11-11-2016

Detector-Free Optimization of Traffic Signal Offsets with Connected Vehicle Data

Cristopher Day

Howell Li

Lucy Richardson

James Howard

Tom Platte

See next page for additional authors

Follow this and additional works at: <https://docs.lib.purdue.edu/civeng>

Part of the [Civil Engineering Commons](#), and the [Transportation Engineering Commons](#)

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries.
Please contact epubs@purdue.edu for additional information.

Authors

Cristopher Day, Howell Li, Lucy Richardson, James Howard, Tom Platte, Jim Sturdevant, and Darcy M. Bullock

---

Detector-Free Optimization of Traffic Signal Offsets with Connected Vehicle Data

Christopher M. Day*

Purdue University
550 Stadium Mall Drive West Lafayette, IN 47906
(765) 494-9601 cmday@purdue.edu

Howell Li

Purdue University
550 Stadium Mall Drive West Lafayette, IN 47906
(765) 494-9601 howell-li@purdue.edu

Lucy Richardson

Purdue University
550 Stadium Mall Drive West Lafayette, IN 47906
(765) 496-7314 richa297@purdue.edu

James Howard

Indiana Department of Transportation
185 Agrico Ln Seymour, IN 47274
(812) 569-2972 jhoward@indot.in.gov

Tom Platte

Indiana Department of Transportation
315 Boyd Blvd Laporte, IN 46350
(219) 325-7492 toplatte@indot.in.gov

James R. Sturdevant

Indiana Department of Transportation
8620 E. 21st St. Indianapolis, IN 46219
(317) 691-9091 jsturdevant@indot.in.gov

Darcy M. Bullock

Purdue University
550 Stadium Mall Drive West Lafayette, IN 47906
(765) 494-2226 darcy@purdue.edu

*Corresponding author.

Word Count: 4945 words + 10 * 250 words/(figure-table) = **7445** words

November 11, 2016

ABSTRACT

It has recently been shown that signal offset optimization is feasible using vehicle trajectory data at low levels of market penetration. This study performs offset optimization on two corridors using this type of data. Six weeks of trajectory splines were processed for two corridors including 25 signalized intersections, in order to create vehicle arrival profiles, using a proposed procedure called “virtual detection.” After processing and filtering the data, penetration rates between 0.09–0.80% were observed, varying by approach. The arrival profiles were statistically compared against those measured with physical detectors, with the majority of the approaches showing statistically significant goodness-of-fit at a 90% confidence level. Finally, the virtual detection arrival profiles were used to optimize offsets, and compared against a solution derived from physical detector arrival profiles. The results demonstrate that virtual detection can produce good quality offsets with current market penetration rates of probe data. The study also includes a sensitivity analysis to the sample period, which shows that two weeks of data may be sufficient for data collection at current penetration rates.

keywords: *traffic signals, connected vehicles, vehicle trajectories, offset optimization*

INTRODUCTION

Connected vehicle (CV) data is expected to transform traffic operations by enabling numerous applications, including enhanced methods of traffic control (1,2). CV data is anticipated to include, among other information, vehicle position, speed, and heading. This would enable applications at intersections such as collision avoidance with red light violation warnings, and improved signal operation by providing suggested speeds to drivers, or by making use of enhanced information to improve signal timing.

One key factor that determines the degree of impact of potential CV applications is the level of market penetration, p , or the proportion of the vehicle fleet that is connected. Another important factor is the time scale of the application. Real-time control applications generally require a high level of penetration in order to achieve benefits. However, other applications that sample traffic patterns over longer periods of time can potentially work at lower levels of penetration. One example is the sampling of vehicle arrivals for evaluating and optimizing offsets in coordinated traffic signal control. In this application, if the arrival patterns are cyclic and remain consistent over a time period, they can be sampled with a low rate of p given a long enough sample period. A previous proof-of-concept study (3) investigated the impact of market penetration and sample period on the feasibility of offset optimization, finding that $p = 1\%$ was sufficient for a 3-hour sample period, with lower values of p potentially feasible if multiple days of data can be layered. The ability to aggregate data from multiple days depends on the consistency of traffic across time periods.

