FINAL REPORT

PROJECT N2

SEPTEMBER 2023

Data Fusion for Signalized Arterial Performance Measurement

Shoaib Samandar, Ph.D., North Carolina State University/ITRE
Nagui Rouphail, Ph.D., North Carolina State University/ITRE
Thomas Chase, PE, North Carolina State University/ITRE
Gyounghoon Chun, Ph.D., North Carolina State University/ITRE

STRIDE | Southeastern Transportation Research, Innovation, Development and Education Center

UF | Transportation Institute UNIVERSITY of FLORIDA

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. Project N2	**2. Government Accession No.**	**3. Recipient's Catalog No.**
4. Title and Subtitle Data Fusion for Signalized Arterial Performance Measurement		**5. Report Date** 9/11/2023
		6. Performing Organization Code
7. Author(s) Shoaib Samandar, Ph.D., North Carolina State University/ITRE Nagui Rouphail, Ph.D., North Carolina State University/ITRE Thomas Chase, PE, North Carolina State University/ITRE Gyounghoon Chun, Ph.D., North Carolina State University/ITRE		**8. Performing Organization Report No.** STRIDE Project N2
9. Performing Organization Name and Address ITRE at North Carolina State University Centennial Campus 909 Capability Drive, Suite 3600 Research Building IV Raleigh, North Carolina 27606-3870		**10. Work Unit No.** **11. Contract or Grant No.** Funding Agreement Number 69A3551747104
12. Sponsoring Agency Name and Address University of Florida Transportation Institute/ Southeastern Transportation Research, Innovation, Development and Education Center (STRIDE) 365 Weil Hall, P.O. Box 116580 Gainesville, FL 32611 U.S Department of Transportation/Office of Research, Development & Tech 1200 New Jersey Avenue, SE, Washington, DC 20590		**13. Type of Report and Period Covered** 8/1/2018 to 9/11/2023 **14. Sponsoring Agency Code**
15. Supplementary Notes N/A		
16. Abstract - This project focused on three applications of multiple data sources, or data fusion, to evaluate traffic operations on signalized arterials. The first application used a data fusion framework to evaluate its utility in predicting travel times on signalized arterials. The second application developed a framework that combined/enhanced the Automated Traffic Signal Performance Measures (ATSPM) generated through data fusion. The third project examined the use of probe vehicle data fusion for obtaining link speeds at signalized arterials by exploring the effects of market penetration rate (MPR), broadcasting frequency (BC), and aggregation level on the accuracy of the reported speeds. Results underscore the importance of BC, MPR, and aggregation levels. Importantly, varying the performance measures tended to yield different requirements in terms of probe MPR and BC.		
17. Key Words ATSPM, approach delay, data fusion, signalized arterials	**18. Distribution Statement** No restrictions	

19. Security Classif. (of this report) N/A	**20. Security Classif. (of this page)** N/A	**21. No. of Pages** 77 pages	**22. Price** N/A

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

ACKNOWLEDGEMENT OF SPONSORSHIP AND STAKEHOLDERS

This work was sponsored by a grant from the Southeastern Transportation Research, Innovation, Development, and Education Center (STRIDE).

Funding Agreement Number - 69A3551747104

LIST OF AUTHORS

Lead PI:

Shoaib Samandar, Ph.D.
Institute for Transportation Research and Education (ITRE) at North Carolina State University
smsamand@ncsu.edu
ORCID # 0000-0002-0018-3105

Nagui M. Rouphail, Ph.D.
North Carolina State University
roupahil@ncsu.edu

ORCID # 0000-0002-2420-9517

Thomas Chase, PE
Institute for Transportation Research and Education (ITRE) at North Carolina State University
rtchase@ncsu.edu

Gyounghoon Chun, Ph.D. Candidate
North Carolina State University
gchun@ncsu.edu

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

According to the USDOT, there are over 330,000 traffic signals in the US. Over 75% of these signals could be improved by updating equipment or timing plans. Poor signal timing accounts for nearly 300 million vehicle hours of delay on major roadways alone. As agencies now have access to multiple real-time data sources for performance measurement, they must consider those sources' cost, availability, and accuracy. This project focuses on three applications of multiple data sources, or data fusion, to evaluate traffic operations on signalized arterials.

The first application used a data fusion framework to evaluate its utility in predicting travel times on signalized arterials. Five fusion approaches were tested: simple averaging, K nearest neighbor (KNN), linear regression, random forest, and artificial neural networks (ANN). Data sources included probe vehicles, Bluetooth (BT)sensors, loop detectors, and signal timing plans. The implementation was on Peachtree signalized arterial in Atlanta, for which detailed trajectories were available from the NGSIM program. Each data source was simulated to generate its travel time estimate using appropriate sampling techniques and signal data in four scenarios: two time periods and two travel directions. Sources were tested individually and in various fusion combinations.

In summary, BT data yielded the lowest % error (MAPE) compared to the ground truth at 17.1%, while the probe data (a 10% sample) had a MAPE of over 30%. The virtual loop sensors placed at midblock generated errors of over 60%. The fusion of multiple data sources did improve the travel time predictions, although this reduction was more substantial in the case of the ANN, yielding a MAPE of 10.6% when combining BT, probe, and signal timing data. A fusion model of signal plans and detector data that develops travel time estimates using virtual trajectories showed promise as a tool for improving those estimates.

The second application developed a framework that combined/enhanced the Automated Traffic Signal Performance Measures (ATSPM) generated through data fusion. The study found that the ATSPM delay estimation method excludes the effect of acceleration, deceleration, and queuing. This study introduced two delay estimation methods to address those limitations by fusing high-resolution detector data and signal plans. The arrival-departure method uses upstream and downstream detector data to estimate delay. The second method uses only detectors measuring departures near the stop bar. The procedures were implemented alongside the current ATSPM approach on a Utah intersection in Salt Lake City. The methods were validated using high-resolution NGSIM data on Peachtree Street in Atlanta. That effort revealed the superiority of the arrival-departure to ATSPM, while the departure-only method was found to be uniquely sensitive to delay in peak periods and only when arrivals are nearly random.

Finally, the project examined the use of probe vehicle data fusion for obtaining link speeds at signalized arterials by exploring the effects of market penetration rate (MPR), broadcasting frequency (BC), and aggregation level on the accuracy of the reported speeds. Results underscore the importance of BC, MPR, and aggregation levels. Importantly, varying the performance measure tended to yield different requirements in terms of probe MPR and BC.

Keywords: ATSPM, approach delay, data fusion, signalized arterials

EXECUTIVE SUMMARY

There are over 330,000 signalized intersections operating in the US. Until recently, performance measurement of the efficiency of a signal timing plan relied mostly on manual observations and the use of floating travel time runs. With new emerging sources of data, including high resolution vehicle detection, and the availability of probe vehicle fleets, researchers and signal timing professionals are beginning to exploit those data sources to generate traffic signal performance measures in near real time. This study takes advantages of both data sources using the concept of data fusion to generate and track signal performance over time. It begins by developing a data fusion framework for signalized arterials at both the link, network and path levels. It then proceeds to apply the framework to signal and high-resolution detector data for estimating control delay, as well as to third party probe data reporting for estimating arterial speed /travel time.

The first data fusion application focused on testing various data fusion methods with multiple data sources of travel time. The context was to evaluate combinations of both sources and methods in predicting travel time on an urban signalized arterial, namely Peachtree road in Atlanta. The arterial had 6 signalized intersections, and because high resolution trajectory data were available from the NGSIM program, it allowed the study to simulate the effect of the availability of various datasets on through travel time prediction. In addition, the actual travel time from those trajectories was available by time period and travel direction. The study tested five fusion algorithms: simple averaging, K nearest neighbor (KNN), linear regression, random forest and artificial neural networks (ANN). The simulated data sources included probe vehicles assuming a 10% sample, Bluetooth sensors on either end of the arterial, loop detectors installed at midblock locations and signal timing plans including cycle, splits and offsets for the 6 signals. The study found that the closest data source to measured travel time was from Bluetooth sensors, yielding an overall Mean Absolute Percent Error (MAPE) of 17.1%. This compares with a MAPE of 30% from probe data and over 60% (underestimation of travel time) from midblock fixed sensor data. The fusion algorithms produced meaningful reductions in those errors, depending on which combination of sources were tested. Across 10 different source combinations, ANN fusion produced the lowest errors in 7 of those cases. The lowest prediction error occurred by fusing BT, Probe and Signal data via ANN. That combination yielded a MAPE of 10.6%, or a 38% error reduction compared to the best single source data.

The second data fusion application evaluates and compares the Automated Traffic Signal Performance Measure (ATSPM) against two alternative methods for control delay estimation. The evaluation took place at a signalized intersection equipped with high resolution sensors in Salt Lake City, Utah. The ATSPM approach was found to generate low values of delays as it does not account for deceleration, acceleration or queuing effects. Only red time delay is considered. The first arrival departure method uses both upstream and stop bar detection and generates higher delays than ATSPM, while the second method, a departure only detection method was found to generate delays which were even lower than ATSPM at that site. Because the true delay was not measured at the Utah site, a separate validation effort took place using an archived, full trajectory and signal plan dataset under the Next Generation Simulation Program (NGSIM) collected on Peachtree St. in Atlanta, similar to the previous application. The validation work confirmed the original findings from the Utah experiment.

That is, ATSPM generally underestimated the true delay, while the arrival departure method generated slightly higher delays, and the departure only method worked well only under heavy congestion, otherwise this method significantly underestimated delays. One contributing factor may have been the arrival patterns at the Peachtree intersection, where most arrivals occurred near the start of the red phase, while the departure only method assumes vehicles uniformly arrive throughout the red phase. Overall, the arrival departure method appears to show much promise, although some additional, minor modifications to the delay estimation model may be in order.

Finally, the study also examined the validity of probe vehicle data fusion for estimating link speeds on a signalized arterial, using high resolution trajectory data from another NGSIM site on Lankershim Blvd. in Los Angeles, CA. Specifically, the research question was to assess the accuracy of probe sample speed reports from third party providers. To this effect, accuracy was tested using the metrics: Average Absolute Speed Error (AASE), Speed Error Bias (SEB) and Root Mean Square Error (RMSE). It was assumed that accuracy would be a function of probe market penetration rate (MPR), probe broadcasting frequency (BC), and provider data aggregation level. The study results showed that depending on the metric used to assess the accuracy, certain combinations of MPR and BC would be required to satisfy the criteria. Generally speaking, a minimum MPR of 10% and broadcasting frequency in the 15-30 seconds range was found to meet most of the performance criteria that were specified in the study.

1.0 INTRODUCTION

Though a vital part of today's traffic control systems, traffic signals can also contribute to chronic congestion and safety hazards when not operated in an efficient manner. There are upwards of 330,000 traffic signals in the United States. Critical assessments of traffic signal operations have revealed several issues contributing to their poor performance. These include poor selection of performance measures, aging control infrastructure, disparate and inconsistent data sources, and a lack of robust methodology to combine available data sources. Two promising solutions to some of these challenges are fusion of available but disparate data sources and Automated Traffic Signal Performance Measures (ATSPM), respectively.

Data fusion combines traffic data collected from multiple discrete sources to enhance the accuracy and reliability of the metrics of interest. Researchers have attempted various data fusion techniques to assess signalized arterials' performance. Fusing fixed and mobile sensor data types has yielded promising results compared to single source datasets. While attractive, fusion practices are scarce, amorphous, and difficult to adopt and replicate. Furthermore, these practices need to consider the data availability, the disparity in geospatial levels of the data, performance measure selection, and intra-source and inter-source fusion precursors. As such, there is a dire need for a generalized data fusion framework for signalized arterials to address these limitations.

This study takes advantage of multiple data sources using the concept of data fusion to generate and track signal performance over time. It begins by developing a data fusion framework for signalized arterials at the link, network, and path levels. It then proceeds to apply the framework to a key signalized arterial performance metric -- arterial travel time. Application of the framework is done using a Next Generation Simulation (NGSIM) dataset and considering various data availability scenarios.

Automated Traffic Signal Performance Measures (ATSPM) collect high-resolution signal and detector data and generate performance metrics for approaches at signalized intersections (*1*). These metrics include but are not limited to total and average delay, split failure, and yellow and red actuations. They help improve decision-making and in selecting more effective operational strategies to reduce congestion and crash risk. However, there are significant limitations in traditional ATSPM methods to generate delay and other performance measures. For example, approach delay reported by ATSPM does not account for acceleration, deceleration, and queueing delays. Moreover, the delay estimation process does not account for any initial queue and/or cycle failures. Hence, it is likely that ATSPM's reported delays will underestimate the actual delays, prompting the need for further algorithmic improvements.

This study has developed two algorithms to estimate approach delay at a signalized intersection using high-resolution signal and detector data. Each of these algorithms addresses at least one limitation in the current ATSPM delay estimation method. High-resolution traffic data from the Utah Department of Transportation were provided at two corridors for the year 2018 (*2*). The signal phase and detector actuation data were fused, and fundamental traffic

flow relationships were applied to develop these methods. The two methods and the existing ATSPM delay estimation methods were applied on a through-movement approach at a signalized intersection located at the 700 East Corridor in Salt Lake City (SLC), Utah. Delays estimated by all three methods are compared, and their variation by time of day is portrayed.

1.1 OBJECTIVES

The main objective of this research project is to improve our understanding of the operational performance of traffic signals and signalized arterials through the fusion of disparate data sources. Principal tasks include:

1) Develop a data fusion framework to incorporate data with different spatiotemporal resolutions, including synthetic data sources.
2) Collect/acquire traffic signal data through Automated Traffic Signal Performance Measures (ATSPM) software, high-resolution publicly available databases, link probe data from HERE and NPMRDS, and individual vehicle data from instrumented vehicles.
3) Application of the developed data fusion framework to a signalized arterial. The framework is employed to assess a signalized arterial performance metric, and the accuracy and effectiveness of the framework are compared to a single-source metric.
4) Prototype and test signalized arterial performance measures using the data fusion set. This objective develops enhanced ATSPM for approach delay using multiple algorithms.
5) Evaluate the effectiveness and accuracy of new performance measures compared to single-source performance measures.
6) Investigate the impact of market penetration rate and broadcasting frequency on probe speed reports.

1.2 SCOPE OF WORK

This project developed a generalized data fusion framework, applies it to a real-world corridor with multiple intersections, proposes algorithms for enhancing the performance of approach delay in the context of ATSPMs, and investigates the impact of market penetration rate and broadcasting frequency of probe vehicle speed reports.

Multiple datasets, including high-resolution signal and detector data from the Utah Department of Transportation, along with vehicle trajectory datasets from Atlanta, Georgia, are used throughout different stages of the fusion process. Using these data, data fusion techniques are employed to develop signalized arterial performance measures, which can be used for the maintenance and operations of traffic signals. Trajectory level data provides the most accurate estimate for individual vehicle operations. However, collecting high resolution data for an entire traffic stream is highly infeasible. When determining how to fuse multiple data sources for a combined set of signalized arterial performance measures, the sample size, geographic and temporal scope, and data quality effects are evaluated.

For measures such as travel time and delay, multiple sources may directly provide an estimated value, with each incorporating the limitations of the data source. Commercial probe data may also provide a link travel time which is assumed to incorporate approach delay; however, data providers are not clear about how or if stopped time is filtered out from this data and if all turning movements are aggregated into the link average. Bluetooth and WiFi, as well as instrumented or connected vehicles, will potentially provide direct measurements of delay; however, sample size and intermediate stops (such as a gas station visit) must be considered. Regarding arterial corridor travel time, we fuse these disparate data sources (loop detector, Bluetooth, and varying market penetration levels of probe vehicle data) to obtain an enhanced estimate of the travel time along a signalized corridor. Regarding delay, ATSPM provides a delay estimate based on the red time remaining after a projected arrival on red (using upstream detection and progression speed). This approach does not consider queuing or other aspects of control delay. We develop two algorithms that address the drawbacks present in the ATSPM approach delay methodology.

Validation and sensitivity testing is performed based on the developed measures using simulated or NGSIM data where all information is known about the entire traffic stream. Scenario-based fusion is also conducted to determine the value of adding additional data sources to track signalized arterial performance.

2.0 LITERATURE REVIEW

The literature review section of this report is organized into two sub-sections focusing on (a) the application of data fusion in transportation engineering and (b) automated traffic signal performance measures – particularly the estimation of approach delay at signalized intersections.

2.1 Application of Data Fusion in Transportation

A multitude of definitions can be found in the literature for data fusion. One of the comprehensive definitions for the process is suggested by Mitchell (*3*) that defines it as the "theory, techniques and tools which are used for combining sensor data, or data derived from sensory data, into a common representational format" and goes on to provide the main goal and purpose of it as "in performing sensor fusion our aim is to improve the quality of the information, so that it is, in some sense, better than would be possible if the data sources were used individually". Ng (*4*) states that "fusion involves the combination of data and information from more than one source". Hall & Llinas (*5*) states "data fusion techniques combine data from multiple sensors and related information to achieve more specific inferences than could be achieved by using a single, independent sensor. Although a variety of definitions are available for data fusion, a close look, however, reveals that these definitions converge on the fact that the process is comprised of a variety of activities that pertain to more than one data source.

Several raison d'être for data fusion has been suggested by (*3*–*5*) and summarized by (*6*):

- **Reliability/robustness/redundancy:** systems that rely on a single input source are prone to operational failures amid malfunction of the input source. However, systems obtaining their input information by fusing multiple data sources have higher resilience since multiple input sources providing redundant information decrease the un-reliability in the case of a sensor malfunction.
- **Accuracy/certainty:** obtaining data from multiple sensors can provide a system with more accurate information. Combining multiple readings from the same sensor makes a system less sensitive to noise and temporary glitches. Increasing the number of input sensors not only boosts accuracy, but also certainty by separating signal from noise.
- **Completeness/coverage/complementarity:** ability to collect multiple data sources on an object/state of interest will result in extended coverage of information which is crucial in spatial and temporal environments for the sake of entirety.
- **Cost effectiveness:** cost of developing and manufacturing a sensor that can take multiple accurate measurements and conduct multiple functions is expensive compared to integrating multiple simple and cheap sensors that are capable of specific functions.
- **Representation**: an important issue that data fusion helps with is reduction of information overload. This is essentially achieved by combining information from multiple sources and clearly presenting the best interpretation of the sensor data to allow for well informed and timely decision.
- **Timeliness**: fused data from multiple sensors provide more timely information compared to a single sensor. This could be achieved either due to the operational speed of each sensor, or the processing parallelism happening in the fusion process.

Data fusion in transportation engineering is relatively new and has received attention parallel with the emergence of intelligent transportation systems (ITS). El Faouzi and Klein (*7*) identified a comprehensive list of transportation domains in which data fusion has previously been employed. They include advanced transportation management systems (ATMS), automatic incident detection (AID), advanced traveler information systems (ATIS), advanced driver assistance systems (ADAS), and commercial vehicle operations (CVO). These domains can gather information from different data sources and apply different data fusion techniques for processing and combining the information to make better decision or provide better knowledge of the situation at hand. Table 1 provides a summary of the data fusion algorithms used in ITS to support its applications and strategies.

TABLE 1: DATA FUSION ALGORITHMS CURRENTLY APPLIED TO ITS ADAPTED FROM (*7*)

No.	Application	Data Fusion Algorithm
1	Ramp Metering	Fuzzy Logic
2	Pedestrian Crossing Time	Fuzzy Logic
3	Automatic Incident Detection	Artificial Neural Network
		Bayesian Inference
		Dempster-Shafer
4	Travel Time Estimation	Inference Rules
		Dempster-Shafer
		Weighted mean of several travel-time estimators. Weights are a function of the variance or covariance of the estimators.
		Weighted mean where weights are function of the data source reliability.
		Fuzzy Logic
5	Vehicle and Object Tracking	Kalman Filter
6	Lane Departure Warning	Image processing using edge detection and extraction of the other features.
7	Traffic State Estimation	Extended Kalman Filter
8	Crash Analysis and Prevention	K-means algorithm
9	Traffic Forecasting and Monitoring	Bayesian Inference
		Artificial Neural Network
		Kalman Filter
		Extended Kalman Filter
		Kernel Estimator
		Particle Filter
10	Vehicle Position Estimation	Unscented Kalman Filter
		Artificial Neural Network

The application of data fusion techniques in the context of signalized arterial corridor is scarce and relatively new. One of the first studies was conducted by Niittymaki and Kikuchi (*8*) who designed a fuzzy system to regulate pedestrian crossing at a signalized intersection. The

fuzzy system had multiple objective functions such as, minimizing pedestrian waiting time, minimizing the delay for the vehicular traffic by not stopping the vehicle flow for longer periods, and maximizing safety for both the pedestrians and vehicular traffic by preserving and passing groups of approaching vehicles. In a separate study Mueck (*9*) developed a model to determine queue length using vehicle count data obtained from traffic flow sensors located at the stop bar in conjunction with signal timing information. Using Muek's model as a quasi-measurement that employs Kalman Filtering techniques, Friedrich and Minciardi (*10*) developed a new approach based on queuing theory models for real time queue length estimation.

Zheng et al. (*11*) used a weighted mean technique for fusing loop detector with floating car data (FCD) to estimate arterial link travel time. The weights were determined using both the sample size of the floating car data and the margin of error of the detectors. The fusion method was verified using VISSIM platform and the authors concluded that the output of the procedure provides better estimate of link travel time compared to a single data source. Ivan et al. (*12*) developed a methodology for real-time arterial street incident detection by fusing loop detector data and probe vehicles reporting their travel times. Artificial neural network was deployed as the main fusion algorithm. Wolfermann et al. (*13*) used floating car data with loop detector and historic data to develop a fusion method using Kalman Filter algorithms for traffic flow estimation at intersections. They found that the concept works well and promises to be useful tool for municipalities with sufficient FCD sample sizes.

2.2 ATSPM Approach Delay Method

Earlier studies on evaluating the performance of signalized intersections employed simplified shockwave theories to estimate queue length and delay based on average vehicle demand data (*14*, *15*). With technological advancements, models that utilize detector data for delay estimation were developed. According to the Highway Capacity Manual (*16*) and other studies (*17–19*), total vehicle count, arrival type, and signal timing events are key inputs for control delay estimation using vehicle actuation based data. Due to the spatial variability of queued and unqueued traffic states, it is difficult to determine a reasonable estimate of vehicle count and arrival type from detectors at a single location. Several studies relied on data from multiple locations to estimate control delay and other performance measures (*20–22*). Vigos et al. (*23*) and Vigos and Papageorgiou (*24*) used data from three sets of loop detectors to estimate the number of vehicles at a metered on-ramp. The use of virtual detectors in video records of an intersection is also common (*20*, *21*). Kebab et al. (*20*) proposed a delay estimation model by placing virtual detectors at two points of video records from three signalized intersection: one at upstream of the expected maximum extent of the queue and another at the stop bar. When compared to ground truth data, the study concluded that the proposed method performed better than the HCM method. Zheng Ma et al. (*21*) also used two sets of detectors to develop a queue length and control delay estimation method and tested that using both field observations and simulated data. Lee et al. (*25*) developed a discriminant model using stop-bar and upstream detectors to estimate the existence of residual queue, and subsequently, to estimate different performance measures.

