Scientific Research Publishing

Journal of Transportation Technologies, 2023, 13, 443-464
https://www.scirp.org/journal/jtts
ISSN Online: 2160-0481
ISSN Print: 2160-0473

A Data-Driven Intersection Geometry Mapping Technique to Enhance the Scalability of Trajectory-Based Traffic Signal Performance Measures

Enrique D. Saldivar-Carranza, Darcy M. Bullock

Lyles School of Civil Engineering, Purdue University, West Lafayette, USA
Email: esaldiva@purdue.edu, darcy@purdue.edu

How to cite this paper: Saldivar-Carranza, E.D. and Bullock, D.M. (2023) A Data-Driven Intersection Geometry Mapping Technique to Enhance the Scalability of Trajectory-Based Traffic Signal Performance Measures. *Journal of Transportation Technologies*, **13**, 443-464.
https://doi.org/10.4236/jtts.2023.133021

Received: June 30, 2023
Accepted: July 18, 2023
Published: July 21, 2023

Abstract

Connected vehicle (CV) trajectory data provides practitioners with opportunities to assess traffic signal performance with no investment in detection or communication infrastructure. With over 500 billion trajectory records generated each month in the United States, operations can be evaluated virtually at any of the over 400,000 traffic signals in the nation. The manual intersection mapping required to generate accurate movement-level trajectory-based performance estimations is the most time-consuming aspect of using CV data to evaluate traffic signal operations. Various studies have utilized vehicle location data to update and create maps; however, most proposed mapping techniques focus on the identification of roadway characteristics that facilitate vehicle navigation and not on the scaling of traffic signal performance measures. This paper presents a technique that uses commercial CV trajectory and open-source OpenStreetMap (OSM) data to automatically map intersection centers and approach areas of interest to estimate signal performance. OSM traffic signal tags are processed to obtain intersection centers. CV data is then used to extract intersection geometry characteristics surrounding the intersection. To demonstrate the proposed technique, intersection geometry is mapped at 500 locations from which trajectory-based traffic signal performance measures are estimated. The results are compared to those obtained from manual geometry definitions. Statistical tests found that at a 99% confidence level, upstream-focused performance estimations are strongly correlated between both methodologies. The presented technique will aid agencies in scaling traffic signal assessment as it significantly reduces the amount of manual labor required.

Keywords

Connected Vehicle, Trajectory, Traffic Signal, Performance, Map, Geometry

1. Introduction

A turning movement describes the course undertaken by a vehicle at an intersection [1]. It usually entails the combination of approaching directions (e.g., northbound, eastbound, southbound, and westbound) and turn types (e.g., through, left, and right). For example, a vehicle approaching an intersection northbound (NB) and continuing through is described as a NB-through movement and a vehicle approaching westbound (WB) and turning left is described as a WB-left movement.

Practitioners need the ability to evaluate the performance of their managed intersections at the movement-level to identify and prioritize signal timing opportunities [1]. Once a poor-performing movement is identified, the signal phase parameters and its associated detection hardware can be evaluated to determine a possible remedy (e.g., maintenance, split rebalance, offset modification, or changes in cycle length).

The estimation of movement-level traffic signal performance measures derived from connected vehicle (CV) trajectory data has previously been discussed in [2] [3] and generally requires:

1) The linear referencing of vehicle waypoints;

2) The movement identification of passing vehicles;

3) The location of intersection centers, and;

4) The identification of the area of each intersection approach upstream of the center where vehicle performance is dependent on the operation of the analyzed intersection.

Techniques to accomplish the first and second points are presented in [3] and [4], respectively. The intersection center serves as a reference to linearize trajectories [3] and is also necessary for the vehicle movement identification technique [4] as it is used to assess entry and exit vehicle heading boundaries. If an inaccurate center is provided, the location of relevant performance events would be incorrect and traversing vehicles may be assigned erroneous movements which would lead to misleading performance estimations.

Furthermore, an area for each approach upstream of the center that covers the space that is directly controlled by the evaluated traffic signal needs to be defined. These areas are used to retrieve vehicle trajectory segments that are impacted by the operations of the intersection being studied and need to be carefully determined to derive accurate performance measures, for example:

- Approach retrieval areas that are too small can miss relevant events that happen far back from the intersection, such as long queues or the occurrence of split failures [2].

- Approach retrieval areas that are too large can include vehicle conditions that result from other control devices, such as four-way stops or adjacent signalized intersections, where vehicle stops and induced delay would be incorrectly assigned to the evaluated signal.

1.1. Literature Review

During the last decade, there has been an increase in the use of crowdsourced data to monitor the state of roadway infrastructure [5]. One of these crowdsourced data sets is CV trajectory data, comprised of journey-based vehicle trajectories, which already provides over 500 billion records monthly in the United States and is expected to significantly grow in the coming years [3] [6].

CV trajectory data has effectively been used to complement existing infrastructure-based traffic signal performance measures [3] [7] without the need for vehicle detection or communication equipment. Zhao *et al.* [8] used trajectory data to estimate queue lengths, penetration rates, and traffic volumes at signalized intersections. Waddell *et al.* [9] and Huang *et al.* [10] calculated delays, and various studies estimated traffic signal operational conditions, such as the level of arrivals on green (AOG), split failures (SF), and downstream blockage (DSB) from CV trajectory data [2] [9] [11].

An open-access report titled *Next Generation Traffic Signal Performance Measures: Leveraging Connected Vehicle Data* [3], which provides a comprehensive suite of trajectory-based techniques to evaluate traffic signal performance and identify maintenance opportunities, has recently been published. Accuracy and scalability benefits of using this crowdsourced data set are discussed and demonstrated by analyzing almost 5000 traffic signals nationwide (around 1% of the total number of signals in the United States). Nevertheless, the effective evaluation of traffic signal performance with CV trajectory data currently requires the manual identification of relevant intersection geometry characteristics, such as intersection centers and approach areas of interest. This can hinder the large-scale implementation of analysis techniques as significant human input is needed.

GPS data has been extensively used to automatically update and create maps. Cao and Krumm utilized raw GPS traces and a custom clustering algorithm to generate routable road nodes and links [12]. Carisi *et al.* used GPS vehicle data to identify the location of stop signs and traffic signals with over 90% accuracy [13]. Jang *et al.* suggested a highly accurate real-time map creation system using a few GPS traces [14]. Erramaline *et al.* designed a convolutional neural network (CNN) that uses speed and acceleration profiles to classify intersections by their control type [15].

1.2. Motivation and Objective

Most proposed GPS-based automatic mapping techniques focus on the identification of roadway characteristics that facilitate vehicle navigation; however, no

studies were found that automatically map intersection areas to scale the estimation of traffic signal performance measures. Currently, the intersection centers and the approach areas of interest required to calculate trajectory-based signal performance are manually defined and usually take 5 minutes per intersection to complete. To carry out the analysis of the over 4700 intersections (Figure 1) discussed in [3], 392 hours, or 49 workdays, were spent on the manual definition of geometry characteristics.

Since most state transportation agencies manage thousands of traffic signals, the intersection geometry characteristics needed to estimate CV-based performance measures can represent an obstacle to reaching agency-wide coverage. The objective of this paper is to provide a heuristic that uses CV trajectory data and open-source map elements to improve the scalability of CV-based signal performance measures by automatically estimating intersections' centers and the approach areas of interest.

It is particularly important to provide accurate trajectory-based techniques to scale the evaluation of traffic signals when large amounts of infrastructure- descriptive data points can be derived from vehicle trajectories. For example, a popular reporting framework that uses trajectory data to evaluate the performance of signalized corridors provides up to 3072 data points over a 24-hour period for each assessed intersection [16].

2. Data Description

CV trajectory data from March 2023 weekdays and open-source map elements are used in the proposed heuristic to automatically map relevant intersection geometry characteristics.