A near-term application for sampling arrival profiles would be to optimize offsets on corridors where no detection is available to measure arriving flow profiles. At present, there are a considerable number of coordinated systems where this is the case. Many central business district (CBD) areas still operate under fixed-time plans and little or no detection is used. Some corridors operate in a semi-actuated manner, with detection available on all movements *except* the coordinated movements. The optimization of offsets in these environments requires either modeling, which might not accurately reproduce field conditions, or manual adjustment in the field. CV data could potentially fill the role of detection in enhancing or automating offset optimization. If it is possible to do so at low levels of market penetration, as in the proof-of-concept study (3), this could be an early application of CV data which could potentially improve signal operations at present, while market levels rise to those needed for more advanced applications.

This paper investigates whether detector-free offset optimization is possible at present, using vehicle trajectory data obtained from private-sector data providers as a proxy for CV data obtained via DSRC or other means. To accomplish this, a “virtual detection” concept is proposed for determining individual arrival times from the private-sector trajectory data. The sampled arrival profiles are statistically compared against those measured with detectors. Finally, the sampled arrival profile data is used to drive an offset optimization process, and the arterial travel times are compared between a solution based on detector data, and one based on CV data.

VIRTUAL DETECTION WITH CONNECTED VEHICLE DATA

Overview of the Concept

“Advance” or “setback” detectors are located several seconds’ travel time upstream from the stop bar on approaches to signalized intersections. These are primarily used for added initial timing and dilemma zone protection by phase extension. They can also be used for measuring arrival profiles, which can be used to adjust signal offsets (4). The availability of such detectors varies considerably by agency and location.

Figure 1 shows a time-space diagram showing several vehicle trajectories on the approach to a signalized intersection. The chart shows the location of an advance detector relative to the stop bar, from which an arrival profile can be determined, as shown at the bottom of the chart. Each record in the arrival profile corresponds to a vehicle detection time.

The vehicle stream contains both connected and non-connected vehicles. In this example, there is one CV out of the seven total vehicles whose trajectories are illustrated. The detector enables the arrival time of each individual vehicle to be directly measured from the time when those vehicles physically pass over the detection zone. If that detector were absent, it normally would not be possible to make such a measurement. The single CV, however, reports its position periodically; Figure 1 shows it having a fairly short reporting interval. Although the CV’s reported locations do not necessarily align with the detection zone, it is possible to interpolate between those timestamped positions to estimate when the CV was likely to have crossed the detector. Further, although the arrival profile determined from CVs does not provide a lot of information about the vehicle fleet during one cycle, aggregating data from many cycles would enable a profile estimating the overall profile shape to be developed.

Figure 1. Comparison of vehicle arrival profile sampling with a physical detector and with connected vehicle (CV) trajectory data (3).

Implementation

To implement the concept, a suitable data source is needed. Many mobile devices, including most smartphones, have the ability to determine their location using GPS signals. Several vendors have developed analytical products based on the analysis of that data, such as minute-by-minute segment speeds. The *Urban Mobility Report*, for example, is based on the analysis of such speed data (5). Some researchers have begun to use this type of data for more detailed analysis of signal performance. A study by the Bavarian road Administration (6) used combination of geofencing and map-matching to measure delays on movements through 2300 traffic signals, while Argote-Cabañero *et al.* (7) measured delay and several other MOEs by sampling trajectories from the Next Generation Simulation dataset.

For this study, probe vehicle trajectories were obtained from a private sector vendor. The data consisted of probe vehicle trajectories observed in the state of Indiana between May 1 and June 11, 2016. The data were organized into individual “trips” for which a single identifier was maintained from the beginning and end of travel from an origin to a destination point. The reporting intervals varied from second-by-second timestamped positions, to one every few minutes. Approximately 20,000,000 timestamped vehicle positions per day were available in the statewide data set.