Only a few studies developed methods to assess signalized intersection performance based on vehicle actuation data from detectors at a single point (*18*, *26*, *27*). Liu et al. (*26*) proposed a queue length estimation process using high-resolution data from upstream detectors obtained from the SMART-SIGNAL database (*28*) by leveraging the Lighthill–Whitham–Richards (LWR) shockwave theory (*29*). The method hinges on identifying the timestamps of three critical events: the arrival of the stop shockwave, departure shockwave, and queue clearance shockwave at the detector location. Zheng et al. (*27*) used the same dataset to develop an algorithm for creating and visualizing trajectories upstream of a signalized intersection. The method is straightforward to apply to cycles during which the queue does not reach the detectors. However, during peak periods when such cycles are rare, the method relies on Newell's simplified car-following theory to estimate the queue length. The major limitations of these two methods are that the queue must be cleared at least up to the detector location and vehicles must not arrive in a platoon; although both scenarios are common in an urban road network. Chang et al. (*18*) used the occupancy and count data to develop a queue length estimation method. It suggested that such methods generate accurate estimates if the setback distance of the upstream detectors is less than the expected queue length. Besides, this method relies on calibrating two parameters namely the average vehicle speed while approaching a queued intersection and an adjustment factor to convert the discharge flow to the arrival flow. Liu et al. (*30*) demonstrated a method to improve the conventional queue length estimation method using advance and stop-bar detectors if one or both sets of detectors give faulty measurements. However, it uses a Kalman filtering technique– the state equation of which is based on a simplified linear regression model between queue length and occupancy of faulty detectors. More recently ATSPM (*31*) has gained popularity in terms of using high-resolution data at signalized intersections for performance estimation. It estimates the delay to each vehicle based on actuation data from the upstream detectors and matching the expected departure time of the vehicle with the signal change timestamp. Details about the ATSPM delay estimation method are discussed later.

It is apparent from the above discussion that several methods exist in the literature to estimate control delay, queue length, and other measures using multiple sets of detectors. Only a few methods addressed the issue of reducing data requirement using vehicle actuation data only from advance upstream detectors. Nonetheless, most of these methods are pre-conditioned on several factors such as the setback distance of the detectors, the presence of under-saturated cycles, and no platooning effects. The potential benefits of improved control delay estimation from single detectors have direct benefits to systems like ATSPM which face limitations based on the equipment installed in the field.

3.0 Data Fusion Framework Development

Several data fusion process representations, such as functional, architectural and mathematical exist. A functional data fusion model represents the associated functions, databases and connections required for the underlying process. The architectural data fusion frameworks, however, best represents the data flow procedures, software or hardware configurations, and external/internal interfaces.

This project uses the *functional data fusion process model* to deliver the framework which utilizes disparate data sources for geospatial and temporal fusion of traffic data. Figure 1 is a flowchart of the proposed data fusion process and the main components in the fusion, including the performance assessment geospatial level, ATSPM performance measure selection, the intra-source fusion module, and the inter-source fusion module.

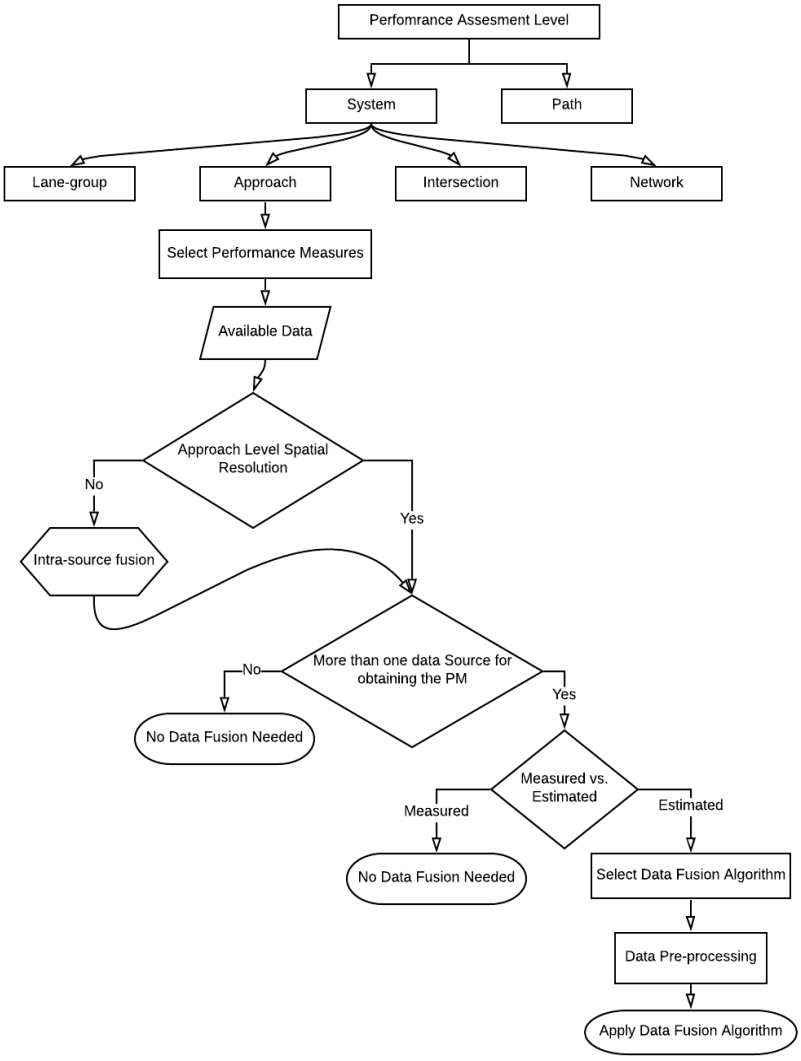


FIGURE 1: DATA FUSION FRAMEWORK

Initially the performance assessment at the geospatial level should be selected. Depending on what type of elements comprise the geospatial extent of the assessment, it can be classified into two groups. The first group represents those that contain only nodes such as, lane-group, approach, intersection, and network (a collection of intersections for the purpose of this project). The second group, however, represents geospatial levels that are composed of both nodes and links such as a corridor, which is a collection of nodes and links.

Second, the performance measure pertaining to the geospatial assessment level chosen in step one needs to be identified. Multiple assessment metrics are present for each geospatial assessment level. Table 22 gives a complete list of available performance measures for each geospatial level. Available data for the selected performance measure will be assessed to determine whether it represents the same granularity as the spatial extent of the measure itself. Depending on the spatial resolution of the data two potential scenarios exist. The first scenario assumes that the collected data and the performance metrics both have the same spatial resolution. As such, no further processing is needed, and the data fusion process moves to the next step. The second scenario, however, deals with the possibility of having different spatial resolutions between the collected data and the performance metrics. For example, evaluating intersection delay while having lane-group data will require multiple (lane-group to approach and approach to intersection level) intra-source fusion of the data to elevate it to the same geospatial resolution as the performance metric. Figure 2 presents the entity relationship diagram for both the system and path levels.

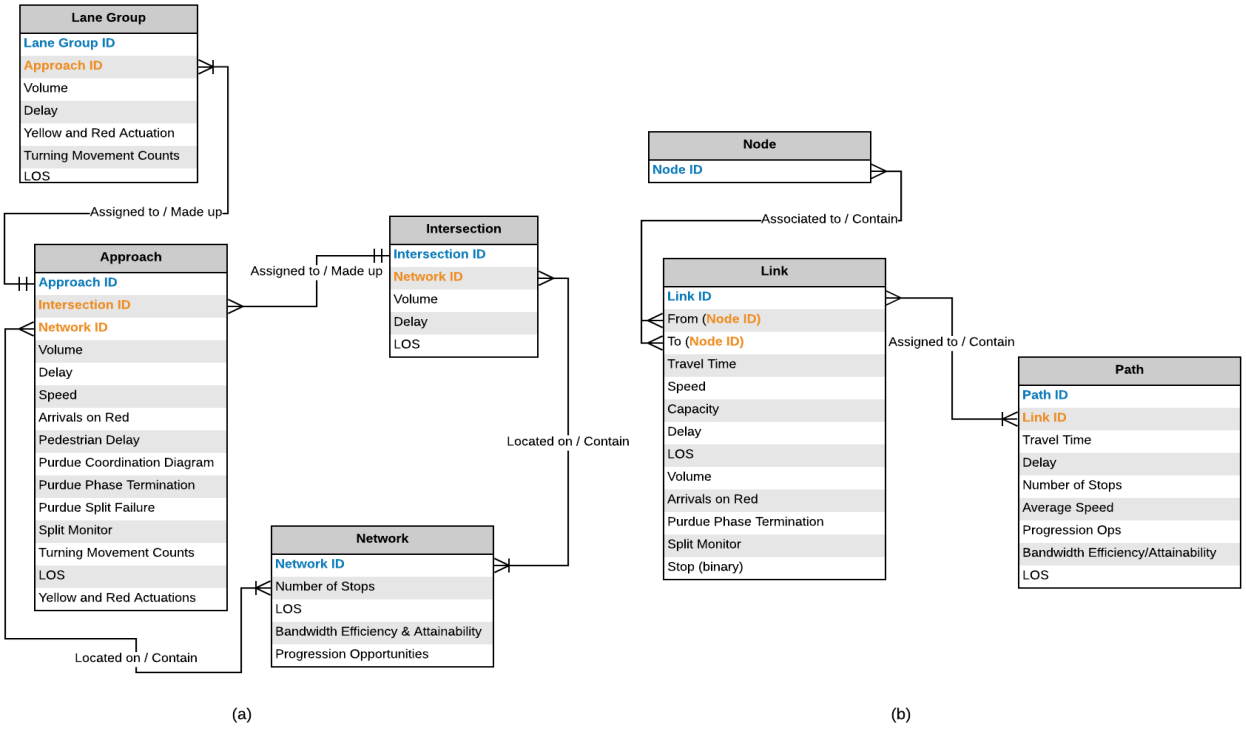


FIGURE 2: ENTITY RELATIONSHIP DIAGRAM FOR ARTERIAL PERFORMANCE MEASURES A) SYSTEM LEVEL B) PATH LEVEL

The last module determines whether data fusion for a specific ATSPM is needed. The initial step under this module determines whether multi-source data for a given ATSPM are available. In the absence of multi-source data, no fusion is feasible and as such the single-source data are used for estimating the performance measure. However, if multi-source data is available for a given ATSPM the process moves to the next step where the nature of data is investigated for potential estimation errors. Given that the measures were calculated using the available data (i.e. no errors are associated with the calculated measures ---as it is not an estimation --- and therefore the measure obtained from different data sources provide the same result), no fusion is needed because all data sources will report the same value and have no inherent errors. However, if the measure was estimated (i.e. inherent errors are associated with the reported measure value from different sources) a data fusion algorithm is selected, and the fusion process is performed for the ATSPM.

TABLE 2: TRAFFIC SIGNAL PERFORMANCE MEASURES AND ASSOCIATED SPATIAL RESOLUTION

Performance Measures	Spatial Resolution				
	Lane-group	Approach	Intersection	Network	Path
Control Delay	✓	✓	✓		
Demand Volume	✓	✓	✓		
Speed		✓		✓	
Arrivals on Red		✓			
Pedestrian Delay		✓			
Purdue Coordination Diagram		✓			
Purdue Phase Termination		✓			
Purdue Split Failure		✓			
Split Monitor		✓			
Turning Movement Counts	✓	✓			
Yellow and Red Actuations	✓	✓			
Level of Service	✓	✓	✓		✓
Travel Time				✓	
Number of Stops				✓	✓
Travel Time Reliability				✓	
Bandwidth Efficiency					✓
Bandwidth Attainability					✓
Progression Opportunities					✓

4.0 Fusion Framework Applied to Signalized Arterial Travel Time

Tracking signalized intersection performance is crucial for agencies in order to identify problematic sites and establish alternative treatments, thereby reducing loss experienced by travelers. Efforts have been made to accurately and reliably estimate performance measures such as travel time by employing various data sources, such as fixed sensors (e.g., inductive loop detectors and surveillance cameras) and mobile sensors (e.g., probe vehicle fleet and Bluetooth units). However, these disparate data sources contain estimation errors that makes it difficult to properly assess the actual performance of signalized arterials. A promising solution to this issue is the fusion of these disparate data sources to achieve a more accurate representation of the signal performance.

This study applies the developed data fusion framework to a widely used signalized arterial performance metric -- arterial travel time. The framework is applied on a dataset from the Next Generation Simulation (NGSIM) study, simulating the availability of various data sources, and enabling a comparison with the ground truth travel time.

4.1 METHODOLOGY

The methodology section is divided into two parts – the fusion framework development and travel time fusion sections. The first part details the development of a generalized data fusion framework for signalized corridors that incorporates disparate data sources with different geospatial and temporal resolutions. That material was outlined in Chapter 3 of this report. The second section details the application of the framework to arterial travel time estimation using different data sources and fusion algorithms. This section is detailed next.

4.1.1 Signalized Arterial Travel Time Fusion

Travel time estimation at signalized arterials is of utmost importance to transportation managers. That measure is considered to be highly variable due to the stochastic effects of signals and is challenging to produce, compared to its estimation on freeways and multilane highways. As such, signalized corridors' travel time estimation stands to benefit the most from the fusion of disparate data sources available to public agencies.

It is expected that the choice of a fusion algorithm will impact the accuracy and reliability of the estimated performance measure because each algorithm is based on different mathematical assumptions and has its individual merits and demerits (*32*). As such, selecting and testing various fusion algorithms and their associated optimization is needed for deriving accurate and reliable output under a given data availability scenario. This study assesses the performance of five algorithms, including a) Simple average, b) Linear regression, c) K-Nearest Neighbor (KNN), d) Random Forest (RF), and e) Artificial Neural Network (ANN), for fusion of travel time information along a signalized corridor. Furthermore, multiple data availability scenarios are explored that shed light on the importance of data types at the agency disposal.

The evaluation of fusion algorithms requires model training, validation, and testing. Among these processes, training and validation are carried out in the same cycle while tests must be conducted separately. As such, test data constituting 20% of the full dataset is

randomly extracted. For the purpose of adjusting hyper-parameters and optimizing models, 10-fold cross-validation is conducted. Since k-fold cross validation is conducted without loss of training data, it is suitable for the small size dataset used in this study.

The following sub-sections provide details of the selected algorithms along with the model tuning processes. Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) are employed for the fusion model evaluation.

SELECTED ALGORITHMS

Simple Average

Simple average is the easiest way to fuse data from multiple sources. The basic assumption here is that the reliability of each data source is identical, and thus the same weights are distributed to all data sources. However, the performance of the model is highly affected by a sensor type that produces systematic under or overestimated performance measurement due. Thus, it is recommended to apply the model only in the event multiple data sources have very similar levels of reliability.

Linear Regression

Linear regression is a straightforward approach for supervised learning. It is a valuable tool for predicting a quantitative response. Linear regression predicts quantitative response Y based on a single or multiple predictor variables X. Mathematically, this linear relationship takes the form:

$$y = \beta_0 + \beta_1 x$$

Where β_0 is the constant or intercept term and β_1 is the regression slope.

Linear regression, by definition, only examines linear relationships between dependent and independent variables. That is, it assumes that a straight-line relationship exists between them. Therefore, it cannot consider non-linear relationships. Since this model does not require hyper-parameter tuning, all training data are used for fitting a regression line, and is evaluated with a test dataset.

K-Nearest Neighbors

The K-Nearest Neighbors (KNN) approach is one of the most extensively utilized data fusion methods (*33*). It enables the construction of a general framework for flexible imputation. It obtains conditional clues on standard variables, which reduces bias while preserving the link between specific and common variables. Furthermore, it can work with any sort of variable (continuous, categorical, or textual). Also, it allows using the same neighbor table for different grafting operations rather than estimating different models for each operation (*34*). It also maintains confidentiality on specific variables because the estimation can be performed independently of the neighbor search. It also assures coherent presumptions to avoid nonsensical results, and it can allow clues from a predictive distribution to duplicate real data, followed by multiple suggestions to account for estimating uncertainty.

KNN uses a lazy and non-parametric learning algorithm for data fusion (*35*). Non-parametric means that no assumptions are made about the underlying data distribution. The dataset determines the KNN model structure, which is extremely useful in practice, as most real-world datasets do not adhere to mathematical theoretical assumptions. A lazy algorithm does not require any training data points for model generation (*35*). In KNN, K is the number of nearest neighbors. The number of neighbors is the core deciding factor. The workflow of KNN is as follows: 1) Calculate distance, 2) Find closest neighbors, 3) Vote for labels.

The number of K is a crucial parameter in the KNN model. If a small number of K is selected, the model relies on a few neighbors' data points, but it can miss essential information in a neighbor in the high number of K (*36*). Thus, a selection of the appropriate number of K is necessary for KNN model. Besides the number of K, the distance unit and weight for distance can be adjusted to improve the model.

Although it is not necessary for KNN to have a separate training dataset, the use of a test dataset for searching for optimal K leads to overfitting in the specific dataset. Thus, cross-validation is conducted with a training dataset in the setting of Euclidean distance, weight for distance, and various K.

Random Forest

Random forest builds multiple decision trees and merges them to get a more accurate and stable prediction. Random forest adds randomness to the model while the trees grow. When splitting a node, it looks for the best feature among a random subset of features rather than the most crucial element. As a result, there is a greater variety, which leads to a better model. The algorithm for splitting a node in a random forest considers only a random subset of the features. Users can make trees more random by using arbitrary thresholds for each element instead of looking for the best possible thresholds (like a normal decision tree does). The hyper-parameters in the random forest are either used to increase the predictive power (number of trees, number of estimators, minimum number of leaves, the maximum number of features random forest considers to split a node) of the model or to make the model efficient (number of iterations, random forest cross-validation).

Among the hyper-parameters mentioned earlier, the number of trees, minimum number of split and leaf, the maximum number of features for split, and the maximum depth of tree were searched with cross-validation to optimize the model through a grid search method.

Artificial Neural Network

An artificial neural network (ANN) is a branch of machine learning (ML) that mimics the functioning of the human brain by processing data with a specified logical structure (*37*). The multiple layers in deep neural networks allow models to learn complex features more efficiently and perform more intensive computational tasks. An ANN model is a feed-forward artificial neural network with more than one hidden layer. ANN models process the information through weighted connections through a series of fully connected layers associated with other layers (*38*). Each node, called a neuron, transforms the input with a nonlinear function to create a decision boundary. Each neuron can be considered a non-linear computational unit that applies an

activation function (e.g., sigmoid, Exponential Linear Unit (ELU), and Rectified Linear Unit (ReLU) function).
Each neuron can be defined as:

$$a^{l+1} = f(W^l a^l + b^l)$$

Where a^l and a^{l+1} denote the activation value in levels l and l+1, respectively, W^l is a weight matrix, b^l is the bias, and $f(.)$ represents the activation function. The particular case is l=1, which denotes the input layer where $a^l = x$.

ANN mainly consists of seven parameters 1) Optimizer 2) Learning rate 3) Initializer 4) Dropout rate 5) Number of neurons 6) Batch size and epochs, and 7) Activation function. For an efficient ANN model, the parameters need to be hyper-tuned. For finding the optimal hyperparameters, cross-validation and grid search method were used.

4.1.2 Study Site and Data Preparation

Transportation operating agencies have access to various traffic data such as vehicle counts, speeds, and travel time collected by single or multiple available sensor types. Some agencies have access to probe vehicle data through a Software-as-a-Service (SaaS) application provided by third-party vendors such as INRIX (*39*) and RITIS (*40*). Although the transportation data spectrum is expanding day by day and novel technologies such as drones and Lidar are now being employed, commonly available travel time data sources available to transportation agencies are a) probe vehicles, b) Bluetooth sensors, c) loop detectors, and d) signal timing data.

Knowledge of ground truth travel time is an essential part of any data fusion practice, without which a robust assessment of the performance of the fusion algorithms is impossible. As such, this study uses on of the widely knowns NGSIM dataset to apply the developed fusion framework and assess the performance of the selected fusion algorithms. The NGSIM data used in this study was collected in Atlanta, GA, in 2006 for two 15-minute periods (e.g., 12:45-1:00 PM and 4:00-4:15 PM). The dataset includes vehicle trajectories at 10 Hz resolution for the half-hour noted above. Of the needed datasets, ground truth travel times and signal timing are directly available. The other employed datasets (Bluetooth, fixed-sensor, and probe) were simulated using information from the vehicle trajectories.

The arterial performance measure selected was travel time in both the northbound and southbound directions of the corridor. **Figure 3** shows a schematic drawing of the site along with approximate (virtual) sensor locations. Considering data aggregation interval and time resolution provided by a SaaS company, 1-minute interval was deemed to be a reasonable aggregation level. In this process, travel times are independently estimated without data leakage between data sources considering single available data source conditions.

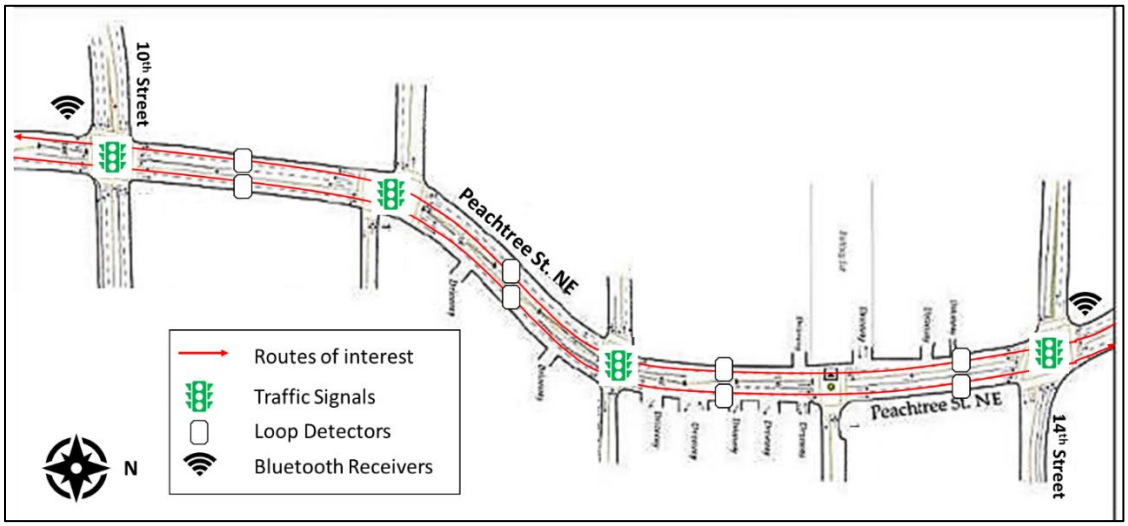


FIGURE 3: SCHEMATIC DRAWING OF THE STUDY SITE ALONG WITH THE SENSOR LOCATIONS

The following sections describe each of the data sources and provide details of the assumptions made in simulating the Bluetooth, Fixed-sensor, and Probe Vehicle datasets. The evaluation of travel time from a single data source is also presented.

Ground Truth Travel Time

Ground truth travel time is arguably the most crucial dataset in a data fusion practice since it is needed to gauge the fusion framework's and algorithms' performance. Vehicles traveling the entire length of the corridor were extracted to calculate the ground truth travel times for the two directions in each minute. This is achieved by subtracting the end travel time from the start travel time and assigning it to intervals based on their start clock time. Table 3 describes aggregated observational ground truth data by direction and time period.

TABLE 3: NUMBER OF OBSERVATIONS AND AVERAGE TRAVEL TIME BY DIRECTION AND TIME PERIOD

	Number of observations (veh)	Average Travel Time (sec/veh)
Northbound Period 1	82	129.48
Northbound Period 2	67	140.17
Southbound Period 1	58	105.6
Southbound Period 2	73	131.97

Probe Vehicles

Probe vehicle data is usually collected by third-party companies that hold intellectual property rights on the algorithms that collect those. Thus, the travel time estimation method attempted to simulate probe vehicle data here is based on publicly available information.

In this study, probe-based travel time estimation starts by dividing the corridor into multiple segments, similar to the TMCs (Traffic Message Channel) that a SaaS company distributes data based on. Since the company generally defines a TMC boundary at

intersections where traffic and signal condition change, each intersection is defined using the segment boundary, consistent with the data that are obtainable by the agency.

Presumably, a deterministic travel time for a specific segment and interval can be estimated using two methods. One is to calculate the difference between the entry and exit time of the segment for each probe vehicle and take the average travel time. The other estimates the average speed rover a given segment and converts it to average travel time by dividing the distance by the average speed. The first method is clear and straightforward, while the travel time can be underestimated by the probe vehicles entering from or exiting to a driveway in the middle of the segment. In order to minimize this bias, the second method is selected to estimate each segment's travel time in this study.

Probe vehicle sample size is another important matter to consider. This paper assumes that the probe vehicle sample size is 10% with a random selection method that considers vehicles with at least one segment travel. **Figure** 4 shows the heatmap of speed for the randomly selected 10% probe vehicles extracted on the Northbound route for the first data collection period along with the population presented by gray points.

