Estimating Turning Movement Counts from Probe Data

Jorge A. Barrios and Ryan Casburn (Kittelson & Associates, Inc.)

Abstract

Probe-data from cell phones and connected vehicles has greatly enhanced transportation engineering and planning. WIthin the past five years, probe data vendors have begun offering estimates of traffic volumes based on extrapolations of their data samples. This paper explores the suitability of probe data-based traffic volume estimates at intersections.

Two intersections along US 192 in Central Florida—one small and one large— were selected as test locations. The findings of a traditional traffic data collection and analysis effort at these two intersections were compared to the findings from a probe data-based effort.

The results of the comparison show that probe data-based estimates were decent substitutes for field-collected traffic counts at the large intersection. At this test location, the intersection delay results from the probe data-based analysis were within 8 percent (AM) and 14 percent (PM) of the results using the field-collected counts. While 8 percent and 14 percent error rates may seem high, it is noted that some of the discrepancy is inherent to comparing a three-month average from probe data against a single day of traffic counts.

On the other hand, major discrepancies on the count estimates for one of the intersection approaches to the small intersection resulted in extremely large differences in overall results at the small intersection.

Additional research on more test locations is necessary to better understand the likelihood of major discrepancies and to develop checks to identify them without a field-collected comparison dataset.

Keywords: *turning movement counts, StreetLight Data, probe data, volume estimates*

Introduction

Probe-data from cell phones and connected vehicles has greatly enhanced transportation engineering and planning, starting with travel time data, passing through origin-destination patterns, and now entering into traffic volume estimates. Traffic volumes are usually measured at either segments or intersections—where they are commonly known as turning movement counts (TMCs).

TMCs are the bread and butter of traffic engineers everywhere. The current standard practice of collecting TMCs involves a person in the loop counting vehicles. This can either be done live or from video footage.

Recently, computer vision has enabled automatic counting of TMCs from video footage with an acceptable level of accuracy. (Miovision—a leading vendor of computer vision-enabled cameras—ensures 95 percent accuracy after the data passes through its quality checks.) In addition, research and experimentation is ongoing to obtain TMCs from detectors in the pavement. Although both of these approaches greatly reduce the labor associated with obtaining TMCs, they still require physical infrastructure to be available at the intersection.

This paper explores another option: estimating TMCs from probe data. Probe data is sourced from connected vehicles, cell phones, fleet telematics, and other mobile sources. Most probe data is obtained from commercial vendors, which process, aggregate, anonymize, and sell data. INRIX, HERE, AirSage, and StreetLight Data are examples of probe data vendors.

Background

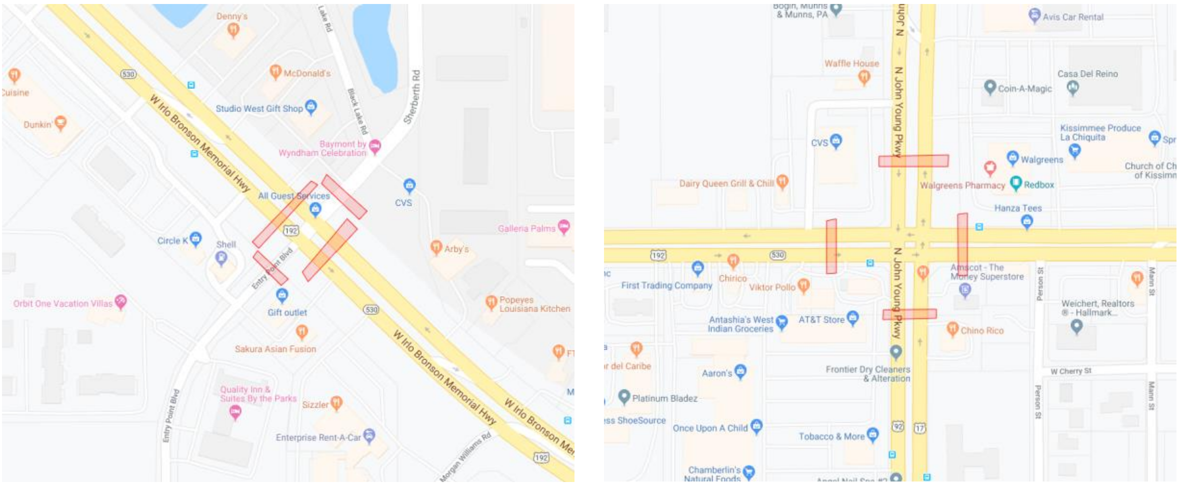
As part of an FDOT corridor study on US 192 in Osceola County (Florida), Kittelson obtained probe data from StreetLight Data. For the same study, Kittelson also obtained TMCs from TEDS—a local transportation data vendor which usually relies on video footage and manual counts for their TMC projects. The availability of these two datasets inspired the authors to test and compare TMC estimates from probe data against field-collected TMCs.