Figure 2a explains how virtual detections were implemented with the trajectory data (8). First, a “detection line” was identified by using two points, D_L and D_R , which represented the left and right sides of a line perpendicular to the roadway and stretching from the centerline to the edge of the right-of-way. Virtual detections were identified when the splines between two consecutive timestamped positions (shown in Figure 2a as P_i and P_{i+1}) intersected this line. The point of intersection is shown as P_X .

Because of variations in the spatial resolution of the individual timestamped locations, it was possible that some intersecting splines would represent travel in the opposite direction. To establish the direction of travel of a spline, a polygon was made by joining points P_i , D_R , P_{i+1} , and P_i . If these points defined a relatively small triangular region, as shown in Figure 2b, the direction of travel for the spline was the same as for the roadway. However, if these points defined a very large region (*i.e.*, approximately the entire surface of the Earth), as shown in Figure 2c, then the direction of travel was opposite to that of the roadway.

The time of intersection (T_X), or the virtual detection time, was determined by interpolating between P_i and P_{i+1} , as follows:

$$T_X = T_i + \frac{d_X}{v}, \quad \text{Equation 1}$$

where T_i is the time when the vehicle was reported at position P_i , and v is the vehicle speed,

$$v = \frac{d}{T_{i+1} - T_i}, \quad \text{Equation 2}$$

where d is the Euclidean distance between P_i and P_{i+1} .

(a) Virtual detection with trajectory splines.

(b) Enclosed area for a spline in the correct direction.

(c) Enclosed area for a spline in the opposing direction.

Figure 2. Extracting virtual detections from vehicle trajectory data.

The intersecting lines were subject to additional filtering to eliminate the inclusion of erroneous splines caused by CVs with long reporting intervals. A filter of 30 seconds was applied for this purpose. The purpose of this filter is illustrated more clearly by example in the next section.

To construct a cyclic arrival profile, the individual detection timestamps were transformed as follows:

$$T_C = (T_x) \bmod C, \quad \text{Equation 3}$$

where T_C is the time in cycle of the arrival and C is the cycle length. The arrival profiles were created by counting the number of vehicles per one-second bin, from 0 to $C - 1$ (9). In this study, all of the controllers synced their clock times using a common Network Time Protocol (NTP) time server, and no additional corrections were applied to accommodate differences between phase times and trajectory data times.

APPLICATION TO FIELD DATA

Two corridors in the greater Indianapolis area were chosen to test the virtual detection concept. Both of these featured advance detection (inductive loops) located approximately 5 seconds upstream from the stop bar on the coordinated movements, which could be used to evaluate the CV arrival profiles for accuracy.

- SR 37 (Figure 3a) is a high speed arterial passing through rural and suburban areas. It is a four-lane divided highway between intersections L and F, and has six lanes between F and A.
- US 36, Pendleton Pike (Figure 3b) passes through urban and suburban areas. It has six lanes between intersections M and S, and four lanes from T through Y.

These two corridors mainly serve commuter traffic into and out of Indianapolis, and tend to have consistent patterns from one day to the next during a typical week. There were no disruptive events during the study period. Detectors are generally well maintained on the corridors, but there were a few individual approaches where detectors were not working during this study.

Figure 4 illustrates the use of the virtual detection technique and the 30-second reporting interval filter, for intersections M through Q on Pendleton Pike. Figure 4a shows all of the trajectory splines from a 48-hour period. This represents about 36 million splines distributed across the state of Indiana, which came from about 460,000 individual trips. The map is almost entirely covered by this collection of lines. However, the locations of the major roadways are visible, as though they had been roughly scribbled over the map. Notice that the heaviest saturation of lines coincides with the location of the Interstate highway.

Figure 4b shows only the splines that intersect the virtual detection lines along the desired direction of travel. This has eliminated many, but not all, of the trajectories that are clearly from travel on other roadways. There are still many splines that clearly seem to originate from travel along the Interstate highway, as well as a few from crossing streets. Also note that many of the splines are very long, with many reaching entirely across the map. Figure 4c shows the effect of the 30-second filter. Now the splines are almost entirely limited to travel along the roadway of

interest, and the longer ones have been excluded; this also reduced the number of intersecting splines from 3,075 to 1,141, or about two-thirds.

