

12-17-2012

Roadway System Assessment Using Bluetooth-Based Automatic Vehicle Identification Travel Time Data

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

Thomas M. Brennan
Purdue University, tmbrenna@purdue.edu

Alexander M. Hainen
Purdue University, ahainen@purdue.edu

Stephen M. Remias
Purdue University, sremias@purdue.edu

Darcy M. Bullock
Purdue University, darcy@purdue.edu

Follow this and additional works at: <http://docs.lib.purdue.edu/jtrpaffdocs>

Recommended Citation

Day, C. M., T. M. Brennan, A. M. Hainen, S. M. Remias, and D. M. Bullock. *Roadway System Assessment Using Bluetooth-Based Automatic Vehicle Identification Travel Time Data*. Purdue University, West Lafayette, Indiana, 2012. doi: 10.5703/1288284314988.

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

Roadway System Assessment Using Bluetooth-Based Automatic Vehicle Identification Travel Time Data



Christopher Day, Thomas Brennan, Alexander Hainen,
Stephen Remias, Darcy Bullock



Roadway System Assessment Using Bluetooth-Based Automatic Vehicle Identification Travel Time Data

Christopher Day, Thomas Brennan, Alexander Hainen,
Stephen Remias, Darcy Bullock

Report Number	M5b—Handbook for Application of Bluetooth Traffic Monitoring on Arterials: Final Guidebook
Contractor	Traffax Inc.
Contract Number	DTRT57-11-C-10009
Contract Term Start	12/23/2010
Contract Term End	12/21/2012
Principal Investigator	Nicholas Ganig

RECOMMENDED CITATION

Day, C. M., T. M. Brennan, A. M. Hainen, S. M. Remias, and D. M. Bullock. *Roadway System Assessment Using Bluetooth-Based Automatic Vehicle Identification Travel Time Data*. Purdue University, West Lafayette, Indiana, 2012. doi: 10.5703/1288284314988.

ACKNOWLEDGMENTS

This work was supported by Traffax/USDOT SBIR DTR57-11-C-10009. The contents of this paper reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein, and do not necessarily reflect the official views or policies of the sponsoring organizations. These contents do not constitute a standard, specification, or regulation.

ABSTRACT

ROADWAY SYSTEM ASSESSMENT USING BLUETOOTH-BASED AUTOMATIC VEHICLE IDENTIFICATION TRAVEL TIME DATA

This monograph is an exposition of several practice-ready methodologies for automatic vehicle identification (AVI) data collection systems. This includes considerations in the physical setup of the collection system as well as the interpretation of the data. An extended discussion is provided, with examples, demonstrating data techniques for converting the raw data into more concise metrics and views. Examples of statistical before-after tests are also provided. A series of case studies were presented that focus on various real-world applications, including the impact of winter weather on freeway operations, the economic benefit of traffic signal retiming, and the estimation of origin-destination matrices from travel time data. The technology used in this report is Bluetooth MAC address matching, but the concepts are extendible to other AVI data sources.

CONTENTS

1. INTRODUCTION	1
1.1 Transportation System Performance Measurement	1
1.2 Direct Measurement of System Performance	1
2. ANALYTICAL TECHNIQUES	3
2.1 Processing and Visualizing Raw Data	3
2.2 Converting Travel Time on Heterogeneous Road Segments	7
2.3 Number of Samples for Before-After Studies	8
2.4 Statistical Tests to Support Before-After Studies	10
3. BEST PRACTICES FOR SENSOR PLACEMENT FOR INTERRUPTED FLOW	16
3.1 Case Study Comparing Intersection and Midblock Sensor Locations	17
3.2 Sensor Antenna Height Impact on Detection Efficiency	18
4. UNINTERRUPTED FLOW CASE STUDY	20
4.1 Background	20
4.2 Data Collection	20
4.3 Winter Weather Impacts	20
4.4 Data Fusion	22
4.5 Conclusion	27
5. INTERRUPTED FLOW CASE STUDIES	27
5.1 Traffic Control Parameters	27
5.2 Management Problem	28
5.3 High Resolution Controller Event Data	30
5.4 Case Study: Synthesis of Controller Event Data and Travel Time Data	32
5.5 Case Study: Alternatives for Assessing Arterial Travel Times	33
5.6 Case Study: Deriving User Benefit from Travel Times	39
6. ORIGIN-DESTINATION MATRIX ESTIMATION	43
6.1 Study Background	43
6.2 Accommodating Routes with Missing Data	43
6.3 Origin-Destination Volume Observations	46
6.4 Critical Links	48
6.5 Conclusions	49
7. CONCLUSIONS	55
REFERENCES	56

LIST OF TABLES

Table	Page
Table 1.1 Test vehicle sample sizes by traffic volume	2
Table 1.2 Test vehicle sample sizes by signal density	2
Table 2.1 Looking up the value of $L(z)$ based on z , for the Kolmogorov-Smirnov statistic	15
Table 2.2 Statistical significance (at the 1% confidence level) based on statistical test for the four different comparisons of distributions in Figure 2.18	16
Table 5.1 Example high-resolution controller event data	31
Table 5.2 Median travel times	41
Table 5.3 Reductions in median travel time compared to baseline	41
Table 5.4 Summary of cost savings for alternative optimization objectives, by section	42
Table 6.1 Sample economic impact from popular special events	43
Table 6.2 Origin-destination matrix on 10/1/2011 between 1200 and 2359	46

LIST OF FIGURES

Figure	Page
Figure 2.1 Data collection on SR 37: Location of intersections and Bluetooth monitoring stations (BMS)	3
Figure 2.2 24 hours of raw data for southbound vehicles (Friday, 5/28/2010)	4
Figure 2.3 Histogram construction in Excel	5
Figure 2.4 Histogram, CFD, and descriptive statistics for southbound vehicles on Friday, 5/28/2010, 1500–1900	5
Figure 2.5 Raw data for southbound vehicles (1500–1900), 5/26/2010–7/19/2010	6
Figure 2.6 Box-whisker plot of southbound travel times (1500–1900), 5/26/2010–7/19/2010	6
Figure 2.7 Map of locations on I-65	7
Figure 2.8 CFDs of southbound <i>travel times</i> on I-65, 2/14/2010, 1400–1800	8
Figure 2.9 CFDs of southbound <i>speeds</i> on I-65, 2/14/2010, 1400–1800	8
Figure 2.10 Map of locations on SR 37	9
Figure 2.11 CFDs of southbound <i>travel time</i> on SR 37, 6/17/2010, 1500–1900	10
Figure 2.12 CFDs of southbound <i>delay</i> on SR 37, 6/17/2010, 1500–1900	10
Figure 2.13 Travel time CFDs for each 5-second offset for entire cycle length	11
Figure 2.14 Number of samples needed to differentiate travel time distributions between a simulation run using an offset of zero, versus other settings	12
Figure 2.15 Conducting a <i>t</i> -test in Excel	12
Figure 2.16 Conducting a Mann-Whitney <i>U</i> -test in Excel	13
Figure 2.17 Conducting a Kolmogorov-Smirnov test in Excel	14
Figure 2.18 Pairwise comparisons with <i>P</i> -values resulting from statistical tests for Friday data	16
Figure 3.1 Example vehicle sensor deployments	17
Figure 3.2 Sensor locations in 2009 field study	18
Figure 3.3 Saturday (0900–1300) travel time cumulative distributions	19
Figure 3.4 Traceable MAC address counts by antenna height and direction	19
Figure 4.1 Location of portable Bluetooth data collection points on I-65 in Indiana	21
Figure 4.2 Scatter plot of AVI travel times	21
Figure 4.3 Visual depiction for storm and clear conditions at 10:20 on each day	22
Figure 4.4 Space mean speed over entire study segment	23
Figure 4.5 Southbound cumulative frequency diagrams of space mean speed	24
Figure 4.6 Comparison between each segment for the corresponding storm and clear periods	25
Figure 4.7 Southbound space mean speed from mm 168 to 178 on 2/5/11	25
Figure 4.8 Fused data capturing storm event from 2/1/2011–2/2/2011	26
Figure 5.1 Definition of cycle length, offset, and split	27
Figure 5.2 An overview of the traffic signal control design process, showing two iterative feedback loops	28
Figure 5.3 Location of intersections on US 30	28
Figure 5.4 Controller database structure for US 30	28
Figure 5.5 Analysis of controller database parameters for 22 intersections	29
Figure 5.6 Timeline showing the schedules of two signal programs used on the same day for two adjacent subsystems on US 30	29
Figure 5.7 Timeline showing the schedules of two signal programs used on the same day for two adjacent subsystems on US 30	30
Figure 5.8 Fusion data approach to signalized arterial analysis	31

Figure 5.9 Explanation of the Purdue Coordination Diagram (northbound at SR 37 and 141st St., 1/18/2012, 0700–0800)	32
Figure 5.10 Example of correlation between percent on green and arrival patterns that can be observed in the coordination diagram (northbound at SR 37 and 141 st St., 1/18/2012)	33
Figure 5.11 Influence of frequent pattern changes on interface between two systems under the same cycle length (eastbound at US 30 and Austin Rd., 2/9/2012)	34
Figure 5.12 AVI travel time data collection on US 30	34
Figure 5.13 Travel times on US 30 between US 55 and Grant (2/9/2012)	35
Figure 5.14 Analysis of progression at Int. 1 (US 30 and SR 55), 2/9/2012	36
Figure 5.15 Bluetooth monitoring stations and subsection segment regimes on SR 37 south	37
Figure 5.16 Southbound SR 37 travel time analysis during the PM peak period	38
Figure 5.17 Map of the SR 37 corridor	39
Figure 5.18 Cumulative frequency diagrams of anonymous AVI travel times for alternative objective functions, Saturday, 1500–1800	40
Figure 6.1 Equipment layout and numbering regime	44
Figure 6.2 Schematic layout of network noting sensor locations	44
Figure 6.3 Bluetooth monitoring station installations	45
Figure 6.4 Interpolation of utilization for unobserved route	46
Figure 6.5 Inbound route utilization	47
Figure 6.6 Outbound route utilization	47
Figure 6.7 Travel time scatter plots for inbound traffic on Route 1	48
Figure 6.8 Travel time scatter plots for outbound traffic on Route 1	49
Figure 6.9 Travel time scatter plots for inbound traffic on Route 2	50
Figure 6.10 Travel time scatter plots for outbound traffic on Route 2	51
Figure 6.11 Travel time scatter plots for inbound traffic on Route 3	52
Figure 6.12 Travel time scatter plots for outbound traffic on Route 3	53
Figure 6.13 Travel time scatter plots for inbound traffic on Route 4	54
Figure 6.14 Travel time scatter plots for outbound traffic on Route 4	55

1. INTRODUCTION

*What gets measured gets done,
what gets measured and fed back gets done well,
what gets rewarded gets repeated.*

– John E. Jones

In recent years, new travel time data collection technologies have emerged, which have greatly increased the power of transportation system operators to automatically measure the performance of roadway and other facilities. The purpose of this book is to serve as a resource for analysts who are tasked with interpreting measured travel times from automatic vehicle identification (AVI) systems to assess the quality of service in transportation systems. While this document emphasizes roadway systems, the analysis techniques presented here could be expanded to any mode. This book is divided into six chapters that divide the material into six modules as follows.

- The introductory chapter provides a review of travel time data analysis development to date and discusses concepts of operations for different types of traffic flow.
- Chapter 2 presents a variety of analysis techniques for reducing the data into useful performance measure graphics, and for statistical before/after comparisons. Detailed descriptions of the necessary computations and data transformations are presented.
- Chapter 3 discusses issues relevant to the location of sensors, with particular emphasis on the impact of interrupted-flow facilities on measured travel times.
- Chapter 4 discusses probe vehicle identification on uninterrupted flow systems.
- Chapter 5 discusses probe vehicle identification on interrupted flow systems.
- Chapter 6 discusses probe vehicle applications to origin-destination matrix estimation.

1.1 Transportation System Performance Measurement

Transportation is the convergence of demand for movement from one location to another, and the supply of the means to make that movement possible, through various modes (passenger cars, buses, etc.). Automatic vehicle identification (AVI) data allows individual travelers to be time stamped at different parts of the transportation system, enabling the quality of service to be measured by calculating travel times between pairs of points in the system. Travel time is the quintessential performance measure, common to all modes and users, that impacts how users decide if, when, and how to travel. In addition, AVI data cast over a wide area enables background demand to be characterized by identifying travelers at various origin and destination locations. Therefore, this type of data can yield extremely valuable information to transportation system operators to better understand demand as well as to better characterize the supply they provide by mode.

There are four ways in which transportation system operational performance can be evaluated:

1. **User experience.** Because so many individuals interact with transportation systems on a daily basis, it is inevitable that a most of them form an opinion about the level of service. While opinion is often inaccurate and biased, it is important to be cognizant that this is the only information available to most system users, and it often has a profound impact on operations.
2. **Modeling.** The oldest and most common type of quantitative performance analysis relies on volume data and parameterized system information (number of lanes, saturation flow rate, etc.) to determine the degree to which the system is loaded, and thereby estimate the level of service according to a performance measure such as delay or speed. This type of modeling continues to be important because traffic volumes are relatively easy to obtain. However, the results of this analysis are highly dependent on the assumptions of the model, and the accuracy that the analyst is able to determine the model parameters.
3. **Simulation.** Similar to modeling, simulation estimates system performance. Rather than the “top-down” formulaic approach of a model, simulations take a “bottom-up” approach that models microscopic traveler (*i.e.*, vehicle and pedestrian) behavior and measures the performance characteristics for these individual data points to collectively depict the level of service in the system. As with quantitative models, the results of simulation are dependent on the assumptions regarding traveler behavior, and the accuracy of the input parameters that control the simulation.
4. **Direct measurement.** Another possibility is that the *performance* of the system (distinct from measures of demand or supply) can be directly measured through observations. A variety of quantities can be measured, with varying degrees of relevance depending on the type of system.

This book is concerned with the last of these, direct measurement, in light of recent technological innovations that have made this type of evaluation more cost-effective, and have made it possible to obtain relatively large volumes of system data. With new data sets comes the need to update analysis techniques to handle them. One of the objectives of this book is to advance some analysis concepts into more common usage, which might be new to some readers.

1.2 Direct Measurement of System Performance

In this document, we are concerned with direct measurement of system performance. The traditional tool for direct measurement is the *floating-car study* (1), where the analyst will get into a vehicle (or send a staff member to do so) and actually drive along the corridor. While this can be done at a very high resolution, such as with GPS tracking (2), the overall system performance essentially consists of the space-time positions at an entry point and an exit point, which yields a travel time, a travel distance, and from these, an average travel speed. The advantage of the floating-car study is that it does not require a substantial investment in data

TABLE 1.1
Test vehicle sample sizes, by traffic volume (3).

Average Daily Traffic (ADT) Volume Per Lane	Average Coefficient of Variation (%)	Sample Sizes		
		90% Confidence, +/- 10% Error	95% Confidence, +/- 10% Error	95% Confidence, +/- 5% Error
< 15000	9	5	6	15
15000–20000	11	6	8	21
> 20000	17	10	14	47

collection infrastructure. However, the number of samples obtained are quite low, and for moderately sized systems (*e.g.*, several miles long) it is difficult to obtain a significant number of data points within a time period.

The *Travel Time Data Collection Handbook* (3) guidelines for the number of floating car runs to be carried out to achieve a given level of statistical confidence in estimating *mean* travel times are shown in Table 1.1 and Table 1.2 respectively for ADT and density of traffic signals. A more detailed discussion is provided elsewhere by Quiroga and Bullock (4). According to Table 1.1, to characterize the travel time on a network with 3 signals per mile at 90% confidence, we would need to have 5 data points. For a short network, say 1 mile, at an average speed of 30 mph (assuming a higher posted speed limit and several stops), each floating car transit would require 2 minutes, so the entire study could be completed in 10 minutes. For a network spanning 5 miles, the amount of time would be 50 minutes, during which it is likely that the level of demand would vary. Adding a second floating car would cut the time by half, but double the cost; and it would take many more floating cars to achieve higher confidence levels in a high-volume or high-signal density environment. More importantly, however, this effort would only characterize the *mean* travel time for one particular time of day and traffic condition.

The floating-car study is the most commonly used travel time measurement technology prior to the development of sensors capable of identifying vehicles. In a floating-car study, the transportation agency drives vehicles on a corridor while tracking the time that it takes to traverse the system. This can provide a very

detailed trajectory for a single vehicle, but is rather limited in that the entire run provides only one overall travel time for the entire segment, and it covers only a limited amount of time.

One strategy to increase the number of data points is to instrument more vehicles for data collection. For example, maintenance trucks or transit buses can be outfitted with GPS devices so that their trajectories can be measured. This increases the amount of exposure of the travel time measurement. However, the number of overall travel time measurements is still rather low, and the data can be limited by the behavior of these types of vehicles. Buses, for example, make stops along the route, introducing a substantial increase in travel time that would have to be subtracted, introducing potential variance in the data.

Another strategy is to identify vehicles in traffic at different locations in the system and measure the travel time from the observation times of vehicles matched at two locations. In the past, this has been possible through license plate matching studies, which were labor intensive. More recently, new technologies have emerged that have enabled *automatic* vehicle identification (AVI), making the acquisition of these data sets much easier and less expensive—as well as enabling 24-hour coverage.

With the ability to generate large data sets, there is now an opportunity to not only compute an average or median travel time, but also to characterize the variability (or reliability) of travel time. It is possible to establish measures of system performance on the basis of not only average values, but more specific criteria such as the 90th percentile of delay.

The objective of the document is to advance the state of the practice in the analysis of this data, by

TABLE 1.2
Test vehicle sample sizes by signal density (3).

Traffic Signal Density (signals per mile)	Average Coefficient of Variation (%)	Sample Sizes		
		90% Confidence, +/- 10% Error	95% Confidence, +/- 10% Error	95% Confidence, +/- 5% Error
< 3	9	5	6	15
3–6	12	6	8	25
> 6	15	9	12	37

demonstrating that the resolution of this data affords a more useful characterization of travel time than measures of central tendency. This book focuses on applications of AVI data sets in travel time measurement applications on non-interrupted-flow (limited-access highways) and interrupted-flow roadways (surface streets), as well as origin-destination studies. The technology used to generate the data for the observations in this study is the matching of Bluetooth MAC addresses, but the principles described in this document can be applied to any AVI travel time measurement methodology.

2. ANALYTICAL TECHNIQUES

This chapter discusses various techniques for describing and analyzing data sets for travel time and travel time reliability, and to facilitate before/after studies. The information in this chapter enables the analyst to conduct basic comparison tests using automatic vehicle identification (AVI) travel time data with spreadsheet tools. Actual spreadsheets that can be used for such analysis are explained in detail to assist in the setup of an analysis tool from a blank slate.

2.1 Processing and Visualizing Raw Data

The example data set used in this chapter is taken from a section of State Road 37 in Fishers, Indiana. Figure 2.1 provides a map of the section. The distance between the two Bluetooth monitoring sensors (BMS) in this study was approximately 3 miles, and the posted speed limit on SR 37 is 55 mph. For simplicity of discussion, we will presume that the free flow speed is also 55 mph; this means that the free flow travel time on this section is 3.3 minutes. Four signalized intersections exist along the route.

Figure 2.2 shows a 24-hour view of measured travel times for southbound vehicles in this system. The vertical scale of the figure has been extended to illustrate the large number of outliers present in the raw data. There are many measured travel times well above 10 minutes, which is about three times as long as the travel time at the speed limit. These outliers almost certainly represent travelers briefly leaving the road (*e.g.*, at a gas station, coffeehouse, etc.) between the two measurement points.

To describe the throughput performance of this roadway section, we are generally not interested in these data points, but instead want to know the travel times of vehicles that did not depart and reenter. In this particular example data set, the range of “typical” travel times is unambiguously represented by the dense point cloud hovering just above the 3.3 minute line, which is distinct from the outliers well above it. However, in longer corridors or congested conditions, the travel times of interest may extend into the domain of outliers. Some travelers may have experienced very poor travel times because their vehicle travelled at a different speed from what was used to coordinate traffic signals (such

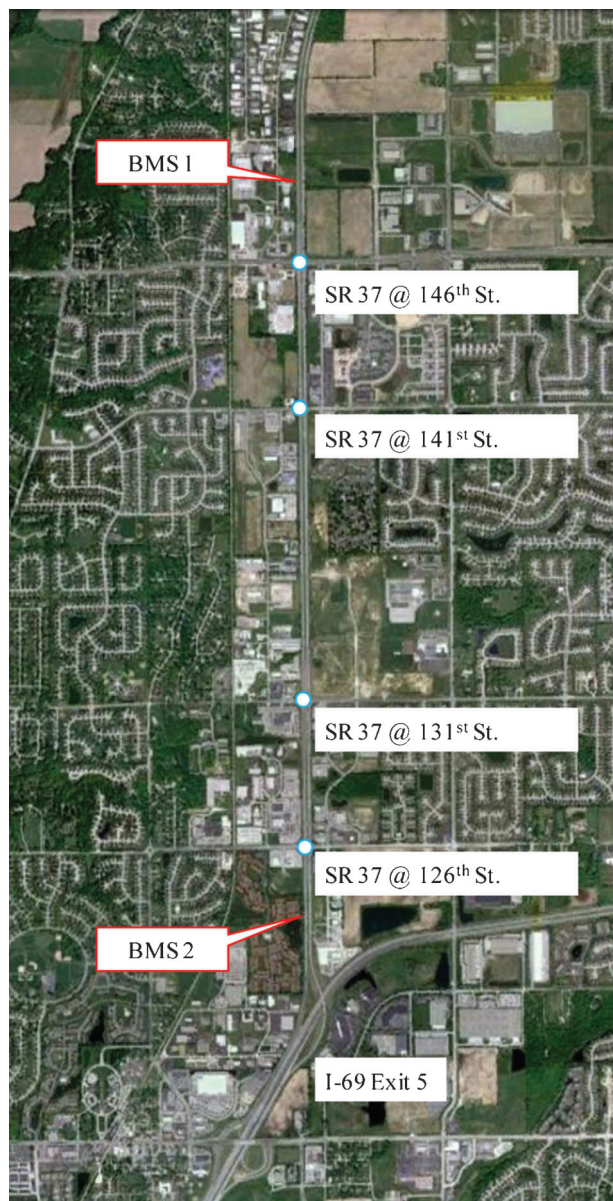


Figure 2.1 Data collection on SR 37: Location of intersections and Bluetooth monitoring stations (BMS).

as a freight truck). Even in these cases, the greater travel times generated are not indicative of the experience of the larger group, and are generally referred to as outliers.¹

In Figure 2.2, there appear to be a convenient cutoff region at approximately six minutes. However, we would like to be able to compare this particular data set

¹Note that the scope of this discussion is throughput performance. Most roadways apart from freeways provide accessibility to goods and services. The authors acknowledge that outliers that do not reflect the throughput characteristics of the roadway may contain vital information concerning its ability to access goods and services, such as for an origin-destination study—as will be investigated in Chapter 6.

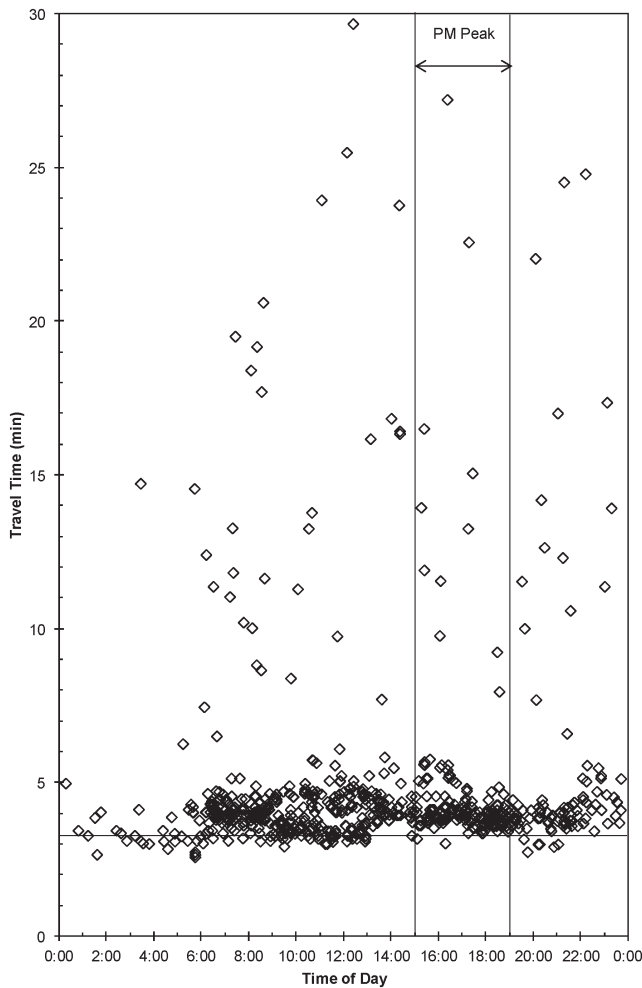


Figure 2.2 24 hours of raw data for southbound vehicles (Friday, 5/28/2010).

with others in which there might be increased delay. It is preferable to use a filter based on a reasonable assumption that can be more generally applied, rather than use arbitrary cutoff points. Although it may be possible to develop some sort of self-filtering methodology that uses properties of the data to exclude outliers, presently a resource is unlikely not immediately available to an analyst that is handed a set of raw travel time data. Rather, some engineering judgment is needed, which should have as its primary consideration *the objective of the study*. In this example, we will leverage information about the signal timing plan.

In this study, let's first define the objective: to measure the performance of the signal timing plan under typical conditions. This means that incidents, which would exceptionally skew the data, are not our primary interest. Therefore, it is not our interest to explore travel times of 20–30 minutes or more. While that data would be interesting to understand the impact of an incident, if we happened to collect such a data set it would not fulfill our need to understand typical conditions.

In this signal system, the longest cycle length (and the one that is used on Friday afternoons) is 116 seconds.² Without getting into details of the control plan, let us assume, for simplicity, that the absolute minimum green time for any arterial movement is 40 seconds. A motorist with exceptionally poor luck who is the first vehicle to stop on red at every signal would encounter an increase in travel time of approximately $(116 - 40) \times 4 = 304$ seconds, or 5.1 minutes. When added to the baseline travel time of 3.3 minutes, this suggests a “worst-case” travel time of 8.4 minutes. Rounding up to 9 minutes provides a useful and rather cautious filter for this signal system.

As with all estimates based upon “engineering judgment,” this filter is attached strongly to its context, namely the study objective and the characteristics of the data set. Recurring oversaturation, special events, or road construction could substantially increase typical travel times beyond 9 minutes. The analyst must be cognizant of such possibilities when considering how to eliminate outliers.

Having filtered the data, it is now possible to use it to characterize travel time and travel time reliability during a given time period. Figure 2.3 gives an overview of how a histogram and CFD can be quickly constructed in a spreadsheet.

The filtered raw data is listed under column A; column B is an (optional) conversion from seconds to minutes. Column C is a function that provides the bin value of each travel time (a function that *round down* would also work). A completely separate range of cells in columns E-H produces the histogram data. The desired bin values are entered in column E; column F contains a function that counts the number of bins in column C that are equal to the bin value (where the range extends down to row 1000 in this example); column G normalizes these by dividing by the total number of entries in that same range; and finally column H produces the CFD by, in each row, adding the current row value in column G to the prior row's value in column H.

The resulting histogram, CFD, and descriptive statistics from this data set are presented in Figure 2.4. All of the statistical values are prepared by applying different functions to column B of the spreadsheet shown in Figure 2.3. Both the histogram and CFD show that the travel time distribution is skewed slightly to the right, with a heavy tail on the upper end of the distribution. This is a rather common shape for travel time distributions to have. Because of the skew, the mean can sometimes be a rather poor measure of central tendency. In Figure 2.4, for example, the core of the distribution appears visually centered around approximately 4 minutes. The mean travel time is 4.12 minutes, which is slightly to the

²If the signal timing data is unknown, or a fixed cycle length is not used, the analyst may choose to make an educated guess. Typical cycle lengths vary by agency and system, but 120–180 seconds would be a reasonable range for most arterial systems, while 60–120 seconds is reasonable for central business district areas.

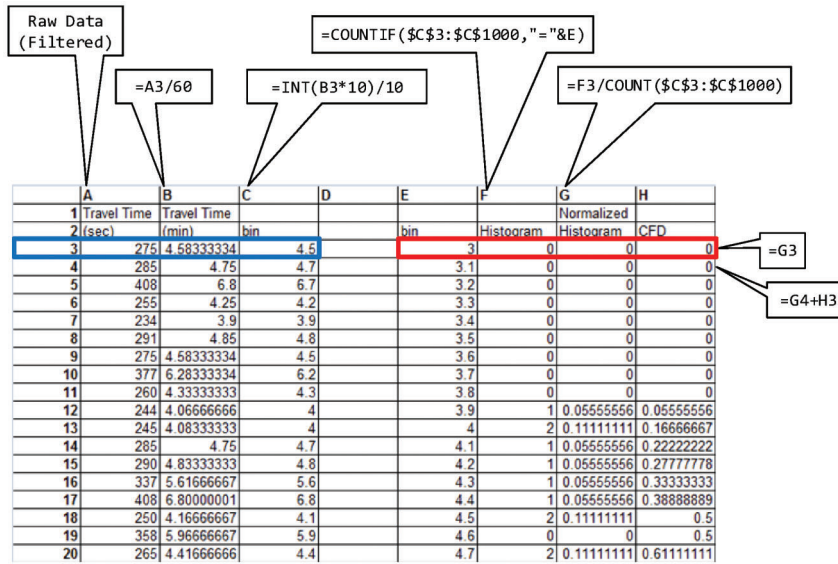


Figure 2.3 Histogram construction in Excel.

right—whereas the median (50th percentile) value of 3.96 minutes is a better reflection of the central tendency of this distribution. In this example, these distinctions might seem rather trivial, but on longer corridors with more variability, the difference can be substantial. The use of median values is similar to how income distributions, which have similarly heavy tails, are usually discussed.

With regard to the *reliability* of the data, this chapter recommends that the analyst consider the interquartile range (IQR). The IQR is defined as the difference between the 75th and 25th percentiles, which is the range of variation experienced by 50% of the sampled vehicles. This value can be somewhat more informative than standard deviations, when the underlying data

does not match a normal (Gaussian) distribution very well.

The 25th percentile, median, and 75th percentile are also helpful to compare large numbers of data sets against each other by using a “box-whisker” plot, as illustrated in the next series of figures. In Figure 2.5, the raw data for southbound travel times are plotted over the entire data collection period spanning from 5/28/2012 through 7/19/2010 (data from the first several days of June were lost because of an equipment malfunction). Only vehicles entering the system between 1500 and 1900 are included in this visualization. While the plot of the raw data gives a general sense of the range of the point cloud, the central tendency of the distribution is rather hard to distinguish.

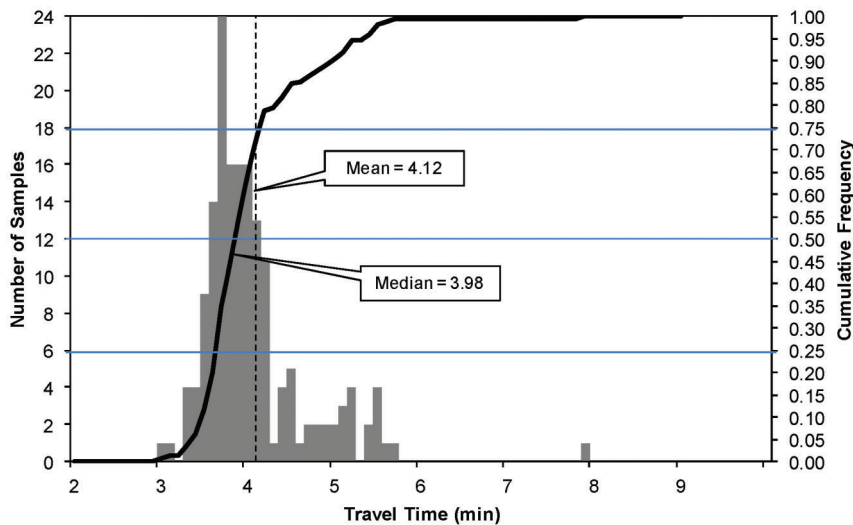


Figure 2.4 Histogram, CFD, and descriptive statistics for southbound vehicles on Friday, 5/28/2010, 1500–1900.

