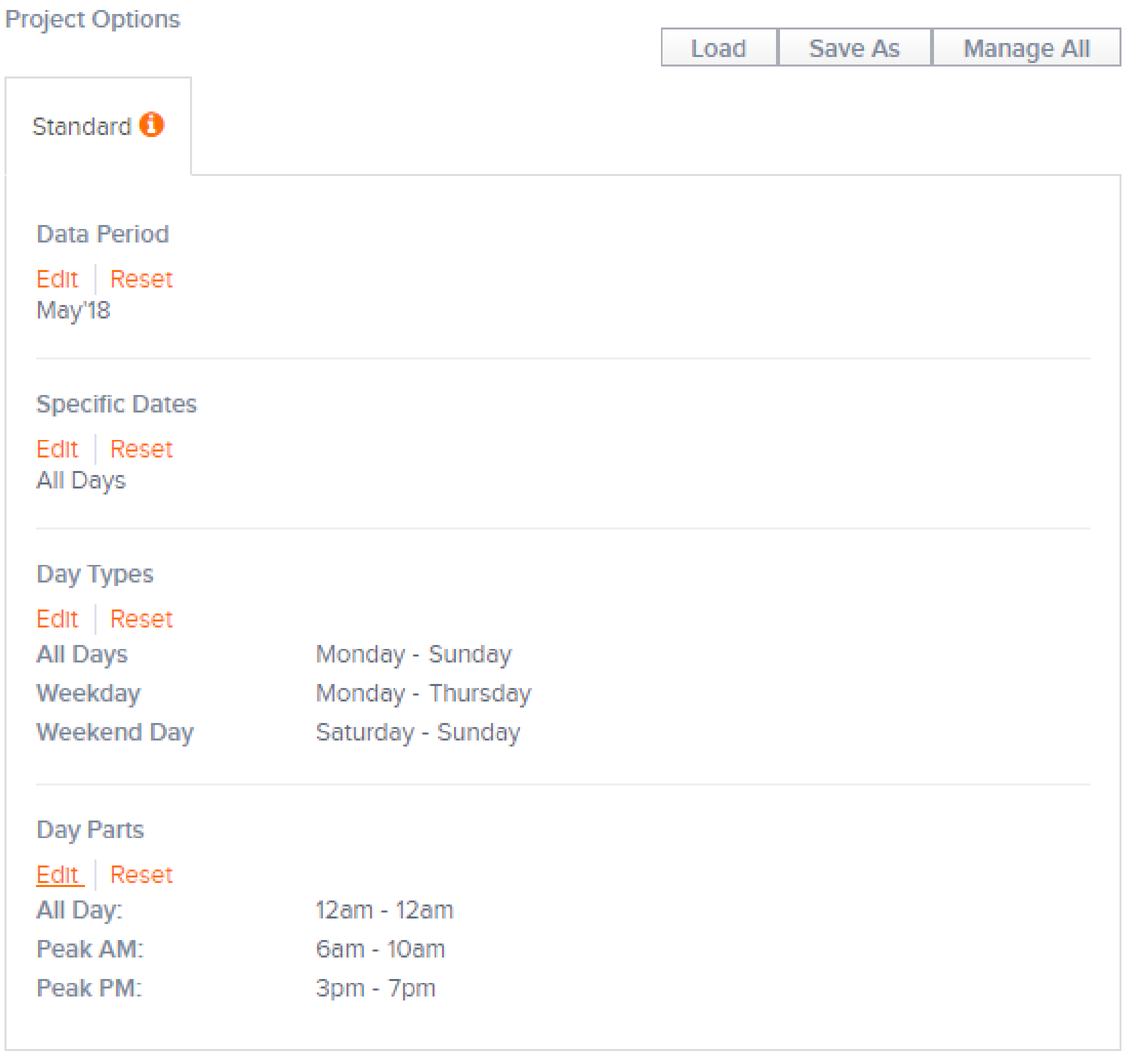
Accessed August 15, 2019. Reprinted With Permission.

