2016

Opportunities for Detector-Free Signal Offset Optimization with Limited Connected Vehicle Market Penetration: A Proof-of-Concept Study

Christopher M. Day
Purdue University, cmday@purdue.edu

Darcy M. Bullock
Purdue, darcy@purdue.edu

Opportunities for Detector-Free Signal Offset Optimization with Limited Connected Vehicle Market Penetration: A Proof-of-Concept Study

Christopher M. Day*
Purdue University
550 Stadium Mall Drive
West Lafayette, IN 47906
(765) 496-9601
cmday@purdue.edu

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

*Corresponding author.

Word Count: 5083 words + 9 * 250 words/(figure-table) = 5220 + 2250 = **7470** words

November 6, 2015

ABSTRACT

Connected vehicle (CV) data has the potential to transform traffic signal operations, but the success of control methods based on CV data will depend on the level of market penetration. Recent studies of real-time operational strategies in CV environments suggest that penetrations exceeding 20% will be required. This study explores the feasibility of using CV data to generate arrival profiles for optimizing arterial progression. Applications to offline (3-hour analysis period) and online (15-minute analysis period) offset optimization are considered. Vehicle arrival profiles obtained from real world measurement are used as a basis for comparison. Subsampled distributions are used to estimate the potential distributions that might be obtained from CVs, and these are statistically analyzed to explore the impacts of penetration rate, analysis period, and traffic volume. For selected penetration rates ranging from 0.1% to 50%, the subsampled distributions are used to optimize the corridor, and the results are evaluated in the complete-data model. The results show that over a 3-hour window, successful offline optimization can be achieved with a CV penetration rate as low as 1%. Layering multiple days of data could potentially allow offline optimization with penetration rates as low as 0.1%. Online optimization with 15-minute windows require somewhat higher penetration rates of at least 5%. The results suggest that early applications of CV data may be possible at very low levels of market penetration. In corridors with high penetration of connected mobile devices, some private sector probe data services may be at the cusp of providing the necessary data to facilitate detector-free optimization.

INTRODUCTION

Information is essential for traffic control systems to respond to changing demands. For example, actuation profoundly changes how signalized intersections operate, and adaptive control is enabled by further increasing the amount of available information. Historically, the only way to obtain more information has been to install more detectors. However, like all infrastructure, a large detector inventory can be expensive to deploy and maintain. Setback (or "advance") detectors, located well upstream of the stop bar, are costly because of the need to connect distant equipment to the local intersection.

Connected vehicles (CVs), which can send and receive information between each other and with traffic infrastructure (*1*), will very likely transform traffic control by providing increasingly detailed information. The possibilities of CV data for signal control have been explored by several researchers recently (*2*–*8*). The outcomes of these studies are summarized in Table 1. This selection of papers includes studies that focus on the influence of the market penetration rate, p. This is the proportion of the vehicle fleet consisting of CVs. As p increases, more benefit is generally attained from the CV-assisted algorithm. A critical value of p may be defined as the lower bound for achieving a benefit. A survey of previous studies puts this lower bound at approximately 20% or so. Methods that specifically incorporate estimates of the non-CV states can reduce this to as low as 10%. Thus, for real time applications, relatively high values of p are needed to begin seeing benefits.

Table 1. A summary of selected literature on signalized intersection control using connected vehicles.

Study	Synopsis	Critical Value of p	Range of p Considered
Priemer and Friedrich (*2*)	Real-time control on a 20-second horizon, compared to TRANSYT-7F actuated-coordinated control	17%[a]	10–100%
He *et al*. (*3*)	Multimodal real-time control, compared to free, actuated-coordinated, and transit signal priority	40%[a]	20–100%
Lee *et al*. (*4*)	Real-time control minimizing cumulative travel time compared to actuated control	30%[a]	10–100%
Goodall *et al*. (*5*)	Real-time control over a 15-second horizon compared to Synchro actuated-coordinated control	25%[b]	10–100%
Goodall *et al*. (*6*)	Real-time control assisted with estimation of unequipped vehicle information	10–25%[a]	5–100%
Feng *et al*. (*7*)	Real-time control assisted with estimation of unequipped vehicle information	25–50%[c]	25–100%
Argote-Cabañero *et al*. (*8*)	Estimation of arterial measures of effectiveness over a 15-minute analysis period	15%[d]	0.01–100%

[a] the critical value identified by the researchers.
[b] lower bound based on delay reductions first occurring at $p = 25\%$ (see Tables 1 and 2 in the referenced paper).
[c] lower bound based on delay increases for $p = 25\%$ and decreases for $p = 50\%$ (see Table 4 in the referenced paper).
[d] taken from the authors' concluding statement regarding this value as a general threshold.

It is difficult to estimate what year any particular value of *p* might be achieved. A 2011 study (*9*), assuming that onboard dedicated short range communication (DSRC) equipment (*10*) would be mandated by 2019, projected that 30% market penetration would be achieved by 2024. If the proliferation of smartphones is a good predictor, CVs may see rapid deployment. However, a substantial amount of equipment upgrades would still be needed to facilitate communications between vehicles and infrastructure, including approximately $350 million in controller upgrades alone (*9*). It seems pragmatic to expect that another decade or more will pass before roadways are ready to launch real-time control schemes based on CV data. However, there may be other opportunities in the shorter term that could be achieved at lower values of *p*, particularly if the "connected vehicle" concept is expanded to include data obtained from mobile phone and vehicle navigation systems. Currently, about 64% of adults in the US own a smartphone (*11*); thus, many vehicles are already carrying devices that report their position over time.

This paper explores how CV data at very low values of *p* could potentially be used to facilitate offline optimization and online adaptive adjustment of coordinated signal timing parameters. Rather than focusing on a horizon of less than one minute, horizons of 3 hours (offline) or 15 minutes (online) are used to investigate the level of *p* needed before the system can make a beneficial control decision. The emphasis of this paper is on the lower ranges of *p* corresponding to the early adoption phase of CV deployment, with values as low as $p = 0.1\%$ considered. The proof of concept presented here uses the cyclic, phase-based paradigm currently widespread in signal control. However, the findings from exploring the sampling problem will likely manifest in other paradigms as well.

REPLACING DETECTION WITH CV DATA

A starting point for exploring how CV data could replace detectors is to understand what types of information detectors collect. Figure 1 presents a series of data visualizations that represent similar types of information at differing levels of aggregation.

Figure 1a presents a time-space diagram showing vehicle arrivals on a particular signalized approach. Here, the stop bar is located at 3000 ft and a setback detector is located at 2600 ft. This is a typical location for dilemma zone protection in a 55 mph zone (at 81 fps, a detector would be located 405 ft upstream to provide a 5-second extension). The intersection operates with a cycle length of 114 seconds; the red bars at 3000 ft indicate when the signal is red. The black lines represent vehicle trajectories. The red and green diagonal lines representing the end of green (EOG) and beginning of green (BOG) project the signal state backward, delineating the range of detection events relevant to the cycle between 114 and 228 seconds. Each dot located at the intersection of a vehicle trajectory and the detector location at 2600 ft represents a detection. The graphic just above the time space diagram represents these detection events relative to the signal state. This represents the data provided by setback detection. It implies the pattern of arrivals and the likely trajectories that occurred.