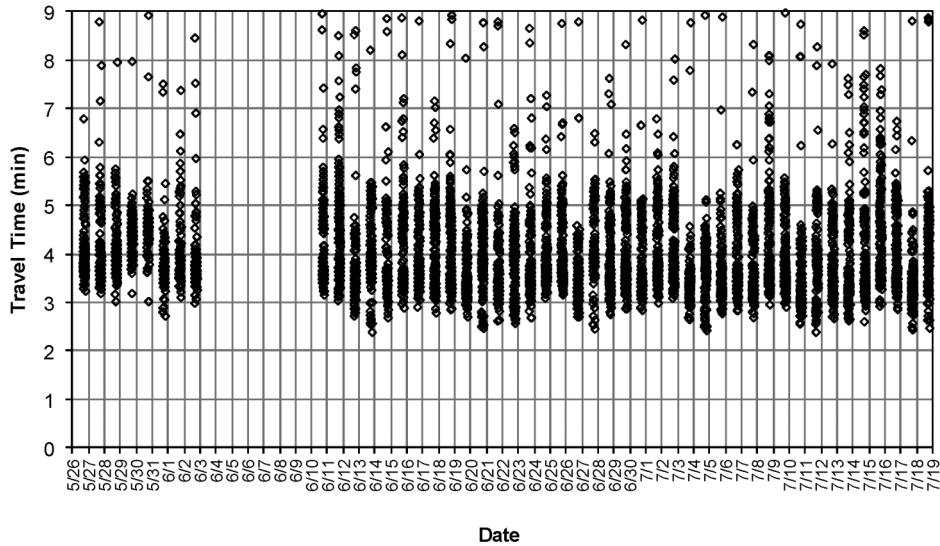


Figure 2.5 Raw data for southbound vehicles (1500–1900), 5/26/2010–7/19/2010.

Figure 2.6 shows a box-whisker plot representing the same data. The “box” in these plots shows the 25th percentile (bottom), median (bar in the middle), and 75th percentile (top). The standard box-whisker plot uses the min and max values for the “whiskers.” However, with this type of data, it is more useful to use the 5th and 95th percentiles since the absolute minimum values tend not to vary from day to day, and the maximum values are not very meaningful in any case, because they are determined primarily by how the data is filtered for outliers. By assigning the 5th and 95th percentile values to the whiskers, they represent the range of 90% of the observations, while the box represents the range of 50% of the observations. This representation of the data makes it easier to visually

detect trends, particularly in the IQR (which corresponds to the height of the boxes), that can only be guessed at from the raw data. Notably, it is clear from Figure 2.6 that on most days, the median travel time is closer to the 25th percentile than the 75th percentile—therefore it is *not* usually in the center of the point cloud (as might be implied by Figure 2.5). Additionally, the upper whisker is almost always longer than the lower whisker. This reveals that the upper tails of the distribution are fairly extended, most of the time.

We are also able to point out certain days of the week in which travel time could be considered “reliable” for this movement. 6/12/2010, 6/19/2010, 6/26/2010, and successive Saturdays (highlighted by arrows) all have not only shorter travel times than the rest of the week, but

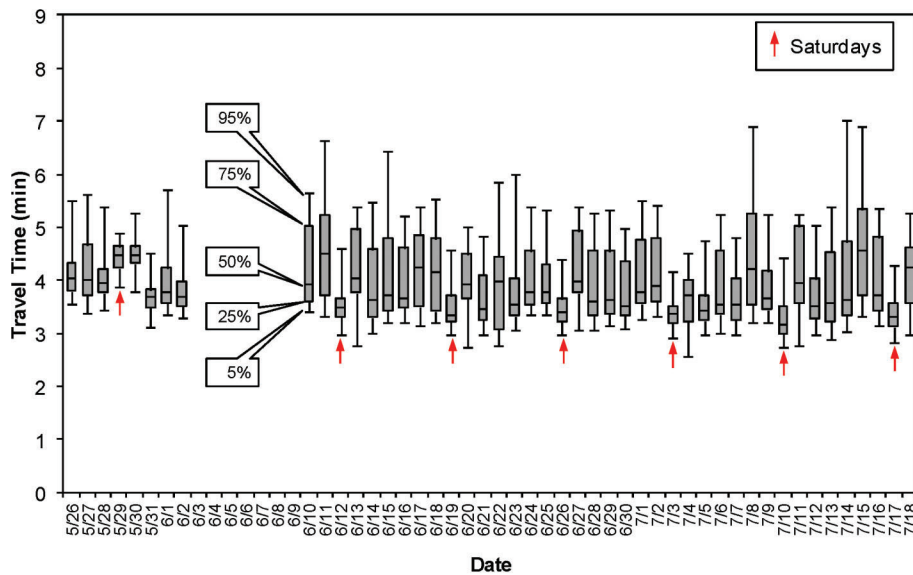


Figure 2.6 Box-whisker plot of southbound travel times (1500–1900), 5/26/2010–7/19/2010.

reduced IQR, as well as reduced ranges spanning the 5th and 95th percentiles. Interestingly, the Sundays that occur just after them have rather different characteristics. This is a good example of how signal control can influence the operating characteristics of a system; a different timing plan is used on Sundays on this corridor.

2.2 Converting Travel Time on Heterogeneous Road Segments

When comparing multiple roadways, different sections on a roadway, or different directions on the same roadway, it is difficult to draw useful conclusions when the free flow travel times for the routes are unequal, and therefore the CFDs of the different routes have some intrinsic differences—namely, the travel time on one route will always be higher or lower, and therefore its central tendency will *always* be to the right or left.

To deal with this problem for comparing heterogeneous routes, there are two possibilities:

- Travel time can be converted to speed by dividing the route distance by the travel time (*e.g.*, dividing the distance in miles by the travel time in hours gives the speed in miles per hour). This conversion is more relevant to uninterrupted-flow facilities (freeways).
- Travel time can be converted to delay per mile by subtracting the free flow travel time (the distance divided by the free flow speed) from the measured travel time, and dividing by the length of the road in miles. This conversion is more relevant to interrupted-flow facilities (surface streets).

An example is presented in the following series of figures. Figure 2.7 shows a map of two sections on I-65 south of Lafayette, Indiana. Three travel time sensors were situated on the corridor at mile marker (mm) 166.0, mm 163.3, and mm 159.3. The distances of the two sections created by these sensors are indicated on Figure 2.7. The example travel time data shown here was measured on 2/14/2010 from 1400–1800.

Figure 2.8 shows the CFDs of travel time on the two sections. As we would expect, the travel time on the shorter section is lower than that of the longer section (*i.e.*, the CFD for the shorter section is to the left). While it may be possible to make some inferences about the shape, this is challenging without putting the travel time into context.

Figure 2.9 shows the CFDs of speeds. This allows the two sections to be compared. The speed limit of 70 mph is shown on the graph. The graphs indicate that the median speed is closer to 75 mph. Speeds are slightly lower on the section between mm 163.3 and 166.0, likely because of the curve on that section (Figure 2.7).

To illustrate the conversion of travel time to delay per mile, an arterial example is more appropriate than a freeway example. Figure 2.10 shows a map of two sections on SR 37 in Noblesville, Indiana. Three travel time sensors are indicated in the figure. Travel time data was collected on 6/17/2010 from 1500 to 1900.

Figure 2.11 shows the CFDs of travel time; As before, vehicles on the longer section have greater travel time. However, the difference is rather substantial at

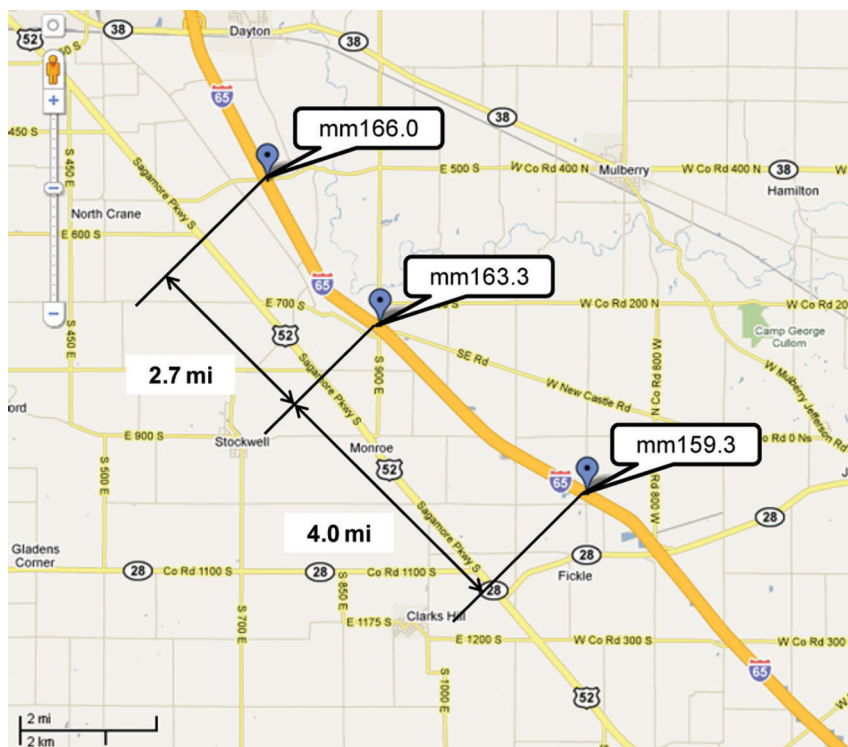


Figure 2.7 Map of locations on I-65.

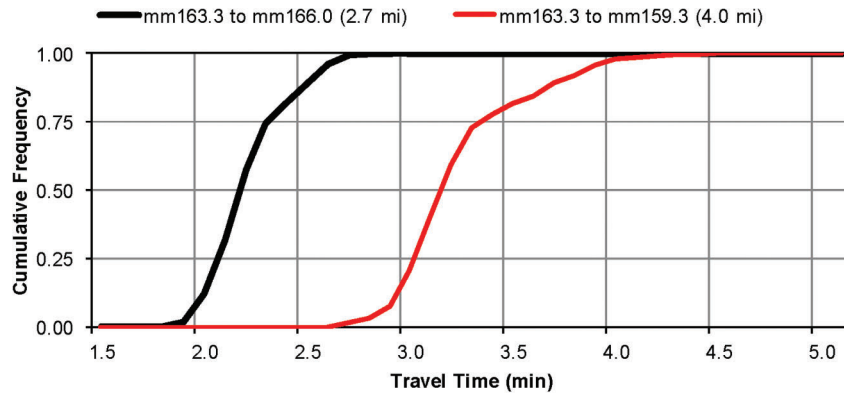


Figure 2.8 CFDs of southbound *travel times* on I-65, 2/14/2010, 1400–1800.

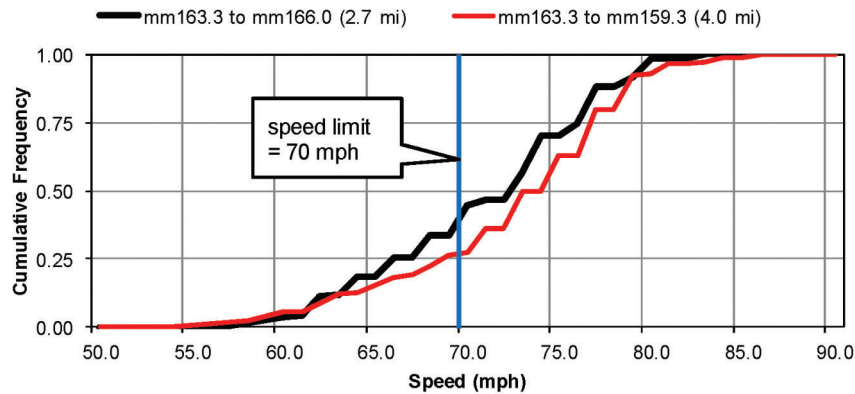


Figure 2.9 CFDs of southbound *speeds* on I-65, 2/14/2010, 1400–1800.

almost two minutes, whereas the difference in the section lengths is only 0.4 miles. Figure 2.12 shows the CFDs of the delay obtained by subtracting the free flow travel time (at the posted speed limit of 55 mph) and dividing by the length of each roadway. The shapes of the curves are almost the same, and although the curves are a bit closer to each other, it is clear that there is considerably more delay on the segment between BMS-1 and BMS-2.

2.3 Number of Samples for Before-After Studies

Often, it is desirable to compare two data sets in order to ascertain the value of an investment: a geometric improvement, traffic signal retiming, signal or roundabout installation, and so forth. An important question relevant to before-after studies is, *how many samples are needed to be able to reasonably differentiate between two different travel time conditions?*

To seek an answer to this question, a study was recently conducted (5) using simulation, where all of the travel times could be exactly known, to generate a family of data sets under varying conditions. The offsets at a particular intersection were transitioned through a range of possible values, producing some conditions that were highly similar to the baseline (when the difference between offsets was small), and

other conditions that were completely different (when the difference was great). This was done in a simulation environment to be able to compare subsamples against the entire population. Concepts from information theory were then applied to the data to determine the number of samples needed to be able to statistically differentiate two different data sets for a given confidence level.

Figure 2.13a shows how the travel time CFDs evolve with the signal offset. As will be discussed in more detail in Chapter 5, it is possible to distinguish between those CFDs that represent platoon arrival prior to the start of green (shown in Figure 2.13b), in which most vehicles experience some queuing delay; and those representing platoons being cut off by the end of green (Figure 2.13b), which take on bimodal distributions that indicate stopped and unstopped vehicles.

The results of the study are shown in Figure 2.14. This plot shows, for a given offset, the number of samples required to distinguish, at a 90% confidence level, the difference between conditions at that offset, and a “baseline” condition where the offset is zero. The results show that the more substantial the difference in conditions, the fewer samples needed. The minimum point of the graph occurs at an offset of 60 seconds. The difference between a 60-second offset and the baseline is equivalent to having all vehicles arriving at a red signal,

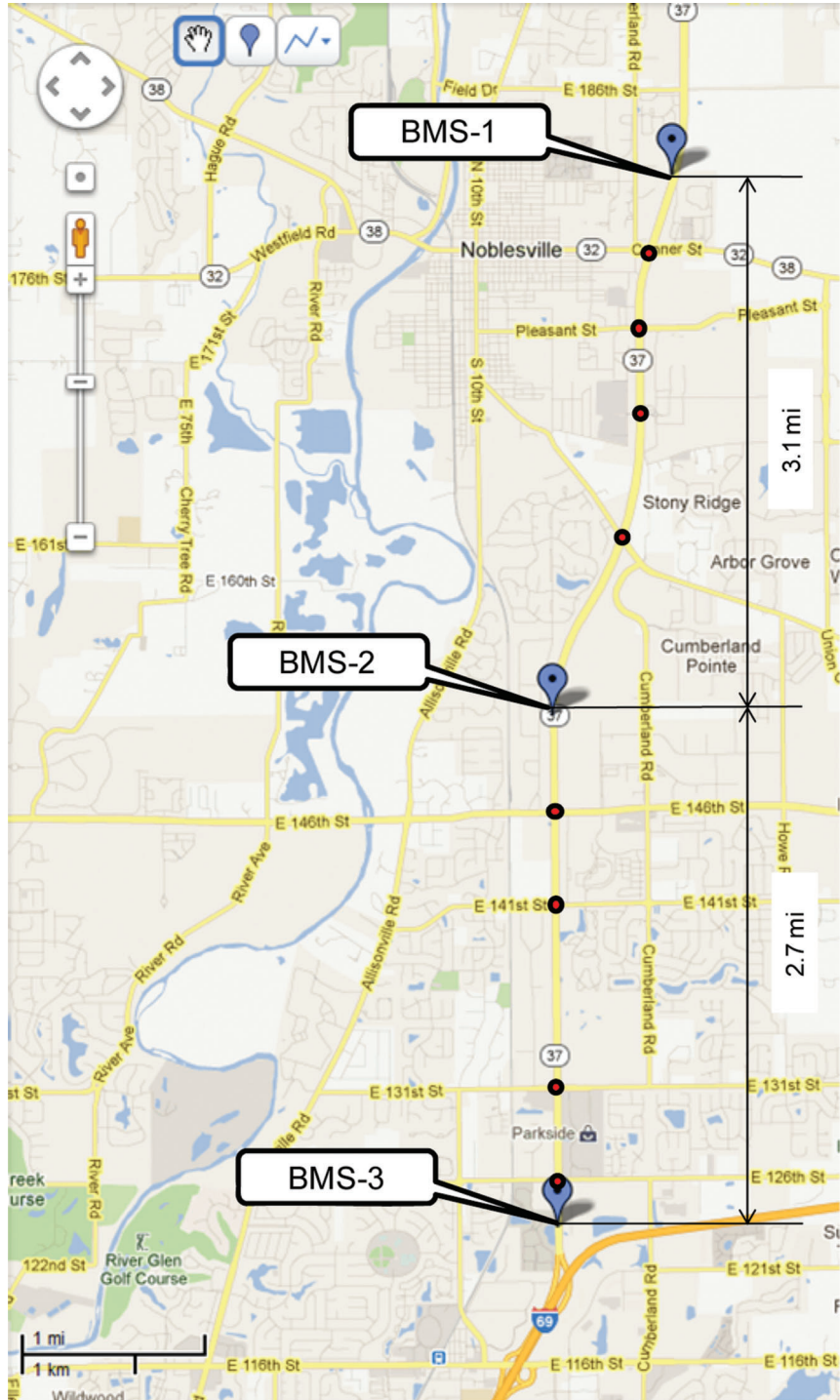


Figure 2.10 Map of locations on SR 37.

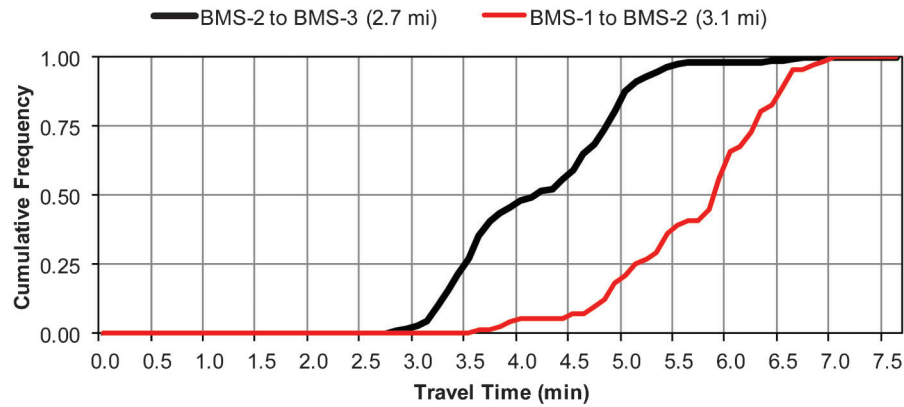


Figure 2.11 CFDs of southbound *travel time* on SR 37, 6/17/2010, 1500–1900.

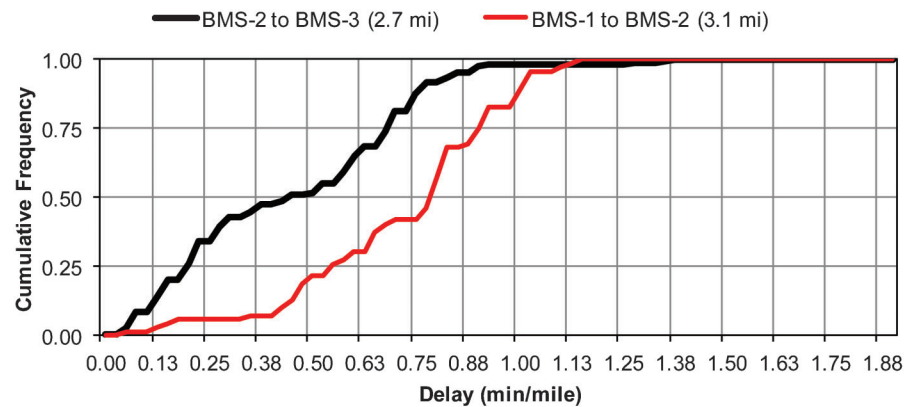


Figure 2.12 CFDs of southbound *delay* on SR 37, 6/17/2010, 1500–1900.

versus having all vehicles arriving on green.³ In this case, a very small number of samples are needed to distinguish the difference because the change in central tendency is very large. For smaller changes in offset, representing more subtle differences, more samples are needed. The *traditional* numbers of samples recommended for travel time studies (Table 1.2) are shown as horizontal lines. Except for the most drastic changes to the offset, these sample sizes are insufficient.

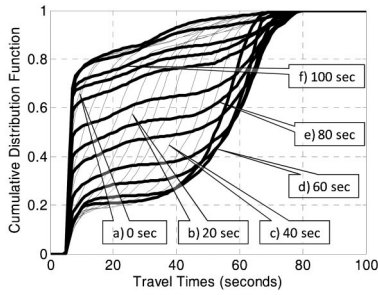
Ultimately, the number of samples needed relates to the desired resolution of changes that the analyst wishes to be able to confidently distinguish. To statistically verify major incidents or other extreme changes in conditions, only a few samples are needed. Changing the offset from 0 to 60 is roughly equivalent to causing vehicles to stop where they did not stop before. Only five samples are needed—a number that can rather easily be obtained in a floating-car study (Table 1.2). To investigate more subtle changes in the facility operation, more samples are needed. For the purpose of determining the travel time impact of changes in

signal control, these results show that around 40–60 samples would generally be needed to detect 10-second shifts in offsets, which would approximately correspond to 10-second shifts in travel time. It would be quite expensive to collect this amount of data using traditional floating car techniques. This illustrates one of the reasons why AVI data collected from Bluetooth MAC address readers is so attractive for collecting operations oriented data.

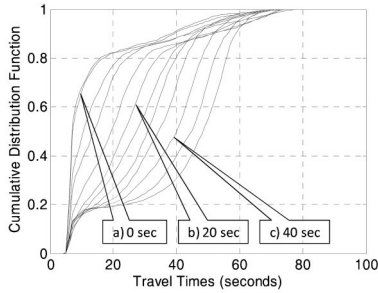
2.4 Statistical Tests to Support Before-After Studies

It is often desired to compute the statistical significance of a test result, which lends additional credibility to the outcome of an action taken to influence travel times. The meaning of a test result should be considered carefully; however, the lack of statistical “significance” is not always synonymous with the lack of substantial impact (6). A statistical test will determine, to a specified degree of confidence, whether the different characteristics of two samples are *unlikely to have resulted from random variation*. A statistical test is based on the concept of the “null hypothesis,” that there is no statistical difference between two distributions. The test is only capable of *rejecting* or *not*

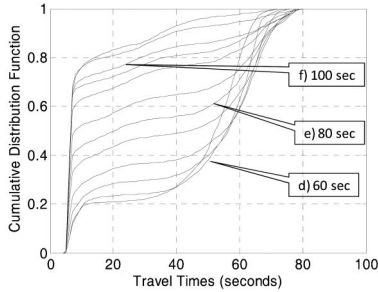
³The curve begins moving upward again above 60 seconds, because the arrivals begin moving back into green. The meaning of signal offset is described in more detail in Chapter 5.



(a) Travel time CFDs for each 5-second incremental offset adjustment.



(b) Platoon arrival before start of green.



(c) Platoon cut off by end of green.

Figure 2.13 Travel time CFDs for each 5-second offset for entire cycle length. The travel time distributions associated with early arrivals is from one family of distributions and the late platoon arrivals are from a different family of distributions (5).

rejecting the null hypothesis. Failure to “pass” the test only shows that the effect is not determined to be statistically significant at that particular confidence level.

Most statistical tests are intended for normally distributed data (*i.e.*, it is assumed that the measured parameter in the underlying populations follows a Gaussian, or normal distribution). Student’s t -test is an example. These tests lose power when the data is increasingly different from normal, meaning that they might not detect a statistically significant difference between two data sets, even when the distributions have visually apparent differences. When data sets are clearly not normally distributed, an alternative would be to use a nonparametric test. In general, however, nonparametric tests are less “powerful” than parametric ones, in the statistical sense, where “power” means the probability that the test correctly rejects the null hypothesis. This means that nonparametric tests generally require larger sample sizes to be able to draw conclusions with the same degree of confidence.

Most spreadsheets have internal functions to carry out a t -test. Figure 2.15 shows an example spreadsheet t -test, showing how to obtain the P -value (cell B2), as well as the degrees of freedom (cell B3) and t -value (cell B1), if desired. In this case, we seek a low P -value to reject the null hypothesis that the two distributions are drawn from the same population. In this case, the P -value of 2.24×10^{-7} is very likely to be significant (meaning that we reject the null hypothesis, and conclude that the two samples are drawn from different population distributions), depending on the choice of the selected α value for significance. For example, if we choose $\alpha = 0.01$, or a 99% confidence level, this result would be considered statistically significant because $2.24 \times 10^{-7} < \alpha = 0.01$.

The Mann-Whitney U -test (7) is a non-parametric test for differentiating between two distributions that does not require the underlying data to be normally distributed. We present this test here because it is relatively straightforward to carry out in a spreadsheet, and is a rather common nonparametric test that is used for similar situations as the t -test—namely, for comparing independent observations that can be rank-ordered. As with the t -test, we seek a low enough P -value to reject the null hypothesis that the two distributions are drawn from the same population. This test relies upon a rank-ordering (sorting of values from least to greatest) of all of the data in each data set with respect to the combined data sets; the sum of the ranks is then compared to the numbers of samples in each group to compute the test statistic. Figure 2.16 explains how to carry out this test in a spreadsheet. Note that the P -value is slightly higher at 1.23×10^{-5} .

Another nonparametric test that is highly sensitive for detecting horizontal differences between two curves is the Kolmogorov-Smirnov test (7). To conduct the test, the D -statistic is first calculated. D is the maximum vertical distance between the two cumulative frequency diagrams, which can be calculated as shown in Figure 2.17. A critical D -value (D_{crit}) is calculated by:

$$D_{crit} = z \sqrt{\frac{n_1 + n_2}{n_1 n_2}} \quad (1)$$

where n_1 and n_2 are the numbers of *bins* in the two CFDs, and z is a quantity that is related to another value called the limiting cumulative distribution function, $L(z)$. The formula to compute $L(z)$ is iterative, and consequently difficult to set up in a spreadsheet. Instead, a lookup table can be used to find the value of $L(z)$. From this, a P -value can be calculated simply by:

$$P = 1 - L(z) \quad (2)$$

Figure 2.17 shows an example of spreadsheet formulas necessary to execute the comparison. The data table that drives the lookup function to find $L(z)$ is shown in Table 2.1.

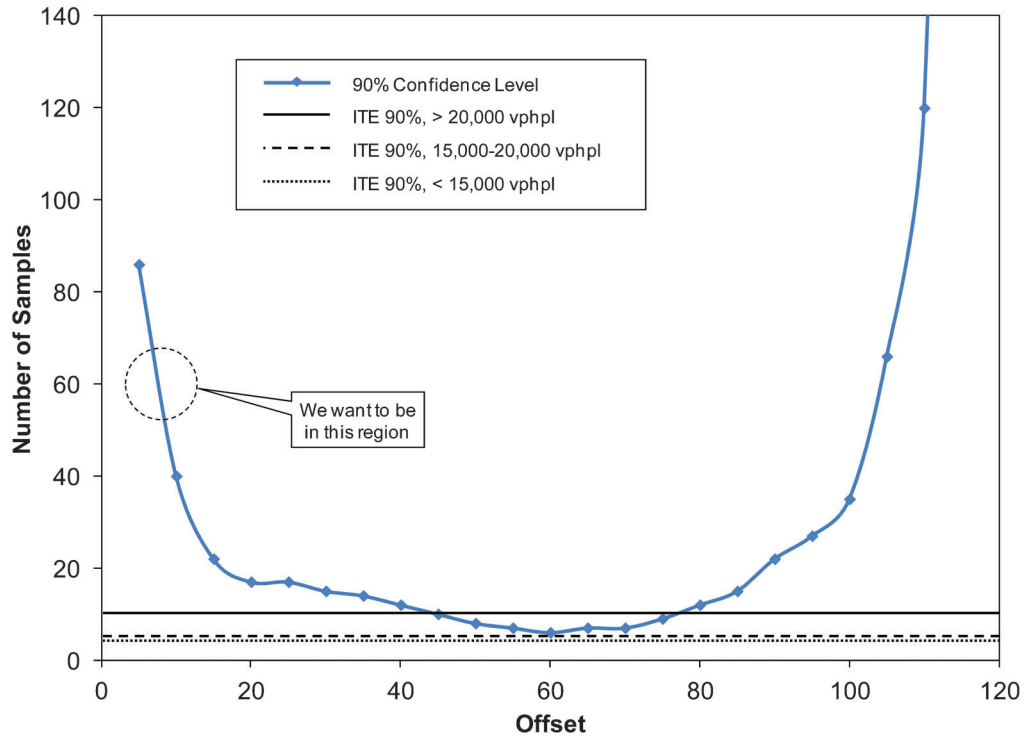


Figure 2.14 Number of samples needed to differentiate travel time distributions between a simulation run using an offset of zero, versus other settings. The cycle length is 116 seconds. (Data source: (5).)

	A	B	C
1	t-value	5.299777113	
2	p-value	2.24575E-07	
3	d.f.	301	
4		count 1	count 2
5		164	139
6		data 1	data 2
7		5.68	6.57
8		5.15	4.25
9		5.58	4.25
10		3.78	3.73
11		3.78	3.92
12		5.15	3.98
13		3.18	4.93
14		5.07	4.77
15		4.22	3.48
16		3.78	3.15
17		4.98	6.63
18		4.13	4.33
19		3.87	3.92
20		4.57	3.30

Figure 2.15 Conducting a *t*-test in Excel.

