Estimation of Connected Vehicle Penetration Rate on Indiana Roadways

Margaret Hunter, Jijo K. Mathew, Ed Cox, Matthew Blackwell, and Darcy M. Bullock

***Abstract*— Over 400 billion passenger vehicle trajectory waypoints are collected each month in the United States. This data creates many new opportunities for agencies to assess operational characteristics of roadways for more agile management of resources. This study compared traffic counts obtained from 24 Indiana Department of Transportation traffic counts stations with counts derived by the vehicle trajectories during the same periods. These stations were geographically distributed throughout Indiana with 13 locations on interstates and 11 locations on state or US roads. A Wednesday and a Saturday in January, August, and September 2020 are analyzed. The results show that the analyzed interstates had an average penetration of 4.3% with a standard deviation of 1.0. The non-interstate roads had an average penetration of 5.0% with a standard deviation of 1.36. These penetration levels suggest that connected vehicle data can provide a valuable data source for developing scalable roadway performance measures. Since all agencies currently have a highway monitoring system using fixed infrastructure, this paper concludes by recommending agencies integrate a connected vehicle penetration monitoring program into their traditional highway count station program to monitor the growing penetration of connected cars and trucks.**

I. Introduction

Commercialized, crowdsourced probe vehicle data has been available for about a decade for assisting individual drivers with traffic and route information and providing agencies with average segment speeds [1]–[3]. In recent years, connected vehicle trajectory data has expanded upon this concept. Now, the speed and location of individual vehicles are available. Monthly, approximately 400 billion passenger vehicle trajectory waypoints are collected in the United States [4]. Combining the information from individual vehicles creates a rich data set that has the potential to revolutionize how agencies evaluate their road networks. However, for this data to be beneficial to agencies, the data must be a representative sample of all the vehicles on the road. To assess the penetration level of connected vehicle trajectory data, this paper compares traffic counts obtained from 24 Indiana Department of Transportation (INDOT) count stations to counts derived from the connected vehicle trajectories for the same periods.

II. Literature Review

The earliest use of GPS based travel time data for systematic assessment of agency infrastructure occurred in Louisiana around 1999 [5]. Around 2007, crowdsourced vehicle probe data began to emerge for use by drivers and agencies through a growing number of providers and smartphone apps [6]–[8]. While this crowdsourced data was largely collected via drivers' smartphones, some providers were also able to incorporate GPS-enabled vehicles [9], [10]. Since then, there have been numerous studies looking extensively into the accuracy of these datasets. In 2008, approximately 2,500 miles of roadway along I-95 were used to evaluate travel time and speed data obtained from a commercial probe data provider [11]. Some additional comparison of commercial probe data studies include Kim and Coifman's two-month study where they compared speeds from probe vehicles to speeds from loop detectors [10], Zhang et al.'s study that compared it to data from Bluetooth sensors for arterials [12], and Ahsani et al.'s 4-year study in Iowa that compared it to Wavetronix smart sensors [13]. More recently, Hoseinzadeh et al. also compared speeds from commercial probe data with Bluetooth sensors for surface streets in Texas [9].

While these crowdsourced vehicle probe datasets have been supported and validated for years, a new type of probe data is emerging. Connected vehicle (CV) trajectory data that contains individual vehicle locations, timestamp, speed, and heading from onboard sensors is now available. Over the past couple of years, there have been many studies presenting methods for evaluating road networks using low-penetration trajectory data. In one study, Zhang et al.

*Research supported in part by the Joint Transportation Research Program and Pool Fund Study (TPF-5(377)).

M. Hunter is with Purdue University, West Lafayette, IN USA (phone: 317-640-6603; e-mail: hunter87@purdue.edu).

J. Mathew is with Purdue University, West Lafayette, IN USA (e-mail: kjijo@purdue.edu).

E. Cox is with the Indiana Department of Indiana, Indianapolis, IN USA (e-mail: ecox.indot.in.gov)

M. Blackwell is with Wejo Data Services, Manchester, UK (e-mail: matthew.blackwell@wejo.com).

D. Bullock is with Purdue University, West Lafayette, IN USA (e-mail: darcy@purdue.edu).

Recommended Citation:

Hunter, Margaret; Mathew, Jijo K.; Cox, Ed; Blackwell, Matthew; and Bullock, Darcy M., "Estimation of Connected Vehicle Penetration Rate on Indiana Roadways" (2021). *JTRP Affiliated Reports.* Paper 37.

evaluated their method for utilizing low penetration enhanced trajectory data to determine queue length using simulations [14]. Another study, by Zhao et al., also tested their method against simulations, as well as real-world data. They proposed a method for estimating queue length and traffic volume at low penetrations without explicitly needing to know the market penetration [15].