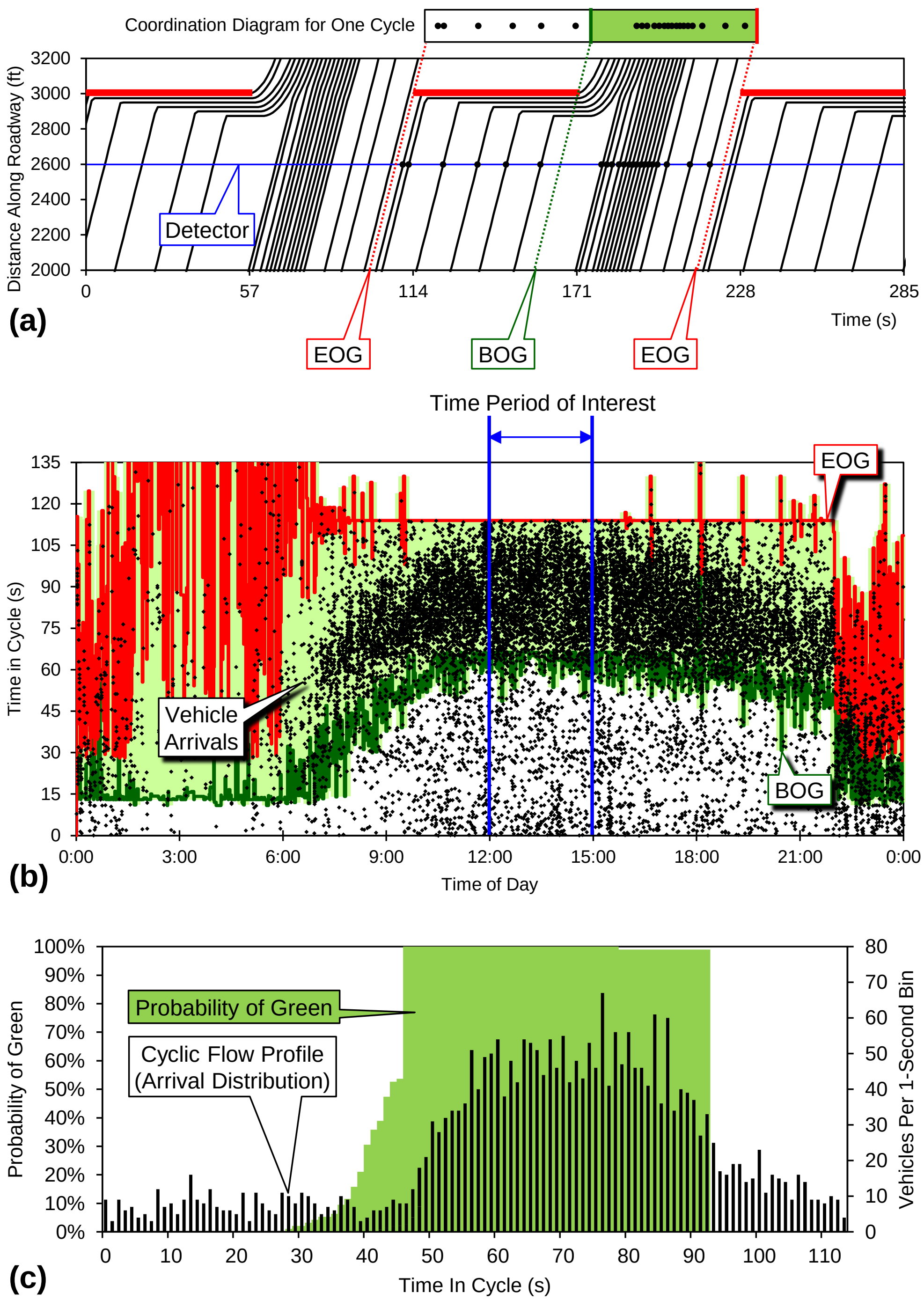


Figure 1. Three graphical representations of vehicle arrival and phase status data based on conventional (complete) detection:
(a) a time-space diagram, and a coordination diagram for one cycle;
(b) a Purdue Coordination Diagram (PCD) covering 24 hours;
(c) a cyclic flow profile showing arrival and probability of green distributions.

This summary cyclic visualization, when repeated for multiple cycles (*12*), was introduced as the "Purdue Coordination Diagram" (PCD). It has been extensively used to evaluate progression in Indiana (*13*), Utah (*14,15*), and elsewhere. Figure 1b shows a PCD covering a 24-hour period. Each column in the data represents one cycle. The time in cycle flows vertically. The horizontal axis represents the previous EOG; the green line shows the BOG during each cycle, and the upper red line shows the EOG and the end of the cycle. In each cycle, the data timeline is truncated at the upper EOG and continues in the next column to the right. The dots represent vehicle detections at the setback detector. Between 6:00 and 22:00, the intersection is operated with a cycle length of 114 seconds. The duration of green and the arrival times of platoons vary throughout the day as a consequence of early return to green when side-street demand is low. During the overnight period (22:00–6:00), the signal sometimes cycles more rapidly, while at other times it dwells in green on the coordinated movement. This chart visually summarizes the relevant information to assess the quality of progression; in this case, during most of the day, most vehicle arrivals are clustered between the BOG and EOG, so progression is favorable.

When an intersection is operated with cyclic coordination, cyclic flow profiles can be developed (*16*). Figure 1c shows an example. For example, at 60 seconds into the cycle, about 50 vehicles arrived during that 1-second bin, and the probability of green was 100%, looking across the entire 3-hour analysis period. Most of the vehicle arrival distribution coincides with a high probability of green, indicating favorable progression. The distribution corresponds to the highlighted region between 12:00–15:00 in Figure 1b.

These data representations enable coordinated signal timing parameters, particularly offsets, to be optimized. By predicting how the distributions will change under a given adjustment, it is possible to systematically optimize all of the offsets in the system (*18*). It is possible to consider "online" and "offline" methods of optimization:

- *Offline* methods use data from longer time periods, and adjust the static offset that is called into service when the corresponding time of day plan comes into service. While many offline methods use model-based estimations of roadway traffic, methods that incorporate measured arrivals have proven to be quite effective (*16,17,18,19*).
- *Online* methods use data from shorter time periods, and dynamically adjust offsets to respond to developing conditions. Such methods are used in adaptive control systems such as SCOOT (*20*) and ACS-Lite (*21*).

Both rely upon the availability of working detection to measure arrivals and determine an appropriate adjustment. If detection systems malfunction, the control system also fails; and detection systems have associated costs for installment and maintenance. If a sufficient sampling of the arriving traffic could be deduced from CV data, the arrival distributions could be developed with no detection at all—meaning that advanced control could be made possible with little to no infrastructure investment.

Figure 2 shows how a detector-free measurement might be facilitated by CV data. This shows a time space diagram of several vehicles on a signalized approach. In this example, one of the vehicles in the stream is a connected vehicle. The true trajectory of the vehicle is reported in the CV data, with a timestamped position being recorded at some reporting interval (which might be regular or irregular). The detector data yields an arrival profile for the example timeline; with only one CV in this vehicle stream, the CV data would only show one vehicle while there were actually seven in total. It is not difficult to imagine that some cycles might contain no CVs at all.

However, if the control scheme remains consistent over a long enough time period, as in the situation illustrated by the PCD in Figure 1b, the arrival distribution could be estimated from a sampled measurement of the vehicles. If the CVs are randomly distributed in the vehicle stream, then this becomes a sampling problem: find the critical value of p needed for a useful estimate of the arrival distribution, over analysis period T. In signal control, the maximum feasible duration T is determined by the use case. For online applications, the upper bound of T would likely be several minutes, whereas for offline applications, multiple hours would be acceptable. For offline applications, multiple days could even be layered if traffic patterns are consistent, similar to how sparse travel time data may be combined to yield a representative distribution (*22*).