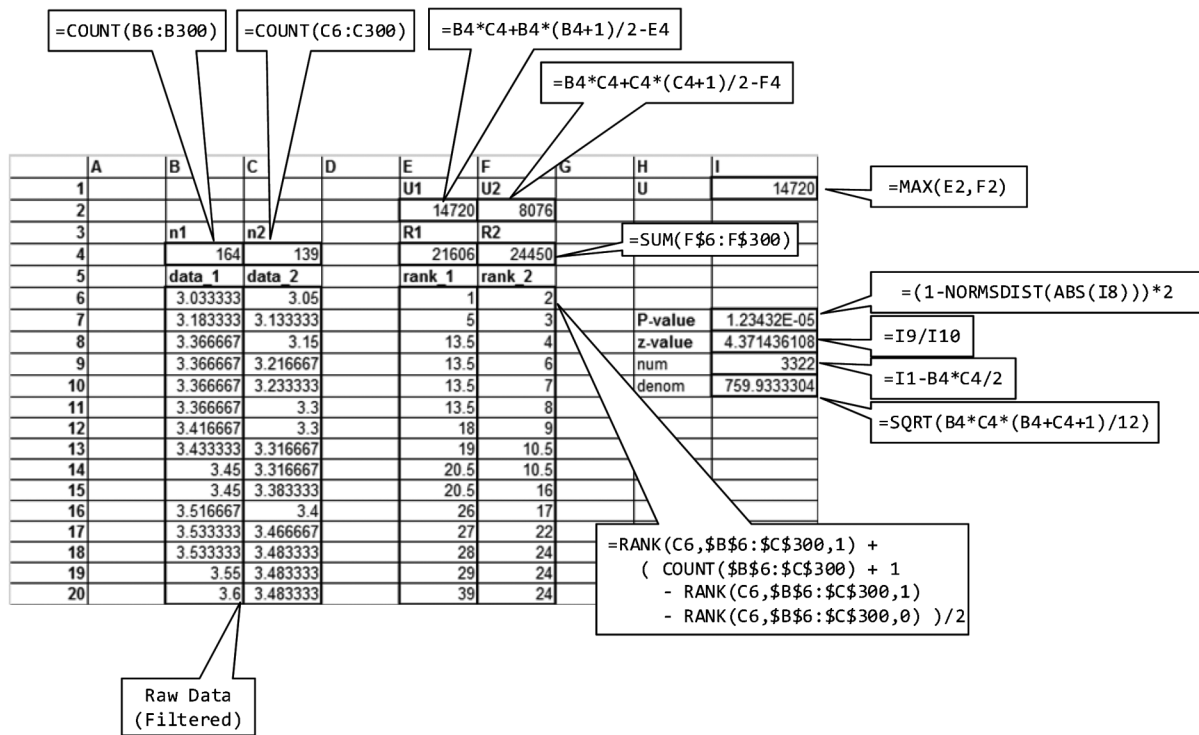


Figure 2.16 Conducting a Mann-Whitney U -test in Excel.⁴

Example test statistics are shown for four different before-after comparisons of travel time in Figure 2.18. The data used in these examples represent travel times in the signalized arterial network shown in Figure 2.1, and were collected over a series of weeks in May through June 2010. The “before” data (D1) is the same as previous shown in Figure 2.4. Four different “after” data sets (D2, D3, D4, D5) were collected on subsequent Friday afternoons during which various alternative signal timing plans were tested. The purpose of this test exercise was to test the impact of various optimization objectives (25).

Each plot in Figure 2.18 shows the resulting P -value obtained from the t -test (P_{TT}), the Mann-Whitney U -test (P_{MW}), and the Kolmogorov-Smirnov test (P_{KS}). Conclusions of statistical significance using a 1% confidence level are summarized in Table 2.2.

- In Figure 2.18a, the distribution shifts to the right (an increase in travel time), and all three tests find a highly statistically significant result.
- Figure 2.18b shows that the distribution changes shape to exhibit bimodal characteristics. The median travel time remains approximately the same. The t -test and Mann-Whitney test do not find a significant result, while the Kolmogorov-Smirnov test finds a result that would be significant at the 1% level.

- Figure 2.18c shows a distribution shift to the left (a decrease in travel time). The t -test does not show a statistically significant result, while the Mann-Whitney test shows a significant result at the 1% level, and the Kolmogorov-Smirnov test shows a significant result at the 5% level.
- Figure 2.18d indicates another transition to a bimodal distribution, except that it is less symmetric about the median. None of the tests concludes a statistically significant change. However, note that P_{KS} is much smaller than P_{TT} and P_{MW} .

In summary, the choice of statistical test is a nontrivial consideration. The most common test, Student’s t -test, is a reasonable choice when the data is normally distributed, which is more likely to occur on longer segments and on continuous-flow facilities. A nonparametric test such as the Mann-Whitney U -test or the Kolmogorov-Smirnov test is more appropriate when the data is *not* normally distributed, which is more likely the case on shorter segments, especially with signal control. The Kolmogorov-Smirnov test is capable of distinguishing significant differences in distributions even when one distribution is bimodal, as in Figure 2.18b.

While statistical significance can be helpful in establishing the effect of a treatment with some certainty, and is sometimes required, there are several pitfalls that the analyst should be careful to avoid (6). The failure of a data set to “pass” a statistical test (*i.e.*, when the null hypothesis cannot be rejected, or when the P -value is greater than the confidence level cutoff) does *not* mean that there was no effect at all. A lack of

⁴Credit for these formulas: Marc Brysbaert. See also: <http://www.palgrave.com/psychology/brysbaert/students/excelguidelines/run-Mann-Whitney-test-Excel.pdf> NOTE: these formulas are not accurate for small sample sizes (under 20 samples).

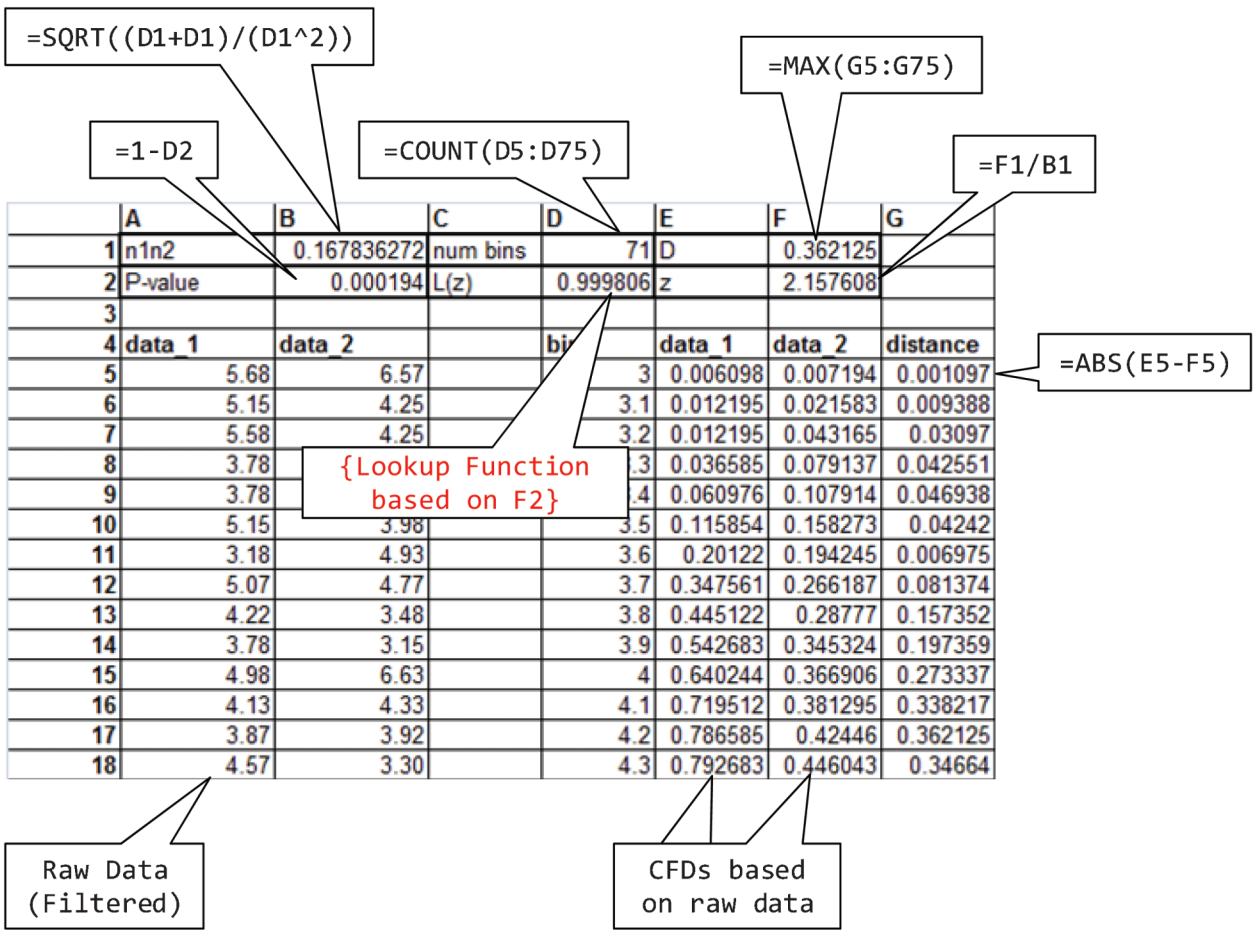
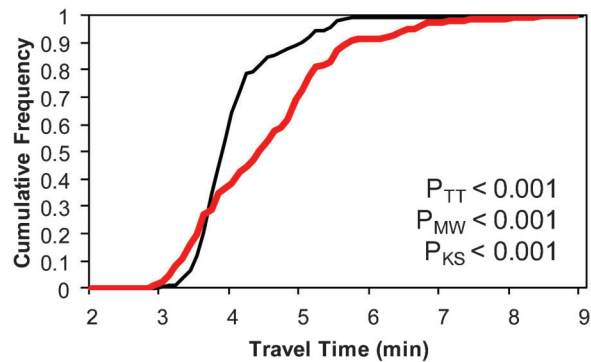


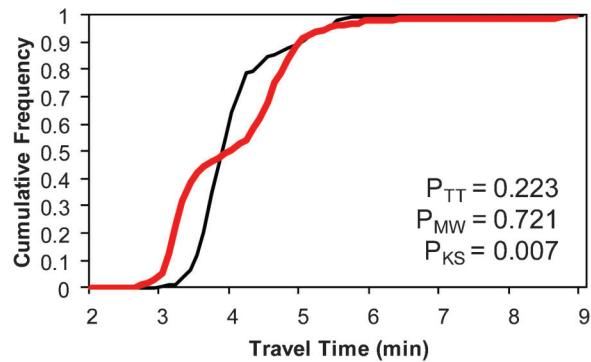
Figure 2.17 Conducting a Kolmogorov-Smirnov test in Excel.

TABLE 2.1
 Looking up the value of $L(z)$ based on z , for the Kolmogorov-Smirnov statistic (δ).

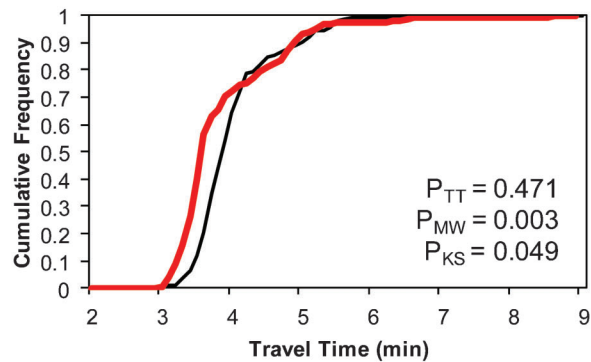
z	$L(z)$	z	$L(z)$	z	$L(z)$	z	$L(z)$
0.280	0.000 001 000	0.870	0.564 546 000	1.460	0.971 846 000	2.050	0.999 552 000
0.290	0.000 004 000	0.880	0.579 070 000	1.470	0.973 448 000	2.060	0.999 588 000
0.300	0.000 009 000	0.890	0.593 316 000	1.480	0.974 970 000	2.070	0.999 620 000
0.310	0.000 021 000	0.900	0.607 270 000	1.490	0.976 412 000	2.080	0.999 650 000
0.320	0.000 046 000	0.910	0.620 928 000	1.500	0.977 782 000	2.090	0.999 680 000
0.330	0.000 091 000	0.920	0.634 286 000	1.510	0.979 080 000	2.100	0.999 705 000
0.340	0.000 171 000	0.930	0.647 338 000	1.520	0.980 310 000	2.110	0.999 728 000
0.350	0.000 303 000	0.940	0.660 082 000	1.530	0.981 476 000	2.120	0.999 750 000
0.360	0.000 511 000	0.950	0.672 516 000	1.540	0.982 578 000	2.130	0.999 770 000
0.370	0.000 826 000	0.960	0.684 636 000	1.550	0.983 622 000	2.140	0.999 790 000
0.380	0.001 285 000	0.970	0.696 444 000	1.560	0.984 610 000	2.150	0.999 806 000
0.390	0.001 929 000	0.980	0.707 940 000	1.570	0.985 544 000	2.160	0.999 822 000
0.400	0.002 808 000	0.990	0.719 126 000	1.580	0.986 426 000	2.170	0.999 838 000
0.410	0.003 972 000	1.000	0.730 000 000	1.590	0.987 260 000	2.180	0.999 852 000
0.420	0.005 476 000	1.010	0.740 566 000	1.600	0.988 048 000	2.190	0.999 864 000
0.430	0.007 377 000	1.020	0.750 826 000	1.610	0.988 791 000	2.200	0.999 874 000
0.440	0.009 730 000	1.030	0.760 780 000	1.620	0.989 492 000	2.210	0.999 886 000
0.450	0.012 590 000	1.040	0.770 434 000	1.630	0.990 154 000	2.220	0.999 896 000
0.460	0.016 005 000	1.050	0.779 794 000	1.640	0.990 777 000	2.230	0.999 904 000
0.470	0.020 022 000	1.060	0.788 860 000	1.650	0.991 364 000	2.240	0.999 912 000
0.480	0.024 682 000	1.070	0.797 636 000	1.660	0.991 917 000	2.250	0.999 920 000
0.490	0.030 017 000	1.080	0.806 128 000	1.670	0.992 438 000	2.260	0.999 926 000
0.500	0.036 055 000	1.090	0.814 342 000	1.680	0.992 928 000	2.270	0.999 934 000
0.510	0.042 814 000	1.100	0.822 282 000	1.690	0.993 389 000	2.280	0.999 940 000
0.520	0.050 306 000	1.110	0.829 950 000	1.700	0.993 823 000	2.290	0.999 944 000
0.530	0.058 534 000	1.120	0.837 356 000	1.710	0.994 230 000	2.300	0.999 949 000
0.540	0.067 497 000	1.130	0.844 502 000	1.720	0.994 612 000	2.310	0.999 954 000
0.550	0.077 183 000	1.140	0.851 394 000	1.730	0.994 972 000	2.320	0.999 958 000
0.560	0.087 577 000	1.150	0.858 038 000	1.740	0.995 309 000	2.330	0.999 962 000
0.570	0.098 656 000	1.160	0.864 442 000	1.750	0.995 625 000	2.340	0.999 965 000
0.580	0.110 395 000	1.170	0.870 612 000	1.760	0.995 922 000	2.350	0.999 968 000
0.590	0.122 760 000	1.180	0.876 548 000	1.770	0.996 200 000	2.360	0.999 970 000
0.600	0.135 718 000	1.190	0.882 258 000	1.780	0.996 460 000	2.370	0.999 973 000
0.610	0.149 229 000	1.200	0.887 750 000	1.790	0.996 704 000	2.380	0.999 976 000
0.620	0.163 225 000	1.210	0.893 030 000	1.800	0.996 932 000	2.390	0.999 978 000
0.630	0.177 753 000	1.220	0.898 104 000	1.810	0.997 146 000	2.400	0.999 980 000
0.640	0.192 677 000	1.230	0.902 972 000	1.820	0.997 346 000	2.410	0.999 982 000
0.650	0.207 987 000	1.240	0.907 648 000	1.830	0.997 533 000	2.420	0.999 984 000
0.660	0.223 637 000	1.250	0.912 132 000	1.840	0.997 707 000	2.430	0.999 986 000
0.670	0.239 582 000	1.260	0.916 432 000	1.850	0.997 870 000	2.440	0.999 987 000
0.680	0.255 780 000	1.270	0.920 556 000	1.860	0.998 023 000	2.450	0.999 988 000
0.690	0.272 189 000	1.280	0.924 505 000	1.870	0.998 145 000	2.460	0.999 989 000
0.700	0.288 765 000	1.290	0.928 288 000	1.880	0.998 297 000	2.470	0.999 990 000
0.710	0.305 471 000	1.300	0.931 908 000	1.890	0.998 421 000	2.480	0.999 991 000
0.720	0.322 265 000	1.310	0.935 370 000	1.900	0.998 536 000	2.490	0.999 992 000
0.730	0.339 113 000	1.320	0.938 682 000	1.910	0.998 644 000	2.500	0.999 992 500
0.740	0.355 981 000	1.330	0.941 848 000	1.920	0.998 744 000	2.550	0.999 995 600
0.750	0.372 833 000	1.340	0.944 872 000	1.930	0.998 837 000	2.600	0.999 997 400
0.760	0.389 640 000	1.350	0.947 756 000	1.940	0.998 924 000	2.650	0.999 998 400
0.770	0.406 372 000	1.360	0.950 512 000	1.950	0.999 004 000	2.700	0.999 999 000
0.780	0.423 002 000	1.370	0.953 142 000	1.960	0.999 079 000	2.750	0.999 999 400
0.790	0.439 505 000	1.380	0.955 650 000	1.970	0.999 149 000	2.800	0.999 999 700
0.800	0.455 857 000	1.390	0.958 040 000	1.980	0.999 213 000	2.850	0.999 999 820
0.810	0.472 041 000	1.400	0.960 318 000	1.990	0.999 273 000	2.900	0.999 999 900
0.820	0.488 030 000	1.410	0.962 486 000	2.000	0.999 329 000	2.950	0.999 999 940
0.830	0.503 808 000	1.420	0.964 552 000	2.010	0.999 380 000	3.000	0.999 999 970
0.840	0.519 366 000	1.430	0.966 516 000	2.020	0.999 428 000		
0.850	0.534 682 000	1.440	0.968 382 000	2.030	0.999 474 000		
0.860	0.549 744 000	1.450	0.970 158 000	2.040	0.999 516 000		



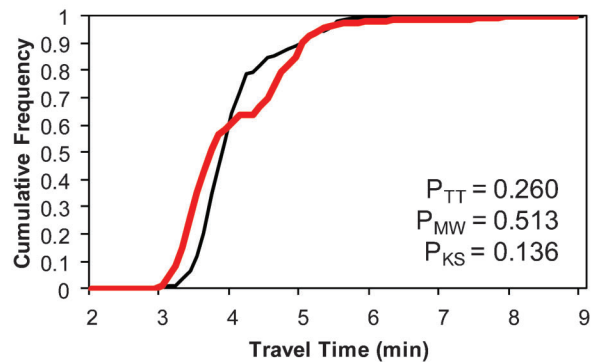
(a) D1 vs. D2.



(b) D1 vs. D3.



(c) D1 vs. D4.



(d) D1 vs. D5.

Figure 2.18 Pairwise comparisons with P -values resulting from statistical tests for Friday data.

TABLE 2.2. Statistical significance (at the 1% confidence level) based on statistical test for the four different comparisons of distributions in Figure 2.18.

Comparison	Student's t -test	Mann-Whitney U -test	Kolmogorov-Smirnov test
D1 vs. D2	✓	✓	✓
D1 vs. D3			✓
D1 vs. D4		✓	
D1 vs. D5			

statistical significance also does not mean that the impact was not substantial. It is also possible that not enough samples are available to be able to make the determination. Finally, the level of confidence severely determines the test outcome, but there is rarely any substantive reason to select a particular value (10%, 5%, 1%) other than to match the P -values that arise from the test. In closing, we would stress that being able to visually compare CDFs is perhaps more important when assessing the impact on travel time.

3. BEST PRACTICES FOR SENSOR PLACEMENT FOR INTERRUPTED FLOW

A major consideration in setting up sensors for vehicle identification is the locations where the sensors are to be installed. Depending on the type of detection technology selected, sensors for capturing vehicle identification characteristics typically require dedicated access to power and perhaps communications. This can be a challenge in many transportation systems, because there are often few locations where roadside infrastructure is available to provide access to these elements. This chapter discusses issues specific to placement of Bluetooth MAC address sensors on interrupted flow facilities.

The high traffic volumes and essential status of most freeways often provides a rationale for investment in the construction of permanent sensing stations. To economize, these stations can often be co-located with dynamic message signs (DMS) or other monitoring infrastructure such as speed detectors, where power and communications access is already available. These existing facilities are typically located away from interchanges (*e.g.*, it is undesirable to present motorists with information on a DMS when they are preparing to merge or exit), meaning that the data from freeway locations is typically high quality.

For surface streets, there are essentially two possible locations for detectors—at intersections and at mid-block locations. There are considerable advantages and disadvantages to each option:

1. At intersections, there are often existing power sources, such as for street lights and traffic signals. Communications infrastructure can sometimes be leveraged at signalized intersections. However, the quality of the data can suffer because of traffic interactions at the intersection, as will be demonstrated in the next section.

- At midblock locations, the data is higher quality because vehicles are interacting less with other vehicles and control devices. It can be more expensive to establish dedicated power and communications, although these costs can be reduced by supporting detectors with solar power and cellular communications.

3.1 Case Study Comparing Intersection and Midblock Sensor Locations

Pictures of alternative sensor deployment locations are shown in Figure 3.1. Battery-powered, temporary sensors can be secured to fixed roadside infrastructure such as road signs (Figure 3.1a, Figure 3.1b) and light poles (Figure 3.1c). Permanent sensors, requiring dedicated power and communications, can be established at intersections using traffic signal cabinets (Figure 3.1d).

Midblock permanent stations might be possible to set up by leveraging agency facilities positioned along the roadside (Figure 3.1e), or accessing other power sources (Figure 3.1f). In absence of such infrastructure, power and communications would need to be installed to establish a permanent station at a midblock location, or rely on solar recharging and/or wireless communications.

Midblock installations are preferable, but in the real world, there are often practical limitations on sensor deployment. It can be difficult, for example, to establish permanent midblock stations on most arterials. The tradeoff in selecting intersection or midblock location is illustrated by results from a field deployment. Figure 3.2 shows an aerial view of the test system, which is a 4-intersection segment of SR 37 in Noblesville, Indiana. Vehicle IDs were measured from



(a) Road sign.



(b) Road sign.



(c) Light pole.



(d) Traffic signal cabinet.



(e) Building interior.



(f) Other roadside electrical equipment.

Figure 3.1 Example vehicle sensor deployments (11).

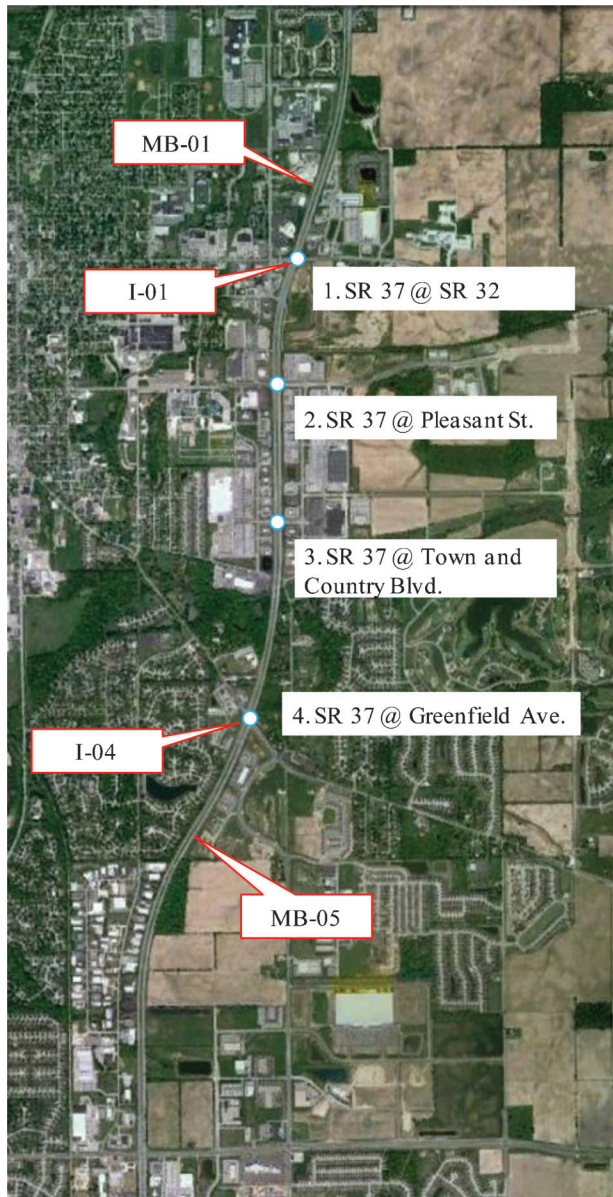


Figure 3.2 Sensor locations in 2009 field study.

two intersection locations (I-01 and I-04) and two “midblock” locations over 1000 ft upstream of those intersections (MB-01 and MB-04), well outside of the maximum queue lengths (9).

Figure 3.3 compares CFDs of the travel times measured from intersection to intersection with those measured from midblock to midblock, for both directions on the arterial on two different dates: Figure 3.3a shows northbound travel times on 6/20/2009; Figure 3.3b shows northbound travel times on 7/25/2009; Figure 3.3c shows southbound travel times on 6/20/2009; and Figure 3.3d shows southbound travel

times on 7/25/2009. In each case, the midblock travel time CFDs are between 1 and 2 minutes to the right of the intersection CFDs. This additional travel time at the midblock locations is attributable in part to the longer distance between MB-01 and MB-04 than between I-01 and I-04. However, the variation in the separation of the intersection and midblock CFDs shows that the difference is also attributable to the fact that intersection delays are not captured by the intersection CFDs. The travel times were measured using the first observation of the vehicle at each sensor location, as described in more detail by Wasson and Bullock (10).

In summary, midblock detector locations are preferable because they will more accurately capture increases in travel time due to queuing at intersections. However, establishing permanent midblock locations is challenging because power and communications is not usually available. Intersection locations may be used when a midblock sensor is infeasible. Intersection-located sensors will capture delay accumulated at any intersections in between the endpoints of the route, but will not capture delay occurring at the endpoints.

3.2 Sensor Antenna Height Impact on Detection Efficiency

Another consideration in sensor installation that is specific to Bluetooth MAC address matching technology is the vertical position of the Bluetooth transceiver. In a prior study (12), temporary sensors with different transceiver antenna heights (0 ft, 2.5 ft, 5 ft, 7.5 ft, and 10 ft) were tested in the field to determine which height would produce the largest sample size. The sensors were stationed at various positions along the southbound direction on I-65 along a segment with no intermediate entrances or exits. The results of the study are shown in Figure 3.4. This figure shows the total number of detected vehicle IDs (MAC addresses) for each antenna height, broken down by whether these vehicles were known to be traveling northbound or southbound (by matching the vehicle ID at another sensor), or whether the direction was unknown (*i.e.*, no match as found). Notably, the sensor that lacked an antenna (*i.e.*, the “0 ft” antenna) produced the least counts. The study concluded that a height of ~7.5 ft was best for roadside location because it detected a high number of vehicle IDs, but the far side lanes were not substantially underrepresented in the data.

In summary, when locating a vehicle sensor along an arterial, it is preferable to use a midblock location when possible, although an intersection location can be used for permanent stations if there are no feasible opportunities at any midblock station. Various possibilities for sensor location are illustrated in Figure 3.1. For Bluetooth MAC address matching, it is recommended to install the Bluetooth transceiver in

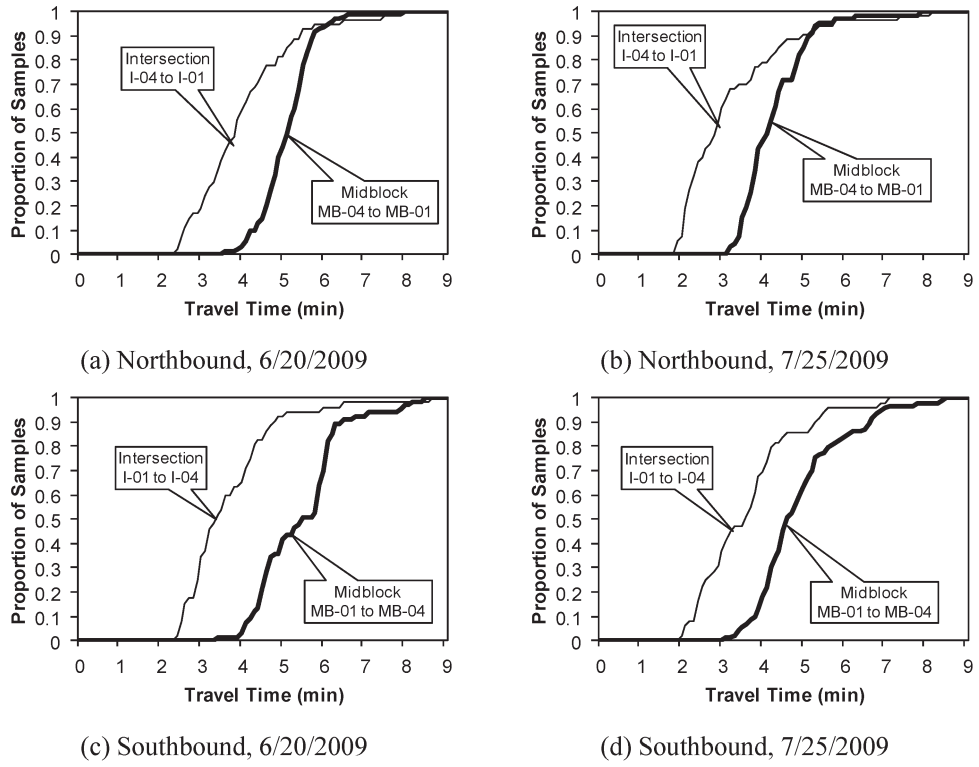


Figure 3.3 Saturday (0900–1300) travel time cumulative distributions (9).

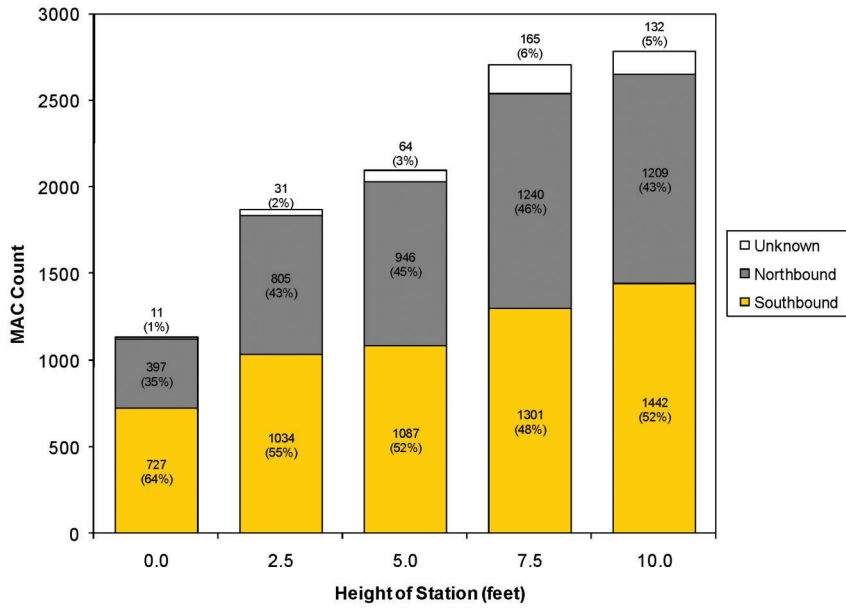


Figure 3.4 Traceable MAC address counts by antenna height and direction (12).

an antenna, and an antenna height of ~ 7.5 ft is recommended.

4. UNINTERRUPTED FLOW CASE STUDY

On facilities such as freeways where traffic flow is unimpeded by intersections, travel time is a largely a function of the traffic volume, incidents such as accidents or inclement weather, and sometimes geometry (*e.g.*, curves that necessitate speed reduction). In general, fluctuations in travel time related to the daily ebb and flow of volume are recurring, and geometry is of course rather permanent. It is important to measure recurring congestion to be able to assess the location and severity of bottlenecks, and make decisions about potential improvements. In a day-to-day operational context, it is also useful to understand the impact of incidents. This chapter presents a detailed case study⁵ investigating the use of AVI data to measure the impact of winter weather on a 62-mile freeway section.

4.1 Background

Performance measures for winter operations management during inclement weather has historically been limited to quantifying the amount of effort expended, such as the number of hours spent plowing or quantity of deicing chemicals applied. All of these performance measures are oriented toward the system input; their impact on the actual outcome (system performance) is based on the *assumed* effects of, for example, an hour of plowing time or 10 tons of magnesium chloride. Actual outcome assessment has typically been limited to qualitative evaluation by the maintenance foreman. In recent years, point speed detection at defined locations has been used in a few trial assessments. However, this data is typically unavailable in rural areas, while urban areas rarely have sufficient sensor density to provide consistent quantitative feedback for evaluating traffic mobility along specific snow and ice maintenance routes.