In 2016, Li et al. compared counts from a loop detector to counts obtained from connect vehicle trajectories. This study found an overall market penetration of 1.1% with a range of 0.2% to 2.0% depending on the time of day [16]. Since then, the number of connected vehicles on the road have and will continue to increase. Zhang et al. found that a minimum of 4% penetration was needed to increase ramp metering performance [17]. In sequential studies, Day et al. determined aggregate data at penetration levels as low as 0.09% - 0.8% would provide adequate representation for corridor retiming [18], [19].

While connected vehicle data offers many new opportunities for the evaluation of road networks [20]–[27], there are no reported evaluations of recent penetration rates. The objective of this paper is to present a methodology for characterizing the penetration of connected vehicles using our highways and apply that methodology to quantitatively characterize selected locations on Indiana interstates and surface streets.

III. Data

A. *Indiana Department of Transportation*

For this study, the counts obtained from Indiana Department of Transportation (INDOT) count stations are considered the ground truth vehicle counts. The study used 24 count stations, shown in Figure 1.. These stations are geographically distributed around the state of Indiana with 13 of the count stations located on interstates and the other 11 locations located on non-interstate roads, such as state roads or US highways. In addition to having a range of average annual daily traffic (AADT) values, the count stations are distributed among rural, suburban, and urban communities in a variety of different regions around Indiana.

INDOT's count stations consist of embedded loop detectors (callout i on Figure 2.) that record the speed and classification of every vehicle to pass over it. The data is aggregated into 15-minute bins and is available online on INDOT's Traffic Count Database System [28].

Figure 1. Locations of INDOT count stations classified by road type and AADT

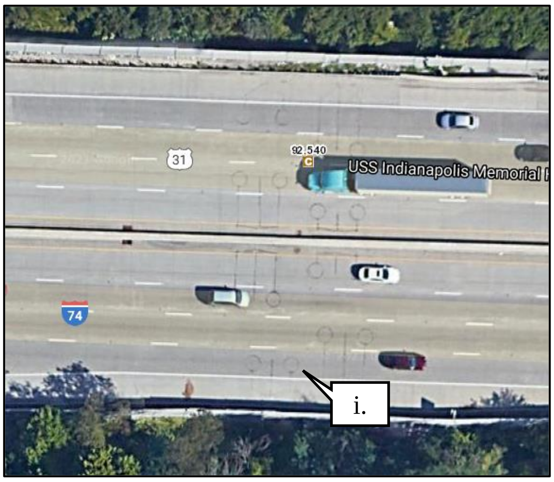
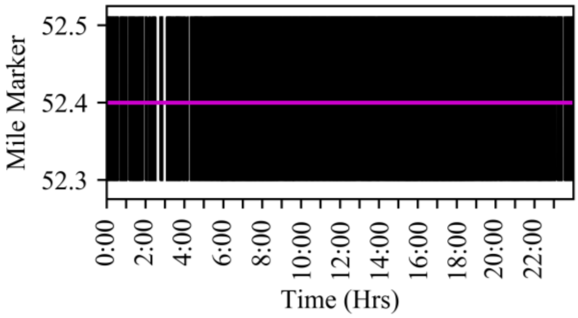


Figure 2. Loop detector count stations on Indiana roadways

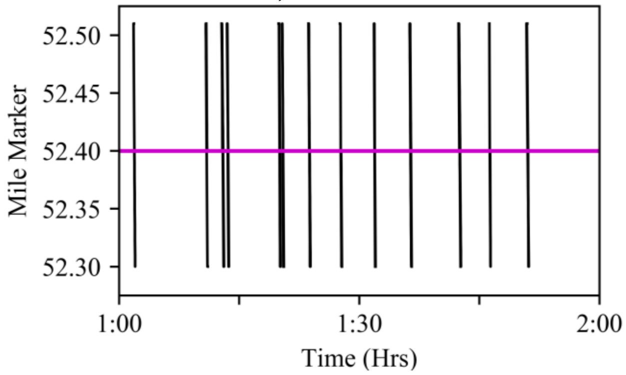
B. *Vehicle Trajectory Data*

The vehicle trajectory data used in this study was provided by a third-party commercial data provider who obtains their data directly from the original equipment manufacturers (OEMs). The data consists of anonymized individual waypoints collected every three seconds that contain a GPS location, date, time, speed, heading, and an anonymized trajectory identifier.