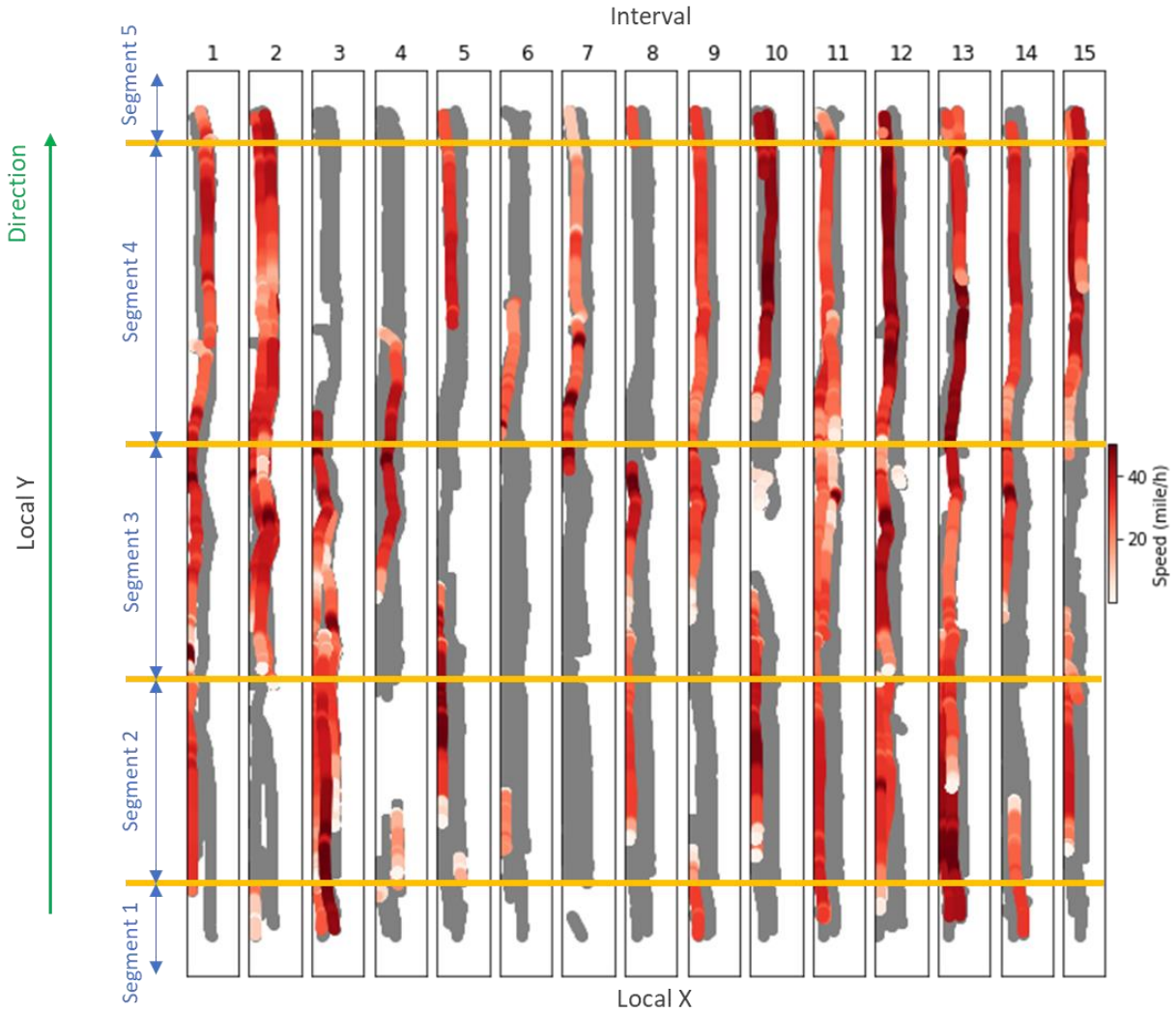


FIGURE 4: POPULATION AND 10% SAMPLE PROBE VEHICLE SPEEDS FOR NB TRAFFIC

Data points from randomly selected trajectories are grouped by time interval and segment, and speed for each vehicle is estimated by averaging the instantaneous speeds. The average speeds are then converted to travel time based on the length of each segment. In this step, the team defined travel times in a segment that exceed two-cycle lengths in duration as outliers and excluded them from further consideration. After excluding outliers, the average travel time for each segment and interval is estimated and the travel time for the same interval is combined to compute route travel time. However, there were missing data for some specific segments and intervals, as evidenced in **Figure 4.** In real world SaaS applications, missing data is substituted with time of day and day of week historical data. In this study, we used interpolated values to replace the missing data due to a lack of historical datasets for the NGSIM site.

Bluetooth

Bluetooth is a suitable medium for measuring travel time on signalized intersection corridors. Since it records the Media Access Control (MAC) ID of Bluetooth-enabled devices in a vehicle along with time, travel times for sampled vehicles are collected by matching the ID and calculating time differences. However, a Bluetooth receiver has inevitable location detection errors. In this study, the detection error is assumed to follow a normal distribution with zero mean and 50 ft. standard deviation based on empirical data (*41*). Another assumption made in simulating the Bluetooth dataset is the market penetration rate or sample size of detected MAC IDs. Even though the detection rate varies by site, this study assumes a 10% rate of reliable travel times, reflecting favorable detection rate conditions.

Based on these assumptions, a 10% observation sample is randomly selected among vehicles traveling the entire corridor in both directions, and recording their timestamps at each Bluetooth receiver. Data are collected considering the detection error. The measured travel time for each sample is then assigned based on its travel start time, and the average travel time for each interval is calculated. Missing travel times caused by having zero samples in a specific time interval are interpolated, similar to what was done with probe vehicle data.

Loop Detectors

Loop detectors are the most widely used sensors for collecting traffic data on roadway networks. Thus, many researchers have attempted to utilize traffic counts and speeds collected at those sensors for estimating various performance measurements such as delay, travel time, and traffic status. In terms of travel time estimation various methods were developed specifically for signalized arterials (*42–44*). However, these methods require supplemental signal data to estimate accurate travel times, in order to account for the effect of signals on travel time in interrupted facilities. Since traffic signals are a primary source of delay on urban and suburban arterials, signal plans effects must be considered in loop-based travel time estimates. The team recognizes the serious drawbacks of loop detector-based travel times in the absence of signal effect consideration. Nevertheless detector data is exclusively used for travel time estimation in this study, in order to replicate the situation when only loop detector data are available, as well as to avoid data leakage before fusion.

The travel time estimation method is based here on the average speed measured at fixed points. The average speed measured for a given interval at a loop detector is converted to travel time based on the distance of the corresponding segment on which the sensor is located. The travel time estimated for each segment is combined to match the spatial assessment level with travel time from other sources.

Signal Data

As described in the previous section, signal data has been used as a supplemental source in travel time estimation method at signalized arterials. However, signal data may be the only available data source at some sites. Thus a novel travel time estimation method for signalized arterial using signal data is developed and implemented in this study.

The proposed method estimates the travel time for a vehicle traveling the route with the assumption of uniform arrival probability at the start of the route. The method is based on virtual trajectories of vehicles traveling from the start to the end of the arterial. The vehicle generation rate at the start of the route is governed by the average headway between vehicles (number of vehicles / hour). Free flow speed upstream of the route, an empty system at the beginning of the first period, and no existing queues are the other assumptions driving this method. **Figure 5** provides a visual illustration of one vehicle and its trajectory. In the figure, effective green and red times for through movements at each signal and the trajectory for the first virtual vehicle are presented. In effect, knowledge of the interval-varying: through green splits, cycle length and offsets in both directions can assist in generating these virtual travel times in each time period.

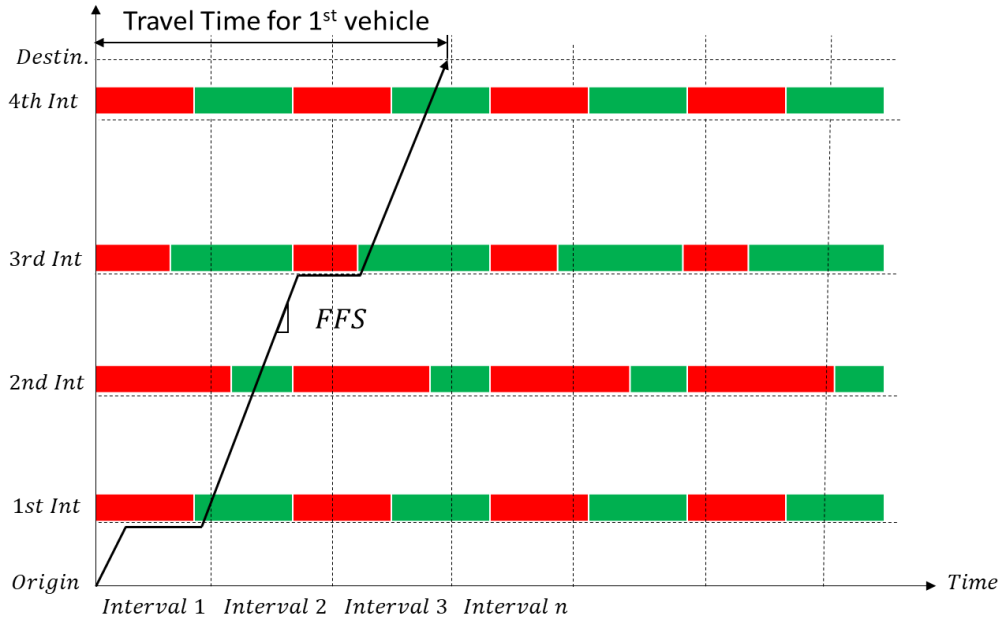


FIGURE 5: VISUAL ILLUSTRATION OF TRAVEL TIME FOR A VEHICLE USING SIGNAL DATA AND HOURLY VOLUME

The virtual trajectories are repeatedly generated at a predefined rate without consideration of the previously generated trajectory. Travel time for each virtual trajectory is then grouped based on the travel start time and averaged for a given interval.

4.2 ANALYSIS AND RESULTS

This section gives the results of the application of the data fusion framework using different algorithms. Travel time estimation using single source estimates is presented first.

4.2.1 Performance Evaluation of Individual Data Sources

As described in previous sections, travel time along the arterial is independently estimated using simulated or measured data from four different sources. **Figure 6** presents these travel times along with the ground truth values. Ground truth travel times (blue lines) range from 85 to 170 seconds on the 2030 feet travel distance. No time-dependent pattern is identified, which is common in time-sequential travel time data.

Bluetooth travel time (shown by the green lines) shows a similar pattern as the ground truth and is potentially closer to it than the other estimates. This may be because Bluetooth sensor travel time is measured for almost the same spatial range as the ground truth. However, the smaller sample size, the inclusion of turning movement vehicles, and missing data interval add to the variance.

Travel time from probe vehicles (indicated by the orange lines in **Figure 6**) shows significant fluctuations in both periods and directions. While probe-based travel time is estimated by summing travel times on all the segments in the same time interval, ground truth is measured by differentiating travel's start and end time. In other words, the travel time from probes cannot take account of the time lag that the actual vehicles experience while traveling the corridor. Thus, the difference in travel time definition would primarily worsen the discrepancy between ground truth and probe vehicle. In addition, some travel time estimates in the second period are significantly higher than ground truth. This is induced by lower average speed of the probe vehicles because speeds measured during a specific interval and in a segment can only include significantly low ones depending on the selected samples as evidenced in segment 1 and interval 4 in **Figure** 4.

Loop detector and signal data (red and purple lines in **Figure 6**, respectively) significantly underestimate travel times. The loop detectors are located in midblock locations where vehicle speeds are usually high and the travel time between the sensors does not include the deceleration, stop time and acceleration delays at the signal. Similarly, the tendency of systematic underestimation in signal-based travel time results from the absence of acceleration, deceleration, and initial queue delays. Interestingly, signal data somewhat captures the ground truth's pattern despite its systematic underestimation of travel time. This underscores the potential of signal-based travel time as an independent data source. In addition, the accuracy of the dataset can be improved by including some of the missing delay components outlined above.

Since four datasets for two different directions and periods are independent and no time-dependent pattern is recognized, they are combined into a dataset for the following application of fusion algorithms. This approach not only increases data size for training fusion

algorithms, but also generalizes them regardless of direction and period. With the combined dataset in hand, the performance of single source datasets is measured as shown presented in **Table 4** using MAE, RMSE, and MAPE. The best and worst-performing datasets were identified as Bluetooth and loop detector datasets, respectively.

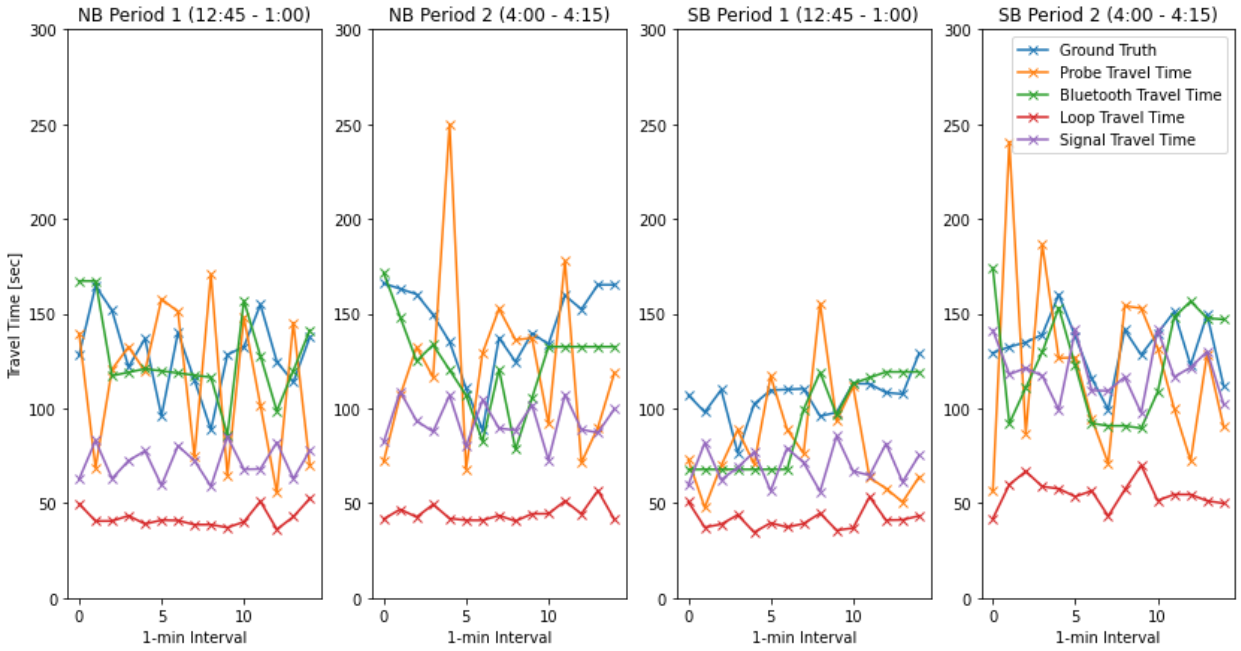


FIGURE 6: GROUND TRUTH AND ESTIMATED TRAVEL TIME DATA BY VARIOUS SENSORS IN THE NB AND SB DIRECTIONS OF THE ARTERIAL DURING TWO PERIODS

TABLE 4 EVALUATION OF TRAVEL TIME ESTIMATES BY DATA SOURCE

Virtual Data Sources	MAE (sec)	RMSE (sec)	MAPE (%)
Bluetooth	21.36	25.84	17.12
Probe Vehicles	40.68	48.79	32.3
Signal Timing Data	40.82	46.69	31.24
Loop Detectors	81.99	84.61	63.41

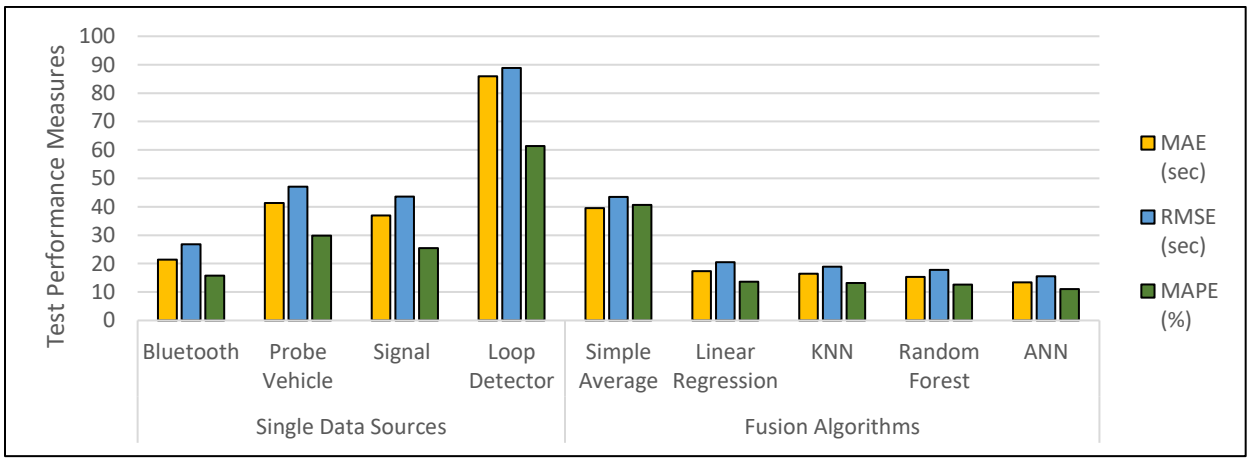
4.2.2 Performance Assessment of the Fusion Framework and Algorithms

The developed fusion framework was applied to the available datasets using the five different fusion algorithms. As described earlier, the team randomly divided the combined dataset into training and test dataset with the ratio of 8 to 2 in this process. The performances of the fusion algorithms were measured by comparing the outputs with the ground truth travel times in the test dataset. First, the data fusion framework was applied assuming all four data sources are available to the operating agency. The performance of the applied fusion algorithms in terms of MAE, MAPE, and RMSE is shown in **Figure** 7.

Figure 7 organizes the performance for various single data sources and fusion processes using the identified algorithms. The fusion process almost consistently outperforms the single source dataset according to the performance metrics identified. The simple average algorithm is the only fusion method that was unable to exceed the performance of the single source travel time data.

Compared to the best performer of single source data (i.e., Bluetooth), the fusion process yielded 19%, 23%, 28%, and 38% improvement for the linear regression, KNN, Random Forest, and ANN, respectively. On the other hand, the simple average method shows poor performance compared to the some single source datasets. Simple average's poor performance stems from the fact that equal weight has been given to all the data sources without consideration of their equivalent utility to the fusion process.

FIGURE 7: COMPARING THE EVALUATION METRICS FOR SINGLE DATA SOURCES AND FUSION ALGORITHMS



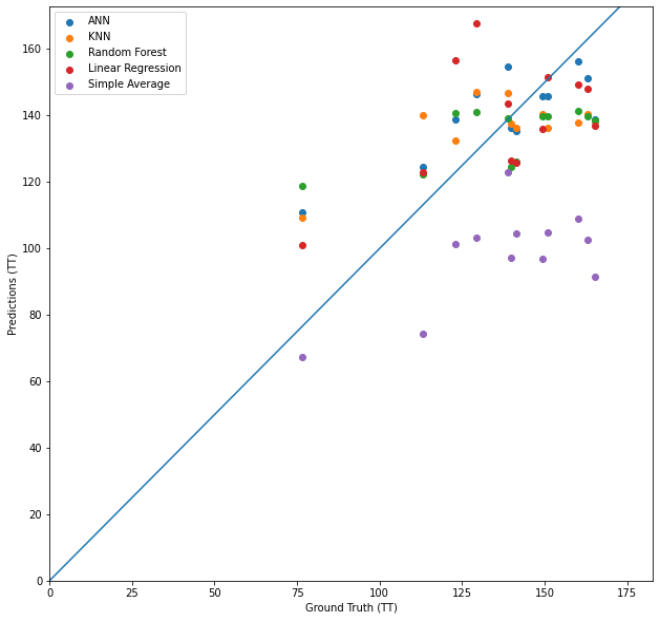


FIGURE 8: GROUND TRUTH VS PREDICTION FROM VARIOUS FUSION ALGORITHMS

Figure 8 shows the ground-truth travel times compared to the five algorithms' fusion outputs. The closer the points are to the diagonal line, the better the fusion process. It is clear that the fusion process using the simple average algorithm is the farthest from the diagonal line compared to other fusion processes. This aligns with the performance measures presented in the previous figure. The outputs of the remaining fusion processes are distributed near the diagonal line. The ANN outputs especially appear to be closest to the diagonal line compared to the remaining fusion outputs.

In addition, the sensitivity of the fusion performance to a varying set of dataset availability was explored where one or more of the datasets is not available. The fusion detailed in the previous section assumed the most favorable condition where multiple data sources are available. However, this favorable condition is rare in reality, where an agency may have access to no more than two data sources. Thus, the robustness and sensitivity of the fusion process to data availability are explored by developing multiple scenarios of data availability. In each scenario, the model is optimized using the same method outlined in the methodology section of this report. MAPE is used to compare the impact of data availability across the scenarios. Observation from Table 5 shows that, the availability of data impacts the output of the fusion process. Generally, and as expected, the more datasets are available the more accurate the outcome of the fusion process becomes, with the exception of the simple average case. In most scenarios, ANN yielded better performance than other algorithms, producing MAPE values ranging from 10.6 % to 16.0%.

Another finding is that the fusion accuracy is significantly related to the available data characteristics. Note that probe and signal data exhibited similar variance. However, the mean of probe data is closer to the ground truth than the signal data, thus pointing to a systematic error in the signal dataset. When scenarios are compered in which the two data sources are swapped, the inclusion of signal data actually yielded better performances. This implies that fusion algorithms somewhat mitigate certain systematic errors by optimizing parameters in the models. Nevertheless, the level of mitigation appears to be different between signal and loop detector data. Although both data sources had a tendency to underestimate travel time, the signal dataset had a *similar trend* with the ground truth. Therefore, the systematic error in loop detectors appears to not be well addressed in fusion algorithms, and a relatively worse performance is produced when the loop data is included in fusion processes.

As evidenced in **Figure 6**, loop detector and signal-based travel times systematically underestimated the ground truth travel times along the corridor. Based on the limitations identified for these data sources, they can be viewed as complementary and the use of their combination is thought to benefit the overall performance of the data fusion process. Thus, the team attempted to generate a single travel time estimate from both data sources, and its value is fused with other mobile sensor-based travel time datasets.

Table 5 Travel Time MAPE: Effect of Data Availability and Fusion Algorithm

Scenario No.	Available Data Source*	Simple Average Estimate	Linear Regression Estimate	KNN Estimate	Random Forest Estimate	ANN Estimate
1	B+P+S+L	40.7	13.66	12.67	13.03	**11.61**
2	P+S+L	59.63	**14.49**	15.58	14.73	14.59
3	B+S+L	46.46	13.58	13.1	13.42	**12.55**
4	B+P+S	22.61	13.67	12.01	11.58	**10.60**
5	B+P+L	42.68	13.68	13.65	**13.48**	15.18
6	B+P	18.53	14.39	13.37	12.88	**12.18**
7	B+S	25.29	13.65	12.86	12.57	**11.75**
8	B+L	53.18	**13.72**	15.71	14.45	14.71
9	P+S	35.98	14.77	15.15	**14.47**	14.64
10	P+L	79.59	**15.63**	16.26	16.25	16.01
11	S+L	79.83	14.71	14.95	14.61	**14.11**

* Bluetooth (B), Probe (P), Signal (S), and Loop Detector (L); **Bold** numbers show the values closest to the ground truth travel time in each scenario

The estimation method applied here shares the concept of signal-based travel time estimation described earlier in this chapter. Thus, virtual trajectories are generated at the origin of the routes. In this process, the generating rate is the flow rate measured in the first loop detector on the route. In the next step, delay polygons are constructed at each signalized intersection using the flow rate measured in each loop detector, and assuming default saturation flow rate as described in HCM 6th (*16*). Once the flow rate and delay polygons are generated, each trajectory is constructed and the time delay that is the horizontal line within each delay polygon encountered while traveling is added to free flow travel time. The concept of the method is illustrated in Figure 9.