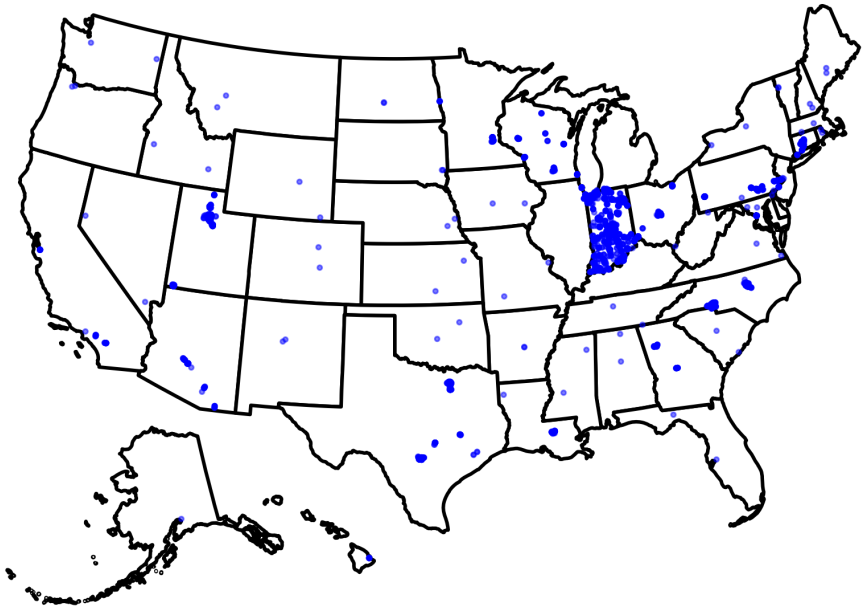

Figure 1. Manually defined intersections to evaluate CV-based traffic signal performance measures in [3] (n: 4703).

2.1. Connected Vehicle Trajectories

CV trajectory data is obtained from a third-party data vendor that works directly with original equipment manufacturers (OEMs). This data set is generated from passenger vehicles that are factory-equipped with the required technology for sampling and transmission. The CV trajectory data consists of sets of waypoints for entire vehicle trips with a reporting interval of 3 seconds and a spatial accuracy of 3 meters (~10 ft.). Every waypoint contains the following information: GPS location, timestamp, speed, heading, and anonymous unique trajectory identifier. It is important to mention that only vehicle information is available from the CV data set and no infrastructure attributes, such as Signal Phase and Timing (SPaT) or map data (MAP) messages [17] [18], are provided.

2.2. OpenStreetMap

Open-source map elements used in the proposed heuristic are obtained from OpenStreetMap (OSM) [19]. OSM uses a peer production model to generate map data that is free to use and edit. An advantage of using OSM data is that it not only contains information on road nodes and links, but also on traffic control devices, railways, buildings, green areas, landmarks, and other map elements that can be useful in the accurate identification of relevant areas near signalized intersections. Map elements can be downloaded as vectors (*i.e.*, point, line, and polygon) or raster data for further processing [20]. Additional information on OSM map elements utilized in the heuristic is discussed in the following sections. Although there is valuable information in OSM data, the quality can vary, and it is important to validate this data set independently.

3. Approach

The intersection mapping methodology is comprised of the steps shown in **Figure 2**. The first step is to derive intersection centers from OSM traffic signal tags. Intersection names are then assigned from road link information in the second step. In the third step, the surroundings of the derived centers from step one are

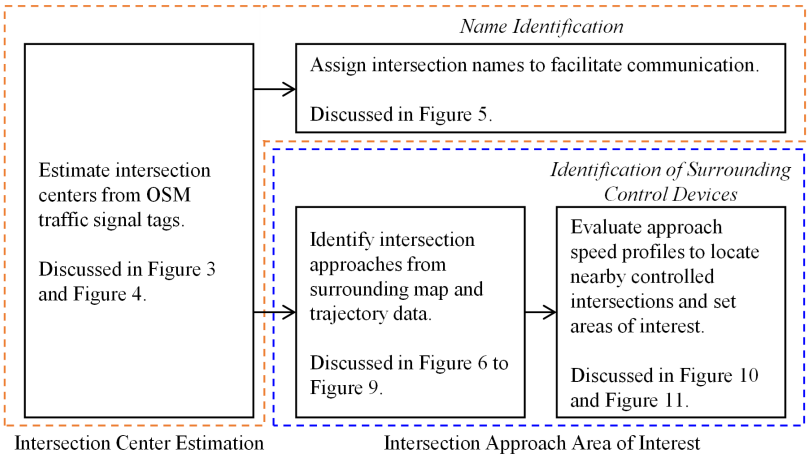

Figure 2. High-level steps to map relevant intersection geometry characteristics.

evaluated to determine intersection approaches. Finally, in the fourth step, speed profiles on the assigned approaches are assessed to identify the location of other nearby controlled intersections.

The outcome of the technique is intersection centers and distances of interest for each relevant approach. The methodology steps presented in Figure 2 are explained in detail in the sections below.

4. Intersection Center Estimation

Accurate intersection centers are critical for the correct calculation of signal performance from trajectory data. These geospatial points are used to both linear reference trajectories and to assign movements to traversing vehicles. Furthermore, the approach area of interest identification discussed in the following section is also significantly influenced by the center of the intersection.

One approach to acquiring intersection centers is to request the Departments of Transportation (DOTs) traffic signal location inventories. Even though this approach requires no manual definitions and no data processing, some important challenges exist. Agencies may or may not be able to share this information, communication delays can slow down the analysis process, and the provided signal locations may not be optimal for performance calculations as DOT traffic signal location inventories are usually not created for this purpose. For example, Figure 3(a) shows an aerial view of a signalized intersection displaying the manually defined intersection center (blue) and one obtained from a DOT signal location inventory (pink). The error, measured as the geodesic distance between the two points, is 140 ft. Traffic signal performance estimations would be inaccurate if the inventory traffic signal location was used to linear reference trajectories and assign intersection movements.

An alternative is to derive signalized intersection centers from OSM data. OSM counts with the location of traffic signals and makes this information readily available to anyone. A challenge of using OSM traffic signals is that there are different ways in which the control device is tagged on the map. At complex intersections, traffic signals appear at all crossings or at all approaches, which results in the assignment of more than one traffic signal tag at a single intersection. For example, Figure 3(b) shows the location of two OSM traffic signal tags (callout i) assigned at the same intersection as Figure 3(a).

Nonetheless, intersection centers can still be obtained from several traffic signal tags. First, groups of traffic signals that belong to the same intersection are identified by clustering together any tag that is closer than 60 meters (196.85 ft.) to another. Then, the location of the intersection center is calculated as the centroid of the cluster that contains n tags as:

$$Lat_c = \frac{1}{n}\sum_{i=1}^{n} Lat_i \tag{1}$$

$$Lon_c = \frac{1}{n}\sum_{i=1}^{n} Lon_i \tag{2}$$

where Lat_c is the latitude and Lon_c is the longitude of the cluster centroid, and