Quarter mile geofence regions were drawn centered at the location of the count station for both travel directions. From the vehicle trajectory waypoints located inside the geofence region, unique vehicle trajectories were identified and counted. Figure 3a. shows the 3,382 trajectories that passed an I-465 count station (location 990312) traveling in the outer loop (OL) on Jan 15, 2020. The large number of trajectories in Figure 3a. obscures the individual trajectories; therefore, Figure 3b. focuses on one hour and shows that 13 vehicle trajectories passed through between 1:00 am to 2:00 am.

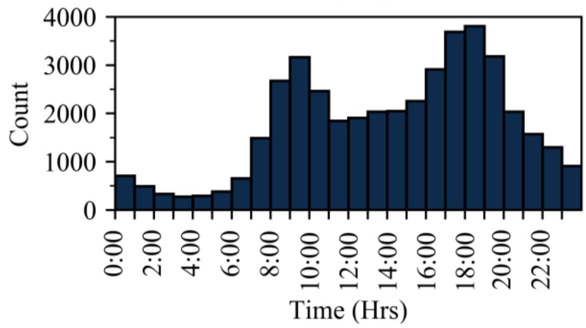


a) 24 hours

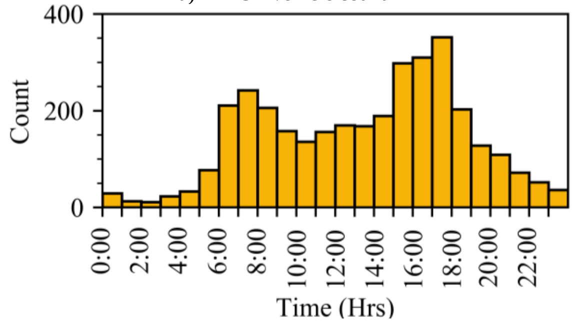


b) 1 hour from 1:00 am to 2:00 am

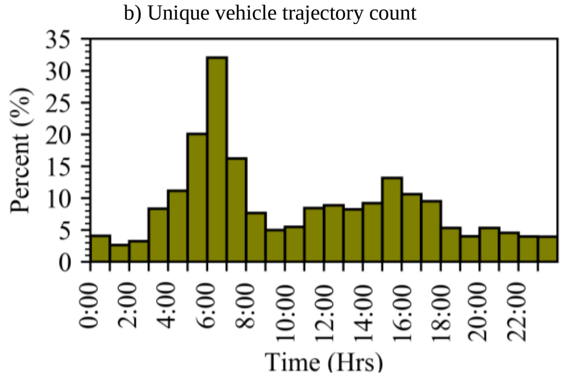
Figure 3. Vehicle trajectories on the outer loop of I-465 (location 990312) on January 15, 2020. Purple line indicates the location of the count station.



a) INDOT vehicle count



b) Unique vehicle trajectory count



c) Percent penetration

Figure 4. Hourly counts and percent penetration for at I-465 (location 990312) OL on Wednesday January 15, 2020

IV. METHODOLOGY

This study looks at a total of six days, three Wednesdays and three Saturdays, over three months, January, August, and September 2020.

- Wednesday, January 15, 2020
- Saturday, January 11, 2020
- Wednesday, August 19, 2020
- Saturday, August 15, 2020
- Wednesday, September 23, 2020
- Saturday, September 26, 2020

First, both the count station counts and vehicle trajectory counts are aggregated by hour and by direction. The hourly, directional percent penetration is calculated by

$$H_p = \left(\frac{V_h}{C_h}\right) \times 100 \qquad (1)$$

where H_p is the hourly percent penetration per direction, V_h is the hourly count of unique vehicle trajectories, and C_h is the hourly count of vehicles to pass the count station. Figure 4. graphically shows the INDOT counts, vehicle trajectory counts, and the resulting percent penetration for the I-465 OL station (location 990312).