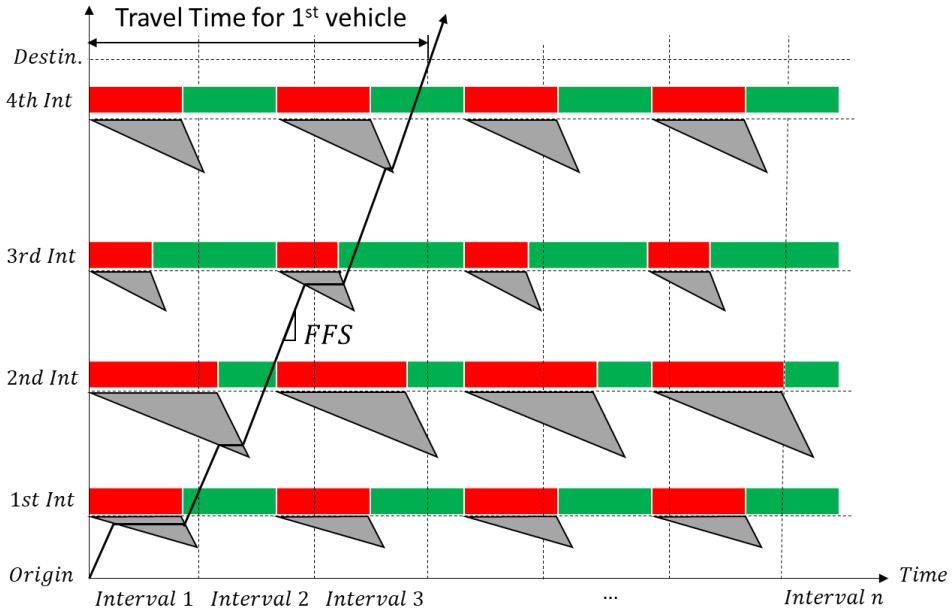


FIGURE 9: CONCEPT OF TRAVEL TIME ESTIMATION FROM LOOP DETECTOR AND SIGNAL DATA

The average travel time for trajectories in the same interval and direction is computed to estimate each interval's travel time. The generated travel times using this methodology is presented along with other data sources in Figure 10. Although the combination of loop detector and signal dataset did not completely eliminate the underestimation problem (as shown in the NB direction in time period 2), the method did produce more reliable travel times compared to standalone signal and loop detector datasets.

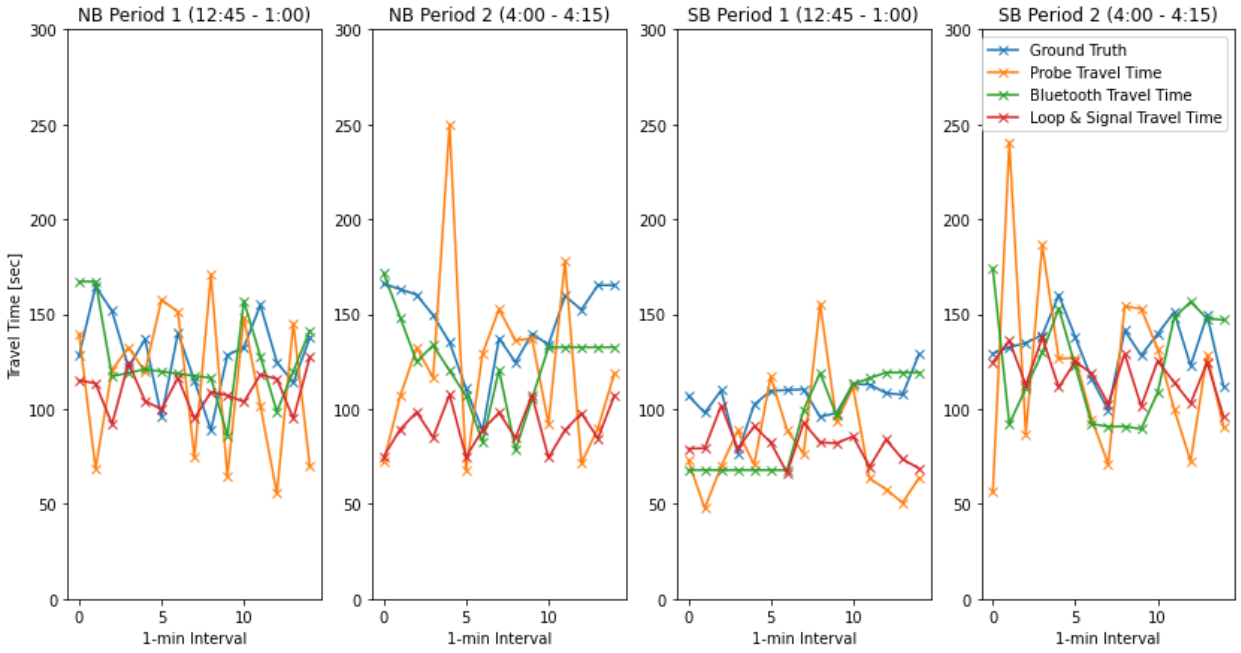


FIGURE 10: LOOP DETECTOR AND SIGNAL TRAVEL TIME ESTIMATES WITH OTHER SOURCES FOR NB AND SB DIRECTIONS OF THE ARTERIAL DURING THE TWO PERIODS

With the new estimates in hands, fusion algorithms were once again applied for various data availability scenarios. The MAPE's for each algorithm and scenario are presented in **Table 4.** In all scenarios, ANN outperformed the other algorithms. A comparison of the four scenarios found in Table 6 with those in Table 5 indicate a marginal improvement in travel time estimation for some of the algorithms, and a slight degradation in others. The simple average method stands to benefit the most from the combination of loop detector with signal data, while linear regression loses some predictive power as a result of this combination – rendering the impact of the initial amalgamation insignificant and unnecessary.

Table 6 Evaluation of Fusion Algorithms in Various Data Availability Scenarios by Combining Signal and Loop Detector Data For Travel Time Estimation

Scenario No.	Available Datasets	Simple Average	Linear Regression	KNN	Random Forest	ANN
1	B+P+(S+L)	23.77	13.92	13.15	12.33	**11.52**
2	P+(S+L)	39.15	14.77	15.21	15.18	**14.18**
3	B+(S+L)	22.57	13.88	13.29	12.24	**11.77**
4	B+P	18.53	14.39	13.37	12.88	**12.18**

5.0 Prototype Testing for Signalized Arterial Performance Measures

Traffic signals, though a very critical part of today's traffic control system, are also a major contributor to chronic congestion and safety hazard on roads if not operated efficiently. In the United States, there are more than 330, 000 traffic signals according to the United States Access Board (*1*). National reviews of traffic signal systems have found that many signal systems are not performing up to standard due to aging control infrastructure and lack of proper assessment. In addition, the lack of tools and frameworks to assess traffic signals based on actual phasing and detector data has been a notable limitation. Consequently, changes in time-variant traffic demand in terms of overall traffic volume, peaking time, and arrival patterns are not incorporated in their assessment. Hence, it is important to construct an advanced framework to evaluate the performance of signals based on the measured traffic signal and detector data.

Automated Traffic Signal Performance Measures (ATSPM) is one such framework. It collects high-resolution signal and detector data and generates performance metrics for approaches at signalized intersections (*1*). These metrics include but are not limited to total and average delay, split failure, and yellow and red actuations. These metrics help improve decision making and the selection of more effective operational strategies to reduce congestion and crash risk. However, there are important limitations in the ATSPM methods for generating delay and other performance measures. The approach delay reported by ATSPM does not account for acceleration, deceleration, and queueing delays. Moreover, the delay estimation process does not account for any initial queue and cycle failures. Hence, it is very likely that ATSPM's reported delays underestimate the true delay, thus prompting the need for further algorithmic improvements.

In this study, two algorithms have been developed to estimate approach delay at a signalized intersection using high-resolution signal and detector data. Each of these algorithms addresses at least one limitation in the current ATSPM delay estimation method. High-resolution traffic data from the Utah Department of Transportation were provided for two corridors for the year 2018 (*2*). The signal phase and detector actuation data were fused, and fundamental traffic flow relationships were applied to develop these methods using some reasonable assumptions. The two methods and the existing ATSPM delay estimation methods were applied to a through movement approach at a signalized intersection located at the 700 East Corridor in Salt Lake City (SLC), Utah. Delay estimated by all the three methods are compared and its variation by time of day are portrayed.

5.1 METHODOLOGY

5.1.1 Method 1: Arrival-Departure Approach

Basic Principles of Incremental Queue Accumulation (IQA)

The Incremental Queue Accumulation (IQA) method, adopted in the 2016 Highway Capacity Manual (*16*), estimates the length of the queue using arrivals and departures information in short time steps. The method adds and subtracts vehicle arrivals and departures during each time step to and from the queue. Queue length is therefore the difference between the cumulative vehicle arrivals and departures. Time steps are defined whenever a vehicle arrives or departs, thus designating a change of status in the size of the queue. Hence, the total delay in each cycle is calculated by summing all the products of queue size and the time steps. The corresponding average vehicle delay per cycle can be computed by dividing the total delay by the number of vehicles departing within the cycle. As such, the IQA method computes the area of trapezoids representing the total time during the cycle where the inflow and outflow rates are not equal.

Application of Arrival-Departure method to the ATSPM Data

Prior to application of the arrival-departure method to the ATSPM data, it is necessary to identify the type of data available for use. The arrival-departure method would provide ideal results if vehicle trajectory data is available. ATSPM data, however, include only vehicle observation records that are detected at the upstream (usually 400 ft from the stop-bar) and at stop-bar detectors. As such, for the arrival-departure method, it is essential to connect the upstream observations to those at the stop-bar detectors.

The arrival-departure method assumes stable operations of the traffic signal (i.e. the departure curve catches up with the arrival curve prior to the green termination). Hence, there will be no queued vehicles present at the end of the green phase. We further assume that arrivals and departures obey the FIFO (first-in-first-out) principle. This means that vehicles depart from the stop bar in the same order they arrived at the upstream detector. These assumptions enable the analyst to synchronize the arrivals at the upstream detectors to the departures from the stop bar.

Using the refined data and the associated assumptions, a cyclic vehicle arrival-departure curve can be drawn as shown in **Figure 11**. In this figure, the orange and blue lines are vehicle arrivals at the upstream and stop-bar detectors, respectively. The horizontal time gap between the two lines reflects the travel time of vehicles between the two detectors. To estimate the delays for individual vehicles, the expected departure time under free-flow conditions is calculated by adding the free-flow travel time to the observed arrival time (see **Figure 11**) to estimate the vehicle arrival times at the stop-line. Free flow travel time is estimated for each vehicle as described below:

$$T_{FF} = \frac{L}{Speed\ Limit} \qquad \text{Eq. (1)}$$

where, T_{FF} is the free flow travel time in seconds, L is the distance between upstream and downstream detectors (ft) and Speed Limit is the posted speed limit at the site (ft/sec).

Thus, delay can be computed by subtracting the expected free-flow departure time from the observed departure time as shown in Figure 3. The total experienced delay within a cycle can be estimated by computing the area between the arrival and departure curves. The average experienced delay per vehicle can be calculated by dividing total delay by the number of departures. Total and average cycle delays are estimated by first calculating delay during any time slice Δ_t, in which delay calculations are made.

$$d_{\Delta_t} = [A(t) - S(t)]\,\Delta t \qquad \text{Eq. (2)}$$

Defining T_Q to be the point at which $d_{\Delta_t} = 0$, total delay per cycle can be estimated by:

$$D = \sum_{t=0}^{T_Q} [A(t) - S(t)]\,\Delta t \qquad \text{Eq. (3)}$$

$$Average\ Delay\ per\ Cycle = \frac{D}{N} \qquad \text{Eq. (4)}$$

Where A(t) is the cumulative arrival to time (t), S(t) is the cumulative departure to time (t), D is the total cycle delay, N is total number of arriving or departing vehicles within a cycle.

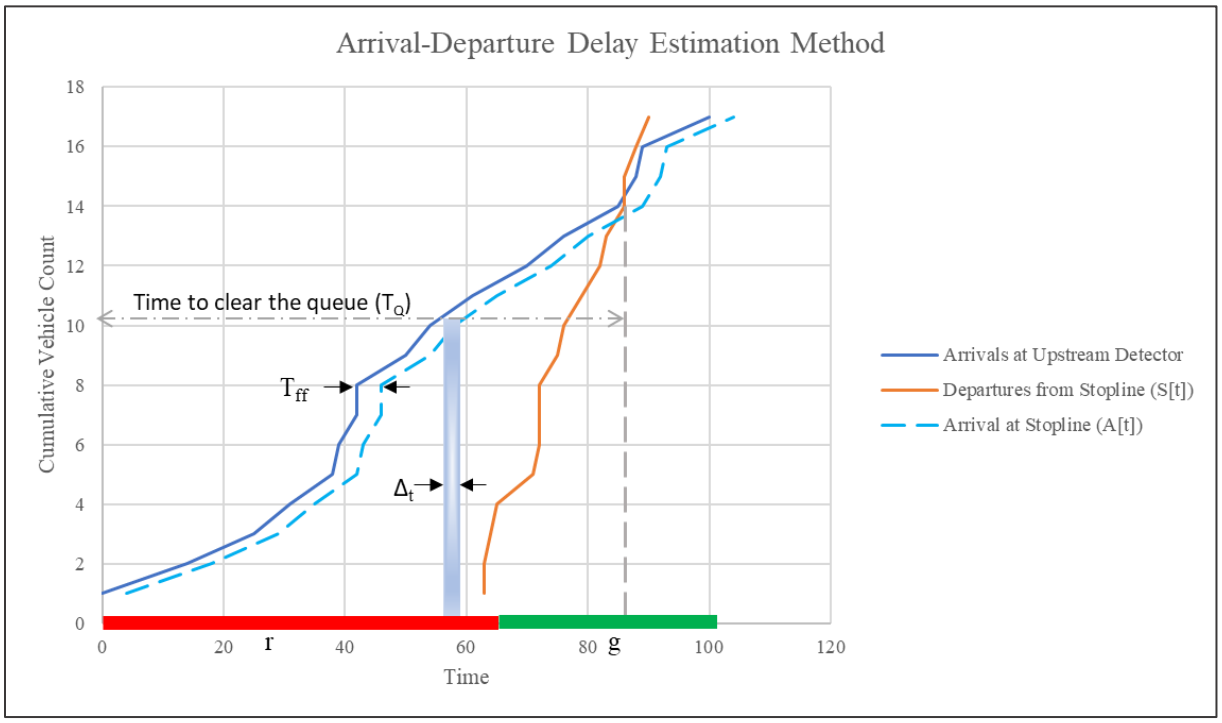


FIGURE 11: CUMULATIVE ARRIVAL – DEPARTURE CURVES IN A CYCLE

5.1.2 Method 2: Delay Based only on Departures Only

This method primarily employs actuation data only from lane-by-lane detectors located at or just downstream of the stop line of an intersection approach. It leverages the changes in the

headway of departing vehicles when all queued vehicles are served, and the un-queued vehicles start arriving and departing with no delays. Two thresholds are implemented to classify queued and un-queued vehicles, which enables the estimation of total delay per cycle and average delay per vehicle based on reasonable estimates of arrival volumes. The following subsections present the underlying assumptions, input data requirement, and the step-by-step process to apply the method.

Assumptions

1. Vehicles departing from a queue are likely to have inter-departure times or headways that are close to the saturation headway.
2. A large gap between two successive departures indicates that the queue *may* have cleared, and the remaining vehicles will not experience any delay.
3. The arrival rate is assumed to be uniform during the green and red periods, except for some special cases that are described separately.
4. Multiple lanes of a movement group operate independently, and their delay can be aggregated at the movement level.
5. There is no residual queue at the start of the first cycle of the analysis period.

Inputs

From the start to the end of each cycle (start of red and end of the green period, respectively) the phase time information and actuation (detector "on") events from the stop bar detectors are archived. Only for a special case described in the latter part of the section, it needs the input of the upstream detectors.

Process

In the subsections below, the steps required to apply this method to each lane of a movement group are described. Following this, the applicability of this method to some special cases are discussed.

Step 1: Queued and Un-queued Vehicle Identification: Figure 12(a) shows a typical cumulative departure plot for a lane within a cycle obtained from a stop bar detector. It is evident that, had the first vehicle departed from a queue, the queue would have cleared with the departure event circled red. That and its preceding vehicles departed from a queue (boxes marked "Q") and the following ones departed in a free-flow condition (boxes marked "UnQ"). However, this classification of queued and un-queued vehicles is not trivial because the saturation headway for queued vehicles is a stochastic parameter. To address this issue, a moving average of headways of the departing vehicles are calculated and compared against a threshold.

In this regard, the first task is to identify whether the first vehicle arrived during green or red. One way to do so is to compare the headway for the first vehicle against a threshold value. The headway of the first departing vehicle is estimated as the time difference between the start of the green time and its departure time. A large value for this headway would indicate that the

first vehicle (which is possibly in a platoon) arrived during green and all the following vehicles in that lane did not experience any delay in that cycle. Note that a cycle ends with the end of its green period. Mathematically, this condition is expressed as:

$$h_1 > t_1 \rightarrow ((v_1, v_2, \dots v_N) \in U) \Lambda (v_1 \in Q)$$

Here, h_1 = time headway of the first departing vehicle from a lane, t_1 = A threshold for h_1 to check whether the first vehicle arrived during green or red, $v_1, v_2, \dots v_N$ = Sequence of departing vehicles; N is the total number of departing vehicles from a lane within a cycle, and Q and U are the sets of queued and un-queued vehicles in a lane, respectively.

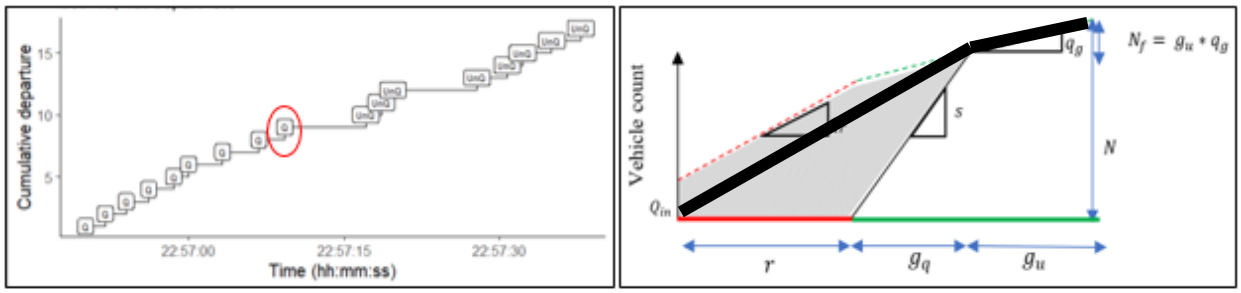
The second task is to identify the last queued vehicle (v_q), which is marked by the red circle in Figure 12(a). To do so, a moving average of headways of the departing vehicles are calculated. Then, the time difference between the headway of a vehicle and the moving average headway of its leader is compared against another threshold (t_2). A vehicle v_i is tagged as the last queued vehicle if that difference is greater than t_2 for that vehicle.

$$h_i - \bar{h}_i > t_2 \rightarrow i = N_q$$

Where Nq is the total number of queued vehicles. $\bar{h}_i$, the average headway of the first (i-1)th vehicles is estimated as follows

$$\bar{h}_i = \frac{h_1 + h_2 + \cdots + h_{i-1}}{i-1} \qquad \text{Eq. (4)}$$

The selection of these two thresholds is very critical to the accuracy of the proposed method of delay estimation. The major issue with these thresholds is that these are very sensitive to any latency in the detector activation. Moreover, the presence of a queue from a closely spaced downstream intersection may affect the selection of this threshold (*45*). The method will then drastically underestimate the delay. If such issues are apparent, a minimum delay criterion is introduced toward the end of this section.



(a) (b)

FIGURE 12: DELAY ESTIMATION USING ONLY DEPARTURE DETECTOR DATA (A) CUMULATIVE DEPARTURE OF QUEUED AND UN-QUEUED VEHICLES FROM A LANE DURING THE GREEN TIME FOR THAT MOVEMENT (B) ESTIMATION OF ARRIVAL RATES DURING RED AND GREEN

Step 2: Arrival Rate Estimation: The next step involves calculating the arrival rate during the red and green period for a lane within a cycle. Figure 12(b) shows the estimated cumulative arrival and departure of vehicles by the dashed and solid lines, respectively. The departure line can be

drawn based on the known departure event data, where the dashed lines' slopes need to be estimated. The cycle length (C), red (r), and green time (g) are assumed to be known from signal timing data. The saturation flow rate(s) can be calculated as the inverse of the average headways for all the queued vehicles identified in Step 1.

If N_q is the number of queued vehicles, then the time to clear the queue from the start of the green (g_q) and the remaining part of the green time (g_u) are estimated using the following equations.

$$g_q = h_{av} + \sum_{i=1}^{N_q} h_i \quad \text{Eq. (5)}$$

$$g_u = g - g_q$$

Where h_{av} is the arithmetic mean headway of the queued vehicles. The arrival rate during green (q_g) is estimated using the following equation.

$$q_g = N_f / g_u \quad \text{Eq. (6)}$$

Where N_f = count of vehicles departing from an un-queued condition. The arrival rate during red (q_r) is estimated using the following equation:

$$q_r = \max\left(\frac{N - Q_{in} - q_g * g}{r}, 0\right) \quad \text{Eq. (7)}$$

Where Q_{in}= initial queue or the residuals from the preceding cycle. Note that Q_{in} = 0 for the first cycle in the analysis period or for any cycle that follows an under-saturated cycle. The estimation of Q_{in} for a cycle following an oversatured cycle is discussed in the "Special Cases" section.

Step 3: Performance Measure Estimation: The shaded area in Figure 12(b) represents the total delay within that cycle, which is estimated using the following equation

$$Delay\ per\ cycle, D = \left(\frac{1}{2}\right) * r * (r * q_r + Q_{in}) + \left(\frac{1}{2}\right) * q_r * r * g_q \quad \text{Eq. (8)}$$

Delay per vehicle and percentage arrivals on red can be estimated as:

$$Delay\ per\ vehicle, d = \frac{D}{N} \quad \text{Eq. (9)}$$

$$AR\% = (q_r * r)/N * 100 \quad \text{Eq. (10)}$$

The above performance measures are calculated for each cycle and each lane of a movement, then averaged over multiple lanes, and weighted by the estimated arrival volumes served in each lane.

Special Cases

The method described above – specifically the ability to estimate variable arrival rates in the red and green phases, will not work if the cycle is oversaturated or if there are no un-queued vehicles. However, with some additional, rational assumptions, the analyst can provide sensible estimates of delays and overflow queues for such cases as described in the next section. Finally, the assumption of uniform arrivals during the red period may or may not be correct in all cases. In other words, if the queued vehicles in a cycle had arrived early in the red period, then the delay estimated by this method will underestimate the true delay. Conversely, if most of the red arrivals occurred in the later part, then the method may overestimate delay. Also as mentioned earlier, any latency in the detector or queue from a downstream intersection may significantly underestimate the delay.