Setup

For this comparison, Kittelson selected two test intersections, one small and one large. The small intersection is US 192 & Sherberth Road; the large intersection is US 192 & John Young Parkway. As can be seen in Figure 1, Kittelson used the StreetLight Data platform to draw rectangular “pass-through” analysis zones on each approach to the intersections.

Figure 1: Analysis Zones for Test Intersections

Small intersection (Sherberth Road) **Large intersection (John Young Parkway)**



The probe data was calibrated based on annual average daily traffic (AADT) counts conducted by FDOT in 2017 at seven locations in the project vicinity. When the calibration feature is selected, StreetLight Data scales up the probe data sample with a single factor or multiplier. Recently, StreetLight Data began offering volume estimates using their own calibration data.

Several combinations of data sources and data periods were entered into the StreetLight Data platform to identify the dataset most comparable to the field-collected TMCs[1]. StreetLight Data offers two data sources: GPS and location-based services (LBS). The GPS data source is based on "fine" location data from connected vehicles, fleet telematics, and other probes using GPS to determine location. LBS data is obtained from mobile devices' "coarse" location data.

GEH statistics were computed for each pair of probe-based and field-collected turning movements, and then averaged across each of the datasets obtained. The GEH statistic is a measure commonly used in travel demand modeling and simulation. This measure is similar to the chi-squared test, and avoids the pitfalls of using a straight percentage comparison due to the wide range of values of hourly traffic. The equation to calculate GEH is shown below. For simulation, a general threshold is a GEH value under five is generally considered adequate and a GEH value between five and ten should be investigated.[2]

$$GEH = \sqrt{\frac{2(M-C)^2}{M+C}}$$

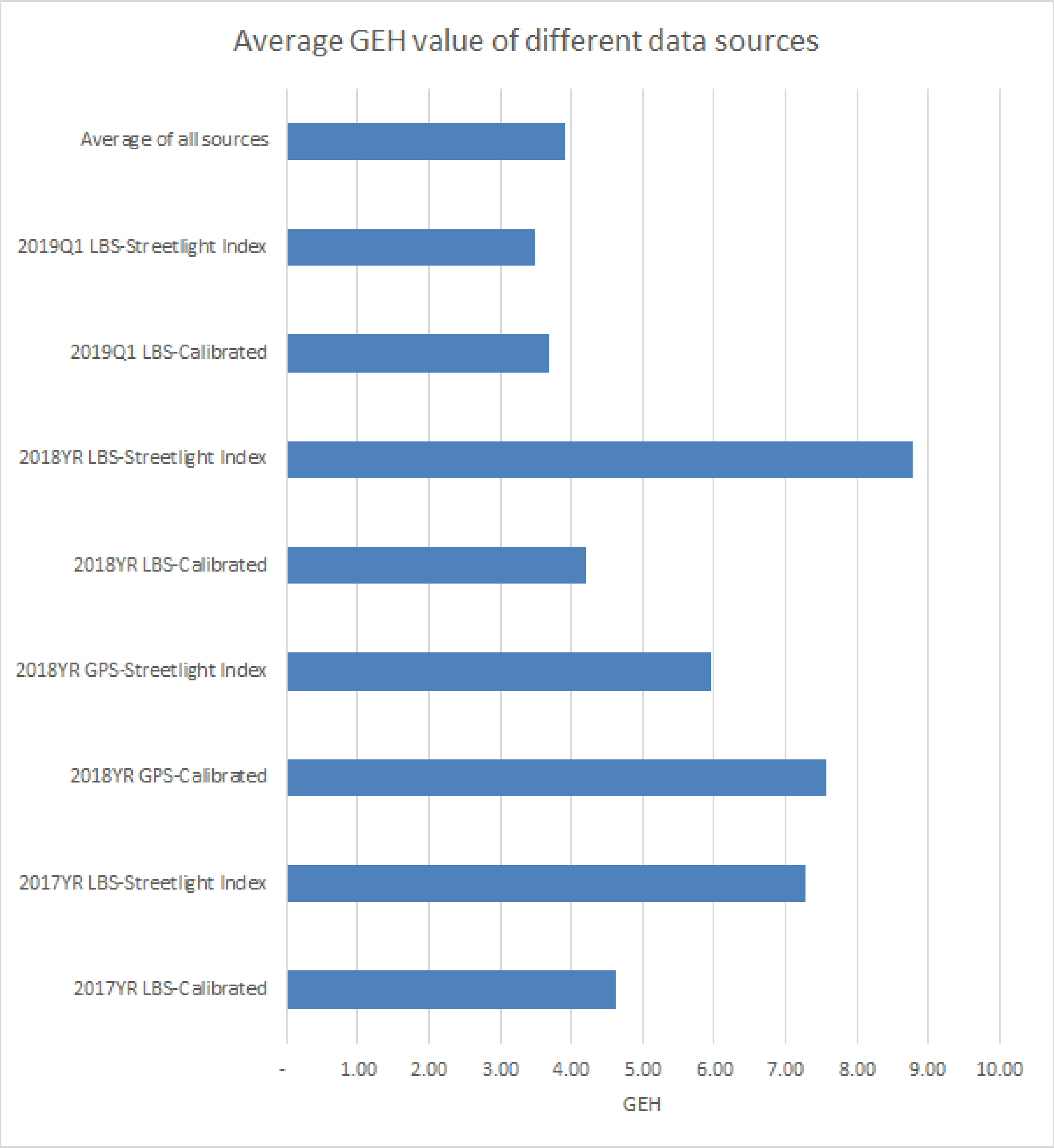
Where M = Modeled volume and C = Counted Volume