(a) SR 37 South.

(b) Pendleton Pike (US 36).

Figure 3. Maps of the corridor locations.

(a) All trajectory splines from a 48-hour period (35,884,046 lines across the state).

(b) Splines that intersect the virtual detectors (3,075 lines at 10 detectors).

(c) Intersecting splines after application of a 30-second filter (1,141 splines at 10 detectors).

Figure 4. Example of how virtual detections are extracted from vehicle trajectories.
Data shown for the westernmost five intersections on Pendleton Pike for May 1 and May 2, 2016.

DETERMINATION OF ARRIVAL PROFILES

Profile Construction from Trajectory Data

The proof-of-concept study (3) speculated that a market penetration of 1% would be sufficient to conduct a measurement for a 3-hour sample period. The unfiltered trajectory data likely approaches this number in terms of overall penetration, but when filtered to include only those trajectories from which useful arrival times can be extracted, the penetration rates fall to values lower than 1% for many approaches. However, layering of multiple days, thereby extending the sample period, was demonstrated to further reduce the penetration needed to capture an arrival profile (3).

Figure 5 shows a comparison of the concept using a Purdue Coordination Diagram (PCD) (4). This graphic shows the arrival time for every cycle in a 24-hour period, and its relationship to the signal phase. The green shaded region represents the green interval; arrivals aligned with this region show vehicles arriving on red, while arrivals below the green line occur during the red interval. The vertical blue lines indicate time-of-day pattern changes.

- Figure 5a shows a PCD based on data from a physical detector. The data shows very good progression during the AM peak, with most of the arrivals aligned with the green interval. The midday and PM plans show many vehicles arriving in red.
- Figure 5b shows a PCD where the detector data has been replaced by virtual detections, from one 24-hour period. There is a very sparse number of points distributed across the day, which are not sufficient enough to get a sense of the overall arrival patterns.
- Figure 5c shows a PCD with six weeks of virtual detections (from May 1 through June 11, 2016), overlaid onto a representative 24-hour period of phase times. The density of points is much higher. At this level it becomes possible to visually distinguish the presence of platoons. They are visually quite similar to those seen in Figure 5a.

The layering concept is illustrated using a cyclic arrival profile in Figure 6; here, the profile grows as the sample period is expanded from one day to six weeks. Similar to the PCD, a one-day period (Figure 6a) does not yield enough data points to begin to confidently characterize the arrival profile. However, with one week of data (Figure 6b), the overall shape generally begins to take form. As each additional week is added, the number of observations increases linearly, and the profile is continually built up.

One consideration worth mentioning is that the identity of each vehicle is not known from one day to the next. The data only goes as far as to give a general category of each vehicle type. It is impossible to determine whether the same vehicles are being sampled multiple times across different days of the week, or not. This methodology assumes that the data represents a random sampling of overall traffic. Further investigation of the data for potential bias would be helpful to validate whether such an assumption is true, and whether the sampled profiles are sensitive to such bias.

(a) Physical detector data.

(b) One day of virtual detections.

(c) Six weeks of virtual detections.

Figure 5. View of Purdue Coordination Diagrams showing real and virtual detections. Data shown for the inbound approach at Intersection Q.

Figure 6. Building an arrival profile by increasing the duration of the sample period. Data is shown for the inbound approach at intersection Q.

Statistical Comparison

To make the profile more useful for optimization, it is smoothed using a 5-second moving average, and normalized so that its magnitude matches the expected volume on the approach. The 5-second value was selected from qualitative comparison of the resulting profiles. There was a substantial difference between raw data, 3-second, and 5-second moving averages, while increasing the moving average beyond 5 seconds did not produce much further change in the profile shapes. Figure 7 compares a normalized physical detector profile to the normalized, smoothed virtual detector profile for the same approach. The physical detector data (Figure 7a) is extracted from one representative week, while the virtual detector data (Figure 7b) was developed from a six-week period. These are also visually similar, with the platoon indicated by the bulk of the distribution ranging from about 80 seconds in cycle through 20 seconds in the cycle (wrapping around the cycle boundary).