Step 4. Click “Create Projects”, following the settings as below.
Project Type: Zone Activity Analysis
Type of Travel: Personal
Mode of Travel: All Modes
Data Source Type: Location-Based Services with Pass-through
Output Type: StreetLight Volume



Accessed August 15, 2019. Reprinted With Permission.

Step 5. Configure the Project Options. Make sure the created zone is set as the Selected Zone Set(s). Then click "Confirm" to build the project.



Accessed August 15, 2019. Reprinted With Permission.

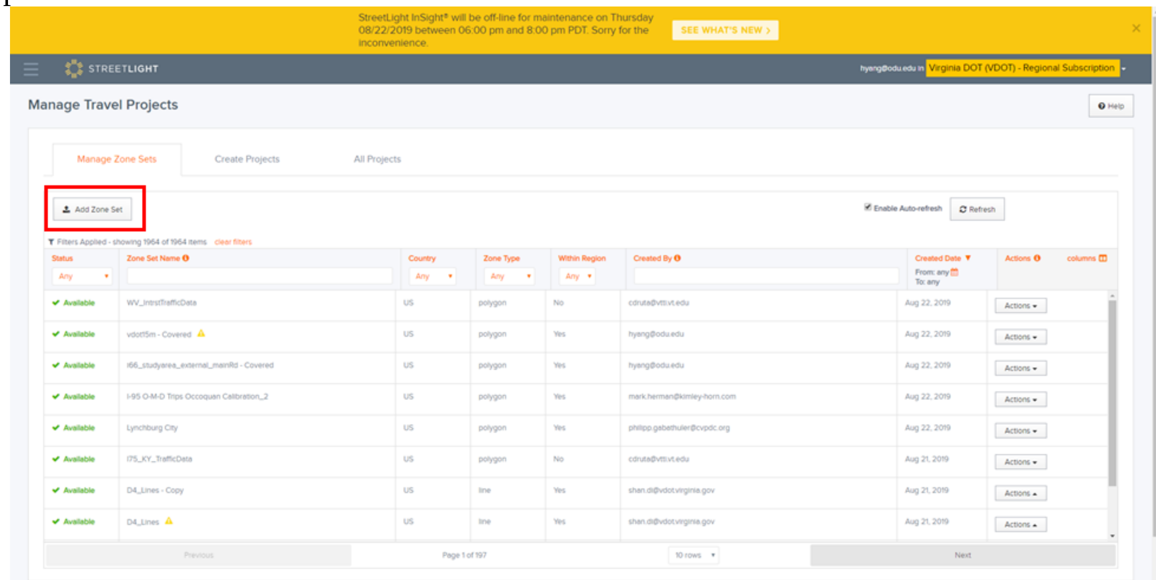
Step 6. Find your project in the "All Projects" panel and download the analysis results. All results will be stored in a .zip file.

Step 7. Analyze the detailed results. See examples shown in Tutorial 4.

Tutorial 3. Segment Analysis

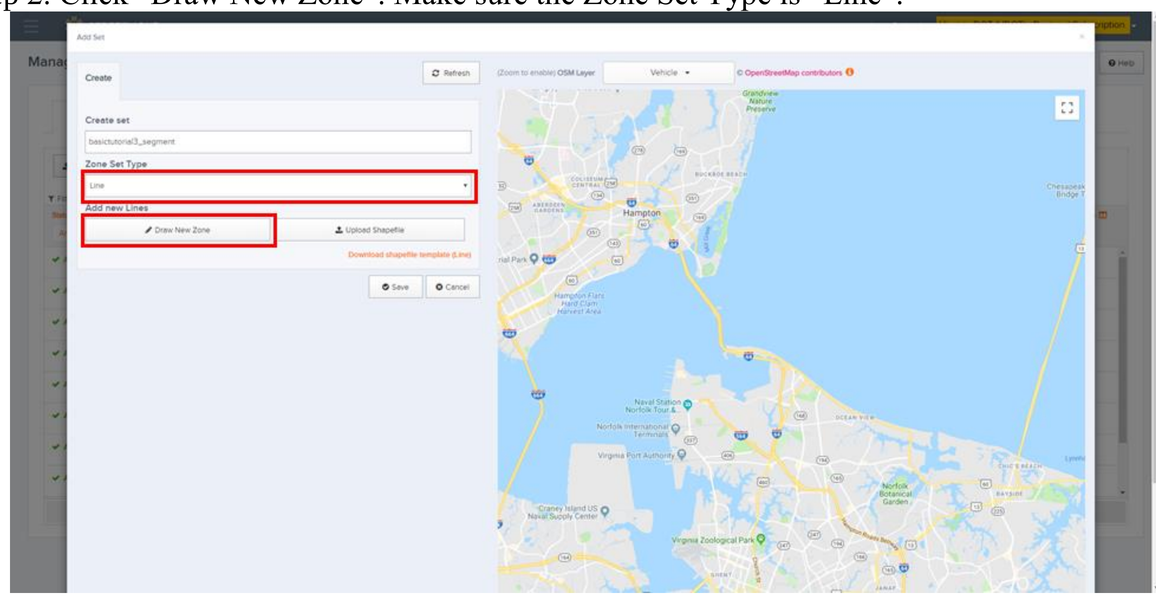
Goal: Explore the communication patterns between ***Norfolk*** *and* ***Hampton*** *on I-66.*