[1] Field-collected TMCs are often considered "ground truth". But this only holds for the snapshot of time that the TMCs covered (e.g., Tuesday, March 26, 2019 from 4 to 6 PM). The number of vehicles turning at an intersection varies on a daily basis.

[2] 2014 FDOT Traffic Analysis Handbook

Figure 2 illustrates the average GEH statistic across the datasets evaluated. The dataset using LBS data and covering the first three months of 2019 has the lowest average GEH statistic, meaning that it is most comparable to the field-collected TMCs. This makes sense given that the field-collected TMCs were conducted in March 2019 and that the 2019 and newer LBS datasets use a much higher sample size than previous years. While the StreetLight Index version had a slightly lower average GEH, the rest of this paper uses the locally-calibrated version as it showed more consistent GEH values over the individual movements.

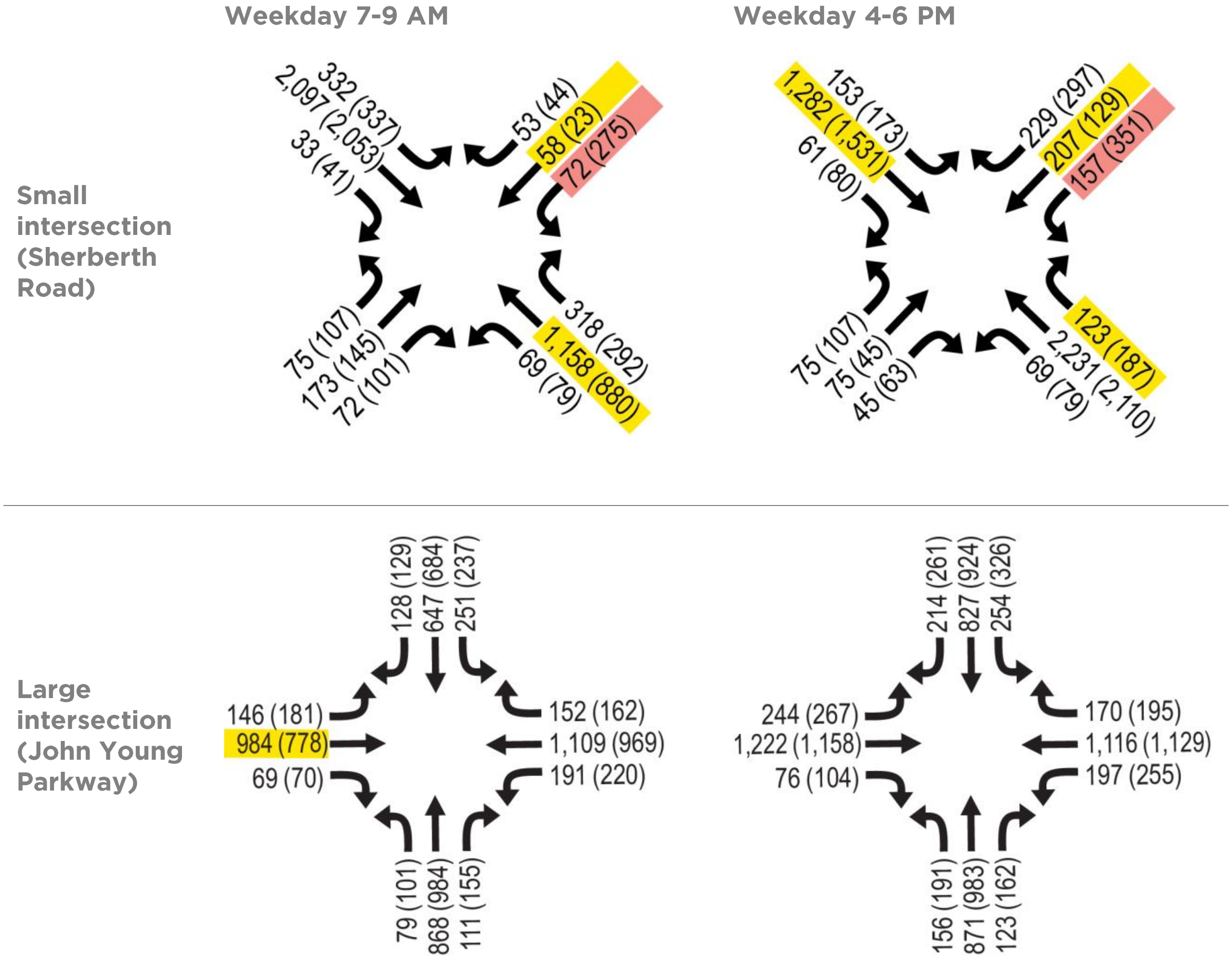
Figure 2: Average GEH Values across Datasets



TMC Comparison

Figure 3 presents a comparison of field-collected counts and probe data-based estimates for each turning movement. Yellow and red are used to highlight turning movements for which the probe data-based estimate deviates from the field-collected counts by 5 GEH and 10 GEH, respectively.

Figure 3: TMC Comparison (Intersection Diagram)