Figure 7c compares the two cumulative cyclic distributions of the two profiles. The maximum absolute difference between the two cumulative distributions is highlighted; this value, called the *D*-statistic, can be used to evaluate the statistical goodness-of-fit between the two cyclic distributions using a Kolmogorov-Smirnov (KS) test. For the example shown here, the *D*-statistic value is 0.046. For 100 bins (representing 1-second bins and a cycle length of 100 second), this represents a significant result at a confidence level of 90%.

Table 1 shows the *D*-statistics for all approaches at the 25 intersections where a comparison could be made. Those where the statistical test yields a significant result at the 90% confidence level are noted. This represents a majority of the approaches on the two corridors. Intersections A, B, and O had relatively poor fits for both directions, perhaps due to their proximity with the Interstate highway, which would seem to increase the likelihood that erroneous detections could be picked up by the detection lines, even after filtering. More comprehensive map-matching of the trajectory points to roadway segments would likely improve the agreement between the field loop detectors and virtual detectors at intersections A, B and O.

The market penetration of each approach is also shown. The penetration rate is found by comparing the number of filtered intersecting splines with the volume measured by the vehicle detectors at each location. The penetration rates are under 1% for all of the approaches in the system, with a few that are lower than 0.1%.

(a) Physical detector profile, tabulated over a one week period.

(b) Virtual detector profile, tabulated over a six week period.

(c) Comparison of cumulative frequency distributions.

**Figure 7. Comparison of arrival profiles from real and virtual detectors.
Data is shown for the inbound approach at Intersection Q.**

Table 1. Market penetration rate p and Kolmogorov-Smirnov D -statistics from comparing physical and virtual detector arrival profiles. Data shown for each approach during the midday time period. Bold entries indicate where the statistical match was significant at the 90% level.

System	Intersection	Inbound		Outbound	
		p	D -statistic	p	D -statistic
SR 37	A. Pilot	0.42%	0.237	0.34%	0.203
	B. I-465 WB Ramp	0.80%	0.099	0.53%	0.224
	C. I-465 EB Ramp	0.56%	0.077	0.48%	0.187
	D. Thompson	0.26%	0.081	0.36%	0.070
	E. Harding	0.27%	0.040	^b N/A	
	F. Epler	0.15%	0.145	0.27%	0.062
	G. Banta	0.08%	0.088	0.32%	0.040
	H. Southport	0.22%	0.073	0.39%	0.020
	I. Wicker	0.10%	0.074	0.25%	0.071
	J. County Line	0.33%	0.073	0.34%	0.056
	K. Fairview	0.31%	0.051	0.35%	0.077
	L. Smith Valley	0.36%	0.142	0.32%	0.063
	M. SR 144	0.39%	0.040	0.25%	0.025
Pendleton	N. I-465 SB Ramp	0.25%	0.072	0.09%	0.062
	O. I-465 NB Ramp	0.18%	0.139	0.19%	0.122
	P. 42nd St	0.26%	0.076	0.22%	0.095
	Q. Franklin	0.20%	0.056	0.18%	0.046
	R. Esquire	0.23%	0.055	0.17%	0.036
	S. Post	0.31%	0.042	^c N/A	
	T. MBC	0.16%	0.152	0.09%	0.171
	U. Mitthoeffer	0.13%	0.151	0.20%	0.062
	V. 56th St	0.10%	0.118	0.18%	0.046
	W. Walmart	0.15%	0.072	0.18%	0.048
	X. Sunnyside	0.17%	0.046	0.22%	0.052
	Y. Oaklandon	^a N/A		0.14%	0.068

^a No detector data available at this approach.

^b This approach is unsignalized.

^c Some lanes had broken detectors, so this approach is excluded from the comparison.

OFFSET OPTIMIZATION

Although not every approach yielded a statistically significant goodness-of-fit result when the virtual detector profiles were compared to the physical detector profiles, this does not necessarily mean that the data is not useful for optimization. The proof-of-concept study (3) showed that good optimization results could still be achieved even though the estimated profiles did not fit the actual ones with high statistical confidence. As long as the part of the cycle where the bulk of the vehicles arrive is identified well by the distribution, small differences in its overall shape which may accumulate into large D -statistics may not impact the optimization results.