The daily, directional percent penetration is determined by

$$D_p = \left(\frac{\sum V_h}{\sum C_h}\right) \times 100 \qquad (2)$$

where D_p is the daily percent penetration per direction, V_h is the hourly count of the vehicle trajectories, and C_h is the hourly count of vehicles to pass the count station. 0contains the hourly counts and percent penetration and the resulting total daily counts and percent penetration. Table Ii. presents the directional, daily counts and resulting penetration for the two days in January 2020 and the overall monthly count and penetration for the I-465 OL location (location 990312).

The monthly percent penetration, shown in Table III, is a non-directional measure calculated from the Wednesday and Saturday of that month using

$$M_p = \left(\frac{\Sigma V_d}{\Sigma C_d}\right) \times 100 \qquad (3)$$

where M_p is the monthly percent penetration, V_d is the daily count of the vehicle trajectories, and C_d is the daily count of vehicles to pass the count station.

The method of aggregating the counts over the day(s) in order to calculate the daily and monthly percent penetration was chosen over simply averaging the hourly and daily percent penetrations to eliminate the effects of a couple high or low hourly percent penetrations.

TABLE I. HOURLY INDOT AND VEHICLE TRAJECTORY COUNTS AND THE RESULTING PERCENT PENETRATION FOR I-456 (LOCATION 990312) OL

Time (hrs)	Count		% Penetration
	INDOT	**Veh. Traj.**	
0:00	708	29	4.10
1:00	491	13	2.65
2:00	336	11	3.27
3:00	275	23	8.36
4:00	295	33	11.19
5:00	383	77	20.10
6:00	658	209	31.76
7:00	1491	242	16.23
8:00	2676	207	7.74
9:00	3167	158	4.99
10:00	2465	137	5.56
11:00	1847	156	8.45
12:00	1908	170	8.91
13:00	2037	167	8.20
14:00	2050	190	9.27
15:00	2259	296	13.10
16:00	2915	312	10.70
17:00	3690	352	9.54
18:00	3808	203	5.33
19:00	3185	128	4.02
20:00	2037	109	5.35
21:00	1575	72	4.57
22:00	1301	52	4.00
23:00	910	36	3.96
Total	42467	3382	7.96

TABLE II. MONTHLY SUMMARY FOR I-465 LOCATION (LOCATION 990312) OL

Day	Dir.	% Pen.	Count	
			INDOT	**Veh. Traj.**
Wed Jan 15, 2020	Outer Loop (OL)	7.96	42467	3382
	Inner Loop (IL)	5.56	59927	3334
Sat Jan 11, 2020	Outer Loop (OL)	8.96	33762	3024
	Inner Loop (IL)	5.95	47582	2832
Jan Avg.		6.84	183738	12572

V. RESULTS

Table Iii. presents the average percent penetration for the 24 count stations. Some INDOT permanent count stations had no available data for the days of interest; therefore, the asterisks and blank boxes indicate that either one day or both days of data were missing, respectively. Table Iv. lists the summary statistics for interstate, non-interstate, and all count stations. Figure 5. graphically depicts the average percent penetration for the 24 count stations, grouped by road type. The orange line represents the average percent penetration for that road type. The overall CV penetration average is 4.7%, with the interstate locations averaging 4.3% and the non-interstate locations averaging 5.0%. The standard deviation ranged from 1.0 to 1.36.

A location along Indiana SR 9 near Anderson, IN (location 990301) stands out for having a percent penetration that is roughly three standard deviations above the non-interstate average. The AADT for the location is the median for the non-interstate roads, and therefore, is likely not a large factor in the percent penetration. This location, however, is 30 miles from an OEM facility, which is less than a mile and half from SR 9 [29]. Vehicles from that particular OEM are a significant contributor to the CV data used in this study.

Overall, the percent penetration generally fluctuates within a 2% range. The AADT and location of the count stations explain little. A potential explanation for the fluctuation is the hourly variation in the proportion of commercial vehicles, which may be underrepresented in the current CV data set. This potentially could explain the higher average percent penetration along non-interstate roads. Future research will examine the relationship between percentage of heavy vehicles and market penetration.