Step 1. Click "Add Zone Sets".



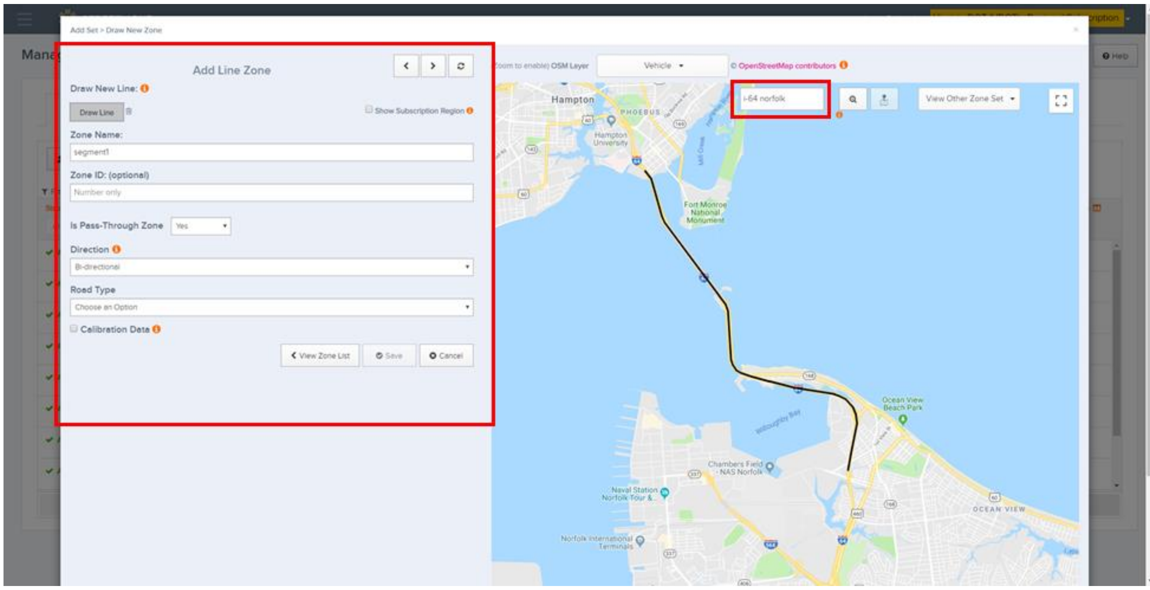
Accessed August 15, 2019. Reprinted With Permission.

Step 2. Click "Draw New Zone". Make sure the Zone Set Type is "Line".