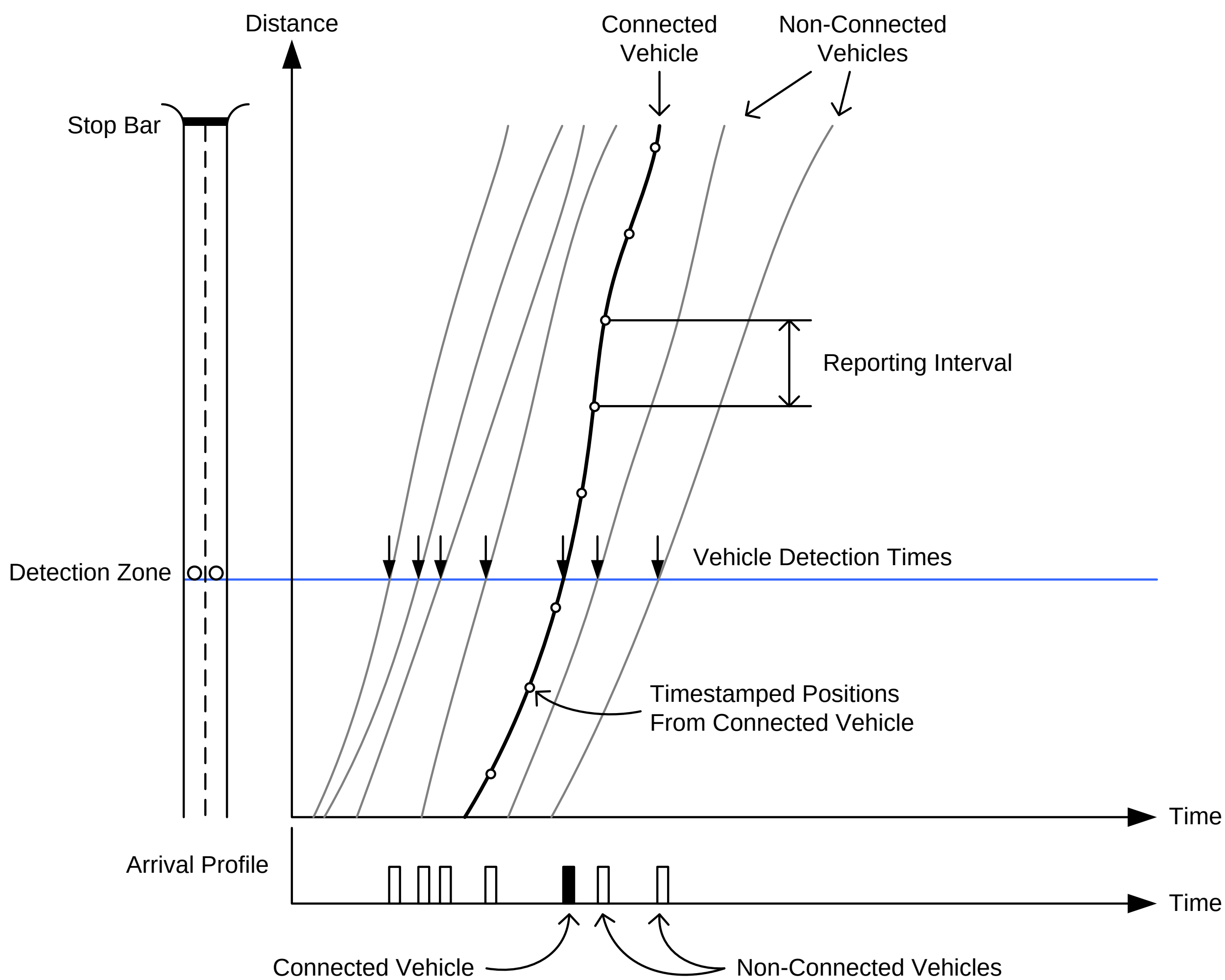


Figure 2. Traditional vehicle arrival measurement with detection, and opportunities for detector-free measurement using connected vehicle data.

ESTIMATING THE IMPACT OF MARKET PENETRATION

Hypothetical CV Data from Sampled Detector Data

To investigate the impact of the market penetration rate *p*, this study considers how vehicle arrival profiles could potentially be estimated with CV data, and then investigates the use of those for optimizing the signal offsets. SR 37, a nine-intersection arterial north of Indianapolis, Indiana, served as the test corridor for this study. This location was previously used for several offset-optimization studies (*12*,*23*). Each intersection is capable of logging phase and detector status using high resolution event data (*24*). Complete field data from Saturday, June 20, 2015 was used for this study.

The complete cyclic flow profile data from the physical detectors, as represented by the flow profile in Figure 1c, was used to represent the scenario where p = 100%. These are measured for analysis periods (T) of various durations. To model the effects of substituting CV data for detector data, a subset of the detector data was constructed by random selection of the detection events to mimic the distribution of CVs in the vehicle stream, using a likelihood of selection equal to p. This yields a sampled arrival profile for given values of T and p.

Figure 3 presents an example of how the sampled flow profiles degrade as p decreases from 100% to 0.1%. The left column shows data for T = 3 h while the right column shows T = 15 min.

- Figure 3a and Figure 3b respectively show the probability of green profiles for T = 3 h and T = 15 min. This study assumes that the green times can be measured without loss of fidelity.
- Figure 3c and Figure 3d respectively show the complete (p = 100%) data, as measured by the actual physical detector. The same profile should be duplicated without a detector, if each vehicle in the traffic stream is a CV. This study ignores the possibility of spatial or temporal variation in the CV data.
- At p = 10%, the 3 h profile (Figure 3e) is still similar to the p = 100% profile (Figure 3c). This is less true of the 15 min profile (Figure 3f).
- At p = 1%, the 3 h profile has a lower fidelity (Figure 3g), yet the vehicles are still clustered in the same part of the cycle as the p = 100% distribution. The 15 min profile (Figure 3h) has only two vehicles.
- At p = 0.1%, the 3 h profile shows only one vehicle (Figure 3i) while the 15 min profile has none at all (Figure 3j).

The bottom row shows the normalized cumulative arrivals for all of the above arrival histograms overlaid in one chart. For the 3 h distribution (Figure 3k), the p = 100% and p = 10% lines are very similar; the p = 1% line has the same general shape but is somewhat displaced from the others. The p = 0.1% takes its shape from the single observed vehicle (Figure 3i). The degradation in the 15 min data is more severe. The degradation in the 15 min data (Figure 3l) is more severe. Here, the p = 10% line substantially diverges from p = 100%, while the p = 1% line is very different. The p = 0.1% line cannot be plotted because there were no observations.