Oversaturated Cycle: If a cycle is over-saturated, then the true traffic demand will be unknown from the stop bar detector data. In such cases, the subsequent cycles should be scrutinized until an under-saturated cycle is found. Then, the total departing vehicles would represent the total demand throughout all those cycles. Assuming that the arrival rate (q) is uniform throughout all such cycles, the delays for each cycle can be estimated using the following equations.

$$Average\ arrival\ flow\ rate, q = N'/(t_2 - t_1) \quad \text{Eq. (11)}$$

Where N' is the total vehicle departing between the start of the first oversaturated cycle (t_1) and the end of the first under-saturated cycle that follows (t_2). For each oversaturated cycle, the overflow queue size, Q_o is estimated as:

$$Q_o = Q_{in} + qC - sg \quad \text{Eq. (12)}$$

Where Q_{in} is the size of the initial queue at the start of each cycle. The delay for each cycle is then estimated as:

$$D = 0.5 * q * r * C + 0.5 * (Q_{in} * r + Q_o * g) \quad \text{Eq. (13)}$$

Under-saturated Cycle with No Un-queued Vehicles: From Figure 12(b), it is apparent that should there be no vehicle departing from an un-queued condition (i.e., $N_f = 0$), the arrival rates during green and red cannot be estimated using Equation (6) and (7). In such cases, we assume that the arrival rate is uniform from the start of the cycle to the end of g_q. The average arrival flow rate is estimated as:

$$q = (s * g_q - Q_{in}) / (r + g_q) \quad \text{Eq. (14)}$$

Delay per cycle is estimated using the following equation.

$$Total\ delay, D = 0.5 * (Q_{in} * r + q * r * (r + g_q)) \quad \text{Eq. (12)}$$

Note that the formula for delay per vehicle and percentage arrival on red remain same for these two special cases. In all cases, the per-vehicle delay is reported as per departing vehicle in the cycle.

Detector Latency and Presence of Queue from a Downstream Intersection: As explained earlier, the proposed threshold-based method would significantly underestimate the delay in a cycle if the stop bar detectors have an unknown latency or if queue spills back from a downstream intersection. In such cases, we may see that the estimated delay is even lower than what is calculated by the ATSPM approach, which is, as discussed earlier, allegedly underestimating the delay. Hence, we propose to use the ATSPM approach to estimate the delay if the proposed method based only on stop bar detector data generates a delay in a cycle that is lower than the ATSPM delay. Here, it is important to explain the mechanism of ATSPM approach of calculating delay. This method estimates approach delay for each vehicle as follows: a) the vehicle arrival time at the stopbar projected as the sum of the upstream detector actuation time plus travel time to the stopbar at the free flow or speed limit; b) If the arrival time occurred in red, then delay is simply the remaining red time, otherwise delay is set to zero. The total approach delay and delay per vehicle for a typical cycle can be obtained as follows:

$$Total\ delay\ in\ a\ cycle, D_{ATSPM} = \sum_i^N \mathrm{Max}[(T_G - T_{FF} - T_{arr_i}),0] \quad \text{Eq. (13)}$$

Where: T_G is the starting time of the next green for a cycle, T_{FF} is the free flow travel time for vehicle i, and T_{arr_i} is the time vehicle (i) activates the upstream detector. For this special case, the adjusted delay is obtained as the maximum of the two delays as shown below.

$$D_{adj} = \max(D_{stop\ bar}, D_{ATSPM}) \quad \text{Eq. (14)}$$

Here, $d_{stop\ bar}\ is$ the delay calculated based on the proposed stop bar detector only method.

5.2 Proposed Algorithms' Data Requirements, Strengths and Weaknesses

As mentioned earlier, the proposed algorithms have been developed recognizing the fundamental limitations of the ATSPM approach delay and with the goal of addressing these limitations. For example, the arrival–departure method is capable of fully incorporating stopped and deceleration delays, but can only partially incorporate acceleration delay (only the portion incurred prior to the downstream detector). The departure-only method, however, has the capability to fully take into consideration the stop component of the approach delay. Furthermore, it includes any initial queue delays – estimated with the assumption of uniform arrival rate that spans through all consecutive cycles with split failure. Table 7 provides a graphical depiction of delay components considered by each methodology.

TABLE 7: ASPECTS OF CONTROL DELAY INCLUDED IN ESTIMATE

Method	Delay Component			
	Stop	Deceleration	Acceleration	Initial
Arrival-Departure	FI	FI	i	NI
Departure-Only	FI	NI	NI	ii
ATSPM	iii	NI	NI	NI

i) Acceleration delays incurred prior to the downstream detector are included
ii) Initial queues are estimated using uniform arrival rate which spans consecutive cycles with split failure
iii) Stopped delay is minimally estimated assuming a "vertical queue"
FI) Fully included
NI) Not included

Furthermore, each method relies upon the availability of certain data elements and involves some critical assumptions or limitations, many of which may be addressed through future research. Table 8 provides a comprehensive list of data needs, assumptions, and limitations for each methodology.

TABLE 8: MODEL LIMITATIONS/ASSUMPTIONS AND DATA NEEDS

Data Need/Assumption/Limitation	Method		
	Arrival-Departure	Departure-Only	ATSPM
Upstream Detections (Arrivals)	✓	-	✓
Downstream Detections (Departures)	✓	✓	-
Traffic Signal Timing	✓	✓	✓
Needs Adjustment of Arrivals	✓	-	-
Fixed Speed	✓	✓	✓
Under-saturated Cycles	✓	-	✓
Uniform Arrival Rate	-	✓	-
First-In-First-Out (FIFO)	✓	-	-

5.3 Application of the developed framework and methodologies to a real-world facility

To adequately assess the presented three delay methods, the project team identified three critical data elements needed at a candidate approach for conducting a case study. First, signal timing data must be available in the ATSPM dataset, which is true for almost all of the sites equipped with ATSPM technology. Second, the approach must have upstream detection in order to count arrivals. Finally, a downstream detector is needed for both the departure-only and for the arrival-departure method to keep track of vehicle departures at stop-bar. The data used in this section was provided as part of the *big data challenge* on signalized intersection organized by the Traffic Signal Systems Committee of the Transportation Research Board (*46*). To determine which intersections had the necessary data to exercise all three methods, all Signal IDs were entered into the UDOT ATSPM website (*2*) to identify the available performance measures and associated detector data for that intersection.

Among the 22 signalized intersections for which the data was provided through the challenge, Signal ID 7181 was the only one that met the data requirements for all three methods. Thus, the analysts selected it based on data availability. The research team then established the event enumerations of interest in the raw data file for the Southbound approach along with turning movement count detectors and the corresponding Phase 6 signal data. Asingle day on Tuesday November 6, 2018 was selected for use in all three methods. In addition to the selection of events of interest in the log, an additional field was added for

advanced processing of data use for each method to establish a common cycle ID based on the start of the red for Phase 6 in each cycle.

The two proposed methods used the data from a number of detectors at this intersection. Detector 4, an advanced count detector located 390ft from the stop bar, was used for keeping track of arrivals. Detectors 43, 42, and 41 are lane-by-lane stop bar count detectors that were used for keeping track of departures from the intersection. Arrival – Departure method used actuations of detectors 4, 41, 42, and 43 for calculation of approach delay, while the departure-only method used detectors 41, 42, and 43.

5.4 ANALYSIS AND RESULTS

This section provides the analysis and results pertaining to all three methods for estimating the average delay per cycle, and in 15-minutes periods. Since the true delay is unknown we limit our discussion to delay trends under different methods and conditions. Figure 13 shows the results of average approach delay per vehicle by all three methods, aggregated at the cycle level. Two-hour periods are highlighted in Figure 13b and Figure 13c to allow for direct comparison during the AM peak and Mid-day periods. A three-hour period is shown in Figure 13d for comparison of PM peak period.

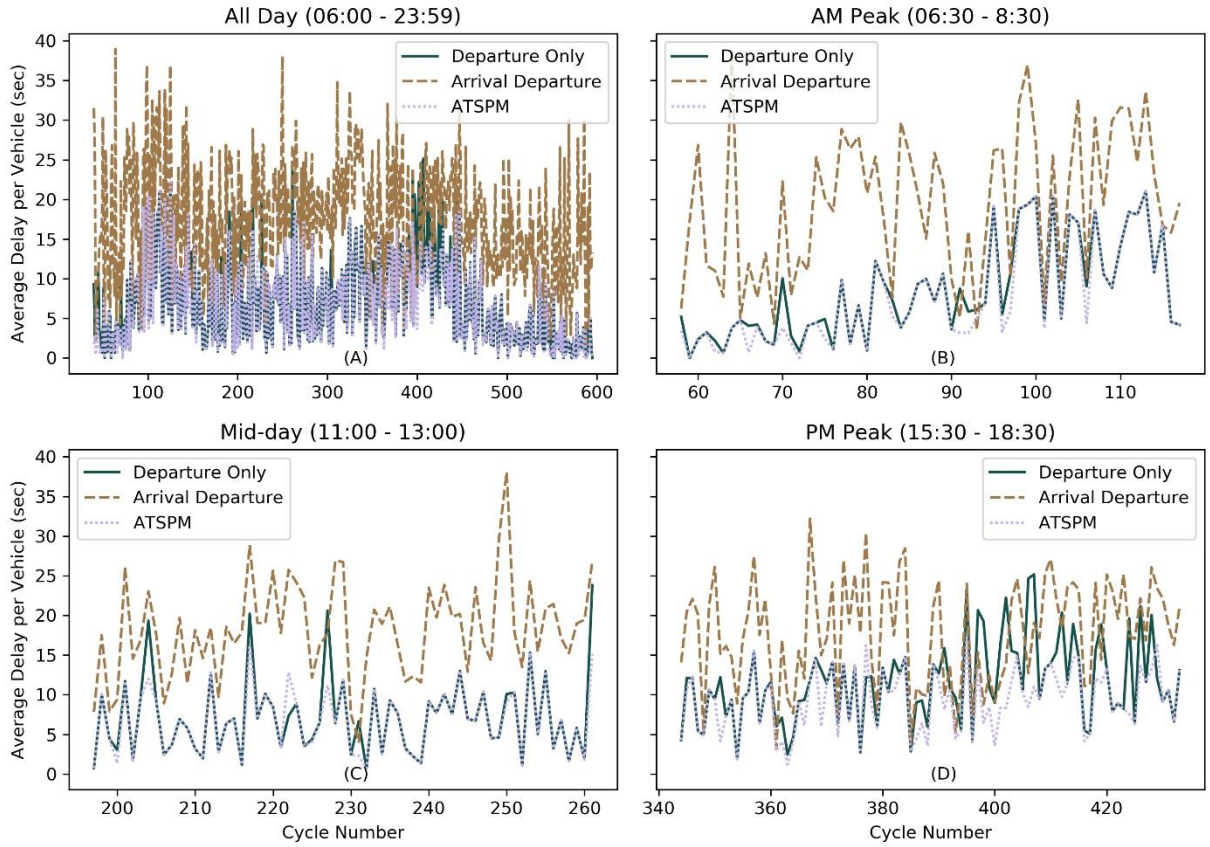


FIGURE 13: AVERAGE DELAY PER VEHICLE ON A CYCLE-BY-CYCLE BASES BY ATSPM AND THE TWO PROPOSED METHODS A) ALL-DAY PERIOD B) AM PEAK PERIOD C) MID-DAY PERIOD D) PM PEAK PERIOD

Visual observations from Figure 13 indicate that the delay estimated by arrival-departure method is almost always larger than that estimated by the departure-only and ATSPM. This is expected because the arrival-departure method, in addition to incorporating stopped delay, includes deceleration delay and a portion of the full acceleration delay profile of vehicles.

The departure-only method closely matches the ATSPM delay during the mid-day period but is relatively higher during the AM and PM peaks. Figure 13a shows that departure-only and ATSPM methods spike during the AM and PM peaks. However, the arrival-departure method does not show similar trends during the PM peak. Although not showing a clear peaking pattern during the PM peak, the arrival-departure method still provides higher estimates of delay during this period compared to ATSPM and departure-only methods.

The violin plot shown in Figure 14 provides further insight into the distribution and range of each group within the four time periods. Each violin shows the median (white dot located roughly in the middle of the black boxes in each violin), the interquartile range (the black box in each violin), minimum and maximum values for each group shown by the leftmost and rightmost points of violins, respectively. Furthermore, the probability density distribution for each group is shown by the upper bound of the violin and mirrored around the interquartile range line for aesthetic purposes.

Close observations of the violin plots reveal that the arrival-departure method consistently estimates higher average cycle delay for all time periods. In addition, the variability of delay estimates through this method is higher than the other two methods as shown by the large interquartile range within each arrival-departure violin. However, the violin plots for ATSPM and departure-only methods are very similar for the AM peak period and slightly different for other periods. The departure-only method's median is consistently higher than the ATSPM method. However, the variance between the two methods is very similar for all time periods as indicated by the size of the interquartile range within each violin.

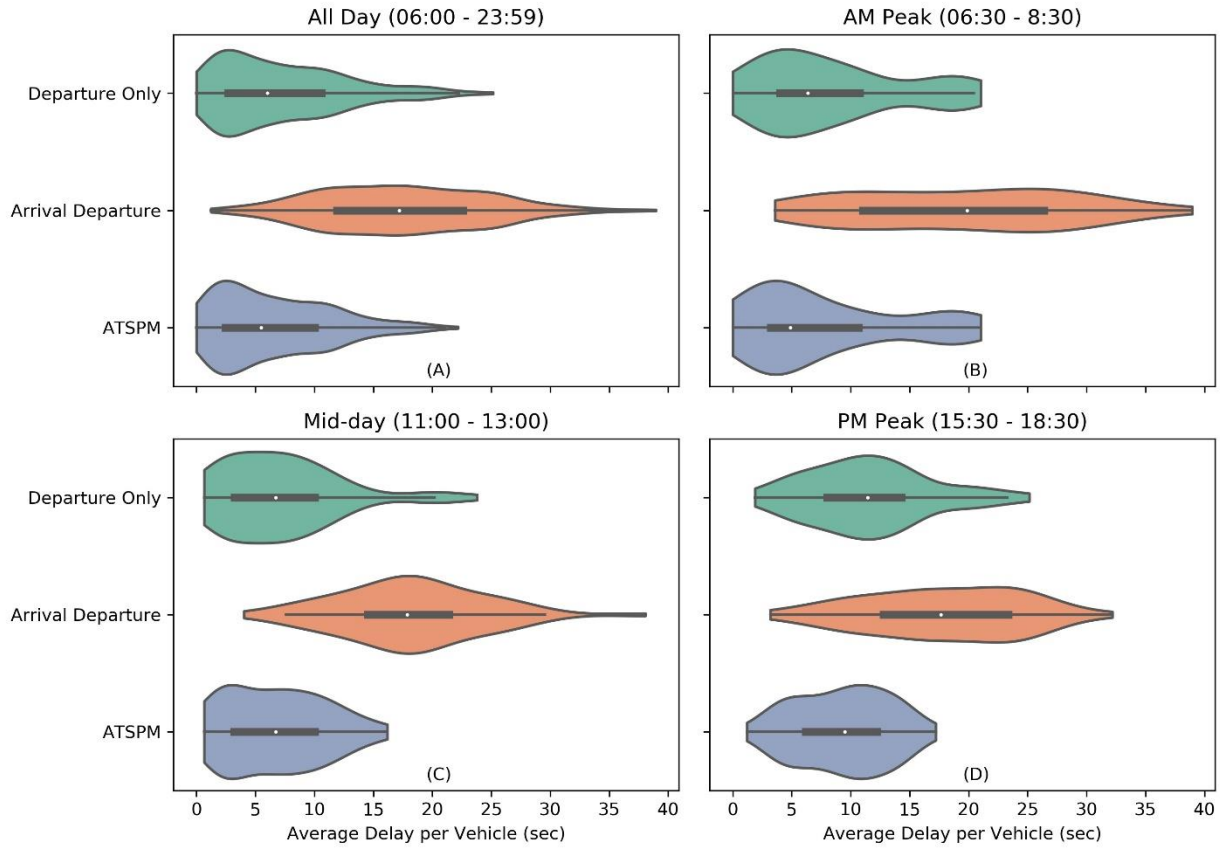


FIGURE 14: VIOLIN PLOT FOR THE AVERAGE DELAY PER VEHICLE WITHIN EACH CYCLE BY THE THREE METHODS - SHOWING MAXIMUM, MINIMUM AND INTERQUARTILE RANGE UNDER EACH METHOD A) VIOLIN PLOTS FOR ALL DAY B) VIOLIN PLOTS FOR AM PEAK C) VIOLIN PLOTS FOR MID-DAY D) VIOLIN PLOTS

An analysis of variance (ANOVA) was conducted on the four time periods depicted in Figure 13. Results for all time periods give p-values less than 0.05, indicating that at least the mean of each method is statistically different from the two others. To identify which methods' means is/are statistically different, we conducted Tukey's Honest Significant Difference (*47*) post-hoc test. Results of Tukey test show that for the All Day, AM Peak, and Mid-Day periods, the means of ATSPM and departure-only methods are not statistically different. However, the means of (ATSPM – arrival-departure) and (arrival-departure – departure-only) pairs were found to be statistically different for the same periods. For the PM peak period, however, the pairwise comparison of group means were all found to be statistically different.

In addition, a covariance analysis was conducted in which the covariance between ATSPM and arrival-departure was approximately 21.73, the covariance between ATSPM and departure-only was 24.09, and the covariance between arrival-departure and departure-only was 19.22. These values indicate that all three relationships are positive.

Aggregation of average cycle delays to 15-minute periods, depicted in Figure 15, show similar trends as depicted in Figure 13. First, the arrival-departure method consistently produces higher estimates of delay compared to the other two methods. Observation of AM,

Mid-day and PM subplots show that arrival-departure method resembles ATSPM's increasing and decreasing patterns. The departure-only method closely matches ATSPM estimates for all periods except PM peak. AM peak is captured well by all three methods and so is the PM peak, except by the arrival-departure method which does not show significantly evident peaking within that time period. This could be because of not meeting the stable operations assumption made under this method. Cross examining the split-failure for the PM peak period revealed that most of the cycles failed during this period. Nonetheless, the method is estimating higher delays for the PM peak compared to the ATSPM method.

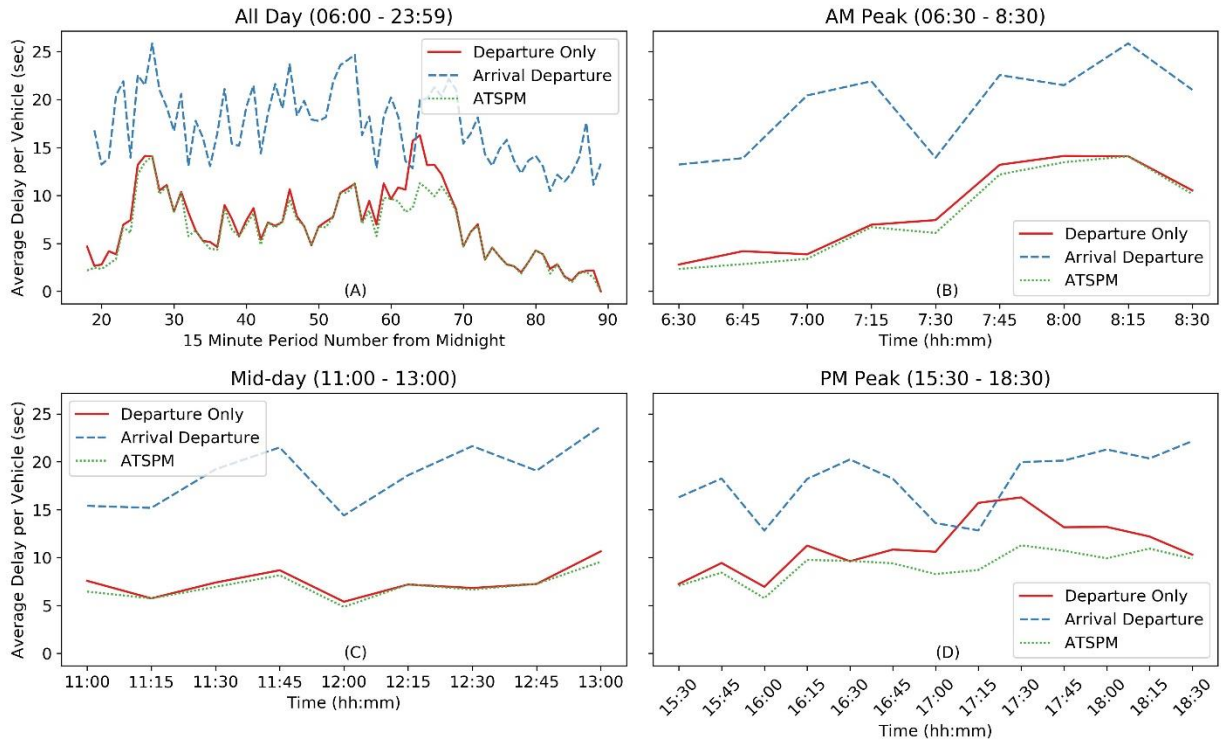


FIGURE 15: FIFTEEN-MINUTES AGGREGATED AVERAGE DELAY PER VEHICLE BY ATSPM AND THE TWO PROPOSED METHODS A) ALL-DAY PERIOD B) AM PEAK PERIOD C) MID-DAY PERIOD D) PM PEAK PERIOD

The violin plots for the average 15-minutes aggregated delays are shown by Figure 16. One can clearly note that the arrival-departure method provides higher estimates at all time periods. Especially, the distribution of AM and Mid-day delay estimates by this method are not overlapping the other two methods' distribution of delays. However, there is overlapping of these distributions in the PM peak period. The departure-only and ATSPM methods have very similar violin plots for AM, and Mid-day periods. However, the departure-only violin plot is elongated during the PM peak compared to ATSPM. Furthermore, the median delay estimated by the departure-only method is consistently above that of the ATSPM for all time periods.

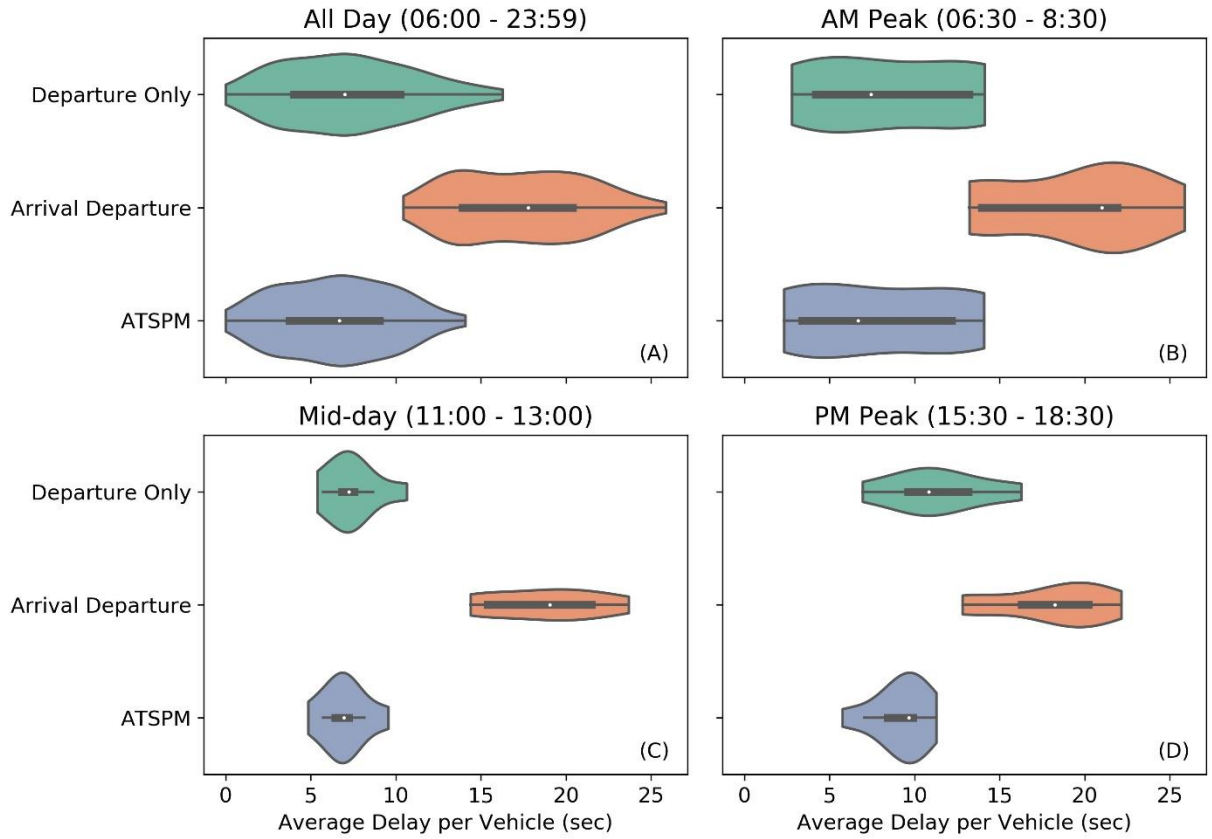


FIGURE 16: VIOLIN PLOT FOR THE AVERAGE DELAY PER VEHICLE WITHIN EACH CYCLE BY THE THREE METHODS - SHOWING MAXIMUM, MINIMUM AND INTERQUARTILE RANGE UNDER EACH METHOD A) VIOLIN PLOTS FOR ALL-DAY B) VIOLIN PLOTS FOR AM PEAK C) VIOLIN PLOTS FOR MID-DAY D) VIOLIN PLOTS

Results of the ANOVA for the 15-minute aggregated average delay within each time period yielded very small p-values indicating that within each time period there is at least one mean in a method that is statistically different from the mean of the other two. Tukey's post-hoc test was conducted to do a pairwise comparison of all three methods. The results at the 5% significance level within all time periods showed that only the (ATSPM – departure-only) pair resulted in p-values greater than 0.05 indicating that the two groups do not have means that are statistically different from each other. The rest of the pairs (ATSPM vs. arrival-departure and departure-only vs. arrival-departure), however, resulted in p-values smaller than 0.05. Thus, indicating a statistically significant difference between the means of these groups.