Two sets of optimized offsets were prepared using the Link Pivot algorithm (9), a formulation of the Combination Method (10) for arterial highways.

- Physical detector offsets were determined by running the optimizer with arrival profiles based on physical detectors, and green times logged by the controllers. One week of data was aggregated on each corridor to support the optimization.
- Virtual detector offsets were determined by running the optimizer with arrival profile based on trajectory data, and green times logged by the controllers. The same green time distributions were used as for the physical detector offsets.

Each corridor ran the physical detector offsets and the virtual detector offsets for one week each. Travel time data was extracted from the analysis of minute-by-minute segment speeds obtained from the same data vendor that provided the trajectory data. This data has been used for arterial travel time analysis in previous studies (11,12). The same methodology as explained by Remias *et al.* (11) was applied in the present study, with the main difference being that the smaller “XD” segment definitions were used rather than the longer Traffic Message Channel (TMC) definitions.

Figure 8 presents the results of that exercise. The figure is organized to show two directions on each corridor for the midday time period, with each chart containing a CFD of the travel times with the existing offsets, with the physical detector offsets, and with the virtual detector offsets.

- On inbound SR 37, the physical detector offsets yielded an approximately 1-minute reduction in travel time at the 50th and 75th percentiles (Figure 8a). The virtual detector offsets also yielded an improvement over the existing offsets, although it was not as great as with the physical detector offsets. The travel times were also more reliable, as shown by the more vertical slopes of the curves.
- Outbound SR 37 (Figure 8b) had virtually no change in travel time between the three data sets.
- On inbound Pendleton Pike (Figure 8c), there was almost no difference between the existing offsets and the physical detector offsets. However, the virtual detector offsets yielded a reduction of approximately half a minute at the 25th percentile.
- On outbound Pendleton Pike (Figure 8d), both the physical detector offsets and the virtual detector offsets yielded reductions in travel time of approximately 1 minute at the 25th percentile and slightly less than a minute at the 50th percentile. Above the 75th percentile, the CFD for the physical detector offsets exhibits a long tail, indicating that there were some occurrences under physical detector offsets where the travel times became longer than before.

Figure 8. Travel time CFDs.

Overall, the travel time results show that the virtual detector offsets are capable of providing results that are similar to physical detector offsets. On SR 37, the physical detector offsets outperformed the virtual detector offsets, but the virtual detector offsets still yielded a worthwhile level of improvement. A travel time reduction and improvement in reliability would have been welcome, for example, if analysis from physical detection was not possible on the corridor. This was the result that was anticipated, given that the physical detectors provide more complete information.

In contrast, on Pendleton Pike, the virtual detector offsets actually yielded results that were a little better than the physical detector offsets. This may be attributable to a few factors. One is that a few of the approaches on Pendleton Pike had poorly performing detection. As seen in Table 1, the outbound approach at intersection S only had one working detector, so the arrival profiles captured by physical detection may have been poor at that location.

It is also possible that, in the more urban environment, with lower speeds (hence shorter distances of the advance detector from the stop bar) and longer queues, there are some approaches where the physical detection might not have captured the true arrival patterns. On the

other hand, the virtual detection, interpolating between reported vehicle movement across a detection line, rather than the rising edge of detector occupancy that creates a “detector on” event in the high-resolution data, might actually have better characterized the arrival patterns.

SHORTENING THE SAMPLE PERIOD

The results demonstrate that the virtual detection concept is feasible, and can provide reasonable improvements in arterial travel times, as shown by results from two corridors where offset improvements were made based on cyclic flow profiles generated from virtual detections. This study was based on an analysis of 6 weeks of data. This begs the question of whether such a long sample period is actually necessary to carry out such an analysis.