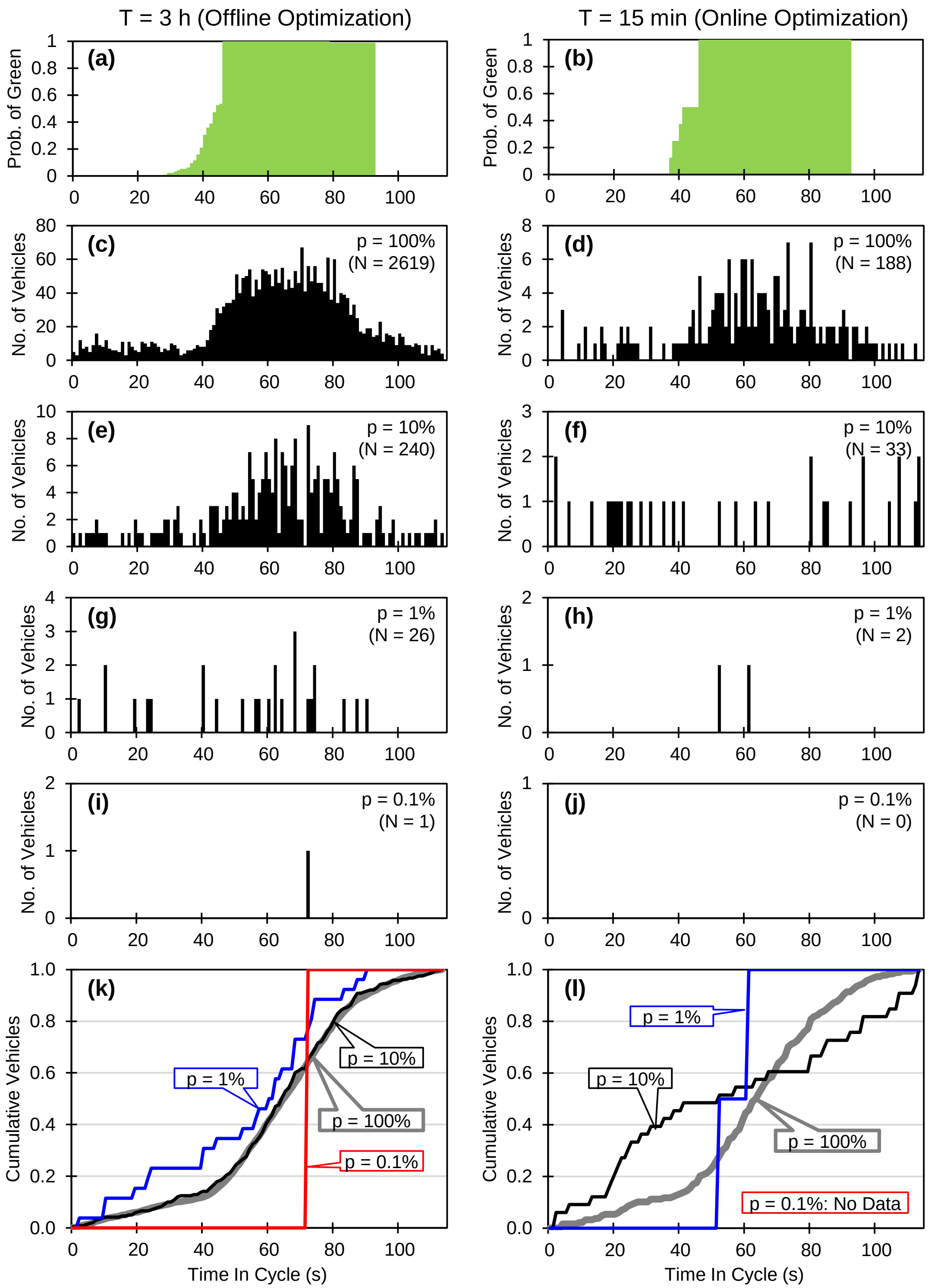


Figure 3. Measured cyclic profiles. Data shown for Northbound at SR 37 and Town and Country Blvd:
(a) distribution of green for $T = 3$ h; (b) distribution of green for $T = 15$ min;
(c) vehicle arrival distribution for $T = 3$ h, $p = 100\%$; (d) vehicle arrival distribution for $T = 15$ min, $p = 100\%$;
(e) vehicle arrival distribution for $T = 3$ h, $p = 10\%$; (f) vehicle arrival distribution for $T = 15$ min, $p = 10\%$;
(g) vehicle arrival distribution for $T = 3$ h, $p = 1\%$; (h) vehicle arrival distribution for $T = 15$ min, $p = 1$;
(i) vehicle arrival distribution for $T = 3$ h, $p = 0.1\%$; (j) vehicle arrival distribution for $T = 15$ min, $p = 0.1\%$;
(k) cumulative profiles for $T = 3$ h, for various p; (l) cumulative profiles for $T = 15$ min, for various p.

Impact of Market Penetration and Analysis Period

The example data in Figure 3 represents one particular random sample for selected values. To further explore the interaction of *T* and *p*, the process was repeated for a range of *T* varying from one signal cycle (114 seconds) to 5 hours, and for *p* ranging from 0.1% to 75%. Because the random sampling process yields a different subset on each iteration, the sampling procedure was repeated 100 times for each combination of *T* and *p*. These were then compared to the complete data set (*p* = 100%) using the Kolmogorov-Smirnov (KS) test to compare cumulative distributions.

The average *P*-values for the 100 KS tests are shown for different combinations of *T* and *p* in Figure 4. The vertical lines indicate the values of *T* that were tested, and the labels indicate the values of *p* that were used for each. As expected, the higher the value of *p*, the smaller the *T* needed to achieve a similar degree of confidence in the estimated profile (i.e., same *P*-value). For real-time applications, small values of *T* would be required. For these, high market penetrations are needed, which agrees with previous studies. For other applications where estimates can be aggregated over a longer observation period, lower market penetrations suffice. For example, for two cycles, *p* = 75% yields a P-value above 0.95, while for 3 hours, similar confidence can be achieved with *p* = 5%.

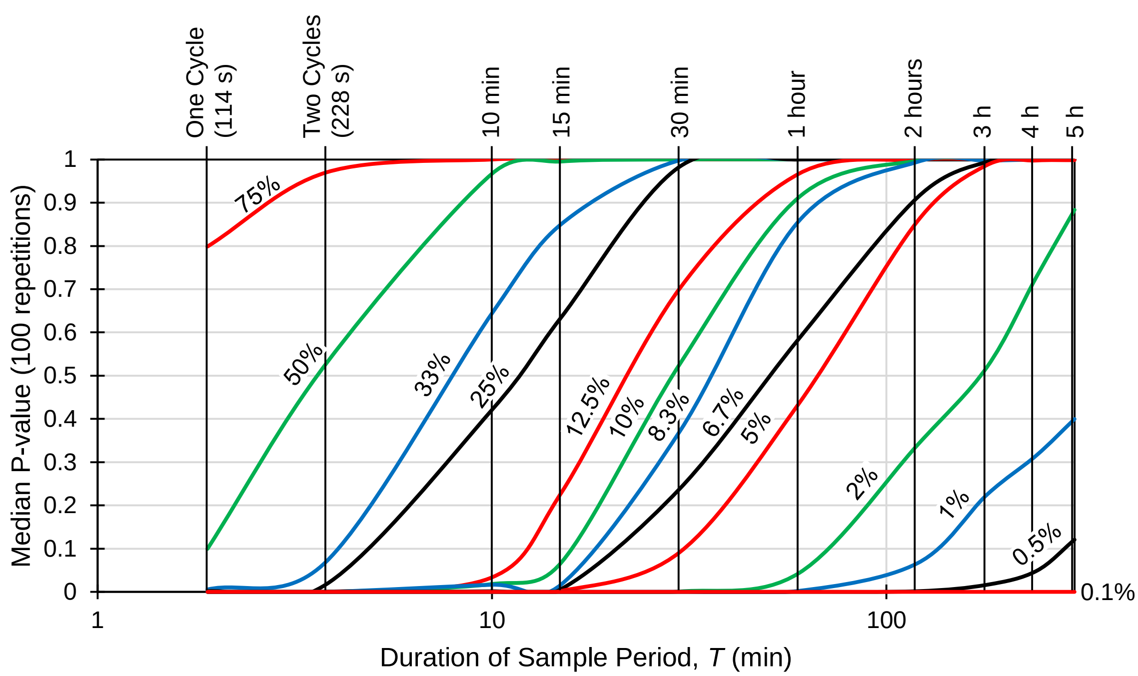
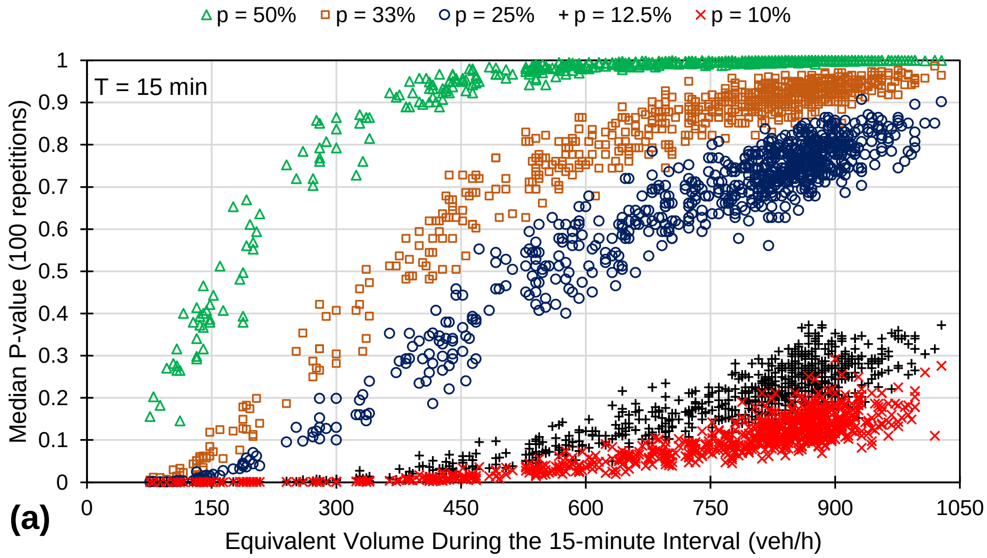