4.2 Data Collection

This case study demonstrates how AVI data can be used to assess the impact of winter events on roadway mobility over a segment. To collect AVI data to quantitatively evaluate the feasibility using travel time or space mean speed as a maintenance performance measure, eight portable data collection devices were used to record discoverable Bluetooth MAC Addresses in Indiana along I-65 north of Indianapolis, between mile marker (mm) 139 and mm 201. Figure 4.1 shows a map of where the sensors were located along the

interstate, with approximately 10-mile spacing between data collection devices. These stations nominally corresponded to the boundaries of snow and ice maintenance units. The devices were placed in the median of the roadway, with the goal of sampling northbound and southbound vehicles. Class 1 radios with high-gain omni-directional antennas were used for maximum detection of Bluetooth signals. Over 10 million MAC address readings were collected and synthesized during the data collection period between 12/20/2010, and 5/12/2011.

4.3 Winter Weather Impacts

Segment travel times during the first substantive storm event of 2011 are documented in Figure 4.2, with Figure 4.2a and Figure 4.2b respectively showing southbound and northbound travel times. The winter weather even occurred on 1/11/2011 and 1/12/2011. Travel times from 1/13/2011 are included to compare with typical operations.

Figure 4.3 shows a visual record of the storm event. From these we can develop a narrative of the storm.

- On 1/11/2011, there was substantial snow on the roadway (Figure 4.3a) and one can see the impact on travel time in Figure 4.2a, callout “i” and Figure 4.2b, callout “iv”.
- On the following day, 1/12/2011, the roads are less snow covered (Figure 4.3b). However, the travel time indicates similar if not worse delay (Figure 4.2a, callout “ii”; Figure 4.2b, callout “v”).
- On 1/13/2011, the roads are clear and dry (Figure 4.3c). Around 10:20 (Figure 4.2a, callout “iii”; Figure 4.2b, callout “vi”), the travel time reflects near free flow conditions.

There are no spikes in delay on 1/13/2011, which suggests that there are no episodes of recurring congestion that might have caused the spikes on 1/11/2011 or 1/12/2011. This example provides a good generalization of the intuitive relationship with inclement weather and its effects on traffic flow.

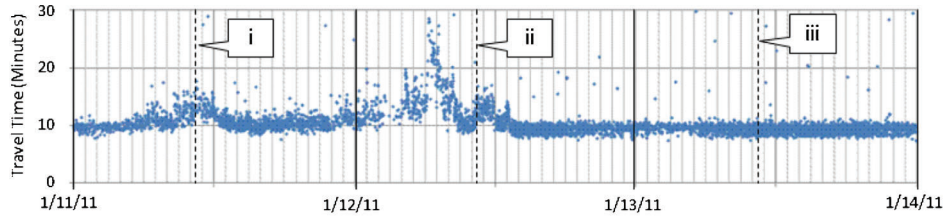
Travel time is dependent upon segment length. One way of normalizing for segment length is to measure segment performance in terms of space mean speed (SMS) for each vehicle, as discussed in Chapter 2.

A more severe storm that affected the study region occurred on 2/1/2011 and 2/2/2011. Figure 4.4 shows the SMS for the entire 62-mile segment from mm 139 to mm 201, along with corresponding data from a variety of other sources. Within the time window represented in these graphics, there was a sustained ice storm on 2/1/2011 and 2/2/2011 that dramatically reduced the number of vehicles on the road as well as their speeds. Other less dramatic events occurred on 2/4/2011 and 2/5/2011. Although Figure 4.4 provides good temporal reference regarding when and how long a storm occurred, it would be preferable to distill the massive amounts of data represented in these plots into summary statistics illustrating the impact of the storm with time.

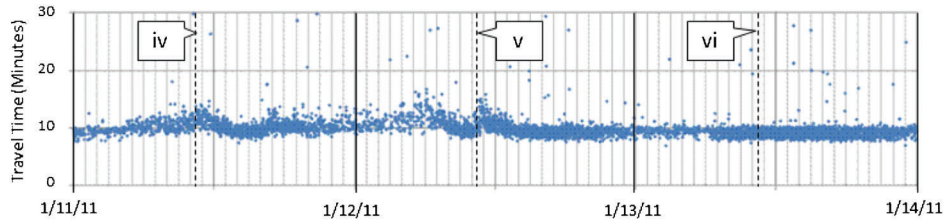
⁵This chapter is adapted from Hainen, A. M., S. M. Remias, T. M. Brennan, C. M. Day, and D. M. Bullock. Probe Vehicle Data for Characterizing Road Conditions Associated with Inclement Weather to Improve Road Maintenance Decisions. *Proceedings, 2012 IEEE Intelligent Vehicles Symposium*, 2012, pp. 730–735.



Figure 4.1 Location of portable Bluetooth data collection points on I-65 in Indiana.

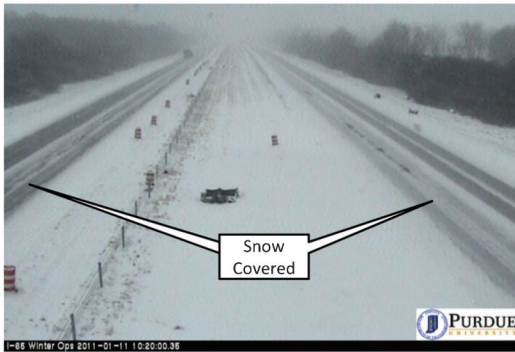


(a) Southbound travel time from mm 168 to mm 158 on I-65.

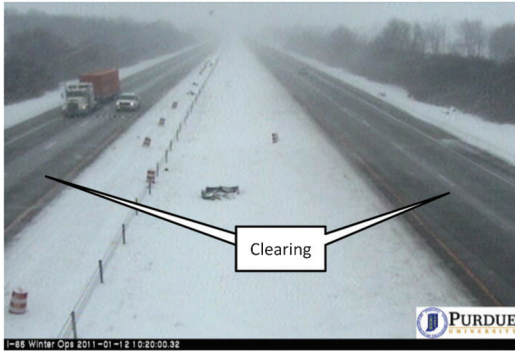


(b) Northbound travel time from mm 158 to mm 168 on I-65.

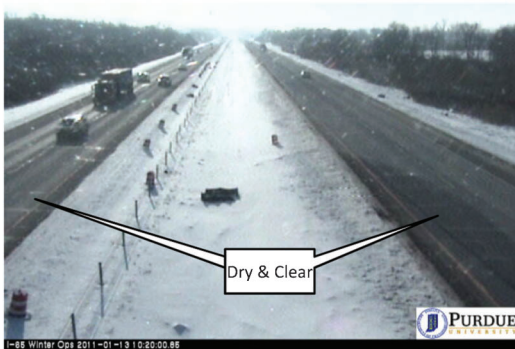
Figure 4.2 Scatter plot of AVI travel times.



(a) Visual conditions at 10:20AM EST on 1/11/2011.



(b) Visual conditions at 10:20AM EST on 1/12/2011.



(c) Visual conditions at 10:20AM EST on 1/13/2011.

Figure 4.3 Visual depiction for storm and clear conditions at 10:20 on each day.

Figure 4.5 shows the CFDs of northbound and southbound SMS for several segments (see Figure 4.1). The two CFDs in each plot are generated from two specific time periods:

- The storm data was collected from 1800 on 2/1/2011 through 1200 on 2/2/2011.
- The clear weather data was collected one week later, from 1800 on 2/7/2011 through 1200 on 2/8/2011.

The median speed during clear or free flow conditions is approximately 65 mph, while the speed during the storm is between 45 and 50 mph. The interquartile range (IQR), or difference between the 75th and 25th percentiles, shows the range of speeds of the middle

50% of the observations, and is an indicator of variability. It is visually correlated with the overall slope of the CFDs. In Figure 4.5, the “storm” CFD is usually a bit more horizontal than the “clear” CFD, indicating that there is a wider range of speeds during storm conditions.

The changes in the CFDs of SMS are similar across the eight segment-direction pairs in Figure 4.5. To efficiently compare both the northbound and southbound directions, during the storm and during the clear conditions, a plot of the median and IQR is presented in Figure 4.6. This plot is useful in two capacities:

The median can be compared directly with other segments, directions, or conditions to understand the magnitude in the difference of speed.

The IQR gives a description of the variation of the middle 50% of the sample. Figure 4.6 shows that SMS decreases by approximately 20 mph during a storm, and the IQR increases from about 8mph to about 12mph.

As mentioned before, the IQR is an indicator of the variability of speed, which is highly relevant to maintenance and operations. This information could potentially be integrated with road maintenance policies to improve strategies.

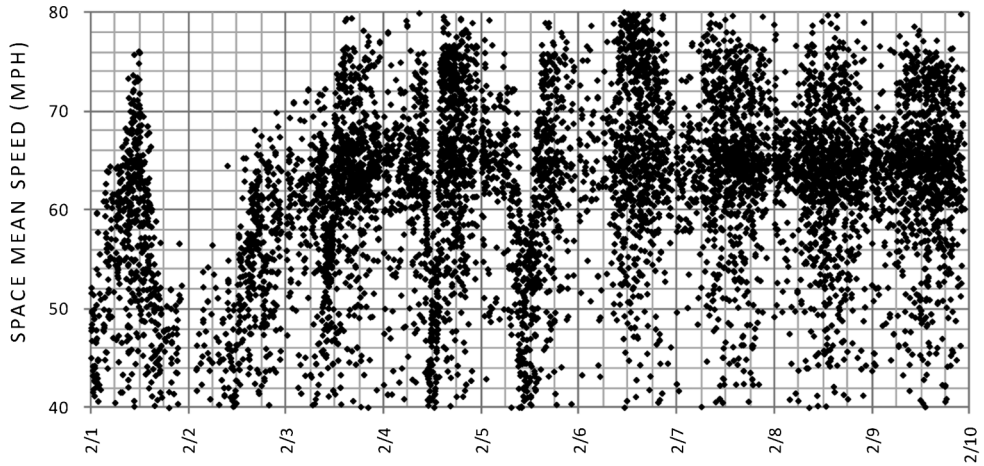
A brief weather event on 2/5/2011 caused a disruption to travel time in both directions. Figure 4.7 statistically quantifies the impact of this event for the northbound segment from mm 168 to mm 178. This event was shorter in duration than the weather events on 2/1/2011 and 2/5/2011, but the impact on both the median speed and IQR are particularly clear during the period between 800 and 1400.

The representation of the data in Figure 4.7 is intellectually sound in terms of characterizing both the central tendency and stochastic variation of the segment speeds; it is also visually intuitive, and can be used by maintenance personnel with a wide range of academic backgrounds to systematically assess the condition of the roads with more operational detail than qualitative descriptions such as “snow covered” or “wheel tracks.”

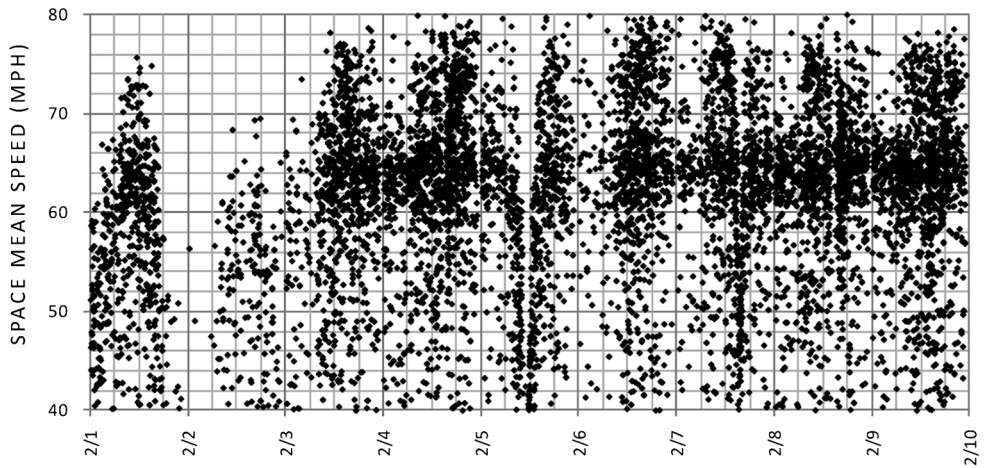
4.4 Data Fusion

AVI travel time data is even more powerful when fused with additional data sets. Figure 4.8 arranges a diverse collection of data sources in an organized graphical view. The SMS from AVI travel times (Figure 4.8a, Figure 4.8b) was combined with side-fire radar spot speeds measured midway through the corridor (Figure 4.8c), traffic counts obtained from the same location (Figure 4.8d), hourly precipitation (Figure 4.8e) componentized wind speed (Figure 4.8f), and temperature/pressure (Figure 4.8g). The seven charts share a common horizontal axis representing time, with divisions representing 6 hour intervals.

It is interesting to compare the SMS obtained from AVI data (Figure 4.8a, Figure 4.8b) with the point speeds from a single location (Figure 4.8c). Figure 4.8c



(a) Northbound space mean speed from mm 139 to mm 201.



(b) Southbound space mean speed from mm 139 to mm 201.

Figure 4.4 Space mean speed over entire study segment.

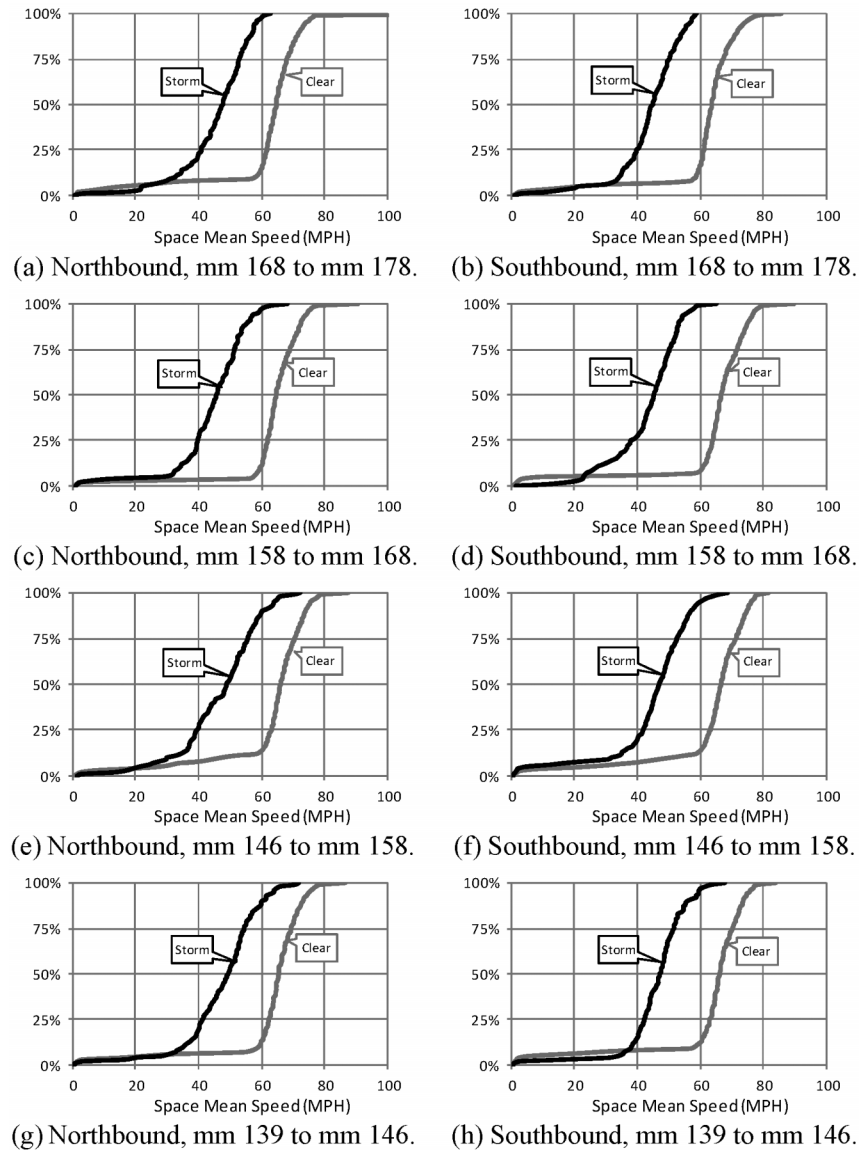


Figure 4.5 Southbound cumulative frequency diagrams of space mean speed. Storm weather data is from 2/1/2011 1800 through 2/2/2011 1200 and clear weather data is from 2/7/2011 1800 through 2/8/2011 1200.

shows an obvious decrease in point speed during the storm of 2/1/2011–2/2/2011 that corresponds to a similar drop in speed in the AVI data just above. However, other decreases in speed that are observable in the AVI data, such as in the northbound direction (Figure 4.8a) around the middle of the day on 2/4/2011, do not have corresponding decreases in point speed in Figure 4.8c. This is not unexpected, because point speeds only capture the performance at the exact location where the measurement occurs. If traffic is slowed at a certain location on the roadway, it cannot be observed in the spot speeds measured from another location.

Figure 4.8 shows an intuitive relationship between the SMS and additional integrated data including

weather conditions, point speed, and volume. The impact of the storm that occurred in the late evening of 2/1/2011 and early morning of 2/2/2011 is clearly defined by the following trends:

- SMS (Figure 4.8a, Figure 4.8b) and spot speeds (Figure 4.8c) fall substantially during the storm, with the degree and duration of the decrease varying by direction.
- Traffic volume is quite low during and following the storm (Figure 4.8d).
- An increase in precipitation is observed (Figure 4.8e), mainly due to snow.
- The wind direction changes from primarily northeast to southwest (Figure 4.8f).

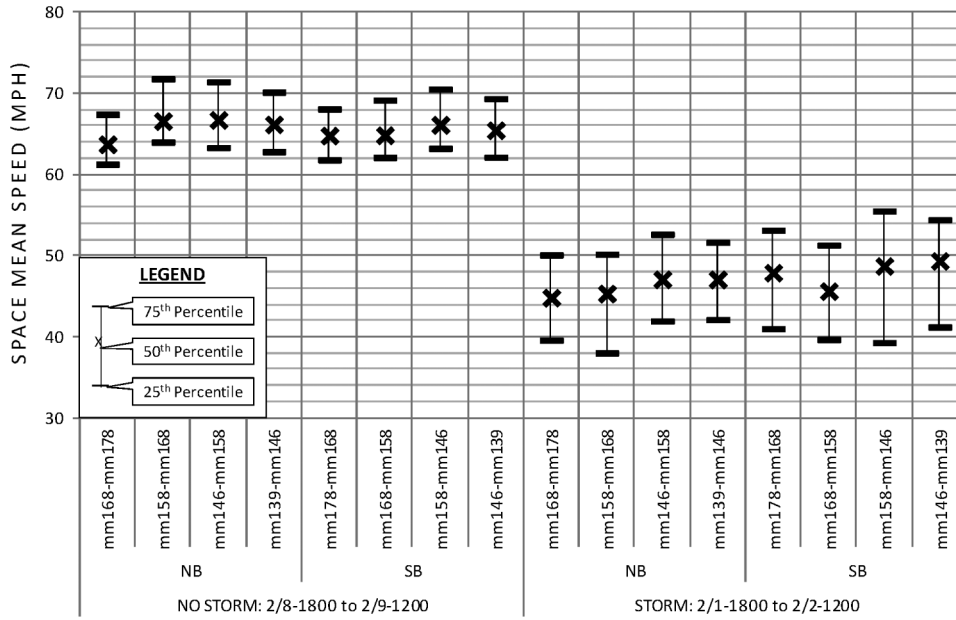


Figure 4.6 Comparison between each segment for the corresponding storm and clear periods.

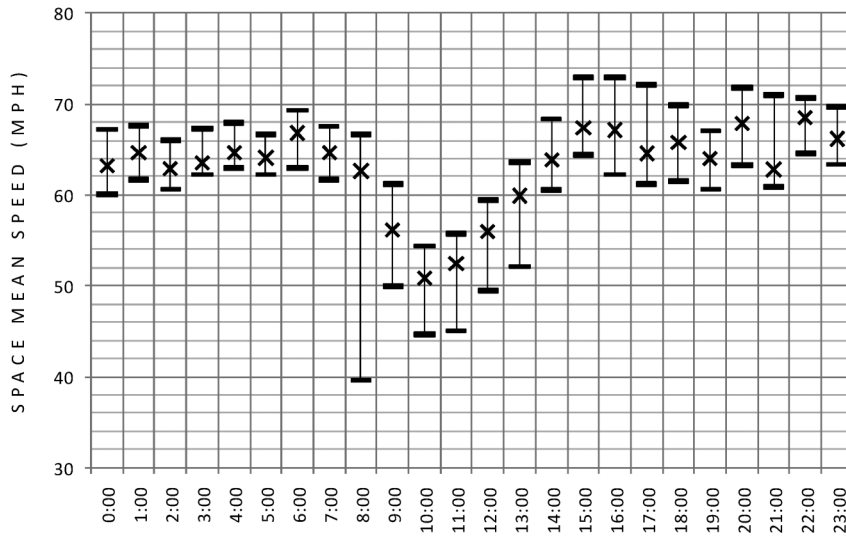


Figure 4.7 Southbound space mean speed from mm 168 to 178 on 2/5/2011.

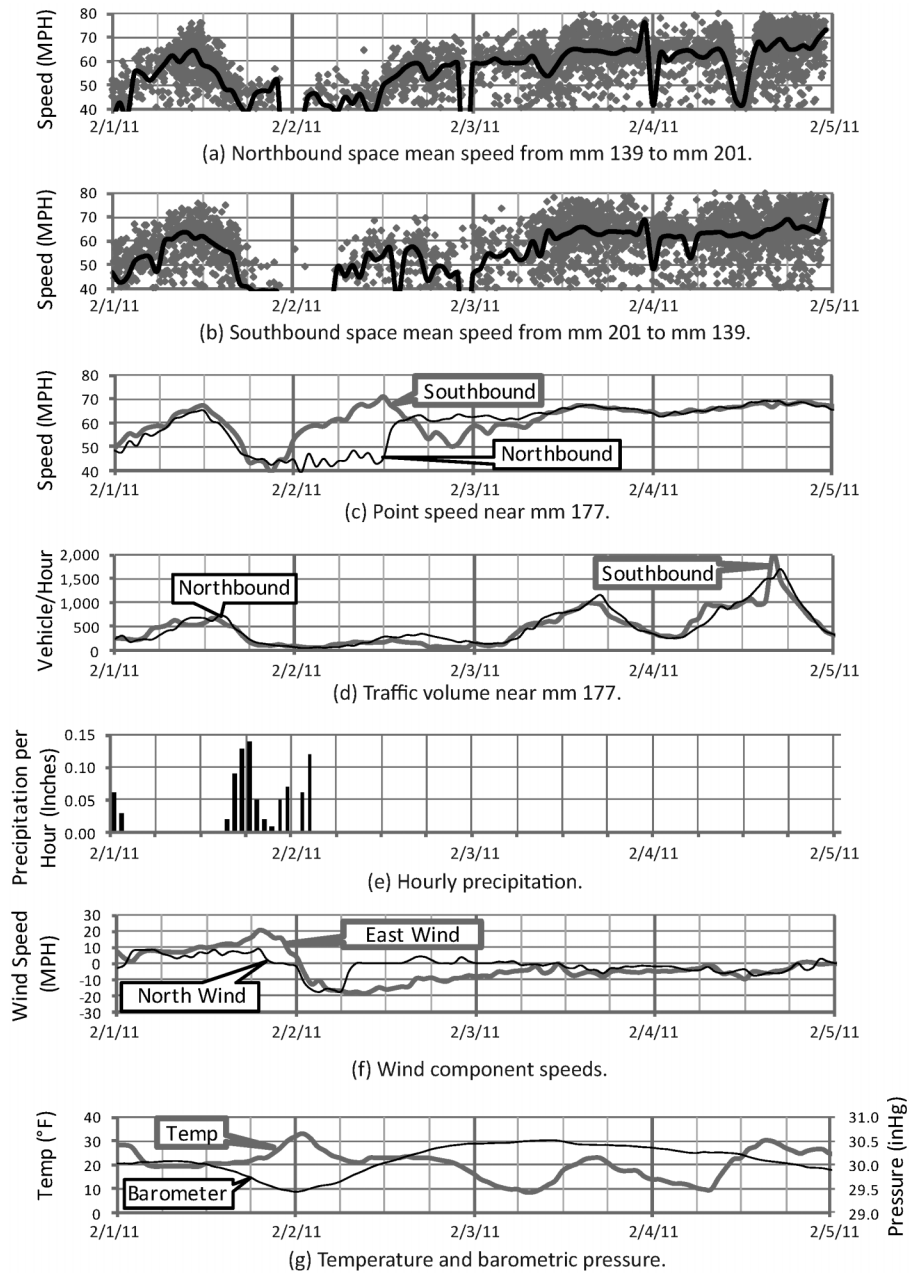


Figure 4.8 Fused data capturing storm event from 2/1/2011 to 2/2/2011.

- A brief spike in temperature, and drop in the barometric pressure can also be observed during the storm (Figure 4.8g).

4.5 Conclusion

Historically, the only quantitative feedback on plowing and deicing/salting operations has been the amount of time that the drivers spent on the road, de-icing chemical application rates, and other similar measures that do not assess the outcomes in system performance. AVI travel time data is an effective performance measure on travel conditions and appears to hold substantial

value for public agencies to develop quantitative mobility performance measures that can be integrated into their winter operations decision support system to provide both real-time and after action reviews of winter operations management and strategies.

5. INTERRUPTED FLOW CASE STUDIES

Interpreting measured travel times on interrupted-flow facilities, including signalized arterial corridors or other surface street networks, is challenging because of the presence of crossing movements, numerous entry and exit points of traffic, and the potential complexity of traffic control. This chapter provides an extended

discussion of traffic control parameters and their impact on measured travel times, and presents several case studies with field measured arterial travel times to show real-world examples of the impact, as well as to demonstrate how travel time data can be used to assess user cost savings in signal retiming and to address the feasibility of advanced control schemes.

5.1 Traffic Control Parameters

The following discussion in this chapter will rely on the definitions of the following terms in conventional signal control, which are illustrated in Figure 5.1:

- The *cycle length* is the amount of time in which the signal controller is scheduled to serve all movements for which there is demand. Sometimes, this is difficult to observe in the field because of phase actuation. The time between two successive ends of green, for example, will approximate the cycle length.
- An intersection's *offset* is the amount of time that one controller's "clock," or mechanism that it uses to determine where it is in the cycle, differs from a "system clock". There are several alternative reference points that can be used to define the "local zero" of a controller; Figure 5.1 shows the end of the coordinated phase green. The offset is an extremely important parameter to adjust in coordinated operations because it controls when the start of green occurs relative to neighboring intersections (thereby defining the "green bands").
- The *split* of a phase is the portion of the cycle that is assigned to that particular phase. During coordination, a phase will be terminated if it runs out of split time (which is called "forcing off"), or, if gap detection is used, when vehicles have cleared from the approach.

Two aspects of operation can be defined that represent competing objectives of traffic control in a series of intersections, such as in an arterial system:

1. **Efficient capacity allocation at the local intersection.** It is most desirable to serve each vehicle as it arrives at an intersection as quickly as possible.

2. **Coordination between green intervals at sequential intersections along a route.** In general, and usually in arterial operations, we want vehicles to progress along the major routes in a system; that is, it is desirable to avoid forcing them to stop numerous times along the route. This is done by linking neighboring signals.

Although certain controller models and operating modes replace the above concepts with terminology other than cycle, offsets, and splits, these basic outcome objectives are relevant to all signal control.

5.2 Management Problem

Figure 5.2 shows an overview of the signal control design process and is used to place this document in the context of the traffic signal design process. In Step I, based on agency objectives, resources are invested in retiming and other activities. In Step II, the necessary data to carry out this action are assembled. Usually this means the collection of manual turning movement counts (although automatic collection of turning movement counts is also possible). The data is typically used to obtain a timing plan, through Steps III–IV, and an iterative process (feedback loop 1) in which software model results are used to fine-tune the plan before it is field deployed. In Step V, the control parameters are assigned to the field equipment. Most operators will do some sort of performance measurement (Step VI), varying in sophistication from limited field observation of traffic to mitigate any obvious problems, to automatic data collection and adaptive control. This is represented as a second iterative process (feedback loop 2) in which the control parameters are tuned, based on observed performance.

Any analysis of travel time on a signalized facility should consider the signal timing that is used on the facility, because it directly influences the delay encountered at the intersections. In fact, signal timing decisions often have a more substantial impact on travel time than the traffic volume, particularly at moderate

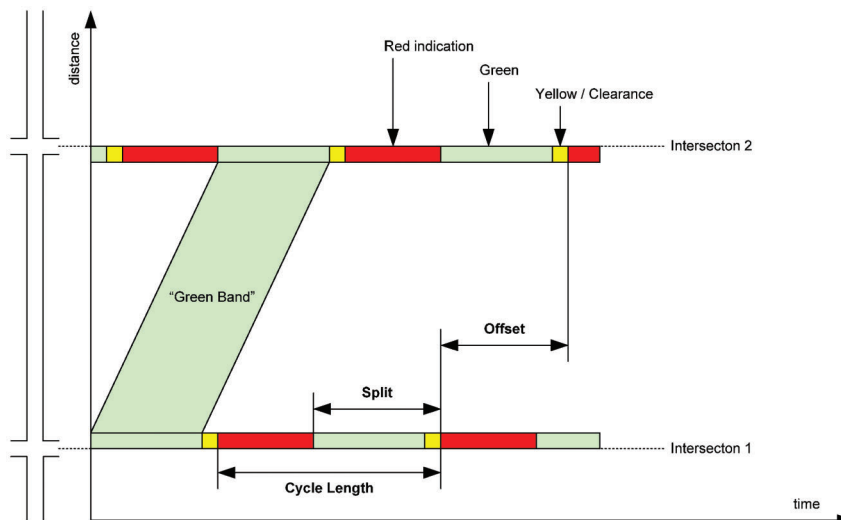


Figure 5.1 Definition of cycle length, offset, and split.

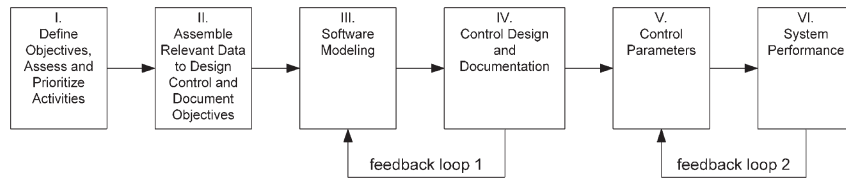


Figure 5.2 An overview of the traffic signal control design process, showing two iterative feedback loops.

volume levels. In Chapter 1, an overview of the signal control design process was presented (Figure 5.2); to support that process, a variety of other operations and maintenance activities are necessary. It is beyond the scope of this document to delve into those procedures in great detail, but one aspect that sheds some light on the challenge of signal operations is in the management of the controller databases that contain all of the necessary information needed to establish intersection control.

To provide an order of magnitude feel for the “design space” when one discusses signal timing, let’s consider the number of controller parameters that require configuration on a corridor level. Figure 5.3 shows a map of US 30, a 22-intersection corridor in north-western Indiana. Intersections 1—5 are operated in isolated “free” (non-coordinated) mode, while the rest form a coordinated system. The controller data contained in this system is represented in Figure 5.4. Each of the 22 controllers has its own database, storing 72,000 programmable parameters (13). The entire corridor contains 1.6 million parameters in total. By comparing these against a default database, it is possible to determine the number of parameters that are typically changed when a controller is deployed in the field. Figure 5.5 shows the results of this comparison. From inspection of Figure 5.5, we see that individual controllers required between 1400 and 2800 parameters to be programmed to be set up for field operations, or between 30,000 and 60,000 for a 22-intersection corridor.