Furthermore, a covariance analysis was conducted in which the covariance between ATSPM and arrival-departure was approximately 7.011, the covariance between ATSPM and departure-only was 13.2, and the covariance between arrival-departure and departure-only was 1.62. These values indicate that all three relationships are positive.

Table 9 provides descriptive statistics for the ATSPM and the two proposed methods at both the cycle-by-cycle and fifteen-minutes aggregated levels. In addition, these statistics have been clustered to four time periods of all day, AM peak, Mid-day and PM peak.

Observations of Table 9 provide further insight on the descriptive statistics for the three methods. It is apparent that the mean delay estimated from the arrival-departure method is always higher than the other two methods by at least 50%. At the cycle-based level, it is found that the standard deviation of the arrival-departure method is much higher than the other two methods for all time periods, indicating high variability of estimates by the method. However, the difference becomes much lower when the cycle-based average delay is aggregated at the 15-minutes level. The remaining statistics, shown in Table 9, are always higher for the arrival-departure method compared to the ATSPM and departure-only methods for all time periods.

TABLE 9: DESCRIPTIVE STATISTICS OF ESTIMATED DELAYS BY ATSPM AND PROPOSED METHODS

Time Period	Stats.	Cycle-Based			15-Minutes Aggregated		
		Departure Only	Arrival Departure	ATSPM	Departure Only	Arrival Departure	ATSPM
All Day	**mean**	**7.19**	**17.44**	**6.53**	**7.09**	**17.36**	**6.43**
	std	5.41	7.05	4.88	3.80	3.79	3.32
	min	0.00	1.23	0.00	0.00	10.44	0.00
	25%	2.71	11.92	2.45	4.11	13.92	3.56
	50%	6.01	17.19	5.50	6.97	17.77	6.67
	75%	10.59	22.61	10.06	10.29	20.39	8.87
	max	25.15	38.96	22.19	16.28	25.86	14.09
AM Peak	**mean**	**8.38**	**19.42**	**7.69**	**8.33**	**19.17**	**7.64**
	std	6.15	9.29	6.44	4.80	4.80	4.91
	min	0.00	3.58	0.00	2.79	13.22	2.33
	25%	4.00	11.04	3.21	4.11	13.93	3.25
	50%	6.38	19.88	4.88	7.19	20.97	6.39
	75%	10.78	26.40	10.67	13.42	22.08	12.51
	max	21.03	38.96	21.03	14.12	25.86	14.09
Mid-Day	**mean**	**7.20**	**18.18**	**6.71**	**7.01**	**18.12**	**6.66**
	std	5.05	6.15	4.14	1.04	2.83	1.01
	min	0.66	4.04	0.66	5.39	14.39	4.85
	25%	3.22	14.53	3.19	6.55	15.34	6.27
	50%	6.72	17.86	6.72	7.22	18.83	6.80
	75%	10.05	21.43	10.05	7.44	19.78	7.21
	max	23.80	38.06	16.19	8.68	21.63	8.15
PM Peak	**mean**	**11.55**	**17.81**	**9.32**	**11.37**	**17.68**	**9.16**
	std	5.20	6.63	3.90	2.93	3.08	1.62
	min	1.89	3.18	1.21	6.95	12.82	5.76
	25%	7.98	12.77	6.18	9.57	15.62	8.39
	50%	11.43	17.64	9.49	11.04	18.24	9.53
	75%	14.32	23.35	12.22	13.17	20.15	10.12
	max	25.15	32.19	17.22	16.28	21.28	11.28

6.0 Validation of Delay Estimation Algorithms

6.1 Description of data

The purpose of this chapter is to assess the validity of the three algorithms presented in the previous sections in terms of their ability to match the true vehicle delays. This of course required a high-resolution dataset with access to individual vehicle trajectories to enable the estimation of the true delay, and also to emulate the data needed by each algorithm.

The research team used data from the Next Generation SIMulation (NGSIM) program which focuses on collecting detailed, high-quality individual vehicle trajectories along with detailed signal data. The dataset chosen for this project contains vehicle trajectories on Peachtree Street in Atlanta, Georgia. Vehicles traveling southbound in the arterial through Peachtree St. at 10th St. NE on June 16, 2005, were tracked. The data were collected for a total time period of 30 minutes segmented into two 15-minute time periods (12:45 to 13:00 and 16:00 to 16:15). The figure below shows geometry at the intersection.

FIGURE 17: AERIAL VIEW OF THE SITE: PEACHTREE STREET AT 10TH ST. NE)

The posted speed limit on the arterial is 35mph. The delay experienced by all the vehicles heading SB from 11th St. to 10th St. arriving in all 3 lanes (a shared through-right lane, a through-only lane, and a left-turn pocket) at Peachtree St and 10th St. NE was selected. The signal timing data for the intersections during the study period is also provided in the NGSIM

Peachtree Street dataset. The signal timing plan includes a leading SB left-turn phase at the intersection. The dataset contains information about every vehicle every tenth of a second in the study area consisting of the clock time, speed, acceleration, direction, movement, intersection number, lane number, and the vehicle id of the following and successive vehicle. A schematic of vehicle trajectory arrivals and departures in Cycle six in the second time period is shown in **Figure 18**.

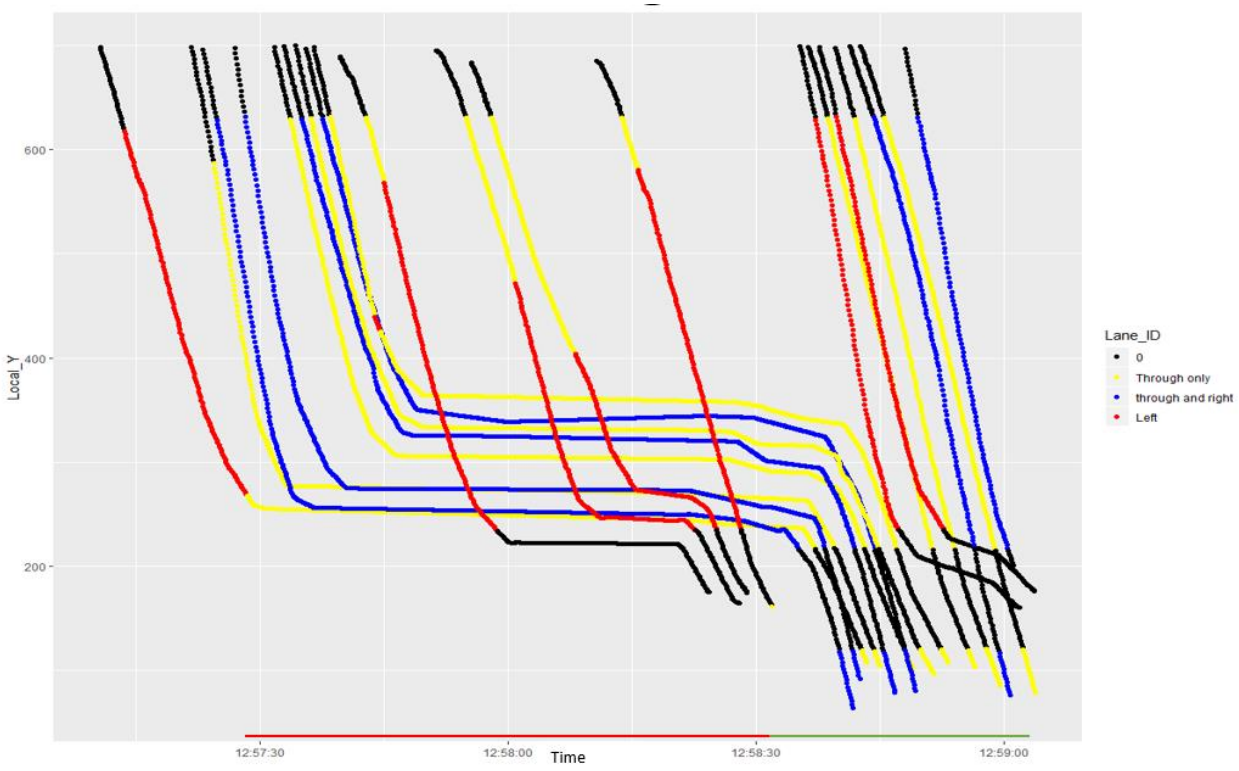


FIGURE 18: VEHICLE TRAJECTORIES IN CYCLE 6 COLORED CORRESPONDING TO THE LANE (16:00 – 16:15 TIME PERIOD)

6.2 Methods used to clean/process the data

A quality check of the trajectory data generated by the NGSIM program revealed several irregularities, and the team applied various filtering methods to address many of the irregularities in the data. Vehicles have been first filtered by class to filter out trucks and motorcycles. Since the vehicle ids are apparently been recycled during the study period, a unique identifier called Vehicle Identification Number (VIN) was created for every vehicle. VIN for a vehicle is obtained by concatenating the vehicle id and the total number of frames for that particular vehicle. The NGSIM data definition sheet states that the global time value is the time elapsed since Jan 1, 1970, in milliseconds. However, a few corrections had to be made to the global time values field to negate the errors present. The first 391,647 rows of the dataset contain trajectory data from the 12:45 PM to 1:00 PM time period and the remaining rows

contain data from the second time period. The minimum global time value for both time periods also needed to be set to the starting time (i.e., 12:45 PM or 4:00 PM). The rest of the fields can be calculated as time elapsed in milliseconds since the start of the time period.

Since the lane id field is set to zero in the vicinity of an intersection, the Local X values are taken to identify the lane in which a vehicle is traveling. Vehicles departing well after the start of red have been identified and removed from the data. A total of 16 cycles (8 in both the time periods) were finally retained for comparing the different methods of calculating delay.

6.3 Implementation of the methods

The study includes vehicles heading southbound from Peachtree St. at 11th St. to Peachtree St. at 10th St. in all three lanes. The starting and ending cycles of each time period are omitted since the information for the entire time period is unavailable. Eight cycles from each time period containing a sample set of 252 vehicles have been used to compare the three methods to estimate delay. The signal timing sheets in the NGSIM Peachtree dataset contain frame ids from the start of red, green, and yellow. The cycle lengths, red times, and green times have been derived by taking the corresponding global time value for each frame. The tables containing the signal parameters for the two time-periods are given below.

TABLE 10 AND TABLE 11 depict the start of red, the start of green, cycle length, red time, and green time for each cycle for comparing ATSPM, departure-only methods, and arrival-departure methods with the true delay. The total number of all the vehicles heading southbound in Peachtree St. towards Peachtree St. at 10th St. in each signal cycle is also listed in the table. The algorithm formulated for the delay estimation has been done using R script. The code is attached to the appendix.

TABLE 10: SIGNAL TIMING PARAMETERS FOR THE FIRST PERIOD (12:45 – 13:00)

Cycle number	Cycle length (s)	Red time (s)	SB Phase Green time (s)	No. of vehicle arrivals
2	95.1	65.5	29.6	9
3	95.1	65.5	29.6	14
4	95.1	65.5	29.6	14
5	95.2	65.6	29.6	13
6	95.1	65.5	29.6	17
7	95.1	65.4	29.7	13
8	95.1	65.4	29.7	14
9	95.1	65.4	29.7	16

TABLE 11: SIGNAL TIMING PARAMETERS FOR THE SECOND PERIOD (16:00 – 16:15)

Cycle number	Cycle length (s)	Red time (s)	SB Phase Green time (s)	No. of vehicle arrivals
2	100.4	61.9	38.5	7
3	100.4	62	38.4	19
4	100.3	62.1	38.2	16
5	100.3	62	38.3	34
6	100.4	62.1	38.3	20
7	100.3	62	38.3	14
8	100.4	62.1	38.3	24
	100.3	62	38.3	18

6.3.1 True delay estimation

The first step in the analysis is to filter all the vehicles passing through Peachtree St. at 10th St. NW from the Peachtree Street dataset. The point at which departure times are clocked is at 10 ft from the stop bar. The stop bar is taken to be at a local Y value of 225 ft. The line of arrival was taken at 400ft upstream of the departure point. These points are somewhat arbitrary but represent typical location of upstream and stop-line sensors. The frame id with Local Y closest to those points has been tracked for each vehicle. The team made sure that the arrival point is taken far enough from the upstream intersection so as to give the vehicles enough time to accelerate and attain free-flow speed. The location of both the points are shown in Figure 9. The clock times at which the filtered vehicles crossed the reference points were noted. The difference between those two times is the actual travel time for each vehicle. The travel time without stopping at the signalized intersection is calculated from the relation between upstream speed, distance and time. There are multiple possible speeds that can be used such as the speed limit, 85th percentile speed, median speed, arrival speed and trimmed mean of the speeds. The speed taken for the calculation is the actual speed at which each vehicle was captured at the upstream point. Taking the arrival speed would give the closest estimate to the true delay experienced by every vehicle assuming they have achieved free flow speed. The difference between time taken with and without the signalized intersection gives the true control delay experienced by each vehicle.

6.3.2 Arrival-departure method

For implementing the arrival-departure method, the vehicle detection times are arranged in ascending order at both the upstream and the downstream point without using the vehicle id. The delay is calculated by subtracting the first downstream detection with the first upstream detection and so on. After calculating individual vehicle delays, vehicles are assigned to the corresponding signal cycle in which they arrived. This is done by comparing the arrival time of the vehicle and the start of the signal cycle. Once the vehicles have been assigned a signal cycle number, the average delay per cycle can be determined by taking the average delay of all the vehicles in the cycle. A delay vs signal cycle graph is then plotted.

6.3.3 ATSPM method

The delays have been calculated for all vehicles including left turning vehicles. However, the signal timing plan for through vehicles is used to calculate delay for all vehicles as it is not possible to determine the movement of a vehicle at the intersection using upstream data. Similar logic has been used to formulate the code for estimation of ATSPM delay.

6.3.4 Departure-Only Method

Two threshold values are used in this method to identify if the first vehicle had arrived in red and to estimate the number of queued vehicles respectively. The headway of the first vehicle (time difference between the start of green and the departure time of that vehicle) is compared against a threshold value (t_1) taken as 4 seconds. If the headway is larger than the threshold value, then it is concluded that the vehicle has arrived in green. To estimate the number of vehicles that are queued in each cycle, the method uses a moving average of the headway of vehicles that arrived before the subject vehicle. If the difference between headway and the moving average is higher than the threshold value (t_2) taken as 5 seconds, it is concluded that the vehicle was not queued. The function written in R-code calculates the area under the average delay line and the x-axis to estimate the average delay per vehicle. This concept is illustrated in **Figure 19**. The average delay is estimated using the average delay line, and the maximum delay assumes all vehicles departing from the queue had arrived at the start of the red phase in a platoon. Simply state, the maximum delay line will generate a delay that is twice the value generated by the average delay line.

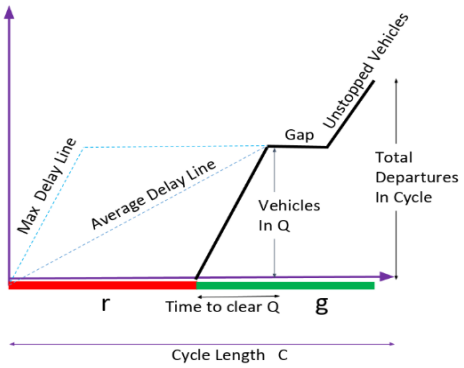


FIGURE 19: AVERAGE AND MAXIMUM DELAY FOR A CYCLE USING THE DEPARTURE ONLY METHOD

6.4 Results

This section provides the results for the analysis conducted on the Peachtree dataset for the estimation of average delay (only) in each cycle using the three methods mentioned against the true delay. There are two case studies: One comparing delay for all vehicles on the approach and the second case in which upstream filtering is carried out to match the downstream through and right vehicle counts.

6.4.1 All Vehicles on SB Approach

The bar plots depicting the average *approach* delay in each cycle in **Figure 20** and **Figure 21** show that the ATSPM method is consistently underestimating the delay in each cycle. This is expected since that method does not capture the acceleration, deceleration and queuing delays for each vehicle. The delay predicted by the arrival-departure method is slightly higher than the ground truth delay. This can be a result of the selection of the speed used to calculate these delay values. It was observed that the 85th percentile speed (taken at the upstream detector) is 10mph lower than the posted speed limit used to estimate delay. The departure only method (using the average delay line) estimated lower delays than the true delay in 13 of the 17 cycles studied. It turned out that the arrival patterns at that intersection exhibited early arrivals in red (see **Figure 19** for the maximum delay example), thus violating the assumption of uniform arrivals in red under that method. This point was made earlier in the previous chapter. Statistical significance tests were performed on the delays predicted by each method. The results from each of the tests are discussed below.

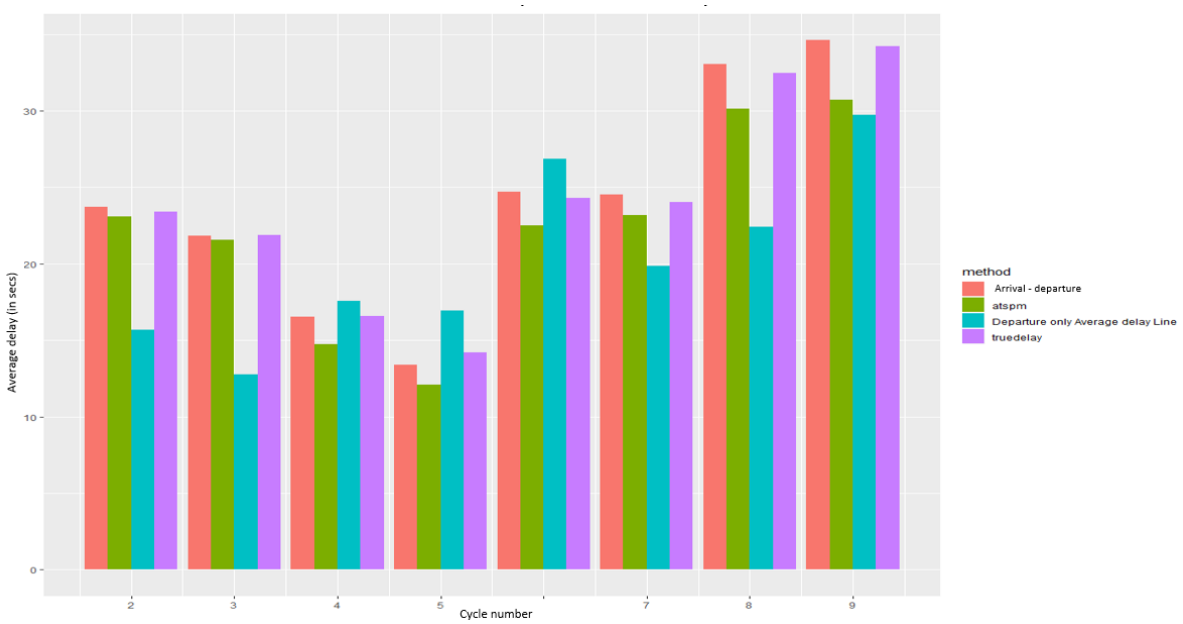


FIGURE 20: COMPARISON OF APPROACH DELAY ESTIMATION USING THE PROPOSED METHODS FOR THE 12:45 -13:00 TIME-PERIOD

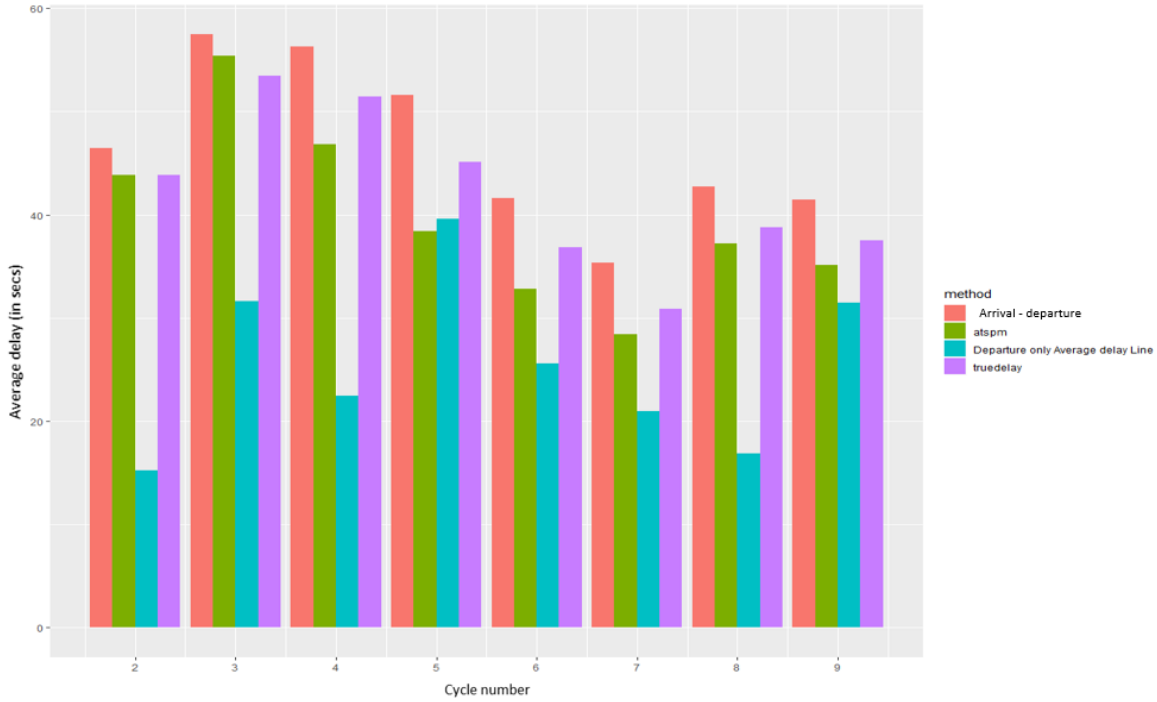


FIGURE 21: COMPARISON OF APPROACH DELAY ESTIMATION USING THE PROPOSED METHODS FOR THE 16:00 - 16:15 TIME-PERIOD

6.4.1.1 Descriptive Statistical tests

The Spearman's test gives a spearman rank coefficient based on the ranking of delays corresponding to the different cycles in a time-period. The rank coefficient is one of the ranks from the method is the same concerning the true delay and it goes towards zero when the rank order is completely different. The Spearman's rank coefficient test yielded a value of 0.93 for ATSPM vs true delay, 1.0 for Arrival-departure vs true delay and 0.79 for the departure-only method for the 12:45 PM – 1:00 PM time-period. This goes to show that the order in which the arrival-departure method predicted from the highest to lowest delays were the same concerning the ground truth. The ATSPM did slightly better than the departure-only method in capturing the higher and the lower delays across all the cycles in this dataset.