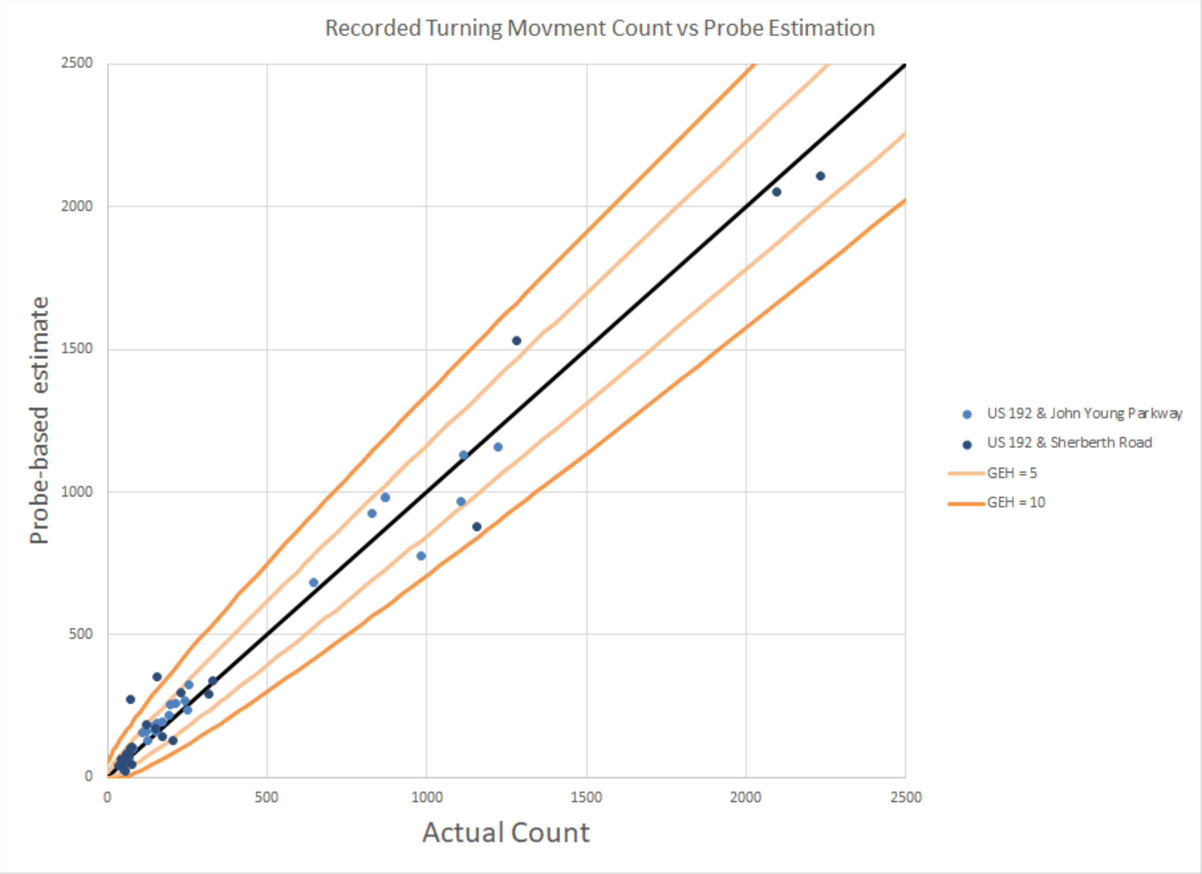


Legend: Field-collected TMCs (probe-based estimates)
GEH > 5 | GEH > 10

Figure 3 shows large differences (GEH > 10) for the southbound left movement from Sherberth Road (the small intersection). Differences of GEH between 5 and 10 were noted for four other movements at the small intersection and for one movement at the large intersection.

Figure 4 presents the same comparison using a scatter plot. Each point represents a movement at the Sherberth Road and John Young Parkway intersections. The value along the x-axis is the field collected data and the data along the y-axis is the probe-based estimate. If these values match exactly, they would fall on the straight black line. A value above the black line represents a probe-based estimate that is high, a value below the black line represents a probe-based estimate that is low. A point within the outer orange line has a GEH below 10, and a point within the inner orange line has a GEH below five.

Figure 4: TMC Comparison (Scatter Plot)



Intersection Performance Comparison

For most traffic engineering projects, TMCs are only interim numbers on the path to performance results—such as average delay or level of service. This section compares two sets of performance results at the two intersections: one with the field-collected TMCs and one with the probe-based TMC estimates. All other inputs, such as lane geometry or signal timing, are identical. Table 1 shows a comparison of results obtained from field-collected counts versus those obtained from probe data-based estimates.