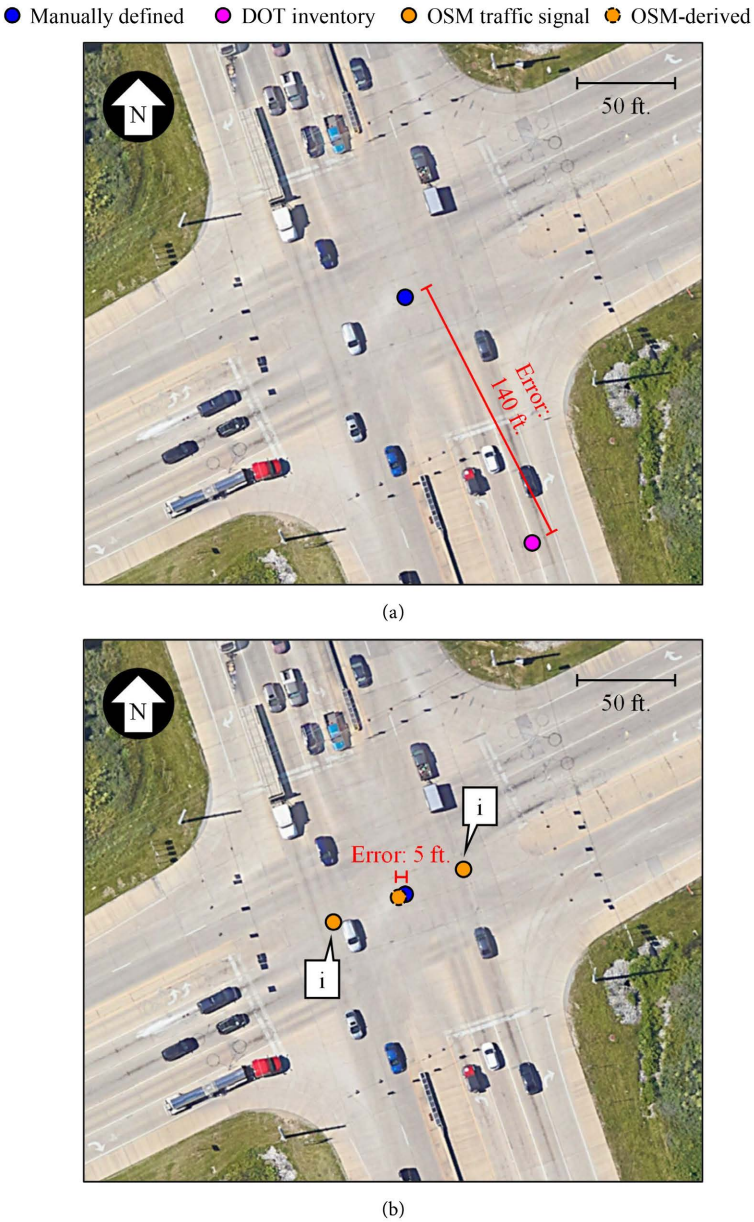

Figure 3. Intersection center location error (map data: Google and OpenStreetMap). (a) DOT inventory; (b) OSM-derived.

Lat_i is the latitude and Lon_i is the longitude of the *i-th* traffic signal tag in the cluster. The OSM-derived intersection center (centroid) at the intersection shown in **Figure 3(b)** is displayed as an orange circle with a dashed boundary, which has an error of only 5 ft. in relation to the manually defined center.

Figure 4 presents a cumulative frequency distribution (CFD) plot of the intersection center errors obtained from a state DOT inventory and from OSM at 500 intersections in relation to the manual approach. The number of studied intersections was chosen to provide statistical significance and the locations were randomly selected from a manually defined inventory to include a geospatially distributed set [3]. **Table 1** provides information on the distribution of this error.

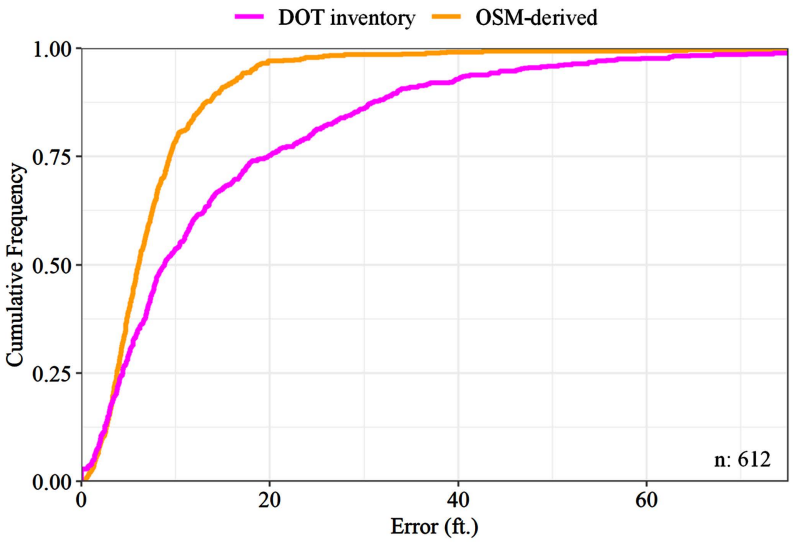

Figure 4. Intersection center location error CFD for 500 signals.

Table 1. Intersection center location error distribution.

Percentile (%)	DOT Inventory (ft.)	OSM-derived (ft.)
25	4.4	3.8
50	8.8	6.0
75	19.7	9.3
90	33.5	14.6
99	75.8	37.9
100	141.5	114.1

Clearly, the OSM-derived intersection centers have smaller errors than those obtained from the DOT inventory. With 90% of the estimated centers being closer than 15 ft. to the manual definitions, the OSM-derived centers are an acceptable approach to scaling the estimation of trajectory-based traffic signal performance measures.

Name Identification

Practitioners usually refer to their managed traffic signals by the road names that intersect at the location. Therefore, it would be useful to include this information when communicating performance estimations. Once an OSM-derived intersection center is estimated, the point can be assigned a name by retrieving nearby OSM links. All links with the same name are treated as the same road. The shortest distance from the intersection center to all nearby roads is then calculated. Finally, the intersection center is assigned the combination of the two closest road names.

This methodology is shown in Figure 5. At this intersection, the West Thompson Road link (green) is closest to the OSM-derived intersection center

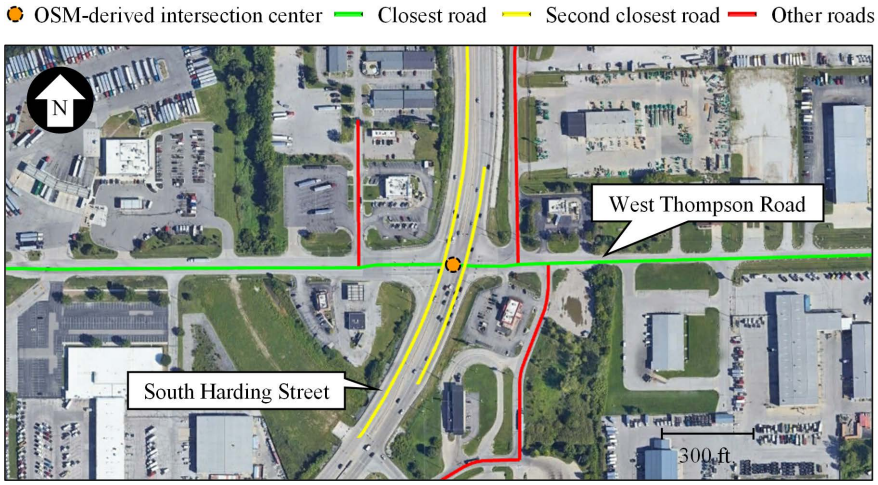

Figure 5. Intersection name identification at West Thompson Road and South Harding Street (map data: Google and OpenStreetMap).

followed by the South Harding Street link (yellow). Therefore, the intersection center is named "West Thompson Road and South Harding Street". This technique was used to assign names to the 500 derived centers and 98% of the labels were correct.

5. Intersection Approach Area of Interest

The other intersection geometry characteristic that is relevant to enhance the scalability of CV-based traffic signal performance measures is how far upstream from the estimated centers vehicle trajectories should be evaluated. This distance needs to be accurately defined for each approach to generate performance estimations that include all relevant events outcome of the traffic signal being studied and that exclude those events that are products of adjacent traffic control devices. Therefore, the objective is to define distances upstream of the center where vehicle behavior is mainly affected solely by the traffic signal at the center.

Figure 6 shows a signalized intersection with its OSM-derived center and other surrounding control devices. The manually defined distances of interest for each approach are also displayed. Since the WB, NB, and eastbound (EB) approaches do not have other traffic control devices that would affect the progression of vehicles, their area of interest is defined as 1320 ft. long, which is a previously tested distance that includes all relevant vehicle events that are the result of traffic signal operations. The southbound (SB) approach is defined as 1000 ft. long to not include trajectory segments affected by the adjacent traffic signal (callout i).