TABLE III. SUMMARY TABLE OF THE PERCENT PENETRATIONS FOR THE 24 COUNT LOCATIONS

Location ID	Route	AADT	Avg. Penetration		
			Jan 2020	Aug 2020	Sept 2020
990104	I65	61790	4.11		
991211	I69	56158	3.79*	3.08	3.01
990511	I265	56431	4.37		3.71*
950507	I65	34932	3.86	3.89	3.87
990329	I70	52737	4.47		3.84
952000	I69	25406		4.35	5.91
954400	I94	97824			3.96
955300	I74	31121	4.28		
950106	I70	30506		3.21	3.06
990611	I64	10794	4.49		
991325	I465	106368		4.52	4.35*
990312	I465	92540	6.84		
991374	I69	114909	6.26*	5.41	5.47
990501	SR37	37738	5.41	5.15*	5.12
990509	SR56	3737	5.30	5.43	4.71
951000	US41	3176	3.83	4.68	5.31
990403	US20	35793	3.10	3.27	3.35
950436	US30	17392	4.67	4.71	
990607	US21	18954	4.80	4.68	4.67
990506	US50	10524	3.62	5.26	4.83
990301	SR9	15529	9.78		8.91
990101	US52	19864	4.77	4.47	4.28
952100	US24	9566	5.66	5.36	5.03
990308	US40	7058	5.38	5.32	3.58

* count station data only available for one day of the month

Note: blank boxes indicate that INDOT counts were unavailable

TABLE IV. PERCENT PENETRATION SUMMARY STATISTICS

	Interstates	Non-Interstates	All
Min	3.01	3.10	3.01
Max	6.84	9.78	9.78
Mean	4.34	4.98	4.70
Standard Deviation	1.00	1.36	1.25

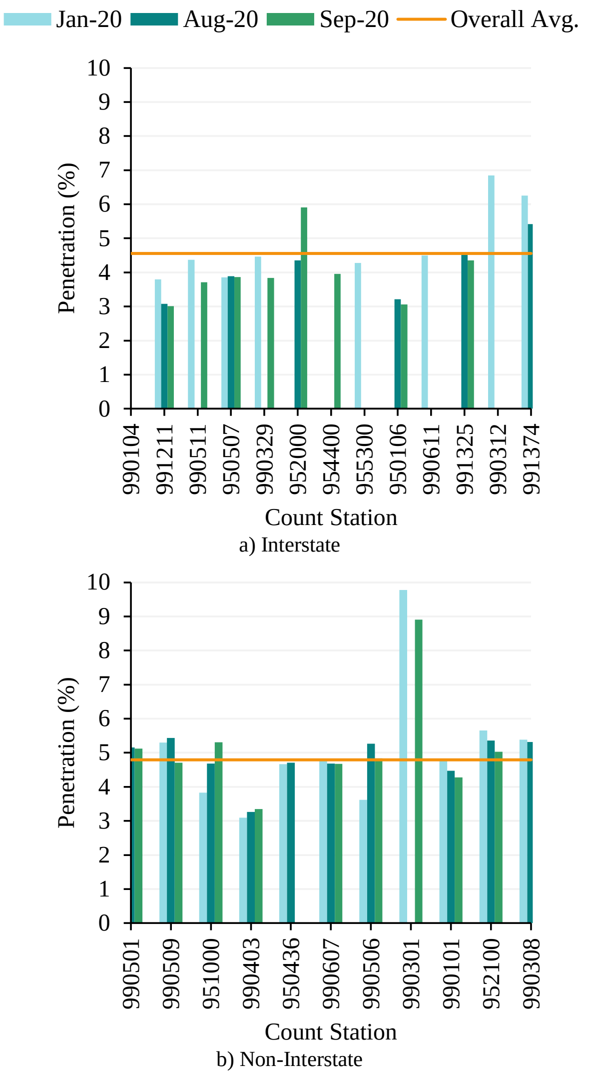


Figure 5. Summary plots of the monthly percent penetrations for the 24 count stations

VI. CONCLUSION

This study compared vehicle counts obtained by INDOT count stations to the number of unique trajectories crossing the count station in order to determine the percent penetration for a vehicle trajectory dataset. A method for calculating the hourly, daily, and monthly percent penetrations was presented. Figure 4. graphically depicted the hourly sample data, while 0and Table Ii. listed the sample data. 24 locations, of varying AADT, location, and road type, were analyzed (Figure 1., Table Iii. and the average percent penetration was determined to be 4.7% with a standard deviation of 1.25. Since all agencies currently have a highway monitoring system using fixed infrastructure, this paper concludes by recommending agencies integrate a connected vehicle penetration monitoring program into their traditional highway count station program to monitor the growing penetration of connected cars and trucks.