While a “timing plan” as defined in modeling software such as Synchro is fully described by a rather short list of parameters (chiefly, cycle, offset, splits; also, phase sequence, gap times, min green times, etc.), there are many more parameters required to establish a family of timing plans in the field. This includes day-of-week and time-of-day schedules, detector configuration

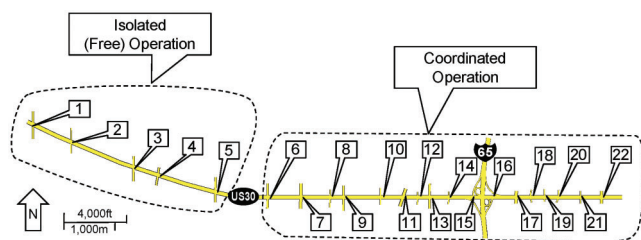


Figure 5.3 Location of intersections on US 30 (13).

(which can be especially time-consuming), communications configuration, preempts, and numerous other controller options. Altogether this means that there are thousands of opportunities for mistakes to be made in each controller, some of which can have detrimental consequences for operations.

Another illustration of the complexity of signal operation is apparent if we consider the boundaries between two different systems that share a link. This situation is not uncommon in practice. Often, two different agencies will have jurisdiction over different parts of a roadway. This is especially typical when a state-managed arterial intersects with a locally managed road with its own signalized intersections close by. Within one agency, sometimes two originally small systems grow into one, as land development occurs along the corridor and new signals are added.

Figure 5.6 shows a calendar view of two time of day (TOD) plans for two adjacent signal systems that operate on different, neighboring sections of US 30. In System 1, Day Plan “A” uses 6 patterns in 9 TOD intervals, while in System 2, Day Plan “B” uses 4 patterns in 6 intervals. The overlap between the two TOD plans is shown in Figure 5.7. Because of the changing patterns in the neighboring systems, the link that is shared between the two systems will have 13 different traffic patterns throughout the day. It would be particularly difficult to draw conclusions about travel times measured along that particular link, especially during a time period in which patterns frequently change, such as around 1100–1400. This illustrates that it is important to be cognizant of system boundaries when establishing travel time data collection zones.

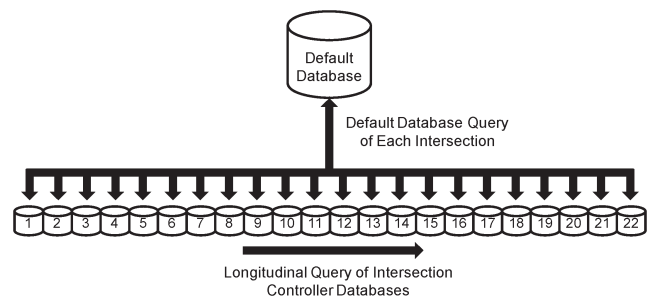


Figure 5.4 Controller database structure for US 30 (13). With 72,000 parameters existing per controller, this corridor encompasses approximately 1.6 million parameters in total.

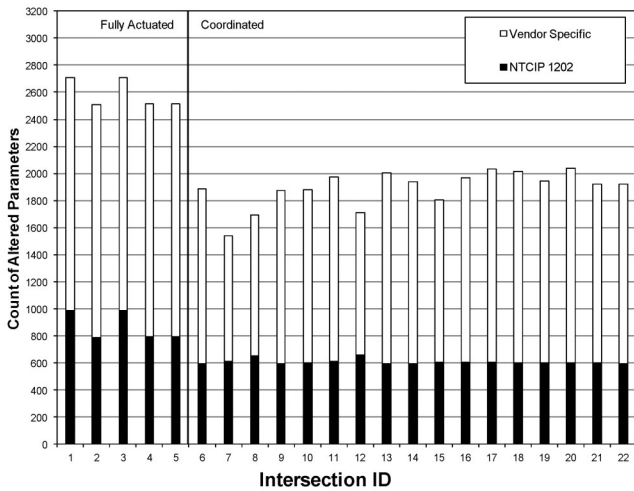


Figure 5.5 Analysis of controller database parameters for 22 intersections: Count of parameters by intersection that were changed from the default controller database (13).

5.3 High Resolution Controller Event Data

One set of data that has emerged as a rich source of information on signal operations is the high-resolution time-stamped record of controller events. In

recent years, several controller manufacturers have begun to develop internal logging utilities for writing this type of data, making it available for analysis. A variety of performance measures have been developed from this data covering a variety of use cases (9,14–21).

These events include:

- Detector on/off times, allowing occupancy and pulse counts to be recorded.
- Phase state changes (green, yellow, red clearance, and “off”).
- Coordinator events, including entry and exit from transition, yield points, and controller local zero times.
- Preemption events, including entry and exit from preempt state.

Table 5.1 provides an example stream of controller data, showing a typical chain of controller events as they are logged in real time. The majority of these events are detector on/off events, with occasional phase events occurring. By pulling this data into a relational database and querying on selected events (22), it is possible to extract a variety of performance measures.

Figure 5.8 shows an overview of an arterial management strategy incorporating these data sets, corresponding to feedback loop 2 in Figure 5.2. After optimizing the signal control parameters, data collection activities

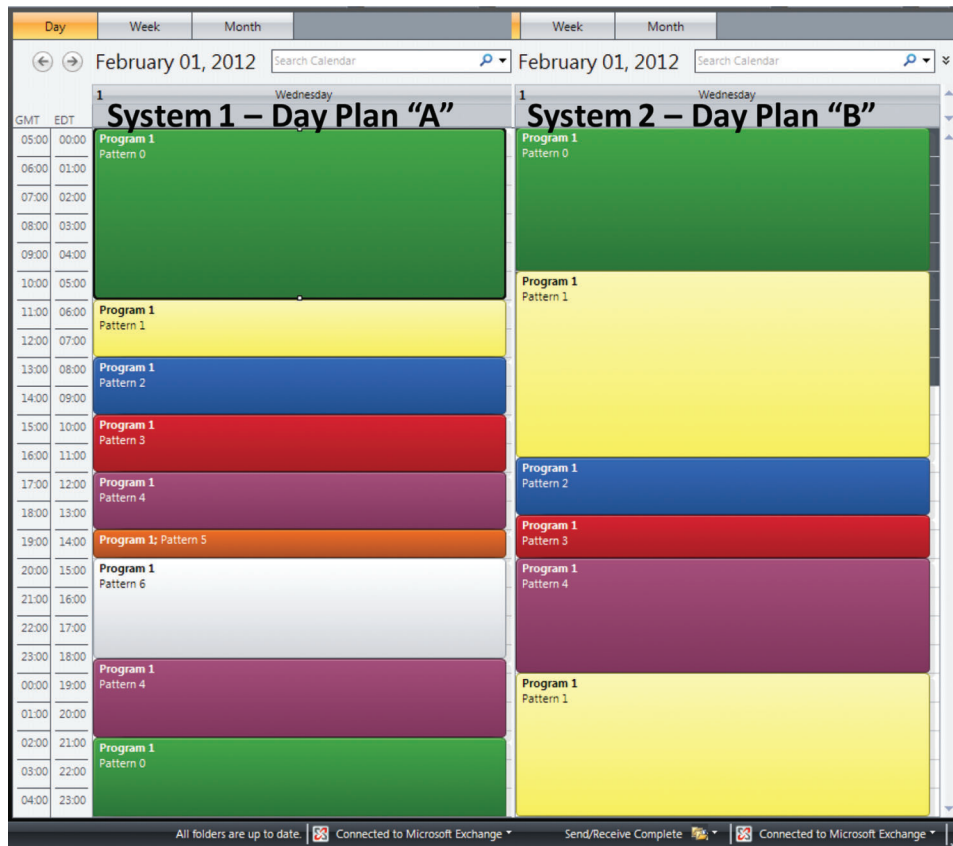


Figure 5.6 Timeline showing the schedules of two signal programs used on the same day for two adjacent subsystems on US 30.

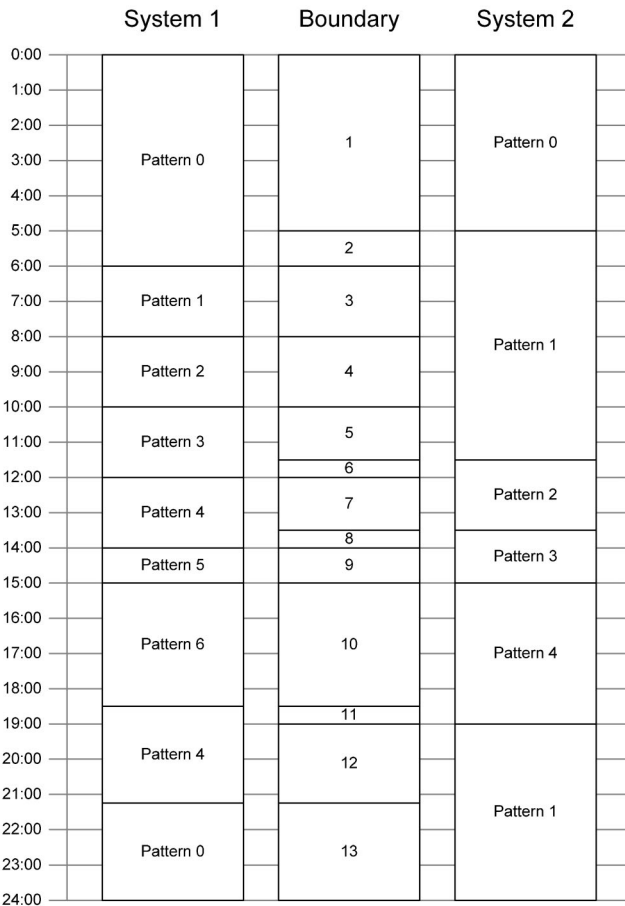


Figure 5.7 Timeline showing the schedules of two signal programs used on the same day for two adjacent subsystems on US 30.

based on various data sets would be regularly or continually monitored to detect operational deficiencies. When a deficiency is found, this would trigger tuning or retiming or activities.

With regards to arterial operations, we are most interested in the quality of progression along a corridor. While delay and travel time directly measure the outcomes of signal control, they do not by themselves give very detailed information about whether vehicle progression was satisfactory. However, by comparing vehicle arrival times with the status of the coordinated phase green, it is possible to determine whether vehicles arrive on green or arrive on red. The higher the proportion of vehicles arriving on green, the better the quality of progression.

One way to visually represent progression data (vehicle arrivals relative to the green state) is to construct a “cyclic flow profile.” Figure 5.9a shows an example of such a profile. The horizontal axis represents the portion of the cycle length, and the two distributions show the number of vehicles crossing the advance detector during that portion of the cycle (adjusted for travel time to the intersection), and the probability that the coordinated phase is green at that time in the cycle. Ideally, we would endeavor to

schedule vehicles to arrive in the portion of the cycle when the probability of green is highest. A number of cycles of data are needed to develop a flow profile.

While the flow profile is useful for showing the quality of progression in an aggregated manner, it is often helpful to examine what happens on a cycle-by-cycle basis, to determine whether the patterns are consistent across a time period. Recently, a visualization tool called the “Purdue Coordination Diagram” (PCD) was developed that allows all the individual coordination events to be viewed in a disaggregate manner (9,16,17,23). Figure 5.9b shows a PCD for 1 hour of coordination data. All of the events in the plot are positioned according to the time of day in which they occur (x-axis) and the time in cycle (or more specifically, the time after the last end of green, y-axis). The green and red lines respectively show the beginning and end of green during each cycle, and each point represents a vehicle crossing the advance detector (as before, adjusted for travel time to the intersection). Platoons of vehicles can be visually identified by strings of detections, as shown in Figure 5.9b.

The PCD also offers the advantage of enabling 24 hours of signal operations to be viewed in one graphic. Figure 5.10 pairs a PCD (Figure 5.10b) with the cycle-by-cycle percent on green (Figure 5.10a) directly corresponding to each cycle in the PCD. Several observations can be made from this figure:

- During the overnight “free” or fully actuated, non-coordinated operations, the intersection tends to dwell in the coordinated green for long periods of time (Figure 5.10b, callout “i”).
- During most of the coordinated portion of the day (0600–2400), platoons of vehicles appear to arrive mostly during green (e.g., Figure 5.10b, callout “ii”), although there are some times of day when some platoons are cut off by the end of green, spilling into the next cycle (e.g., Figure 5.10b, callout “iii”). These times of day correspond to reductions in the percent on green in Figure 5.10a.
- Toward the end of the day, reduced minor phase extension causes the controller to return to the coordinated phase early more often (Figure 5.10b, callout “iv”), while platoons become thinner (Figure 5.10b, callout “v”). The underutilized green bands suggest that a shorter cycle length might better serve this time period.

Figure 5.11a shows the percent on green, and Figure 5.11b shows a PCD for a signal on one of the links on the boundaries of the two subsystems described in Figure 5.7, where either local or upstream intersection patterns change 13 times during the day. These are reflected in the strongly varying percent on green and vehicle arrival patterns. The vertical lines show when a pattern changes on one of the two systems.

5.4 Case Study: Synthesis of Controller Event Data and Travel Time Data

Now let us consider these high resolution event-based arrival distributions relative to concurrent travel

TABLE 5.1
Example high-resolution controller event data (16).

Timestamp	Event Code ID	Parameter	Explanation
04/08/09 14:10:49.6	8	22	Detector 22 off
04/08/09 14:10:49.9	9	7	Detector 7 on
04/08/09 14:10:50.1	8	7	Detector 7 off
04/08/09 14:10:51.1	63	2	Phase 2 yield point
04/08/09 14:10:51.3	2	2	Phase 2 yellow state
04/08/09 14:10:51.3	33	2	Phase 2 termination: gap out
04/08/09 14:10:51.8	9	10	Detector 10 on
04/08/09 14:10:51.9	9	28	Detector 28 on
04/08/09 14:10:52.0	8	10	Detector 10 off
04/08/09 14:10:52.0	9	12	Detector 12 on
04/08/09 14:10:52.0	8	28	Detector 28 off
04/08/09 14:10:52.1	8	12	Detector 12 off
04/08/09 14:10:52.4	9	17	Detector 17 on
04/08/09 14:10:52.8	9	19	Detector 19 on
04/08/09 14:10:53.0	8	19	Detector 19 off
04/08/09 14:10:56.1	9	21	Detector 21 on
04/08/09 14:10:56.4	3	2	Phase 2 red clearance state
04/08/09 14:10:57.0	9	1	Detector 1 on
04/08/09 14:10:57.2	9	3	Detector 3 on
04/08/09 14:10:57.3	8	1	Detector 1 off
04/08/09 14:10:57.4	1	1	Phase 1 green state
04/08/09 14:10:57.4	0	2	Phase 2 off state
04/08/09 14:10:57.4	8	3	Detector 3 off
04/08/09 14:10:57.8	9	10	Detector 10 on
04/08/09 14:10:58.0	8	10	Detector 10 off
04/08/09 14:10:58.0	9	12	Detector 12 on
04/08/09 14:10:58.1	8	12	Detector 12 off
04/08/09 14:10:58.5	9	4	Detector 4 on
04/08/09 14:10:58.6	8	4	Detector 4 off
04/08/09 14:10:58.6	32	6	Phase 6 min green complete

time measurements. Figure 5.12 shows a map of three intersections along a section of US 30. In February 2009, travel time measurements were carried out across this 3-intersection section by deploying two Bluetooth cases as shown in the map (“BT Case”). The resulting eastbound and westbound travel times are respectively shown by Figure 5.13a and Figure 5.13b. The travel time distributions in both directions show strong bimodal characteristics, similar to what we observed earlier in simulation for platoons being cut off by the end of green.

- In the eastbound direction, there are two distinct groups of vehicles; we can speculate that one group represents those who were able to progress through the three intersections without stopping (Figure 5.13a, callout “i”) and the other group represents vehicles experiencing perhaps one stop (Figure 5.13a, callout “ii”).
- In the westbound direction, similar patterns are visible. Again, it is likely that the lower group (Figure 5.13b, callout “iii”) represent vehicles that are not stopped, and they have approximately the same travel time as the corresponding vehicles in the other direction. The other

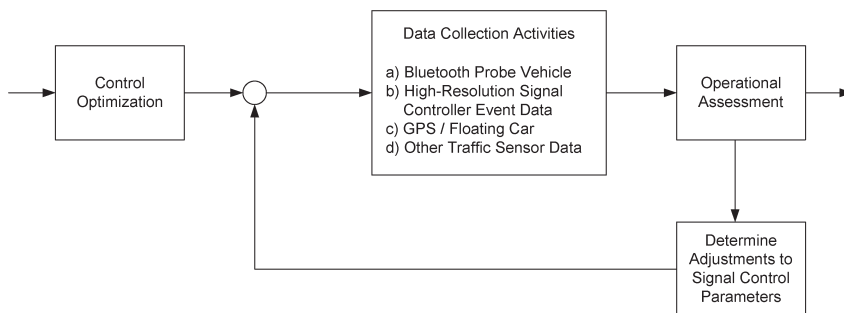
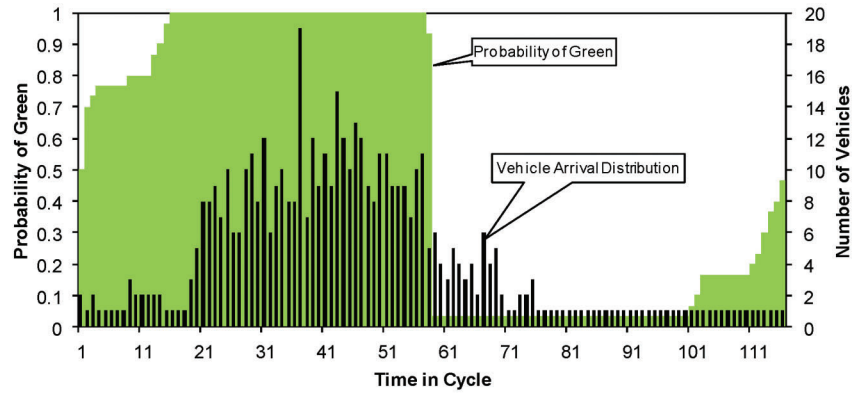
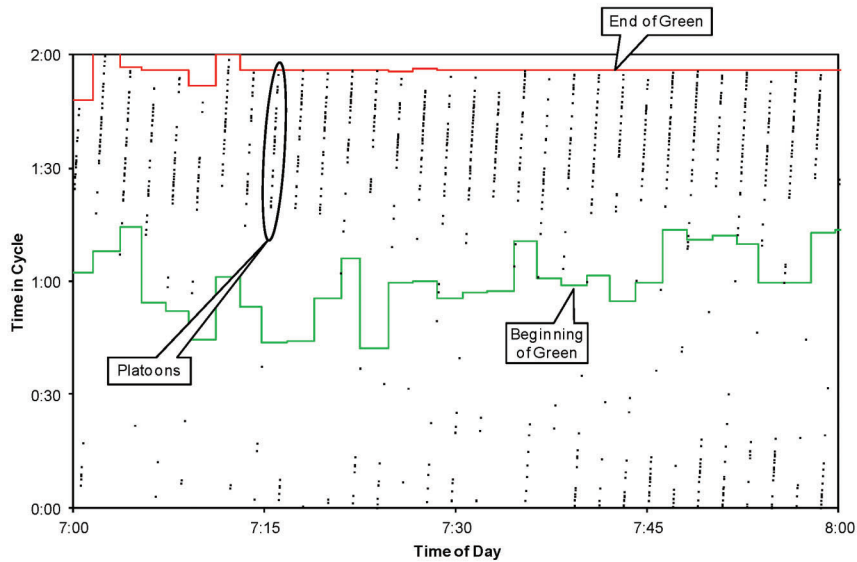


Figure 5.8 Fusion data approach to signalized arterial analysis.



(a) Flow profile and probability of green distribution.



(b) Coordination diagram for 1 hour.

Figure 5.9 Explanation of the Purdue Coordination Diagram (northbound at SR 37 and 141st St., 1/18/2012, 0700–0800).

group (Figure 5.13b, callout “iv”) probably represents vehicles that stopped at least once.

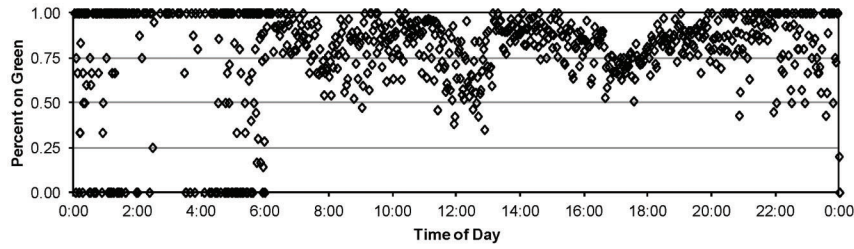
To see whether these travel time observations correlate well with the arrival patterns, we can examine the PCDs at each of the three intersections in some detail. Figure 5.14 contains six PCDs that correspond to the three eastbound and three westbound approaches on this section of US 30.

- In the eastbound direction at Int. 1 (Figure 5.14a), many of the coordinated vehicles arrive during green, but a substantial number of these platoons are cut off by the end of green (as shown by the heavy clustering of vehicles at the bottom of the PCD), and must wait until the next green to move forward. The coordinated platoons at Int. 2 (Figure 5.14c) and Int. 3 (Figure 5.14e) arrive almost entirely within the beginning and end of green.
- Westbound vehicles are similarly unimpeded at Int. 3 (Figure 5.14f) and Int. 2 (Figure 5.14d), judging by the arrival patterns that show most of the coordinated

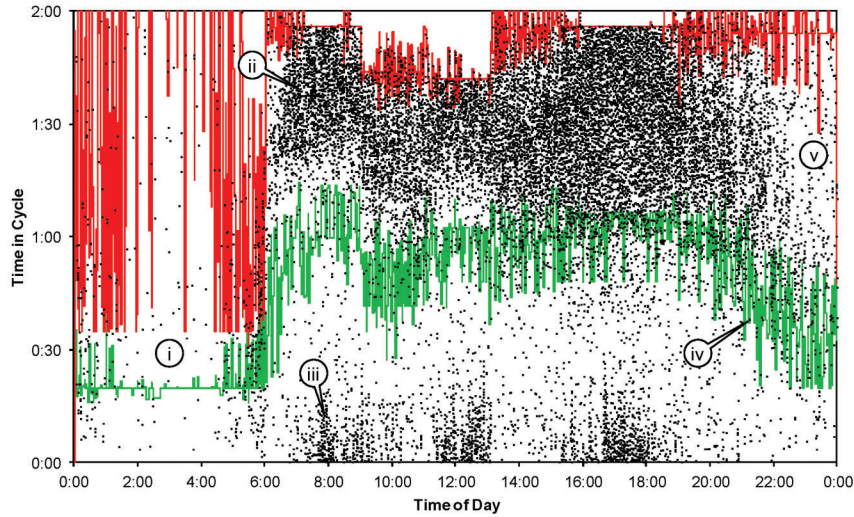
platoons arriving between the start and end of green. Many of these vehicles are also cut off at Int. 1 (Figure 5.14b), similar to the eastbound traffic.

In this case, the travel time patterns in Figure 5.13 are easily explained in terms of the arrival patterns in Figure 5.14, similar to the examples shown in Chapter 2. In general, when travel times are expanded over longer corridors including more intersections, regular patterns can become less distinct depending on how many times platoons of vehicles moving along the path of interest encounter red signals.

In this chapter, we provided a look at challenges in arterial operations from the perspective of data management and system boundaries, and introduced a way to examine the quality of progression along the arterial by using PCDs derived from high-resolution controller event data. An example of how travel time relates to vehicle progression was presented in which travel time measurements from the field were compared to concurrent signal event data. This example was very



(a) Percent on green.



(b) Coordination diagram.

Figure 5.10 Example of correlation between percent on green and arrival patterns that can be observed in the coordination diagram (northbound at SR 37 and 141st St., 1/18/2012).

similar to earlier observations from simulation where bimodal travel time distributions were observed when the stream of coordinated vehicles was cut off early by the end of green.

5.5 Case Study: Alternatives for Assessing Arterial Travel Times

In this case study,⁶ we examine the use of AVI travel times in assessing arterial performance from a variety of sensor location perspectives.

While agencies will often want to use corridor travel time as a system metric, especially in the cases of a coordinated system, identifying the relative delay at each signal is also important. SR 37 south of Indianapolis (Figure 5.15) is a 10-mile commuter corridor linking Indianapolis, Indiana and Bloomington, Indiana. This map indicates the locations of sensor locations (labeled as Bluetooth Monitoring Station, or “BMS”) along the

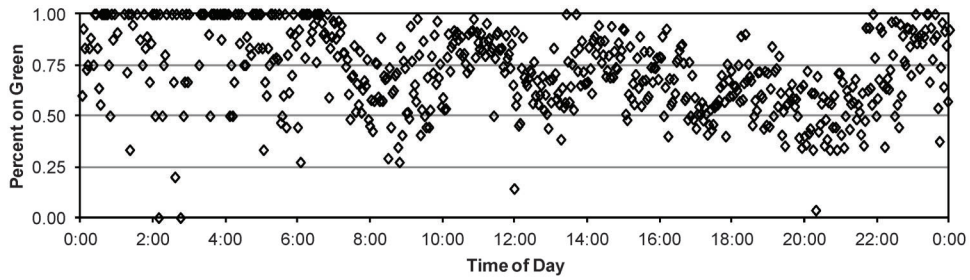
roadway as well as three alternative methods of matching travel times between pairs of sensors to characterize travel time. It is important to note that the same data set from any given BMS is used for all three methods, but each method tells a different story.

5.5.1 Origin-Based Travel Time

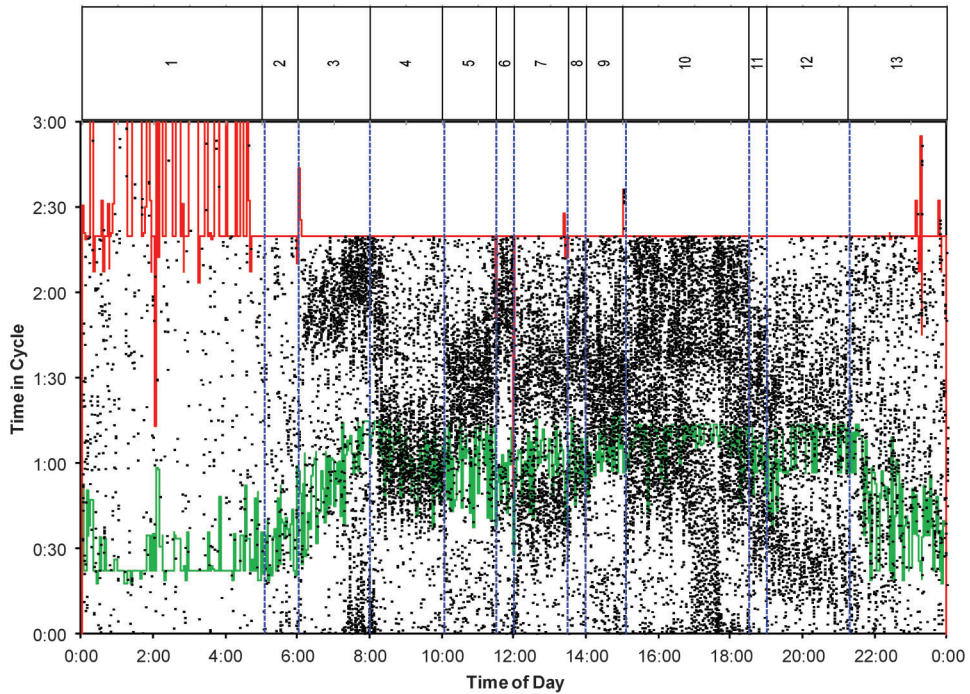
The first data reduction technique is called “origin-based travel time” (Figure 5.15a, Figure 5.16a). In this method, we consider all of the travel times between the starting point and each of the sensing locations to the south in a sequence of BMS pairs. During the PM peak period from 1600–1900, commuters from Indianapolis leave the city and the SB traffic is quite heavy (AADT exceeds 30,000 vehicles per day). The northern most BMS labeled “BMS-1” in Figure 3a is where each origin-based travel time segment originates. A family of sequentially larger travel time segments is formed with BMS-1 to BMS-2, BMS-1 to BMS-3, BMS-1 to BMS-4, and so on.

Rather than assuming a normal distribution and calculating an average and standard deviation, the distribution for each travel time segment is plotted in a cumulative distribution of the field-collected sample as shown in Figure 5.16a. In a more traditional study, typically only the average travel time for the entire

⁶The material in this section is adapted from Remias, S. M., A. M. Hainen, C. M. Day, T. M. Brennan, H. Li, E. Rivera-Hernandez, J. R. Sturdevant, S. E. Young, and D. M. Bullock. Characterizing Performance of Arterial Traffic Flow Using Probe Vehicle Data. Submitted to *Transportation Research Record, Journal of the Transportation Research Board*, Paper No. 13-1812, Transportation Research Board of the National Academies, Washington, D.C., 2012.



(a) Percent on green.



(b) Coordination diagram.

Figure 5.11 Influence of frequent pattern changes on interface between two systems under the same cycle length (eastbound at US 30 and Austin Rd., 2/9/2012).

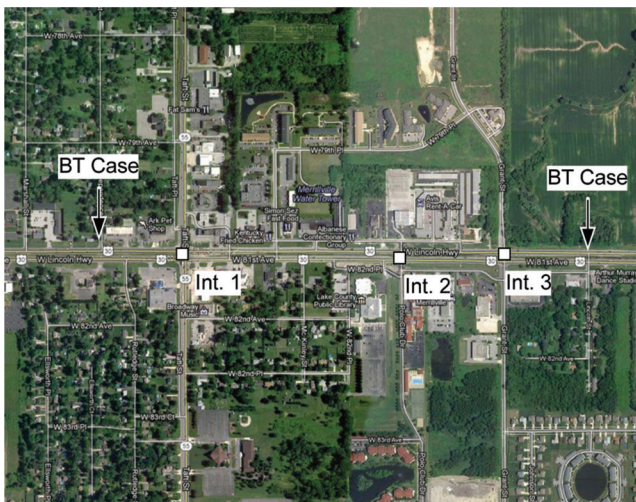
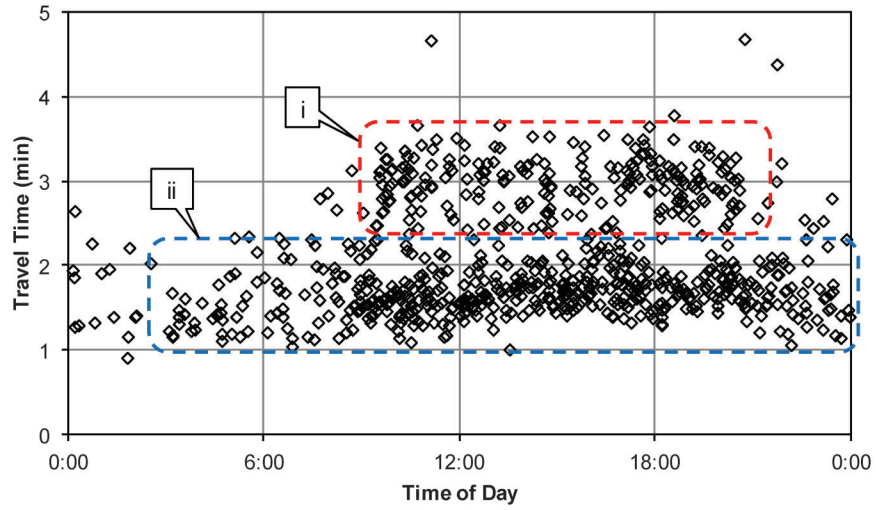


Figure 5.12 AVI travel time data collection on US 30.