To automatically approximate the manually defined approach distance of interest shown in Figure 6, information on the intersection center surroundings needs to be extracted from the CV trajectory waypoints in the area (Figure 7). Intersection approaches may be connected to other actively controlled intersections, local roads, or driveways within the 1320 ft. radius of interest as displayed in Figure 8(a).

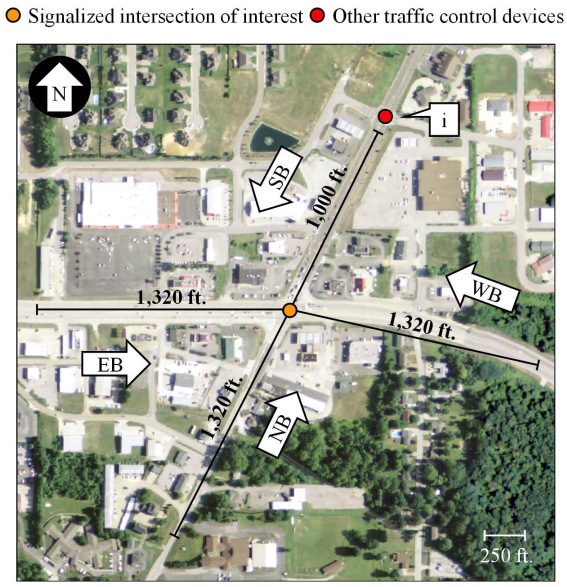

Figure 6. Approach distances of interest (map data: Google).

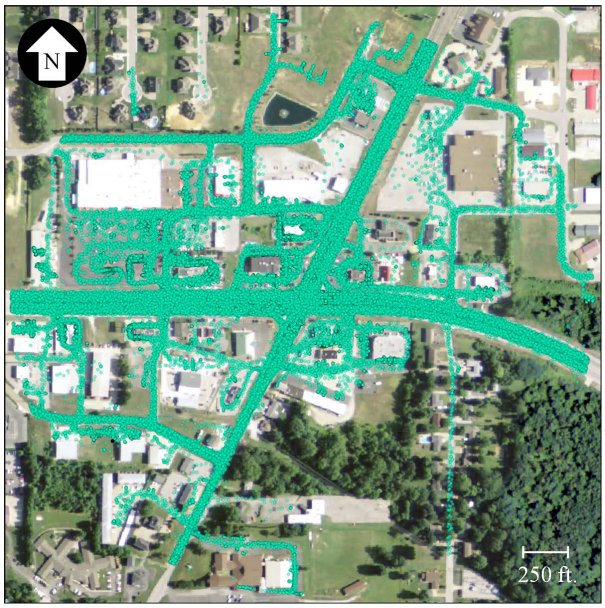

Figure 7. Vehicle waypoints around the signalized intersection (n: 268,549) (map data: Google).

To differentiate the intersection approaches from other roadway elements and identify surrounding traffic control devices, CV trajectory data is processed and rasterized with a 5-meter (16.40 ft.) resolution to obtain the following information:

- Actively approaching cells (Figure 8(b)): an actively approaching cell is any raster cell where most of its contained waypoints' headings are within ±25˚ from a vector drawn from the waypoint itself to the OSM-derived intersection center. This means that most of the waypoints within that cell head towards the center of the intersection.

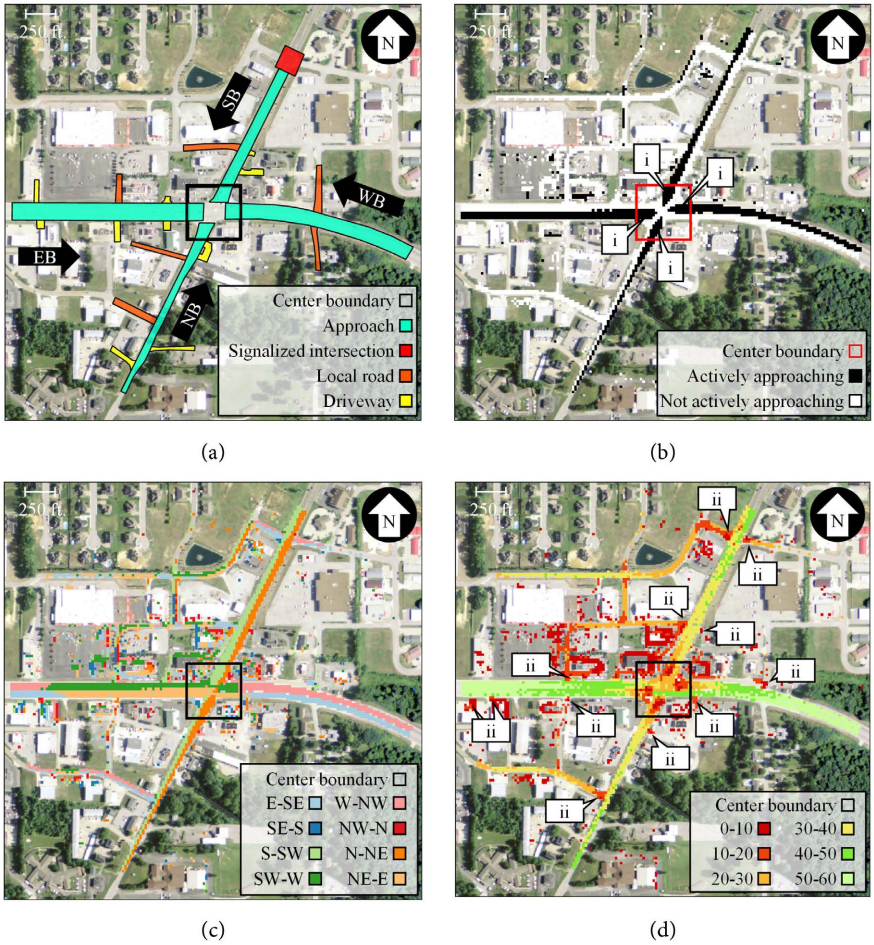

Figure 8. Relevant data for approach area of interest identification. (a) Surroundings; (b) Actively approaching cells; (c) Cell direction; (d) 90th percentile cell speed (mph).

- Cell direction of travel (Figure 8(c)): each raster cell is assigned two possible directions of travel based on its contained waypoints' headings. A cell can travel N (337.5˚ to 360˚ or 0˚ to 22.5˚), NE (22.5˚ to 67.5˚), E (67.5˚ to 112.5˚), SE (112.5˚ to 157.5˚), S (157.5˚ to 202.5˚), SW (202.5˚ to 247.5˚), W (247.5˚ to 292.5˚), or NW (292.5˚ to 337.5˚). If most of the waypoints within a cell have headings between 67.5˚ and 112.5˚, that cell is said to be able to travel to the cell in the E direction. Additionally, the NE and SE directions of travel are compared and the one with most samples is also assigned to the evaluated cell.
- Ninetieth percentile cell speed (Figure 8(d)): the 90th percentile speed of the waypoints contained within the cell is assigned as the cell speed.

These waypoint-derived rasters can be evaluated to define the intersection approaches. First, the cells actively approaching within the 100 by 100-meter (328.08 by 328.08 ft.) center boundary are identified as part of the intersection approach (Figure 8(b), callout i). These cells likely belong to the correct approaches due to their proximity to the intersection center. Then, the rest of the cells' direction of travel (Figure 8(c)) is reviewed starting from the center

boundary neighbors and continuing outbound. If a cell has an option to travel to a cell that is already part of the intersection approach, and is an actively approaching cell or has a 90th percentile speed greater than 15 mph, it is also included as part of that approach. The 90th percentile speed greater than 15 mph is used to not include local roads or driveways that join the main approaches (Figure 8(d), callout ii).