ACKNOWLEDGMENT

Trajectory data used in this study, was provided by Wejo Data Services, Inc. The vehicle counts were obtained from the Indiana Department of Transportation. 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.

REFERENCES

[1] M. S. Lasky, "Real-time traffic info gets you past jams," *PC World (San Fr. CA)*, vol. 25, no. 2, p. 26, 2007, [Online]. Available: https://www.proquest.com/magazines/real-time-traffic-info-gets-you-past-jams/docview/231401823/se-2?accountid=13360.

[2] James Chase, "There ' s an App for That," *Mob. Guid.*, pp. 10–11, 2010.

[3] S. Remias *et al.*, "Performance characterization of arterial traffic flow with probe vehicle data," *Transp. Res. Rec.*, vol. 2380, no. 2380, pp. 10–21, Jan. 2013, doi: 10.3141/2380-02.

[4] R. Downing, "5 Reasons Why Companies Choose Connected Vehicle Data Over Mobile Data," 2020. https://www.wejo.com/press/5-reasons-why-companies-choose-connected-car-data-over-mobile-data.

[5] C. A. Quiroga and D. Bullock, "Travel time information using global positioning system and dynamic segmentation techniques," *Transp. Res. Rec.*, vol. 1660, no. 1660, pp. 48–57, Jan. 1999, doi: 10.3141/1660-07.

[6] D. Wang, "Stuck in Traffic?," *Google Official Blog*, 2007. https://googleblog.blogspot.com/2007/02/stuck-in-traffic.html.

[7] U. Levine, "From the Co-Founder of Waze, A Blueprint To Eliminate Traffic Jams," *Forbes*, Jan. 2019.

[8] INRIX, *INRIX: A History of Innovation*. .

[9] N. Hoseinzadeh, Y. Liu, L. D. Han, C. Brakewood, and A. Mohammadnazar, "Quality of location-based crowdsourced speed data on surface streets: A case study of Waze and Bluetooth speed data in Sevierville, TN," *Comput. Environ. Urban Syst.*, vol. 83, p. 101518, Sep. 2020, doi: 10.1016/j.compenvurbsys.2020.101518.

[10] S. Kim and B. Coifman, "Comparing INRIX speed data against concurrent loop detector stations over several months," *Transp. Res. Part C Emerg. Technol.*, vol. 49, pp. 59–72, Dec. 2014, doi: 10.1016/j.trc.2014.10.002.

[11] A. Haghani, M. Hamedi, and K. F. Sababadi, "I-95 Corridor Coalition Vehicle Probe Project : Validation of INRIX data July-September 2008 Final Report," *I-95 Corridor Coalition*, 2009. http://www.i95coalition.org. (accessed Mar. 27, 2021).

[12] X. Zhang, M. Hamedi, and A. Haghani, "Arterial travel time validation and augmentation with two independent data sources," *Transp. Res. Rec.*, vol. 2526, pp. 79–89, 2015, doi: 10.3141/2526-09.

[13] V. Ahsani, M. Amin-Naseri, S. Knickerbocker, and A. Sharma, "Quantitative analysis of probe data characteristics: Coverage, speed bias and congestion detection precision," *J. Intell. Transp. Syst. Technol. Planning, Oper.*, vol. 23, no. 2, pp. 103–119, 2019, doi: 10.1080/15472450.2018.1502667.

[14] H. Zhang, H. X. Liu, P. Chen, G. Yu, and Y. Wang, "Cycle-Based End of Queue Estimation at Signalized Intersections Using Low-Penetration-Rate Vehicle Trajectories," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 8, pp. 3257–3272, Aug. 2020, doi: 10.1109/TITS.2019.2925111.