Figure 4. Relationship between analysis period *T* and the quality of arrival profile estimation (average P-value from KS tests for 100 separate random samples) at various values of market penetrations, *p*. The percentage displayed on each curve shows the value of *p*.

Impact of Traffic Volume

To investigate the impact of traffic volume, KS tests were repeated for selected values of *p* for every possible interval starting from the beginning of every minute from 6:00 to 19:00 during Saturday, June 20, 2015. This was done for T = 15 min and T = 3 h. For every interval, KS tests for 100 random samples were conducted. Figure 5 shows plots of the average P-values of those tests against the equivalent hourly volume. The range of volumes is smaller in Figure 5b (T = 3 h) because of the larger analysis period. Lower values of *p* are investigated for the T = 3 h analysis.

As the charts indicate, the higher the traffic volume, the more likely the sampled distributions exhibit good statistical fits to the true distribution. This is as expected; as volume decreases, the number of CVs also decreases (presuming they are randomly distributed in the vehicle fleet). As Figure 5a shows, for a high degree of confidence in matching arrival distribution shapes with T = 15 min, rather high values of *p* are needed, but even high penetration rates face problems at low volume intervals. Figure 5b shows that the longer analysis period lessens the required value of *p* considerably. For example, in Figure 5a, p = 10% results in P-values of 0.2 or less, but in Figure 5b, p = 10% yields P-values higher than 0.9 for all volumes. This shows that a low *p* can be combined with a high *T* to better sample the traffic characteristics; the tradeoff would be in whether traffic conditions would remain constant over the 3-hour period in order to be useful. The next section considers uses of the sampled profiles for offset optimization.



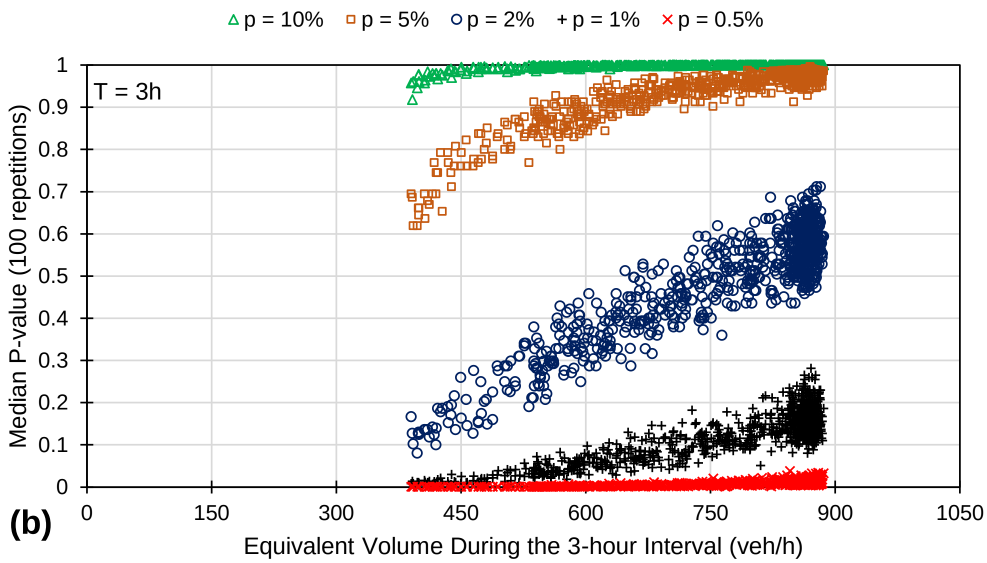


Figure 5. Relationship between traffic volume and the quality of arrival profile estimation. Each point is the average *P*-value of KS tests for 100 separate random samples for a different starting time. Data shown for (a) T = 15 min; and (b) T = 3 h.

CV Penetration Impact on Optimizing a Single Approach

In both offline and online methods of offset optimization, the goal is to minimize or maximize a performance measure calculated using measured traffic characteristics (*16*,*20*,*21*,*25*). Vehicle arrival profiles can be used to construct an offset-performance curve for an individual signalized approach. This offers an opportunity to develop insights on the influence of p on optimization, if CV data are used to measure traffic.

Figure 6 shows the plot of percent on green versus offset adjustment for the example data presented in Figure 3, for T = 3 h (Figure 6a) and T = 15 min (Figure 6b). Each curve represents the combination of the example subsampled flow profile with the observed green. Note that since the vehicle arrivals are already coincident with green, an adjustment close to 0 is optimal.

- With T = 3 h (Figure 6a), the p = 100% and p = 10% curves are very close, which occurs because their distributions are similar (Figure 3k). The p = 1% curve is slightly displaced yet tracks the others fairly closely.
- With T = 15 min (Figure 6b), the p = 10% curve and p = 100% curves are slightly different, yet exhibit the same general trend. Both attain their maximum value within 10 seconds of each other. The p = 1% curve is very different from the others because only two vehicles were observed (Figure 3h).

These results demonstrate that, despite the apparent poor quality of some subsampled flow profiles (Figure 3), and rather poor statistical fit to the sampled arrival curves (Figure 4), there remains a possibility of achieving near-optimal offset values when using the subsampled data for optimization (Figure 6), for at least some levels of market penetration.