To answer that question, a sensitivity analysis of the goodness-of-fit with respect to the duration of the sample period was carried out. Figure 6 shows the raw arrival profiles for various sample periods ranging from one day to 6 weeks, for a particular approach. In Figure 9, these profiles have been smoothed, normalized, and converted into cumulative frequency diagrams, and compared with detector data aggregated over the exact same time periods. For each of these, the KS test was carried out, and the D -statistic is shown in the lower right corner. Note that the D -statistic for the 6-week result differs from that in Table 1 because the detector data is aggregated across all 6 weeks rather than one representative week (data was not available from all 25 intersections across the entire 6-week period to be able to do so for all approaches).

Although there is good visual agreement between the cumulative distributions for even one day of data, only those data sets with data from 2 weeks or longer pass a statistical goodness-of-fit test at a 90% confidence level. Note, however, that the D -statistic actually reaches its lowest value when 4 weeks of data are aggregated. Beyond this, the D -statistic starts to increase as more data is added—which indicates that there are differences in the arrival profiles which are accumulating over time. This demonstrates that there is an upper bound beyond which aggregating further data becomes counterproductive, because the traffic patterns changes associated with memorial day weekend introduce some additional noise into the data set.

This brief analysis shows that about 2 weeks of data may be an appropriate sample period for many corridors. However, the sample period could potentially be reduced if the market penetration of probe vehicle data increases. The arrival profiles that can be measured tomorrow are likely to be more accurate than those that could be measured yesterday. Also, the availability of true CV data from vehicles with DSRC-based or other such communication, which may come to incorporate a substantial amount of the overall vehicle fleet, would likely both increase the accuracy of the arrival profiles and decrease the duration of the sample period needed to characterize them—and indeed enable more active, real-time control schemes once the penetration rate is high enough. Privacy concerns may limit the ability of agencies to collect trajectory information by this means. If few vehicles decide to “opt in” to provide that data within the envisioned CV environment, data from the private sector might prove to be a valuable resource to obtain anonymized trajectories at a network level. The results shown in this paper indicate that early improvements are possible with conventional types of control using private sector data at low levels of market penetration.

Figure 9. Comparison of virtual detector data from different sampling intervals and detector data from the same date range. Each figure shows the KS test D -statistic; those marked with an asterisk show statistically significant goodness of fit at the 90% level. Data is shown for the inbound approach at Intersection Q.

CONCLUSIONS

This study examined the feasibility of offset optimization using connected vehicle (CV) data at low level of market penetration. Vehicle trajectory data was obtained from a private sector vendor. The splines of the trajectories underwent a geometric analysis to find the times when they would have crossed lines defined as arrival detectors. These were considered to be “virtual detections” from which arrival profiles were created. Data was collected for two corridors including 25 coordinated signalized intersections. After sampling and filtering, penetration rates between 0.09–0.80% were observed, varying by approach (Table 1).

The arrival profiles were analyzed using two tests. First, a statistical goodness-of-fit matching test was made between arrival profiles aggregated over 6 weeks, and detector data aggregated over a representative week, to determine whether the virtual detection arrival profiles closely matched those measured from detectors. The majority of approaches did exhibit statistical significant fit between the virtual and physical detector arrival profile distributions, using a Kolmogorov-Smirnov test at a 90% confidence level (Table 1). Some approaches did not have a good statistical fit, particularly those at intersections with freeway ramps.

Next, the arrival profiles were used to optimize the offsets. These were compared against the performance of offsets determined from the physical detector arrival profiles. Corridor travel times were obtained for the existing offsets, physical detector offsets, and virtual detector offsets. On one corridor, improvements were obtained from both physical and virtual detector offsets; the travel times were slightly lower with the physical detector offsets. On the second corridor, the virtual detector provided slightly lower travel times. The results demonstrate that virtual detection is feasible for optimizing offsets with current penetration rates of probe data.

Lastly, a sensitivity analysis of the sampling period was conducted to determine whether shorter intervals than 6 weeks can be used. The analysis showed that 2 weeks may be a sufficient sample period for current market penetration rates, and in fact that the quality of the data (in terms of statistical goodness-of-fit for arrival profiles) does not necessarily increase linearly as the sample period is extended, because of shifting in traffic patterns.