Figure 9 shows the approach identification outcome as vector polygons of the methodology discussed above. All the intersection approaches are correctly delineated as no driveways or local roads that join upstream are included in the result. Nevertheless, the SB approach still extends past the adjacent traffic signal (callout i). Therefore, additional processing is needed to identify the location of adjacent traffic control devices.

Identification of Surrounding Control Devices

Traffic control devices that affect the progression of vehicles which should not be included in the trajectory retrieval area when evaluating traffic signal performance, such as other traffic signals or stop signs, can significantly reduce vehicle speeds and induce stops. By evaluating the distribution of vehicle speeds on each intersection approach, adjacent control devices can be identified.

Figure 10 shows 5th percentile speed profiles at 50 ft. bins obtained by evaluating vehicle waypoints found within the 10-meter (32.81 ft.) buffered estimated approaches (Figure 9). Fifth percentile speeds are used to locate where on the approaches some vehicle must travel slowly. Callout i points to the first 50 ft. from the intersection center. Since these bins are located past the stop bar where

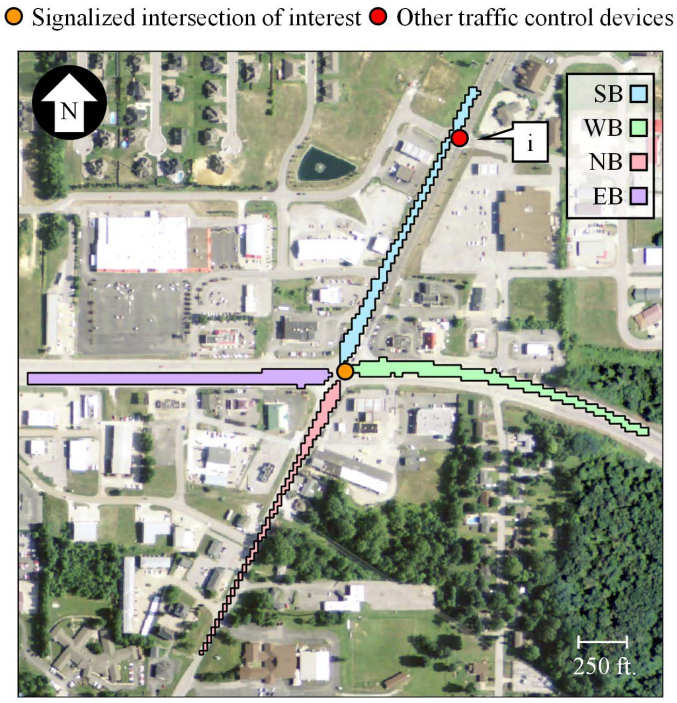

Figure 9. Approach identification outcome.

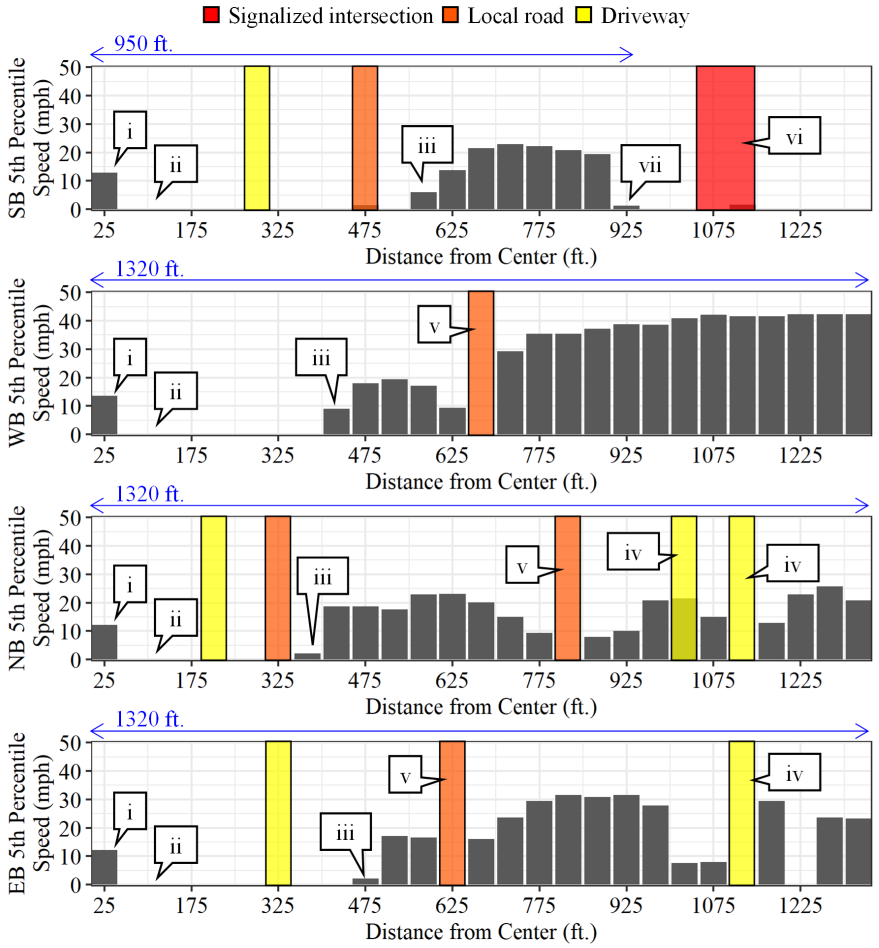

Figure 10. Speed profiles evaluation to identify approach areas of interest.

vehicles start accelerating, the 5th percentile speeds are above zero. Callout ii indicates where standing queues start to form. These queues stretch all the way until callout iii, where vehicles upstream of the intersection do not have to stop. Once the speed profiles become non-zero, only the connection of driveways, local roads, or other intersections with traffic control devices to the approaches significantly reduce vehicle speeds. Driveways (callout iv) and local roads (callout v) only reduce the speed profiles for short distances. In contrast, traffic control devices affect vehicle speeds for hundreds of feet. For example, the adjacent traffic signal in the SB approach (callout vi) significantly reduces vehicle speeds starting 900 ft. upstream of the intersection (callout vii).

From the speed profiles shown in Figure 10, relevant approach distances can be estimated. At the evaluated intersection, the WB, NB, and EB approaches are assigned the maximum 1320 ft. retrieval distance since no prolonged speed reductions are seen after callout iii. The SB approach is assigned a 950 ft. retrieval distance since starting 900 ft. upstream of the intersection speeds are reduced over a long stretch. These results are similar to the manual definitions shown in Figure 6.

In addition to the relevant approach identification methodology presented thus far, the location of map elements that would disrupt the progression of vehicles, such as traffic signals, stop signs, roundabouts, and railways, can be obtained from OSM and help select accurate trajectory retrieval distances. For this reason, if any of these roadway elements are found near any of the intersection approaches, retrieval distances are stopped before the element.

The distance of interest for all approaches at the 500 intersections is estimated with the proposed technique and the absolute difference with the manual results is presented in the CFD in Figure 11. Table 2 provides information on the distribution of this difference. It is shown how 75% of the estimated approaches have a difference smaller than 250 ft. compared to the manual definitions. Furthermore, even if large discrepancies exist for some approaches, it is possible that human error is the cause and the data-driven approach would still generate accurate traffic signal performance measures.