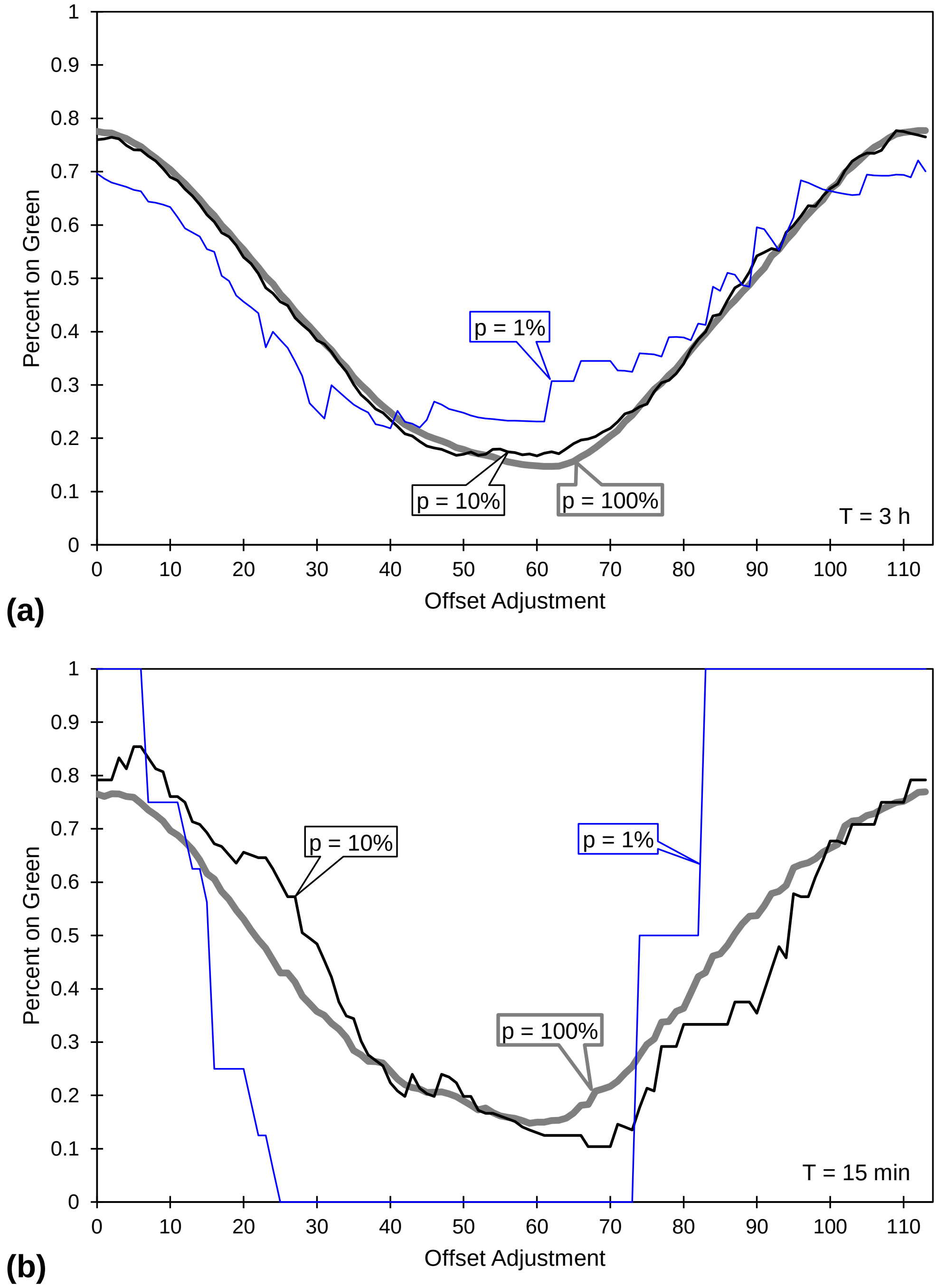


Figure 6. Offset-performance curves for decreasing market penetration p for (a) $T = 3$ h and (b) $T = 15$ min. Data shown for Northbound at SR 37 and Town and Country Blvd.

Optimizing the Corridor

Next, offset optimization with CV data arrival profiles was explored. Profiles were developed for the entire corridor by the sampling process described earlier. For each corridor profile, offsets were optimized using a previously developed algorithm (*18*); the resulting offsets were then re-entered into a model with complete data to estimate the performance of the solution based on the subset data. This is similar to testing trial offsets in a mesoscopic simulation tool such as TRANSYT (*25*).

To study the impact of market penetration, the procedure was executed for p = {50%, 10%, 5%, 1%, 0.5%, and 0.1%}, and for T = 3 h (using data from 12:00–15:00) and T = 15 min (using data from 13:45–14:00). This respectively modeled the performance for offline and online use cases. To mitigate random effects, this entire procedure was repeated 100 times for every combination of T and p.

Figure 7 presents the results for optimizing the 9-intersection corridor. Figure 7a and Figure 7b respectively show the overall system percent on green for T = 3 h and T = 15 min. The leftmost point represents the optimal offsets with complete data (p = 100%), which gives the best possible performance as indicated by the upper green line. The lower red line represents the *worst* possible performance, found by running the optimization algorithm to minimize arrivals on green. The red and green lines thus delineate the range of possible values. The blue line shows the existing percent on green for the existing offsets. Values above this line are system improvements while those below the line are degradations.

The black lines show the distribution of outcomes for 100 iterations at each value of p; the median value and the interquartile range are indicated by the 25th and 75th percentiles, while the dashed lines show the minimum and maximum values within each group. As p decreases, the optimization results tend to gravitate toward the middle of the range, resembling the performance of randomly selected offsets. However, the degree to which the curve begins to diverge differs considerably depending on T.

- For T = 3 h (Figure 7a), representing the offline optimization use case, the median outcome for p = 1% is only a few percentage points lower than the best possible outcome; even p = 0.5% produces a reasonable result, although it is becoming less reliable. Thus, for offline optimization, replacement of detection by CV data seems likely to provide a reasonable source of data even at extremely low rates of market penetration.
- For T = 15 min (Figure 7b), the median outcome begins diverging much more quickly. While the results for p = 10% are still within an acceptable range, while p = 5% is slightly more degraded and less reliable, while p = 1% is much worse. At p = 0.5% and p = 0.1%, there was not enough data available to run the optimizer, because most of the flow profiles had no data.

These results imply a critical value of p for offline and online applications in current signal control practice. Using the existing system performance (the blue line in Figure 7) as the threshold, consider a value of p to be unacceptable if the 25th percentile of its outcome distribution falls beneath the existing offsets. This means that estimated performance is considered acceptable as long as the offsets would be improved 75% of the time. Based on this threshold, the critical value for p for offline applications is 1.0%, and the critical value for online applications (with a 15-minute horizon) is 5%.