For the 4:00 PM – 4:15 PM time-period, the Spearman's rank coefficients were 0.98, 0.98, 0.33 for ATSPM, arrival-departure and departure-only method respectively with respect to ground truth delays. The performance of the departure only method in this time-period is worse than the previous time-period which could be due to the higher percentage of arrivals during the start of the red. The departure only method assumes uniform arrivals in red. This assumption leads to the method underestimating delay for a cycle if the percentage of arrivals at the start of the red are higher and similarly overestimating delay for a cycle if the percentage of arrivals is higher in the green or the later part of the red.

The Root Mean Squared Error (RMSE) test is done to estimate the differences in the values of delays predicted for each cycle by the three methods. Smaller the RMSE, better is the performance of the method when compared against true delay. The RMSE for ATSPM is 1.9 and 3.6 for the 12:45 to 1:00 PM and 4:00 to 4:15 PM periods respectively. Similarly, the RMSE for the arrival-departure method is 0.4 and 4.5, the RMSE for the departure-only method is 6.1 and 19.1. The values show that there is little error between the predicted delay and the true delay for a particular cycle except for the departure-only method for the PM peak (19.1).

The Mean Absolute Percentage Error (MAPE) measures the deviation of a predicted value from the actual value in terms of percentage. The MAPE for ATSPM is 7.1 and 6.2 while the MAPE for the arrival-departure method is 1.7 and 10.6. Similar to the RMSE, we find that the arrival-departure method has a smaller error as to the MAPE value of ATSPM in the first time-period while ATSPM outperforms the arrival-departure method in the PM peak. The MAPE for the departure-only method is 1.9 and 38.8. The error for the first time period is very negligible and better than that of ATSPM.

As discussed in earlier chapters, the arrival-departure method captures more components of a signal delay than the ATSPM method. It, therefore, outperforms the ATSPM method for this particular dataset as expected. The sample taken for analysis is small hence, a larger sample could be used in the future to further validate the mentioned methods.

6.4.2 Matching Upstream Arrivals and Downstream Through and Right Departures

This analysis focuses on estimating delays for through and right turning vehicles, to the exclusion of left turns at the stop-line. However, it should be noted that the analyst cannot

identify the vehicle destination from the upstream location. In this analysis, the total count of through and right vehicles in a cycle at the downstream detector is made, and a random sample from the upstream vehicles is taken in order to match the upstream count with the downstream count. Of course, there is no guarantee that the generated random sample will only yield through and right turning vehicles. The true delay, arrival-departure approach delay and the ATSPM delay is calculated for the selected set of vehicles. The average delay per cycle from the above-mentioned methods is compared with departure only delay for all the through and right vehicles. The edited dataset is not needed in the departure only method as that method does not use any upstream detections in its algorithms.

Figure 22 and **Figure 23** show the approach delay estimated by ATSPM and the two proposed methods against the true delay for the first and second time periods, respectively. As expected, the method yielding delays closest to the ground truth is the arrival departure in both periods followed by the ATSPM and the departure only methods. These observations are in line with the findings reported previously. The arrival departure method slightly overestimates the delay in all of the cycles except for one cycle in the second period. The ATSPM method, on the other hand systematically underestimates the delay throughout all the cycles of the two periods. As commented earlier, the main reason for this underestimation is the exclusion of the acceleration, deceleration and queue delays associated with the approach delay. Finally, the departure only methods still underestimated delays in all 17 cycles, due to the early arrival in red pattern for SB vehicles at that intersection.

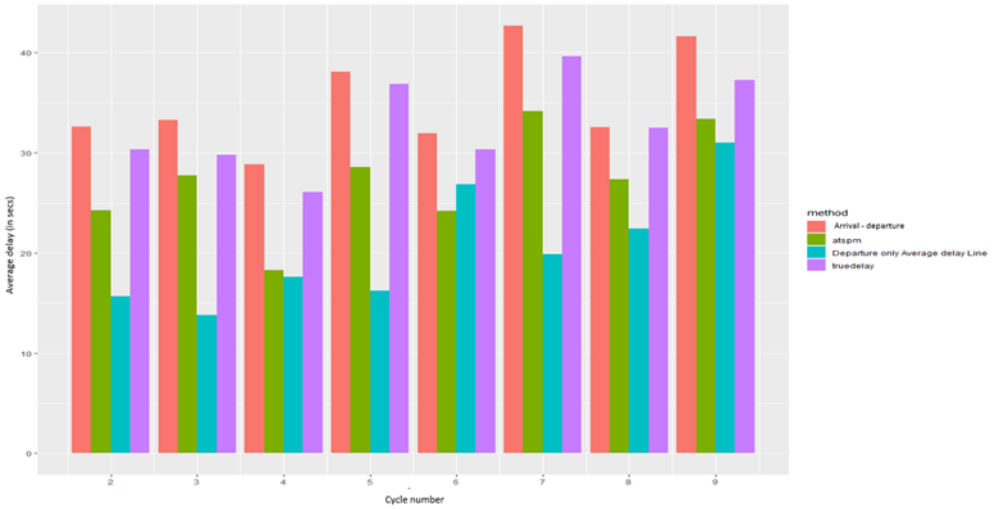


FIGURE 22: COMPARISON OF ESTIMATED CYCLE DELAY BY THE PROPOSED METHODS FOR THE 12:45-13:00 TIME-PERIOD

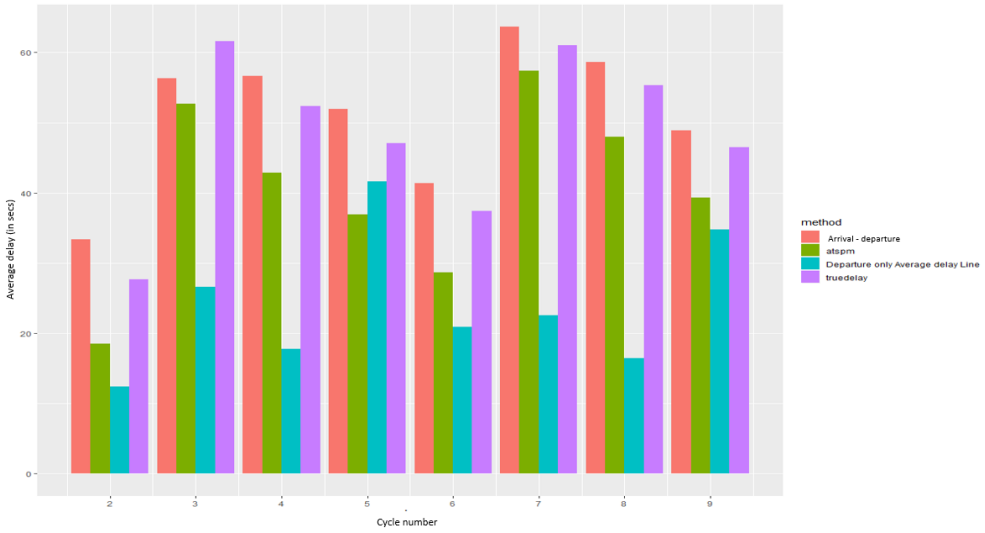


FIGURE 23: COMPARISON OF ESTIMATED CYCLE DELAY BY THE PROPOSED METHODS FOR THE 16:00 - 16:15 TIME-PERIOD

6.4.2.1 Descriptive Statistics tests

For the 12:45 to 1:00 PM time-period, the ATSPM rank coefficient (0.83) is lower than that observed in the previous analysis (0.93). This is expected as the method may have included some left turn vehicles in the expected through or right-turning vehicle sample.

The departure only method underestimates the delay for each vehicle as the percentage of arrivals at the start of the red is high. In cycles where the arrivals are uniform throughout the red, the departure only method can predict the delays with high accuracy. The departure does not account for the right turns on red which can lead to errors in estimating the number of vehicles that are queued. This accuracy in estimating the queue is critical to the performance of this method. The arrival-departure method performs better than other methods. The RMSE values for the method are 2.7 and 4.2 respectively for the two time periods. The percentage error for the arrival departure method is 7.3 and 9.2 which translates to a difference of around 3 seconds between the ground truth and predicted value. Decreasing the speed taken to calculate the delay to the 85th percentile value of the upstream approach value decreases the average delay by around 3 seconds. Thus, the method can further be improved if the approach speeds at the upstream detector are known while calculating the delay.

7.0 IMPACT OF MARKET PENETRATION RATE AND BROADCAST FREQUENCY ON REPORTED PROBE VEHICLE SPEEDS

7.1 Introduction

The successful wide-scale deployment of the Automated Traffic Signal Performance Measures (ATSPM) depends on many factors, one of which is the ability to obtain/acquire information that accurately reflects network speed and travel time. Many different media are available to collect speed and travel time on a network. One that has widely been used by public and private agencies are probe vehicle speeds or travel times. This method enables a representative sample of speeds experienced by all vehicles passing through a link to be collected. This section investigates the use of probe vehicles for obtaining link speeds on signalized arterials and reports on the effect of factors such as market penetration rate, broadcasting frequency, and aggregation level on the accuracy of the reported speeds.

A review of the literature revealed multiple efforts to determine the market penetration level of probe vehicles needed to accurately estimate link travel times for different facility types. Van Aerde et al. (*48*) investigated the reliability of probe vehicle travel times in the context of signalized arterials and found that as the number of probe vehicle reports increases, the reported probe vehicle mean travel times approach the ground truth travel times. Sen et al. (*49*) explored field data acquired through the ADVANCE project to investigate how the increase in probe vehicle sample size impacts the accuracy of travel times. They found through rigorous statistical analyses that probe vehicle speed reports are not independent and that an increase in probe vehicle sample size does not necessarily bring the sample mean to approach the population mean. Cheu et al. (*50*) found that a minimum probe market penetration rate of 5% results in estimates of average speed links that on average is within 5 km/h at a 95% confidence level.

Despite the plethora of recent efforts in the area of probe vehicle measures of effectiveness, none of the studies investigated the impact of *broadcasting frequency* in tandem with the market penetration rates and aggregation levels. In this section, we address this issue. These efforts employ visual and statistical methods to investigate the impact of these factors but do not involve developing in-depth analytical methods as this would distract from the main scope of the project. This investigation is intended to *roughly* determine whether these factors—and at what level-- would have a significant impact on the probe vehicle speed accuracy and if so to what extent.

7.2 Methodology

This section details the methods used for investigating the impact of market penetration rate, and broadcasting frequency on the accuracy of probe vehicle reported speeds. In addition, it presents details on the sampling method, market penetration, and broadcasting frequency levels considered.

Three distinct metrics are employed to determine how market penetration rates and broadcasting frequency impact the accuracy of reported probe vehicle speeds. In this study, the I-95 corridor coalition probe vehicle speed validation metrics (*51*) are used to determine the minimum market penetration rate that would yield accurate estimates of the speed on arterial links. This method uses two measures namely the Average Absolute Speed Error (AASE) and the Speed Error Bias (SEB). AASE is defined as the average absolute value of the difference between the mean speed reported from the probe vehicles and the ground truth mean speed for a specified time period. The limit at which probe vehicle data is said to not represent real-world conditions is identified to be greater 10 mph. The SEB, however, is defined as the average speed error. This metric is a measure of whether the reported probe vehicle speeds consistently under or overestimate speed compared to the ground truth. The limit for this metric is set to +/- 5 mph. These two metrics are reported in four-speed bins: 0-15 mph, 15-25 mph, 25-35 mph, and > 35 mph. Root mean square error (RMSE), the third metric, is used to explore the effect of broadcasting frequency on the accuracy of probe vehicle reported average speeds compared to the ground truth average speeds.

Multiple levels of probe market penetration rate (1%, 3, 5, 10, 15, 20, 30, 50, 70, and 90%) and broadcast frequency (15 seconds, 30 seconds, 45 seconds, and 60 seconds) are considered in this exercise. Broadcasting frequency is defined as the interval at which the probe vehicle collects and transmits its data to a central server. Considering a 15 seconds broadcasting frequency, for example, a probe vehicle will produce four readings in a minute. For each combination of market penetration rate and broadcasting frequency identified, 30 different sampling runs (with replacement) are generated. In each run a random process with a probability equivalent to the market penetration rate is performed to determine if a vehicle is a probe. This process reduces the number of vehicles to a random sample that matches the market penetration rate of probe vehicles. Speed readings from the randomly selected vehicles are collected at intervals that coincide with the broadcasting frequency and then used to determine the MOE estimate.

7.3 Study Site

The impact of market penetration rate (MPR) and broadcasting frequency (BC) on the accuracy probe vehicle reported speeds was estimated with the use of real-world vehicle trajectories. The site is an arterial namely Lankershim Boulevard, Los Angeles, CA which is a part of the Next Generation Simulation (NGSIM) program. This dataset contains vehicle trajectories at four signalized intersections starting with the Lankershim Boulevard Hollywood freeway off-ramp at the south end and ending at the James Stewart Avenue on the north end. The dataset includes information on vehicle ID, length, width, class, velocity, acceleration, intersection ID, coordinates, and origin-destination. Trajectory data are available at 10-hertz resolution. **Figure 24** shows the schematic of the site.

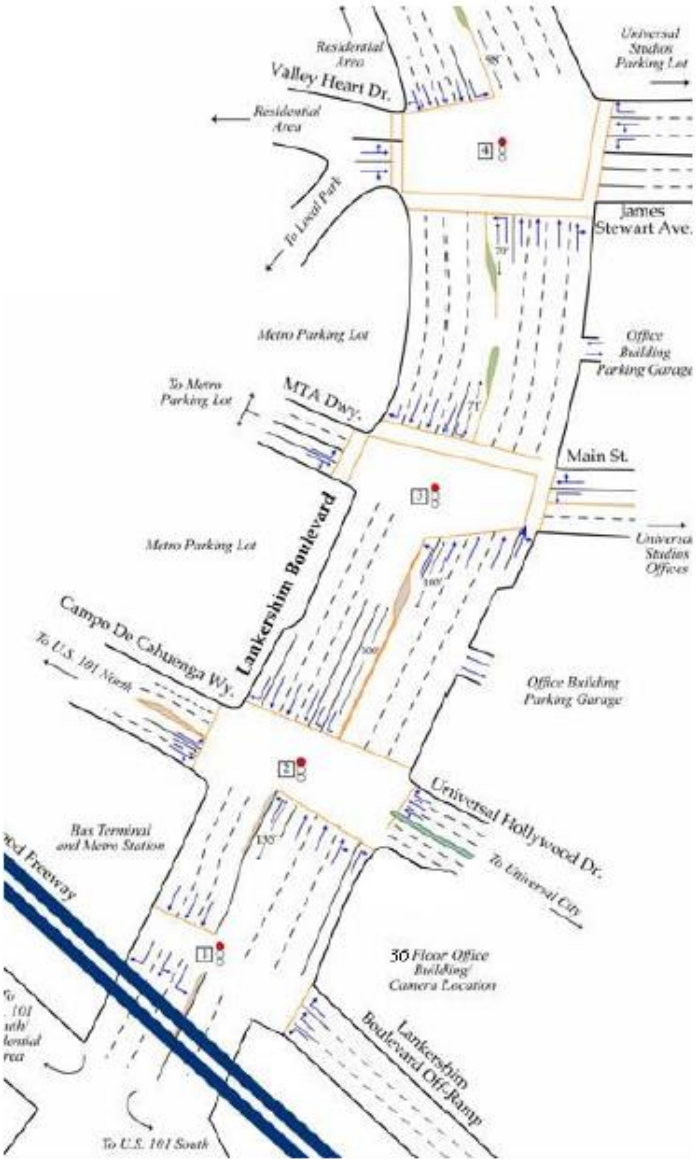


FIGURE 24: CONFIGURATION OF LANKERSHIM BOULEVARD STUDY SITE

7.4 Results and Analysis

Figure 25 and **Figure 26**: Effect of Market Penetration Rate on AASE and SEB: 45 seconds Broadcast Frequency

show the impact of market penetration rate on the Average Absolute Speed Error (AASE) and Speed Error Bias (SEB) under 15 and 45 seconds broadcasting frequencies, respectively. The three solid broken lines represent the SEB for the three different speed bins. The two dashed red lines indicate the upper and lower acceptable limits of the SEB. The three bars for each market penetration rate represent the AASE values for the three speed bins. The purple dashed-dotted line shows the acceptable limit for AASE.

Visual observations of **Figure 25** show that at low market penetration rates (below 5%) the AASE and the SEB are outside the acceptable range. Once the market penetration rate is at or above 5%, both AASE and SEB are within the accepted range. Additionally, an increase in MPR results in a decrease in both the AASE and SEB at all speed bins. This improvement is significant going from 3% to 5%. The rate of this improvement diminishes, however, at MPRs above 15%. Another observation that can be made, is the deviation of AASE and SEB from ground truth mean speed in bins 1 and 3. Both measures benefit from an increase in the MPR. The rate of improvement, however, decreases beyond 30% MPR range. The second speed bin, however, has a relatively low error compared to the first and third bins.

Figure 26: Effect of Market Penetration Rate on AASE and SEB: 45 seconds Broadcast Frequency

shows the result for the 45 seconds broadcasting frequency. A drastic departure of estimated speeds can be observed for all the speed bins at the low market penetration rates (lower than 10%). Particularly, estimates for the first and third speed bins show a significant departure from the acceptable limits of AASE and SEB. Probe vehicle speed estimates for the 45 seconds broadcasting scenario are acceptable only at or above 10% market penetration rates. Comparing the results for the 15 seconds and 45 seconds broadcasting frequencies, it becomes obvious that the broadcasting frequency and speed accuracy are indirectly related – the higher the broadcasting frequency (60 BC vs 20 BC in 15 minutes) the higher the speed accuracy.

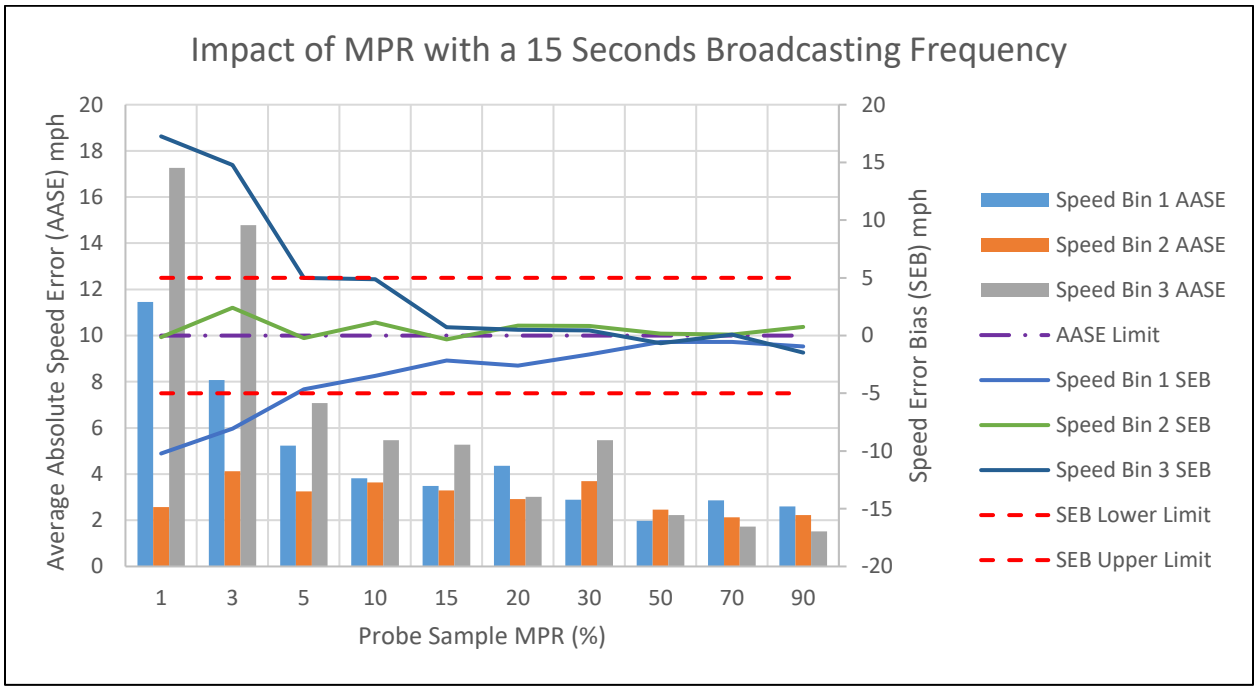


FIGURE 25: EFFECT OF MARKET PENETRATION RATE ON AASE AND SEB: 15 SECONDS BROADCAST FREQUENCY

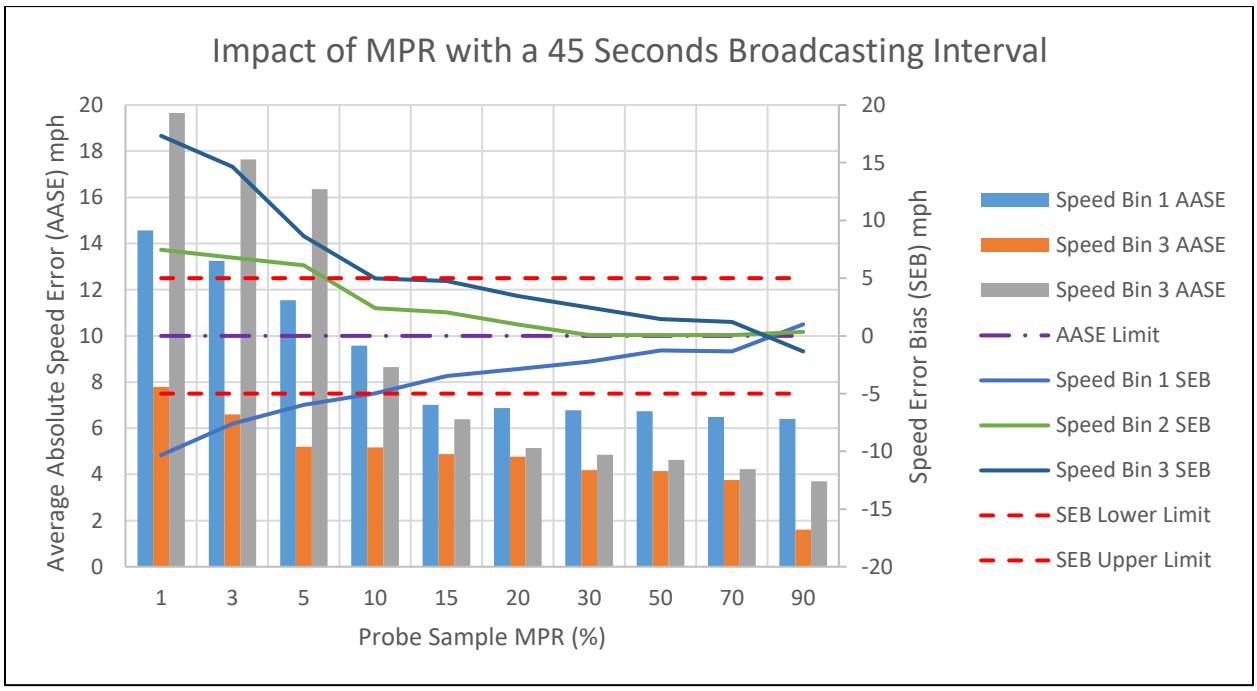


FIGURE 26: EFFECT OF MARKET PENETRATION RATE ON AASE AND SEB: 45 SECONDS BROADCAST FREQUENCY

Figure 27 shows the impact of broadcasting frequency and market penetration rate on the root mean square error (RMSE) of the reported speed. There are four bars at each market penetration rate representing the four broadcasting frequencies. Visual observations of the figure reveal the impact of the broadcasting interval. The more frequent the broadcasting interval the lower the root mean squared error (i.e. the lower the deviation of reported probe vehicle mean speed from the ground truth average speed in each time epoch). Furthermore, an increase in the market penetration rate of probe vehicles results in a decrease in the root mean squared error. This observation holds for all broadcasting frequencies. The level of decrease in root mean squared error as the market penetration rate increases is different, however, and depends on the broadcasting frequency. Those with higher broadcasting frequencies (15 and 30 seconds) tend to show a quicker decrease in the root mean squared error for the higher market penetration rates compared to those with lower broadcasting frequencies. While the root mean squared error decreases with increases in broadcasting frequency and market penetration rate, it levels off after a certain market penetration rate. For most of the broadcasting frequencies, this translates to a 15% market penetration rate.