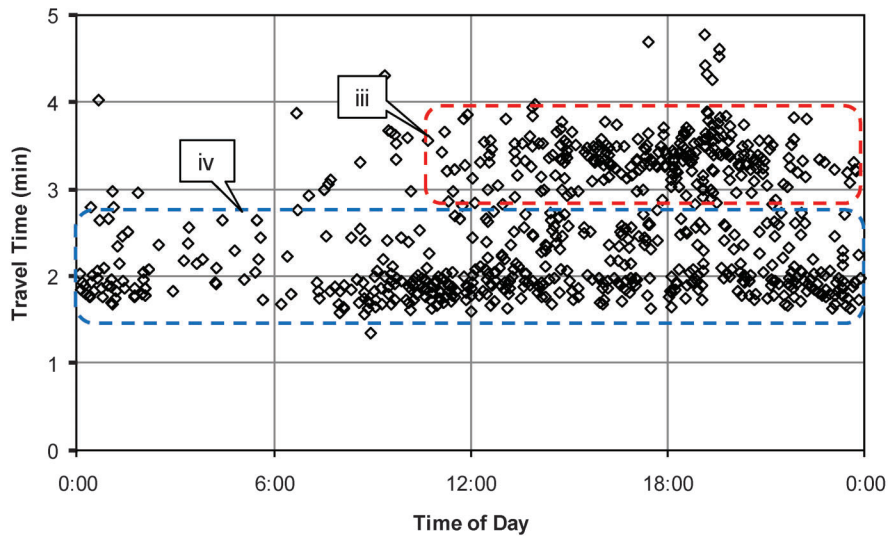
corridor (*i.e.*, from BMS-1 to BMS-8) would be considered. The distributions in Figure 5.16a capture some interesting details that are not visible when only endpoint-to-endpoint travel times are compared. For example, if we scan across the figure from left to right we see that the slope of the CFDs gradually decreases. As the comparison segments grow larger, the amount of variability tends to naturally increase, as variation in driver speeds translate into greater variation in travel time as more distance is accumulated.

5.5.2 Destination-Based Travel Time

The next data reduction technique is called “destination based travel time” (Figure 5.15b, Figure 5.16b). Similar to origin-based travel time, a sequence of BMS pairs is considered, only in this case they share a common exit point rather than a common entry. While some cars exit the arterial north of BMS-8, and others

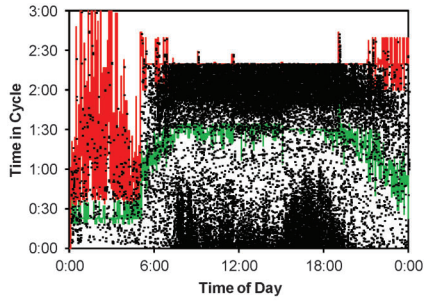


(a) Eastbound.

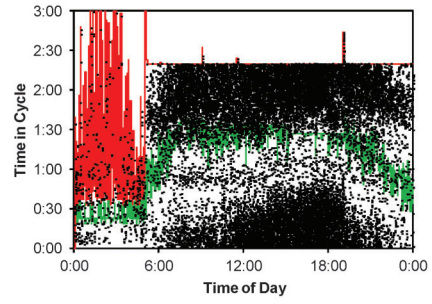


(b) Westbound.

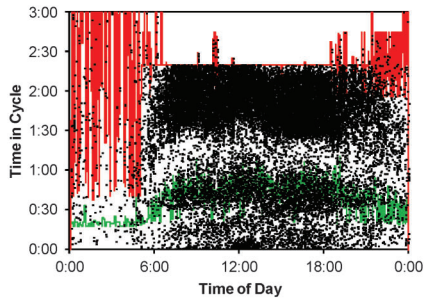
Figure 5.13 Travel times on US 30 between US 55 and Grant (2/9/2012).



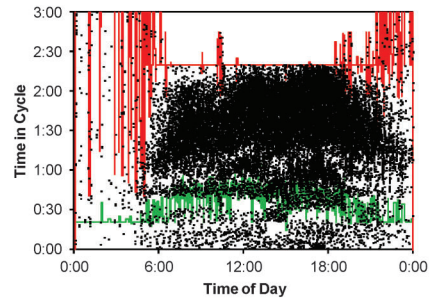
(a) Eastbound, Int. 1.



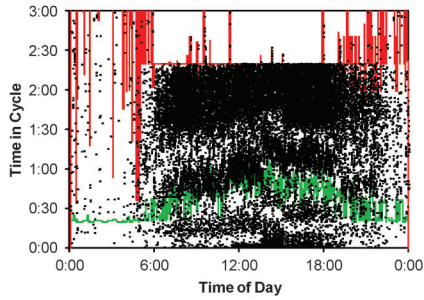
(b) Westbound, Int. 1.



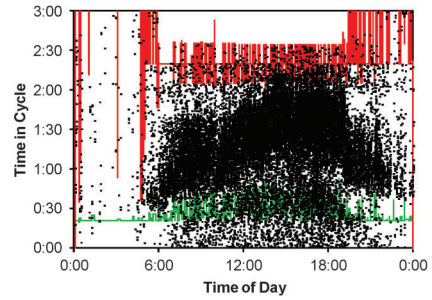
(c) Eastbound Int. 2.



(d) Westbound, Int. 2.



(e) Eastbound, Int. 3.



(f) Westbound, Int. 3.

Figure 5.14 Analysis of progression at Int. 1 (US 30 and SR 55), 2/9/2012.

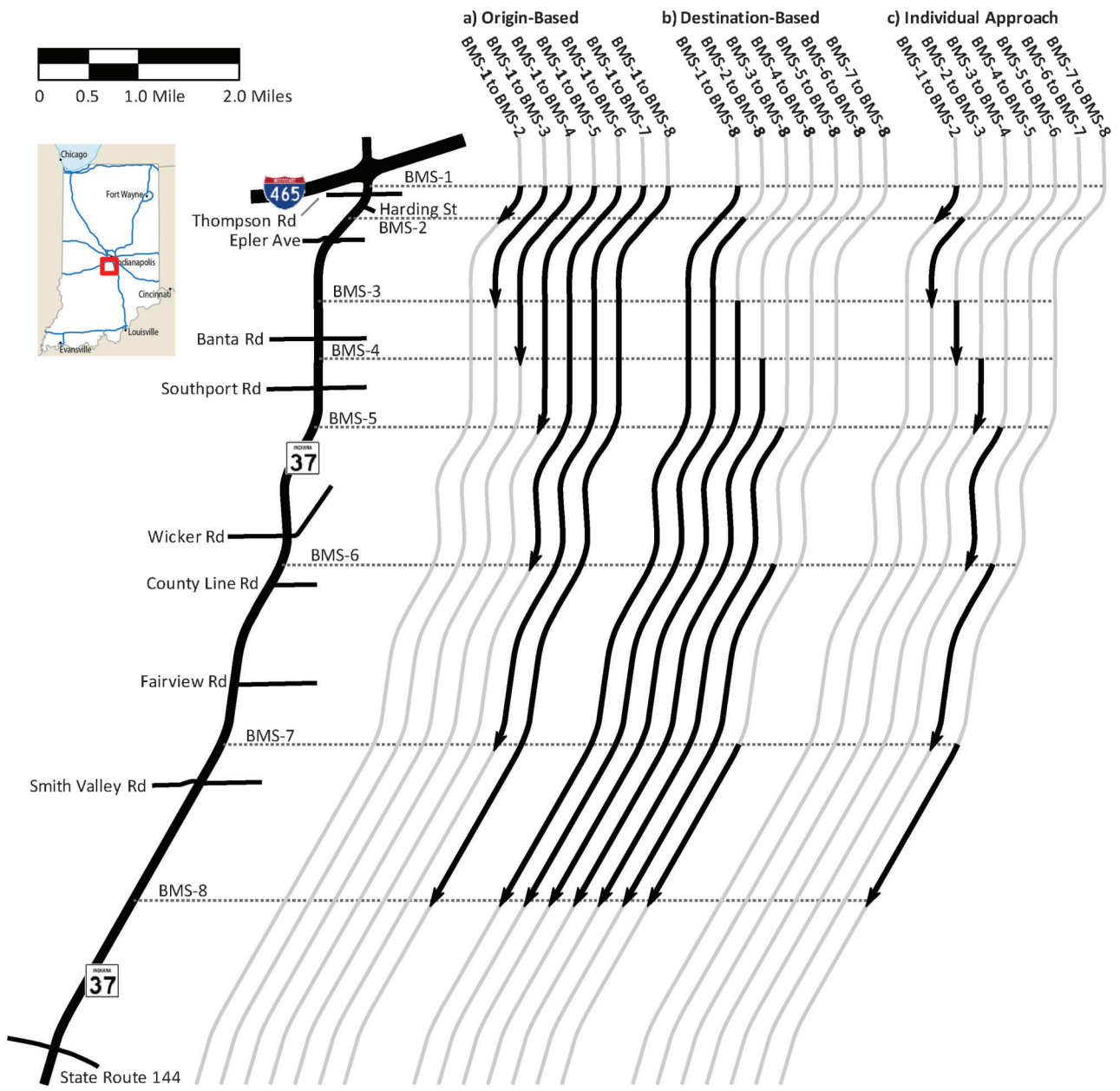
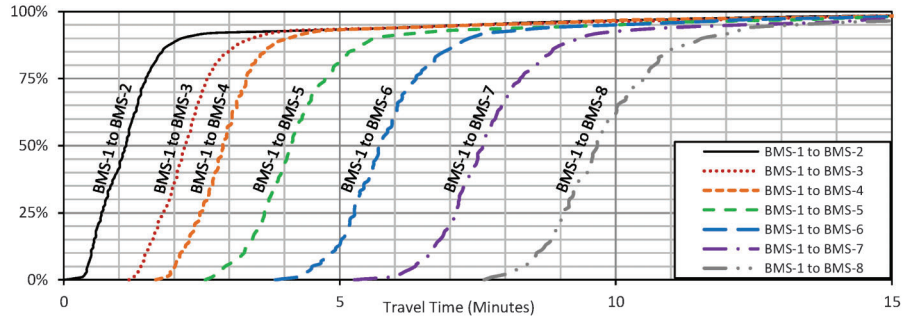
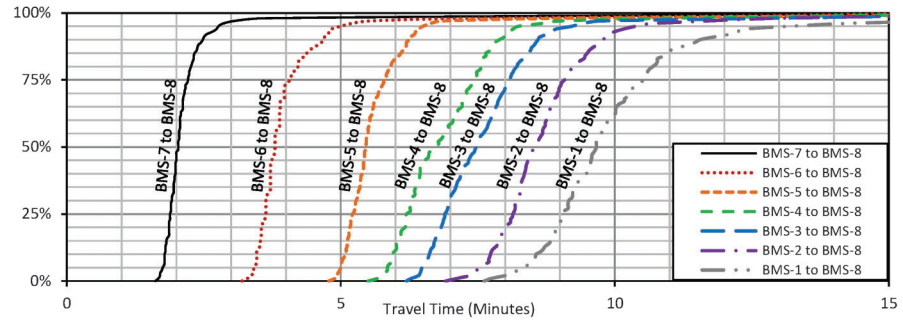


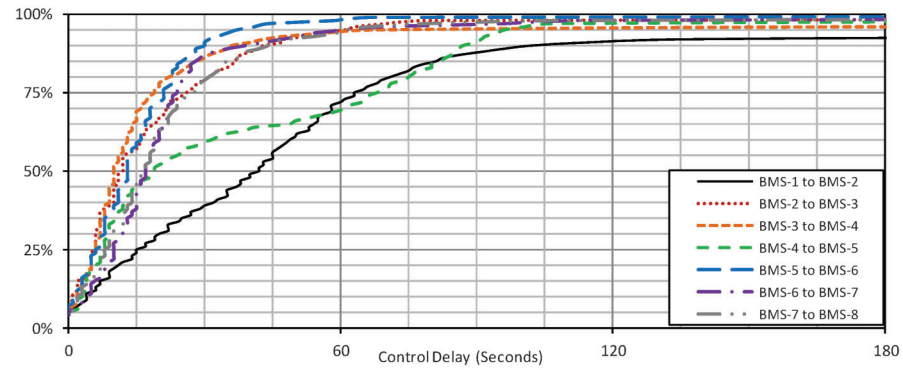
Figure 5.15 Bluetooth monitoring stations and subsection segment regimes on SR 37 south.



(a) Origin-based travel time (minutes).



(b) Destination-based travel time (minutes).



(c) Segment control delay (seconds) based on 5th percentile free flow travel time.

Figure 5.16 Southbound SR 37 travel time analysis during the PM peak period.

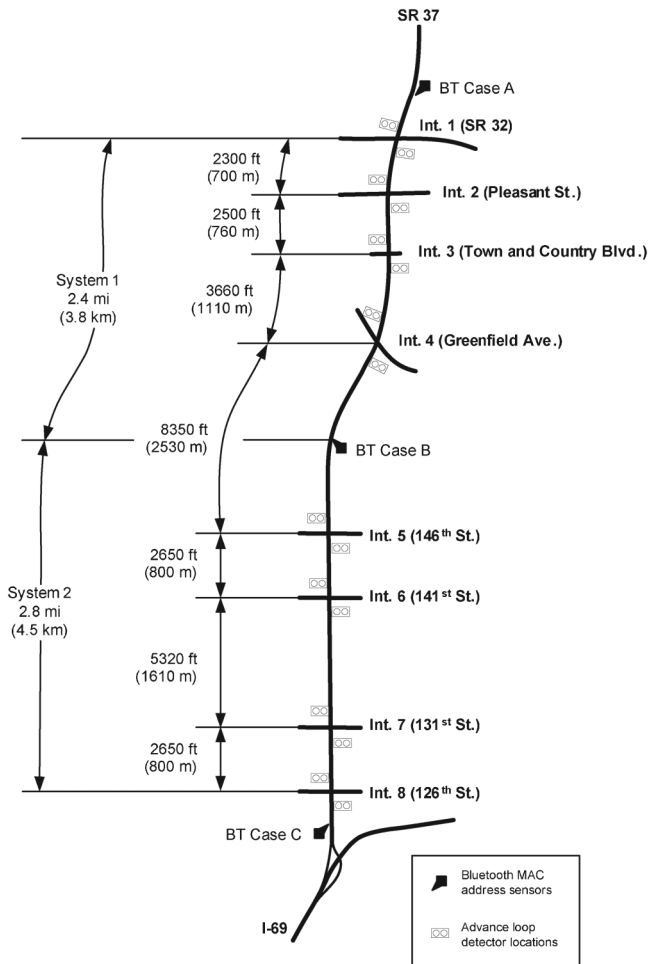


Figure 5.17 Map of the SR 37 corridor (25).

enter the system south of BMS-1, it is still important to examine how the southern portion of the corridor operates. The curves of the destination-based cumulative frequencies shown in Figure 5.16b appear similar to the origin-based travel time cumulative frequencies curves shown in Figure 5.16a. However, as discussed in the following section, a slight anomaly is discernible between BMS-5 to BMS-8 and BMS-4 to BMS-8.

5.5.3 Individual Intersection Approach Delay Segmentation

The third data reduction technique is called “individual intersection approach delay segmentation” (Figure 5.15c, Figure 5.16c). This method examines the southbound traffic on individual links within the corridor. This creates a series of links including BMS-1 to BMS-2, BMS-2 to BMS-3, BMS-3 to BMS-4, and so forth. Since the length of each link varies, control delay will be estimated. Control delay is here defined as the travel time, less the time required to proceed through the segment at the free flow speed. The 5th percentile of measured travel times was used as the free flow speed; vehicles traveling more rapidly than this percentile were

excluded from the analysis. Figure 5.16c shows the CFDs of delay for each sample.

The curve labeled “i” for BMS-1 to BMS-2 (which corresponds to the curve as shown in Figure 5.16a, shifted leftward by the 5th percentile travel time) and the curve labeled “ii” for BMS-4 to BMS-5 both appear to warrant further examination. The segment between BMS-1 and BMS-2 is near the exit ramp of I-465, and is influenced by random arrivals of traffic with high volume and turning movements. Curve “ii” is much more interesting; this curve represents a segment (BMS-4 to BMS-5) in the middle of the corridor. Recall that in the destination-based regime (Figure 5.16b), the discontinuity in the slope between the 25th and 75th percentiles was an anomaly that was hard to distinguish; it is quite apparent in terms of control delay (Figure 5.16c). This indicates a bimodal distribution, which results from traffic interaction with the only signal in that part of the corridor (the intersection of SR 37 and Southport Rd, Figure 5.15). While travel time may not be able to explicitly identify the cause of the anomaly, the characteristics of this curve suggest there is an offset that could be improved.

5.6 Case Study: Deriving User Benefit from Travel Times

In this chapter, we present results from a case study (25) in which field measurements of travel time were used to assess signal retiming activities on an eight-intersection corridor. Combining the reduction in corridor travel time with the traffic volumes enables us to compute an annual user benefit of \$500,000 resulting from signal retiming.

Figure 5.17 shows a map of the test corridor, which is SR 37 in Noblesville, Indiana. The southern group of four intersections (5–8) are the same as those presented in the previous chapter. Between May and July of 2010, several different signal timing optimization strategies were tested on this network, with travel time measurements collected from the Bluetooth cases as indicated on the map in Figure 5.17.

The four optimization objectives tested in this study were as follows (25):

- I. **Minimize delay.** The arrival and expected departure data contained in a flow profile (Figure 5.9a) was used to estimate the delay incurred for given arrival patterns relative to the probability of green. The signal offsets were adjusted to minimize the total delay for the system.
- II. **Minimize delay and the number of stops.** Rather than only delay, a composite objective function using delay, plus the number stops multiplied by 20 seconds of delay per stop, was minimized.
- III. **Maximize the number of arrivals on green.** This objective sought to maximize the number of arrivals in the green band. Rather than estimating delay or stops, the arrival data was directly used to locate the greatest portion of the arrival distributions where the probability of green was greatest.
- IV. **Maximize the number of arrivals on green, excluding the first 10 seconds of green.** The mechanism for achieving

objective was identical to III, except that the first 10 seconds of green were not considered part of the green distribution for the purpose of optimization. The idea in this case was to allow some time for queued vehicles to clear before the arrival of coordinated vehicles.

After computing four sets of optimal offsets, four new timing plans were deployed in the field on four

successive Saturdays between June and July 2010. Travel time measurements were obtained for each of these Saturdays using the sensor locations in Figure 5.17. The resulting CFDs for southbound and northbound vehicles across the three paths in the system are shown in Figure 5.18. The baseline travel time CFD and the four optimized travel time CFDs are labeled by callouts.

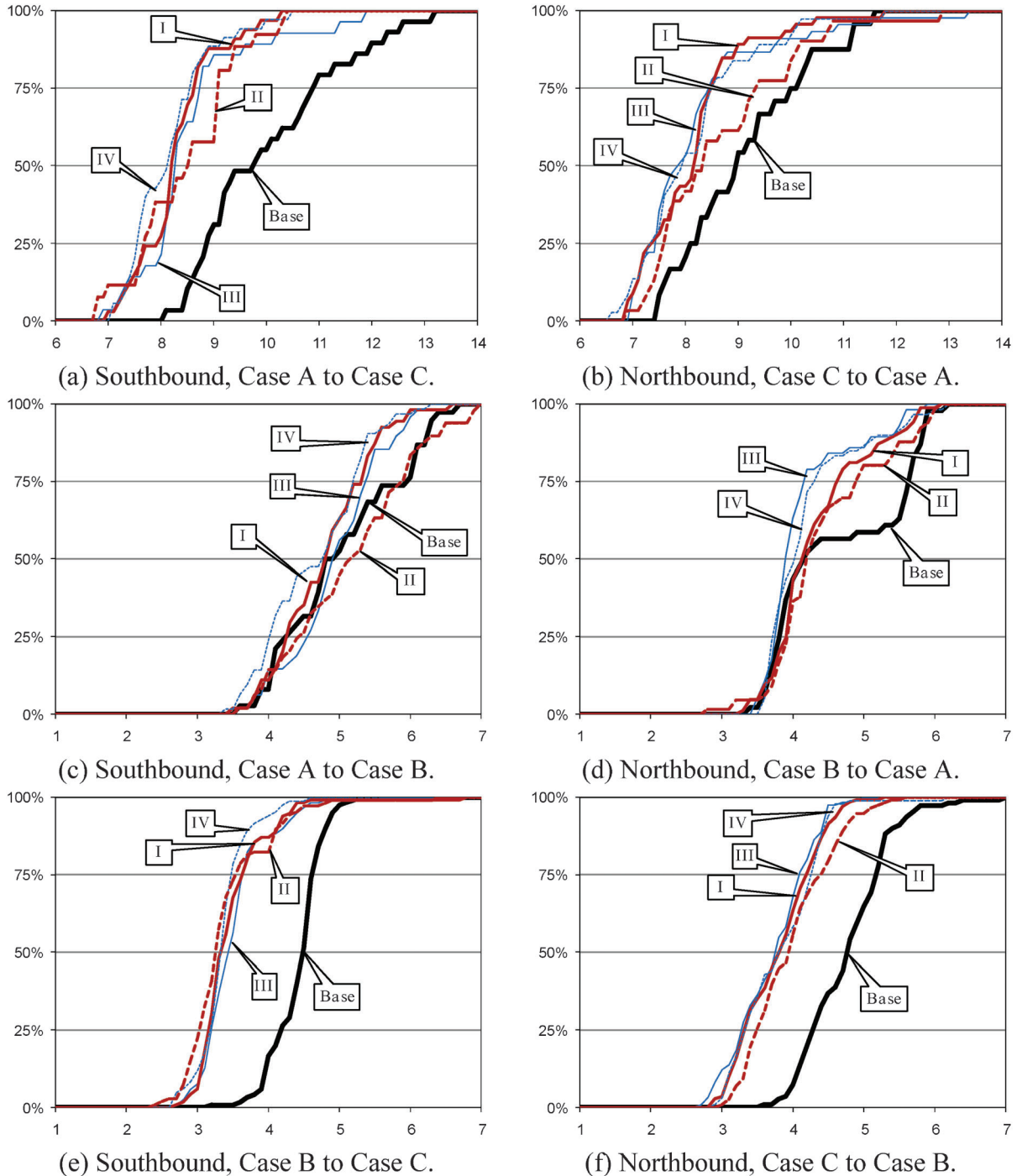


Figure 5.18 Cumulative frequency diagrams of anonymous AVI travel times for alternative objective functions, Saturday, 1500–1800 (25).

- Figure 5.18a and Figure 5.18b respectively show CFDs of southbound and northbound travel times through all eight intersections. Most of the optimized offsets led to overall improvements in travel time along the corridor of approximately 1–1.5 minutes in the southbound and 1 minute in the northbound.
- Figure 5.18c and Figure 5.18d show CFDs of travel time through the four intersections on the northern subsystem (Ints. 1–4). There was little change in southbound travel times on this subsystem, and slight improvements for northbound travel times. Although the median northbound travel time improved by only a slight amount, the upper portion of the curves shifted to the left, indicating a reduction in stops somewhere within the system.
- Figure 5.18e and Figure 5.18f show CFDs of travel time through the four intersections on the southern subsystem (Ints. 5–8). Substantial improvements in travel time were achieved in both directions, for all of the tested objective functions.

The changes in the median travel times are summarized in Table 5.2, while changes in the median compared to the baseline travel time are shown in Table 5.3. These tables show the changes in travel times broken down into 3-hour analysis periods, which shows some interesting trends by time of day.

TABLE 5.2
Median travel times.

System	Time	Southbound Travel Time (min)					Northbound Travel Time (min)				
		Base	I	II	III	IV	Base	I	II	III	IV
1	6–9	3.70	3.60	3.70	3.90	3.70	3.80	4.00	3.90	3.75	3.80
	9–12	4.40	4.00	4.30	4.00	4.20	5.20	4.95	5.10	4.25	5.05
	12–15	4.90	5.30	4.40	4.70	4.40	5.50	6.20	5.60	5.55	5.60
	15–19	4.35	3.90	4.20	4.10	4.20	5.60	4.95	4.85	4.80	5.20
	19–21	4.30	4.00	3.95	3.80	4.00	4.50	4.55	4.10	4.00	4.50
2	6–9	3.60	3.60	3.90	3.60	3.50	4.70	3.50	3.50	3.45	3.45
	9–12	4.40	3.60	3.70	3.80	3.80	4.70	3.60	3.40	3.50	3.40
	12–15	4.50	3.90	3.90	3.55	4.10	4.60	3.50	3.40	3.40	3.40
	15–19	5.20	3.80	3.90	3.90	3.85	4.45	3.50	3.40	3.40	3.30
	19–21	4.35	3.60	4.00	3.70	3.80	4.50	3.40	3.30	3.30	3.20

TABLE 5.3
Reductions in median travel time compared to baseline.*

System	Time	Southbound Travel Time (min)				Northbound Travel Time (min)			
		I	II	III	IV	I	II	III	IV
1	6–9	0.10	0.00	–0.20	0.00	–0.20	–0.10	0.05	0.00
	9–12	0.40	0.10	0.40	0.20	0.25	0.10	0.95	0.15
	12–15	–0.40	0.50	0.20	0.50	–0.70	–0.10	–0.05	–0.10
	15–19	0.45	0.15	0.25	0.15	0.65	0.75	0.80	0.40
	19–21	0.30	0.35	0.50	0.30	–0.05	0.40	0.50	0.00
2	6–9	0.00	–0.30	0.00	0.10	1.20	1.20	1.25	1.25
	9–12	0.80	0.70	0.60	0.60	1.10	1.30	1.20	1.30
	12–15	0.60	0.60	0.95	0.40	1.10	1.20	1.20	1.20
	15–19	1.40	1.30	1.30	1.35	0.95	1.05	1.05	1.15
	19–21	0.75	0.35	0.65	0.55	1.10	1.20	1.20	1.30

*Negative numbers represent increases in travel time.

The largest reductions occurred during the 1500–1800 time period.

The changes in median travel times are useful for reporting improvement in system performance, but it is much more compelling, particularly when justifying the investment in retiming activities, to convert these savings into dollar amounts. To do so, we adopted the methodology of the 2009 Transportation Urban Mobility Report (26) to convert travel time changes into estimated user benefit and environmental savings.

The change in travel times (Table 5.3) are first computed by section :

$$\Delta TT = TT_{Base(section)} - TT_{Objective(section)} \quad (3)$$

where $TT_{Base(section)}$ is the arterial travel time (minutes) for System 1 or System 2 (Figure 5.17) and direction (northbound or southbound) measured during the baseline conditions, and $TT_{Objective(section)}$ is the travel time for each section, measured when new offsets were deployed that had been optimized according to the given objective.

The user costs for commercial vehicles ($USER_c$) are given by

TABLE 5.4
Summary of cost savings for alternative optimization objectives, by section.

Objective	Daily				Annual				
	Total Time Saved (veh-min)	CO ₂ Emission Reduction (tons)	CO ₂ Savings	User Benefits	Multiplier	CO ₂ Emission Reduction (tons)	CO ₂ Savings	User Benefits	
(a) System 1, Northern Section									
I	Min Delay	5032	0.71	\$16	\$1,697	52	37	\$810	\$88,233
II	Min Delay & Stops	3813	0.54	\$12	\$1,286	52	28	\$614	\$66,864
III	Max N_g	1760	0.25	\$5	\$593	52	13	\$283	\$30,855
IV	Alt. Max N_g	7883	1.11	\$24	\$2,658	52	58	\$1,268	\$138,229
(b) System 2, Southern Section									
I	Min Delay	24386	3.43	\$75	\$8,223	52	178	\$3,924	\$427,614
II	Min Delay & Stops	25327	3.56	\$78	\$8,541	52	185	\$4,075	\$444,111
III	Max N_g	25147	3.54	\$78	\$8,480	52	184	\$4,046	\$440,962
IV	Alt. Max N_g	26338	3.70	\$81	\$8,882	52	193	\$4,238	\$461,845
(c) System 1 and System 2, Arterial									
I	Min Delay	29418	4.14	\$91	\$9,920	52	215	\$4,733	\$515,847
II	Min Delay & Stops	29140	4.10	\$90	\$9,826	52	213	\$4,689	\$510,976
III	Max N_g	26907	3.78	\$83	\$9,073	52	197	\$4,329	\$471,817
IV	Alt. Max N_g	34221	4.81	\$106	\$11,540	52	250	\$5,506	\$600,073

$$USER_t = \Delta TT * Vol * \%T * PPV_t * \frac{\$102.12}{hr} * \frac{1hr}{60min} \quad (4)$$

where Vol is the volume (number of vehicles) during the analysis period, $\%T$ is the percentage of commercial vehicles (“trucks”, estimated at 2% for Saturdays), and PPV_t is the number of passengers per vehicle (1). The \$102.12 amount used for the time value of money for commercial vehicles is taken from the 2009 Transportation Urban Mobility Report (26). Where ΔTT is positive, the outcomes of the equation are savings.

User costs for passenger cars ($USER_c$) are given by

$$USER_c = \Delta TT * Vol * \%C * PPV_c * \frac{\$15.47}{hr} * \frac{1hr}{60min} \quad (5)$$

where $\%C$ is the assumed percentage of passenger vehicles (estimated at 98% for Saturdays), PPV_c is assumed to be 1.2, and a lower time value of money at \$15.47 per hour (26) is applied.

Reductions in fuel consumption and CO₂ emissions are derived as follows:

$$FUEL = \Delta TT * Vol * \frac{0.87gal}{hr} * \frac{1hr}{60min} \quad (6)$$

In Equation 6, $FUEL$ is the change in the amount of fuel consumed (gallons), which is a reduction when

ΔTT is positive. Using conversion factors from Argonne National Laboratory, a passenger car that idles at 1,000 rpm, while using air conditioning 50% of the time, consumes 0.87 gallons of gasoline per hour, or 0.0145 gallons per minute (26). The corresponding reductions in CO₂ emissions are calculated from the following two equations:

$$CO_2 = FUEL * \frac{19.4lbs}{gal} * \frac{1ton}{2000lbs} \quad (7)$$

$$CC = CO_2 * \frac{\$22}{ton} \quad (8)$$

Here, CC represents the “CO₂ cost.” According to the EPA, the amount of CO₂ emitted when a gallon of gasoline burns is approximately 19.4 lbs/gallon (27). The monetary equivalent of the CO₂ is assumed to be approximately \$22/ton of CO₂ produced (28).

Table 5.4 shows the final tabulation of the overall user and environmental benefits, converted into dollars, based on the analysis described above. The total savings were tabulated by section and by time period, which were then summed. The number of vehicle minutes saved, for example, were calculated by multiplying the reduction in median travel time (Table 5.3) by the expected volumes (computed as an average across all five days in the sampled dates), for each time period, then summing across time periods to get the benefit for all of Saturday. Table 5.4 also shows these numbers when they are multiplied by 52, to produce an annual

benefit. These results show that user cost reductions ranging from \$470,000 to \$600,000 can be realized, varying by objective.

6. ORIGIN-DESTINATION MATRIX ESTIMATION

Origin destination information and quantitative performance of critical routes is often challenging to acquire for special events or other non-recurring activities. This chapter⁷ examines the feasibility of utilizing anonymous high-resolution automatic vehicle identification (AVI) travel time data to assess route distribution during a nighttime football game attended by approximately 65,000 fans. Twelve monitoring stations were used to collect anonymous MAC addresses from Bluetooth-enabled devices. That data was subsequently processed to provide quantitative data documenting:

- Inbound route choice
- Outbound route choice
- Location and magnitude of inbound and outbound congestion.