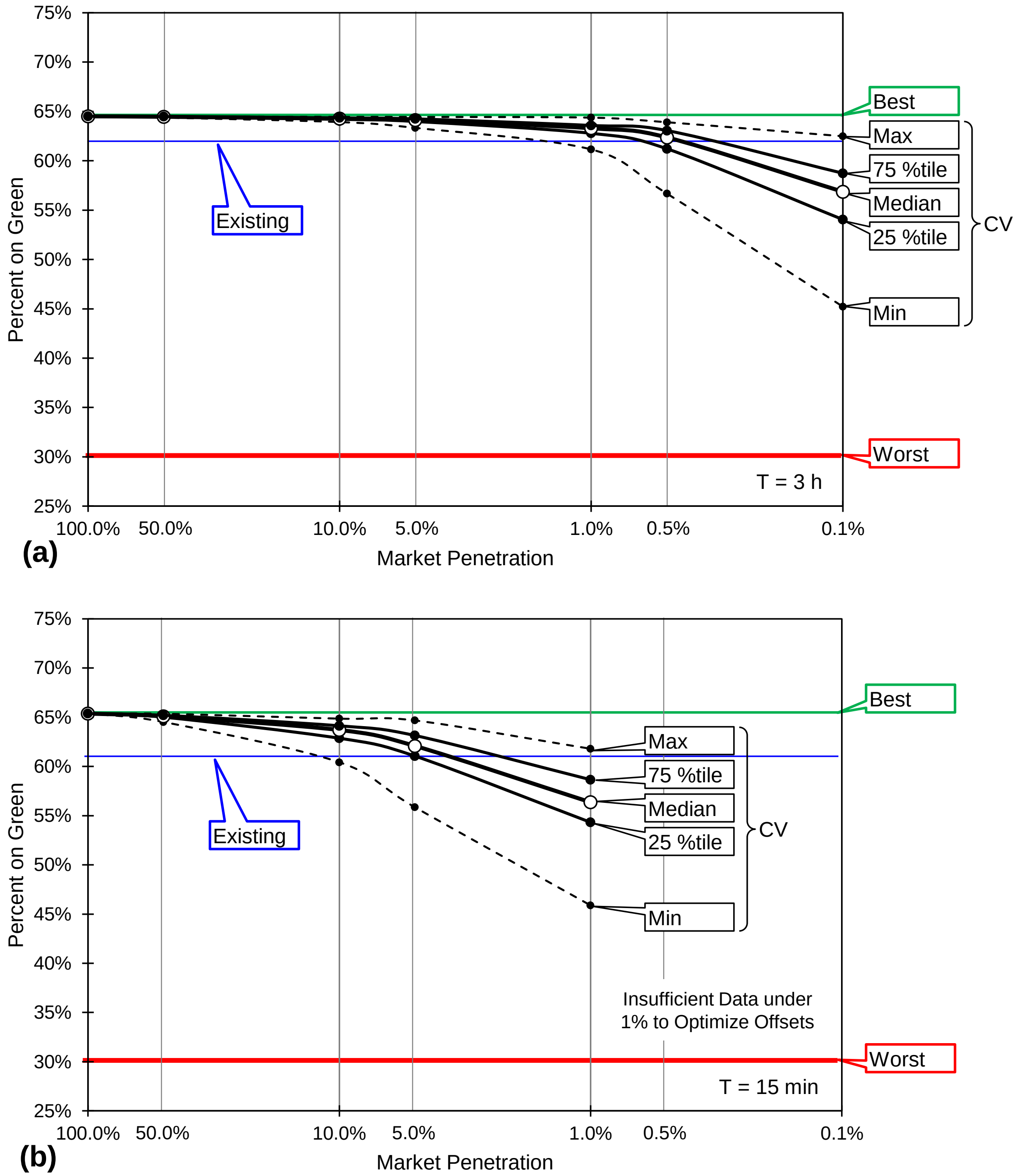


Figure 7. Sensitivity of optimization outcome to market penetration p for (a) T = 3 h and (b) T = 15 min. Lines marked "CV" show the distribution of results for 100 iterations at each trial value of p.

Layering Multiple Days

For offline applications, it may be possible to compile multiple days of operation, increasing the analysis period and increasing the amount of data that can be collected. In Figure 6a, the offset-performance curve for T = 3 h showed that the three performance curves track each other closely. However, there was not enough data at p = 0.1% data to allow for a meaningful adjustment.

Figure 8 repeats this analysis, this time combining data from three separate 3-hour periods from three consecutive Saturdays: June 6, June 13, and June 20, 2015. Therefore, T = 9 h. Such a combination is possible because the same timing plan was in operation and traffic conditions were similar throughout. This is similar to how sparse travel time data can be combined from multiple instances of a time of day plan to yield a more informative distribution of the travel characteristics (*22*).

The 9 h offset-performance curves show that p = 10% and p = 1% both follow the p = 100% line more closely. Also, there is now sufficient data for p = 0.1% to develop a curve for that data set. While the p = 0.1% curve does diverge from the p = 100% curve, the minimum and maximum points lie within similar regions of the cycle.

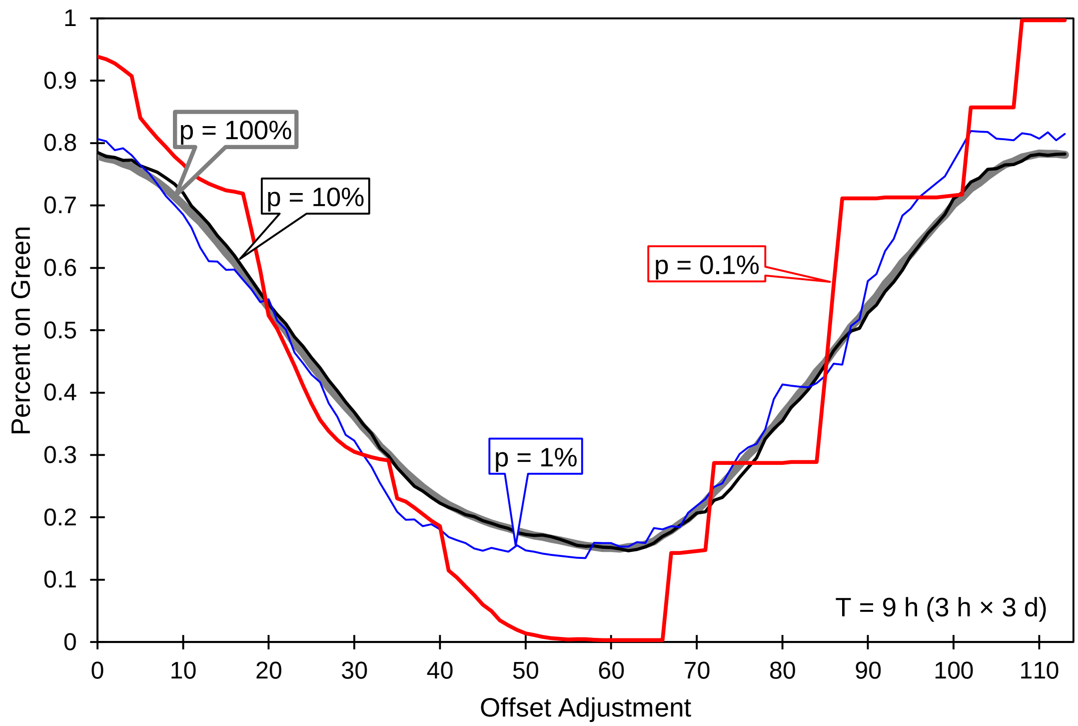


Figure 8. Impact of layering multiple days of data: Offset-performance data for three aggregated days of data (9 hours total), at different levels of *p*.

Implications for Early Stage CV Applications

The results demonstrate that detector-free optimization may be feasible with very low market penetration rates of CV, if the CV location data are of enough fidelity to provide vehicle arrival times as accurate as current physical detectors. Consequently, some initial applications of CV data would be possible during early stages of CV deployment, before penetration can enable more advanced applications.

Currently, the infrastructure for CV data through DSRC (*10*) is under development. However, detector-free optimization could still be implemented using vehicle positions measured by other means. While low-latency transmission via DSRC would be needed for collision avoidance and real-time control, applications such as offline optimization could tolerate latencies of minutes or longer. Vehicle positions obtained from probe vehicle data sources could potentially serve this purpose. While public use of private vehicle trajectories might raise concerns about privacy, the distillation of that data into virtual detections—one anonymous data point per vehicle—would be far less intrusive.

Such "CV-like" data, to propose a name, is already in common use for visualizing real-time traffic conditions. In the US, several data vendors are selling aggregations of such data, which have been used to analyze highway performance (*22,26,27*). In Germany, some researchers have developed performance measures for individual movements using individual vehicle trajectories (*28*). A possible next step would be to fuse vehicle position data with signal phase data, which would require a method to synchronize clocks between the two systems.

While this would not provide more data existing detection systems, it would enable optimization where detection is faulty, inadequate, or nonexistent. The huge inventory of semi-actuated and fixed-time systems could be optimized without having to install any detection. As this conceptual analysis demonstrates, tangible results might be possible with as little as 1% of fleet coverage for offline applications (and possibly less if multiple days of week can be combined), and 5% for rudimentary elements of adaptive control.

CONCLUSIONS

This study investigated the possibility of using connected vehicle (CV) data at very low rates of market penetration p (as low as p = 0.1%) to serve as a substitute for detector data in offline and online offset optimization. The study used distributions of real world detector data as the basis for comparison, with subsampled distributions used to estimate CV data. Relationships between p, analysis period T, vehicle volume, and quality of the observations were investigated using statistical analysis (Figure 4, Figure 5). As expected, higher p or longer T yielded better quality measurements.