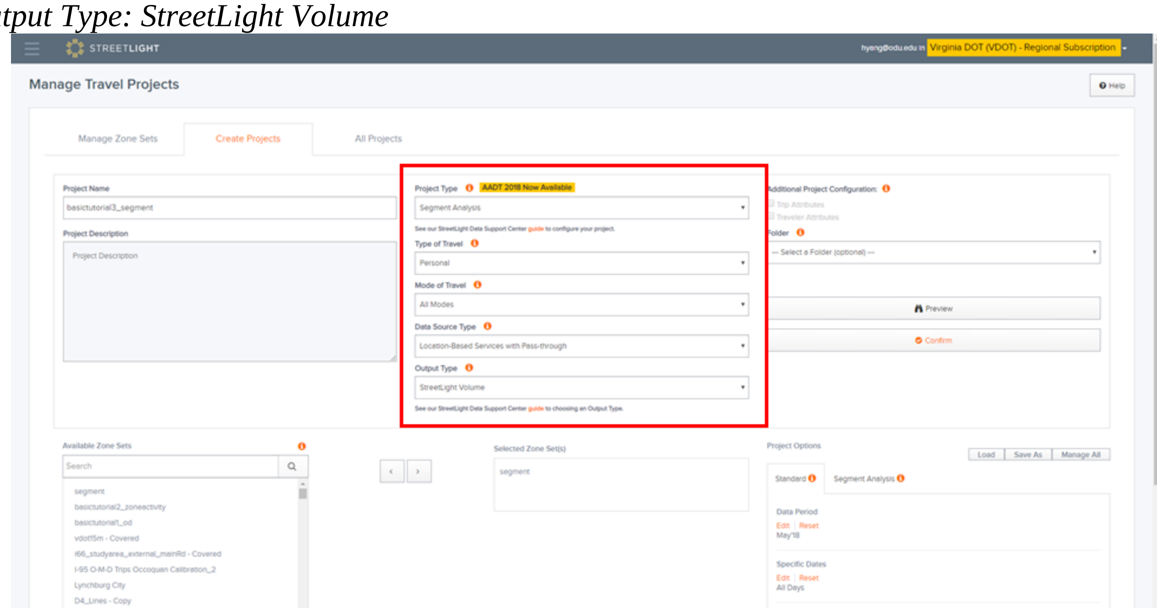
Accessed August 15, 2019. Reprinted With Permission.

Step 3. Firstly, type in “I-64, Norfolk” in the search box. Secondly, draw the line segment along I-66 and set it as a “Yes” pass-through zone. Then, set it as “Bi-directional”. Click “Save” to save the result.