Table 1: Intersection Performance Comparison

Intersection	Approach		7-9 AM Average Hour		4-6 PM Average Hour	
			Delay [sec/veh]	LOS	Delay [sec/veh]	LOS
Small intersection (Sherberth Road)	SB	R	71.9 (71.5)	E(E)	99.1(156.2)	F(F)
		T	71.7 (70.1)	E(E)	82.5 (76.4)	F(E)
		L	215.6 (1581.5)	F(F)	123.3 (496.0)	F(F)
	WB	R	2.8 (1.8)	A(A)	3.5 (2.7)	A(A)
		T	1.4 (0.9)	A(A)	1.8 (1.4)	A(A)
		L	108.1 (98.2)	F(F)	108.5 (107.0)	F(F)
	NB	R	98.3 (101.7)	F(F)	76.3 (75.9)	E(E)
		T	0.0 (0.0)	A(A)	0.0 (0.0)	A(A)
		L	77.9 (75.4)	E(E)	116.7 (109.5)	F(F)
	EB	R	1.3 (1.2)	A(A)	16.1 (14.9)	B(B)
		T	0.7 (0.6)	A(A)	15.7 (14.4)	B(B)
		L	73.2(73.0)	E(E)	132.1 (150.1)	F(F)
	Int.		19.4 (116.5)	B (F)	27.9 (60.6)	C (F)
Intersection	**Approach**		**7-9 AM Average Hour**		**4-6 PM Average Hour**	
			Delay [sec/veh]	**LOS**	**Delay [sec/veh]**	**LOS**
Large intersection (John Young Parkway)	SB	R	50.0 (47.4)	D(D)	74.7 (70.9)	E(E)
		T	52.8 (50.5)	D(D)	77.4 (71.8)	E(E)
		L	89.0 (88.0)	F(F)	113.7 (118.9)	F(F)
	WB	R	34.4 (38.7)	C(D)	41.0 (53.1)	D(D)
		T	33.6 (37.7)	C(D)	40.1 (51.5)	D(D)
		L	26.4 (28.1)	C(C)	33.9 (54.4)	C(D)
	NB	R	84.8 (88.8)	F(F)	110.8 (115.6)	F(F)
		T	76.4 (78.8)	E(E)	100.4 (103.8)	F(F)
		L	88.2 (87.6)	F(F)	111.8 (111.4)	F(F)
	EB	R	34.4 (37.6)	C(D)	38.6 (50.9)	D(D)
		T	33.8 (37.0)	C(D)	38.1 (50.0)	D(D)
		L	27.9 (31.4)	C(C)	36.2 (54.5)	D(D)
	Int.		49.6 (53.6)	D (D)	63.4 (72.1)	E (E)

Legend: Results using field-collected TMCs (probe-based estimates)
% Error > 10% | % Error > 20%

Table 1 shows that probe data-based estimates were decent substitutes for field-collected traffic counts at the large intersection (John Young Parkway). When compared to results obtained from field-collected counts, the overall intersection delays from the probe data-based analysis were within 8 percent for the AM peak hour and 14 percent for the PM peak hour.

On the other hand, the major discrepancies on the count estimates for one of the intersection approaches to the small intersection (Sherberth Road) resulted in extremely large differences in delay. However, if the erroneous approach was excluded, the average intersection delay would be within 1 percent and 17 percent for the AM and PM peak hours, respectively.

Practical Considerations

For practicing traffic engineers, the accuracy of probe-based TMC estimates is only part of the equation. The cost and availability of data are also important factors when selecting datasets for analyses. Table 2 summarizes the key advantages and disadvantages of probe-based TMC estimates when compared to traditional methods of obtaining TMCs.