Unlike many other special event data collection efforts that either require intensive manual labor or extensive permanent surveillance technology, the Bluetooth AVI travel time data procedures described in this paper employs low-cost, battery power data collection devices that can be cost-effectively deployed for virtually any special event drawing a substantial crowd.

6.1 Study Background

Special events often have substantial impacts on a local or a regional area. Examples of estimated economic impacts of major special events are shown in Table 6.1. Several resources exist to assist agencies with dealing with special events, such as the FHWA *Special Events Handbook* (29). Agencies often cooperate and counsel each other for exceptional events planned well ahead of time. With each special event that takes place, there is an opportunity to deploy data collection to build a library of quantitative data to assess the effectiveness of special event operations. These after-action reviews can then help support future event planning and facilitate peer exchange.

The case study in this chapter is based on a Purdue-Notre Dame football game that was played at the Purdue campus in West Lafayette, Indiana on Saturday, 10/1/2011. This game had an attendance of 61,555 fans, which is approximately 150% of the attendance of most home games, likely because Purdue University and Notre Dame are both large universities in Indiana with popular football programs.

⁷This chapter is adapted from Hainen, A. M., E. M. Rivera-Hernandez, S. R. Mitkey, S. M. Remias, and D. M. Bullock. *Use of Bluetooth Technology to Estimate Travel Time and Route Choice for Special Events*. White paper for SBIR DTR57-8-SBIR2.

TABLE 6.1
Sample economic impact from popular special events (29).

Event	Annualized Economic Impact (million USD)
Summer Olympics	700
NAIAS*	580
Daytona 500*	554
Super Bowl XXXVII	367
Indianapolis 500*	337
New York International Auto Show*	187
Ryder Cup	114
Oscars*	63
Kentucky Derby*	53
Major League Baseball All-Star Game	53
World Series	40

*Estimated annual economic impact.

The travel demand generated by events such as this significantly loads the road network in the cities of Lafayette and West Lafayette, Indiana, well above normal levels. This case study demonstrates how AVI data can be used to characterize route choice preferences and changes in travel time during special events.

To collect data, twelve Bluetooth Monitoring Stations (BMS) were deployed along the major routes to and from Ross-Ade Stadium in West Lafayette, Indiana. The BMS were deployed on 9/22/2011 in the locations shown Figure 6.1. The setup was designed to monitor incoming and outgoing traffic along the major ingress/egress corridors: US 52, US 231, SR 43, SR 26, and SR 38 as well as the local traffic along West Lafayette city roads near the stadium, as shown in Figure 6.2. Figure 6.3 provides a pictorial review of sensor locations. As demonstrated in these photos, these temporary installations consisted of equipment cases securely attached to local signs and utility poles. Additional sensors were located along SR 25 and SR 38 near I-65, but these sensors failed during the data location.

6.2 Accommodating Routes with Missing Data

Table 6.2 shows the number of samples from point to point in the sensor network. While this can be useful for data validation, macro-scale route choice is more valuable. Because the sensors located on SR 26 and SR 38 failed, some imputation was required to estimate these volumes. Figure 6.4 illustrates this concept for SR 38. Figure 6.4a presents a map of the physical configuration of sensors and roadways of interest, while Figure 6.4b shows a conceptual map. Callout “i” indicates the composite origin zone used as the origin of traffic. Matched volumes recorded by BMS at STA05, STA09, and STA10, and on routes between STA05-STA10 and STA05-STA09 are shown in Figure 6.4b.

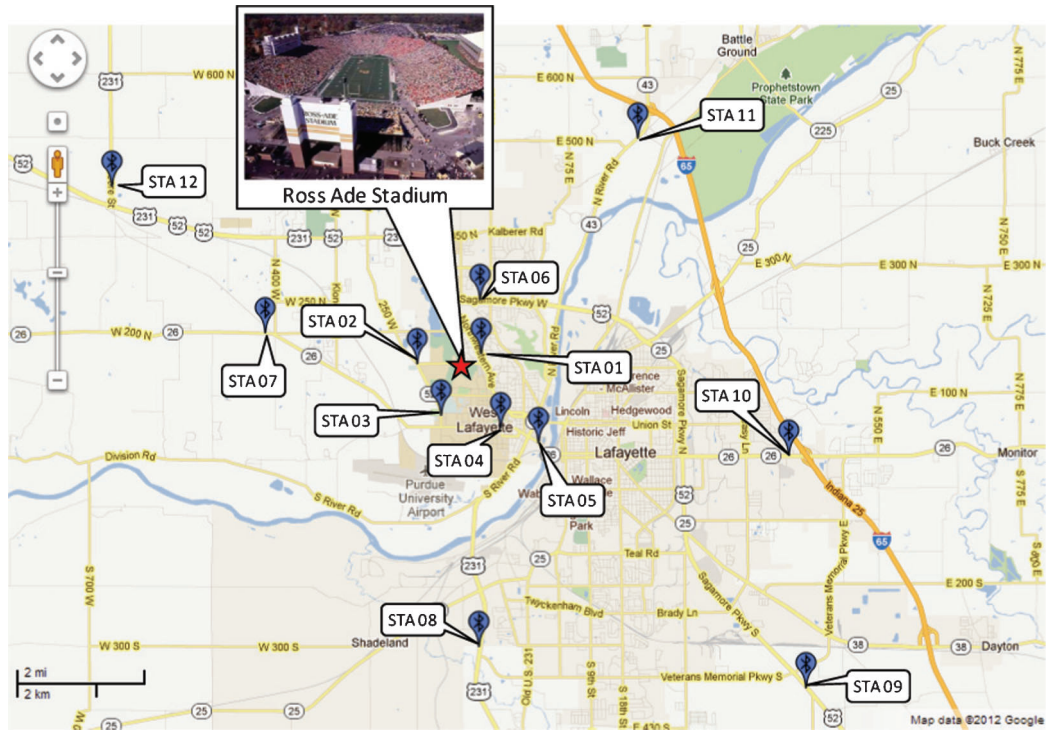


Figure 6.1 Equipment layout and numbering regime.

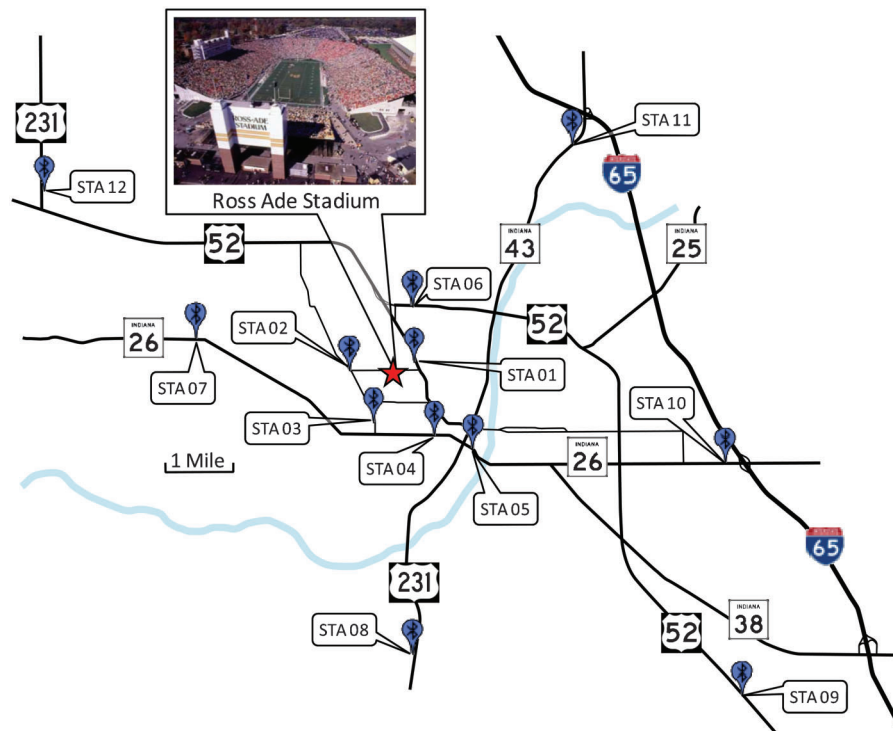


Figure 6.2 Schematic layout of network noting sensor locations.



Figure 6.3 Bluetooth monitoring station installations.

TABLE 6.2.
Origin-destination matrix on 10/1/2011 between 1200 and 2359.

Origin Station	STA	Destination Station											
		01	02	03	04	05	06	07	08	09	10	11	12
01			225	135	63	47	190	5	14	5	14	13	30
02		89		305	92	42	76	11	11	1	5	13	27
03		49	199		237	109	23	16	17	6	17	14	17
04		75	79	263		360	16	18	46	14	69	26	1
05		90	50	152	455		55	17	35	18	135	47	1
06		362	123	72	28	47		11	12	8	41	23	11
07		6	21	25	29	23	11		3	2	4	1	5
08		14	4	29	72	72	14	1		18	5	41	1
09		6	1	11	20	27	9	0	26		26	0	0
10		23	15	33	82	186	44	4	6	9		22	0
11		56	27	40	78	62	63	1	42	0	23		1
12		46	34	30	5	0	10	5	0	0	0	1	

289 matches were observed between the origin zone and STA05. Additionally, 175 matches and 37 matches were also respectively observed between that location and STA09 and STA10. If it is assumed that all 289 matches between the origin zone and STA05 are vehicles leaving the region, then the volume of traffic leaving on SR 38 can be estimated from $289 - 175 - 37 = 77$. A few notes on this imputation are in order:

The 289 vehicles matched between the origin and STA05 are not the same exact vehicles as the 175 measured between STA05 and STA10, or the 32 measured between STA05 and STA09. Assuming that match rates are roughly equal for all routes spanning any two sensors in the network, the volume of matches is expected to be proportional to the traffic volumes between those locations.

Matching directly between the origin zone to STA09 or STA10 finds higher sample rates; this suggests that some traffic may be using a route other than through STA05 to reach these destinations (for example, the local bridge just north of STA05 in Figure 6.4a). Therefore, the conceptual network may not include all available routes. This means there may be additional traffic heading to SR 38 that used alternate routes; This could lead to underassignment of traffic using SR 38.

It is possible that some of the imputed volume for SR 38 was local traffic, which could lead to overassignment of traffic.

Although the imputation method may introduce some variances in the O-D volumes, this technique shows that missing data can be reasonably estimated from other detector data. These contingencies should be taken into consideration when planning sensor locations. Importantly, there is some value to having intermediate sensors in the system, rather than having only sensors near the origin and at cordon points.

6.3 Origin-Destination Volume Observations

Figure 6.5 and Figure 6.6 respectively document the proportion of matched vehicles on the inbound and

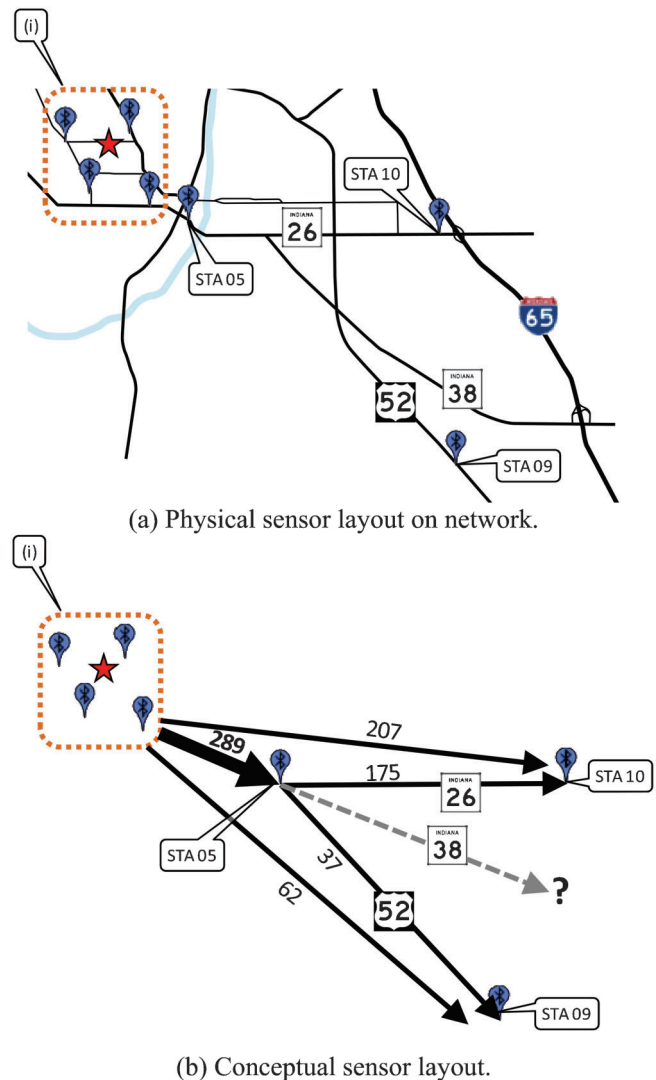


Figure 6.4 Interpolation of utilization for unobserved route.

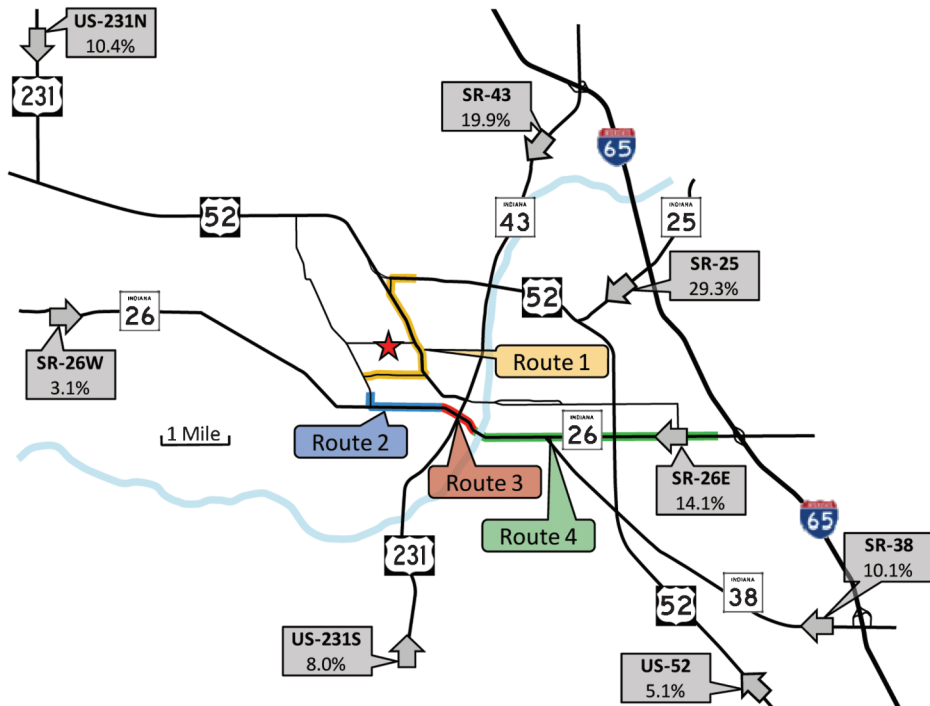


Figure 6.5 Inbound route utilization.

outbound routes. Because the visiting team (Notre Dame) is located in northern Indiana, it is not surprising that approximately 50% of the inbound traffic came from SR 25 and SR 43. Interestingly, these routes accounted for a smaller percentage of the outbound traffic, perhaps because many of the

attendants stayed in local hotels after the late night game or continued their travel toward Indianapolis.

With regards to the routes leading east and south, there were modest differences in route choice between inbound and outbound, with the most substantial change along SR 26. This may be attributable to the

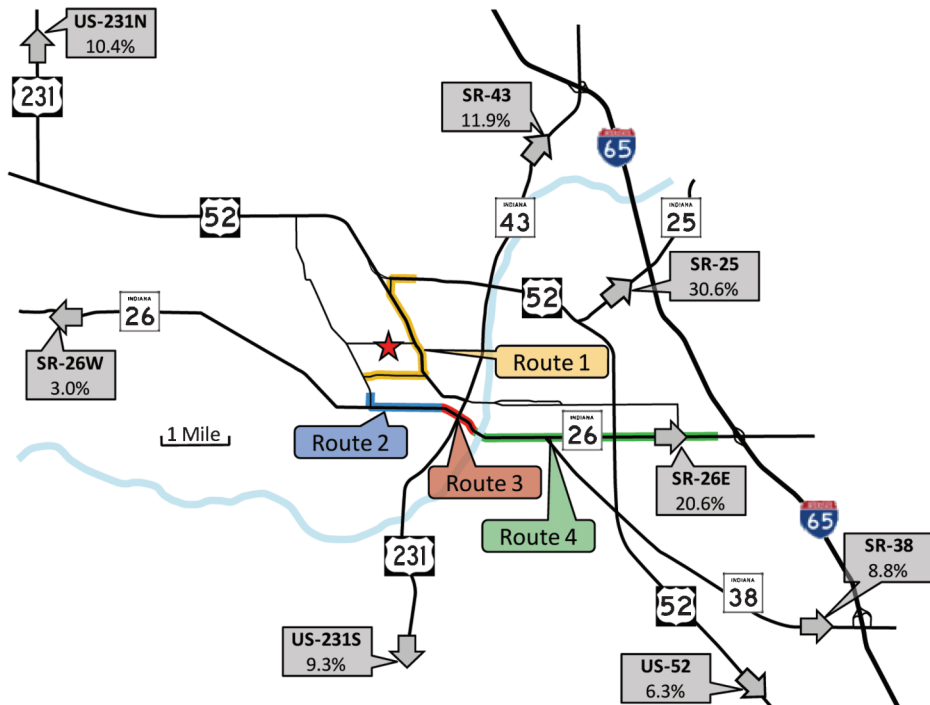


Figure 6.6 Outbound route utilization.

land use on SR 26, which is often congested during typical daytime hours, so motorists may have actively avoided it while traveling in bound. When the game ended (around midnight), departing fans appeared to make heavier use of SR 26, perhaps anticipating that it would not be congested at that time of day.

The routes leading to the west of campus (US 231N and SR 26 West) has almost perfectly balanced proportion of traffic between inbound and outbound. These routes serve rural areas to the west of campus; they are the only feasible routes for many locations in that direction.

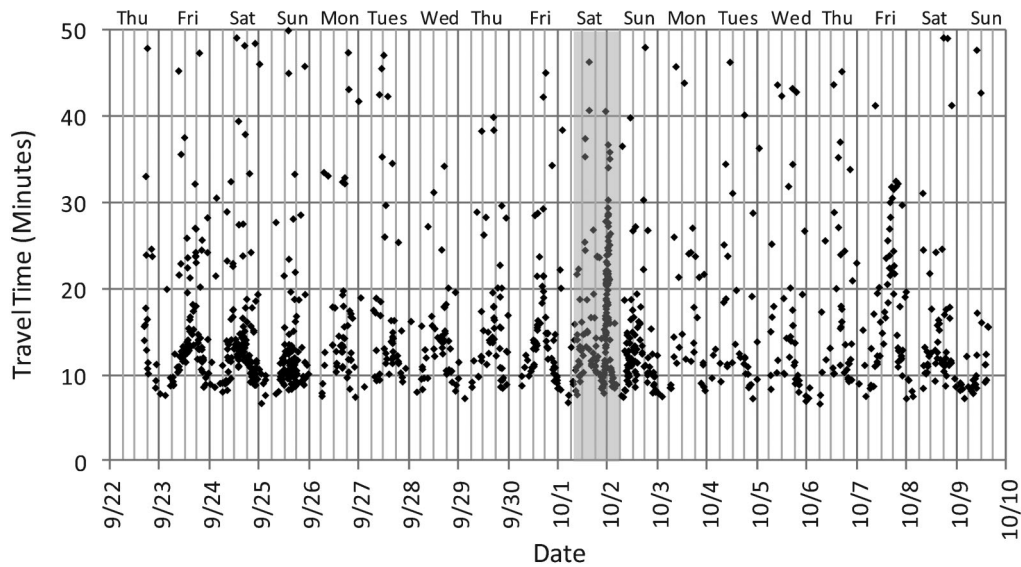
6.4 Critical Links

Figure 6.5 and Figure 6.6 have four callouts marking four routes (Route 1, Route 2, Route 3, and Route 4)

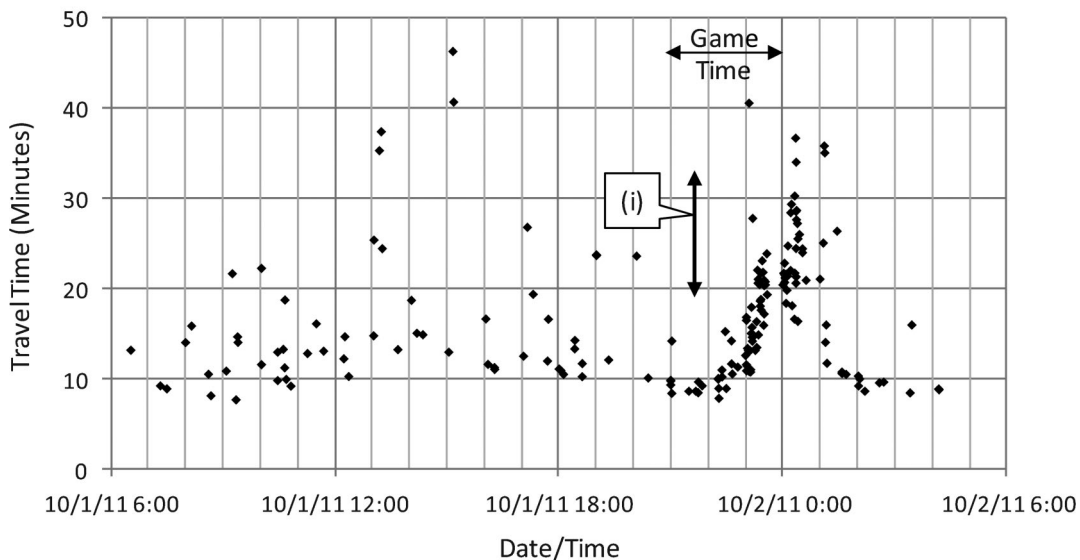
that tend to experience congestion during certain periods of the pre and postgame events.

Figure 6.7a and Figure 6.8a show the two week travel time characteristics for Route 1. Figure 6.7b suggests there is perhaps some short term congestion just before noon when the parking lots open and approximately 7–8 minutes of delay along this corridor in the late afternoon around 1700. However, the more substantial congestion occurred during the outbound Figure 6.8b period starting at approximately 11pm and lasting until about 1am the next day. The peak delay was approximately 20 minutes.

Figure 6.9a and Figure 6.10a show the two week travel time characteristics for Route 2. In general there was no inbound congestion observed. The outbound congestion characterized in Figure 6.10b indicates



(a) Two week travel time.



(b) Game day travel time.

Figure 6.7 Travel time scatter plots for inbound traffic on Route 1.

congestion occurred from just after 11pm to just before 1am, but delay was less than ten minutes.

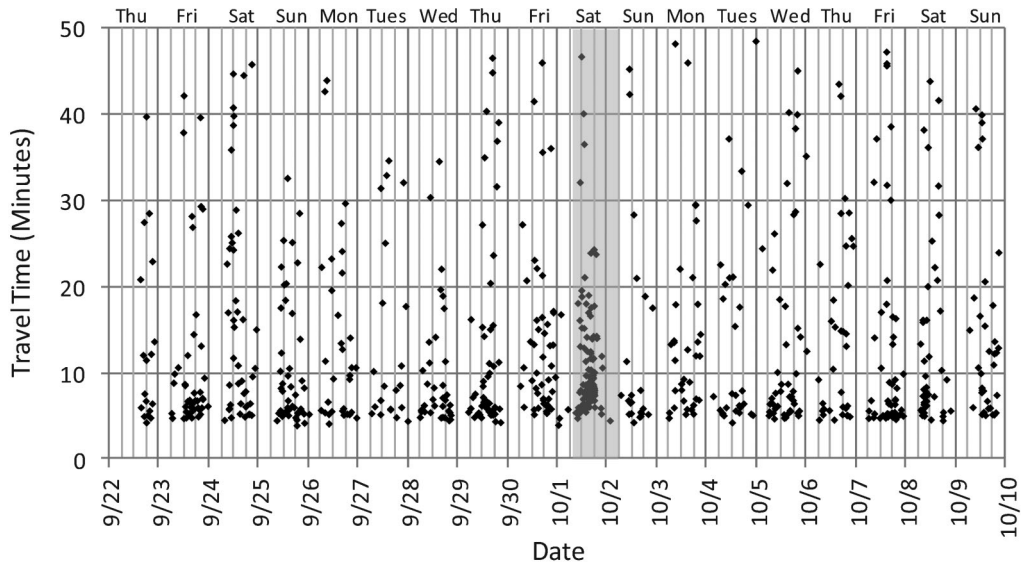
Figure 6.11a and Figure 6.12a show the two week travel time characteristics for Route 3 through the village area of campus where there are restaurants and bars. In general there is very little inbound congestion. The outbound congestion shown in Figure 6.12b again show congestion lasting from approximately 11pm to 1am. Peak delays during this period were 20–25 minutes along this segment.

Figure 6.13a and Figure 6.14a show the two week travel time characteristics for Route 4. Figure 6.13b indicates about 10 minutes of delay occur between 4pm and 6pm and perhaps 8 minutes of delay around

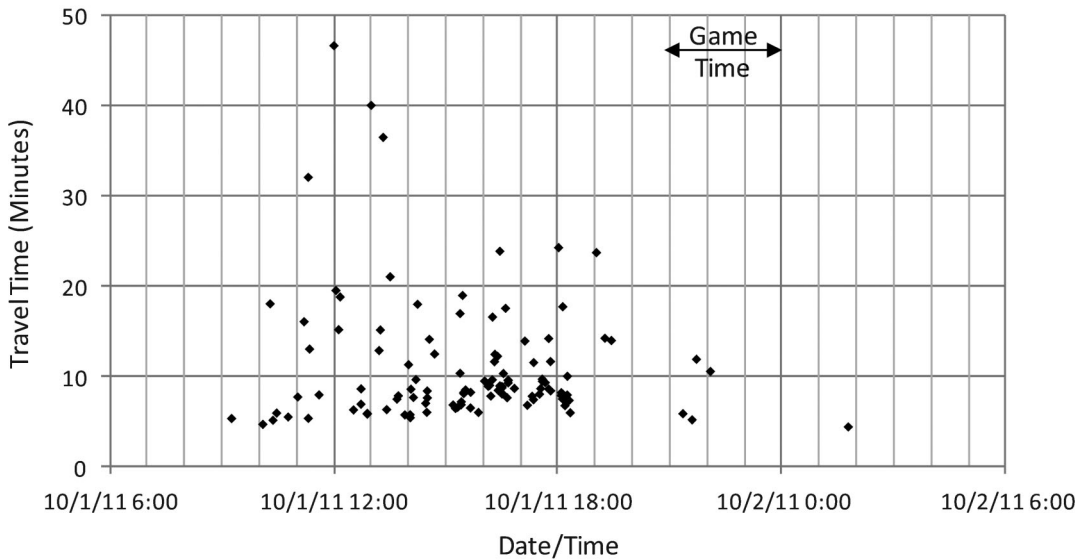
midnight in the outbound direction (Figure 6.14b). It is also clear that the outbound peak is over a shorter interval.

6.5 Conclusions

This case study used data collected during a special event to demonstrate how to collect quantitative special event data documenting route choice and location of congestion. Unlike many other special event data collection efforts that either require intensive manual labor or extensive permanent of surveillance technology, this paper documents a methodology that employs low-cost, battery power data collection devices that can

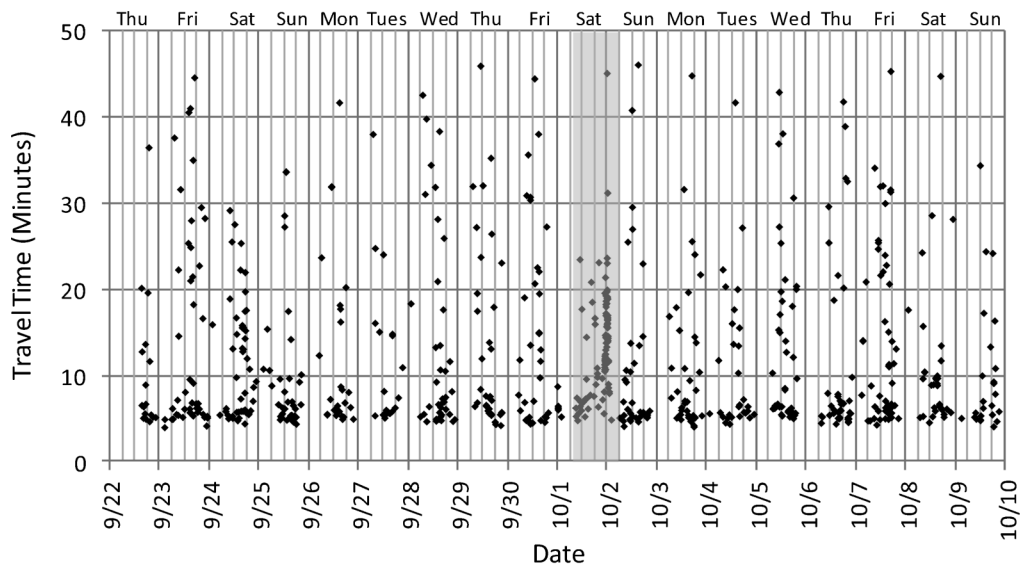


(a) Two week travel time.

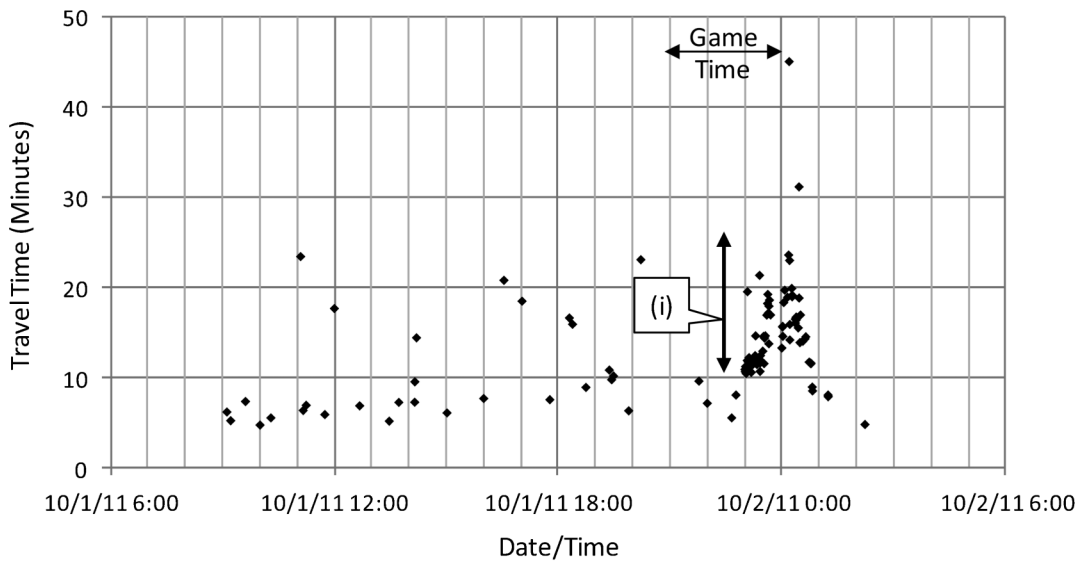


(b) Game day travel time.

Figure 6.8 Travel time scatter plots for outbound traffic on Route 1.