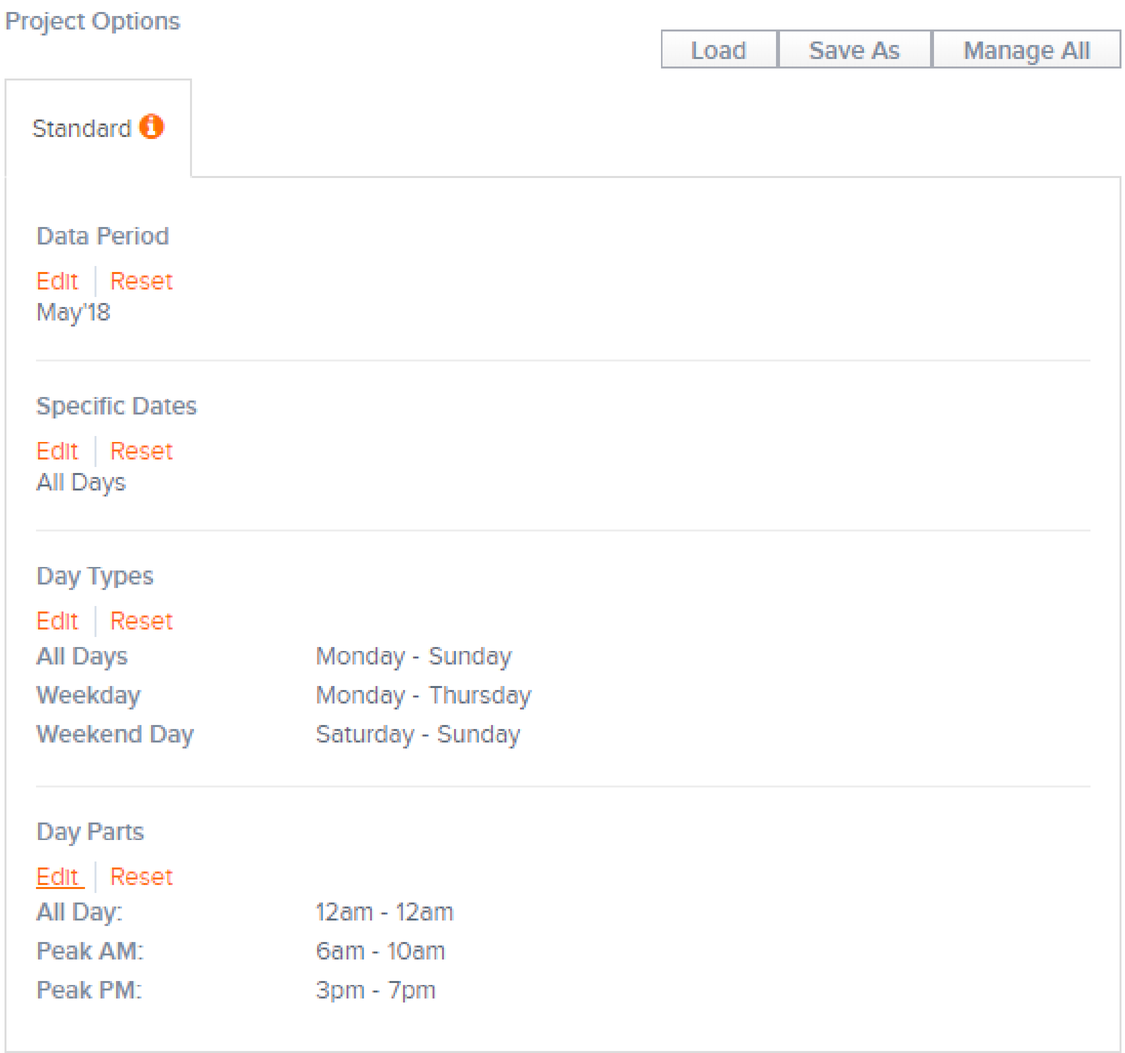
Accessed August 15, 2019. Reprinted With Permission.

Step 4. Click “Create Projects”, following the settings as below.
Project Type: Segment Analysis
Type of Travel: Personal
Mode of Travel: All Modes
Data Source Type: Location-Based Services with Pass-through
Output Type: StreetLight Volume



Accessed August 15, 2019. Reprinted With Permission.

Step 5. Configure the Project Options. Make sure the created zone is set as the Selected Zone Set(s). Then click “Confirm” to build the project.