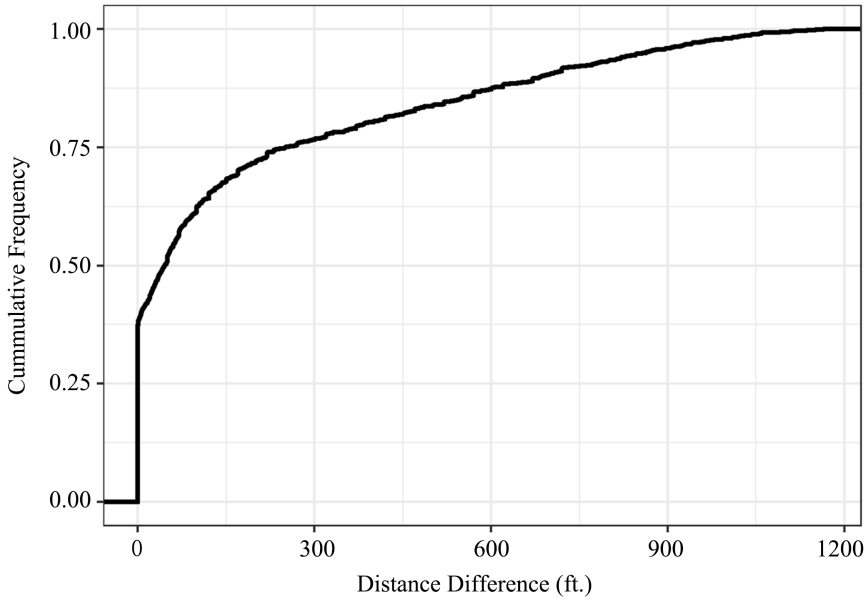

Figure 11. Manual and automatic approach of interest absolute distance difference CFD.

Table 2. Manual and automatic approach of interest absolute distance difference distribution.

Percentile (%)	Distance Difference (ft.)
25	0
50	45
75	250
90	685
99	1054
100	1170

6. Results

To evaluate the accuracy of the developed heuristic, movement-level control delay, AOG, and SF [3] are calculated at the 500 intersections based on the mapping results for all weekdays in March 2023. The estimated upstream-focused performance measures are compared to those derived from manual intersection geometry definitions for different time-of-day (TOD) periods in Figures 12-14. Each dot represents the performance estimation for a particular intersection movement. A blue line with no offset and a slope of one is plotted for reference. The closer points are located to the blue line, the smaller the discrepancy between estimations. Additionally, a linear least-squares regression line is plotted in red to show the overall trend of the relation. Similar results with both techniques are shown as regression lines have offsets close to zero and slopes close to one.

Correlation between Automatic and Manually Derived Traffic Signal Performance Measures

In addition to the graphical visualizations in Figures 12-14 highlighting similar patterns between upstream-focused performance estimations, several tests are performed to determine if linear correlation exists. The performance estimations derived from the automatic geometry mapping technique are compared with those obtained from the manual methodology.

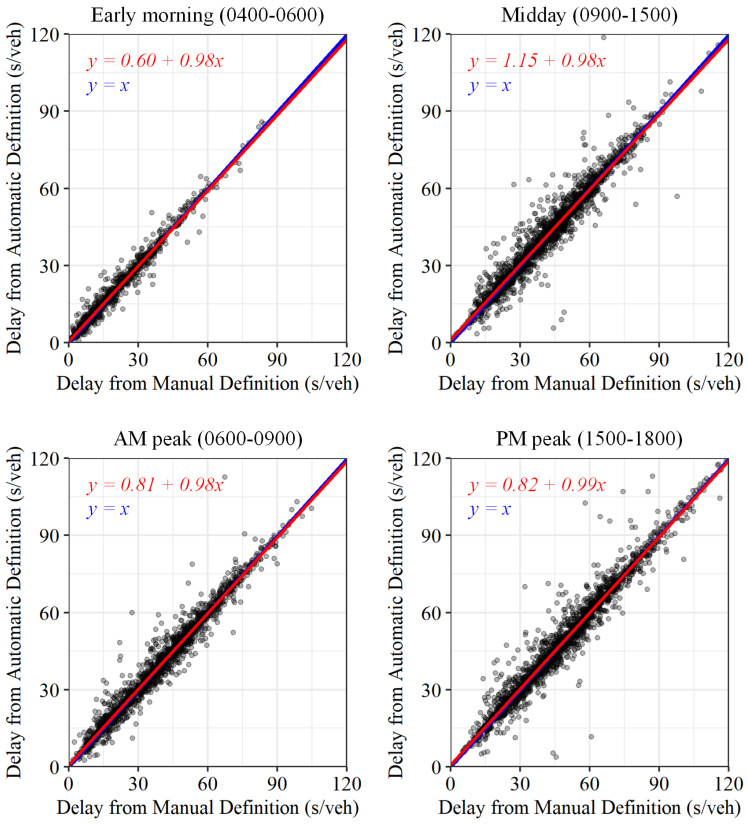

Figure 12. Average control delay estimation comparison.

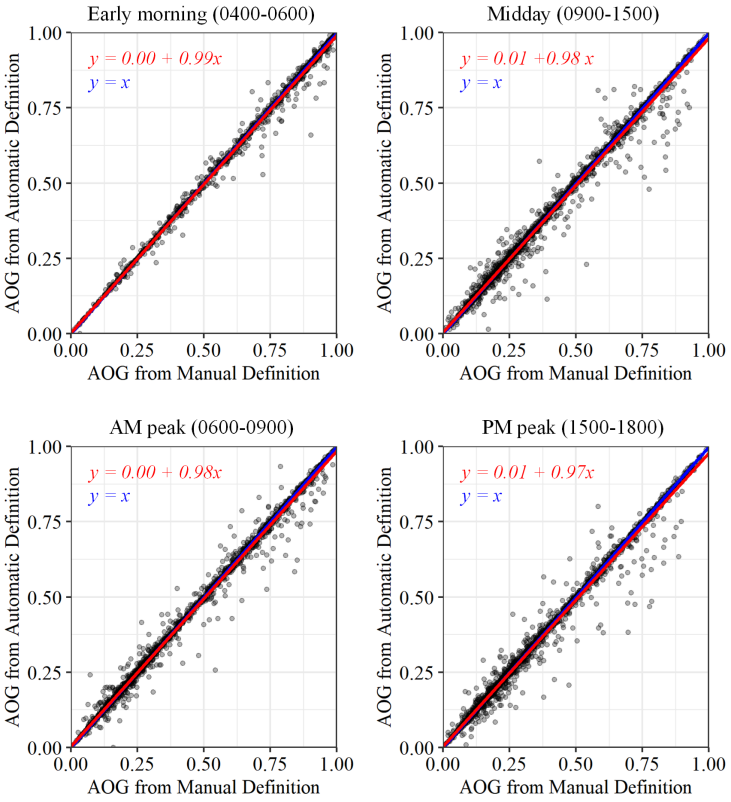

Figure 13. AOG estimation comparison.

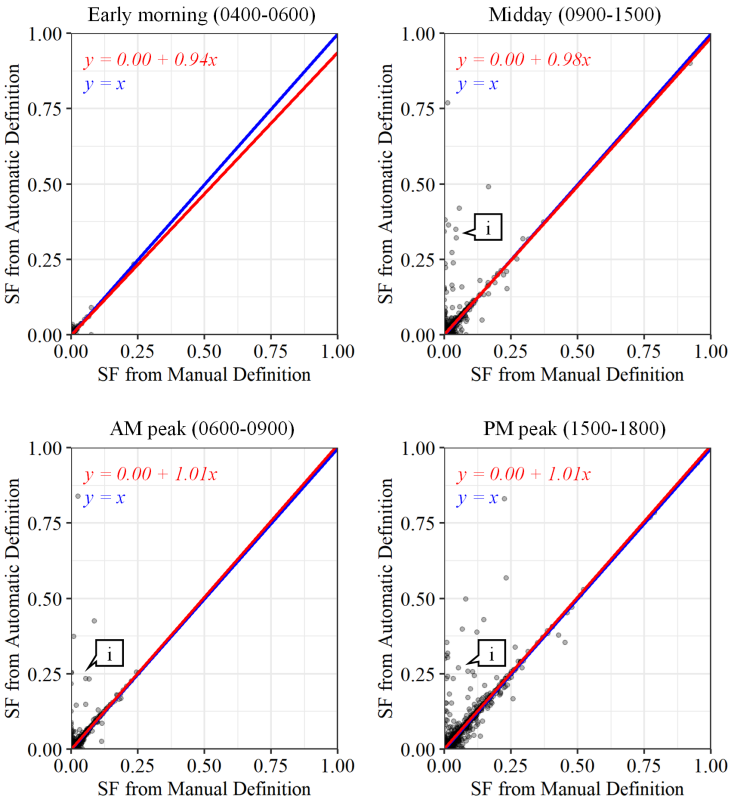

Figure 14. SF estimation comparison.