The subsampled vehicle distributions were then used to optimize offsets, and the resulting parameters were re-entered into the complete-data model to evaluate their performance. As expected, the corridor performance degraded as p decreased (Figure 7). However, acceptable performance was obtained even at relatively low values of p. For a 3-hour analysis period, representing offline offset optimization, 75% of the solutions improved upon existing conditions at p = 1% or higher (Figure 7a). For a 15-minute analysis period, representing online offset optimization, this was true for p = 5% or higher (Figure 7b). Thus, a critical value of p = 1% was found for data aggregated over 3 hours and p = 5% for data aggregated over 15 hours.

The results suggest uses for CV data with low rates of market penetration in the early stages of CV deployment. In particular, would become possible to begin measuring vehicle arrival profiles at locations without any detection infrastructure. Since the applications discussed here occur in time frames of minutes or longer, the use of vehicle position data from external sources such as smartphones or other mobile devices might be an appropriate data set to explore in the near term, which would yield “CV-like” data.

FUTURE IMPLEMENTATION

Future work would test the concepts explored in this paper by combining real world phase status and vehicle position data (from a CV or CV-like data source). One unknown is whether the vehicle positon data would be accurate enough to be effectively translated into detector data. Possible avenues of exploration include simulation (*7*), or possibly probe vehicle data. For a real-world setting, time synchronization will be important. The goal would be to see whether the estimated distributions shown in this paper are borne out in a real-world setting. The concept would eventually be validated by using CV arrival distributions to optimize signal settings on a corridor without detection.

Lastly, while this study touched upon one particular aspect of signal operations (offset optimization), similar analyses could be done for different applications. For example, measurements of capacity utilization, queue length, or vehicle delay based on detector data might be achievable, which would make it possible to implement those concepts and related control strategies at locations that lack adequate physical detection.

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. 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. USDOT. Connected Vehicle Research in the United States. Available online at http://www.its.dot.gov/connected_vehicle/connected_vehicle_research.htm. Accessed July 27, 2015.

2. Priemer, C. and B. Friedrich. A Decentralized Adaptive Traffic Signal Control Using V2I Communication Data. Proc., 12th International IEEE Conference on Intelligent Transportation Systems, St. Louis, MO, 2009.

3. He, Q., K.L. Head, and J. Ding. PAMSCOD: Platoon-Based Arterial Multi-Modal Signal Control with Online Data. *Transportation Research Part C*, 20, 164-184, 2012.

4. Lee, J., B. Park, and I. Yun. Cumulative Travel-Time Responsive Real-Time Intersection Control Algorithm in the Connected Vehicle Environment. *Journal of Transportation Engineering*, 139, 1020-1029, 2013.

5. Goodall, N.J., B.L. Smith, and B. Park. Traffic Signal Control with Connected Vehicles. *Transportation Research Record No. 2381*, 65-72, 2013.

6. Goodall, N.J., B. Park, and B.L. Smith. Microscopic Estimation of Arterial Vehicle Positions in a Low-Penetration-Rate Connected Vehicle Environment. *Journal of Transportation Engineering*, 04014047-1, 2014.

7. Feng, Y., K.L. Head, S. Khoshmagham, and M. Zamanipour. A Real-Time Adaptive Signal Control in a Connected Vehicle Environment. *Transportation Research Part C*, 55, 460-473, 2015.

8. Argote-Cabañero, J., E. Christofa, and A. Skabardonis. Connected vehicle penetration rate for estimation of arterial measures of effectiveness. *Transportation Research Part C*, 60, 298-312, 2015.

9. Hill, C.J. and J.K. Garrett. *AASHTO Connected Vehicle Deployment Analysis.* Report FHWA-JPO-11-090. USDOT, 2011.

10. SAE International. *Dedicated Short Range Communications Message Set Dictionary*. Standard J2735. SAE, 2015. Available online at http://standards.sae.org/j2735_201504/. Accessed July 27, 2015.

11. Pew Research Center. *US Smartphone Use in 2015*. Available online at http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/. Accessed July 27, 2015.

12. Day, C.M., R. Haseman, H. Premachandra, T.M. Brennan, J. 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*, 37-49, 2010.

13. Day, C.M., D.M. Bullock, H. Li, S.M. Remias, A.M. Hainen, R.S. Freije, A.L. Stevens, J.R. Sturdevant, and T.M. Brennan. *Performance Measures for Traffic Signal Systems: An Outcome-Oriented Approach*. Purdue University, West Lafayette, Indiana, 2014.

14. "UDOT Signal Performance Metrics." http://udottraffic.utah.gov/signalperformancemetrics/. Accessed November 1, 2015.

15. AASHTO. "Automated Traffic Signal Performance Measures." http://aii.transportation.org/Pages/AutomatedTrafficSignalPerformanceMeasures.aspx. Accessed November 1, 2015.

16. Hillier, J.A. and R. Rothery. The Synchronization of Traffic Signals for Minimum Delay. *Transportation Science*, Vol. 1, pp. 81-94, 1967.

17. Gartner, N. Microscopic Analysis of Traffic Flow Patterns for Minimizing Delay on Signal-Controlled Links. *Highway Research Record No.445*, 12-23, 1973.

18. Day, C.M. and D.M. Bullock, Computational Efficiency of Alternative Algorithms for Arterial Offset Optimization. *Transportation Research Record No. 2259*, 37–47, 2011.

19. Hu, H. and H.X. Liu. Arterial Offset Optimization using Archived High-Resolution Traffic Signal Data. *Transportation Research Part C*, 37, 131-144, 2013.

20. Hunt, P.B., D.I. Robertson, R.D. Bretherton, and R.I. Winton. *SCOOT- A Traffic Responsive Method of Coordinating Signals*. Report No. LR 1014, Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1981.

21. Luyanda, F., D. Gettman, S. Shelby, L. Head, D. Bullock, P. Mirchandani, "ACS-Lite Algorithmic Architecture: Applying Adaptive Control System Technology to Closed-Loop Traffic Signal Control Systems," *Transportation Research Record No. 1856*, pp. 175-184, 2004.

22. Young, S. Use of Probe and Bluetooth Data for Arterial Performance Measures in the I-95 Corridor Coalition. Presented at the NATMEC Workshop on Arterial Data and Performance Monitoring, Chicago, IL, June 2014.

23. Day, C.M., T.M. Brennan, A.M. Hainen, S.M. Remias, H. Premachandra, J.R. Sturdevant, G. Richards, J.S. Wasson, and D.M. Bullock. Reliability, Flexibility, and Environmental Impact of Alternative Objective Functions for Arterial Offset Optimization. *Transportation Research Record No. 2259*, 8-22, 2011.

24. Smaglik E.J., A. Sharma, D.M. Bullock, J.R. Sturdevant, and G. Duncan. Event-Based Data Collection for Generating Actuated Controller Performance Measures. *Transportation Research Record No. 2035*, 97-106, 2007.

25. Robertson, D.I. *Transyt: a Traffic Network Study Tool.* Report No. LR 253, Road Research Laboratory, Crowthorne, Berkshire, England, 1969.

26. Schrank, D., B. Eisele, and T. Lomax. *2012 Urban Mobility Report and Appendices*. Texas A&M Transportation Institute, College Station. http://mobility.tamu.edu/ums/report/.

27. Day, C., S. Remias, H. Li, M. Mekker, M. McNamara, E. Cox, D. Horton, and D. Bullock. *2013-2014 Indiana Mobility Report – Full Version.* West Lafayette, Indiana: Purdue University, Indiana Mobility Reports, 2014.

28. 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.