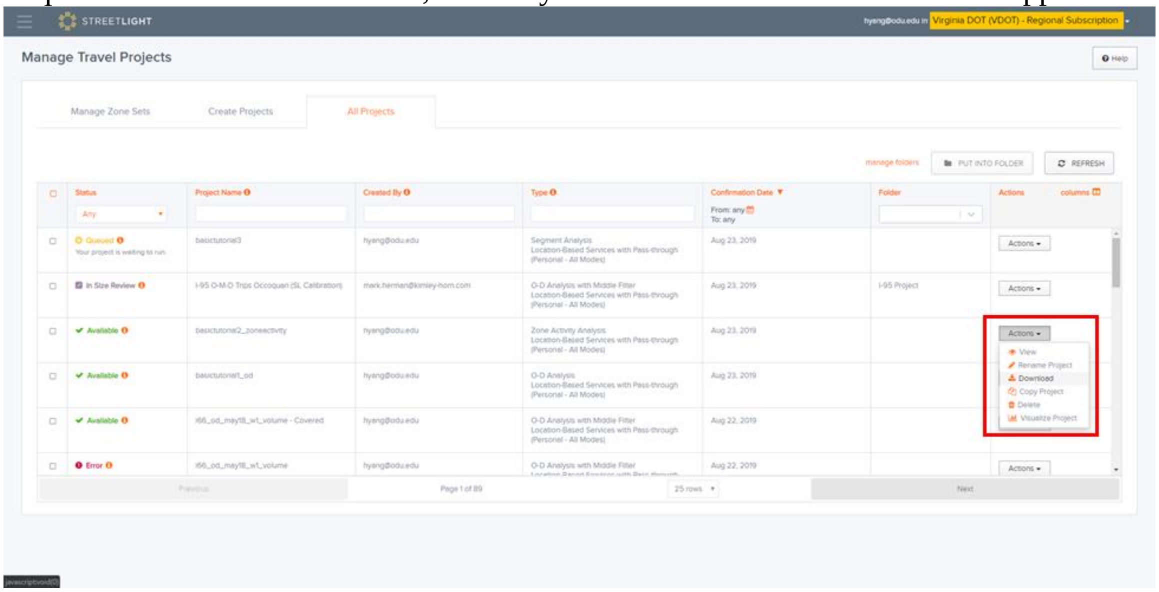


Accessed August 15, 2019. Reprinted With Permission.

Step 6. Find your project in the “All Projects” panel and download the analysis results. All results will be stored in a .zip file.
Step 7. Analyze the detailed results. See examples shown in Tutorial 4.

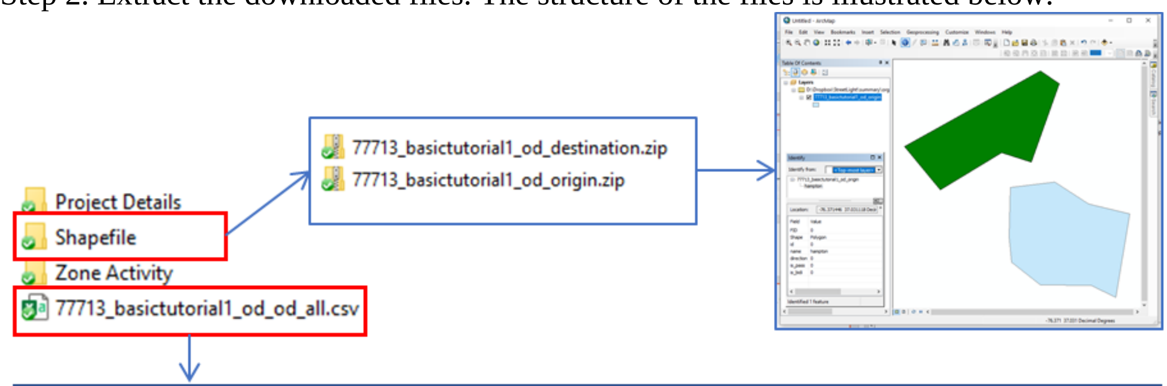
Tutorial 4. Examples of Downloaded Results and Interpretations

Step 1. Click “download” button, the analysis results can be downloaded as a zipped file.



Accessed August 15, 2019. Reprinted With Permission.

Step 2. Extract the downloaded files. The structure of the files is illustrated below.



Origin Zone ID	Origin Zone Name	Destination Zone ID	Destination Zone Name
1	hampton	2	norfolk
1	hampton	2	norfolk
1	hampton	2	norfolk
1	hampton	2	norfolk
...	...	...	...

Day Type	Day Part	Average Daily O-D Traffic (StL Volume)
0: All Days (M-Su)	0: All Day (12am-12am)	307772
0: All Days (M-Su)	1: Peak AM (6am-10am)	50666
0: All Days (M-Su)	2: Peak PM (3pm-7pm)	93660
1: Weekday (M-Th)	0: All Day (12am-12am)	305007
1: Weekday (M-Th)	1: Peak AM (6am-10am)	55567
1: Weekday (M-Th)	2: Peak PM (3pm-7pm)	96159
...	...	...

Accessed August 15, 2019. Reprinted With Permission.

Note:

The shapefiles contain the feature layers of the zone sets involved in this project. Users can use a spatial analysis tool such as ArcMap to visualize these shapefiles.

The detailed StreetLight metrics (data) are stored in the CSV file. Users can use Excel to open and perform the basic analysis of the data. These data (metrics) include information such as origin and destination zone names, StreetLight Index, time information, etc.

APPENDIX G

WEBLINKS FOR EXPLORING THE SL INDEXES AND BENCHMARK DATA

A site hosted by the Old Dominion University Transportation Research Lab enables one to perform additional analysis. The weblink to this site is: http://senselane.com/streetlight/