Pearson's (r) [21], Kendall's (τ) [22], and Spearman's (ρ) [23] correlation tests are performed to evaluate the relationship between the data. Table 3 presents the coefficient interpretations used for the Pearson's and Kendall's tests [24], and Table 4 shows the coefficient interpretations used for the Spearman's test [25]. Tables 5-7 show the results of the tests for different TOD periods and highlight how all performance estimations have at least high positive or strong correlation (SC).

Table 3. Interpretation of correlation coefficient—Pearson and Kendall.

Correlation Coefficient	Correlation Significance
0.90 - 1.0	Very high positive correlation (SC)
0.70 - 0.90	High positive correlation (SC)
0.50 - 0.70	Moderate positive correlation
0.30 - 0.50	Low positive correlation
0.00 - 0.30	Negligible correlation

Table 4. Interpretation of correlation coefficient—Spearman.

Correlation Coefficient	Correlation Significance
0.80 - 1.0	Very strong (SC)
0.60 - 0.79	Strong (SC)
0.40 - 0.59	Moderate
0.20 - 0.39	Weak
0.00 - 0.19	Very weak

Table 5. Correlation between average control delay estimations.

TOP Period	Pearson			Kendall			Spearman		
	r	p-value	SC	τ	p-value	SC	ρ	p-value	SC
Early morning	0.99*	<0.001	✓	0.89*	<0.001	✓	0.98*	<0.001	✓
AM peak	0.98*	<0.001	✓	0.90*	<0.001	✓	0.98*	<0.001	✓
Midday	0.96*	<0.001	✓	0.87*	<0.001	✓	0.97*	<0.001	✓
PM peak	0.98*	<0.001	✓	0.88*	<0.001	✓	0.97*	<0.001	✓

Note: *significant at 99% confidence level.

Table 6. Correlation between AOG estimations.

TOP Period	Pearson			Kendall			Spearman		
	r	p-value	SC	τ	p-value	SC	ρ	p-value	SC
Early morning	1.00*	<0.001	✓	0.96*	<0.001	✓	1.00*	<0.001	✓
AM peak	1.00*	<0.001	✓	0.96*	<0.001	✓	1.00*	<0.001	✓
Midday	0.99*	<0.001	✓	0.95*	<0.001	✓	0.99*	<0.001	✓
PM peak	0.99*	<0.001	✓	0.94*	<0.001	✓	0.99*	<0.001	✓

Note: *significant at 99% confidence level.

7. Discussion and Future Research

In general, SF estimations from both techniques are similar (Figure 14 and Table 7); nonetheless, there are few cases where performance results from the automatic definitions overestimate results (Figure 14, callout i). This usually occurs when adjacent traffic control devices are undetected as they have not yet been included in OSM or when large queues stretch over the entire approach, making entirely flat speed profiles (Figure 10). There is debate on whether intersections with long queues that extend to adjacent locations should be evaluated as one system, or if they should be assessed individually. This phenomenon should be further studied in future research where best practices to identify improvement opportunities are discussed. For the time being, practitioners need to be aware that performance estimations derived from the presented mapping technique may stem from approaches that include adjacent control devices and should review these definitions before making system changes.

Additionally, DSB [3] is also evaluated to assess the accuracy of downstream-focused performance measures derived from the proposed mapping technique. Figure 15 compares performance results from the automatic and manual methodologies, and Table 8 evaluates the correlation between the estimations. Some discrepancies are observed since the technique focuses on the upstream limits of the intersection. In the manual approach, the locations of the intersections' far sides are individually determined to accurately identify where the downstream segments start. In contrast, a far side location of 100 ft. was defined for all the evaluated movements with the automatic technique. Future research will focus on the improvement of DSB calculations by estimating the far side locations from CV data.

Table 7. Correlation between SF estimations.

TOP Period	Pearson			Kendall			Spearman		
	r	p-value	SC	τ	p-value	SC	ρ	p-value	SC
Early morning	0.94*	<0.001	✓	0.93*	<0.001	✓	0.94*	<0.001	✓
AM peak	0.70*	<0.001	✓	0.86*	<0.001	✓	0.90*	<0.001	✓
Midday	0.74*	<0.001	✓	0.79*	<0.001	✓	0.85*	<0.001	✓
PM peak	0.90*	<0.001	✓	0.86*	<0.001	✓	0.92*	<0.001	✓

Note: *significant at 99% confidence level.

Table 8. Correlation between DSB estimations.

TOP Period	Pearson			Kendall			Spearman		
	r	p-value	SC	τ	p-value	SC	ρ	p-value	SC
Early morning	0.83*	<0.001	✓	0.54*	<0.001		0.54*	<0.001	
AM peak	0.58*	<0.001		0.83*	<0.001	✓	0.86*	<0.001	✓
Midday	0.65*	<0.001		0.76*	<0.001	✓	0.81*	<0.001	✓
PM peak	0.67*	<0.001		0.84*	<0.001	✓	0.87*	<0.001	✓

Note: * significant at 99% confidence level.

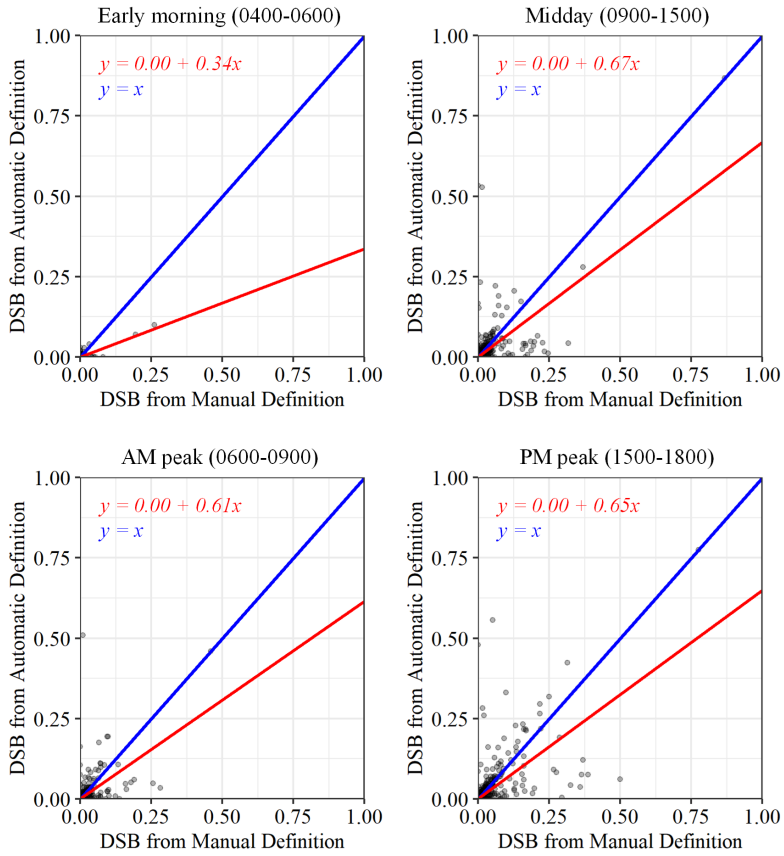

Figure 15. DSB estimation comparison.

Further improvements can be accomplished by solely using CV trajectory data to derive the location of map elements that would disrupt the progression of vehicles, such as traffic signals, stop signs, roundabouts, and railways. With this approach, more updated and complete surrounding information would be available which would surely improve the technique's performance.