Importantly, the various performance measures used will tend to yield different requirements in terms of probe market penetration and broadcast frequency. For example, both AASE and SEB requirements are met with a 10% market penetration rate. For AASE that is a 10 mph speed error, and for SEB a ± 5 mph speed error. To achieve the same error using RMSE requires a MPR of about 15% (with BC = 15 sec) and close to 50% (with BC =30 sec). In other words, the minimum sampling requirement depend on the selected measure of effectiveness. This is an important finding to consider in future implementations.

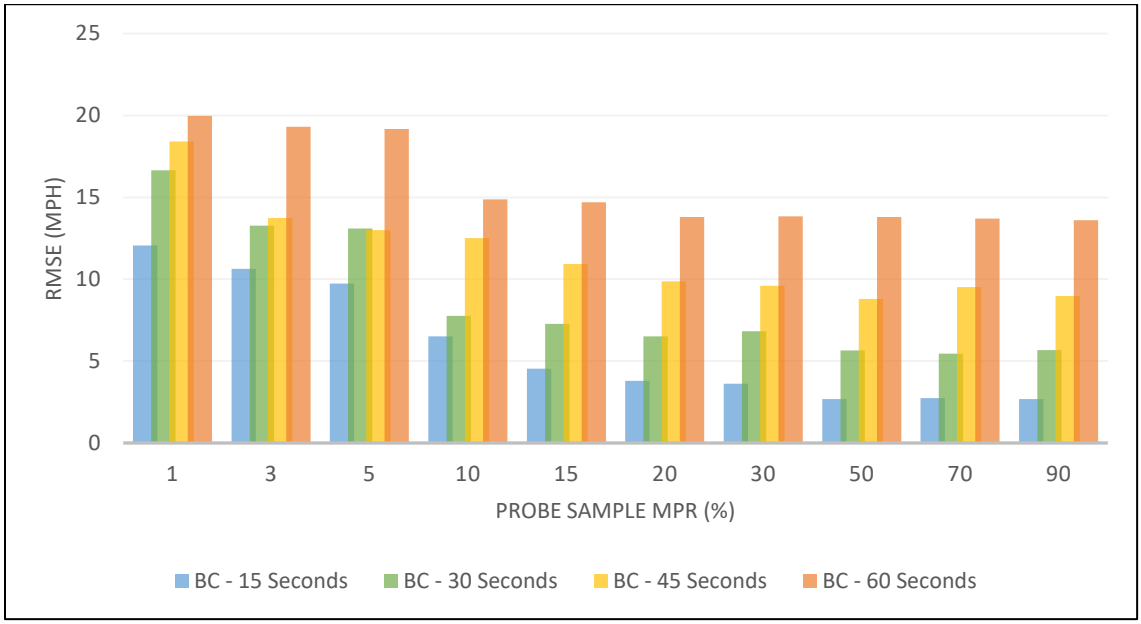


FIGURE 27: IMPACT OF MARKET PENETRATION RATE AND BROADCASTING FREQUENCY ON SPEED RMSE

8.0 SUMMARY AND CONCLUSIONS

There are over 330,000 traffic signals in the United States, more than 75% of which could improve their operations by updating their equipment or timing plans. Traditional traffic signal maintenance and operations, used by most states, rely on local knowledge, driver feedback/complaints, and long-range planning. More recent feedback methods employ performance-based strategies, such as automated traffic signal performance measures (ATSPM). As agencies gain access to more data sources for performance measurement, there is difficulty ensuring that each source is appropriately used. As such, this project focused on developing a framework in which combined performance measures can be generated through data fusion of disparate data sources. Two alternative methods to estimate approach delay at signalized intersections sprang up from the application of this framework.

The successful deployment of ATSPM depends on many factors, one of which is the ability to obtain/acquire information that accurately reflects network speed and travel time. Many mediums are available to collect speed and travel time on a network. One widely used by public and private agencies are probe vehicle speeds or travel times. Despite the rich literature on probe vehicle measures of effectiveness, the impact of broadcasting frequency coupled with the market penetration rate and report aggregation levels has not been investigated. This project investigates the use of probe vehicles for obtaining link speeds at signalized arterials and explored the combined effect of factors such as market penetration rate, broadcasting frequency, and aggregation level on the accuracy of reported speeds. Multiple metrics were used to conduct this investigation including the absolute average speed error, speed error bias, and root mean squared error.

Employing the data fusion framework in tandem with the diagnoses conducted on ATSPM metrics, two algorithms, namely arrival-departure and departure-only were developed for estimating approach delay at signalized intersections. Application of the algorithms to a Utah intersection equipped with ATSPM equipment found that both proposed methods produced higher approach delays than ATSPM, at both the cycle and 15-minutes aggregated levels. One can expect that because the proposed methods accounts for additional delay components on the approach compared to ATSPM, it will produce higher delays. The delay calculated by the Arrival-Departure usually peaked with those calculated by ATSPM and the departure-only methods. However, it is not as effective as the latter two in capturing the peak periods, especially the PM peak period. This may result from assuming stable operations during the period while it is not (almost 90% of cycles fail during this period). Furthermore, departure-only follows ATSPM very closely during the off-peak periods but gives higher delay estimates during the peak periods.

A Next-Generation Simulation (NGSIM) trajectory dataset for Peachtree Street in Atlanta, Georgia, was used to validate the two proposed methods. The validation exercise results indicate that the arrival-departure method provides estimates of approach delay that are closest to the ground truth delays compared to the ATSPM or the departure only method. The departure-only method was found to underestimate the delay systematically. The cause for

this drawback was identified as the deviation of the validation data from the required assumption for this method. The Departure-only method assumes uniform arrival throughout the cycle. However, the arrival pattern in the validation dataset was far from uniform throughout the cycle – they were mostly clustered at the beginning of the red. For cycles with uniform arrival, departure-only provided delays with high accuracy.

This report presented a framework in which enhanced signal performance measures were generated through a process of data fusion. The framework was applied to produce enhances travel time estimates for a signalized arterial. The site referred to earlier is on Peachtree Avenue in Atlanta, was part of the Next Generation Simulation (NGSIM) and included complete individual vehicle trajectories along a corridor of four signalized intersections. The detailed NGSIM data enabled the research team to develop simulated datasets acquired from multiple sources. The developed fusion framework was applied and tested with five algorithms; simple average, linear regression, K-nearest neighbors, random forest, and artificial neural networks. The simulated datasets used in the fusion process included Bluetooth, Loop Detector, Probe Vehicle, and arterial Signal Plans. Of the single source datasets, travel times obtained using Bluetooth sensors and loop detectors were the closest to and farthest from the actual travel time, respectively.

Application of the framework shows that the fusion process almost unfailingly outperforms any single source travel-time estimate. The best performer of single source data were Bluetooth sensors, producing the lowest travel time Mean Absolute Percent Error (MAPE) at 17.12%. The corresponding data fusion MAPE were 13.92%, 13.15%, 12.33% and 11.52% when using linear regression, KNN, Random Forest, and ANN, respectively.

Furthermore, the study explored the sensitivity of the fusion performance to a varying set of dataset availability where one or more of the datasets was not available. This effort found that the availability of data sources effects the output of the fusion process – the more datasets available, the more accurate the outcome of the fusion process. However, the choice of fusion algorithm played a more significant role than data availability. In fact, the ANN algorithm proved to be superior in producing better travel time estimates in most of the scenarios tested.

The key results of the sensitivity analysis, which investigates the impact of market penetration rate, broadcasting frequency, and aggregation level on the accuracy of probe vehicle speeds) indicate that both market penetration rate and broadcasting frequency significantly affect the accuracy of the reported probe vehicle speeds. Furthermore, the required minimum MPR depends on both broadcasting frequency and the measure of effectiveness used. In essence, the design requirements to meet a specific error target depends on the metric used to define the target.

In the course of our research, we utilized synthetically generated data to lay the groundwork for our methodology. This approach, while valuable for creating a controlled environment for model testing and development, does have inherent limitations. Specifically, synthetic data might not capture all biases and idiosyncrasies found in field data. Biases inherent in certain data sources, such as those related to the specific technology used for data

collection (e.g., active cellular phones or vehicle-integrated GPS units), can influence the outcome significantly.

Moreover, while our synthetic datasets are designed to capture select aspects of variability, fully representing the multifaceted nature of real-world data is inherently challenging. The intricacies and unpredictable variabilities found in actual field scenarios might not always be mirrored in our synthetic constructs.

Practitioners are urged to exercise caution when considering the market penetration rate presented in this study. While our findings offer a foundation for understanding, it's imperative to conduct additional field validation before adopting these rates in practical applications. This precaution ensures more accurate and relevant outcomes in real-world scenarios.

In light of these considerations, especially concerning the probe data study, we urge practitioners to tread with caution. It is imperative to conduct comprehensive field validations before applying our methodologies and results to real-world operational scenarios. Such validation is essential to ensure that the insights derived truly align with, and are relevant to, the complexities of real-world transportation environments.

9.0 RECOMMENDATIONS

This project developed a data fusion framework based on a functional fusion process. While this method has widely been used in the literature, exploring other fusion processes, such as architectural and mathematical, are deemed essential and could be investigated as part of future research efforts.

Furthermore, the application of the developed fusion framework to signalized arterial travel time underscored the importance of the fusion algorithm on the performance of the fusion process. Several mathematical, statistical, and data-intensive algorithms were tested. The latter class was found to outperform the first two. Expanding the list of fusion algorithms would be another potential direction for future research. Implementation of the developed fusion framework indicates that the two items influencing its effectiveness are the choice of fusion algorithm and data availability. Future research can expand the fusion algorithms to include complex artificial intelligence and machine learning algorithms and include carefully engineered features of the available datasets. In addition, the application of the framework to a site with ground truth travel-time data and field-observed fixed and mobile datasets instead of the use of simulated datasets is highly recommended.

Pertaining to ATSPM, the two developed algorithms include some critical assumptions or limitations, many of which may be addressed through future research. The arrival-departure method relies heavily on consistent actuations upstream and downstream, which cannot be achieved when the upstream detector includes additional turning movements. Thus, future work on this method should consist of more defensible and realistic addition or subtraction of upstream actuations as well as consider split failure conditions where departures are lower than arrivals for one or more cycles.

Departure-only is the only method that includes initial queue delay, though the current method assumes a constant arrival rate for consecutive cycles with split failure. This provides a minimal estimate of the initial queue delay that may be better calibrated using upstream arrival distributions. This is also the only method that requires solely downstream detection. Other candidate improvements identified include detection of platooned vs. discharging flows, handling downstream spillback, and initial headway thresholds during low flow conditions.

It is speculated that signal timing plans, traffic conditions (congested vs. uncongested), number of intersections, roadway friction, and intersection density are other factors that may impact the accuracy of reported probe vehicle speeds. Future research can investigate the impact of these factors along with factors explored in this project (market penetration rate, broadcasting frequency, and aggregation level) using a robust analytical or simulation method.

10.0 REFERENCE LIST

1. Federal Highway Administration. Automated Traffic Signal Performance Measures (ATSPMs). https://www.fhwa.dot.gov/innovation/everydaycounts/edc_4/atspm.cfm. Accessed May 4, 2020.
2. UDOT (Utah Department of Transportation). Automated Traffic Signal Performance Measures. https://udottraffic.utah.gov/atspm/. Accessed May 4, 2020.
3. Mitchell, H. B. *Multi-Sensor Data Fusion An Introduction*. 2007.
4. Wah, N. Intelligent Systems: Fusion, Tracking and Control. 2003.
5. Llinas, J., D. Hall, and M. Liggins. Handbook of Multisensor Data Fusion: Theory and Practice. 2009.
6. Bachmann, C. Multi-Sensor Data Fusion for Traffic Speed and Travel Time Estimation. 2011, pp. 29–32. https://doi.org/10.1063/1.3515207.
7. Faouzi, N. E. El, and L. A. Klein. Data Fusion for ITS: Techniques and Research Needs. No. 15, 2016, pp. 495–512.
8. Niittymaki, J., and S. Kikuchi. Application of Fuzzy Logic to the Control of a Pedestrian Crossing Signal. *Transportation Research Record*, No. 1651, 1998, pp. 30–38. https://doi.org/10.3141/1651-05.
9. Control, J. M.-T. E. and, and undefined 2002. Estimation Methods for the State of Traffic at Traffic Signals Using Detectors near the Stop-Line.
10. Friedrich, B., I. Matschke, E. Almasri, and J. Mück. Data Fusion Techniques for Adaptive Traffic Signal Control. *IFAC Proceedings Volumes*, Vol. 36, No. 14, 2002, pp. 61–66. https://doi.org/10.1016/S1474-6670(17)32396-0.
11. Zheng, L., H. Ma, B. Wu, Z. Wang, and Q. Cai. Estimation of Travel Time of Different Vehicle Types at Urban Streets Based on Data Fusion of Multisource Data. 2014.
12. Ivan, J. N., J. L. Schofer, F. S. Koppelman, and L. L. E. Massone. Real-Time Data Fusion for Arterial Street Incident Detection Using Neural Networks. *Transportation research record*, No. 1497, 1995, pp. 27–35.
13. Wolfermann, A., B. Mehran, M. Kuwahara, and B. Mehran. Data Fusion for Traffic Flow Estimation at Intersections. *International Workshop on Traffic Data Collection and its Standardisation*, No. May 2011, 2011, pp. 1–19.
14. Michalopoulos, P., … G. S.-… R. P. B., and undefined 1981. An Application of Shock Wave Theory to Traffic Signal Control. *Elsevier*.
15. Michalopoulos, P., V. P.-T. R. P. B, and undefined 1981. Derivation of Delays Based on Improved Macroscopic Traffic Models. *Elsevier*.
16. Manual, H. C. *A Guide for Multimodal Mobility Analysis*. WASHINGTON, D.C, 2016.

17. Robertson, D. TRANSYT: A Traffic Network Study Tool. 1969.

18. Chang, J., S. Muthuswamy, J. Chang, M. Talas, and S. Muthuswamy. Simple Methodology for Estimating Queue Lengths at Signalized Intersections Using Detector Data. *Transportation Research Record: Journal of the Transportation Research*, Vol. 2355, No. 2355, 2013, pp. 31–38. https://doi.org/10.3141/2355-04.

19. Kim, S. O., and R. F. Benekohal. Comparison of Control Delays from CORSIM and the Highway Capacity Manual for Oversaturated Signalized Intersections. *Journal of Transportation Engineering*, Vol. 131, No. 12, 2005, pp. 917–923. https://doi.org/10.1061/(ASCE)0733-947X(2005)131:12(917).

20. Kebab, W., M. P. Dixon, and A. Abdel-Rahim. Field Measurement of Approach Delay at Signalized Intersections Using Point Data. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2027, No. 1, 2007, pp. 37–44. https://doi.org/10.3141/2027-05.

21. Zheng, J., X. Ma, Y. J. Wu, and Y. Wang. Measuring Signalized Intersection Performance in Real-Time with Traffic Sensors. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, Vol. 17, No. 4, 2013, pp. 304–316. https://doi.org/10.1080/15472450.2013.771105.

22. Barkley, T., R. Hranac, K. Fuentes, and P. Law. Heuristic Approach for Estimating Arterial Signal Phases and Progression Quality from Vehicle Arrival Data. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2259, No. 1, 2011, pp. 48–58. https://doi.org/10.3141/2259-05.

23. Vigos, G., M. Papageorgiou, and Y. Wang. Real-Time Estimation of Vehicle-Count within Signalized Links. *Transportation Research Part C: Emerging Technologies*, Vol. 16, No. 1, 2008, pp. 18–35. https://doi.org/10.1016/j.trc.2007.06.002.

24. Vigos, G., M. P.-I. T. on intelligent, and undefined 2010. A Simplified Estimation Scheme for the Number of Vehicles in Signalized Links. *ieeexplore.ieee.org*.

25. Lee, S., S. Wong, Y. L.-T. R. P. C. Emerging, and undefined 2015. Real-Time Estimation of Lane-Based Queue Lengths at Isolated Signalized Junctions. *Elsevier*.

26. Author, 7, H. X. Liu, W. Ma, X. Wu, and H. Hu. *Development of a Real-Time Arterial Performance Monitoring System Using Traffic Data Available from Existing Signal Systems*. 2008.

27. Zheng, J., H. X. Liu, S. Misgen, K. Schwartz, B. Green, and M. Anderson. Use of Event-Based Traffic Data in Generating Time–Space Diagrams for Evaluation of Signal Coordination. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2439, No. 1, 2014, pp. 94–104. https://doi.org/10.3141/2439-09.

28. Liu, H. X., and W. Ma. Real-Time Performance Measurement System for Arterial Traffic Signals. 2008.

29. On Kinematic Waves II. A Theory of Traffic Flow on Long Crowded Roads. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, Vol. 229, No. 1178, 1955, pp. 317–345. https://doi.org/10.1098/rspa.1955.0089.

30. Liu, H. X., X. Wu, and P. G. Michalopoulos. Improving Queue Size Estimation for Minnesota's Stratified Zone Metering Strategy. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2012, No. 1, 2007, pp. 38–46. https://doi.org/10.3141/2012-05.

31. Federal Highway Administration Center for Accelerating Innovation. Automated Traffic Signal Performance Measures (ATSPMs). 2018.

32. Kashinath, S. A., S. A. Mostafa, A. Mustapha, H. Mahdin, D. Lim, M. A. Mahmoud, M. A. Mohammed, B. A. S. Al-Rimy, M. F. M. Fudzee, and T. J. Yang. Review of Data Fusion Methods for Real-Time and Multi-Sensor Traffic Flow Analysis. *IEEE Access*, Vol. 9, 2021, pp. 51258–51276. https://doi.org/10.1109/ACCESS.2021.3069770.

33. Şengül, G., E. Ozcelik, S. Misra, R. Damaševičius, and R. Maskeliūnas. Fusion of Smartphone Sensor Data for Classification of Daily User Activities. *Multimedia Tools and Applications*, Vol. 80, No. 24, 2021, pp. 33527–33546. https://doi.org/10.1007/S11042-021-11105-6/TABLES/6.

34. Qi, W., S. E. Ovur, Z. Li, A. Marzullo, and R. Song. Multi-Sensor Guided Hand Gesture Recognition for a Teleoperated Robot Using a Recurrent Neural Network. *IEEE Robotics and Automation Letters*, Vol. 6, No. 3, 2021, pp. 6039–6045. https://doi.org/10.1109/LRA.2021.3089999.

35. Ali, U., M. H. Shamsi, F. Alshehri, E. Mangina, and J. O'Donnell. Application Of Intelligent Algorithms For Residential Building Energy Performance Rating Prediction. *The 16th International Building Simulation Association, Rome, Italy, 2-4 September 2019*, Vol. 5, 2020, pp. 3177–3184. https://doi.org/10.26868/25222708.2019.210232.

36. Li, Y., Y. Yang, J. Che, and L. Zhang. Predicting the Number of Nearest Neighbor for KNN Classifier.

37. Sarker, I. H. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *SN Computer Science*, Vol. 2, No. 6, 2021, pp. 1–20. https://doi.org/10.1007/S42979-021-00815-1/FIGURES/13.

38. Saridemir, M. Prediction of Compressive Strength of Concretes Containing Metakaolin and Silica Fume by Artificial Neural Networks. *Advances in Engineering Software*, Vol. 40, No. 5, 2009, pp. 350–355. https://doi.org/10.1016/J.ADVENGSOFT.2008.05.002.

39. INRIX. INRIX Website.

40. RITIS. RITIS Webpage.

41. Samandar, M., B. Williams, and A. Wagner. Validation of Travel Time Reliability Prediction from Probe Data. 2018.

42. Gipps, P. G. The Estimation of a Measure of Vehicle Delay from Detector Output (No. Monograph). 1977.

43. Gault, H. E., & Taylor, I. G. The Use of the Output from Vehicle Detectors to Assess Delay in Computer-Controlled Area Traffic Control Systems (No. Res Rpt. 37 Monograph). 1981.

44. Sisiopiku, V. P. *Travel Time Estimation from Loop Detector Data for Advanced Traveler Information Systems Applications*. University of Illinois at Chicago, 1994.

45. Prosser, N., and M. Dunne. A PROCEDURE FOR ESTIMATING MOVEMENT CAPACITIES AT SIGNALISED PAIR INTERSECTIONS. 1994.

46. Commiittee, T. S. S., and UDOT. Big Data Challenge on Signalized Intersections. https://github.com/TSSC2019/Big_Data_Challenge_on_Signalized_Intersections. Accessed May 5, 2020.

47. Abdi, H., and L. J. Williams. *Tukey's Honestly Signiflcant Difierence (HSD) Test*. 2014.

48. Aerde, M. Van, B. Hellinga, L. Yu, and H. Rakha. *VEHICLE PROBES AS REAL-TIME ATMS SOURCES OF DYNAMIC O-D AND TRAVEL TIME DATA*.

49. Sen, A., P. (Vonu) Thakuriah, X.-Q. Zhu, and A. Karr. Frequency of Probe Reports and Variance of Travel Time Estimates. *Journal of Transportation Engineering*, Vol. 123, No. 4, 1997, pp. 290–297. https://doi.org/10.1061/(ASCE)0733-947X(1997)123:4(290).

50. Long Cheu, R., C. Xie, and D.-H. Lee. Probe Vehicle Population and Sample Size for Arterial Speed Estimation. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 17, No. 1, 2002, pp. 53–60. https://doi.org/10.1111/1467-8667.00252.

51. Hamedi, M., and S. Aliari. I-95 Corridor Coalition Vehicle Probe Project: HERE, INRIX and TOMTOM Data Validation. No. September, 2017.

11.0 APPENDICES

11.1 Appendix A – Big Data Challenge on Signalized Intersections

The project team participated in the big data challenge on signalized intersections hosted by the traffic signal systems committee of the transportation research board. The research team was selected as one of the three teams to present their findings during the mid-year meeting of the traffic signal systems committee held at Falmouth, MA during August 6-8, 2019. The products of this competition can be found in the Github repository at https://github.com/0C34N5/NCSU_TSSC2019.

11.2 Appendix B – Summary of Accomplishments

Date	Type of Accomplishment	Detailed Description
10/18	Conference Presentation	*ITS5C STRIDE Workshop:* ATSPM and other Data Sources for Performance-based Prioritization of NCDOT's Statewide Signal Retiming Program, Thomas Chase
10/18	Conference Presentation	*6th Annual UTC Conference for the Southeast Region*: Making Sense of Arterial Data: Pilot Testing and Data Fusion, Thomas Chase
1/19	Conference Presentation	*STRIDE Poster Competition, TRB*: AN INTEGRATED DATA FUSION FRAMEWORK FOR SIGNALIZED ARTERIAL PERFORMANCE MEASURES, Shoaib Samandar, Thomas Chase, Nagui Rouphail
5/19	Publication	*Report*: IMPROVING ATSPM'S APPROACH DELAY CALCULATION, Ishtiak Ahmed, Shoaib Samandar, Sumit Toshniwal, Pravek Dwivedi, Taehun Lee, Thomas Chase, Nagui Rouphail
8/19	Conference Presentation	*TRB Traffic Signal Systems Committee Midyear Meeting:* Improving ATSPM's Approach Delay Calculation, Shoaib Samandar
1/20	Conference Presentation	*STRIDE Poster Competition, TRB*: Alternative Algorithms for Improving ATSPM's Approach Delay Estimation, Shoaib Samandar, Thomas Chase, Nagui Rouphail

Planned	Publication	Alternative Methods for Improving ATSPM's Approach Delay Calculation; Shoaib Samandar, Ishiak Ahmed, Thomas Chase, Nagui Rouphail
10/21	Conference Presentation	*NCDOT's Research and Innovation Summit:* Multi-Sensor Data Fusion for Signalized Arterial Travel Time and Delay, Shoaib Samandar