Table 2: Advantages/Disadvantages for Application

Advantages of Probe-Based TMCs	Disadvantages of Probe-Based TMCs
• Easily scalable to obtain TMCs across large areas or even regions without field deployments • Estimates can be obtained using months' worth of data, not just one or two days • Can be used to easily produce weekend and off-peak estimates • Data can be obtained quickly and retroactively—as opposed to scheduling a data collection window in the future	• Supplementary information such as heavy vehicle percentages, crosswalk volumes, etc. are not yet obtainable using this approach • Can be costlier than field-collected TMCs, especially for small projects • May require recent daily counts in the project vicinity for calibration purposes • Not yet a widely accepted practice

Conclusion and Future Work

Based on the limited sample in this study, it appears that the probe data-based TMC estimates would be suitable for planning-level analyses of large intersections. While 8 percent (AM) and 14 percent (PM) error rates may seem high, it is noted that some of the discrepancy is inherent to comparing a three-month average from probe data against a single day of traffic counts.

Looking ahead, StreetLight Data is planning to offer TMC estimates as part of their product offering in the near future. This will reduce the level of effort associated with obtaining these estimates and will likely expand their usage throughout the profession.

As evidenced by the comparison in this paper, there are situations in which the estimates' errors are too large for use even at a planning-level. Additional research is necessary to better understand the nature of those situations, such as traffic levels or intersection geometries. The profession would also benefit from the identification of tell-tale signs of large errors in TMC estimates, especially when field-collected data are not available for comparison. For example, the large error in the small intersection's southbound approach became readily apparent from the performance results. The southbound left turn at Sherberth Road showing a volume-to-capacity (v/c) ratio of 4.2 and 1.9 in the AM and PM peak hours, respectively. A v/c ratio that much over one would be unexpected as theoretically v/c ratios should be less than or equal to one. However, more work is needed to determine movements that cannot be accurately determined from probe data. While an estimate that is too high can be found using v/c ratios, this approach cannot find estimates which are too low.

KITTELSON & ASSOCIATES