[15] Y. Zhao, J. Zheng, W. Wong, X. Wang, Y. Meng, and H. X. Liu, "Estimation of Queue Lengths, Probe Vehicle Penetration Rates, and Traffic Volumes at Signalized Intersections using Probe Vehicle Trajectories," *Transp. Res. Rec.*, vol. 2673, no. 11, pp. 660–670, Nov. 2019, doi: 10.1177/0361198119856340.

[16] H. Li, C. M. Day, and D. M. Bullock, "Virtual detection at intersections using connected vehicle trajectory data," in *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, Dec. 2016, pp. 2571–2576, doi: 10.1109/ITSC.2016.7795969.

[17] C. Zhang, J. Wang, J. Lai, X. Yang, Y. Su, and Z. Dong, "Extracting Origin-Destination with Vehicle Trajectory Data and Applying to Coordinated Ramp Metering," *J. Adv. Transp.*, vol. 2019, 2019, doi: 10.1155/2019/8469316.

[18] C. M. Day and D. M. Bullock, "Detector-free signal offset optimization with limited connected vehicle market penetration: Proof-of-concept study," *Transp. Res. Rec.*, vol. 2558, no. 1, pp. 54–65, Jan. 2016, doi: 10.3141/2558-06.

[19] C. M. Day *et al.*, *Detector-free optimization of traffic signal offsets with connected vehicle data*, vol. 2620. 2017.

[20] J. Desai, H. Li, J. K. Mathew, Y.-T. Cheng, A. Habib, and D. M. Bullock, "Correlating Hard-Braking Activity with Crash Occurrences on Interstate Construction Projects in Indiana," *J. Big Data Anal. Transp.*, vol. 3, no. 1, pp. 27–41, 2020, doi: 10.1007/s42421-020-00024-x.

[21] E. Saldivar-Carranza, H. Li, J. Mathew, M. Hunter, J. Sturdevant, and D. M. Bullock, "Deriving Operational Traffic Signal Performance Measures from Vehicle Trajectory Data," *Transp. Res. Rec. J. Transp. Res. Board*, In press, 2021, doi: 10.1177/03611981211006725.

[22] M. Hunter, E. Saldivar-Carranza, J. Desai, J. K. Mathew, H. Li, and D. M. Bullock, "A Proactive Approach to Evaluating Intersection Safety Using Hard-Braking Data," *J. Big Data Anal. Transp.*, pp. 1–14, Apr. 2021, doi: 10.1007/s42421-021-00039-y.

[23] H. Li, J. Mackey, M. Luker, M. Taylor, and D. M. Bullock, "Application of High-Resolution Trip Trace Stitching to Evaluate Traffic Signal System Changes," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2673, no. 9, pp. 188–201, Sep. 2019, doi: 10.1177/0361198119841043.

[24] H. Li, J. Mathew, W. Kim, and D. Bullock, "Using Crowdsourced Vehicle Braking Data to Identify Roadway Hazards," *JTRP Affil. Reports*, no. 34, 2020, doi: https://doi.org/10.5703/1288284317272.

[25] J. K. Mathew, J. C. Desai, R. S. Sakhare, W. Kim, H. Li, and D. M. Bullock, "Big Data Applications for Managing Roadways," *Inst. Transp. Eng. ITE J.*, vol. 91, no. 2, pp. 28–35, 2021.

[26] J. M. Waddell, S. M. Remias, J. N. Kirsch, and T. Trepanier, "Utilizing Low-Ping Frequency Vehicle Trajectory Data to Characterize Delay at Traffic Signals," *J. Transp. Eng. Part A Syst.*, vol. 146, no. 8, p. 04020069, Aug. 2020, doi: 10.1061/JTEPBS.0000382.

[27] W. Ma, L. Wan, C. Yu, L. Zou, and J. Zheng, "Multi-objective optimization of traffic signals based on vehicle trajectory data at isolated intersections," *Transp. Res. Part C Emerg. Technol.*, vol. 120, p. 102821, Nov. 2020, doi: 10.1016/j.trc.2020.102821.

[28] INDOT and MS2, "Traffic Count Database System." https://indot.public.ms2soft.com/tcds/tsearch.asp?loc=Indot&mod .

[29] GM, "Marion Metal Center." https://media.gm.com/media/us/en/gm/company_info/facilities/stamping/marion.html.