This study demonstrated that offsets can be optimized using probe vehicle data as a proxy for CV data. Future work in this area would include further refinement of the geometric analysis to better filter erroneous data (such as, potentially, vehicles traveling on freeway overpasses or underpasses near interchanges), and additional comparisons of match quality over different sample periods to better tune the data collection requirements. Another area of potential study would be consideration of multimodal extensions of the methodology. The present study makes the assumption that each data point represents a vehicle, but particularly in more urban environments, bicycles and pedestrians may be captured by probe data. The impact of different operating environment (e.g., urban versus rural) on the data characteristics would be worthwhile a topic of future research. Finally, the application to corridors that lack physical detection, such fixed-time operation in a central business district area, would be a potential application for the proposed methodology.

ACKNOWLEDGMENTS

This work was supported in part by the Joint Transportation Research Program and the Pooled Fund Study (TPF-5(258)) led by the Indiana Department of Transportation (INDOT) and supported by the state transportation agencies of California, Georgia, Kansas, Minnesota, Mississippi, New Hampshire, Pennsylvania, Texas, Utah, and Wisconsin, the Federal Highway Administration Arterial Management Program, and the Chicago Department of Transportation. Probe vehicle data was provided by INRIX. 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. Hill, C.J. and J.K. Garrett. *AASHTO Connected Vehicle Deployment Analysis*. Report FHWA-JPO-11-090. USDOT, 2011.
2. Kathis, J., E. Papapanagiotou, and F. Busch. "Traffic Signals in Connected Vehicle Environments: Chances, Challenges, and Examples for Future Traffic Signal Control." Proc., 18th IEEE Conference on Intelligent Transportation Systems, 2015.
3. Day, C.M. and D.M. Bullock. "Opportunities for detector-free signal optimization with limited connected vehicle market penetration: a proof-of-concept study." *Transportation Research Record*, in press, 2016.
4. Day, C.M., R. Haseman, H. Premachandra, T.M. Brennan, J.S. Wasson, J.R. Sturdevant, and D.M. Bullock. "Evaluation of Arterial Signal Coordination: Methodologies for Visualizing High-resolution Event Data and Measuring Travel Time." *Transportation Research Record No. 2192*, 2010, 37–49.
5. Schrank, D., Eisele, B., & Lomax, T. *2015 Urban Mobility Scorecard*. Texas A&M Transportation Institute, 2015.
6. Wünsch, G., F. Bölling, A. von Dobschütz, and P. Mieth. "Bavarian Road Administration uses Probe Data for Large-Scale Traffic Signal Evaluation Support." Presented at Transportation Research Board Annual Meeting, Paper No. 15-0697, 2015.
7. Argote-Cabañero, J., E. Christofa, and A. Skabardonis. "Connected Vehicle Penetration Rate for Estimation of Arterial Measures of Effectiveness." *Transportation Research Part C*, 60, 2015, 298–312.
8. Li, H., C.M. Day, and D.M. Bullock. "Virtual Detection at Intersections Using Connected Vehicle Trajectory Data." Presented at IEEE Intelligent Transportation Systems Conference, 2016, 2571–2576.
9. Day, C.M. and D.M. Bullock. "Computational Efficiency of Alternative Algorithms for Arterial Offset Optimization." *Transportation Research Record No. 2259*, 2011, 37–47.
10. Hillier, J.A. "Appendix to Glasgow's Experiment in Area Traffic Control." *Traffic Engineering and Control*, Vol. 7, 1965, 569–571.
11. Remias, S.M., T.M. Brennan, A.M. Hainen, C.M. Day, and D.M. Bullock. "Characterizing Urban Mobility and Travel Time Reliability along Signalized Corridors using Probe Data." Proceedings, International Scientific Conference on Mobility and Transport, Munich, Germany, 2013.
12. Hu, J., M.D. Fontaine, and J. Ma. "Quality of Private Sector Travel-Time Data on Arterials." *Journal of Transportation Engineering*, 142, 2016, 04016010.