8. Conclusions

This paper proposed a methodology based on CV trajectory and OSM data to improve the scalability of traffic signal performance measures by automatically mapping intersection centers (**Figure 3(b)**) and relevant approach areas (**Figure 9** and **Figure 10**). With the proposed technique, the manual effort required to estimate traffic signal performance for 1000 intersections can be reduced from over 80 hours to just 15 minutes.

First, intersection centers are approximated from OSM traffic signal tags. Then, intersection approaches are estimated by assessing relevant information about the surroundings of the OSM-derived centers extracted from trajectory heading and speed data. Speed profiles for the generated approaches and additional OSM map elements are then used to identify how far upstream from the intersection center trajectories should be used to derive movement-level traffic signal performance measures.

To evaluate the proposed technique, traffic signal performance at 500 intersections was estimated from the generated maps and was compared to results obtained from a manual methodology. Figures 12-14 and Tables 5-7 show that upstream-focused performance estimations from both techniques are strongly linearly correlated, suggesting that accurate performance measures can be derived from the proposed automatic methodology. The presented technique will aid agencies monitor statewide traffic signal performance with CV trajectory data.

Acknowledgements

March 2023 weekdays connected vehicle trajectory data used in this study was provided by Wejo Data Services, Inc. Map data copyrighted OpenStreetMap contributors and available from https://www.openstreetmap.org. This work was supported in part by the Joint Transportation Research Program and Pooled Fund Study (TPF-5(377)) led by the Indiana Department of Transportation (INDOT) and supported by the state transportation agencies of California, Connecticut, Georgia, Minnesota, North Carolina, Ohio, Pennsylvania, Texas, Utah, Wisconsin, plus the City of College Station, Texas, and the FHWA Operations Technical Services Team. The contents of this paper reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein, and do not necessarily reflect the official views or policies of the sponsoring organizations. These contents do not constitute a standard, specification, or regulation.

Conflicts of Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

[1] Urbanik, T., *et al.* (2015) Signal Timing Manual—Second Edition. National Academy of Sciences, Washington DC. https://doi.org/10.17226/22097

[2] Saldivar-Carranza, E., Li, H., Mathew, J., Hunter, M., Sturdevant, J. and Bullock, D. (2021) Deriving Operational Traffic Signal Performance Measures from Vehicle Trajectory Data. *Transportation Research Record: Journal of the Transportation Research Board*, **2675**, 1250-1264. https://doi.org/10.1177/03611981211006725

[3] Saldivar-Carranza, E.D., *et al.* (2023) Next Generation Traffic Signal Performance Measures: Leveraging Connected Vehicle Data. Purdue University, West Lafayette. https://doi.org/10.5703/1288284317625

[4] Saldivar-Carranza, E.D., Li, H. and Bullock, D.M. (2021) Identifying Vehicle Turning Movements at Intersections from Trajectory Data. 2021 *IEEE International Intelligent Transportation Systems Conference* (*ITSC*), Indianapolis, 19-22 September 2021, 4043-4050. https://doi.org/10.1109/ITSC48978.2021.9564781

[5] FHWA (2021) Every Day Counts: Innovation for a Nation on the Move. Federal Highway Administration, Washington DC.

[6] Mathew, J.K., Desai, J.C., Sakhare, R.S., Kim, W., Li, H. and Bullock, D.M. (2021) Big Data Applications for Managing Roadways. *ITE Journal*, **91**, 28-35.

[7] Day, C.M., *et al*. (2014) Performance Measures for Traffic Signal Systems: An Outcome-Oriented Approach. Purdue University, West Lafayette. https://doi.org/10.5703/1288284315333

[8] Zhao, Y., Zheng, J., Wong, W., Wang, X., Meng, Y. and Liu, H.X. (2019) Estimation of Queue Lengths, Probe Vehicle Penetration Rates, and Traffic Volumes at Signalized Intersections Using Probe Vehicle Trajectories. *Transportation Research Record*: *Journal of the Transportation Research Board*, **2673**, 660-670. https://doi.org/10.1177/0361198119856340

[9] Waddell, J.M., Remias, S.M. and Kirsch, J.N. (2020) Characterizing Traffic-Signal Performance and Corridor Reliability Using Crowd-Sourced Probe Vehicle Trajectories. *Journal of Transportation Engineering*, *Part A*: *Systems Archive*, **146**, Article ID: 04020053. https://doi.org/10.1061/JTEPBS.0000378

[10] Huang, J., Li, G., Wang, Q. and Yu, H. (2013) Real Time Delay Estimation for Signalized Intersection Using Transit Vehicle Positioning Data. 2013 13*th International Conference on ITS Telecommunications*, Tampere, 5-7 November 2013, 216-221. https://doi.org/10.1109/ITST.2013.6685548

[11] Day, C.M., *et al*. (2017) Detector-Free Optimization of Traffic Signal Offsets with Connected Vehicle Data. *Transportation Research Record*: *Journal of the Transportation Research Board*, **2620**, 54-68. https://doi.org/10.3141/2620-06

[12] Cao, L. and Krumm, J. (2009) From GPS Traces to a Routable Road Map. *Proceedings of the* 17*th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, Seattle, 4-6 November 2009, 3-12. https://doi.org/10.1145/1653771.1653776

[13] Carisi, R., Giordano, E., Pau, G. and Gerla, M. (2011) Enhancing in Vehicle Digital Maps via GPS Crowdsourcing. 2011 *Eighth International Conference on Wireless On-Demand Network Systems and Services*, Bardonecchia, 26-28 January 2011, 27-34. https://doi.org/10.1109/WONS.2011.5720196

[14] Jang, S., Kim, T. and Lee, E. (2010) Map Generation System with Lightweight GPS Trace Data. *The* 12*th International Conference on Advanced Communication Technology* (*ICACT*), Gangwon, 7-10 February 2010, 1489-1493.

[15] Erramaline, A., Badard, T., Côté, M.-P., Duchesne, T. and Mercier, O. (2022) Identification of Road Network Intersection Types from Vehicle Telemetry Data Using a Convolutional Neural Network. *ISPRS International Journal of Geo-Information*, **11**, Article No. 475. https://doi.org/10.3390/ijgi11090475

[16] Saldivar-Carranza, E.D., Li, H., Mathew, J.K., Gayen, S., Malackowski, H. and Bullock, D. M. (2023) Reporting Framework for Arterial-Level Traffic Signal Performance Measures Estimated from Connected Vehicle Trajectory Data. Purdue University, West Lafayette. https://doi.org/10.5703/1288284317617

[17] SAE International (2016) J2735D: Dedicated Short Range Communications (DSRC) Message Set Dictionary™. SAE International, Warrendale.

[18] Abernethy, B., Andrews, S. and Pruitt, G. (2012) Signal Phase and Timing (SPaT) Applications, Communications Requirements, Communications Technology Potential Solutions, Issues and Recommendations. National Transportation Library, McLean.

[19] OpenStreetMap Contributors (2023) Planet Dump. https://planet.osm.org

[20] Haklay, M. and Weber, P. (2008) OpenStreetMap: User-Generated Street Maps. *IEEE Pervasive Computing*, **7**, 12-18. https://doi.org/10.1109/MPRV.2008.80

[21] Pearson, K. (1895) Note on Regression and Inheritance in the Case of Two Parents. *Proceedings of the Royal Society of London*, **58**, 240-242.

https://doi.org/10.1098/rspl.1895.0041

[22] Kendall, M.G. and Gibbons, J.D. (1990) Rank Correlation Methods. Oxford University Press, New York.

[23] Spearman, C. (1904) The Proof and Measurement of Association between Two Things. *The American Journal of Psychology*, **15**, 72-101. https://doi.org/10.2307/1412159

[24] Mukaka, M. (2012) Statistics Corner: A Guide to Appropriate Use of Correlation Coefficient in Medical Research. *Malawi Medical Journal*, **24**, 69-71.

[25] Evans, J. (1996) Straightforward Statistics for the Behavioral Sciences. Thomson Brooks, Pacifc Grove.