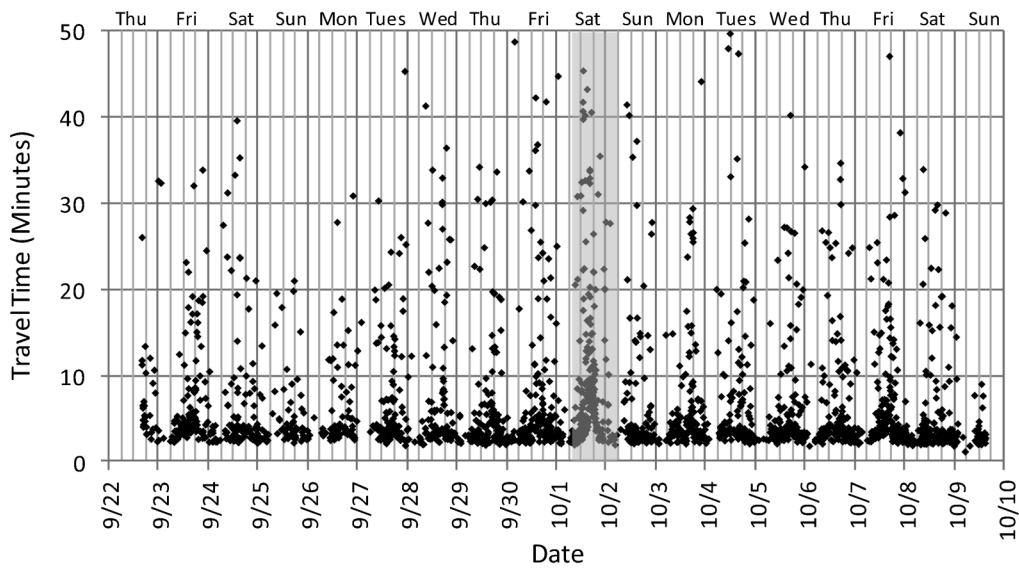


(a) Two week travel time.

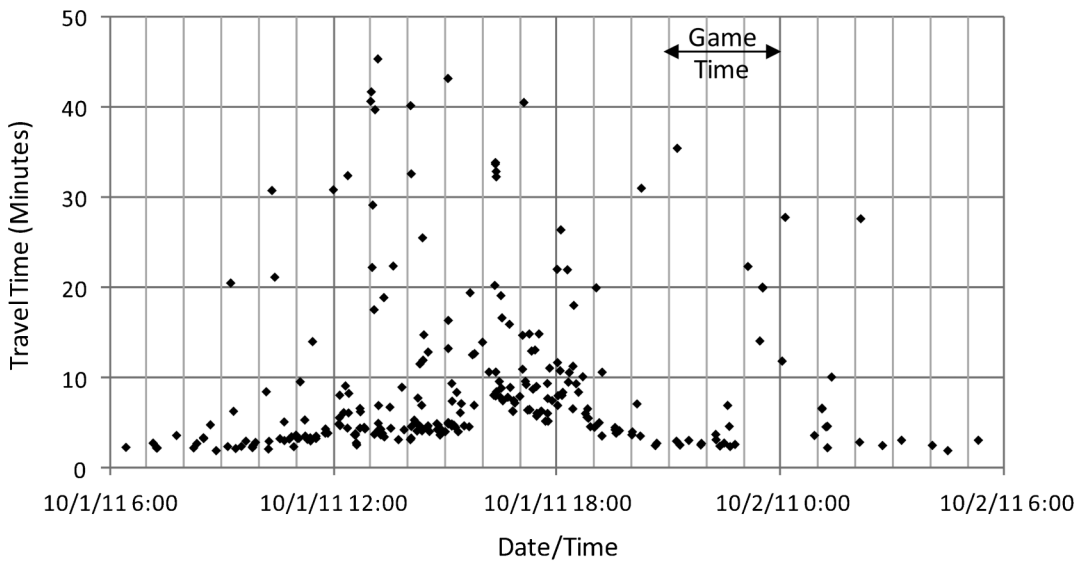


(b) Game day travel time.

Figure 6.9 Travel time scatter plots for inbound traffic on Route 2.

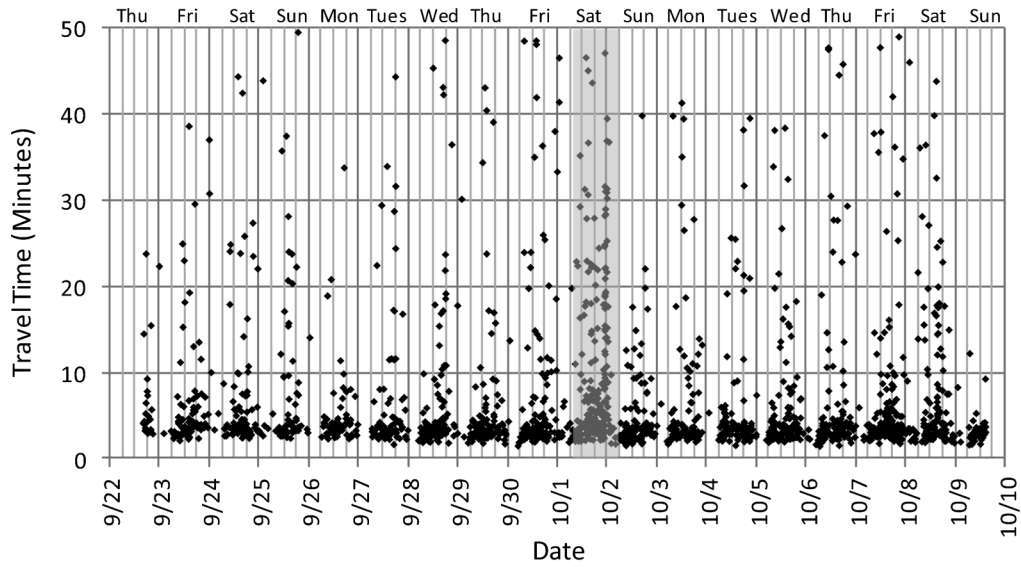


(a) Two week travel time.

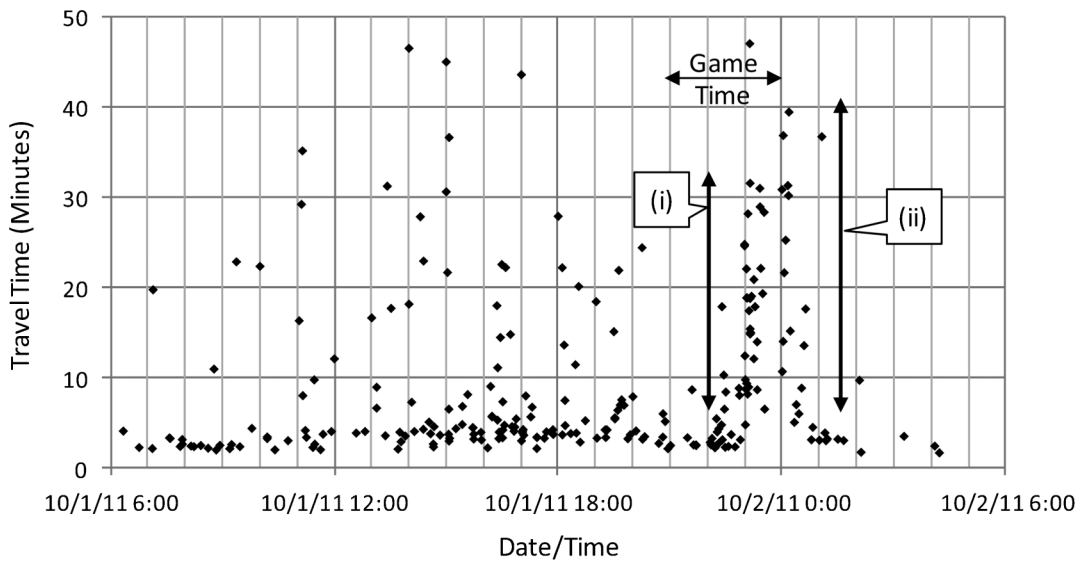


(b) Game day travel time.

Figure 6.10 Travel time scatter plots for outbound traffic on Route 2.

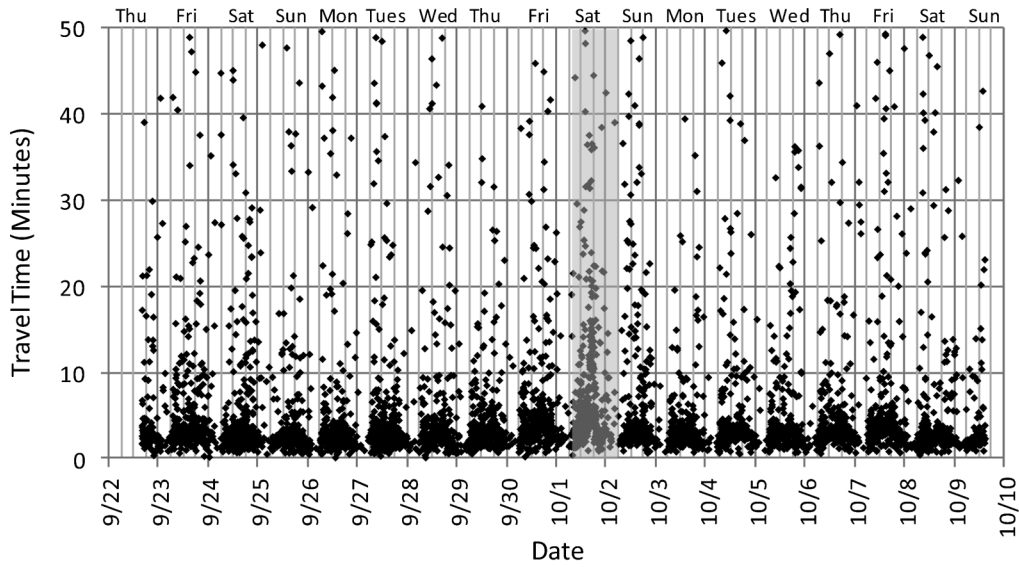


(a) Two week travel time.

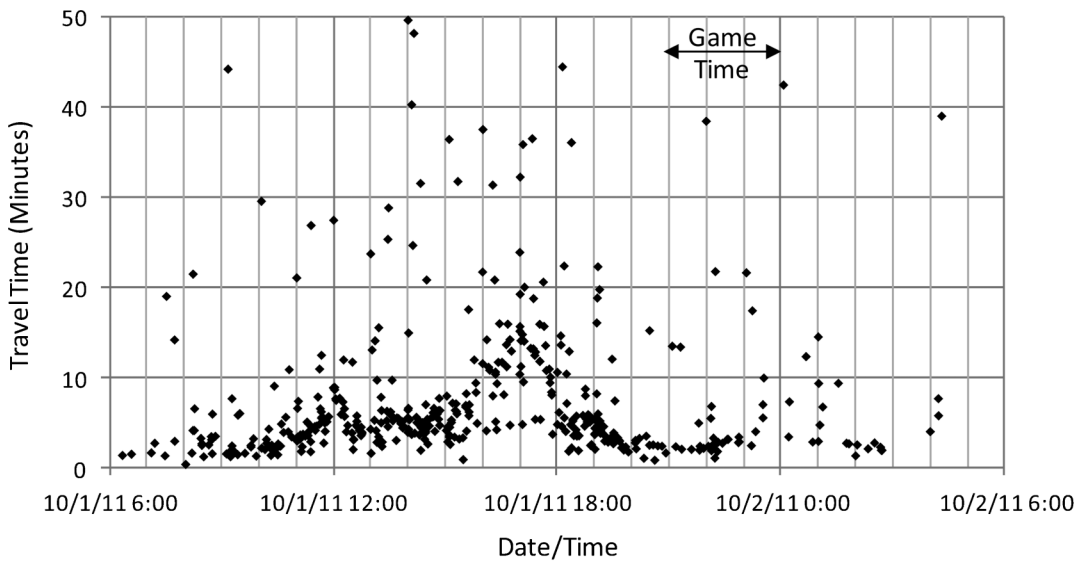


(b) Game day travel time.

Figure 6.11 Travel time scatter plots for inbound traffic on Route 3.

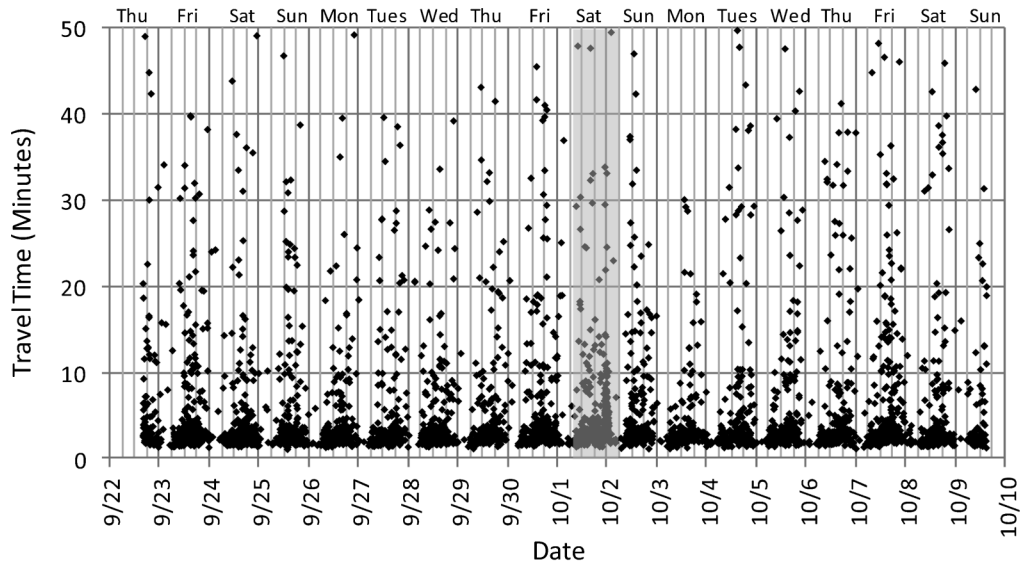


(a) Two week travel time.

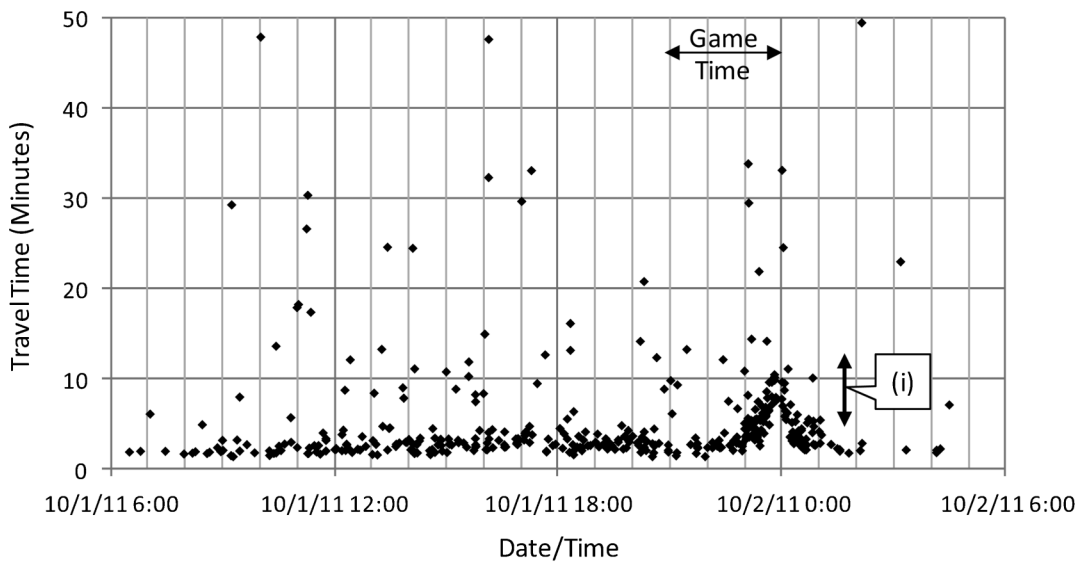


(b) Game day travel time.

Figure 6.12 Travel time scatter plots for outbound traffic on Route 3.

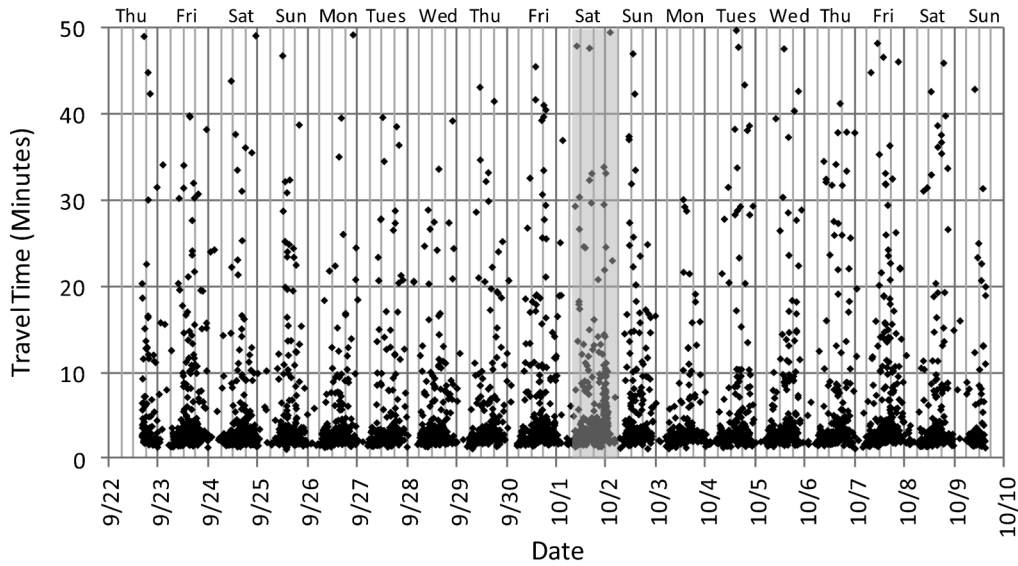


(a) Two week travel time.

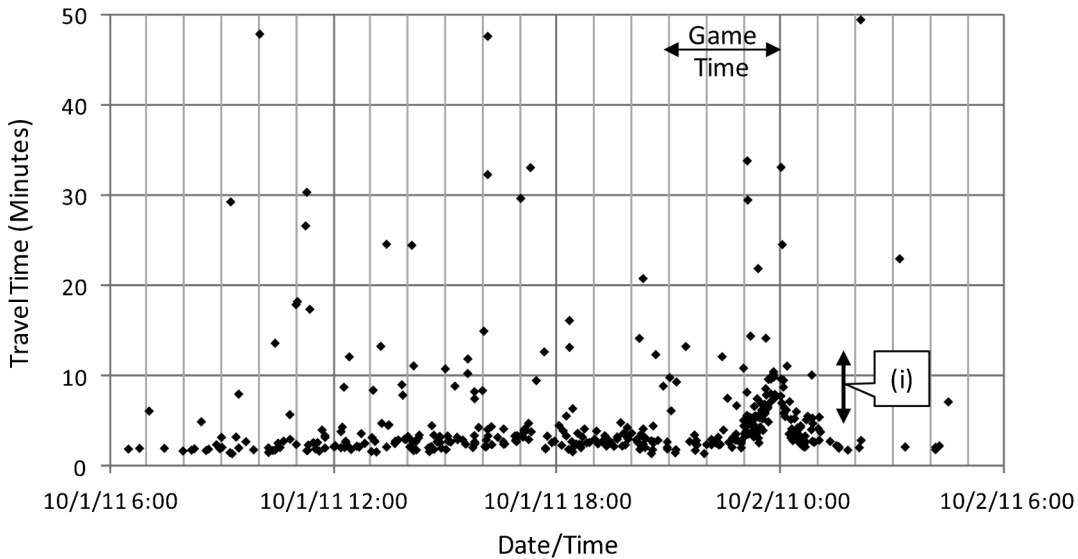


(b) Game day travel time.

Figure 6.13 Travel time scatter plots for inbound traffic on Route 4.



(a) Two week travel time.



(b) Game day travel time.

Figure 6.14 Travel time scatter plots for outbound traffic on Route 4.

be cost-effectively deployed for virtually any special event drawing a substantial crowd. Although only twelve stations were deployed in this study, the techniques presented can be easily scaled to denser sensor deployments as other activities such as hurricane evacuation, NASCAR races, or other similar events where it is important to assess traffic management activities to facilitate future improvement.

7. CONCLUSIONS

This document presented a collection of methodologies to assist a traffic analyst in the design of AVI

travel time data collection systems, and interpretation of travel time data.

Chapter 2 described procedures needed to convert the raw travel time data into information. The central concept of this chapter and the subsequent text focused on the construction of *cumulative frequency diagrams as an informative trace of the data characteristics (as opposed to aggregated average values that mask higher resolution characteristics)*. There was also a discussion of the filtering of outliers, conversions of travel time to speed or delay; a discussion of the number of samples needed for statistical tests, as well as statistical procedures that can be used to compare different case studies.

Chapter 3 discussed sensor placement, with a discussion of the impact of sensor location on what information can be retrieved from the data in the context of arterial systems, as well as the impact of antenna height on the sample rate, as field-tested for Bluetooth monitoring systems.

Chapter 4 presented case studies related to uninterrupted flow (*i.e.*, freeways) including a discussion of the impact of winter weather on travel characteristics, and a prototype for fusing this data with other data sets such as weather information to create a rich operational dashboard.

Chapter 5 focused on the interrupted flow (*i.e.*, signalized corridors). An extended discussion of signal operations concepts was included to assist the discussion. Three case studies were then presented to compare various signal operating conditions with the resulting travel time characteristics. The impact of arrival characteristics was shown to relate to bimodal travel time distributions. The impact of entry-exit detector pairing on the data characteristics was discussed. Finally, it was demonstrated how the travel time data can be used to understand the economic benefit of signal retiming, by converting the travel time difference into monetary measures of impact.

Finally, Chapter 6 presented a case study on the use of matched travel time data to develop an origin-destination matrix at a city-wide level during a special event in Lafayette, Indiana. This discussion included a proposed methodology for filling in missing observations. A detailed discussion on the impact of the special event was also presented.

REFERENCES

1. Kinzel, C. D. Traffic Studies. In J. Pline (ed.), *Traffic Engineering Handbook* (4th edition), Institute of Transportation Engineers, Washington, D.C., 1992, 59–93.
2. Quiroga, C. A., and D. Bullock. Measuring Control Delay at Signalized Intersections. *Journal of Transportation Engineering*, Vol. 125, pp. 271–280, 1999.
3. Turner, S., W. L. Eisele, R. J. Benz, and D. J. Holdener. *Travel Time Data Collection Handbook*. Publication No. FHWA-PL-98-035. Federal Highway Administration, Washington, D.C., 1998.
4. Quiroga, C. A., and D. Bullock. Determination of Sample Sizes for Travel Time Studies. *ITE Journal*, August 1998, 92–98.
5. Ernst, J. M., C. M. Day, J. V. Krogmeier, D. M. and Bullock. Probe Data Sampling Guidelines for Characterizing Arterial Travel Time. In *Transportation Research Record, Journal of the Transportation Research Board*, Paper No. 12-0441, Transportation Research Board of the National Academies, Washington, D.C., accepted for publication, in press, 2012.
6. Hauer, E. The Harm Done by Tests of Significance. *Accident Analysis and Prevention*, Vol. 36, 2004, pp. 495–500.
7. Siegel, S. *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill, New York, 1956.
8. Smirnov, N. Table for Testing the Goodness of Fit of Empirical Distributions. *The Annals of Mathematical Statistics*, Vol. 19, 1948, pp. 279–281.
9. Day, C. M., R. Haseman, H. Premachandra, T. M. Brennan, J. S. Wasson, J. R. Sturdevant, and D. M. Bullock. Evaluation of Arterial Signal Coordination: Methodologies for Visualizing High-Resolution Event Data and Measuring Travel Time. In *Transportation Research Record, Journal of the Transportation Research Board*, No. 2192, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 37–49.
10. Wasson, J. S., and D. M. Bullock. Database Architecture and Query Structures for Probe Data Processing. Appendix A. In Day, C. M., et al., *Application of Travel Time Information for Traffic Management*. Publication FHWA/IN/JTRP-2012/06. Joint Transportation Research Program, Indiana Department of Transportation and Purdue University, West Lafayette, Indiana, 2012. doi: 10.5703/1288284314666.
11. Hainen, A. M., J. S. Wasson, S. M. L. Hubbard, S. M. Remias, G. D. Farnsworth, and D. M. Bullock. Estimating Route Choice and Travel Time Reliability with Field Observations of Bluetooth Probe Vehicles. In *Transportation Research Record, Journal of the Transportation Research Board*, No. 2258, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 43–50.
12. Brennan, T. M., J. M. Ernst, C. M. Day, D. M. Bullock, J. V. Krogmeier, and M. Martchouk. Influence of Vertical Sensor Placement on Data Collection Efficiency from Bluetooth MAC Address Collection Devices. *Journal of Transportation Engineering*, Vol. 136, 12, 2010, pp. 1104–1109.
13. Brennan, T. M., B. D. Griggs, G. Grimmer, A. M. Hainen, C. M. Day, J. R. Sturdevant, and D. M. Bullock. An Empirical Approach for Defining the Design Space for Traffic Signal Timing Parameters. In *Transportation Research Record, Journal of the Transportation Research Board*, Paper No. 12-0685, Transportation Research Board of the National Academies, Washington, D.C., accepted for publication, in press, 2012.
14. Smaglik, E. J., A. Sharma, D. M. Bullock, J. R. Sturdevant, and G. Duncan. Event-Based Data Collection for Generating Actuated Controller Performance Measures. In *Transportation Research Record, Journal of the Transportation Research Board*, No. 2035, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 97–106.
15. Day, C. M., E. J. Smaglik, D. M. Bullock, and J. R. Sturdevant. Quantitative Evaluation of Fully Actuated Versus Non-Actuated Coordinated Phases. In *Transportation Research Record, Journal of the Transportation Research Board*, No. 2080, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 8–21.
16. Day, C. M. *Performance Based Management of Arterial Traffic Signal Systems*. PhD Thesis, Purdue University, West Lafayette, Indiana, May 2010.
17. Brennan, T. M., C. M. Day, J. R. Sturdevant, and D. M. Bullock. Visual Education Tools to Illustrate Coordinated System Operation. In *Transportation Research Record, Journal of the Transportation Research Board*, No. 2259, Transportation Research Board of the National Academies, Washington D.C., 2011, pp. 59–72.
18. Hubbard, S. M. L., D. M. Bullock, and C. M. Day. Integration of Real-Time Pedestrian Performance Measures

- into Existing Traffic Signal System Infrastructure. In *Transportation Research Record, Journal of the Transportation Research Board*, No. 2080, Transportation Research Board of the National Academies, Washington D.C., 2008, pp. 37–47.
19. Day, C. M., J. R. Sturdevant, and D. M. Bullock. Outcome Oriented Performance Measures for Management of Signalized Arterial Capacity. In *Transportation Research Record, Journal of the Transportation Research Board*, No. 2192, Transportation Research Board of the National Academies, Washington D.C., 2010, pp. 24–36.
 20. Brennan, T. M., C. M. Day, J. R. Sturdevant, E. Raamot, and D. M. Bullock. Track Clearance Performance Measures for Railroad Preempted Intersections. In *Transportation Research Record, Journal of the Transportation Research Board*, No. 2192, Transportation Research Board of the National Academies, Washington D.C., 2010, pp. 64–76.
 21. Day, C. M., J. M. Ernst, T. M. Brennan, C.-S. Chou, A. M. Hainen, S. M. Remias, A. P. Nichols, B. D. Griggs, and D. M. Bullock. Adaptive Signal Control Performance Measures: A System-in-the-Loop Simulation Case Study. In *Transportation Research Record, Journal of the Transportation Research Board*, Paper No. 12-0060, Transportation Research Board of the National Academies, Washington, D.C., accepted for publication, in press, 2012.
 22. Day, C. M., T. M. Brennan, J. M. Ernst, J. Sturdevant, and D. M. Bullock. *Procurement Procedures and Specifications for Performance Measure Capable Traffic Infrastructure Data Collection Systems*. Publication FHWA/IN/JTRP-2011/18. Joint Transportation Research Program, Indiana Department of Transportation and Purdue University, West Lafayette, Indiana, 2011. doi: [10.5703/1288284314642](https://doi.org/10.5703/1288284314642).
 23. Day, C. M., and D. M. Bullock. Computational Efficiency of Alternative Algorithms for Arterial Offset Optimization. In *Transportation Research Record, Journal of the Transportation Research Board*, No. 2259, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 37–47.
 24. Brysbaert, M. *Basic Statistics for Psychologists*. Macmillan Publishers Ltd., Basingstoke, Hants, UK, 2011.
 25. Day, C. M., T. M. Brennan, A. M. Hainen, S. M. Remias, H. Premachandra, J. R. Sturdevant, G. Richards, J. S. Wasson, and D. M. Bullock. Reliability, Flexibility, and Environmental Impact of Alternative Objective Functions for Arterial Offset Optimization. In *Transportation Research Record, Journal of the Transportation Research Board*, No. 2259, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 8–22.
 26. Schrank, D., and T. Lomax. *2009 Urban Mobility Report*. Texas Transportation Institute, College Station, Texas, July 2009.
 27. EPA. *Emission Facts: Greenhouse Gas Emissions from a Typical Passenger Vehicle*. Publication EPA420-F-05-004, U.S. Environmental Protection Agency, Washington, D.C., February 2005.
 28. EPA. *EPA Analysis of the American Clean Energy and Security Act of 2009: H.R. 2454 in the 111th Congress*. U.S. Environmental Protection Agency, Washington, D.C., June 23, 2009. Available online at http://www.epa.gov/climatechange/Downloads/EPAactivities/HR2454_Analysis.pdf. Retrieved July 27, 2010.
 29. Federal Highway Administration. *Managing Travel for Planned Special Events Handbook: Executive Summary*. Report No. FHWA-HOP-07-108, U.S. Department of Transportation, Federal Highway Administration, June 2007. Available online at http://www.ops.fhwa.dot.gov/publications/fhwahop07108/plnd_spcl_evnts.pdf. Retrieved January 30, 2012.

Report Sponsor

The “Small Business Innovation Development Act of 1982” (Pub. L. No. 97-219), along with reauthorizing legislation (Pub. L. No. 99-443 and Pub. L. No. 102-564, the “Small Business Research and Development Enhancement Act of 1992”), seeks to encourage the initiative of the private sector and to use small business effectively to meet federal research and development objectives. To comply with statutory obligations of the Act, the U.S. Department of Transportation established the Small Business Innovation Research (SBIR) Program, which conforms to the guidelines and regulations provided by the Small Business Administration. Annually, small businesses are solicited to submit innovative research proposals that address the high-priority requirements of the U.S. Department of Transportation and that have potential for commercialization.

In 2010 Traffax Inc. was awarded U.S. Department of Transportation Contract Number DTRT57-11-C-10009, with Nicholas Ganig serving as principal investigator. This publication was prepared as part of Contract Task M5b—Handbook for Application of Bluetooth Traffic Monitoring on Arterials.

Publication

This report was published in collaboration with Traffax Inc., the Joint Transportation Research Program, and Purdue University. The full content of this technical report is available for download at dx.doi.org/10.5703/1288284314988. In addition, a print-on-demand bound version of this report can be purchased at that location.

Open Access and Collaboration with Purdue University

The Indiana legislature established the Joint Highway Research Project in 1937. In 1997, this collaborative venture between the Indiana Department of Transportation and Purdue University was renamed as the Joint Transportation Research Program (JTRP) to reflect state and national efforts to integrate the management and operation of various transportation modes. Since 1937, the JTRP program has published over 1500 technical reports. In 2010, the JTRP partnered with the Purdue University Libraries to incorporate these technical reports in the University’s open access digital repository and to develop production processes for rapidly disseminating new research reports via this repository. Affiliated publications have also recently been added to the collection. As of 2012, the JTRP collection had over 450,000 downloads, with some particularly popular reports having over 20,000 downloads.

